Dimensions, Conditions, and Improvement of Teacher–Toddler Interaction Quality in Early Childhood Education and Care: Teachers' Role in the Complex Interplay of Process–Person–Context–Time

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität
München

vorgelegt von

Franka Luise Baron

aus
Wiesbaden

2024

Erstgutachterin: Prof. Dr. Susanne Kuger, Ludwig-Maximilians-Universität

Zweitgutachterin: Prof. Dr. Simone Lehrl, Pädagogische Hochschule Weingarten

Tag der Disputation: 8. Juli 2024

Zusammenfassung

Institutionelle Betreuungseinrichtungen spielen bei der Entwicklung von Kindern eine zentrale Rolle, da in dieser Lernumgebung wichtige Kompetenzen und Fähigkeiten erworben werden. Fachkraft–Kind–Interaktionen in der frühkindlichen Bildung, Betreuung und Erziehung (FBBE) sind dabei entscheidend: Ihre Qualität bestimmt maßgeblich, inwieweit Interaktionen zu einer positiven Entwicklung beitragen. Diese Interaktionen werden in globale und domänenspezifische Merkmale unterteilt, da sich erwiesen hat, dass beide in besonderer Weise mit dem kindlichen Lernen und der spezifischen Kompetenzentwicklung zusammenhängen. Die mathematikspezifische Entwicklung gewinnt dabei in der frühkindlichen Bildung und Forschung an Bedeutung. Trotz der zunehmenden Nutzung institutioneller Betreuung von Kindern unter drei Jahren ist jedoch wenig über die Qualität und das Potenzial zur Qualitätssteigerung dieser globalen und mathematikspezifischen Interaktionen in Krippen bekannt. In Gruppen mit älteren Kindern haben sich Fortbildungen für pädagogische Fachkräfte als vielversprechend für die Steigerung von Interaktionsqualität erwiesen. Erkenntnisse darüber, ob und inwieweit solche Fortbildungen auch in Krippen effektiv sind, stehen jedoch aus. Auch welche Rolle verschiedene personenbezogene, strukturelle und kontextuelle Rahmenbedingungen für die Qualität und ihre Steigerung in Krippengruppen spielen, bleibt bisher, vor allem in Bezug auf die FBBE in Deutschland, offen. Die vorliegende Dissertation hat daher **das übergeordnete Ziel, das komplexe Zusammenspiel verschiedener Systeme (Prozess, Person, Kontext und Zeit) zu veranschaulichen und die Rolle der pädagogischen Fachkraft innerhalb dieser Systeme zu untersuchen**. Dafür werden die Fragen beantwortet (1) welche Instrumente zur Beobachtung von globalen und mathematikspezifischen Fachkraft–Kind–Interaktionen verwendet werden können, (2) welche Zusammenhänge zwischen personenbezogenen und strukturellen Merkmalen und der Qualität von Fachkraft–Kind–Interaktionen bestehen und (3) inwieweit sich globale und mathematikspezifische Fachkraft–Kind–Interaktionen während verschiedener Aktivitäten durch Fortbildungen für pädagogische Fachkräfte verbessern lassen.

In der frühpädagogischen Forschung wird eine Vielzahl von Instrumenten verwendet, um globale und mathematikspezifische Interaktionsqualität festzustellen. Dabei bleibt jedoch unklar, welche Instrumente vorrangig verwendet werden und wie sich diese unterscheiden. Instrumente zur Beobachtung von mathematikspezifischen

Interaktionen mit Kindern unter drei Jahren sind außerdem noch wenig betrachtet worden. In der **ersten Studie** wurde daher eine systematische Literaturarbeit durchgeführt, die zum Ziel hatte, Instrumente zur Erfassung globaler und mathematikspezifischer Interaktionsqualität zu identifizieren, zu analysieren und zu diskutieren. Als Methode wurde SALSA angewendet, was verschiedene Schritte umfasst: den Such- (Search) und Auswahlprozess von Artikeln (AppraisaL), die Zusammenfassung der identifizierten Instrumente (Synthesis) und die Diskussion der Ergebnisse (Analysis). In mehreren Datenbanken wurden englisch- und deutschsprachige Artikel aus Fachzeitschriften, Dissertationen, Handbüchern und Konferenzbeiträgen gesucht. Eingeschlossen wurden Artikel, die ein Beobachtungsinstrument beschreiben, das von einer objektiven Person durchgeführt wird, das Interaktionen in Krippen- und/oder Kindergartengruppen erfasst, das einen Aspekt von Interaktionsqualität berücksichtigt und das globale und/oder mathematikspezifische Interaktionen einschließt. Aus insgesamt 4211 Publikationen verblieben nach dem Auswahlprozess 148 Artikel, in denen insgesamt 55 Beobachtungsinstrumente verwendet wurden. Diese wurden hinsichtlich ihrer Skalen, Items, Bewertungsmethode, Beobachtungsverfahren, des Alters der Kinder, der Reliabilität und der Validität analysiert. Bezüglich der Skalen bestand ein wesentlicher Unterschied darin, dass sie auf unterschiedlichen theoretischen Annahmen basieren. Zwei Annahmen konnten hierbei voneinander abgegrenzt werden: Interaktionen werden als Teil der Gesamtqualität von FBBE und somit als ein Aspekt von vielen betrachtet, oder der Fokus liegt auf Interaktionen und es werden verschiedene Facetten unterteilt. Die Items der Instrumente stellen die kleinste zu beobachtende konzeptionelle Einheit dar. Einige Instrumente legten Indikatoren innerhalb dieser Items fest, um qualitativ hoch- oder niedrigwertige Interaktionen ausführlicher zu definieren. Je nach Instrument erfüllten sie unterschiedliche Zwecke. In einigen Fällen wurden Indikatoren nicht direkt bewertet, sondern dienten als Beschreibung zur genaueren Definition des Items. In anderen Fällen wurden sie direkt bewertet und bildeten die Grundlage für die Einschätzung des Ausgangsitems. Der Hauptunterschied zwischen den Items bestand darin, inwieweit die beobachtende Person Fachkraft–Kind–Interaktionen auf einem abstrakten Level einschätzt (Inferenzlevel). Die meisten Instrumente verwendeten eine mehrstufige Bewertungsskala, um die Qualität von Interaktionen zu beurteilen. Andere, seltenere Methoden gaben Prozente des beobachteten Verhaltens an oder verwendeten eine dichotome Skala. Dies war vor allem bei Instrumenten zur Beobachtung von mathematikspezifischen Interaktionen der Fall. Die Instrumente unterschieden sich auch darin, wie lange die Beobachtung dauerte und ob sie live oder anhand einer Videoaufnahme erfolgte. Die meisten Bewertungen basierten auf mehrstündigen Beobachtungen, die im Laufe eines Tages durchgeführt wurden. Nur wenige Instrumente teilten die Beobachtung auf mehrere Tage auf. In Bezug auf das Alter der Kinder boten manche Instrumente unterschiedliche Versionen für jüngere und ältere Kinder, die aufeinander aufbauen. Die Angaben zur Reliabilität und Validität waren sehr unterschiedlich, weshalb hierzu nur grobe Einschätzungen getroffen werden können. Für das Verwenden mancher Instrumente wird durch eine offizielle Schulung eine bestimmte Inter-Rater Reliabilität vorausgesetzt, bei den meisten gibt es dafür jedoch kein Standardverfahren. Viele Studien verwendeten Cronbachs Alpha als Maß der internen Konsistenz, wobei die Ergebnisse hierzu unterschiedlich ausfielen. Die Validität der Instrumente wurde meistens durch explorative oder konfirmatorische Faktorenanalysen bewertet, wobei auch diese zu unterschiedlichen Erkenntnissen führten. Insgesamt erfassten mehr Instrumente globale (n = 39) als mathematikspezifische (n = 3) Aspekte von Fachkraft–Kind–Interaktionen, was insbesondere auf Krippengruppen zutraf. Mit manchen Instrumenten ließen sich beide Facetten von Interaktionsqualität beobachten (n = 13). Dieser Überblick über die in der frühkindlichen Forschung verwendeten Beobachtungsinstrumente bietet die Möglichkeit, Besonderheiten einzelner Instrumente hervorzuheben und zu identifizieren, welche am besten für einen bestimmten Forschungszweck geeignet sind. Somit trägt die durchgeführte systematische Literaturarbeit zur Ermöglichung einer effektiven Forschung bei. Es konnte gezeigt werden, dass trotz einer Vielzahl an Instrumenten oftmals dieselben verwendet wurden. Außerdem deckte die Bandbreite der vorhandenen Instrumente nicht alle Inhaltsbereiche und Altersgruppe im gleichen Maße ab, denn vor allem Instrumente, welche die Qualität mathematikspezifische Interaktionen zwischen Fachkräften und Kindern unter drei Jahren erheben, waren unterrepräsentiert.

Forschungsergebnisse zeigen, dass verschiedene Domänen von Fachkraft–Kind–Interaktionen differenziert werden können und sich diese bezüglich ihrer Qualität unterscheiden. Je nach Alter der Kinder werden diese Domänen unterschiedlich betont und gruppiert. In Bezug auf Kinder von ein bis drei Jahren wird die emotionale und verhaltensbezogene Unterstützung und die aktive Lernunterstützung als wichtige Grundlagen für das Lernen und die Entwicklung definiert. Dazu, ob diese Unterteilung von Domänen der Interaktionen auf empirischer Grundlage tatsächlich getroffen werden kann, gibt es international jedoch unterschiedliche Erkenntnisse. In Bezug auf Krippen in

Deutschland bleibt außerdem unklar, wie sich deren Qualität darstellt und wie verschiedene Domänen mit personellen und strukturellen Merkmalen in Verbindung stehen. Die **zweite Studie** untersucht deshalb die Domänen und die Qualität von Fachkraft–Kind–Interaktionen und deren Zusammenhänge mit Merkmalen von pädagogischen Fachkräften und Krippengruppen. Für die Beobachtung der Interaktionen wurde das Classroom Assessment Scoring System Toddler Version (CLASS Toddler; La Paro et al., 2012) verwendet. Die Krippengruppen wurden an einem regulären Morgen besucht und die Fachkraft–Kind–Interaktionen wurden während verschiedener Aktivitäten beobachtet. Pro Gruppe wurden vier Zyklen von 15–20 Minuten beobachtet und mit CLASS Toddler eingeschätzt, wobei drei davon live und einer im Nachgang per Videoaufzeichnung bewertet wurden. Gemäß CLASS Toddler wurde die Qualität der Interaktionen auf einer 7-stufigen Likert-Skala von niedrig (1, 2) über mittel (3, 4, 5) bis hoch (6, 7) eingestuft. Dabei wurden acht Dimensionen eingeschätzt: positives Klima, negatives Klima, Sensibilität der Fachkraft, Berücksichtigung der kindlichen Perspektiven, Verhaltenslenkung, Förderung von Lernen und Entwicklung, Qualität des Feedbacks und Sprachmodellierung. Anhand von Notizen während der Beobachtungen und Fragebögen für die teilnehmenden Fachkräfte wurden personelle und strukturelle Merkmale erfasst. Bei der Analyse wurden Angaben zum Alter und zur Ausbildung der pädagogischen Fachkraft, zur Gruppengröße und zur Alterszusammensetzung der Kinder in der Gruppe berücksichtigt. Mit den generierten Daten von insgesamt 95 pädagogischen Fachkräften aus je einer bayerischen Krippengruppe wurden konfirmatorische Faktorenanalysen (CFA) und Strukturgleichungsmodelle (SEM) in R berechnet. Die Ergebnisse der CFA unterstützten eine zweifaktorielle Struktur der Interaktionsqualität mit einer Domäne der emotionalen und verhaltensbezogenen Unterstützung (EBS; umfasst positives Klima, negatives Klima, Sensibilität der Fachkraft, Berücksichtigung der kindlichen Perspektiven, Verhaltenslenkung) und einer Domäne der aktiven Lernunterstützung (ESL; umfasst Förderung von Lernen und Entwicklung, Qualität des Feedbacks, Sprachmodellierung). Die Modellpassung kann als annehmbar, wenn auch nicht perfekt bewertet werden: $\chi^2(19) = 56.51$, $p < .001$, CFI = .91, TLI = .87, RMSEA = .14 (CI90: .10–.19), SRMR = .07. Das zweifaktorielle Modell wurde mit einem ein- und dreifaktoriellen Modell verglichen, die jeweils eine schlechtere Modelpassung erzielten. Die Domäne EBS hatte eine höhere Qualität ($M = 5.33$, $SD = 0.59$) als die Domäne ESL ($M = 3.23$, $SD = 0.70$). Das Resultat des SEM zeigte, dass das Alter der pädagogischen Fachkraft, ihre Ausbildung und die Alterszusammensetzung der Kinder in der Gruppe mit der Qualität von Interaktionen zusammenhängt. Die Modellpassung fiel hierbei etwas besser aus: $\chi^2(43) = 81.65$, $p < .001$, CFI = .91, TLI = .88, RMSEA = .10 (CI90: .06–.13), SRMR = .06. Weniger EBS wurde beobachtet, wenn die Fachkraft älter war ($B = –.02$, $SE = .01$, $\beta = –.26$, $p = .01$) und wenn die Gruppen altersgemischt waren ($B = –.35$, $SE = .15$, $\beta = –.25$, $p = .02$). Mehr ESL wurde von Fachkräften mit höherem Abschluss bereitgestellt ($B = .12$, $SE = .07$, $\beta = .16$, $p = .10$). Darüber hinaus waren altersgemischte Gruppen negativ mit ESL assoziiert ($B = –.33$, $SE = .16$, $\beta = –.26$, $p = .04$). Die Gruppengröße hing mit keiner der Domänen von Fachkraft–Kind–Interaktionsqualität signifikant zusammen. Insgesamt trägt diese Studie zum Verständnis über unterschiedliche Domänen von Interaktionen sowie deren Zusammenspiel mit personellen und strukturellen Merkmalen bei. Somit können Aspekte identifiziert werden, die zur Verbesserung der Qualität von Fachkraft–Kind–Interaktionen in Krippengruppen beitragen können. Sowohl Fachkraftmerkmale (Alter und Ausbildung), als auch Gruppenmerkmale (Alterszusammensetzung der Kinder) spielen dabei eine Rolle. Dies zeigt, dass solche Aspekte sorgfältig unterschieden werden sollten und Zusammenhänge mit verschiedenen Domänen der Interaktionsqualität differenziert betrachtet werden müssen.

Die Unterscheidung verschiedener Facetten von Interaktionen, insbesondere auch die Unterscheidung von globalen und mathematikspezifischen Interaktionen, ist vor allem in Bezug auf die Steigerung von deren Qualität sinnvoll. Studien haben gezeigt, dass pädagogische Fachkräfte diese Bereiche nicht unbedingt gleichermaßen unterstützen und deren Steigerung unterschiedliche Herausforderungen mit sich bringen kann. Hochwertige Interaktionen in allen Domänen und Dimensionen zu fördern ist jedoch zentral, da globale und mathematikspezifische Interaktionen mit verschiedenen Kompetenzentwicklungen von Kindern zusammenhängen. Die **dritte Studie** hat daher das Ziel, Fortbildungen für pädagogische Fachkräfte zu konzipieren und durchzuführen, um die Frage zu beantworten, ob diese zu einer gesteigerten Qualität von globalen und mathematikspezifischen Fachkraft–Kind–Interaktionen führen. Hierbei erscheint es wichtig, verschiedene Aktivitäten in den Blick zu nehmen, da kontextuelle Faktoren, in denen Interaktionen stattfinden, einen Einfluss auf deren gelingen haben können. Auch Fortbildungen könnten deshalb, abhängig von der durchgeführten Aktivität, unterschiedliche Wirkung entfalten. Erkenntnisse aus Krippengruppen gibt es dazu aber kaum. Diese Studie bezieht somit das Freispiel und strukturierte Situationen wie den Morgenkreis ein, indem Interaktionsqualität und deren potenzielle Steigerung differenziert nach diesen Aktivitäten betrachtet wird. Zur Untersuchung von Interventionseffekten wurden 95 pädagogischen Fachkräfte quasi-randomisiert zu drei Gruppen zugewiesen: zur mathematikspezifischen Interventionsgruppe (MIG), zur globalen Interventionsgruppe (GIG) oder zur Kontrollgruppe (CG). Für die Interventionsgruppen waren die Fortbildungen in Bezug auf Methoden und Umfang identisch, aber in Bezug auf Fokus und Inhalt verschieden. Die zweitägigen, virtuellen Fortbildungstage umfassten insgesamt acht Module, die Interaktionen in alltäglichen Situationen betonten, wobei die MIG neben generellen auch Strategien zur mathematikspezifischen Anregung behandelte, während die GIG generelle Strategien zur Förderung globaler Anregung fokussierte. Als zentrale Methoden für beide Interventionsgruppen wurden Videoanalysen mit Beobachtungsbögen sowie Rollenspiele angewendet. Den Fachkräften wurden standardisiert ausgestattete Kisten mit Büchern und Spielen bereitgestellt, um die Fortbildungsinhalte zu veranschaulichen. Diese wurden anschließend in der Gruppe belassen. Zur weiteren Implementation wurden die pädagogischen Fachkräfte im Zeitraum von acht Wochen darum gebeten, die Materialien und erlernten Strategien mindestens dreimal pro Woche zu verwenden. Darüber hinaus wurden wöchentliche Erinnerungen zur Selbstbeobachtung und Newsletter mit anregenden Ideen versendet. Die CG hatte lediglich die Möglichkeit, an einer kurzen Informationsveranstaltung teilzunehmen. Diese Veranstaltung beinhaltete keinerlei interaktiven Elemente. Die Qualität von Fachkraft–Kinder–Interaktionen wurde vor und nach der Fortbildung bei allen Gruppen mit CLASS Toddler erhoben. Während der Einschätzung globaler Interaktionen wurde auch die Häufigkeit mathematikspezifischer Interaktionen erfasst – dies ist ein projektspezifischer Zusatz. Wenn Interaktionen im Bereich Mathematik beobachtet wurden, wurden diese zu sechs Kategorien zugeordnet: Mengen/Zahlen/Ziffern, Größe/Messen, Raum/Form, Zeit, Muster/Strukturen und Sortieren/Klassifizieren. Für die folgenden Analysen wurde ein Summenscore aus der Häufigkeit aller beobachteten mathematikspezifischen Interaktionen pro Zyklus gebildet. Die Aktivitäten wurden ebenfalls während der CLASS Toddler Beobachtung notiert. Unterschieden wurde in Freispiel/Interessenbereiche, Übergänge, Routinen und Gruppenzeit. Hierbei wurde für jeden Zyklus angegeben, wie viel Prozent der Zeit die jeweilige Aktivität einnahm – auch dies ist ein projektspezifischer Zusatz. Zur Beantwortung der Fragestellung wurden Varianzanalysen mit Messwiederholungen (RMANOVA) in R mit Daten von 89 Fachkräften durchgeführt. Die Ergebnisse zeigten, dass sich durch die Fortbildung für die MIG deren mathematikspezifischen Interaktionen während des Freispiels marginal signifikant verbesserte ($F(2, 75) = 2.82$, $p = .07$, $\eta_p^2 = .03$), jedoch nicht während strukturierter Aktivitäten. Die Qualität der globalen Interaktion von MIG und GIG verbesserte sich während beiden Aktivitäten im Vergleich zur Kontrollgruppe nicht signifikant. Allerdings konnte sowohl im Freispiel, als auch während strukturierten Beobachtungen ein Qualitätsanstieg bei allen Gruppen über die Zeit hinweg beobachtet werden (EBS: $F(1, 81) = 4.17$, $p = .04$, $\eta_p^2 = .02$; ESL: $F(1, 81) = 3.24$, $p = .08$, $\eta_p^2 = .01$; Mathematik: $F(1, 75) = 13.12$, $p = .001$, $\eta_p^2 = .07$). Diese Studie verdeutlicht das Potenzial von Fortbildungen, zeigt aber auch, dass die Steigerung von Interaktionsqualität während unterschiedlicher Aktivitäten besser gelingen oder schwieriger sein kann. Vor allem bezüglich Krippengruppen in Deutschland ist diese differenzierte Betrachtung lohnend, da Freispiel vergleichsweise viel Zeit in Anspruch nimmt und ein positiver Effekt von Fortbildungen für mathematikspezifische Interaktionen in diesen Situationen identifiziert werden konnte.

Neben den Erkenntnissen aus den einzelnen Studien wird in der **Gesamtbetrachtung** deutlich, dass sich die analysierten Aspekte in das Process–Person–Context–Time (PPCT) Modell von Bronfenbrenner und Morris (1998, 2006) einbetten lassen. Dies ist eine umfassende Theorie, die vor allem die gegenseitigen Beeinflussungen der Systeme und deren Verbindungen in Hinblick auf Interaktionen als Teil von Prozessen betrachtet. Deutlich wird dabei aber, dass die Rolle der Fachkraft – auch wenn sie einen aktiven und zentralen Part bei der Gestaltung von Interaktionen einnimmt – teilweise unbeleuchtet bleibt. Dem könnte entgegengewirkt werden, indem die pädagogische Fachkraft stärker in den Mittelpunkt gerückt wird. Ein Ansatz dafür könnte sein, ihre Überzeugungen, Selbstwirksamkeit und Kompetenzen sowie deren Veränderungen genauer zu untersuchen. Daneben sind Aspekte wie die Zufriedenheit im Beruf, das Klima innerhalb des Teams oder die Möglichkeit zur beruflichen Weiterqualifizierung maßgeblich für gelingende Interaktionen mit Kindern. Dies sollte sowohl in der Forschung als auch in der Praxis und bei politischen Maßnahmen stärker berücksichtigt werden. Eine Studie zu konzipieren, die alle wesentlichen Systeme des PPCT Modells beinhaltet ist herausfordernd aber lohnenswert, weil dadurch nicht nur einzelne Komponenten, sondern auch deren Zusammenspiel dargestellt werden kann. Voraussetzung dafür ist, dass das PPCT Modell reflektiert eingesetzt wird. Dies umfasst vor allem diejenigen Systeme zu diskutieren, die nicht einbezogen wurden und sich damit auseinanderzusetzen, welche Auswirkungen das für Resultate und Schlussfolgerungen haben könnte.

Summary

Institutional childcare takes a central part in children's development, as important competencies and skills are acquired in this learning environment. The quality of teacher–child interactions in early childhood education and care (ECEC) is crucial in determining the extent to which interactions contribute to positive development. These interactions are divided into global and domain-specific characteristics, as it has been proven that both are specifically associated with children's learning and specific competency development. In this regard, math-specific development is gaining importance in early childhood education and research. However, despite the increasing use of institutional childcare for children under the age of three, little is known about the quality and potential for enhancing the quality of global and math-specific interactions. In classrooms with older children, teacher trainings are promising tools to improve interaction quality, but it remains unclear whether and to what extent such trainings are also effective in toddler classrooms. Moreover, the role of various personal, structural, and contextual conditions for quality and its improvement remains open, especially concerning ECEC and toddler classrooms in Germany. Therefore, **the overarching goal of this thesis is to illustrate the complex interplay of different systems (process, person, context, and time) and to examine teachers' role within these systems**. This includes addressing the questions (1) which instruments can be used to assess global and math-specific teacher–child interactions, (2) what associations exist between personal and structural characteristics and the quality of teacher–child interactions, and (3) to what extent do global and math-specific teacher–child interactions improve across different activity settings through teacher trainings.

Early childhood research uses a variety of instruments to assess global and math-specific interaction quality. However, it remains unclear which instruments are primarily used and how they differ. Further, instruments for assessing math-specific interactions with children under three years have received little attention. Hence, a systematic literature review was conducted in the **first study** to identify, analyze, and discuss instruments for assessing global and math-specific interaction quality. The method SALSA was applied, which includes various steps: the search (Search) and selection process of articles (AppraisaL), the synthesis of identified instruments (Synthesis), and the discussion of results (Analysis). English and German articles from scientific journals, dissertations, handbooks, and conference papers were searched in multiple databases.

Articles describing an observation instrument conducted by an objective person, capturing interactions in toddler and/or preschool classrooms, considering an aspect of interaction quality, and including global and/or math-specific interactions were included. Out of a total of 4211 publications, 148 articles remained after the selection process, using a total of 55 observation instruments. These were analyzed regarding their scales, items, rating method, observation procedure, children's age, reliability, and validity. There was a significant difference in scales, as they were based on different theoretical assumptions. Two assumptions could be distinguished: interactions are considered as part of the overall quality of ECEC, or the focus is on interactions and various facets are divided. The items of the instruments represent the smallest conceptual unit to be observed. Some instruments established indicators within these items to define high- or low-quality interactions more precisely. Depending on the instrument, they served different purposes. In some cases, indicators were not directly evaluated, but served as a description to further define the item. In other cases, they were directly rated and formed the basis for assessing the item of interest. The main difference between the items of instruments was the extent to which the observing person assessed teacher–child interactions at an abstract level (level of inference). Most instruments used a multi-level rating scale to assess interaction quality. Other, more rare methods indicated percentages of observed behavior or used a dichotomous scale. This was particularly the case with instruments for assessing math-specific interactions. The instruments also differed in the duration of observation and whether it was performed in person or based on video recordings. Most assessments were based on observations lasting several hours that were conducted throughout one day. Only few instruments divided the observation into several days. Regarding the children's age, some instruments offered different versions for younger and older children, building up on each other. As the information on reliability and validity largely varied across the articles, only non-conclusive statements could be made. For the use of some instruments, a certain inter-rater reliability is assumed through official training, but there is no standard procedure for most of them. Many studies used Cronbach's Alpha as a measure of internal consistency, with varying results. The validity of the instruments was mostly assessed through exploratory or confirmatory factor analyses, which also led to different findings across studies. Overall, more instruments captured global (n = 39) than math-specific (n = 3) aspects of teacher–child interactions, especially in toddler classrooms. Some instruments allowed observation of both facets of interaction quality (n = 13). This overview of observation instruments used in early childhood research provides the opportunity to highlight particularities of instruments and identify which ones are best suited for a specific research interest. Thus, the systematic literature review conducted contributes to enabling effective research. It could be shown that despite a variety of instruments, often the same were used. Moreover, the range of existing instruments did not cover all content areas and all age groups to the same extent. Especially instruments assessing the quality of math-specific interactions between teachers and children under three years were underrepresented.

Research results show that different domains of teacher–child interactions can be differentiated and that they vary in their quality. Depending on the children's age, these dimensions are emphasized and grouped differently. Regarding children aged one to three, emotional and behavioral support and engaged support for learning are defined as important foundations for learning and development. However, internationally, there are different findings on whether this separation of domains can actually be made on an empirical basis. Furthermore, regarding toddler classrooms in Germany, it remains unclear how the quality of domains is represented and how these relate to personal and structural characteristics. Therefore, the **second study** examines domains and quality of teacher–child interactions and their associations with characteristics of teachers and toddler classrooms. For the observation of interactions, the Classroom Assessment Scoring System Toddler Version (CLASS Toddler; La Paro et al., 2012) was used. Toddler classrooms were visited on a regular morning, and teacher–child interactions were observed during various activities. Per classroom, four cycles of 15–20 minutes were observed and rated with CLASS Toddler, with three being rated live and one through video recording afterwards. According to CLASS Toddler, the quality of interactions was rated on a 7-point Likert scale from low (1, 2), middle (3, 4, 5) to high (6, 7). Eight dimensions were assessed: positive climate, negative climate, teacher sensitivity, regard for child perspectives, behavior guidance, facilitation of learning and development, quality of feedback, and language modeling. Personal and structural characteristics were recorded through notes during observations and through questionnaires for participating teachers. Age and education of the teacher, group size, and age composition of the children in the classroom were considered in the analysis. Confirmatory factor analyses (CFA) and structural equation modeling (SEM) were conducted in R with data from a total of 95 teachers, each from one Bavarian classroom. The results of the CFA supported a two-factor structure of interaction quality with a domain of emotional and behavioral support (EBS; including positive climate, negative climate, teacher sensitivity, regard for child perspectives, and behavior guidance) and a domain of engaged support for learning (ESL; including facilitation of learning and development, quality of feedback, and language modeling). The model fit can be considered acceptable, though not perfect: $\chi^2(19) = 56.51$, $p < .001$, CFI = .91, TLI = .87, RMSEA = .14 (CI90: .10–.19), SRMR = .07. The two-factor model was compared to a one- and three-factor model, each achieving a worse model fit. The EBS domain demonstrated higher quality (*M* = 5.33, *SD* = 0.59) than the ESL domain (*M* = 3.23, *SD* = 0.70). The SEM showed that teachers' age, their education, and the age composition of children in the classroom were associated with interaction quality. The model fit was slightly better here: $\chi^2(43) = 81.65$, $p < .001$, CFI = .91, TLI = .88, RMSEA = .10 (CI90: .06–.13), SRMR = .06. Less EBS was observed when the teacher was older (*B* = –.02, *SE* = .01, β = –.26, *p* = .01) and when the group was of mixed-age (*B* = –.35, *SE* = .15, β = –.25, *p* = .02). More ESL was provided by teachers with higher educational degrees (*B* = .12, *SE* = .07, β = .16, *p* = .10). Additionally, mixed-age groups were negatively associated with ESL (*B* = –.33, *SE* = .16, β = –.26, *p* = .04). Group size was not significantly associated with any domain of teacher–child interaction quality. Overall, this study contributes to understanding different domains of interactions and their interplay with personal and structural characteristics. Thus, aspects can be identified that can contribute to improving the quality of teacher–child interactions in toddler classrooms. Both teacher (age and education) and classroom characteristics (age composition of children) play a unique role. This shows that such aspects should be carefully distinguished, and associations with different domains of interaction quality must be differentiated.

The distinction between different facets of interactions, especially the distinction between global and math-specific interactions, is particularly useful for enhancing their quality. Studies have shown that teachers do not necessarily support these areas equally and that improving them can pose different challenges. However, promoting high-quality interactions in all domains and dimensions is crucial, as global and math-specific interactions are associated with different competence developments of children. Therefore, the **third study** aims to design and implement teacher trainings to address the question of whether they lead to increased quality of global and math-specific teacher–child interactions. It appears important to consider various activities, as contextual factors in which interactions take place can influence their effectiveness. Also, depending on the conducted activity, trainings could have different impacts. However, there is hardly any knowledge about this concerning toddler classrooms. Therefore, this study includes free play and structured activities, such as morning circles, examining interaction quality and its potential improvement, differentiated according to these situations. For the examination of intervention effects, 95 teachers were quasi-randomly assigned to three groups: the math-specific intervention group (MIG), the global intervention group (GIG), or the control group (CG). For the intervention groups, the trainings were identical in terms of methods and scope, but different in terms of focus and content. The two-day, virtual training sessions comprised a total of eight modules emphasizing interactions in everyday situations. The MIG covered strategies for global and math-specific stimulation, while the GIG focused on general strategies. Video analyses with observation sheets and role-plays were applied as central methods for both intervention groups. Teachers were provided with standardized boxes containing books and games to illustrate the training content. These were then left in the classrooms. For further implementation, teachers were asked to use the materials and learned strategies at least three times a week over a period of eight weeks. In addition, weekly reminders for self-observation and newsletters with stimulating ideas were sent out. The CG only had the opportunity to participate in a short information event. This event did not include any interactive elements. The quality of teacher–child interactions was assessed before and after the training in all groups using CLASS Toddler. During the assessment of global interactions, the frequency of math-specific interactions was also recorded – this is a project-specific addition. If interactions in the field of mathematics were observed, they were assigned to six categories: quantity/numbers/digits, size/measurement, space/shape, time, patterns/sequences, and sorting/classifying. For the following analyses, a sum score was derived from the frequency of all observed math-specific interactions per cycle. The activities were also noted during the CLASS Toddler observation. A distinction was made between free choice/interest areas, transition, routine, and group time. For each cycle, it was indicated how much percent of the time the respective activity took up – this is also a project-specific addition. To address the research question, repeated measures analyses of variance (RMANOVA) were conducted in R with data from 89 teachers. The results showed that for the MIG, math-specific interactions during free play improved marginally significantly ($F(2, 75) = 2.82$, $p = .07$, $\eta_p^2 = .03$), but not during structured activities. The quality of global interactions of both MIG and GIG did not significantly improve compared to the control group during both activities. However, an increase in quality was observed over time for all groups during both free play and structured activities (EBS: $F(1, 81) = 4.17$, $p = .04$, $\eta_p^2 = .02$; ESL: $F(1, 81) = 3.24$, $p = .08$, $\eta_p^2 = .01$; mathematics:

F(1, 75) = 13.12, p = .001, η_p^2 = .07). This study illustrates the potential of teacher trainings, but also shows that enhancing interaction quality during different activities can be more or less successful. Especially concerning toddler classrooms in Germany, this differentiated consideration is worthwhile, as free play takes up comparatively much time and a positive effect of the training for math-specific interactions during this activity could be identified.

In addition to the findings from each individual study, in the **overall consideration** it becomes evident that the analyzed aspects can be embedded into the Process–Person–Context–Time (PPCT) model of Bronfenbrenner and Morris (1998, 2006). This is a comprehensive theory that primarily considers the mutual influences of systems and their connections regarding interactions as part of processes. However, it becomes clear that the role of teachers – although they play an active and central part in shaping interactions – remains partly unexplored. This could be countered by placing the teacher more into the spotlight. One approach could be to examine their beliefs, self-efficacy, and competencies as well as changes of these attributes more closely. In addition, aspects such as job satisfaction, team climate, or opportunities for further professional qualification are crucial for successful interactions with children. This should be more strongly considered in research as well as in practice and in policy making. Designing a study that includes all essential systems of the PPCT model is challenging but meaningful, since it would allow not only individual systems, but also their interplay to be illustrated. A requirement for this is that the PPCT model is used reflectively. This primarily involves discussing systems that were not included and considering what impact this could have on results and conclusions.

Personal Contribution

General Contribution to the research project EarlyMath

- Co-development of instruments (questionnaires for teachers and parents, observation sheets, competence tests for children)
- Co-development and implementation of the teacher trainings
- Co-planning and coordination of the data collections
- Data collection in early childcare centers (CLASS Toddler observations and competence tests with children)
- Preparation and analyses of data with SPSS and R
- Supervision of student and research assistants
- General project administration (research planning, recruitment of student assistants, research assistants and study participants, communication with childcare centers and other stakeholders)

Contribution to Article 1 (first author)

- Literature search and screening
- Conceptualization and methodology
- Analysis and visualization of results
- Writing: original draft, editing, and revision

Contribution to Article 2 (first author)

- Conceptualization and methodology
- Analysis and visualization of results
- Writing: original draft, editing, and revision

Contribution to Article 3 (first author)

- Conceptualization and methodology
- Analysis and visualization of results
- Writing: original draft, editing, and revision

Overall, there was close collaboration within the research project EarlyMath. Analyses and results were discussed with project members and with colleagues from the German Youth Institute. All co-authors have contributed to the conceptualization and methodology of the articles. They have reviewed the articles and have approved the final drafts for submission.

Table of Contents

List of Abbreviations

ABC Checklist	Adults Behaviors in Caregiving Checklist
AWO QM	Arbeiterwohlfahrt Qualitätsmanagement
BETA-Gütesiegel	Bundesvereinigung Evangelischer Tageseinrichtungen für Kinder Gütesiegel
BG	Behavior Guidance
BGB	Bürgerliches Gesetzbuch
BiKS	Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Schulalter
BMFSFJ	Bundesministerium für Familie, Senioren, Frauen und Jugend
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CG	Control Group
CIP Training	Caregiver Interaction Profile Training
CIS	Caregiver Interaction Scale
CK	Content Knowledge
CLASS	Classroom Assessment Scoring System
COVID-19	Coronavirus Disease 2019
DIN EN ISO	Deutsches Institut für Normung, Europäische Norm, International Organization for Standardization
EarlyMath Project	Mathematical Development and the Impact of Interaction Quality in Early Childcare Project
EBS	Emotional and Behavioral Support
ECCOM	Early Childhood Classroom Observation Measure
ECEC	Early Childhood Education and Care
ECERS	Early Childhood Environment Rating Scale
EQOS	Educational Quality Observation Scale
ESL	Engaged Support for Learning
FIML	Full Information Maximum Likelihood
FLD	Facilitation of Learning and Development
GIG	Global Intervention Group
ITERS	Infant/Toddler Environment Rating Scale
KES	Kindergarten Einschätz Skala
KiföG	Kinderförderungsgesetz
KiQuTG	KiTa-Qualitäts- und Teilhabeverbesserungsgesetz
KJHG	Kinder- und Jugendhilfegesetz
KRIPS	Krippen Skala

KTK-Gütesiegel	Katholische Tageseinrichtungen für Kinder Gütesiegel
LM	Language Modeling
MIG	Math-Specific Intervention Group
ML	Maximum Likelihood
MLERS	Mediated Learning Rating Scale
NC	Negative Climate
NCKO Sensitivity Scale	Nederlands Consortium Kinderopvang Onderzoek Sensitivity Scale
NEPS	National Educational Panel Study
NICHD	National Institute of Child Health and Human Development
NUBBEK	Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit
OECD	Organisation for Economic Co-operation and Development
PC	Positive Climate
PCK	Pedagogical Content Knowledge
PITC	Program for Infant/Toddler Care
PK	Pedagogical Knowledge
PPCT	Process–Person–Context–Time
PQ-Sys	Paritätisches Qualitätssystem
QF	Quality of Feedback
QM	Quality Management
RCP	Regard for Child Perspectives
RCT	Randomized Controlled Trial
REACH	Reaching Educators and Children
RMANOVA	Repeated Measures Analyses of Variance
RMSEA	Root Mean Square Error of Approximation
SALSA	Search, AppraisaL, Synthesis, Analysis
SCCC	Secure Child in Child Care
SEM	Structural Equation Modeling
SRMR	Standardized Root Mean Square Residual
SSTEW	Sustained Shared Thinking and Emotional Well-being
TBRS	Teacher Behavior Rating Scale
TLI	Tucker-Lewis Reliability Index
TS	Teacher Sensitivity
US / USA	United States / United States of America
VIPP-CC	Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline for Child Care

List of Tables

Appendix – Study 1

Appendix – Study 2

Appendix – Study 3

Further Appendices

List of Figures

1 Introduction

In recent years, a substantial progress has been made in research regarding early childhood education and care (ECEC), linked with its value at various levels (e.g., OECD, 2021; Viernickel et al., 2016). Early childcare settings are not only an important infrastructure, as they are increasingly in demand due to the need for a double household income and the growing desire for gender equality and work-life balance (Anders & Roßbach, 2020), but they also serve as an important environment to promote children’s developmental processes and as a compensatory factor for unequal conditions for learning and growth (Burchinal et al., 2011; Kluczniok & Schmidt, 2020; Passaretta et al., 2022). From an economic perspective, ECEC is argued to promote long-term benefits, for both individual outcomes (e.g., improved school performance, career success, better health), as well as for societal outcomes (e.g., economic growth, social and healthcare expenses; Rammstedt, 2013). National and international studies on the effectiveness of ECEC suggest that stimulating environments with meaningful interactions can indeed support positive short-, medium-, and long-term child outcomes (Duncan et al., 2023; Lehrl et al., 2020; Suchodoletz et al., 2023; Ulferts et al., 2019).

In regard to its multifunctional role, the German ECEC system has undergone changes over the past decades and attendance has become standard (still with differences between East and West Germany; Autor:innengruppe Bildungsberichterstattung, 2022; BMFSFJ, 2021). Although attending is voluntary, there has been a legal entitlement to a childcare place for children over and under three years of age since 1996[1] and 2013[2], respectively. Alongside the quantitative expansion, the focus has progressively been placed on the quality of childcare settings, as they have developed into central learning environments where competencies are acquired, shaping the future paths of children (e.g., Klinkhammer et al., 2022; Love et al., 2003). Thus, various efforts were made to describe, improve, and evaluate the quality of ECEC. This includes the development and

[1] Gesetz zur Neuordnung des Kinder- und Jugendhilferechts (Kinder- und Jugendhilfegesetz – KJHG) vom 26. Juni 1990 (BGBl. I S. 1163).

[2] Gesetz zur Förderung von Kindern unter drei Jahren in Tageseinrichtungen und in Kindertagespflege (Kinderförderungsgesetz – KiföG) vom 10. Dezember 2008 (BGBl. I S. 2403).

implementation of educational plans across federal states, professionalization initiatives such as offering bachelor's and master's degrees in early childhood education, as well as enhancing ongoing teacher training. Moreover, measures have been designed to monitor and evaluate implementations into the ECEC system, which are constantly revised and further developed (Bryant et al., 2011; Thorpe et al., 2022). These progresses let to enhanced significance of research in the field of ECEC and – despite the strong socio-pedagogical tradition in Germany – early childcare settings are more and more seen as institutionalized education environments before the start of formal schooling. Nowadays, various scientific disciplines are active in this field (e.g., early childhood pedagogy, empirical educational sciences, developmental psychology), leading to diverse theoretical and methodological perspectives and approaches.

Building upon an introduction that illustrates the theoretical assumptions of this study, the present thesis aims to examine which different forms of instruments can be used for quality assessment in early childcare settings, how distinct dimensions of quality are related to other characteristics, and how certain practices of teachers can be strengthened. In the following, these considerations and findings are placed into an overall context along the described theory.

1.1 Theoretical Framework

1.1.1 Process–Person–Context–Time Model

Bronfenbrenner's bio-ecological theory, specifically the Process–Person–Context–Time (PPCT) model (Bronfenbrenner & Morris, 1998, 2006), provides a comprehensive framework for understanding human development. Emphasizing process, person, context, and time, this model investigates the complex impact the environment can have on an individual. The holistic approach considers the dynamic interplay of components – from immediate interactions to broader relations – to examine how individuals learn and develop over time.[3] As a guiding concept, the PPCT model comprises four interrelated systems shaping individual growth.

[3] Please note that the present thesis does not refer to systems theory (Luhmann, 1984; Luhmann & Schorr, 1979; Schmidt, 2005) and not to sectors other than ECEC, even though the components of the PPCT model are referred to as 'systems' in the following.

The main component of the PPCT model is the ***process system***, encompassing immediate, diverse day-to-day interactions, relationships, and engagements that children have within their direct environments. Those interactions encompass relationships with adults, experiences with peers, and interactions with objects and symbols. Children are seen as active part in shaping their progress, rather than being a passive recipient. Proximal processes are described as being the 'engines of development' and serve as primary mechanism to promote cognitive growth, emotional regulation, and social encounters. Specific interactions within the period of proximal processes have the capacity to affect child development by getting progressively more complex. The PPCT model particularly emphasizes the significance of daily interactions in shaping developmental trajectories: regular reciprocal interactions are needed to affect a child's well-being, which determines the ability to gain and retain competencies and knowledge. While proximal processes were initially discussed in a positive light, recent perspectives consider both constructive and destructive proximal processes, acknowledging that development can be negatively associated with under-complex interactions or by interactions that lead to dysfunction (Bronfenbrenner & Evans, 2000; Merçon-Vargas et al., 2020). As the nature of proximal processes varies, to evaluate their functionality and quality is essential: they should be estimated based on progressing complexity, frequency, duration, and interactional reciprocity. Examine processes to understand this system within the PPCT model is essential for comprehending the multifaceted nature of development through interactions.

The PPCT model describes the ***person system*** as being both input and outcome of systems. Personal characteristics crucially influence proximal processes, but they are also influenced by interactions and contextual factors over time. Personal attributes encompass individual innate characteristics such as personality traits, cognitive abilities, and genetic predispositions. Likewise, it comprises variable characteristics such as convictions and beliefs. Understanding the role of these personal characteristics in educational contexts is essential. Personal attributes are categorized into demand, resource, and force characteristics, each playing a fundamental role in influencing proximal processes:

- ***Demand characteristics*** are observable attributes such as age, gender, and physical appearance. Appearances like occurring happy or worried are also summarized under these characteristics. They are immediately apparent and stimulate reactions from the surroundings, meaning that they might either initiate or hamper social interactions in proximal processes.

- ***Resource characteristics*** are defined as mental, emotional, and social resources, including certain skills, knowledge, experiences, as well as (material) capital, which determines the access to education and the achieved degree, housing, health care, et cetera. These characteristics contribute to the ability to interact successfully during proximal processes, but they are less immediately visible.
- ***Force characteristics*** are dynamic personality traits such as motivation, self-efficacy, and beliefs that can either support or restrain interactions in proximal processes. Developmentally disruptive characteristics (e.g., impulsiveness, distractibility) may interfere with processes. The personal tendency for active initiatives (e.g., start and engage in activities) fit into those characteristics as well.

All these characteristics contribute to an ongoing child development through interactions and the PPCT model highlights the multi-directional connections between personal characteristics, contextual factors, and proximal processes. The dynamic nature of growth is brough into focus: the person system constantly evolves and changes because of the relation to the other systems (process, context, and time).

Understanding the role of the ***context system*** for child development is highlighted by the PPCT model, as contexts profoundly impact proximal processes. Different contextual factors are connected across various interrelated ecological systems. Those systems are represented as different layers around an individual, which is often described by the analogy of nested dolls (Rosa & Tudge, 2013). Each system broadens the framework by encompassing the complex environments in which the individual is embedded:

- The ***micro-system*** is the closest and most immediate layer, involving direct interactions of individuals, such as within the family, early childcare settings, or peer groups.
- The ***meso-system*** focuses on connections between multiple micro-systems (e.g., interactions between parents and teachers that in turn might affect the interactions with the child, and vice versa).

- The ***exo-system*** includes external influences and contexts in which individuals are not directly situated, such as parents' workplaces, impacting processes in the child's micro-system.
- The ***macro-system***, the outmost layer, evolves effects on child development by encompassing broader (sub-)cultural, social, and historical impacts that shape all other systems. This contains shared values, norms, beliefs, and resources, which have an impact on processes.

Contexts in which young children participate are usually limited to the home environment and early childcare settings. Personal characteristics of both the child and adults interfere with those environments. Distal contexts are also linked to proximal processes by third parties or external factors. All these contexts are dynamic over time: this can include internal and external changes due to critical events (e.g., birth or death, severe illness, or emergency situations). The interrelation of the systems and their associations with proximal processes are therefore fundamental for the PPCT model.

The PPCT model incorporates the ***time system*** as its final element, highlighting its role as important consideration for child development. With reference to the context systems, the time system is also referred to as *chrono-system*. Hereby, it becomes evident that the context and time systems are strongly interrelated. For a nuanced view, the time system divides several sub-components with a focus on different aspects of time:

- ***Micro-time*** refers to short and specific occurrences during a period of interactions. Interactions are described as continuous or discontinuous episodes, determining how long the proximal process lasts.
- ***Meso-time*** incorporates the stability and reliability of interactions over days, weeks, months, or years, providing insights into the consistency of experiences of an individual. It covers a longer period in which proximal processes take place.
- ***Macro-time***, or historical time, focuses on the broader societal changes within and across generations that shape the context in which development occurs. Historical or personal life events can be assets for further development.

Recognizing the temporal dimension of development, the time system highlights dynamic processes – short-, medium- and long-term – during growth across the lifespan. It emphasizes that development is continuous and not only influenced by personal and contextual characteristics, but also by time in which proximal processes take place, beyond a single direct interaction. Therefore, a longitudinal view is important for understanding development and its trajectories.

Overall, the PPCT model is based on the idea that development is not isolated, but occurs within different systems, which in turn are connected and influence each other. Events in one system can have impacts on others, either directly or indirectly influencing the child. Development can only take place over time, through a process with other persons, objects, or symbols within a context. It is necessary to understand how these components interact, stating that children can actively shape and are shaped by environments. This comprehensive framework offers a nuanced view to foster progress and growth in early childhood and across the lifespan.

However, the focus of the PPCT model lies on the child, whereas the teacher, although considered as a component, is not included as comprehensive mechanism. In the present thesis, the focus is intentionally placed on teachers to examine their specific role. Figure 1 (p. 7) illustrates the theoretical framework with the teacher as explicit element in Process–Person–Context–Time, although the child is still in the center.[4] This perspective allows to identify gaps that arise concerning the inclusion of relevant systems at the teacher level in research and practice, as well as in the conducted studies that are presented in the Appendix.

1.1.2 Quality in Early Childhood Education and Care

Early childhood education and care (ECEC) experienced substantial transformations throughout recent decades, driven by requirements and expectations of the public, politics, and academia. Hence, ‘good quality’ in ECEC came into focus, as it is nowadays not only seen as possibility to enable balance between work and family, but also to facilitate optimal child development and learning (Cadima et al., 2020; Melhuish et al., 2015). This progress emphasizes the profound role of quality in early childcare settings, but also raises the question of what ‘good quality’ in early childcare settings means and how it can be operationalized and archived.

[4] Please note that data from the research project EarlyMath was used for the present thesis. The project also investigated child development. The focus on the child is examined elsewhere and is therefore not included in this thesis.

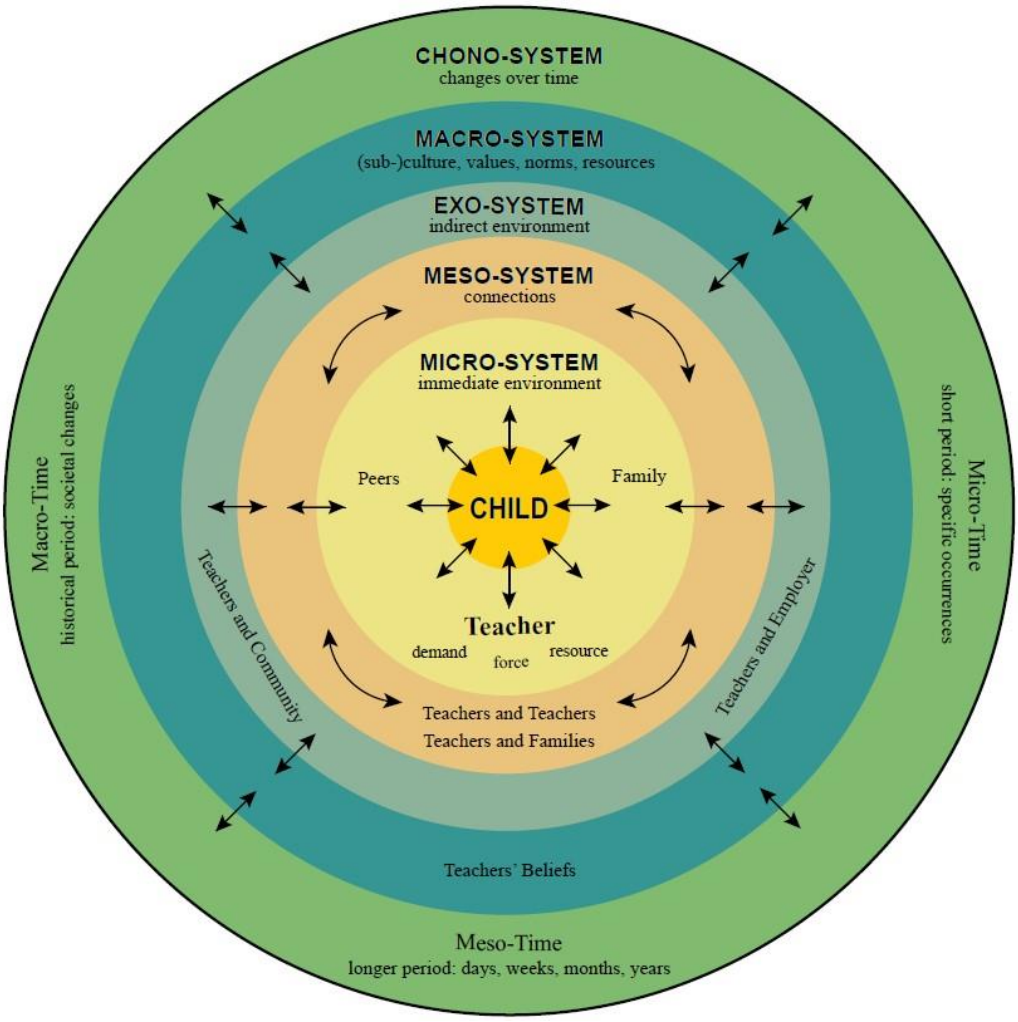


Figure 1. Bio-ecological Process–Person–Context–Time (PPCT) model (Bronfenbrenner & Morris, 1998, 2006; own illustration).

In Germany, discussions about the quality of early childcare settings have evolved around the turn of the millennium (Preissing & Heller, 2003; Roßbach, 2005; Tietze, 1998). Due to the rapid expansion of childcare centers (Berth & Grgic, 2020), this discussion focused on descriptive and quantitative aspects such as attendance rate, duration of attendance per day or week, or child age at entry. Initially, the debate was shaped by economic perspectives: concepts of quality management (QM) focused on guidelines and evaluations with a high degree of standardization. The norm DIN EN ISO 9000ff was adapted to ECEC settings and in many childcare centers, QM systems were implemented (e.g., 'KTK-Gütesiegel', 'BETA-Gütesiegel', 'PQ-Sys', 'AWO QM'; Esch et al., 2006). However, with time-delay, the debate increasingly shifted toward the essence of ECEC quality (Anders, 2013). Gradually, large-scale studies were initiated to find out more about quality in German early childcare settings: some of the most recognized are 'BiKS' (Bildungsprozesse, Kompetenzentwicklung und

Selektionsentscheidungen im Vorschul- und Schulalter, Maurice & Weinert, 2009), ‘NEPS’ (National Educational Panel Study, Blossfeld et al., 2009), and ‘NUBBEK‘ (Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit, Tietze et al., 2013). Each of them has a different focus and methodological approach, but they all aim at understanding what it means to attend institutional childcare and how this environment can unfold positive effects. The quality debate also had an impact on politics: the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (BMFSFJ), in cooperation with federal states, municipals, and youth welfare organizations, supported the development of a so-called ‘Nationaler Kriterienkatalog’ for quality in early childcare settings, which was updated and is still used, but is not mandatory (Tietze & Viernickel, 2023). Moreover, recently, there is a political will to ensure ECEC quality by law: from 2019 to 2022, the ‘Gute-KiTa-Gesetz’ was implemented and from 2023 to 2024, the ‘KiTa-Qualitätsgesetzt’ has been realized.[5] Both provide the opportunity for federal states to call for funds for specific quality topics to improve areas such as demand-oriented offers (e.g., extended opening hours), strong leadership (e.g., more time for management-related tasks), better teacher–child ratio, or more teacher trainings. For children, topics such as balanced nutrition, promotion of physical activity, health, and language education are included (Klinkhammer et al., 2021). In this regard, quality is broadly defined with different target groups, one of whom is the teacher.

Ongoing changes in the ECEC system warrants investigation. To evaluate processes within early childcare centers and their impact on child outcomes, a multidimensional construct is utilized, traditionally divided into *process quality* (***process system***), *structural quality* (***context system***), and *orientation quality* (***person system***) (Kluczniok & Roßbach, 2014). Process quality involves daily experiences, for example interactions between teachers and children, but also between teachers and parents as well as within peer groups. Children’s interactions with provided materials is also included. Going into more detail, process quality is further categorized into global and domain-specific interaction quality (see chapter 1.1.3; Sylva et al., 2006). Structural quality refers

[5] Gesetz zur Weiterentwicklung der Qualität und zur Verbesserung der Teilhabe in Tageseinrichtungen und in der Kindertagespflege (KiTa-Qualitäts- und Teilhabeverbesserungsgesetz – KiQuTG) vom 19. Dezember 2018 (BGBl. I S. 2696), das durch Artikel 1 des Gesetzes vom 20. Dezember 2022 (BGBl. I S. 2791) geändert wurde.

to institutional conditions (e.g., size and equipment of the early childcare center, group size, teacher–child ratio, or the teachers' qualification), and sets the framework in which daily processes take place. These characteristics can be changes by political action or by other stakeholders (Wertfein et al., 2009). Orientation quality encompasses teachers' values, norms, and beliefs, guiding their pedagogical work and behavior towards children (Anders, 2013). In some publications, further aspects of quality are categorized, such as team quality (Eling, 2022). Teachers are often the focus of investigation, as they are the primary attachment figure for children in early childcare settings and an important part of quality dimensions. Current research is eager to closely investigate various aspects of teacher–child interactions, the driving force of development, aiming to examine them with more accuracy and depth (e.g., Oppermann et al., 2024). This requires operationalizing various dimensions of interaction quality and comparing different ways of assessments. Two broad frameworks of how to capture this quality can be distinguished: one considers quality as a holistic construct and includes many aspects, for example the equipment of the early childcare setting, children's access to materials, and proximal processes in the classrooms, including interactions (Gordon et al., 2015). The other defines quality in more detail and focuses entirely on proximal processes, more specifically on teacher–child interactions (Pianta et al., 2020).

Internationally, there is a consensus that ECEC, as central learning environment (besides the home learning environment), is essential for fostering children's developmental trajectories (e.g., Peisner-Feinberg et al., 2001; Siraj-Blatchford et al., 2002; Sylva et al., 2011; Vandell et al., 2016). However, conceptualizations of its quality are often influenced by US-American perspectives (e.g., Environment Rating Scale Institute, North Carolina; Teachstone, Virginia). In the US, studies on early childcare and preschools initiated in the 1960s: 'Head Start' (Resnick & Zill, 1999), the 'Perry Preschool Project' (Schweinhart & Weikart, 1980), and the 'Study of Early Child Care and Youth Development' (NICHD, 2002) are worth mentioning. To date, these are some of the largest projects in the field of ECEC. This is why they have also profoundly shaped the quality debate in Germany, even though many pedagogical approaches are largely based on European scientists (e.g., Fröbel, Steiner, and Montessori). The insight into these different perspectives on quality reveals that it is a broad construct. It encompasses various aspects and can be conceptualized differently. Moreover, determine 'good quality' depends on the view of the observer (Katz, 1993). In this thesis, the quality of interactions from a scientific view is the focus, as they describe the direct contact between teachers and children. Subsequently, these interactions are further differentiated.

1.1.3 Global and Math-specific Teacher–Child Interactions

Among various facets of quality, interactions between teachers and children stand out (***process system***), significantly influencing (academic) outcomes such as literacy and mathematics, as well as socioemotional skills (Bleses et al., 2020; Clements et al., 2011). Findings highlight the diverse impact of global and domain-specific interactions on child development, making it necessary to distinguish them (Hamre et al., 2014; Yang & Liang, 2022). Recognizing the unique contributions of each type of interaction can lead to more effective educational strategies tailored to specific needs. Given the growing body of research emphasizing the significance of early mathematics in both school and later life (Hooper et al., 2010; Nguyen et al., 2016), it is essential to explore how children can develop their mathematical ability through effective teacher–child interactions.[6]

Global interactions encompass a range of engagements simultaneously, including emotional, behavioral, and general learning-related strategies. These sensitive, sufficient, and stimulating interactions play a fundamental role in children's development and learning (Kasüschke, 2010). Among different didactic styles, *co-construction* has established itself – in western cultural circles – as guiding principle. It is characterized by an encouraging exchange between the teacher and the child, which is seen as constructive learning process in which both individuals construct knowledge and solutions collaboratively (Gasteiger, 2012). An important pedagogical approach is *sustained shared thinking,* where teacher and child act out the meaning of learning experiences by jointly developing strategies to solve problems, exchanging thoughts and ideas about activities, and coordinating their actions (Siraj-Blatchford et al., 2015). Teachers can apply *scaffolding* to support the child in approaching the next (developmental) step, which does not mean that the teacher instructs the child to do something or leaves a task entirely to the child, but rather provides support so that the child can achieve it with the teachers' guidance on his or her own (Berk & Winsler, 1995).

---

[6] Please note that the research project EarlyMath had a special focus on mathematics and therefore, the present thesis investigates math-specific interactions as part of domain-specific interactions.

Math-specific interactions target particular competencies related to mathematical skills, which gives children the chance to experience mathematical meanings and constructs. These interactions offer guidance and experiences related to mathematical constructs, including mathematical language and activities such as counting and measuring. Despite explicit math-related teaching units, *teachable moments* are crucial for math-related learning, particularly for young children in ECEC (Hyun & Dan Marshall, 2003). Teachers should observe and utilize moments suitable for mathematical prompting (‘everyday mathematics’; Ginsburg et al., 2008). Such moments can arise spontaneously during play, book reading, or other activities, making it difficult for teachers to perceive these moments and to use them promptly. This requires a high degree of motivation, flexibility, and awareness (Torquati et al., 2007). Mathematical competencies and skills that can be addressed in such moments can be understood in a narrow and broader sense. A *narrow understanding of mathematics* covers numerical terms and the number system (e.g., counting). In contrast, a *broad understanding of mathematics* also involves spatial, temporal, and quantitative elements (e.g., small, tall, short, long, less, more; Ramani et al., 2015). Especially in early childhood, it is crucial to encompass all aspects of mathematical understandings, as a comprehensive approach and the resulting broad child competencies serve as foundation for later math-specific skills (Klibanoff et al., 2006; Ramani & Siegler, 2008).

This description highlights that there are different forms of interactions that require individual consideration. Teachers may potentially stimulate global and math-specific interactions more or less effectively and support respective areas accordingly. Consequently, it is crucial to distinguish between these two facets of interaction quality.

1.2 Current State of Research

In the following, teachers and the quality of their interactions with children are the center of investigation. Furthermore, the focus is primarily placed on teachers in toddler classrooms, as this is an important yet still largely unexplored environment in many respects. An increasing number of toddlers take part in German early childcare setting (mainly since 2006; Berth & Grgic, 2020), making it an important environment to consider. Despite its importance, there is limited understanding of proximal processes such as global and math-specific interactions that occurs in toddler classrooms. Therefore, how interactions in toddler classrooms can be assessed, what associations exist between interactions and other characteristics, and how teacher–child interactions can be enhanced is presented subsequently.

1.2.1 Assessing Interactions (Process)

Regarding the large area of early childhood research, a diverse range of instruments is available for assessing the global quality of interactions and those specific to mathematics (***process system***). Among these, observations are considered the ‘gold standard’, as the accuracy of self-reports is questioned (Linberg et al., 2017). However, observational instruments vary in criteria related to the instrument’s methods, for example whether they apply macro- or micro-analytic approaches (Fassnacht, 1995). A closer look at how a specific instrument works is important to be able to differentiate and to be aware of which proximal processes are assessed, if and which personal or structural characteristics are included, and what role time plays in this context.[7]

In the current international research on ECEC, two assessments (supplemented by a few others) are predominantly utilized for assessing global quality in early childcare settings (Halle et al., 2010). The most frequently used observational instruments that can be used in relation to toddlers are the Infant/Toddler Environment Rating Scale (ITERS; Harms et al., 2006; Harms et al., 2017) and the Classroom Assessment Scoring System Toddler Version (CLASS Toddler; La Paro et al., 2012). Both were originally developed in the USA, but – unlike CLASS Toddler – there are German versions of ITERS, aiming to adapt some elements to the national context (KRIPS; Tietze et al., 2007; Tietze & Roßbach, 2019). ITERS is a comprehensive instrument to assess overall quality that also includes elements for rating structural quality (***process–context–time system***), while CLASS Toddler assesses interactions in detail across two domains (***process–time system***; emotional and behavioral support and engaged support for learning). Despite their widespread use, there is critique on both instruments: the ‘stop-scoring approach’ used in ITERS/KRIPS is criticized, since aspects of higher scale points might not be rated even if they would have met quality standards. Moreover, items from different areas sometimes seems to be mixed within one domain (Mayer & Beckh, 2018). Regarding CLASS Toddler, it is criticized that profound distinctions can be made only between very good

[7] Please refer to the Introduction section ‘Instruments for assessing interaction quality in ECEC’ in the Appendix – Study 1 (pp. 68–69) for further description.

and very poor quality and that a more nuanced assessment is challenging. Furthermore, various indicators of dimensions are captured simultaneously and summarized to one global rating, leading to a less differentiated and partly subjective view on interaction quality (Mayer & Beckh, 2018). Although there are also other established instruments existing (e.g., CIS; Arnett, 1989), and the availability of instruments continues to change (e.g., SSTEW; Siraj-Blatchford et al., 2015), research with and on the two listed instruments is currently the most extensive that is being conducted.

Concerning math-specific aspects of quality in early childcare settings, ITERS-3 (Harms et al., 2017) – the latest version of these scales – offers ‘math/number’ as a new item in the section ‘activities’. This was not included in previous versions, making ITERS-3 an instrument that assesses a mix of global and math-specific interactions. This only partly applies for the latest German version, KRIPS-RZ (Tietze & Roßbach, 2019), which at least assesses the activity of building blocks, but not explicitly ‘math/number’. CLASS Toddler offers no math-specific (or other domain-specific) views on interactions. There are no other instruments that can be used for toddlers, which include mathematics as an independent aspect. This is different for older children, where there are instruments that assess math-specific characteristics of interactions and the environment (e.g., ECERS-E/KES-E; Roßbach & Tietze, 2018; Sylva, Siraj-Blatchford, & Taggart, 2011).

To assess interaction quality, there are various methods and types of instruments available (whereby this is very limited for math-specific interactions in toddler classrooms). They differ mainly in the level of detail in which interactions are captured and which other aspects are included. With reference to the PPCT model, one can conclude that there is a substantial difference of which systems are included in the assessment (***process–time system*** vs. ***process– context–time system***).

1.2.2 Associations between Interactions and Person–Context

Interactions (***process system***) are dependent on various influencing factors regarding personal (***person system***) and structural (***context system***) characteristics. Several conditions within early childcare settings are shaped by governmental regulations and they may change over time.[8] Investigating associations between personal and structural

[8] Please refer to the Introduction section ‘The German ECEC system’ in the Appendix – Study 2 (pp. 112–113) for further description.

characteristics and teacher–child interactions is crucial for identifying indicators for beneficial ECEC (e.g., Locasale-Crouch et al., 2007).

Regarding teachers (***process–person system***), findings about the impact of their qualification, professional experiences, and further training are inconsistent (***resource characteristics***). Some studies indicate (partly) positive associations between teacher' education and quality in toddler classrooms (Burchinal et al., 2002; Goelman et al., 2006; Slot et al., 2015), while others show no such relations (Vermeer et al., 2008). Comparing the effects of teacher education on teacher–toddler interactions in different European countries revealed country-specific difference: a negative effect of teacher education was found only for one country (Portugal; Cadima et al., 2022). Like those results, the findings concerning effects of teachers' professional experience on quality are inconsistent (positive associations: Jamison et al., 2014; Manning et al., 2019; null or negative associations: Justice et al., 2008; Suchodoletz et al., 2014). Only when it comes to the impact of further teacher training, there is consensus that it has positive effects on interaction quality (Egert et al., 2020; Slot et al., 2015). Variability is observed in studies examining the impact of teachers' age on quality in toddler classrooms (***demand characteristics***): Pessanha et al. (2007) found a positive effect of being a young teacher, while van IJzendoorn et al. (1998) reported a negative effect. Contrary, there is a clear picture when it comes to pedagogical beliefs (***force characteristics***): findings suggest that teachers who follow child-centered approaches positively affect the quality of interactions (Eckhardt & Egert, 2018; Pianta et al., 2005). A high degree of teachers' self-efficacy also indicates beneficial influences, as it seems to lead to higher motivation and better well-being of teachers, which in turn lead to successful interactions (Guo et al., 2010; Hu et al., 2021). The insight provided in this section demonstrates that teacher characteristics significantly influence interactions, although it is not always conclusively clarified how this is the case. Furthermore, it becomes evident that ***resource characteristics*** are more prominently examined in research than ***demand*** or ***force***.

Besides teacher characteristics, classroom characteristics play a crucial role for interactions in ECEC as well (***process–context system***; ***micro-system***). Regarding group size and teacher–child ratio, consensus exists across age groups and different interaction domains that quality is higher in smaller classrooms with lower teacher–child ratios (e.g., Barros & Aguiar, 2010; Løkken et al., 2018). However, some studies also found no effect of teacher–toddler ratio and group size on emotional and behavioral support and engaged support for learning, possibly due to already existing high-standard regulations on those parameters (Cadima et al., 2022; Slot et al., 2015). The composition of children in the classroom seems to be more decisive, since responding to children's needs with high-quality interactions may be more challenging in a diverse classroom (Diebold & Perren, 2020; Schipper et al., 2007). However, these mechanisms are complex, and it is not always clear which component has the greatest impact on teacher–toddler interactions.[9] Moreover, another aspect of context does not refer to the childcare center or the classroom, but to the activity setting in which interactions take place. Recent findings give insights that the quality of teacher–toddler interactions might vary depending on specific activity settings: structured activities seem to display more learning supportive interactions compared to other activities (Cabell et al., 2013; Wildgruber et al., 2016) and mealtimes are described as situations where interactions demonstrate the lowest quality (Guedes et al., 2020). These insights are partly based on research in preschools, as research in toddler classrooms is sparse, although recent studies are conducted on this topic.

Overall, it is evident that attributes in the ***person system*** significantly influence interactions with toddlers, although it is not yet always clear in which ways. The teacher, in turn, is influenced by structural conditions in the ***context system*** within which s/he is embedded. Consequently, the quality of interactions cannot be examined without considering and incorporating both the personal characteristics of teachers and the structural characteristics surrounding them.

1.2.3 Improving Interactions in Context–Time

Mainly, there are two ways to adjust conditions and increase the quality in early childcare centers: on the one hand, indirectly by improving relevant aspects of the above-mentioned context (e.g., through policy making) and, on the other hand, through further training carried out directly with teachers. How to effectively raise the quality of global teacher–toddler interactions is a particularly important issue as several studies across different countries reveal low-quality levels (Bichay-Awadalla & Bulotsky-Shearer, 2021; Bücklein et al., 2017; Cadima et al., 2022; Salminen et al., 2021). The same applies to

[9] Please refer to the Introduction section 'Structural characteristics of quality' in the Appendix – Study 2 (pp. 111–112) for further description.

math-specific interactions: minimal time is placed on these interactions compared to other specific areas (Early et al., 2010). Furthermore, teachers tend to focus on a limited range of math-specific content, predominantly emphasizing counting small numbers (Pelkowski et al., 2019).[10] This indicates that there is a need for the quality of interactions to improve over the course of the ***time system***.

When considering teacher trainings for improving interactions (***process–time system***), trainings targeting teachers in toddler classrooms show different results regarding the improvement of global quality (partly effective: 'SCCC', Biringen et al., 2012; 'REACH', Conners-Burrow et al., 2017; 'Responsive Early Childhood Program', Landry et al., 2014; 'Expanding Quality for Infants and Toddlers', Moreno et al., 2015; 'VIPP-CC', Werner et al., 2018; not effective: 'CIP Training', Helmerhorst et al., 2017; 'PITC', Weinstock et al., 2012).[11] Concerning trainings for math-specific teacher–toddler interactions, only one teacher training, the 20-week intervention 'Play and Learn', provides evidence (Bleses et al., 2020). Teachers in toddler classrooms were equipped with strategies and tools to be more explicit and intentional in daily interactions and activities, both globally and in terms of specific content, including math vocabulary and numeracy skills. After a two-day training workshop, teachers were instructed to create activities of their choice to promote children's active engagement in learning occasions, for example by using math talk. This intervention revealed positive training effects on child competencies, however, effects on the teacher-level are not described. This shows that when it comes to teacher–toddler interaction quality, there has been limited implementation and evaluation of teacher trainings targeting math-specific interactions and activities, and results on global interactions differ.

In addition to considering the variability of interactions over time, it is also of interest to consider different contexts in which improvement of interactions occurs (***process–context–time system***). Free play and structured activities are uniquely beneficial for fostering effective teacher–toddler interactions (Nores et al., 2022). In ECEC settings, free play is initiated by the child and not necessarily led by teachers, while structured situations are planned and facilitated by teachers with a high degree of guidance (Fuligni

[10] Please refer to the Introduction section 'Global and math-specific interactions' in the Appendix – Study 3 (pp. 140–142) for further description.

[11] Please refer to the Introduction section 'Improving interactions in toddler classrooms' in the Appendix – Study 3 (pp. 143–145) for further description.

et al., 2012; Goble & Pianta, 2017). Teachers may find it easier or more challenging to address different domains and dimensions of interactions depending on these activity setting. Few studies explicitly examined this interplay (Guedes et al., 2020; Slot et al., 2015).[12] Only one study explored the significance of activity settings for interactions and child development (excluding mathematics): it demonstrated that the intervention 'VIPP-CC' (six 1.5 hour long visits scheduled two to four weeks apart) was particularly effective in structured activities (d = .64), but not in free play (in this setting, the control group even scored higher; Werner et al., 2018). It is the only study about quality improvement which includes all three systems in toddler classrooms: ***process system*** (interactions), ***context system*** (activity settings), and ***time system*** (pretest, posttest).

To summarize, interactions can be improved through approaches such as teacher training (although it is critical which methods are applied). This improvement can be noticed at different levels: with the teachers themselves (e.g., beliefs), with the interactions (e.g., quality), and with the children (e.g., competencies). The effectiveness of trainings can be crucially influenced by where interactions take place. Different contexts may require different expertise from teachers to meet children's needs and use interactions effectively, and trainings may also have different effects in these settings. The specific role of teachers has also not been conclusively investigated. Overall, the comprehension of all these factors is still rudimentary as research in toddler classrooms is inconsistent or lacking. Hence, it is crucial to further examine the interplay of the systems process, person, context, and time.

1.3 Present Thesis

Recognizing the significance of comprehensive perspectives, it is essential to incorporate all systems of the PPCT model into these considerations. Therefore, **the aim of the present thesis is to illustrate the complex interplay of systems (Process–Person–Context–Time) and to examine the role of the teacher within the systems**. Specifically, the focus is placed on the interactions between the teacher and toddlers. This allows for the identification of areas that have been underexplored or overlooked by practice and research.

---

[12] Please refer to the Introduction section 'The role of activity settings' in the Appendix – Study 3 (pp. 142–143) for further description.

Systematic reviews are a helpful method to gain an overview of the current state of knowledge about how interaction quality as part of the ***process system*** can be assessed. Taking the emerging focus on teacher–toddler interactions into account, it is necessary to conduct a systematic review to consider which instruments can be applied to assess these interactions and how they are designed. As reviews of instruments often lack a math-specific focus (especially regarding toddlers), it is also important to consider this aspect in more detail. Contemplating the ***process–person–context system***, it appears that different personal and structural characteristics have an impact on teacher–toddler interaction quality. However, conclusions about associations of personal and structural characteristics are inconsistent: mainly, this applies to whether teachers' qualification and professional experiences or their age affects their interactions. Furthermore, the significance of children's composition in the classrooms remains uncertain. Investigating these aspects is essential to better understand impacts of personal and contextual factors on teacher–toddler interactions. Examining the ***process–context–time system*** reveals that contexts might influence the improvement of teacher–child interaction quality. However, regarding toddler classrooms, studies on the role of different activity settings for interactions in toddler classrooms are limited, and findings on preschool classrooms are mixed. This highlights the need for further exploration in this area. A varied perspective on the meaning of activity settings in enhancing high-quality global and math-specific interactions is needed to promote their quality and the abilities of teachers effectively.

In light of the overall research question of teachers' role in the Process–Person–Context–Time interplay, and regarding the identified research gaps in toddler classrooms, the present thesis addresses the following research questions:

(1) In consideration of the ***process system*** in the PPCT model: Which instruments can be used to assess global and math-specific teacher–toddler interactions? (Study 1)

(2) In consideration of the ***process–person–context system*** in the PPCT model: What associations exist between personal and structural characteristics and the quality of teacher–toddler interactions? (Study 2)

(3) In consideration of the ***process–context–time system*** in the PPCT model: Do global and math-specific teacher–toddler interactions improve across different activity settings through teacher trainings? (Study 3)

2 Method

2.1 Analyzing Instruments

To address the first research question, we investigated how the ***process*** can be assessed. For doing so, we conducted a systematic review of literature using the SALSA method (Grant & Booth, 2009).[13] We initially identified 4211 publications and reduces them to 148 publications meeting our inclusion criteria after the removal of duplicates (articles appearing across several databases). Figure 1 in the Appendix – Study 1 shows the screening process involving title, abstract, and full-text screening. In the systematic review, the included publications were analyzed and evaluated according to various analysis criteria: (sub)scales, items, method, procedure, children's age, reported reliability, and reported validity. Referring to these findings, the present results section reports on which instruments capture what type of interactions (global vs. math-specific), to what extent they include other aspects (focus on processes vs. inclusion of additional aspects), and to what extent the teacher is explicitly considered.

2.2 EarlyMath Project

To address the second and third research questions, data from the project EarlyMath[14] ('Mathematical Development and the Impact of Interaction Quality in Early Childcare') was used. The quasi-randomized controlled trial (RCT) examined the mathematical development of children aged two to four years and the impact of global and math-specific interaction quality in ECEC settings. EarlyMath specifically targeted the variability of teacher–child interaction quality in toddler classrooms through the implementation of teacher trainings. An experimental, multi-cohort, pre-post-follow-up

[13] Please refer to the Method sections 'Data sources and searching process', 'Screening process', and 'Sample description' in the Appendix – Study 1 (pp. 70–72) for further description.

[14] EarlyMath was conducted as a collaboration involving the Otto-Friedrich University of Bamberg (Prof. Dr. H.-G. Roßbach, Prof. Dr. S. Weinert, Prof. Dr. S. Lehrl, Dr. D. Dornheim) and the German Youth Institute (Dr. A. Linberg). It was funded by the German Research Foundation under the grants DO 2304/1-1, LE 3245/1-1, LI 3487/1-1, RO 820/18-1, and WE 1478/13-1. The Ethics Committee of the Otto-Friedrich University of Bamberg approved the project.

intervention design was implemented, capturing teacher–child interactions, child competencies, general childcare characteristics as well as family characteristics (see Figure A1 in the Further Appendices for the overall study design). The present thesis focusses on 95 teachers[15] and gives insights into teacher–child interactions in 95 toddler classrooms. Information about teacher and classroom characteristics is provided in Tables 1 and 3 in the Appendix – Study 1 as well as in Table 1 in the Appendix – Study 2.

2.2.1 Analyzing Associations

To address the second research question, we assessed teacher–child interactions – as part of the ***process–time system*** (***micro-time***) – in all toddler classrooms. This was done through the Classroom Assessment Scoring System Toddler Version (CLASS Toddler, La Paro et al., 2012). This instrument typically captures global interaction, but to also consider math-specific interactions, we added the category 'math' on the observation sheet. Ten research assistants were considered reliable for data collection and were blinded regarding teachers' study conditions. To assess global interaction quality, research assistants rated the interactions on a 7-point Likert scale across eight dimensions: positive climate (PC), negative climate (NC), teacher sensitivity (TS), regard for child perspectives (RCP), behavior guidance (BG), facilitation of learning and development (FLD), quality of feedback (QF), and language modeling (LM). To rate math-specific interactions, research assistants noted their instances and categorized the frequency of teachers' engagement with mathematical content into three levels: (1) one time, (2) two to three times, or (3) four times and more. Those math-specific interactions were classified into six mathematical categories: quantity/numbers/digits, size/measurement, space/shape, time, patterns/sequences, and sorting/classifying (Benz et al., 2015; Fthenakis, 2009; Gasteiger, 2014). When no interaction related to mathematical content was observed, it was recorded as zero. Each classroom was observed on a regular morning across various activities, including morning routines, free play, and structured activities. The research assistant conducted three 15–20 minutes live-observation cycles and one

[15] Please refer to the Method sections 'Sample' in the Appendix – Study 2 (p. 114) and 'Participants and randomization' in the Appendix – Study 3 (pp. 146–147) for further description. Data collection for the first cohort took place from December 2020 to April 2021 (pretest) and from April to June 2021 (posttest), and for the second cohort from January to March 2022 (pretest) and from May to August 2022 (posttest).

video-recorded 20 minutes cycle (subsequently rated), where each teacher interacted with up to three children using materials provided by the research assistants. These materials – picture books and a board game – and the instructions remained standardized across all teachers: they were instructed to use the materials as they typically would do. In total, 730 cycles were assessed (pretest: 377 cycles, posttest: 353 cycles).

Teachers' personal characteristics (***person system***; ***demand characteristics*** and ***resource characteristics***) and structural characteristics (***context system***; ***micro-system***) of the toddler classrooms were assessed via questionnaires for the participating teacher (e.g., teachers' education, age of children in the classroom) and throughout the observations (e.g., number of children, number of teachers). For the analysis of associations, we intentionally selected personal and structural characteristics showing variability to prevent multicollinearity issues (see Table 4 in the Appendix – Study 2). Consequently, we examined teacher education and age, alongside the classroom characteristics group size and children's age composition. Teachers' education level was categorized into six groups: (1) incomplete education, (2) education with an internship component, (3) completion of a two-year vocational training, (4) completion of a five-year vocational training, (5) attainment of a bachelor's degree, or (6) attainment of a master's degree. To determine teachers' age, they provided their birth year, subtracted from the year in which they filled out the questionnaire. Group sizes were derived from the average number of children present during all four observation cycles. Regarding age composition within classrooms, teachers specified the number of children falling within specific age ranges: 0–12 months, 12–24 months, 24–36 months, and over 36 months. Subsequently, a dichotomous variable was computed to assess if all children in a classroom were within a two-year span.[16]

For the analysis, structural equation modeling (SEM) was conducted in R (lavaan, Rosseel, 2012; semPlot, Epskamp, 2019). SEM enables simultaneous analysis of observed and latent variables, offering an advantage over traditional regression analysis. Model fit relied on various parameters: comparative fit index (CFI), Tucker-Lewis reliability index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). CFI and TLI > .95 indicated good fit,

[16] Please refer to the Method sections 'Measures' in the Appendix – Study 2 (pp. 115–116) and 'Measures' in the Appendix – Study 3 (pp. 150–152) for further description.

while > .90 suggested adequate fit. RMSEA < .06 signified close fit, while < .08 reflected reasonable fit. For SRMR, < .08 indicated adequate fit (Bentler & Bonett, 1980; Hu & Bentler, 1999; Steiger, 1990). Employing the full information maximum likelihood (FIML) to handle missing data (5.3%) enabled to include the complete sample of 95 observations with the maximum likelihood estimator (ML) for the analysis.[17]

2.2.2 Analyzing Improvement

To answer the second research question, we conducted teacher trainings to improve global and math-specific teacher–toddler interactions relating to the ***process–time system (meso-time)***. All teachers were randomly assigned to three groups: experimental group 1 (math-specific intervention, $n_{\text{posttest}} = 30$), experimental group 2 (global intervention, $n_{\text{posttest}} = 29$), or control group (business as usual, $n_{\text{posttest}} = 30$). The randomization was conducted at the childcare center level, with matching based on location, center size, percentage of multilingual and socially disadvantaged children. Teachers in the experimental groups underwent a two-day training with eight modules and implementations lasting nine weeks in total. The implementation mainly consisted of a repetition module, newsletters, and reminders for self-observation with observation sheets that were already used during the trainings. The math-specific intervention group examined global as well as math-specific teacher–child interactions, while the global intervention group concentrated on global teacher–child interactions. Despite the content variations, both experimental groups followed the same methodology and scope. Throughout the trainings, great emphasis was placed on day-to-day interactions in toddler classrooms. The main approaches used, were shared book readings and playing (board) games, but also everyday situations were addressed. An overview of the training procedure and content is available in Table A1 in the Further Appendices. Teachers in the control group attended a 90-minute virtual information session covering project details and brief theoretical insights, but without any interactive components or materials.[18] For an overview of global and math-specific teacher–toddler interactions in

---

[17] Please refer to the Method section 'Data analyses' in the Appendix – Study 2 (pp. 116–117) for further description.

[18] Please refer to the Method section 'Intervention' in the Appendix – Study 3 (pp. 147–150) for further description. The intervention for the teachers assigned to experimental groups was conducted in March/April 2021 and March/April 2022.

the pre- and posttest categorized by groups, see Table 2 in the Appendix – Study 3.

During the observations in the toddler classrooms, research assistants categorized the activity settings that occurred in each cycle (***context system***; ***mico-system***). Aligning with the CLASS Toddler manual, four activity settings were assessed: free choice/interest area, transition, routine, and group time. Throughout their observations, research assistants determined the duration of each setting by indicating whether it occurred 1–25%, 26–50%, 51–75%, or 76–100% of each cycle or not at all. These indicated percentages are a study-specific addition to the CLASS Toddler observation sheet, enabling a more precise investigation of activity settings. This makes it possible to investigate whether improvement through the teacher trainings might only appear in a specific activity setting. For the subsequent analyses, activity settings were categorized as following: interactions where free choice persisted for more than half the cycle time were considered instances of free play, and those where group time persisted for more than half of the cycle time were considered structured activities. Among the 661 observed cycles in pre- and posttest where activity settings were indicated, free play was noted in 283 cycles (42.8%), while structured activities were observed in 363 cycles (54.9%).

For the analysis, we employed repeated measures analyses of variance (RMANOVA) in R using rstatix (Kassambara, 2023). To evaluate the effect sizes, we relied on partial eta-squared (ηp^2), a commonly used metric in RMANOVA. As per Cohen (1977), effect size benchmarks were categorized as follows: $\eta p^2 \sim .01$ = small, $\eta p^2 \sim .06$ = medium, and $\eta p^2 \sim .14$ = large. Upon detecting significant group*time effects, post-hoc tests with Bonferroni adjustments were conducted through multiple pairwise comparisons (also using rstatix). The graphical illustration of intervention effects was done with gglot2 in R (Wickham et al., 2023). Dropout rate among teachers was 6.3%. Refer to Figure 1 in the Appendix – Study 3 for participant flow and measurement timepoints. Our dataset included complete data for all posttest observations (N = 89).[19]

[19] Please refer to the Method section ‘Data analyses’ in the Appendix – Study 3 (pp. 152–153) for further description.

3 Results

3.1 Process: Examining Different Types of Instruments

In the systematic literature review, in total 55 instruments that assess global and/or math-specific interaction quality were identified (see Table 2 and Table 3 in the Appendix – Study 1 for an overview). The most frequently used instrument was ECERS-R (22.4%), followed by CLASS Pre-K (15.7%) and CIS (7.6%) – only the latter can also be applied for toddlers. Concerning toddler classrooms, of the 55 instruments, 34 instruments remain that focus on interactions with children younger than 36 months. Here, despite CIS, ITERS-R was the most frequently used (5.7%; please refer to Table 5 in the Appendix – Study 1 for authors, year, and country of the identified studies). This shows that ECERS/ITERS and CLASS are the most used instruments, but it also shows that there are several others that can be used for assessing interactions.

Regarding the frameworks underlying these instruments, there is a notable difference in the conceptual assumptions on which they are based. One framework considers global interactions as part of overall ECEC quality and thus as one aspect of many, while the other focuses on interactions and distinguishes different facets of this behavior. There was a wide variety of scales per instrument, ranging from one (CLASS Infant, NCKO Sensitivity Scale) to twelve (MLERS), reflecting the huge differences in style and level of detail in which interactions are captured – but also if and how other aspects are addressed (e.g., context). This also applies to the items: here was a range from four items (CLASS Infant) to over 100 (EQOS) per instrument. Despite the great variation of each instrument, some similarities can be noticed. ‘Instrument-families’, like ECERS/ITERS and CLASS, use different instruments to assess interactions (and other aspects) in different age groups, but they follow the same framework and are therefore similar to each other. That could be one reason why they are frequently used, as it enables progress to be tracked with instruments having the same design and method.

Likewise, instruments assessing math-specific interactions show different ways of conceptualizing mathematical aspects. Most instruments capture math-specific interactions as one construct with a clear mathematical focus, which means that those interactions are assessed separately from global or other specific interactions (e.g., ITERS-3). More recently developed instruments differentiate among several mathematical aspects and therefore have higher levels of specificity, however, this mostly applies to preschool classrooms. For toddles, seven instruments are assessing math-specific interactions versus thirteen instruments for older children, reflecting the lack of consideration of these specific activities in toddler classrooms. Nonetheless, there is already variety in capturing mathematical aspects within these instruments. Worth noting is the 'ABC Checklist' (Adults Behaviors in Caregiving Checklist, Honig & Lally, 1973): although developed many years ago, it considers a broad understanding of mathematics by using the items 'arranges learning of number', 'arranges learning of seriation, categorization, and polar concepts', and 'arranges learning of space and time'. Such interactions are rated as percentage of appearance and the ratings underlie live observations across different times of the day and days of the week, taking time-dependent variation of interactions into account. The 'ECCOM' (Early Childhood Classroom Observation Measure, Stipek & Byler, 2004) rates the item 'math instruction', also by indicating a percentage of appearance during a three-hour live observation. The 'TBRS' (Teacher Behavior Rating Scale, Landry et al., 2002) rates 'math quality' and 'math quantity' on a scale ranging from 0 (activity not present) to 3 (activity with high quality/activity happened often or many times). However, a clear definition of quality in this context is not further described. ITERS-3 is the only instrument that assesses the quality of math-specific interaction by not indicating the percentage or number of appearances, but by rating on a 7-point scale from (1) inadequate to (7) excellent interactions with anchors and stop-scoring approach.[20]

Overall, regarding the PPCT model, one of the biggest differences across all instruments is the degree of inference in which interactions in the ***process system*** are assessed. Some scales do not have concrete indicators for rating items. Instead, the observer must rate them on an abstract level by summarizing various aspects simultaneously. In contrast, with math-specific interactions, a percentage is often used as a quantity of interactions and activities. That might make it easier to capture these interactions, but on the other hand, it does not necessarily provide insight into their quality. Furthermore, it is noticeable that personal characteristics are less considered in such observations. Some instruments include a subsequent interview where these aspects can be identified. This is probably because most personal traits (e.g., beliefs) are not

[20] Please refer to the 'Results' section in the Appendix – Study 1 (pp. 72–77) for further description.

observable. With structural characteristics, this is different: some instruments explicitly include such aspects and assess their quality while observing (e.g., accessible material). To be more precise, with ECERS/ITERS and its German versions the quality of interaction cannot be measured without assessing the structural quality. In others, the number of teachers and children is noted to consider these structural characteristics when analyzing interactions (e.g., CLASS). Yet, some instruments do not include any structural characteristics at all. The time system is interesting as it is an important component when assessing interactions: most observations occur at a single instance, usually during a multi-hour observation at one day. Only few observations are spread over several days or weeks (in our review observations with one instrument), although there are indications that interactions have time- and day-dependent variations (e.g., Buell et al., 2017; Vitiello et al., 2012). While observations can also be conducted on multiple days with all other instruments, the concept of most instruments does not explicitly indicate it. Nevertheless, many instruments are used to observe changes in interactions over a period of time. It is not clear to what extent the instruments are designed to capture and represent a longer developmental process of interaction quality. Principally, most of them are designed to capture a current state.

To address the first research question, which instruments can be used to assess global and math-specific teacher–toddler interactions, it can be summarized that depending on the type of instrument, a varying number of aspects of the PPCT model are included, which in turn are examined with varying levels of detail. The teacher plays a key role in this process as their interactions with the children are observed. However, the focus is on the entire group, and all teachers and interactions within the group are evaluated. There are other instruments that specifically focus on the child (e.g., inCLASS, Downer et al., 2010), but no instrument specifically focusing on a teacher could be identified. Even though the focus could lie on one teacher during the observation, this is not explicitly intended in the method of the instruments. Moreover, the range of existing instruments does not cover all content areas and age groups to the same extent: math-specific instruments focus primarily on interactions with preschool children. This demonstrates that the choice of instrument must be weighed depending on the goal of the observation, and that there are different selection options depending on the area that should be observed (e.g., preschool vs. toddler classroom, global vs. math-specific interactions).

3.2 Process–Person–Context: Investigating Interactions and Associations with Personal and Structural Characteristics

The structural equation modeling (SEM), revealed an adequate fit, although not perfect: $\chi^2(43) = 81.65$, $p < .001$, CFI = .91, TLI = .88, RMSEA = .10 (CI90: .06–.13), SRMR = .06.[21] When examining emotional and behavioral support (EBS) provided by teachers, it was noted that less support was observed among older teachers ($B = –.02$, $SE = .01$, $\beta = –.26$, $p = .01$) and in mixed-age toddler classrooms ($B = –.35$, $SE = .15$, $\beta = –.25$, $p = .02$). Teachers with higher educational levels tended to provide more engaged support for learning (ESL), but this finding was only marginally significant ($B = .12$, $SE = .07$, $\beta = .16$, $p = .10$). Furthermore, mixed-age classrooms were negatively associated with teachers' ESL ($B = –.33$, $SE = .16$, $\beta = –.26$, $p = .04$). These associations all demonstrated medium effect sizes (Ellis, 2012). Notably, group size showed no significant association with either domain of teacher–toddler interaction quality. A visualization of the SEM can be found in Figure 2 in the Appendix – Study 2.

When examining the relationship between teacher–child interaction quality and structural characteristics, noteworthy findings could be gained. There is a negative association between teacher age and their ability to provide supportive interactions for children's emotions and behavior, suggesting that older teachers may exhibit decreased effectiveness in this regard. This may be attributed to the demanding nature of working in toddler classrooms, which can lead to sustained stress and fatigue. However, age can also reflect professional experience ($r = .69$, $p < .001$ in this study), although it may not be intuitive why interaction quality decreases with increasing experience. Similar reasons (stress, fatigue) could have an influence here. Conversely, our analysis indicates a modest positive influence of teachers' level of education on the quality of engaged support for learning with children. Teachers with higher qualifications demonstrate greater engagement in facilitating diverse opportunities for children's participation and exploration. Their enhanced understanding of child development and broader instructional strategies likely contributes to this. It is important to note that while the association between teacher education and learning support was marginally significant in our model, the standardized nature of German teacher education may limit variance, with

[21] Please refer to the Discussion section 'Limitations and future directions' in the Appendix – Study 2 (pp. 123–124) for a discussion of these parameters.

most teachers reporting to have five years of vocational training (67.0%). Additionally, the composition of children's age in the classroom was a significant factor negatively impacting teacher–toddler interaction quality in both domains. Teachers may encounter challenges in meeting the diverse emotional, behavioral, and learning needs of children across varying developmental stages. Addressing these challenges requires equipping teachers with effective strategies tailored to the dynamic environments of childcare settings to ensure optimal support for children's growth and development.[22]

Considering the PPCT model, it becomes evident that teachers' personal and structural characteristics are associated with the quality of teacher–toddler interactions (***process system***). Teachers' age and education were significant factors of the ***person system***, and children's age composition in the classroom was a significant factor of the ***context system***. However, a crucial aspect of teachers' characteristics was not examined: the force characteristics. It would be interesting to examine how various attitudes, beliefs, or self-efficacy affect the quality in toddler classrooms. Additionally, math-specific interactions were not considered, although the mentioned force characteristics could have a particular impact on these interactions. Moreover, other contexts besides information on the toddler classroom were not contemplated in this analysis, such as the extent to which interaction quality varies across different activity settings.

To address the second research question, what associations exist between personal and structural characteristics and the quality of teacher–toddler interactions, it has been demonstrated that both personal and structural characteristics are essential. Thus, the characteristics of the teacher are of great significance, but they are sometimes not fully assessed and thus cannot be fully examined (force characteristics). Furthermore, it remains unclear what impact the time system has: it can be a crucial factor in assessing the quality of interactions, but is often overlooked (e.g., whether observations are conducted in the morning or afternoon, or during winter or summer). Summing up, there is further potential to investigate associations between interaction quality and other characteristics – beyond classical variables.

[22] Please refer to the 'Results' section in the Appendix – Study 2 (pp. 118–119) for further description.

3.3 Process–Context–Time: Improving Interactions during Free Play and Structured Activities

In terms of global teacher-child interactions during free play, the repeated measures analyses of variance (RMANOVA) showed no significant effects for time, group, or time*group, for neither emotional and behavioral support (EBS) nor engaged support for learning (ESL). However, during structured activities, there were (marginally) significant changes observed over time with small effect sizes for both domains: EBS: $F(1, 81) = 4.17$, $p = .04$, $\eta p^2 = .02$; ESL: $F(1, 81) = 3.24$, $p = .08$, $\eta p^2 = .01$. All teachers, regardless of intervention or control group, exhibited qualitatively higher global interactions with toddlers at posttest compared to pretest. There were no significant group effects or time*group effects concerning EBS and ESL in structured activities. Descriptive statistics of pre- and posttest ratings of teacher–toddler interactions divided by activity settings and groups are presented in Table 2 in the Appendix – Study 3. Table 3 in the Appendix – Study 3 displays the results of the RMANOVA, including significant as well as tested but not significant main and interaction effects. Regarding math-specific interactions, there was a significant increase over time observed during free play, with a medium effect size ($F(1, 75) = 13.12$, $p = .001$, $\eta p^2 = .07$). At posttest, all teachers showed more math-specific interactions during free play compared to the pretest. The group effect was not significant. A marginally significant small interaction effect between time and group was observed during free play ($F(2, 75) = 2.82$, $p = .07$, $\eta p^2 = .03$). Post-hoc tests with Bonferroni adjustment revealed a significant effect only for the math-specific intervention group (MIG; $p < .001$): teachers in the MIG stimulated toddlers more frequently with math-specific content compared to teachers in the other groups (global intervention group and control group). Refer to Figure 2 in the Appendix – Study 3 for the visual representation of the time*group effect of math-specific teacher–toddler interactions during free play. No significant effects for time, group, or time*group were observed for math-specific interactions during structured activities. Table 2 and Table 3 in the Appendix – Study 3 show the pre- and posttest descriptive statistics and all results of the RMANOVA.

These results demonstrate that there is potential to improve teacher–toddler interactions through further training for teachers, particularly math-specific interactions. Nonetheless, in the analyses, there were only minor hints, and it seems difficult for positive effects to unfold. There could be various reasons for this, such as initially high ratings in some domains of interactions, the nature of the chosen observational instrument, or insufficient time for intervention effects to manifest.[23] Furthermore, the methodology of the teacher trainings has some considerable limitations (e.g., treatment of control group).[24]

Concerning the PPCT model, it still becomes evident that when examining the effectiveness of teacher training through the assessment of interactions (***process system***), it is important to consider the ***context system***: a positive impact on math-specific interactions was only observed during free play. This underlines the importance of considering where interactions take place. Therefore, it may also be crucial to incorporate the aspect of specific activity settings into teacher trainings (this was not addressed in this study). A more detailed examination of which teachers specifically require support for successful interactions in which dimension and in which activity setting. Thus, a more tailored offering could be helpful in promoting sustainably high-quality interactions. Regarding the ***time system***, it can be argued that displaying the development of interactions over time is important. However, it remains unclear whether the chosen instrument is suitable for this purpose. While CLASS Toddler captures interactions at a certain level of detail (although with a high degree of inference), it does so at the group level, considering all teachers and children present. In this study, although instructions were given to specifically focus on the participating teacher, it is unclear how well this was actually implemented. An instrument that specifically focuses on the interactions of the target teacher might have been more likely to detect intervention effects. Furthermore, for math-specific interactions, the frequency was observed, which may not necessarily indicate quality. Overall, it would be desirable to use an approach that considers the quality of math-specific teacher–toddler interactions over time.

To address the third research question, if global and math-specific teacher–toddler interactions improve across different activity settings through teacher trainings, it could be revealed that it is possible, however, it is a challenging attempt with many points to consider (e.g., design of the teacher training, selection of the observational instrument). Furthermore, one aspect that has not yet been reflected, but is central when exploring teachers' role in the PPCT model, is the impact of the training on different levels: in this

---

[23] Please refer to the 'Discussion' section in the Appendix – Study 3 (pp. 154–155) for further description.

[24] Please refer to the Discussion section 'Limitations and perspectives' in the Appendix – Study 3 (pp. 155–157) for further description.

study, the impact of the training was examined directly on the level of interactions. However, it would be of interest to investigate the ***person system***. At this level, it would be interesting to examine the force characteristics, as the teacher training may initially influence certain attitudes and beliefs before leading to changes in interactions. To summarize, the teacher should not be viewed as ‘a means to an end’ for successful and high-quality interactions, but rather as an active and individual part of the PPCT model.

4 General Discussion

4.1 Discussion of the Process–Person–Context–Time Model

The presented results provide insight into each of the systems in the PPCT model and their relations to each other, with varying degrees of depth. **The aim of the present thesis was to illustrate the complex interplay of systems (Process–Person–Context–Time) and to examine the role of the teacher within the systems**. It could be shown that the complexity of the systems is high, and that teachers’ role is more or less emphasized depending on which interplay of systems the focus is placed on. For example, the role of the teacher becomes more evident within the Process–Person–Context interplay when examining how teachers’ personal characteristics (in combination with structural characteristics) relate to the quality of teacher–toddler interactions. It is less clear when investigating how interaction quality could be enhanced through training within the Process–Context–Time interplay, even though the teacher has an integral role. This will be addressed in the subsequent discussion. Since the teacher is the focus of investigation, the person system will be discussed first and separately. Given the emphasis on the interplay, the discussion will then come to Process–Person–Context–Time as a whole.[25]

4.1.1 Person System

Many research findings underline the importance of the quality of processes – especially of interactions – in early childhood education and care (ECEC) (e.g., Ulferts et al., 2019). In this context, the teacher is highlighted as an important figure, shaping these

[25] Please refer to the ‘Discussion’ sections of each individual study in the Appendix – Study 1 (pp. 77–80), Appendix – Study 2 (pp. 119–124), and Appendix – Study 3 (pp. 154–157) for further discussion.

interactions: they are responsible for determining which activities are undertaken with toddlers, what materials are used for interactions, and how the toddler's day is structured. They also facilitate interactions among toddler's peer groups. Despite these numerous crucial responsibilities, the teacher remains relatively unrecognized in the PPCT model, and is seen as one aspect among many in the micro-system. However, it seems to be unseen that teachers also represent their own system, with particular characteristics, experiences, or changes that all influence their interactions with toddlers.

Here, a closer examination of teacher's personal characteristics is interesting: the demand, resource, and force characteristics are not depicted in other representations of the PPCT models but have been added in the own PPCT model illustration (see Figure 1 in the Introduction, p. 7). Teacher characteristics include, among others, teachers' age (demand) and teachers' education (resource), both of which were significant variables for interaction quality in this study. Here, force characteristics remained disregarded, for example, how important math-specific experiences are considered for toddlers and what motivation the teacher has to enable such experiences. This misses important aspects, which could significantly change over time (e.g., through further training, but also otherwise; Chen et al., 2014). These aspects, namely motivation, self-efficacy, and beliefs, are crucial for shaping and ensuring the quality of global and math-specific interactions (Besser et al., 2024; Guo et al., 2010; Wieduwilt et al., 2023). Thus, they are a part of the PPCT model that is important in relation to the teacher but have not been considered within the analyses of this thesis. Another aspect concerning these personal characteristics, which the PPCT model (implicitly) incorporates, is that they appear twice in the model: first as the teachers' characteristics at the present moment, influencing proximal processes, and then as a developed outcome, either targeted directly or indirectly – this thesis only took the first into account. The difficulty that arises from this is that there is no concrete measure of teachers' development. Instead, interactions have been examined, which in turn could already be seen as an output, thus as one step further. Even though these interactions directly influence the child and are therefore crucial to examine, it would be important to investigate the teacher 'at the other end' more closely.

One way to do so is to include *force characteristics* and improve as well as evaluate them over time (e.g., Çiftçi & Topçu, 2022; Suchodoletz et al., 2018). That would be feasible for the present study, as this information is available from the teacher questionnaires. Another approach could be to take a closer look at *teachers' competencies* *(resource characteristics)*, which was not done in this study. This component is often only covered through the indication of teachers' education, such as in this study. However, it encompasses much more, as described in the following (Anders & Oppermann, 2024).

Professional competence of pedagogical action can be distinguished into three facets: individual factors (e.g., professional knowledge), situation-related skills (e.g., action planning), and performance (carrying out the action) (Fröhlich-Gildhoff et al., 2011; Gasteiger & Benz, 2018). Going into more detail, the professional knowledge is a critical part of this concepts. It consists of *content knowledge (CK)*, *pedagogical knowledge (PK)*, and *pedagogical content knowledge (PCK)* (Blömeke & Delaney, 2012; Shulman, 1987). In this context, CK refers to teachers' knowledge of content in a specific area (e.g., knowledge about mathematics in early childhood), and PK refers to teachers' knowledge about toddlers' development and pedagogical support for them (e.g., knowledge about emotion regulation and cognitive stimulation). PCK integrates these two aspects and encompasses teacher's knowledge of how to make the specific content accessible to toddlers (knowledge about strategies for promoting competencies, e.g., through math talk). The way in which this knowledge is developed in turn influences how action can be planned and implemented. Indeed, studies could demonstrate that teachers' math-related actions are associated with pedagogical content knowledge (Dunekacke et al., 2016; McCray & Chen, 2012; Oppermann et al., 2016) – but other studies have shown that it can be challenging to enhance this knowledge through interventions and that an improvement does not necessarily last a longer period (Goldschmidt & Phelps, 2010). Findings on the effectiveness of interventions aimed at increasing professional knowledge are predominantly conducted with teachers of older children (e.g., Evens et al., 2015; Weitzel & Blank, 2020). There is no comprehensive understanding of how this takes place in toddler classrooms. It remains uncertain how teachers' knowledge can be successfully enhanced in this environment and if it necessarily leads to an increased interaction quality, as the pedagogical work with toddlers is much more dynamic. To note, there are also additional conceptualizations of professional competence and knowledge that are not listed here (e.g., Guerrero & los Ríos, 2012). They as well as the presented concepts should be included in further considerations and examinations as this enables to understand requirements for successful interactions.

Overall, professional competencies and knowledge, motivation, self-efficacy, and beliefs can be seen as preconditions for high-quality teacher–toddler interactions. Besides other aspects of the teacher that are essential as well (e.g., age and professional experience), especially these should be considered in the context of progressing time. A closer examination of these parameters makes it possible to depict the relations of various factors of the teacher to effective interactions. All of this is present in the PPCT model, but needs to be explicitly addressed and investigated, giving the teacher an active role of shaping interactions in the complex systems of Process–Person–Context–Time.

4.1.2 Interplay of Process–Person–Context–Time

When conducting research based on the PPCT model, it is suggested to include at least two of the systems with at least two of their components (except for process, which does not have sub-components; Tudge et al., 2009; Xia et al., 2020). This means, in terms of measurement instruments – at first glance – it would be most suitable to use ITERS-3 (Harms et al., 2017): it considers not only global but also math-specific aspects and incorporates not only interactions but also structural characteristics. However, these advantages are also offset by disadvantages, as previously described (e.g., mixing of domains, stop-scoring approach). Interactions as part of the ***process system*** are not as differentiated as in other instruments, which might cause a bias if they are the main focus of investigation. Analyses cannot be presented in such a distinguished manner, and implications of the results may remain unclear. An instrument such as CLASS Toddler (La Paro et al., 2012) provides a deeper insight into interactions, but, on the other hand, it does not cover as many systems of PPCT and has other drawbacks (e.g., high degree of inference). It becomes evident: there are many choices, and it is a matter of weighing the options carefully. By using an instrument such as CLASS Toddler, which attempts to precisely assess interaction quality (although it is not clear to what extent it actually does this), one might hope to be able to see changes in these interactions more accurately. However, in the present study, intervention effects for global interactions could not be revealed. Reasons for this could be the study design and the conception of the teacher trainings, as already briefly discussed.[26] However, if the training would have had an

[26] Please refer to the 'Discussion' section in the Appendix – Study 3 (pp. 154–157) for further description.

effect, what could be the reason that it cannot be measured with the chosen instrument?

The answer to this question is difficult, but the following points are briefly outlined. First, CLASS Toddler is designed to be both an assessment instrument and a training tool (Hamre et al., 2012; Rodriguez & Garza, 2014). Also in the present study, the CLASS Toddler observation sheet was used to design self-observation sheets for teachers. While these do not incorporate the identical dimensions and are additionally supplemented with domain-specific aspects, they are still closely aligned with CLASS Toddler. The same instrument is therefore used to assess the current quality status (process in micro-time), to achieve improvement goals (training), and to assess changes in quality (process in meso-time). It remains unclear whether this blending is effective (Naumann et al., 2016; Naumann et al., 2019), even though it seems that it should lead to teachers receiving better ratings in the desired quality dimensions. Again, insights on this are more prevalent in classrooms with older children. Second, CLASS Toddler claims to be able to assess a latent construct such as quality objectively, reliably, and validly (La Paro et al., 2012). However, empirical findings suggest that this may not entirely be the case, and the CLASS Toddler authors have been criticized to provide limited published research with this information (Lopez Boo et al., 2019; Rodriguez & Garza, 2014). Even though Cronbach's alpha as a measure of internal consistency was good in the present study (pretest: .90 for EBS, .83 for ESL; posttest: .87 for EBS, .84 for ESL), the confirmatory factor analysis (CFA) as a measure of construct validity revealed some difficulties[27] and other criteria were not investigated in more detail. A closer examination might expose weaknesses that, in turn, could affect the interpretability of the results (Thorpe et al., 2022). Third, CLASS Toddler is a macro-analytic instrument, because it assesses interaction quality as an aggregated score across the entire classroom, including all present teachers and children. This allows for a resource-effective assessment of interaction quality in terms of time and costs. However, this method does not necessarily allow for detailed statements about the interactions of a particular teacher with toddlers.

The last point is here taken up in order to discuss it in more depth. The criticism of standard instruments, capturing interactions in a macro-analytical manner and thus not encompassing all aspects of quality, leads to the consideration of how missing aspects

[27] Please refer to the Results section 'Domains of teacher–toddler interaction quality' (pp. 118–119) and to the Discussion section 'Limitations and future directions' in the Appendix – Study 2 (pp. 123–124) for further description.

could be covered (Cents-Boonstra et al., 2022; Hestenes et al., 2004). One approach for this are *micro-analytical assessments*: they often concentrate on specific aspects of interactions through time- or event-sampling approaches and can be either quantitative or qualitative (Fassnacht, 1995). Quantitative methods typically seek to answer the question of what happens during interactions (e.g., whether math-specific interactions occur, for example, by counting their frequency). Qualitative methods address the question of how these interactions are shaped (e.g., how the teacher addresses math-specific content, for example, through math talk). Usually, a selected sequence of interactions from a small sample is analyzed. This analysis is detailed and can be technically mediated, for example, by recording interactions on video and rating them afterwards (Greve & Wentura, 1997). The EarlyMath project has the possibility to conduct such a micro-analysis as video recordings were conducted. In the context of the present thesis, if mathematical content occurred, math-specific interactions were quantified by counting their frequency. However, this approach was still macro-analytic since these interactions were not specifically looked at but rated during the CLASS Toddler observation. Re-analyzing the video-recorded interactions through a micro-analytic method, both globally and math-specifically, would be beneficial for describing them in more detail (Haber et al., 2021). This could make it possible to reveal intervention effects over time that were previously unnoticed. An example for a mico-analytic event-sampling approach is the feedback category system of Aumann et al. (2024). In their study, the authors video-recorded one-to-one play-based situation and analyzed teachers' feedback toward children. They coded whether it was specific or non-specific and whether it referred to general, personal, or process-related aspects, or to outcomes. This approach seems promising for gaining deeper insights into interactions (Aumann et al., 2023). However, these results concern older children and in this age group, findings showed that achieving sufficient inter-rater agreement with mico-analytic methods is challenging (Wadepohl & Mackowiak, 2013). One reason could be that it is difficult to establish clear coding rules, because micro-analytic observations are very detailed and assess a huge number of situations closely (Kucharz & Mackowiak, 2014). Considering the advantages and disadvantages of macro- and micro-analytic instruments, a good approach might be to integrate both procedures in a *mixed-methods approach* (e.g., Ertesvåg et al., 2022; König, 2009). In doing so, inference quantitative observations can be supplemented with quantitative and qualitative in-depth coding, which allows for a comprehensive assessment of teacher–toddler interactions.

While interactions between teachers and children may be well described with this approach, teachers' ***person system*** is not represented. As already described, the competence and knowledge of teachers could be incorporated to give more weight to this system of the PPCT model. Also not represented are important parts of the ***context system***, such as cultural aspects in the macro-system (Pagani, 2019). There is increasing criticism that CLASS Toddler was developed in a US-context and that it is challenging to transfer certain underlying assumptions to other cultural contexts (e.g., teachers in the USA are more instructional compared to teachers in other countries, but the certification process is based on videos recorded in US-classrooms; Ishimine & Tayler, 2014). Using the same instrument in different countries makes it seemingly easier to compare results, but it can lead to pitfalls if the complexity of cross-cultural aspects is not considered (Vandenbroeck & Peeters, 2014). For example, in Germany, due to the strong socio-pedagogical orientation, free play has a significant role, which is not considered in CLASS Toddler. If the teacher takes a step back – for example, to actively observe a child during free play – this might be unrecognized or even rated negatively in the CLASS Toddler observation, whereas it is positively connotated in German classrooms (Faas et al., 2017; Vogt et al., 2020; Weiser, 2022). Perhaps cultural differences in the pedagogical practice could have also led to intervention effects being found only in a specific context (in free play) in the present study. Even instruments with a robust theoretical and empirical background and widespread international use cannot be considered culture neutral, as they reflect the cultural values of their origin. Despite its significance, this issue has received limited attention so far. Only one study by Pastori and Pagani (2017) reflected this topic in toddler classrooms. They collected video-data from early childcare centers in Reggio Emilia and Milan, Italy, rated them with CLASS Toddler and additionally conducted a qualitative analysis with an *ethnographic approach*. Together with the teachers, they discussed cultural critical aspects of CLASS Toddler and reflected them with pedagogical coordinators and national experts. This way, they spotted missing elements and differences: approaches for learning, pre-existing relational histories with children, and the significance of provided activities are not included or weighted differently in CLASS Toddler. Moreover, co-constructed observations by observers and teachers are not possible, but participatory approaches of evaluations *with* teachers rather than *of* teachers seem important in the Italian context. Ethnographic approaches that closely examine CLASS Toddler in the German context could be helpful in gaining better insights into the extent to which the instrument is suitable for use in German toddler classrooms, or which additional points need to be considered. Furthermore, the ethnographic approach could help in determining how an instrument (independent of CLASS Toddler) would need to be designed to assess global and math-specific teacher–toddler interaction quality and to integrate important aspects of the PPCT model within the German context.

It becomes evident that the consideration of how interaction quality can be adequately addressed in the complex interplay of Process–Person–Context–Time is substantial. Reflecting the ***time system*** in this regard is perhaps the most challenging aspect. The consideration of micro- and meso-time is comparably easier, although challenges still appear (e.g., quality variations within these time systems, combining these time systems; Buell et al., 2017; Suchodoletz et al., 2014; Thorpe et al., 2020; Vitiello et al., 2012).[28] However, little attention is given to ***macro-time*** – neither in this thesis, nor in other empirical research. This may be because macro-time encompasses historical or personal life events and societal changes that are attributed to time. When assessing temporal aspects, it might not be clear what is such an event or what is the uniqueness about the given time. Hence, integrating such perspectives into research is challenging, as these events might not be regognized, and historical aspects can often only be incorporated retrospectively. However, during the EarlyMath project, there was a time-changing event that quickly became apparent: the COVID-19 pandemic. This became evident – alongside many other restrictions in various areas – in ECEC primarily through the closure of early childcare centers or the offering of emergency care due to the pandemic situation (Loss et al., 2021). The EarlyMath project was affected as well, since data collection had to be postponed, and even when visits to the childcare settings were possible, the classrooms may have been composed differently than usual, or certain safety precautions such as wearing masks were required. In terms of the teacher trainings, the most impactful action was certainly to conduct it virtually, although there are few insights if such trainings are effective (Baumgartner et al., 2022; Locasale-Crouch et al., 2023).

[28] Please refer to the Discussion sections 'Limitations and future directions' in the Appendix – Study 2 (pp. 123–124) and 'Limitations and perspectives' in the Appendix – Study 3 (pp. 155–157) for further description.

Macro-time events are often unpredictable, and thus situation dependent actions must be taken. In the context of the EarlyMath project, while the questionnaires were adapted and both, parents and teachers, were asked COVID-19 related questions, these aspects were not examined within the scope of this thesis. In addition to the point that they should be investigated more closely in relation to interaction quality, it is important to consider what impact this historical and personal life event had within the PPCT model: it had affected the micro-system, as interactions in ECEC settings were not possible as usual (Diefenbacher et al., 2023). It had affected the meso-system, as teachers and parents could no longer exchange information as before (Grgic et al., 2022). It had affected the eco-system, as some parents were required to work from home (Bernhardt et al., 2022). And it had affected the macro-system, as values and norms have changed (Ratten, 2022). It may have also affected personal characteristics, and teachers' might have acted differently, because of changed beliefs or increased stress (Steigleder et al., 2023). Even though these are just few examples of a complex issue, it is obvious that this macro-time event of the chrono-system affected every other system and processes at every level. More in-depth considerations are needed to understand what this means for the results of the present thesis.

The overall PPCT model appears almost too complex to isolate specific aspects, as everything is interconnected and important for the holistic framework. Tobin (2005) sums this up in his work: early childhood pedagogies, ECEC settings, and teacher–child interactions *"are not universal or culture free but instead are reflections of values and concerns of particular people in a particular time and place"* (Tobin, 2005, p. 426). This should not mean that the complex PPCT model cannot be applied in empirical research: it depends on the outcome of interest and how the relevant systems are operationalized and assessed to achieve the research aim. Based on the suggestion to primarily concentrate on proximal processes and include two systems with two components to reflect the interplay (Tudge et al., 2009; Xia et al., 2020), the present thesis has fulfilled this requirement. Teacher–toddler interactions were observed and analyzed over time, incorporating structural and personnel characteristics. Especially, since it is not possible to examine every system with all components within the scope of a study, it is important to discuss which systems were not included and what that might mean for further directions and implications.

4.2 Limitations and Further Directions

In respect to the studies presented in this thesis, the limitations and further directions are discussed in detail in the Appendix.[29] Moreover, some points were already briefly mentioned, which will not be repeated here: for example, limitations regarding the study design, the conception and conduction of the teacher training, the chosen observational instrument, as well as aspects that were not included in the present thesis, but should be analyzed in the future. Instead, limitations and directions will be discussed on a broader scale. Referring back to quality in ECEC, which was introduced in the Theoretical Framework of this thesis (pp. 6–10), different perspectives on quality can be distinguished (Katz, 1993): top-down perspectives (structural conditions), bottom-up perspectives (children's experiences), outside-insight perspectives (parents' experiences), insight perspectives (teachers' experiences), and outside perspectives (societal conditions). Since the teacher is the main focus of investigation in the present thesis, only the *insight perspective* will be elaborated in the following. It is a considerable restriction that teacher–toddler interaction quality was examined, and teacher trainings were designed to improve it, but the perspective of teachers was not included. Findings of other studies have indeed shown that experiences teachers make within the ECEC system can influence their interactions – and even their intention to stay in the profession (Cassidy et al., 2017; Mischo et al., 2022; Torquati et al., 2007). These experiences are made within colleague relationships, teacher–parent relationships, and teacher–employer relationships (Katz, 1993). In terms of the PPCT model, these aspects can be situated in the meso- and exo-system. How could they have been incorporated in the present thesis?

Colleague relationships and the atmosphere within the team is an aspect that is already considered in some publications following the structure–process–orientation model described in the Introduction (pp. 8–9; Eling, 2022; Kluczniok & Roßbach, 2014). The functionality of a team significantly depends on the quality of social relationships among teachers in the workplace (Viernickel et al., 2017). This encompasses the perception of team members regarding the social environment, such as communication within the team, which shapes the team climate. Typically, these aspects are assessed

[29] Please refer to the 'Discussion' sections of each individual study in the Appendix – Study 1 (p. 80), Appendix – Study 2 (pp. 123–124), and Appendix – Study 3 (pp. 155–157) for further discussion of the limitations and further directions.

through questionnaires or interviews (Brodbeck et al., 2000). This approach could have been followed in the EarlyMath project to understand more about impacts of team climate on interaction quality and the effectiveness of teacher trainings. Findings emphasize the importance of team climate for teachers' well-being, including physical and mental health (Trauernicht et al., 2022). This in turn influences how well teachers can shape their interactions (Resa et al., 2017; Wertfein et al., 2013). Since exhaustion is associated with longer periods of illness or fluctuation (Rentzou, 2012; Schaack et al., 2020), a positive team climate in a supportive environment also contributes to the functionality of the whole early childcare center.

Good *teacher–parent relationships* are seen as characteristic that promotes positive child development: alongside the institutional setting, the home learning environment is essential for global and math-specific child development (Burghardt et al., 2020; Daucourt et al., 2021; Miller et al., 2023; Salminen et al., 2021). It is particularly important for teachers and parents to exchange information about the child's developmental steps, events, and information through their good relationship with each other. This is also highlighted by the structure–process–orientation model (Kluczniok & Roßbach, 2014). This way, synergies can be utilized to support the full potential for child development (Cohen & Anders, 2020; Powell et al., 2010). Informal conversations, developmental discussions, and parent information evenings provide a good context for strengthening teacher–parent relationships. Their quality lies in the provision of a diverse range of offers as well as the suitability of these offers to the needs of parents. In this context, accessibility of teachers for parents and active listening are especially valuable (Fröhlich-Gildhoff et al., 2006; Viernickel et al., 2017). In addition to such general quality criteria, specific quality criteria must be defined for each individual offer (Friederich, 2011; Textor & Blank, 2004). However, definitions of good quality criteria are still lacking in some areas (e.g., for developmental discussions about mathematical development of toddlers). The quality of such offers and teacher–parent relationships are usually assessed through questionnaires, but there are recent approaches to observe the quality of teacher–parent interactions during child drop-off (Hummel et al., 2022). In the EarlyMath project, both teachers and parents were asked questions about their collaboration in terms of global and math-specific topics, but interactions have not been observed. These aspects were also not evaluated within the scope of this thesis and including this information in future analyses would therefore be an interesting approach.

Teacher–employer relationships may seem to have little to do with teacher–child interactions or child development at first glance. However, the extent to which teachers are satisfied with their working conditions can largely impact their pedagogical work (Schreyer et al., 2015; Viernickel et al., 2017). The satisfaction depends on several factors: on leadership, on quality management, and on implementation quality (Anders & Oppermann, 2024). Here, not only the employer is considered, but also the leading staff. In addition to administrative and organizational tasks, a childcare center leader undertakes pedagogical leadership (Anders et al., 2021). On the one hand, how much importance the leader places on maintaining and developing quality is an essential aspect. On the other hand, it is also substantial to what extent s/he involves teachers in leadership decisions and duties: depending on the style of leadership, it can be handled differently how much responsibility teachers have for specific tasks (Ballaschk et al., 2017). Additionally, the extent to which leaders create opportunities for teacher' further development is important (Viernickel et al., 2017). Moreover, effective quality management (QM) in the childcare center aims to continuously evaluate quality (see QM systems in the Introduction, p. 7). In such QM systems, ways of working, structural conditions, and processes are jointly reviewed by employers, leaders, and teachers. Despite optimizing the intern management, this should make QM aspects transparent to outside groups (Amerein & Amerein, 2011; Pohl, 2013). Successful QM can in turn ensure the satisfaction of teachers and create a good environment for positive interactions. For this, the quality of implementation is crucial. If aspects are identified that hinder quality, it is essential to what extent counteracting attempts are implemented. In addition, the implementation must be continuously revised. Indicators for this include, among others, the integration and development of quality aspects in pedagogical concepts and the constant dissemination of important aspects to the team (Anders & Oppermann, 2024). Similar to teacher–teacher and teacher–parent relationships, the teacher–employer relationship is generally characterized by good communication and participative processes, in turn influencing teachers' satisfaction. Teacher–employer relationships are not included in the EarlyMath project, and its consideration demands effort: this topic can be strongly influenced by political action (e.g., 'KiTa-Qualitätsgesetzt', see Introduction, p. 8), and integrating it into a project may require revising the study design.

The description of limitations and further directions shows that relationships at various levels are important for teachers' interactions with toddlers. For example, the observed interaction quality could be affected if teachers just had a conflict with other teachers and cannot engage as usual due to increased stress. Or teachers may need to prepare a meeting with parents, for which they should ideally have additional time, but this task might still distract them. Or teachers may be dissatisfied with their opportunities for professional development and seek for better perspectives elsewhere, thus not involving them as much anymore. Once again, it becomes evident that further aspects influencing teacher–toddler interactions cannot be discussed without having the PPCT model as theoretical framework in mind. All the mentioned aspects are already incorporated in Figure 1 in the Introduction (p. 7) but were not included in the present thesis.

4.3 Implications

The implication of each individual study is presented in the Appendix.[30] Overall, the detailed examination of the PPCT model reveals that it has general implications for research, practice, and policy. In early childhood research, Bronfenbrenner's theory is often used as the basis for study design. However, it is rarely discussed which aspects under investigation belong to which system. As a result, the gaps – the systems that have not been considered – remain unrecognized. In empirical research, it is not critical not to include all systems (following the recommendation of Tudge et al., 2009 and Xia et al., 2020), but it is critical not to reflect on its impact on study results and implications. When examining the quality of childcare centers and teacher–child interactions, it becomes apparent that the role of the teacher is sometimes not actively considered, even though it is an important part of the PPCT model. This is also important for practice: teachers should be given the best possible support in their multifunctional role to facilitate good interactions with children. Teacher trainings as one way to achieve this were laid out in the present thesis. Further possibilities, such as positive team climate and job satisfaction, were briefly mentioned but were not the focus of investigation. Politicians should intervene here: they can create suitable conditions (e.g., good teacher–child ratio and equipment), but also actively create incentives to improve quality through legislation

[30] Please refer to the 'Discussion' sections of each individual study in the Appendix – Study 1 (pp. 78–80), Appendix – Study 2 (pp. 120–123), and Appendix – Study 3 (p. 157) for further discussion of the implications.

(e.g., funding for certain subjects). In this context, it is important to consider the teacher and to cover as many systems of the PPCT model as possible. Especially in a federal country like Germany, where conditions may vary greatly between states, regions, and even centers, depending on the autonomy given to them, this could contribute to ensure good conditions for high quality in all ECEC settings.

5 Conclusion

The PPCT model is a comprehensive theory with its three types of person characteristics (demand, resource, and force), four types of contexts (micro-, meso-, eco-, and macro-system), and three types of time (micro-, meso-, and macro-time), which each play a role in shaping processes. Through this interplay, human development is influenced, and the model can be applied to various outcomes at different developmental stages. Consequently, it is widely used in research across different fields. Hence, it cannot be determined which specific aspects of the systems need to be included in a study to reference this theory. This would be possible with another theory that focuses on a particular aspect of interest. However, regarding the PPCT model, the complex interplay between systems is crucial and the center of investigation. Representing this complexity in a study is challenging but can be achieved by theoretically reflecting on the systems' interplay and empirically operationalizing the variables important for the outcome of interest. In the present thesis, the focus was placed on the impact of different systems on the development of teachers' interactions.[31] Various systems and their interplay were included and analyzed, and missing systems and how they could have been incorporated were discussed. Thereby, insights and findings about proximal processes (teacher–toddler interactions) and teachers' development through teacher trainings were generated using the PPCT model. To summarize, the PPCT model provides researchers with a shared framework, but it must be properly utilized by reflecting on each system and – more importantly – by reflecting on their interplay and on the potential impact of systems that are not included.

[31] Each individual study has its own conclusion. Please refer to the 'Conclusion' sections in the Appendix – Study 1 (pp. 80–81), Appendix – Study 2 (p. 125), and Appendix – Study 3 (pp. 157–158).

References

Amerein, B., & Amerein, K. (2011). *Qualitätsmanagement in Arbeitsfeldern der Frühen Bildung*. Braunschweig: Bildungsverlag EINS.

Anders, Y. (2013). Stichwort: Auswirkungen frühkindlicher institutioneller Betreuung und Bildung. *Zeitschrift Für Erziehungswissenschaft*, *16*(2), 237–275. https://doi.org/10.1007/s11618-013-0357-5

Anders, Y., Daniel, H.-D., Hannover, B., Köller, O., Lenzen, D., McElvany, N., Seidel, T., Tippelt, R., Wilbers, K., & Wößmann, L. (Eds.). (2021). *Aktionsrat Bildung*. *Führung, Leitung, Governance: Verantwortung im Bildungssystem: Gutachten*. Münster: Waxmann. https://doi.org/10.31244/9783830994008

Anders, Y., & Oppermann, E. (2024). Frühpädagogische Qualität in Kindertageseinrichtungen: Eine Erweiterung des Struktur-Prozess-Modells. *Zeitschrift für Erziehungswissenschaft*, 1–27. https://doi.org/10.1007/s11618-024-01218-7

Anders, Y., & Roßbach, H.-G. (2020). Empirische Bildungsforschung zu Auswirkungen frühkindlicher, institutioneller Bildung: Internationale und nationale Ergebnisse. In R. Braches-Chyrek, C. Röhner, H. Sünker, & M. Hopf (Eds.), *Handbuch Frühe Kindheit* (pp. 341–356). Leverkusen: Barbara Budrich.

Arnett, J. (1989). *Caregiver Interaction Scale.* Chapel Hill, NC: University of North Carolina at Chapel Hill.

Aumann, L., Gasteiger, H., & Puca, R. M. (2024). Early Childhood Teachers' Feedback in Natural Mathematical Learning Situations: Development and Validation of a Detailed Category System. *Acta Psychologica*, *244*, 104175. https://doi.org/10.1016/j.actpsy.2024.104175

Aumann, L., Graf-König, N., Puca, R. M., & Gasteiger, H. (2023). Feedbackverhalten frühpädagogischer Fachkräfte in mathematischen Spielsituationen – Ergebnisse einer clusteranalytischen Typisierung. *Zeitschrift für Grundschulforschung*, *16*(2), 321–338. https://doi.org/10.1007/s42278-023-00170-5

Autor:innengruppe Bildungsberichterstattung. (2022). *Bildung in Deutschland 2022: Ein indikatorengestützter Bericht mit einer Analyse zum Bildungspersonal*. *Bildung in Deutschland*. Bielefeld: wbv. https://doi.org/10.3278/6001820hw

Ballaschk, I., Anders, Y., & Flick, U. (2017). Führung als Thema deutscher Kindertageseinrichtungen. *Zeitschrift für Erziehungswissenschaft*, *20*(4), 670–689. https://doi.org/10.1007/s11618-017-0761-3

Barros, S., & Aguiar, C. (2010). Assessing the Quality of Portuguese Child Care Programs for Toddlers. *Early Childhood Research Quarterly*, *25*(4), 527–535. https://doi.org/10.1016/j.ecresq.2009.12.003

Baumgartner, E., Kaplan-Rakowski, R., Ferdig, R. E., Hartshorne, R., & Mouza, C. (Eds.). (2022). *A Retrospective of Teaching, Technology, and Teacher Education during the COVID-19 Pandemic*. Waynesville, NC: Association for the Advancement of Computing in Education.

Bentler, P. M., & Bonett, D. G. (1980). Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Benz, C., Peter-Koop, A., & Grüßing, M. (Eds.). (2015). *Frühe mathematische Bildung: Mathematiklernen der Drei- bis Achtjährigen*. *Mathematik Primarstufe und Sekundarstufe I + II Didaktik der Mathematik*. Berlin: Springer Spektrum. https://doi.org/10.1007/978-3-8274-2633-8

Berk, L. E., & Winsler, A. (1995). *Scaffolding Children's Learning: Vygotsky and Early Childhood Education. Naeyc Research into Practice Series. Volume 7*. Washington, DC: National Association for the Education of Young Children.

Bernhardt, J., Recksiedler, C., & Linberg, A. (2022). Work From Home and Parenting: Examining the Role of Work-Family Conflict and Gender during the COVID-19 Pandemic. *The Journal of Social Issues*, *79*, 935–970. https://doi.org/10.1111/josi.12509

Berth, F., & Grgic, M. (2020). Wie kam die Bildung in die Krippe? Frühe Kindertagesbetreuung im Spiegel von Wissenschaften, Recht und individuellen Einstellungen in Westdeutschland seit den 1960er-Jahren. In I. van Ackeren, H. Bremer, F. Kessl, & H.-C. Koller (Eds.), *Schriften der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE). Bewegungen: Beiträge zum 26. Kongress der Deutschen Gesellschaft für Erziehungswissenschaft* (pp. 447–460). Leverkusen: Barbara Budrich.

Besser, N., Kaiser, C., Dornheim, D., & Lehrl, S. (2024). Die Bedeutung von Selbstwirksamkeitserwartungen frühpädagogischer Fachkräfte für die Qualität unterstützender Fachkraft-Kind-Interaktionen und die Häufigkeit mathematischer Aktivitäten. *Frühe Bildung*, *13*(1), 20–28. https://doi.org/10.1026/2191-9186/a000646

Bichay-Awadalla, K., & Bulotsky-Shearer, R. J. (2021). Examining the Factor Structure of the Classroom Assessment Scoring System Toddler (CLASS-T) in Early Head Start and Subsidized Child Care Classrooms. *Early Education and Development*, 1–17. https://doi.org/10.1080/10409289.2021.1887700

Biringen, Z., Altenhofen, S., Aberle, J., Baker, M., Brosal, A., Bennett, S., Coker, E., Lee, C., Meyer, B., Moorlag, A., & Swaim, R. (2012). Emotional Availability, Attachment, and Intervention in Center-Based Child Care for Infants and Toddlers. *Development and Psychopathology*, *24*(1), 23–34. https://doi.org/10.1017/S0954579411000630

Bleses, D., Jensen, P., Slot, P. L., & Justice, L. (2020). Low-Cost Teacher-Implemented Intervention Improves Toddlers' Language and Math Skills. *Early Childhood Research Quarterly*, *53*, 64–76. https://doi.org/10.1016/j.ecresq.2020.03.001

Blömeke, S., & Delaney, S. (2012). Assessment of Teacher Knowledge across Countries: A Review of the State of Research. *ZDM – Mathematics Education*, *44*(3), 223–247. https://doi.org/10.1007/s11858-012-0429-7

Blossfeld, H.-P., Schneider, T., & Doll, J. (2009). Methodological Advantages of Panel Studies. Designing the New National Educational Panel Study (NEPS) in Germany. *Journal for Educational Research Online*, *1*(1), 10–32.

BMFSFJ. (2021). *Kindertagesbetreuung Kompakt: Ausbaustand und Bedarf 2020*. Berlin: Referat Öffentlichkeitsarbeit des Bundesministeriums für Familie, Senioren, Frauen und Jugend.

Brodbeck, F. C., Anderson, N., & West, M. (2000). *Das Teamklima-Inventar*. Heidelberg: Hogrefe.

Bronfenbrenner, U., & Evans, G. W. (2000). Developmental Science in the 21st Century: Emerging Questions, Theoretical Models, Research Designs and Empirical Findings. *Social Development*, *9*(1), 115–125.

Bronfenbrenner, U., & Morris, P. A. (1998). The Ecology of Developmental Processes. In W. Damon, & R. M. Lerner (Eds.), *Handbook of Child Psychology: Theoretical Models of Human Development* (pp. 993–1028). Hoboken, NJ: John Wiley & Sons.

Bronfenbrenner, U., & Morris, P. A. (2006). The Bioecological Model of Human Development. *Handbook of Child Psychology*, *1*. Hoboken, NJ: John Wiley & Sons.

Bryant, D. M., Burchinal, M., & Zaslow, M. (2011). Empirical Approaches to Strengthening the Measurement of Quality: Issues in the Development and Use of Quality Measures in Research and Applied Settings. *Quality Measurement in Early Childhood Settings*, 33–47.

Bücklein, C., Hoffer, R., & Strohmer, J. (2017). Measuring the Quality of Interactions in Early Childhood Education and Care Settings for Toddlers – An Exploratory Comparison of the Observation Tools GInA and CLASS Toddler. In H. Wadepohl, K. Mackowiak, K. Froehlich-Gildhoff, & D. Weltzien (Eds.), *Psychologie in Bildung und Erziehung. Interaktionsgestaltung in Familie und Kindertagesbetreuung* (pp. 83–114). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-10276-0_4

Buell, M., Han, M., & Vukelich, C. (2017). Factors Affecting Variance in Classroom Assessment Scoring System Scores: Season, Context, and Classroom Composition. *Early Child Development and Care*, *187*(11), 1635–1648. https://doi.org/10.1080/03004430.2016.1178245

Burchinal, M., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver Training and Classroom Quality in Child Care Centers. *Applied Developmental Science*, *6*(1), 2–11. https://doi.org/10.1207/S1532480XADS0601_01

Burchinal, M., McCartney, K., Steinberg, L., Crosnoe, R., Friedman, S. L., McLoyd, V., & Pianta, R. C. (2011). Examining the Black-White Achievement Gap among Low-Income Children using the NICHD Study of Early Child Care and Youth Development. *Child Development*, *82*(5), 1404–1420. https://doi.org/10.1111/j.1467-8624.2011.01620.x

Burghardt, L., Linberg, A., Lehrl, S., & Konrad-Ristau, K. (2020). The Relevance of the Early Years Home and Institutional Learning Environments for Early Mathematical Competencies. *Journal for Educational Research Online, 12*(3), 103–125. https://doi.org/10.25656/01:21188

Cabell, S. Q., DeCoster, J., Locasale-Crouch, J., Hamre, B. K., & Pianta, R. C. (2013). Variation in the Effectiveness of Instructional Interactions across Preschool Classroom Settings and Learning Activities. *Early Childhood Research Quarterly*, *28*(4), 820–830. https://doi.org/10.1016/j.ecresq.2013.07.007

Cadima, J., Aguiar, C., Guedes, C., Wysłowska, O., Salminen, J., Slot, P. L., Barata, M. C., & Lerkkanen, M.-K. (2022). Process Quality in Toddler Classrooms in Four European Countries. *Early Education and Development*, *34*(7), 1565–1589. https://doi.org/10.1080/10409289.2022.2139548

Cadima, J., Nata, G., Barros, S., Coelho, V., & Barata, C. (2020). *OECD Education Working Papers No. 243: Literature Review on Early Education and Care for Children Under the Age of 3*. Paris: Organization for Economic Co-operation and Development.

Cassidy, D. J., King, E. K., Wang, Y. C., Lower, J. K., & Kintner-Duffy, V. L. (2017). Teacher Work Environments are Toddler Learning Environments: Teacher Professional Well-Being, Classroom Emotional Support, and Toddlers' Emotional Expressions and Behaviours. *Early Child Development and Care*, *187*(11), 1666–1678. https://doi.org/10.1080/03004430.2016.1180516

Cents-Boonstra, M., Lichtwarck-Aschoff, A., Lara, M. M., & Denessen, E. (2022). Patterns of Motivating Teaching Behaviour and Student Engagement: A Microanalytic Approach. *European Journal of Psychology of Education*, *37*(1), 227–255. https://doi.org/10.1007/s10212-021-00543-3

Chen, J.-Q., McCray, J., Adams, M., & Leow, C. (2014). A Survey Study of Early Childhood Teachers' Beliefs and Confidence about Teaching Early Math. *Early Childhood Education Journal*, *42*(6), 367–377. https://doi.org/10.1007/s10643-013-0619-0

Çiftçi, A., & Topçu, M. S. (2022). Improving Early Childhood Pre-Service Teachers' Computational Thinking Teaching Self-Efficacy Beliefs in a STEM Course. *Research in Science & Technological Education*, *41*(4), 1215–1241. https://doi.org/10.1080/02635143.2022.2036117

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics Learned by Young Children in an Intervention Based on Learning Trajectories: A Large-Scale Cluster Randomized Trial. *Journal for Research in Mathematics Education*, *42*(2), 127–166. https://doi.org/10.5951/jresematheduc.42.2.0127

Cohen, F., & Anders, Y. (2020). Family Involvement in Early Childhood Education and Care and Its Effects on the Social-Emotional and Language Skills of 3-Year-Old Children. *School Effectiveness and School Improvement*, *31*(1), 125–142. https://doi.org/10.1080/09243453.2019.1646293

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (Revised Edition). New York, NY: Academic Press.

Conners-Burrow, N. A., Patrick, T., Kyzer, A., & McKelvey, L. (2017). A Preliminary Evaluation of REACH: Training Early Childhood Teachers to Support Children's Social and Emotional Development. *Early Childhood Education Journal*, *45*(2), 187–199. https://doi.org/10.1007/s10643-016-0781-2

Daucourt, M. C., Napoli, A. R., Quinn, J. M., Wood, S. G., & Hart, S. A. (2021). The Home Math Environment and Math Achievement: A Meta-Analysis. *Psychological Bulletin*, *147*(6), 565–596. https://doi.org/10.1037/bul0000330

Diebold, T., & Perren, S. (2020). The Impact of Childcare-Group Situational Age Composition on Caregiver-Child Interactions. *European Journal of Developmental Psychology*, *17*(4), 598–615. https://doi.org/10.1080/17405629.2019.1699050

Diefenbacher, S., Grgic, M., Neuberger, F., Maly-Motta, H., Spensberger, F., & Kuger, S. (2023). Pedagogical Practices in ECEC Institutions and Children's Linguistic, Motor, and Socio-emotional Needs during the COVID-19 Pandemic: Results from a Longitudinal Multi-Perspective Study in Germany. *Early Child Development and Care*, *193*(11-12), 1317–1334. https://doi.org/10.1080/03004430.2022.2116431

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary Reliability and Validity of a System for Observing Preschoolers' Competence in Classroom Interactions. *Early Childhood Research Quarterly*, *25*(1), 1–16. https://doi.org/10.1016/j.ecresq.2009.08.004

Duncan, R. J., Nordgren, I., Schmitt, S. A., & Vandell, D. L. (2023). Additive and Synergistic Relations of Early Mother–Child and Caregiver–Child Interactions and Socioeconomic Outcomes in Adulthood. *Infant and Child Development*, 1–24. https://doi.org/10.1002/icd.2406

Dunekacke, S., Jenßen, L., Eilerts, K., & Blömeke, S. (2016). Epistemological Beliefs of Prospective Preschool Teachers and Their Relation to Knowledge, Perception, and Planning Abilities in the Field of Mathematics: A Process Model. *ZDM – Mathematics Education*, *48*(1-2), 125–137. https://doi.org/10.1007/s11858-015-0711-6

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D.-M. C., Crawford, G. M., Frome, P. M., Clifford, R. M., Burchinal, M., Howes, C., Bryant, D. M., & Pianta, R. C. (2010). How do Pre-Kindergarteners Spend Their Time? Gender, Ethnicity, and Income as Predictors of Experiences in Pre-Kindergarten Classrooms. *Early Childhood Research Quarterly*, *25*(2), 177–193. https://doi.org/10.1016/j.ecresq.2009.10.003

Eckhardt, A. G., & Egert, F. (2018). Differences in Childcare Quality – A Matter of Personality Traits, Socialization Goals and Pre-Service Curriculum? *Early Child Development and Care*, *188*(12), 1726–1737. https://doi.org/10.1080/03004430.2016.1278372

Egert, F., Dederer, V., & Fukkink, R. G. (2020). The Impact of In-Service Professional Development On the Quality of Teacher-Child Interactions in Early Education and Care: A Meta-Analysis. *Educational Research Review*, *29*, 100309. https://doi.org/10.1016/j.edurev.2019.100309

Eling, V. (2022). Zur Bedeutung von Teamkonflikten für die Prozessqualität in Kindertageseinrichtungen. *Zeitschrift Für Pädagogik, 68*(5), 673–690.

Ellis, P. (2012). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results* (4th Edition). Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

Epskamp, S. (2019). *semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output: R package semPlot version 1.1.6*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/semPlot/index.html

Ertesvåg, S. K., Vaaland, G. S., & Lerkkanen, M.-K. (2022). Enhancing Upper Secondary Students' Engagement and Learning through the INTERACT Online, Video-Based Teacher Coaching Intervention: Protocol for a Mixed-Methods Cluster Randomized Controlled Trial and Process Evaluation. *International Journal of Educational Research*, *114*, 102013. https://doi.org/10.1016/j.ijer.2022.102013

Esch, K., Klaudy, E., Micheel, B., Stöbe-Blossey, S. (2006). *Qualitätskonzepte in der Kindertagesbetreuung: Ein Überblick*. Wiesbaden: VS.

Evens, M., Elen, J., & Depaepe, F. (2015). Developing Pedagogical Content Knowledge: Lessons Learned from Intervention Studies. *Education Research International*, *2015*, 1–23. https://doi.org/10.1155/2015/790417

Faas, S., Wu, S., & Geiger, S. (2017). The Importance of Play in Early Childhood Education: A Critical Perspective on Current Policies and Practices in Germany and Hong Kong. *Global Education Review*, *4*(2), 75–91.

Fassnacht, G. (1995). *Systematische Verhaltensbeobachtung.* München: Reinhardt.

Friederich, T. (2011). *Zusammenarbeit mit Eltern – Anforderungen an frühpädagogische Fachkräfte: Eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF)*. München: Deutsches Jugendinstitut.

Fröhlich-Gildhoff, K., Kraus, G., & Rönnau, M. (2006). Gemeinsam auf dem Weg. Eltern und ErzieherInnen gestalten Erziehungspartnerschaft. *Kindergarten Heute*, *10*, 6–15.

Fröhlich-Gildhoff, K., Nentwig-Gesemann, I., & Pietsch, S. (2011). *Kompetenzorientierung in der Qualifizierung frühpädagogischer Fachkräfte: Eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF)*. München: Deutsches Jugendinstitut.

Fthenakis, W. E. (Ed.). (2009). *Natur-Wissen schaffen: Band 2. Frühe mathematische Bildung*. Köln: Bildungsverlag EINS.

Fuligni, A. S., Howes, C., Huang, Y., Hong, S. S., & Lara-Cinisomo, S. (2012). Activity Settings and Daily Routines in Preschool Classrooms: Diverse Experiences in Early Learning Settings for Low-Income Children. *Early Childhood Research Quarterly*, *27*(2), 198–209. https://doi.org/10.1016/j.ecresq.2011.10.001

Gasteiger, H. (2012). Fostering Early Mathematical Competencies in Natural Learning Situations. Foundation and Challenges of a Competence-Oriented Concept of Mathematics Education in Kindergarten. *Journal Für Mathematik-Didaktik*, *33*(2), 181–201. https://doi.org/10.1007/s13138-012-0042-x

Gasteiger, H. (2014). Professionalization of Early Childhood Educators with a Focus on Natural Learning Situations and Individual Development of Mathematical Competencies: Results from an Evaluation Study. In *Early Mathematics Learning: Selected Papers of the POEM 2012 Conference* (pp. 275–290). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-4678-1_17

Gasteiger, H., & Benz, C. (2018). Enhancing and Analyzing Kindergarten Teachers' Professional Knowledge for Early Mathematics Education. *The Journal of Mathematical Behavior*, *51*, 109–117. https://doi.org/10.1016/j.jmathb.2018.01.002

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). *Mathematics Education for Young Children: What It Is and How to Promote It. Social Policy Report. Volume 22, Number 1.* Ann Arbor, MI: Society for Research in Child Development.

Goble, P., & Pianta, R. C. (2017). Teacher–Child Interactions in Free Choice and Teacher-Directed Activity Settings: Prediction to School Readiness. *Early Education and Development*, *28*(8), 1035–1051. https://doi.org/10.1080/10409289.2017.1322449

Goelman, H., Forer, B., Kershaw, P., Doherty, G., Lero, D., & LaGrange, A. (2006). Towards a Predictive Model of Quality in Canadian Child Care Centers. *Early Childhood Research Quarterly*, *21*(3), 280–295. https://doi.org/10.1016/j.ecresq.2006.07.005

Goldschmidt, P., & Phelps, G. (2010). Does Teacher Professional Development Affect Content and Pedagogical Knowledge: How Much and For How Long? *Economics of Education Review*, *29*(3), 432–439. https://doi.org/10.1016/j.econedurev.2009.10.002

Gordon, R. A., Hofer, K. G., Fujimoto, K. A., Risk, N., Kaestner, R., & Korenman, S. (2015). Identifying High-Quality Preschool Programs: New Evidence on the Validity of the Early Childhood Environment Rating Scale–Revised (ECERS-R) in Relation to School Readiness Goals. *Early Education and Development*, *26*(8), 1086–1110. https://doi.org/10.1080/10409289.2015.1036348

Grant, M. J., & Booth, A. (2009). A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies. *Health Information and Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Greve, W., & Wentura, D. (1997). *Wissenschaftliche Beobachtung: Eine Einführung* (2. Auflage). Weinheim: Beltz.

Grgic, M., Neuberger, F., Kalicki, B., Spensberger, F., Maly-Motta, H., Barbarino, B., Kuger, S., & Rauschenbach, T. (2022). Interaktionen in Kindertageseinrichtungen während der Corona-Pandemie – Elternkooperation, Fachkraft-Kind-Interaktionen und das Zusammenspiel der Kinder im Rahmen eingeschränkter Möglichkeiten. *Diskurs Kindheits- und Jugendforschung / Discourse. Journal of Childhood and Adolescence Research*, *17*(1), 27–56. https://doi.org/10.3224/diskurs.v17i1.03

Guedes, C., Cadima, J., Aguiar, T., Aguiar, C., & Barata, C. (2020). Activity Settings in Toddler Classrooms and Quality of Group and Individual Interactions. *Journal of Applied Developmental Psychology*, *67*, 101100. https://doi.org/10.1016/j.appdev.2019.101100

Guerrero, D., & de los Ríos, I. (2012). Professional Competences: A Classification of International Models. *Procedia - Social and Behavioral Sciences*, *46*, 1290–1296. https://doi.org/10.1016/j.sbspro.2012.05.290

Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among Preschool Teachers' Self-Efficacy, Classroom Quality, and Children's Language and Literacy Gains. *Teaching and Teacher Education*, *26*(4), 1094–1103. https://doi.org/10.1016/j.tate.2009.11.005

Haber, A. S., Puttre, H., Ghossainy, M. E., & Corriveau, K. H. (2021). "How Will You Construct a Pathway System?": Microanalysis of Teacher-Child Scientific Conversations. *Journal of Childhood, Education & Society*, *2*(3), 338–363. https://doi.org/10.37291/2717638X.202123117

Halle, T., Vick Whittaker, J. E., & Anderson, R. (2010). *Quality in Early Childhood Care and Education Settings: A Compendium of Measures.* Washington, DC: Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Hamre, B. K., Hatfield, B., Pianta, R. C., & Jamil, F. (2014). Evidence for General and Domain-Specific Elements of Teacher-Child Interactions: Associations with Preschool Children's Development. *Child Development*, *85*(3), 1257–1274. https://doi.org/10.1111/cdev.12184

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., Locasale-Crouch, J., Downer, J. T., Howes, C., La Paro, K., & Scott-Little, C. (2012). A Course on Effective Teacher-Child Interactions. *American Educational Research Journal*, *49*(1), 88–123. https://doi.org/10.3102/0002831211434596

Harms, T., Cryer, D., & Clifford, R. M. (2006). *Infant/Toddler Environment Rating Scale - Revised*. New York, NY: Teachers College Press.

Harms, T., Cryer, D., Clifford, R. M., & Yazejian, N. (2017). *Infant/Toddler Environment Rating Scale – Third Edition*. New York, NY: Teachers College Press.

Helmerhorst, K. O. W., Fukkink, R. G., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., & Tavecchio, L. W. C. (2017). Improving Quality of the Child Care Environment through a Consultancy Programme for Centre Directors. *International Journal of Early Years Education*, *25*(4), 361–378. https://doi.org/10.1080/09669760.2017.1321528

Hestenes, L. L., Cassidy, D. J., & Niemeyer, J. (2004). A Microanalysis of Teachers' Verbalizations in Inclusive Classrooms. *Early Education and Development*, *15*(1), 23–38. https://doi.org/10.1207/s15566935eed1501_2

Honig, A. S., & Lally, J. R. (1973). *Assessing the Behaviors of Caregivers. ABC-I and ABC-II*. Washington, DC: Department of Health, Education & Welfare, National Institute of Education.

Hooper, S. R., Roberts, J., Sideris, J., Burchinal, M., & Zeisel, S. (2010). Longitudinal Predictors of Reading and Math Trajectories through Middle School for African American versus Caucasian Students across Two Samples. *Developmental Psychology*, *46*(5), 1018–1029. https://doi.org/10.1037/a0018877

Hu, B. Y., Li, Y., Wang, C., Wu, H., & Vitiello, G. (2021). Preschool Teachers' Self-Efficacy, Classroom Process Quality, and Children's Social Skills: A Multilevel Mediation Analysis. *Early Childhood Research Quarterly*, *55*, 242–251. https://doi.org/10.1016/j.ecresq.2020.12.001

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hummel, T. G., Cohen, F., Gessulat, J., & Anders, Y. (2022). Measuring Interaction Quality between Parents and Professionals and Its Relation to Preschool Characteristics. *International Journal of Educational Research Open*, *3*, 100195. https://doi.org/10.1016/j.ijedro.2022.100195

Hyun, E., & Dan Marshall, J. (2003). Teachable-Moment-Oriented Curriculum Practice in Early Childhood Education. *Journal of Curriculum Studies*, *35*(1), 111–127. https://doi.org/10.1080/00220270210125583

Ishimine, K., & Tayler, C. (2014). Assessing Quality in Early Childhood Education and Care. *European Journal of Education*, *49*(2), 272–290. https://doi.org/10.1111/ejed.12043

Jamison, K. R., Cabell, S. Q., Locasale-Crouch, J., Hamre, B. K., & Pianta, R. C. (2014). CLASS–Infant: An Observational Measure for Assessing Teacher–Infant Interactions in Center-Based Child Care. *Early Education and Development*, *25*(4), 553–572. https://doi.org/10.1080/10409289.2013.822239

Justice, L. M., Mashburn, A., Hamre, B. K., & Pianta, R. C. (2008). Quality of Language and Literacy Instruction in Preschool Classrooms Serving At-Risk Pupils. *Early Childhood Research Quarterly*, *23*(1), 51–68. https://doi.org/10.1016/j.ecresq.2007.09.004

Kassambara, A. (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests: R package rstatix version 0.7.2*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/rstatix/index.html

Kasüschke, D. (Ed.). (2010). *Didaktik in der Pädagogik der frühen Kindheit. Grundlagen der Frühpädagogik: Band 3*. Köln: Carl Link.

Katz, L. G. (1993). *Five Perspectives on Quality in Early Childhood Programs. Perspectives from ERIC/EECE: A Monograph Series, No. 1*. Urbana, IL: ERIC Clearinghouse on Elementary and Early Childhood Education.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool Children's Mathematical Knowledge: The Effect of Teacher "Math Talk". *Developmental Psychology*, *42*(1), 59–69. https://doi.org/10.1037/0012-1649.42.1.59

Klinkhammer, N., Kuger, S., Kalicki, B., Riedel, B., Schacht, D. D., Rauschenbach, T., & Meiner-Teubner, C. (Ed.). (2021). *ERiK-Forschungsbericht I: Konzeption und Befunde des indikatorengestützten Monitorings zum KiQuTG*. Bielefeld: wbv. https://doi.org/10.3278/6004862w

Klinkhammer, N., Schacht, D. D., Meiner-Teubner, C., Kuger, S., Kalicki, B., & Riedel, B. (Ed.). (2022). *ERiK Forschungsbericht II: Befunde des indikatorengestützten Monitorings zum KiQuTG*. Bielefeld: wbv. https://doi.org/10.3278/9783763972999

Kluczniok, K., & Roßbach, H.-G. (2014). Conceptions of Educational Quality for Kindergartens. *Zeitschrift Für Erziehungswissenschaft*, *17*, 145–158. https://doi.org/10.1007/s11618-014-0578-2

Kluczniok, K., & Schmidt, T. (2020). Socio-Cultural Disparities in the Quality of Children's Interactions in Preschools. *European Early Childhood Education Research Journal*, *28*(4), 519–533. https://doi.org/10.1080/1350293X.2020.1783926

König, A. (2009). *Interaktionsprozesse zwischen Erzieherinnen und Kindern: Eine Videostudie aus dem Kindergartenalltag*. Wiesbaden: VS. https://doi.org/10.1007/978-3-531-91412-1

Kucharz, D., & Mackowiak, K. (2014). Gesamtdiskussion. In D. Kucharz, K. Mackowiak, S. Ziroli, A. Kauertz, E. Rathgeb-Schnierer, M. Dieck (Eds.), *Professionelles Handeln im Elementarbereich (PRIMEL). Eine Deutsch-Schweizerische Videostudie* (pp. 205–215). Münster: Waxmann.

La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2012). *Classroom Assessment Scoring System (CLASS) Manual, Toddler*. Baltimore, MD: Paul H. Brookes Publishing Company.

Landry, S. H., Crawford, A., Gunnewig, S., & Swank, P. R. (2002). *The CIRCLE-Teacher Behavior Rating Scale*. Unpublished Research Instrument.

Landry, S. H., Zucker, T. A., Taylor, H. B., Swank, P. R., Williams, J. M., Assel, M., Crawford, A., Huang, W., Clancy-Menchetti, J., Lonigan, C. J., Phillips, B. M., Eisenberg, N., Spinrad, T. L., de Villiers, J., de Villiers, P., Barnes, M., Starkey, P., & Klein, A. (2014). Enhancing Early Child Care Quality and Learning for Toddlers at Risk: The Responsive Early Childhood Program. *Developmental Psychology*, *50*(2), 526–541. https://doi.org/10.1037/a0033494

Lehrl, S., Ebert, S., Blaurock, S., Roßbach, H.-G., & Weinert, S. (2020). Long-Term and Domain-Specific Relations between the Early Years Home Learning Environment and Students' Academic Outcomes in Secondary School. *School Effectiveness and School Improvement, 31*(1), 102–124. https://doi.org/10.1080/09243453.2019.1618346

Linberg, A., Kluczniok, K., Burghardt, L., & Freund, J.-D. (2017). Quality of Toddler Childcare – Can It be Assessed with Questionnaires? *Early Child Development and Care*, *189*(8), 1369–1383. https://doi.org/10.1080/03004430.2017.1380636

Locasale-Crouch, J., Konold, T., Pianta, R. C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2007). Observed Classroom Quality Profiles in State-Funded Pre-Kindergarten Programs and Associations with Teacher, Program, and Classroom Characteristics. *Early Childhood Research Quarterly*, *22*(1), 3–17. https://doi.org/10.1016/j.ecresq.2006.05.001

Locasale-Crouch, J., Romo-Escudero, F., Clayback, K., Whittaker, J., Hamre, B. K., & Melo, C. (2023). Results from a Randomized Trial of the Effective Classroom Interactions for Toddler Educators Professional Development Intervention. *Early Childhood Research Quarterly*, *65*, 217–226. https://doi.org/10.1016/j.ecresq.2023.06.005

Løkken, I. M., Bjørnestad, E., Broekhuizen, M. L., & Moser, T. (2018). The Relationship between Structural Factors and Interaction Quality in Norwegian ECEC for Toddlers. *International Journal of Child Care and Education Policy, 12*(9), 1–15. https://doi.org/10.1186/s40723-018-0048-z

Lopez Boo, F., Dormal, M., & Weber, A. (2019). Validity of Four Measures of Child Care Quality in a National Sample of Centers in Ecuador. *PloS One*, *14*(2), e0209987. https://doi.org/10.1371/journal.pone.0209987

Loss, J., Kuger, S., Buchholz, U., Lehfeld, A.-S., Varnaccia, G., Haas, W., Jordan, S., Kalicki, B., Schienkiewitz, A., & Rauschenbach, T. (2021). Infektionsgeschehen und Eindämmungsmaßnahmen in Kitas während der COVID-19-Pandemie – Erkenntnisse aus der Corona-KiTa-Studie. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, *64*(12), 1581–1591. https://doi.org/10.1007/s00103-021-03449-z

Love, J. M., Harrison, L., Sagi-Schwartz, A., van IJzendoorn, M. H., Ross, C., Ungerer, J. A., Raikes, H., Brady-Smith, C., Boller, K., Brooks-Gunn, J., Constantine, J., Eliason Kisker, E., Paulsell, D., & Cha, R. (2003). Child Care Quality Matters: How Conclusions May Vary with Context. *Child Development, 74*(4), 1021–1033. https://doi.org/10.1111/1467-8624.00584

Luhmann, N. (1984). *Soziale Systeme: Grundriß einer allgemeinen Theorie*. Berlin: Suhrkamp.

Luhmann, N., & Schorr, K. E. (1979). *Reflexionsprobleme im Erziehungssystem.* Stuttgart: Klett.

Manning, M., Wong, G. T. W., Fleming, C. M., & Garvis, S. (2019). Is Teacher Qualification Associated with the Quality of the Early Childhood Education and Care Environment? A Meta-Analytic Review. *Review of Educational Research*, *89*(3), 370–415. https://doi.org/10.3102/0034654319837540

Maurice, J. v., & Weinert, S. (2009). Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen: Die interdisziplinäre Längsschnittstudie BiKS im Überblick. *SAL-Bulletin, 131*, 5–20.

Mayer, D., & Beckh, K. (2018). Erfassung pädagogischer Qualität in Kindertageseinrichtungen. *Frühe Bildung*, *7*(2), 67–76. https://doi.org/10.1026/2191-9186/a000370

McCray, J. S., & Chen, J.-Q. (2012). Pedagogical Content Knowledge for Preschool Mathematics: Construct Validity of a New Teacher Interview. *Journal of Research in Childhood Education*, *26*(3), 291–307. https://doi.org/10.1080/02568543.2012.685123

Melhuish, E. C., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., Tawell, A., Slot, P. L., Broekhuizen, M., & Leseman, P. P. M. (2015). *A Review of Research on the Effects of Early Childhood Education and Care (ECEC) upon Child Development. CARE Project.* Brussels: European Early Childhood Education and Care Publications.

Merçon-Vargas, E. A., Lima, R. F. F., Rosa, E. M., & Tudge, J. (2020). Processing Proximal Processes: What Bronfenbrenner Meant, What He Didn't Mean, and What He Should Have Meant. *Journal of Family Theory & Review*, *12*(3), 321–334. https://doi.org/10.1111/jftr.12373

Miller, P., Elliott, L. E., Podvysotska, T., Ptak, C., Duong, S., Fox, D., Coulanges, L., Libertus, M., Bachman, H. J., & Votruba-Drzal, E. (2023). Toddler Home Math Environment: Triangulating Multi-Method Assessments in a U.S. Sample. *Frontiers in Psychology*, *14*, 1105569. https://doi.org/10.3389/fpsyg.2023.1105569

Mischo, C., Wolstein, K., & Peters, S. (2022). Professional Vision of Early Childhood Teachers: Relations to Knowledge, Work Experience and Teacher Child-Interaction. *Early Years*, *43*(4–5), 828–844. https://doi.org/10.1080/09575146.2022.2028129

Moreno, A. J., Green, S., & Koehn, J. (2015). The Effectiveness of Coursework and Onsite Coaching at Improving the Quality of Care in Infant–Toddler Settings. *Early Education and Development*, *26*(1), 66–88. https://doi.org/10.1080/10409289.2014.941260

Naumann, A., Hochweber, J., & Klieme, E. (2016). A Psychometric Framework for the Evaluation of Instructional Sensitivity. *Educational Assessment*, *21*(2), 89–101. https://doi.org/10.1080/10627197.2016.1167591

Naumann, A., Rieser, S., Musow, S., Hochweber, J., & Hartig, J. (2019). Sensitivity of Test Items to Teaching Quality. *Learning and Instruction*, *60*, 41–53. https://doi.org/10.1016/j.learninstruc.2018.11.002

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which Preschool Mathematics Competencies are Most Predictive of Fifth Grade Achievement? *Early Childhood Research Quarterly*, *36*, 550–560. https://doi.org/10.1016/j.ecresq.2016.02.003

NICHD Early Child Care Research Network (2002). CHILD-CARE STRUCTURE → PROCESS → OUTCOME: Direct and Indirect Effects of Child-Care Quality on Young Children’s Development. *Psychological Science (13)*3, 199–206. https://doi.org/10.5194/gi-2016-11-RC2

Nores, M., Friedman-Krauss, A., & Figueras-Daniel, A. (2022). Activity Settings, Content, and Pedagogical Strategies in Preschool Classrooms: Do these Influence the Interactions We Observe? *Early Childhood Research Quarterly*, *58*, 264–277. https://doi.org/10.1016/j.ecresq.2021.09.011

OECD. (2021). *Starting Strong VI: Key OECD Indicators on Early Childhood Education and Care*. Paris: Organization for Economic Co-operation and Development. https://doi.org/10.1787/f47a06ae-en

Oppermann, E., Anders, Y., & Hachfeld, A. (2016). The Influence of Preschool Teachers’ Content Knowledge and Mathematical Ability Beliefs on Their Sensitivity to Mathematics in Children’s Play. *Teaching and Teacher Education*, *58*, 174–184. https://doi.org/10.1016/j.tate.2016.05.004

Oppermann, E., Barenthien, J., Burghardt, L., Steffensky, M., & Anders, Y. (2024). Beobachtungsverfahren zur Erfassung der Interaktionsqualität in Kindertageseinrichtungen. *Frühe Bildung*, *13*(1), 29–35. https://doi.org/10.1026/2191-9186/a000647

Pagani, V. (2019). A Critical-Cultural Approach to Standardized Evaluation Instruments. A Mixed Methods Study of the CLASS Toddler Application in Italy. In *Early Childhood Care and Education (ECCE) Proceedings* (pp. 435–437). Moscow: MGIMO University Press.

Passaretta, G., Skopek, J., & van Huizen, T. (2022). Is Social Inequality in School-Age Achievement Generated before or during Schooling? A European Perspective. *European Sociological Review*, *38*(6), 849–865. https://doi.org/10.1093/esr/jcac005

Pastori, G., & Pagani, V. (2017). Is Validation always Valid? Cross-Cultural Complexities of Standard-Based Instruments Migrating out of Their Context. *European Early Childhood Education Research Journal*, *25*(5), 682–697. https://doi.org/10.1080/1350293X.2017.1356545

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through Second Grade. *Child Development*, *72*(5), 1534–1553. https://doi.org/10.1111/1467-8624.00364

Pelkowski, T., Herman, E., Trahan, K., Winters, D. M., Tananis, C., Swartz, M. I., Bunt, N., & Rodgick, C. (2019). Fostering A “Head Start” in Math: Observing Teachers in Early Childhood Mathematics Engagement. *Journal of Early Childhood Teacher Education*, *40*(2), 96–119. https://doi.org/10.1080/10901027.2018.1522397

Pessanha, M., Aguiar, C., & Bairrão, J. (2007). Influence of Structural Features on Portuguese Toddler Child Care Quality. *Early Childhood Research Quarterly*, *22*(2), 204–214. https://doi.org/10.1016/j.ecresq.2007.02.003

Pianta, R. C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of Pre-Kindergarten Programs, Classrooms, and Teachers: Do They Predict Observed Classroom Quality and Child-Teacher Interactions? *Applied Developmental Science*, *9*(3), 144–159. https://doi.org/10.1207/s1532480xads0903_2

Pianta, R. C., Whittaker, J. E., Vitiello, V., Ruzek, E., Ansari, A., Hofkens, T., & DeCoster, J. (2020). Children's School Readiness Skills across the Pre-K Year: Associations with Teacher-Student Interactions, Teacher Practices, and Exposure to Academic Content. *Journal of Applied Developmental Psychology*, *66*, 101084. https://doi.org/10.1016/j.appdev.2019.101084

Pohl, M. (2013). Qualität und Qualitätsmanagement – Definitionen. In M. Schmidt (Ed.), *Qualitätsmanagement in Kindertagesstätten: Von der Norm zur Haltung; ein konstruktiv-kritischer Diskurs: Kitapraxis* (pp. 49–52). Wiesbaden: Kommunal- und Schulverlag.

Powell, D. R., Son, S.-H., File, N., & San Juan, R. R. (2010). Parent-School Relationships and Children's Academic and Social Outcomes in Public School Pre-Kindergarten. *Journal of School Psychology*, *48*(4), 269–292. https://doi.org/10.1016/j.jsp.2010.03.002

Preissing, C., & Heller, E. (2003). *Qualität im Situationsansatz: Qualitätskriterien und Materialien für die Qualitätsentwicklung in Kindertageseinrichtungen*. Weinheim: Beltz.

Ramani, G. B., Rowe, M. L., Eason, S. H., & Leech, K. A. (2015). Math Talk during informal learning activities in Head Start Families. *Cognitive Development*, *35*, 15–33. https://doi.org/10.1016/j.cogdev.2014.11.002

Ramani, G. B., & Siegler, R. S. (2008). Promoting Broad and Stable Improvements in Low-Income Children's Numerical Knowledge through Playing Number Board Games. *Child Development*, *79*(2), 375–394. https://doi.org/10.1111/j.1467-8624.2007.01131.x

Rammstedt, B. (Ed.). (2013). *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich: Ergebnisse von PIAAC 2012*. Münster: Waxmann.

Ratten, V. (2022). Coronavirus (COVID-19) and Social Value Co-Creation. *International Journal of Sociology and Social Policy*, *42*(3/4), 222–231. https://doi.org/10.1108/IJSSP-06-2020-0237

Rentzou, K. (2012). Examination of Work Environment Factors Relating to Burnout Syndrome of Early Childhood Educators in Greece. *Child Care in Practice*, *18*(2), 165–181. https://doi.org/10.1080/13575279.2012.657609

Resa, E., Groeneveld, I., Turani, D., & Anders, Y. (2018). The Role of Professional Exchange in Improving Language-Related Process Quality in Daycare Centres. *Research Papers in Education*, *33*(4), 472–491. https://doi.org/10.1080/02671522.2017.1353671

Resnick, G., & Zill, N. (1999). *Is Head Start Providing High-Quality Educational Services? "Unpacking" Classroom Processes*. Rockville, MD: Westat.

Rodriguez, B. J., & Garza, S. (2014). Test Review: Classroom Assessment Scoring System Manual: Toddler. *Journal of Psychoeducational Assessment*, *32*(6), 573–576. https://doi.org/10.1177/0734282914523085

Rosa, E. M., & Tudge, J. (2013). Urie Bronfenbrenner's Theory of Human Development: Its Evolution from Ecology to Bioecology. *Journal of Family Theory & Review*, *5*(4), 243–258. https://doi.org/10.1111/jftr.12022

Roßbach, H.-G. (2005). Effekte qualitativ guter Betreuung, Bildung und Erziehung im frühen Kindesalter auf Kinder und ihre Familien. In H.-G. Roßbach, H. G., & L. Ahnert (Ed.), *Materialien zum zwölften Kinder- und Jugendbericht. Bildung, Betreuung und Erziehung von Kindern unter sechs Jahren* (pp. 55–174). München. Deutsches Jugendinstitut

Roßbach, H.-G., & Tietze, W. (2018). *Kindergarten-Skala-Erweiterung (KES-E)*. Kiliansroda: das netz.

Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Salminen, J., Guedes, C., Lerkkanen, M.-K., Pakarinen, E., & Cadima, J. (2021). Teacher–Child Interaction Quality and Children's Self-Regulation in Toddler Classrooms in Finland and Portugal. *Infant and Child Development*, *30*(3), 1–23. https://doi.org/10.1002/icd.2222

Salminen, J., Khanolainen, D., Koponen, T., Torppa, M., & Lerkkanen, M.-K. (2021). Development of Numeracy and Literacy Skills in Early Childhood – A Longitudinal Study on the Roles of Home Environment and Familial Risk for Reading and Math Difficulties. *Frontiers in Education*, *6*, 725337. https://doi.org/10.3389/feduc.2021.725337

Schaack, D. D., Le, V.-N., & Stedron, J. (2020). When Fulfillment is Not Enough: Early Childhood Teacher Occupational Burnout and Turnover Intentions from a Job Demands and Resources Perspective. *Early Education and Development*, *31*(7), 1011–1030. https://doi.org/10.1080/10409289.2020.1791648

Schipper, E. J. de, Riksen-Walraven, J. M., & Geurts, S. A. (2007). Multiple Determinants of Caregiver Behavior in Child Care Centers. *Early Childhood Research Quarterly*, *22*(3), 312–326. https://doi.org/10.1016/j.ecresq.2007.04.004

Schmidt, V. H. (2005). Die Systeme der Systemtheorie: Stärken, Schwächen und ein Lösungsvorschlag. *Zeitschrift für Soziologie*, *34*(6), 406–424. https://doi.org/10.1515/zfsoz-2005-0601

Schreyer, I., Krause, M., Brandl-Knefz, M., & Nicko, O. (2015). Arbeitsbedingungen, Arbeitszufriedenheit und das Auftreten von beruflichen Gratifikationskrisen bei Kita-Mitarbeiter/innen in Deutschland. *Frühe Bildung*, *4*(2), 71–82. https://doi.org/10.1026/2191-9186/a000201

Schweinhart, L. J., & Weikart, D. P. (1980). *Young Children Grow Up: The Effects of the Perry Preschool Program on Youths through Age 15*. *Monographs of the High/Scope Educational Research Foundation: No. 7*. Ypsilanti, MI: High/Scope Press.

Shulman, L. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Siraj-Blatchford, I., Kingston, D., & Melhuish, E. (2015). *Assessing Quality in Early Childhood Education and Care: Sustained Shared Thinking and Emotional Well-being (SSTEW) Scale for 2-5-year-olds Provision.* London: Trentham Books.

Siraj-Blatchford, I., Sylva, K., Muttock, S., Gilden, R., & Bell, D. (2002). *Researching Effective Pedagogy in the Early Years (REPEY). Research Report 356.* London: Department for Education and Skills.

Slot, P. L., Leseman, P. P. M., Verhagen, J., & Mulder, H. (2015). Associations between Structural Quality Aspects and Process Quality in Dutch Early Childhood Education and Care Settings. *Early Childhood Research Quarterly*, *33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, *25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Steigleder, J., Buhr, L., Ehm, J.-H., Gawrilow, C., & Suchodoletz, A. v. (2023). Changes in Subjective Stress Experiences and Self-Efficacy Beliefs of Preschool Teachers in Germany: A Longitudinal Study during 12 Months of the COVID-19 Pandemic. *Teaching and Teacher Education*, *124*, 104015. https://doi.org/10.1016/j.tate.2023.104015

Stipek, D., & Byler, P. (2004). The Early Childhood Classroom Observation Measure. *Early Childhood Research Quarterly*, *19*(3), 375–397. https://doi.org/10.1016/j.ecresq.2004.07.007

Suchodoletz, A. v., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A Typical Morning in Preschool: Observations of Teacher-Child Interactions in German Preschools. *Early Childhood Research Quarterly*, *29*(4), 509–519. https://doi.org/10.1016/j.ecresq.2014.05.010

Suchodoletz, A. v., Jamil, F. M., Larsen, R. A., & Hamre, B. K. (2018). Personal and Contextual Factors Associated with Growth in Preschool Teachers' Self-Efficacy Beliefs during a Longitudinal Professional Development Study. *Teaching and Teacher Education*, *75*, 278–289. https://doi.org/10.1016/j.tate.2018.07.009

Suchodoletz, A. v., Lee, D. S., Henry, J., Tamang, S., Premachandra, B., & Yoshikawa, H. (2023). Early Childhood Education and Care Quality and Associations with Child Outcomes: A Meta-Analysis. *PloS One*, *18*(5), e0285985. https://doi.org/10.1371/journal.pone.0285985

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2011). Pre-School Quality and Educational Outcomes at Age 11: Low Quality has Little Benefit. *Journal of Early Childhood Research*, *9*(2), 109–124. https://doi.org/10.1177/1476718X10387900

Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2011). *ECERS-E: The Four Curricular Subscales Extension to the Early Childhood Environment Rating Scale (ECERS-R)* (4th Edition). New York, NY: Teachers College Press.

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing Quality in Early Childhood through Environmental Rating Scales. *Early Childhood Research Quarterly*, *21*(1), 76–92. https://doi.org/10.1016/j.ecresq.2006.01.003

Textor, M. R., & Blank, B. (2004). *Elternmitarbeit: Auf dem Wege zur Bildungs- und Erziehungspartnerschaft* (2. überarbeitete, aktualisierte Fassung). München: Bayerisches Staatsministerium für Arbeit und Sozialordnung, Familie.

Thorpe, K., Houen, S., Rankin, P., Pattinson, C., & Staton, S. (2022). Do the Numbers Add Up? Questioning Measurement that Places Australian ECEC Teaching as ‘Low Quality’. *The Australian Educational Researcher, 50*, 781–800. https://doi.org/10.1007/s13384-022-00525-4

Thorpe, K., Rankin, P., Beatton, T., Houen, S., Sandi, M., Siraj, I., & Staton, S. (2020). The When and What of Measuring ECE Quality: Analysis of Variation in the Classroom Assessment Scoring System (CLASS) across the ECE Day. *Early Childhood Research Quarterly*, *53*, 274–286. https://doi.org/10.1016/j.ecresq.2020.05.003

Tietze, W. (Ed.). (1998). *Wie gut sind unsere Kindergärten? Eine Untersuchung zur pädagogischen Qualität in deutschen Kindergärten*. Neuwied: Luchterhand.

Tietze, W., Becker-Stoll, F., Bensel, J., Eckhardt, A. G., Haug-Schnabel, G., Kalicki, B., Keller, H., & Leyendecker, B. (Ed.). (2013). *Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit (NUBBEK)*. Weimar/Berlin: das netz.

Tietze, W., Bolz, M., Grenner, K., Schlecht, D., & Wellner, B. (2007). *Krippen-Skala (KRIPS-R): Feststellung und Unterstützung pädagogischer Qualität in Krippen* (revidierte Fassung). Berlin: Cornelsen.

Tietze, W., & Roßbach, H.-G. (2019). *Krippen-Skala (KRIPS-RZ): Revidierte Fassung mit Zusatzmerkmalen* (3. erweiterte Auflage). Kiliansroda: das netz.

Tietze, W., & Viernickel, S. (Eds.). (2023). *Pädagogische Qualität in Tageseinrichtungen für Kinder: Ein Nationaler Kriterienkatalog* (6. Auflage). Weimar/Berlin: das netz.

Tobin, J. (2005). Quality in Early Childhood Education: An Anthropologist's Perspective. *Early Education and Development*, *16*(4), 421–434. https://doi.org/10.1207/s15566935eed1604_3

Torquati, J. C., Raikes, H., & Huddleston-Casas, C. A. (2007). Teacher Education, Motivation, Compensation, Workplace Support, and Links to Quality of Center-Based Child Care and Teachers' Intention to Stay in the Early Childhood Profession. *Early Childhood Research Quarterly*, *22*(2), 261–275. https://doi.org/10.1016/j.ecresq.2007.03.004

Trauernicht, M., Besser, N., & Anders, Y. (2022). Burnout in der Kita und der Zusammenhang zu Aspekten der Arbeitszufriedenheit. *Frühe Bildung*, *11*(2), 85–93. https://doi.org/10.1026/2191-9186/a000566

Tudge, J. R. H., Mokrova, I., Hatfield, B. E., & Karnik, R. B. (2009). Uses and Misuses of Bronfenbrenner's Bioecological Theory of Human Development. *Journal of Family Theory & Review*, *1*(4), 198–210. https://doi.org/10.1111/j.1756-2589.2009.00026.x

Ulferts, H., Wolf, K. M., & Anders, Y. (2019). Impact of Process Quality in Early Childhood Education and Care on Academic Outcomes: Longitudinal Meta-Analysis. *Child Development*, *90*(5), 1474–1489. https://doi.org/10.1111/cdev.13296

Van IJzendoorn, M. H., Tavecchio, L. W. C., Stams, G.-J. J., Verhoeven, M. J. E., & Reiling, E. J. (1998). Quality of Center Day Care and Attunement between Parents and Caregivers: Center Day Care in Cross-National Perspective. *The Journal of Genetic Psychology*, *159*(4), 437–454. https://doi.org/10.1080/00221329809596163

Vandell, D. L., Burchinal, M., & Pierce, K. M. (2016). Early Child Care and Adolescent Functioning at the End of High School: Results from the NICHD Study of Early Child Care and Youth Development. *Developmental Psychology*, *52*(10), 1634–1645. https://doi.org/10.1037/dev0000169

Vandenbroeck, M., & Peeters, J. (2014). Democratic Experimentation in Early Childhood Education. In D. M. de Bie & D. Wildemeersch (Eds.), *Civic Learning, Democratic Citizenship and the Public Sphere* (pp. 151–165). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7259-5_11

Vermeer, H. J., van IJzendoorn, M. H., Kruif, R. E. L. de, Fukkink, R. G., Tavecchio, L. W. C., Riksen-Walraven, J. M., & van Zeijl, J. (2008). Child Care in The Netherlands: Trends in Quality Over the Years 1995-2005. *The Journal of Genetic Psychology*, *169*(4), 360–385. https://doi.org/10.3200/GNTP.169.4.360-385

Viernickel, S., Fuchs-Rechlin, K., Strehmel, P., Preissing, C., Bensel, J., & Haug-Schnabel, G. (2016). *Qualität für alle: Wissenschaftlich begründete Standards für die Kindertagesbetreuung*. Freiburg im Breisgau: Herder.

Viernickel, S., Voss, A., & Mauz, E. (2017). *Arbeitsplatz Kita: Belastungen erkennen, Gesundheit fördern*. Weinheim: Beltz Juventa.

Vitiello, V. E., Booren, L. M., Downer, J. T., & Williford, A. (2012). Variation in Children's Classroom Engagement throughout a Day in Preschool: Relations to Classroom and Child Factors. *Early Childhood Research Quarterly*, *27*(2), 210–220. https://doi.org/10.1016/j.ecresq.2011.08.005

Vogt, F., Hauser, B., Stebler, R., Rechsteiner, K., & Urech, C. (2020). Learning Through Play – Pedagogy and Learning Outcomes in Early Childhood Mathematics. In O. Thiel & B. Perry (Eds.), *Innovative Approaches in Early Childhood Mathematics* (pp. 127–141). London: Routledge. https://doi.org/10.4324/9780429331244-10

Wadepohl, H., & Mackowiak, K. (2013). Entwicklung und Erprobung eines Beobachtungsinstruments zur Analyse der Beziehungs- bzw. Bindungsgestaltung von frühpädagogischen Fachkräften in Freispielsituationen. In K. Fröhlich-Gildhoff, I. Nentwig-Gesemann, A. König, U. Stenger, & D. Weltzien (Eds.), *Materialien zur Frühpädagogik: Band 12. Forschung in der Frühpädagogik, Interaktion zwischen Fachkräften und Kindern* (pp. 87–118). Freiburg im Breisgau: FEL.

Weinstock, P., Bos, J., Tseng, F., Rosenthal, E., Ortiz, L., Dowsett, C., Huston, A., & Bentley, A. (2012). *Evaluation of Program for Infant/Toddler Care (PITC): An On-Site Training of Caregivers. Final Report*. Washington, DC: National Center for Education Evaluation and Regional Assistance.

Weiser, L. E. (2022). Young Children's Free Play in Nature: An Essential Foundation for STEM Learning in Germany. In S. D. Tunnicliffe & T. J. Kennedy (Eds.), *Play and STEM Education in the Early Years: International Policies and Practices* (pp. 85–103). Cham: Springer International. https://doi.org/10.1007/978-3-030-99830-1_4

Weitzel, H., & Blank, R. (2020). Pedagogical Content Knowledge in Peer Dialogues between Pre-Service Biology Teachers in the Planning of Science Lessons. Results of an Intervention Study. *Journal of Science Teacher Education*, *31*(1), 75–93. https://doi.org/10.1080/1046560X.2019.1664874

Werner, C. D., Vermeer, H. J., Linting, M., & van IJzendoorn, M. H. (2018). Video-Feedback Intervention in Center-Based Child Care: A Randomized Controlled Trial. *Early Childhood Research Quarterly*, *42*, 93–104. https://doi.org/10.1016/j.ecresq.2017.07.005

Wertfein, M., Müller, K., & Danay, E. (2013). Die Bedeutung des Teams für die Interaktionsqualität in Kinderkrippen. *Frühe Bildung, 2*(1), 20–27. https://doi.org/10.1026/2191-9186/a000073

Wertfein, M., Spies-Kofler, A., & Becker-Stoll, F. (2009). Quality Curriculum for Under-Threes: The Impact of Structural Standards. *Early Years*, *29*(1), 19–31. https://doi.org/10.1080/09575140802685297

Wickham, H., Chang, W., Henry, L., Lin Pedersen, T., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2023). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics: R package ggplot2 version 3.4.4*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/ggplot2/index.html

Wieduwilt, N., Lehrl, S., & Anders, Y. (2023). Preschool Teachers' Language-Related Pedagogical Beliefs and Their Relation to Observed Classroom Quality. *Early Childhood Research Quarterly, 62*, 175–185. https://doi.org/10.1016/j.ecresq.2022.08.001

Wildgruber, A., Wertfein, M., Wirts, C., Kammermeier, M., & Danay, E. (2016). Situative Unterschiede der Interaktionsqualität im Verlauf des Kindergartenalltags. *Frühe Bildung*, *5*(4), 206–213. https://doi.org/10.1026/2191-9186/a000283

Xia, M., Li, X., & Tudge, J. R. (2020). Operationalizing Urie Bronfenbrenner's Process-Person-Context-Time Model. *Human Development*, *64*(1), 10–20. https://doi.org/10.1159/000507958

Yang, M., & Liang, J. (2022). Early Number Word Learning: Associations with Domain-General and Domain-Specific Quantitative Abilities. *Frontiers in Psychology*, *13*, 1024426. https://doi.org/10.3389/fpsyg.2022.1024426

Appendix – Study 1

Assessing Global and Math-Specific Teacher–Child Interaction Quality in Early Childcare Settings: A Systematic Literature Review of Instruments Used in Research

Baron, F., Linberg, A., & Kuger, S. (2023). Assessing Global and Math-Specific Teacher–Child Interaction Quality in Early Childcare Settings: A Systematic Literature Review of Instruments Used in Research, *European Early Childhood Education Research Journal, 31*(4), 640–659. https://doi.org/10.1080/1350293X.2022.2154817

Studies show that processes like teacher–child interactions are a key dimension of childcare quality, and that global and domain-specific interactions seem to have different effects on child development. A variety of instruments have been used to assess quality in the field of early childhood research. The aim of this study was to identify and discuss instruments that assess global and/or math-specific interaction quality in childcare settings (birth to school-entry). In the systematic literature review, we found 4211 publications using instruments that assess global and/or math-specific interaction quality, of which 148 remained after a screening of titles, abstracts, and full-texts. In all, 55 instruments assessing interaction quality were identified and analyzed regarding their (sub)scales, items, method, procedure, children's age, reliability, and validity. We identified more instruments measuring global aspects of teacher–child interactions than measuring math-specific aspects. In general, some instruments might be over- or underrepresented, due to linguistic and publication bias.

Keywords: early childhood education and care (ECEC); systematic review; overview; instruments; interaction quality; math-specific interaction quality

Introduction

Institutional childcare settings are critical learning environments during early childhood where children acquire and develop competencies and skills (Peisner-Feinberg et al. 2001; Sylva et al. 2011). In addition to other quality factors, such as structural characteristics of childcare centers, process quality is strongly related to child development and learning (Anders et al. 2013; Burchinal et al. 2014). In particular, teacher–child interaction quality, an aspect of process quality, is associated with different facets of child development: a large number of studies point to the effects of high-quality interactions on the development of academic outcomes, such as literacy and mathematics, as well as socioemotional skills (Belsky et al. 2007; Burger 2010; Mashburn et al. 2008; Melhuish et al. 2015). Those high-quality interactions are predictive not only for short-term outcomes but also for later academic achievement, emphasizing their importance (Lehrl, Kluczniok, and Roßbach 2016; Vandell et al. 2010).

Given the relevance of domain-specific competencies to school success and educational careers, interactions are often differentiated into global and domain-specific characteristics (Kluczniok et al. 2013; Sylva et al. 2006). Global facets of interactions are characterized as contributing to all areas of children’s development and are well-described in research (Hamre et al. 2014; Hooper 2017; Hu et al. 2019). With regard to domain specificity, interest in mathematical development and math-specific support has increased in current early childhood research (Burghardt et al. 2020; Braak et al. 2022; Möhring et al. 2021; Pelkowski et al. 2019; Reikerås and Salomonsen 2019; Salminen et al. 2021). Associations between math-specific interactions and children’s mathematical competencies have been identified in preschoolers aged three to six years (Anders et al. 2012) and can be traced to elementary school (Anders et al. 2013; Lehrl, Kluczniok, and Roßbach 2016). This impact of math-specific interactions is particularly important, as math-specific competencies appear to be related to children’s abilities as well as socio-economic status and income later in life, somewhat more than general or other specific domains (Braak et al. 2022; Rammstedt 2013). Consequently, this study focuses on both global and math-specific interactions in early childhood education and care (ECEC) and provides an overview of the instruments used in research to assess the quality of those interactions.

Global and math-specific interactions in ECEC

According to Bronfenbrenner’s bio-ecological theory (Bronfenbrenner and Morris 1998), proximal processes are central to human beings’ developmental outcomes, especially in early phases. Those processes include reciprocal interactions between the individual and the environment or other persons. Another theory highlighting the importance of interactions to child outcomes is Vygotsky’s zone of proximal development (Vygotsky 1977). According to this theory, adult guidance is seen as an essential mechanism for children's competence development, and through joint problem-solving, the child reaches the next stage of development, thereby gaining cognitive growth. Both theories point to the importance of high-quality interactions between children and their caregivers. Studies show that global interactions and domain-specific interactions can have different effects on child development: global interactions positively influence different areas of child learning, and domain-specific interactions address particular learning areas (Bleses et al. 2020; Hamre et al. 2014). Therefore, global and domain-specific (including math-specific) interactions should be differentiated in research, policy, and practice.

In the context of process quality, global interactions refer to activities and interactions that are not targeted to a specific competence domain but rather address and support several competencies and skills at the same time (Berk and Winsler 1995; Remsperger 2013; Siraj-Blatchford et al. 2002). According to previous studies, global teacher–child interactions can be designated into several categories. In line with theoretical assumptions made by La Paro, Hamre and Pianta in their CLASS manuals, which are empirically supported by other researchers, two broad domains seem to be essential to the development and learning of young children (Bichay-Awadalla and Bulotsky-Shearer 2021; La Paro, Hamre, and Pianta 2012; van Schaik, Leseman, and Haan 2018). One domain involves addressing emotional needs and consists of interactions that promote a warm atmosphere with expressions of emotions, physical proximity, and teacher sensitivity (Jamison et al. 2014). The other domain involves adressing learning occasions, where the teacher provides diverse opportunities for children to explore and learn, such as by scaffolding them during a challenging activity (Berk and Winsler 1995). For older children, organizing their learning environment gets to be more important, which becomes an additional domain for that age group (Pianta, La Paro, and Hamre 2008; Thomason and La Paro 2009). Behavior in relation to this domain includes effective management of time and activities, so that all children in the classroom can participate (Curby and Chavez 2013). The quality of these domains of global interactions predicts both short- and long-term socio-emotional and academic (including math-specific) outcomes (Lehrl et al. 2017; Vandell, Burchinal, and Pierce 2016).

In contrast to global interactions, domain-specific interactions are activities and interactions that foster specific competencies, such as mathematical skills (Hamre et al. 2014; Lehrl et al. 2020; Purpura et al. 2017). Math-specific interactions give children experiences with mathematical meanings and constructs, for example through encouraging children to engage in mathematical activities, like counting and measuring, or through mathematical language provided by teachers. Several studies emphasize that the quality and amount of mathematical language positively affects children's mathematical competencies (Boonen, Kolkman, and Kroesbergen 2011; Klibanoff et al. 2006). Furthermore, activities such as board games can stimulate certain domain-specific developmental areas, including mathematical competencies (LeFevre et al. 2010; Lehrl et al. 2020). In general, math-specific interactions can address different mathematical aspects: in a narrow sense, math-specific interactions involve numerical terms and the number system (e.g., counting), while a broader understanding also includes spatial, temporal, and quantitative contents (e.g., small, tall, short, long, less, more; Ramani et al. 2015). Both types of math-specific interactions seem to have a positive influence on children's mathematical outcomes. In this study, we consider the broad understanding of math-specific interactions.

Instruments for assessing interaction quality in ECEC

In early childhood research, a wide variety of instruments exists to measure global and math-specific interaction quality. Among these, observations are depicted as the 'gold standard', as self-reports may be inaccurate or biased (Linberg et al. 2017). However, observational instruments differ in their criteria regarding the observer, the observation material, and the observation method (Fassnacht 1995). In terms of the observer, most instruments are designed to be applied by an objective observer who was trained to use the instrument, as self-observation by staff evaluating their own work might be biased. The observation material can be categorized into technically mediated vs. non-mediated observations (Greve and Wentura 1997): instruments often use technically non-mediated live observation in the childcare setting, which takes several hours. Sometimes, particularly for the observation of math-specific interactions, material is technically mediated (i.e., videos are recorded in childcare settings and coded afterwards). In terms of the observation method, instruments can be differentiated into macro-analytic or micro-analytic instruments (Fassnacht 1995; Greve and Wentura 1997; Zechmeister, Zechmeister, and Shaughnessy 2009). Macro-analytic rating instruments are the most common and time-effective in assessing interaction quality, as they aggregate the observed behavior over the total observation time, while micro-analytic instruments often focus on more specific aspects of (interaction) behavior using time- or event-sampling approaches. Another way observations can be conducted is through the thin-slice method, where raters see a short excerpt of teachers' behavior and are asked to not take notes throughout the observation. When coding, they quickly rate each behavior based on their first impressions of the teacher–child interactions (Ambady, Bernieri, and Richeson 2000; Murphy 2005).

In the literature, some reviews of instruments assessing interaction quality in ECEC have already been conducted, but they are either not systematic (Mathers, Singler, and Karemaker 2012) or somewhat outdated (Halle, Vick Whittaker, and Anderson 2010; Ishimine and Tayler 2014), predating the recent increase in focussing on assessing teacher–child interaction quality. Some overviews take a specific aspect of interactions into account, such as dialogues between teachers and children (García-Carrión and Villardón-Gallego 2016) or language support (Zimmer et al. 2020). Other overviews have a special focus on a certain group such as multilingual children (Langeloo et al. 2019; Peisner-Feinberg et al. 2014), or do not include group-based interactions with all children (Riedmeier 2019). Another way overviews are conducted is to select one instrument and only include studies using this specific instrument (Brunsek et al. 2017; Perlman et al. 2016). Moreover, most overviews do not focus on the instruments themselves but rather on the impact of interaction quality on child outcomes (Ulferts, Wolf, and Anders 2019). With regard to math-specific interactions, to the best of our knowledge the only overviews that exist concern home mathematical or numerical environments and math competencies of children (Daucourt et al. 2021; Mutaf-Yıldız et al. 2020). In systematic reviews of instruments used in early childhood research, math-specific interactions are often missing.

Present study

Both global and math-specific teacher–child interactions seem important to child development and learning in early childhood and beyond. Thus, it is of great interest how the quality of those interactions can be assessed in ECEC. An overview of instruments used in early childhood research offers the opportunity to highlight the specialties of individual instruments. This, in turn, enables the identification of which instruments can best be used for which research purposes (i.e., for tracing effects of teacher–child interaction quality on children's competencies (or other characteristics)). An overview compiles different instruments, thereby contributing to effective research economy. However, considering the state of literature, no current overview exists with a focus on instruments assessing global and/or math-specific interaction quality in childcare settings.

To address this gap, this study aims to systematically review research articles on the available instruments that assess global and/or math-specific interaction quality. The review is guided by the following research questions:

(1) Which instruments are used in research to assess global and/or math-specific teacher–child interaction quality in early childcare settings?
(2) How do the instruments differ on the following criteria: (sub)scales, items, method, procedure, children's age, reported reliability, and reported validity?

Method

Data sources and searching process

For the systematic literature review, we applied the SALSA method (Grant and Booth 2009), which includes different steps: the searching process of articles (Search), the selection process of articles (AppraisaL), the summary of the identified instruments (Synthesis), and the discussion of the results (Analysis). We searched for research articles in April 2021 in three databases: EBSCOhost, Web of Science, and FIS Bildung. Table 1 provides an overview of the search terms. All English- and German-language articles published before the search date (April 11, 2021) were considered, and we did not define a lower limit of publication year. We also included articles in non-scientific journals, dissertations, manuals, and conference contributions. As observations are considered the 'gold standard' for assessing interaction quality (Linberg et al. 2017), only observational instruments were included in the review. Additionally, we only included instruments that consider all children in group-based interactions and that do not focus on specific children in the classroom (such as inCLASS (Downer et al. 2010); see Riedmeier (2019) for an overview of child-specific instruments).

Screening process

All publications were reviewed and coded by the first author. We used inclusion and exclusion criteria that corresponded with our research questions, and we screened each article in light of these criteria. Inclusion criteria required that the instrument involved (1) third-person observations and (2) childcare group settings before school-entry, assessing

(3) at least one aspect of interaction quality and (4) at least one aspect of global and/or math-specific interaction quality. Exclusion criteria required that, concerning interaction quality, the study described: an assessment by a non-objective person (e.g., self-report) (exclusion criteria 1), an assessment focussing on school children or non-pedagogical persons (e.g., parents) (exclusion criteria 2), an assessment of a quality dimension other than interaction quality (e.g., structural quality) (exclusion criteria 3), or an assessment of an interaction type other than global or math-specific (e.g., language-specific interactions) (exclusion criteria 4). We also excluded (systematic) reviews and meta-analyses (exclusion criteria 5), and during the screening process, we noted whether the article met more than one exclusion criteria (exclusion criteria 6). If the coding was not clear, the first author discussed the in- or exclusion of an article with the other authors.

Overall, we found 4211 publications. After duplicates were removed (a duplicate was defined as the same article (identical author(s), title and publication date) found through several databases), 3923 publications remained (see Figure 1). All publications were screened and if an article's status for in- or exclusion was not distinct, it was included in the next screening step. First, we conducted a screening of title, and after this step, 3440 studies were excluded, leaving 483 studies to be screened for their abstracts. The same selection procedure was followed, and 217 studies were removed in the screening of the abstracts. In the final step, we screened the remaining 266 studies for their full texts. We further excluded 102 studies, and 16 full texts were still missing after a hand search, which resulted in 148 studies that met the inclusion criteria and were available for further analysis in our study. As some information was missing about instruments found in the process, we added a hand search after the initial search and screening procedures, in those cases where information about the instruments was essential to answering our research questions. For an overview of the screening process, see Figure 1.

Sample description

From the 148 studies we identified, 145 were written in English and three in German. The authors of all identified articles were mainly from the USA (46.6%), followed by Canada (8.1%) and the Netherlands (5.8%). In total, authors from 28 different countries are represented in this review. With the duplicates taken into account, EBSCOhost was the database with the most articles (*n* = 2407), followed by Web of Science (*n* = 1646) and then FIS Bildung (*n* = 158). EBSCOhost provided 188 articles that remained in our study after the screening process, Web of Science identified 50 articles, and FIS Bildung identified 53 articles, making EBSCOhost the database that provided the most suitable studies for our research questions (64.8%).

Results

Overview of the reviewed studies and instruments

The current review investigates which instruments are used in research to assess global and/or math-specific interactions and how they differ according to certain analysis criteria. In the following section, we will address the first research question: which instruments are used in research to assess global and/or math-specific teacher–child interaction quality in early childcare settings?

In total, 55 instruments that assess global and/or math-specific interaction quality were identified in 148 different studies (see Table 2; for information on original author(s) and full name of the instruments, see Table 3 in the Appendix; for information on author(s), year, and country of identified studies, see Table 5 in the Appendix). The most frequently used instrument was ECERS-R (22.4%), followed by CLASS Pre-K (15.7%) and CIS (7.6%). ECERS-R assesses both global and math-specific interactions between children and their teachers, and the other two instruments assess global interactions. Overall, we identified 39 instruments assessing exclusively global interaction quality (70.9%), 13 instrument assessing both global and math-specific interaction quality (23.6%), and only three instruments assessing on math-specific interactions without explicit global facets (ECERS-E, KES-E, TBRS) (5.5%).

Instruments for assessing math-specific interactions

The instruments assessing math-specific interactions show different ways of conceptualizing mathematical aspects. Most instruments capture math-specific interactions as one construct with a clear mathematical focus, which means that those interactions are assessed separately from global or other specific interactions (e.g., EAS, ECCOM, ECERS-R, KES-R, ITERS-3). More recently developed instruments differentiate among several math-specific aspects and therefore have higher levels of specificity (e.g., COEMET, ECERS-3, ECERS-E, KES-E). A huge difference can be noticed regarding the understanding of what kind of interactions do belong to math-specific interactions: some instruments have a detailed view on mathematics as they are based on a broad understanding of mathematics, which includes interactions concerning, for example, shape, space, time, matching, sorting, and comparing (among math-specific interactions in a narrow sense, such as counting) (e.g., ABC checklist, ECERS-E, KES-E). Other instruments capture math-specific aspects at a more general level, as they contain domain-specific interactions that include some mathematical activities, without focusing on them specifically (e.g., IMCEIC). The (sub-)scales and items of the instruments are described in the respective section below.

Analysis criteria

The following section presents the results regarding the second research question: how do the instruments differ on the following criteria: (sub)scales, items, method, procedure, children's age, reported reliability, and reported validity? Table 2 gives a short overview of the instruments, for detailed information on the reliability and validity, see Table 3 and Table 4 in the Appendix.

(Sub)scales

The identified instruments include different aspects that, depending on the instrument, are described as domains, scales, subscales, or categories. In this section, we refer to them as 'scales' (in the extended overview in Table 3 in the Appendix, the term that was used in the studies is given). With respect to the scales, a notable difference between the instruments is that they draw on different theoretical and conceptual assumptions and therefore assess global and/or math-specific interactions in different ways. Two frameworks that are reflected in the scales can be delineated here: one considers interaction behavior as part of overall ECEC quality and thus as one aspect of many (e.g., ITERS, ECERS, EQOS), while the other focuses on interaction behavior and distinguishes different facets of this behavior (e.g., CLASS, ECCOM, PICCOLO). We were able to identify a wide variety of scales per instrument, ranging from one (CLASS Infant, NCKO Sensitivity Scale) to twelve (MLERS). Despite the great variability of each instrument's scales, some similarities can be noticed. 'Instrument-families', like ITERS/ECERS and CLASS, use different instruments to assess interaction behavior (and other scales) in different age groups, but they follow the same logic and are therefore similar to each other. There are also similarities between different instruments, for example, between CLASS Pre-K and PICCOLO, which both assess parallel aspects of global interaction quality (i.e., emotional supportive behavior and cognitively stimulating behavior) without belonging to an 'instrument-family'. Regarding instruments that assess math-specific interactions, ECERS-E and KES-E contain the scale 'Mathematics' and COEMET contains the scale 'Specific math activities' (with several subscales).

Items

The identified instruments define certain items or indicators to assess the aspect of interest. An item represents the smallest conceptual unit of teacher–child interaction that can be observed. Some instruments also establish indicators within their items to define possible ways to realize high-quality interactions. As the terms differ depending on the instrument, we will use the term 'item' for interaction behavior that is directly assessed by the instrument (the term that was used by the authors is provided in the extended overview in Table 3 in the Appendix). The number of items per instrument ranges from four (CLASS Infant) to over 100 (EQOS). Some items are provided with indicators, which serve different purposes depending on the instrument: in some cases (e.g., CLASS), they are not assessed but rather serve as guidance to describe the item in more detail, in other cases (e.g., ITERS/ECERS), they are directly assessed and are a basis to rate the initial item. Depending on the scales that the instrument utilizes, the items are different in some cases and similar in others. The main difference between the items of the instruments is the extent to which the observer has to rate interactions at an abstract level. However, in some studies, the items are not illustrated and could therefore not be evaluated (e.g., COS, EQOS, TBRS). In terms of math-specific interactions, most instruments include items labeled 'Mathematics' (EAS), 'Math instruction' (ECCOM), 'Math/number' (ECERS-R, KES-R, ITERS-3), or 'Mathematical-cognitive challenges' (DO-RESI-E-KiGs). ECERS-3 differentiates among math-specific aspects by using the items 'Math materials and activities', 'Math in daily events', and 'Understanding written numbers'. ECERS-E and KES-E contain comprehensive items concerning math-specificity: 'Counting and the application of counting', 'Reading and writing simple numbers', 'Mathematical activities: shape/space', and 'Mathematical activities: matching/sorting/comparing'. The ABC checklist also consideres a broader understanding of mathematics by using the items 'Arranges learning of number' and

‘Arranges learning of space and time’. COEMET has a huge math-specific focus with detailed items concerning math-specific interactions. TBRS differentiates the quality and quantity of math-specific interactions by using the items ‘Math quality’ and ‘Math quantity’. However, in some studies, the items are not illustrated and could therefore not be evaluated (e.g., COS, EQOS, TBRS).

Method

In most cases, a rating scale is used to assess teacher–child interaction quality. Usually, interaction behavior is rated on a 7-point (45.5%), a 5-point (18.2%), or a 4-point (12.7%) scale. These rating methods typically indicate the quality of interactions, ranging from ‘low/inadequate/negative behavior’ to ‘high/excellent/positive behavior’. Other, relatively rare methods indicate percentages of the observed behavior or use a dichotomous yes-no rating scale. In some cases, a stop-scoring approach is used, which means that items in a higher quality range are not assessed if items in a lower-quality range do not apply (e.g., ITERS/ECERS). Most of the identified instruments provide a quality description for each rating number (e.g., CIP, CIS, COS, EQOS, PICCOLO). However, other rating scales do not have concrete indicators for items, and instead the observer has to rate them on an abstract level. We noticed that the level of inference, meaning the ability of perceptiveness to go from the observed teacher–child interaction to a concrete decision on the rating, is particularly high for the CLASS instruments. Despite their high abstraction, those instruments do not offer quality descriptions for every possible rating number that the observer can apply. Although the rating methods of the instruments initially seem similar, a closer look at their precision of descriptions and level of inference reveals notable differences.

Procedure

The instruments differ in how long the observations last and whether they are conducted live in the childcare setting or afterward through video recordings. Most instruments base their scoring on observation periods ranging from ten minutes to six or more hours and are conducted over the course of one day. Only two instruments divide the observation over more days (ABC Checklist, CCIS), taking day-to-day variation in interactions into account. Most instruments involve live assessment (67.3%), whereas 23.6% involve observation afterward via video and 7.3% involve both. One instrument uses an event-coding approach (EmoReg), and one instrument uses a thin-slice methodology by assessing interaction behavior from a 5-minute video (ECS). All but four of the instruments assess interaction quality during a regular day. Of those four, one instrument observes a semi-structured situation (Multi-categorical coding scheme), and three instruments assess a structured situation (Evaluation of the Mentalization of the Significant Caregiver, Language samples of educator-child interactions, TC-SPT). An additional interview with teachers, directors, and/or parents can be conducted with 17 instruments.

Age of children in the early childcare setting

In the studies included in our analysis, all ages from birth to school-entry were represented. Three instruments offer different versions of the same instrument for younger and older children (ABC checklist, AQI, Evaluation of the Mentalization of the Significant Caregiver), and three instruments also assess interactions between school-teachers and school-children (COS, DO-RESI-E-KiGs, SBTR). Thirty-four instruments (61.8%) focus on interaction quality with children younger than 36 months, and 37 instruments (67.3%) assess interaction quality between teachers and children above 36 months, as twenty instruments (36.4%) are suitable for both, children younger and older than three years.

Reported reliability

Inter-rater agreement was reported in most of the studies, although the reported statistical parameters differed. Inter-rater reliability ranged from $r = .74$ (ECCOM; Lerkkanen et al. 2012) to $r = .99$ (Caregiver-Child Interaction Behaviors; Beller et al. 1996). Some studies used Cohen’s kappa, reporting results of $\kappa = .55$ (Evaluation of the Mentalization of the Significant Caregiver; Farkas 2019) to $\kappa = .96$ (CIS; Manlove, Vazquez, and Vernon-Feagans 2008). Other studies used intra-class correlations, reporting results of ICC = .63 (TC-SPT; Locasale-Crouch et al. 2018) to ICC = .94 (CIP; Bjørnestad et al. 2020). However, some studies reported the average, and others reported the range; some reported the exact agreement, and others reported the agreement within one point of error on their respective rating scale. Some studies did 100% double-coding, and some used a lower sample size (10%), making it difficult to compare inter-rater reliabilities. Some studies required an inter-rater reliability of at least 80% before starting data collection, but they did not report reliability for the study sample (e.g., studies using DO-RESI-E-KiGs and SSTEW). For a few instruments, it was obligatory to obtain a certificate prior to data collection that confirmed an 80% agreement within one point with the rating 'gold standard' of experts (e.g., CLASS, ITERS/ECERS). Internal consistency was mostly reported using Cronbach's alpha for the scales. For detailed information on the reported reliability, see Table 4 in the Appendix.

Reported validity

In this section, we focus on the construct validity reported in the studies. As the information on convergent validity is extensive, it can be found in the Appendix. The instruments were mostly validated with other instruments that are widely used in this field of research (namely CLASS and ITERS/ECERS). This leaves these two instruments with the most detailed information on convergent validity. The statistical methods that were used to examine construct validity were exploratory or confirmatory factor analysis. For CLASS Toddler, the results were mixed: one study revealed a good model fit of the proposed two factors in the manual (van Schaik, Leseman, and Haan 2018), whereas another study suggested a three-factor solution instead (Slot et al. 2017). Although some studies indicate that the three-factor solution of CLASS Pre-K proposed in the manual did not fit the data well (Hanno et al. 2020), most studies replicated a good model fit of the three factors, at least with a few modifications (Bihler et al. 2018; Hu et al. 2016; Pakarinen et al. 2010). For ITERS-R, one study revealed that the scale measured one global aspect rather than different subscales (Bisceglia et al. 2009). The proposed factors of ECERS-R showed a feasible to good model fit in most studies (Bishop 2013; Gordon et al. 2013). In contrast, some studies revealed that a two-factor solution (Hadeed 2014) or a three-factor solution (Perlman, Zellman, and Le 2004) fit the data better. For ECERS-3, one study revealed a four-factor solution instead of the proposed six-factor solution (Early et al. 2018). For detailed information about the reported validity of these and other instruments, see Table 4 in the Appendix.

Discussion

This review provides an overview of the instruments used to assess quality of teacher–child interactions in early childcare settings. We focused on instruments that are used in research to assess global and/or math-specific interaction quality (first research question).

Overall, we identified 55 instruments in 148 different studies. Thirteen instruments assessed both global and math-specific interaction quality, 39 instruments assessed exclusively global interactions, and three instruments assessed exclusively math-specific interactions. Although few instruments exclusively measured math-specific behavior, recent years have seen a trend of more newly developed instruments that more frequently capture math-specific interactions, at least together with global behavior. Our findings show that even though a number of different instruments are used to assess global and math-specific interactions, the most commonly used instruments by far are ECERS-R and CLASS Pre-K. Assessing both global and (math-)specific aspects of interactions is essential to differentiating between the effects of different aspects of teachers' behavior on various child outcomes (Bleses et al. 2020; Hamre et al. 2014; Kluczniok et al. 2013; Lehrl et al. 2020; LeFevre et al. 2010; Sylva et al. 2006). Identifying beneficial interactions through appropriate instruments is important for policy and practice, as it offers an opportunity to best support children in different domains of development and learning.

Math-specific instruments that already exist primarily assess teacher–child interactions with children older than three years (thirteen instruments for children older than three years vs. seven instruments for children younger than three years). This might be because the focus of research on math-specificity in the past few years has primarily been on children before school-entry (Boonen, Kolkman, and Kroesbergen 2011; Braak et al. 2022; Curby and Chavez 2013; Klibanoff et al. 2006; Lehrl et al. 2017; Möhring et al. 2021; Pelkowski et al. 2019). Despite the fact that not many instruments measure math-specific interactions with very young children yet, an increasing interest in mathematical development and math-specific support in this age group can be seen in current research (Burghardt et al. 2020; Reikerås and Salomonsen 2019; Salminen et al. 2021). Still, instruments with a math-specific component are underrepresented. Although it is possible to contemplate math-specific interactions during an assessment, an instrument incorporating math-specific interactions in its theoretical-conceptual framework (along with global facets of interactions) is needed, especially for math-specific interactions between teachers and very young children.

The second research question addressed how the instruments differ along certain criteria ((sub-)scales, items, method, procedure, children's age, reported reliability, and reported validity). Most instruments assess either interaction quality together with other aspects as one part of overall ECEC quality, or focus particularly on interaction behavior and measure different (mostly global) aspects, such as emotional support and cognitive stimulation. Some instruments also adress the domain-specificity of interactions and take math-specific (and other domain-specific) interactions into account. This demonstrates the large variety of theoretical-conceptual frameworks for assessing interaction quality, resulting in each instruments having a (slightly) different level of detail. While one group of instruments generally locates interaction quality within the framework model of overall quality (and, in line with Bronfenbrenner's bio-ecological theory, also includes, e.g., available materials in the classroom), the other group of instruments focuses more specifically on interaction quality as a global construct and differentiates between various facets. In particular, in terms of the instruments ITERS-R and ECERS-R, which are widely used to assess aspects of quality in ECEC, one critique has been that they only assess teacher–child interaction quality to a limited degree and that they take other areas of quality into greater account (Bisceglia et al. 2009; Fujimoto et al. 2018; Gordon et al. 2013; Gordon et al. 2015).

As for the reliability of the identified instruments, inter-rater agreements seem to be at a good level in most cases. However, not every study reports reliability, and if they do, the reported statistical parameters differ between studies, making it difficult to compare them. A consistent way of reporting reliabilities would have the advantage of enabling better evaluation of statistical parameters. As for construct validity, the identified studies revealed mostly consistent (e.g., CLASS Pre-K, ECERS-R) but also mixed results regarding the factor structure (e.g., CLASS Toddler, ITERS-R, ECERS-3). This reflects the empirical diversity of the theoretical and conceptual foundations of the instruments. In terms of convergent validity, we noticed that most instruments are validated with the ITERS/ECERS or CLASS instruments. This could be because these instruments are the most widely used, which makes it easier to compare convergent validity among instruments. At the same point, this could result in ITERS/ECERS and CLASS being used more frequently than other instruments, making them benchmarks in early childhood research.

Even though reviews on this topic differ in their searching and screening processes, in our review we noticed an increase in instruments identified, in comparison to other overviews from recent years (e.g., CKEQRS-TV (2019) and MELE (2017) are relatively new instruments). This could be an indicator of the growing research interest in instruments assessing teacher–child interaction quality in ECEC, as it is considered a necessary precondition for research on child development (Anders et al. 2013; Belsky et al. 2007; Burchinal et al. 2014; Peisner-Feinberg et al. 2001; Lehrl, Kluczniok, and Roßbach 2016; Mashburn et al. 2008; Sylva et al. 2011; Vandell et al. 2010). This also provides the opportunity to use current instruments in research, since, for example, the still-widely-used CIS has been criticized as somewhat outdated (published in 1989) (Colwell et al. 2013).

Nevertheless, the limitations of this systematic literature review must be discussed. The searching process was somewhat restricted, as this study only included instruments with an observational approach and that address interactions between teachers and all children in group-based situations. Additionally, no articles in other languages than English or German were included, resulting in a selection bias of studies, as articles written in other languages are not considered in the overview. From the studies we identified, a great number were from the USA. One possible reason could be that well-known instruments were initially developed in the USA in the context of large-scale studies (e.g., Head Start, NICHD-SECCYD), thus, several publications using those instruments are published by US-American authors. Consequently, publication bias must be taken into account, as studies conducted in US-American childcare settings might be overrepresented in our review, giving disproportionate weight to US-based views on quality in ECEC and on best practice in quality observations.

Conclusion

This systematic literature review highlights the variety of potential instruments for capturing interaction quality, each with a different methodological approach and (slightly) different content. Two points are worth noting here. First, despite the variety of instruments, the same instruments are often used in early childhood research (e.g., CLASS, ECERS). It is not clear whether this is always done for reasons of comparability and the good quality criteria of these instruments, or whether other instruments might be more suitable for certain research questions. Second, the range of existing instruments does not cover all content areas and age groups to the same extent: for example, math-specific instruments focus primarily on interactions with children before they enter school and not so much on children younger than three years old. Even though it has its limitations, this review offers a detailed overview of the instruments assessing global and/or math-specific interaction quality. These instruments make it possible to identify beneficial interactions and thus to investigate their predictors and their (differential) effects on child development. This provides the opportunity to learn more about how children can best be supported in their development.

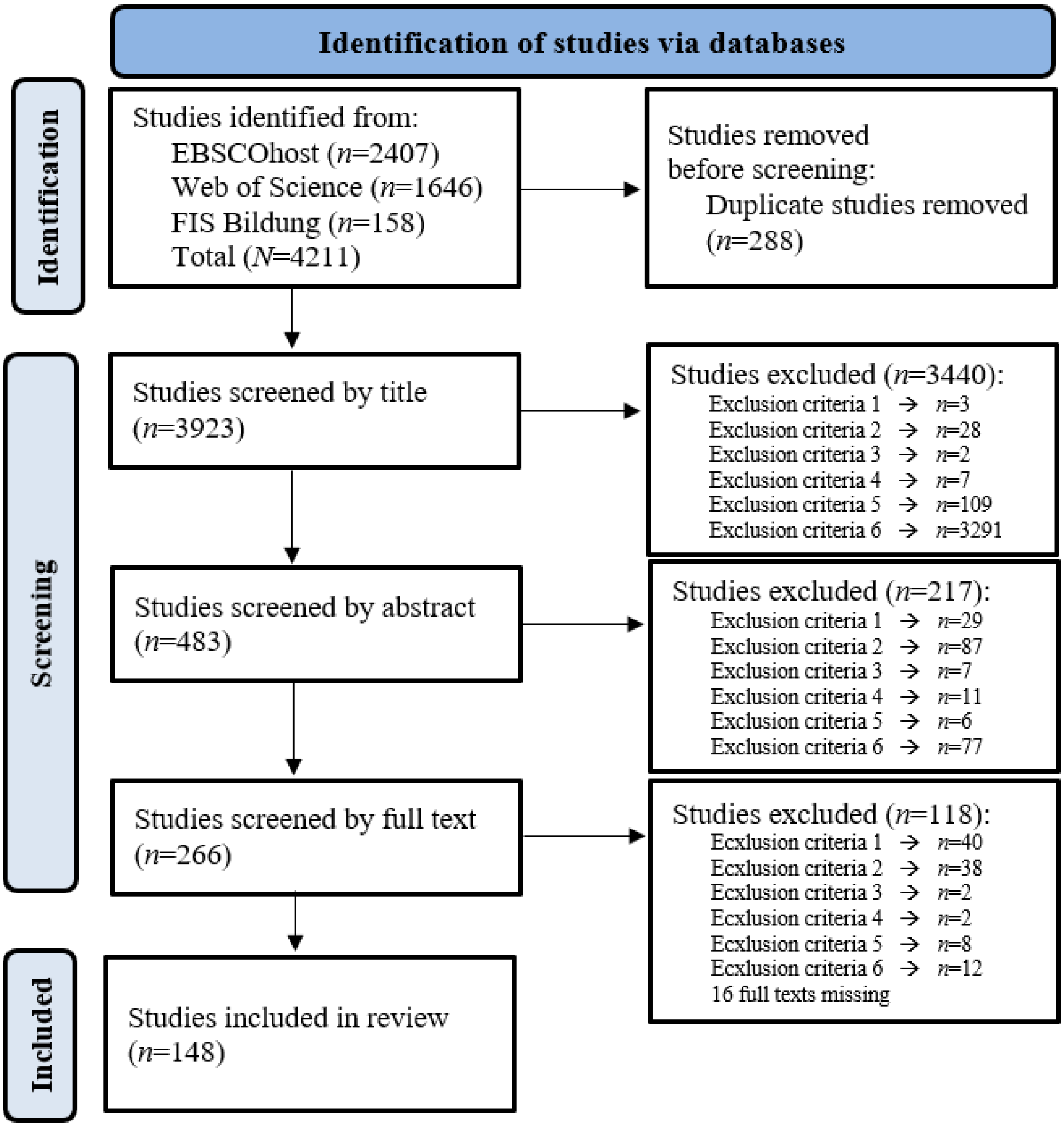
Disclosure statement

The authors report that there are no competing interests to declare.

Table 1. Search terms.

	AND	AND	AND	AND	AND	NOT
OR	instrument*	measur*	crèche*	interact*	stimulat*	disorder
OR	manual*	assess*	classroom*	quality*	support*	syndrome
OR	scale*	scor*	kindergarten	process*	enrich*	autism
OR	inventory	rat*	preschool		sensitiv*	ADHD
OR			infant		responsiv*	deaf
OR			toddler		global*	derma*
OR			child*		domain*	medic*
OR			"early childhood"		math*	pediatr*
OR			"child care"		specific*	neuro*
OR			"day care center"			health*
OR						gene*
OR						school*
OR						youth*
OR						parent*
OR						mother
OR						famil*
OR						home

Note. For FIS Bildung, terms were translated into German and the ‘NOT’ category was not used.



Note. Exclusion criteria 1=assessment by non-objective persons; Exclusion criteria 2=assessment focussing on school children/non-pedagogical persons; Exclusion criteria 3=assessment of non-process quality dimensions; Exclusion criteria 4=assessment of non-global/non-math-specific interaction quality; Exclusion criteria 5=(systematic) reviews/meta-analysis; Exclusion criteria 6=met more than one criterion.

Figure 1. Flowchart of screening process for identified studies.

Table 2. Overview of instruments (short version).

	Instrument	Full name	Targeted domain
1	ABC Checklist	Adults Behaviors in Caregiving Checklist	G and M
2	Adult Engagement Scale*	Adult Engagement Scale	G
3	Adult Interaction Scale	Adult Interaction Scale	G
4	AMSS*	Ainsworth's Maternal Sensitivity Scales	G
5	AQI	Assessment for Quality Improvement	G and M
6	Caregiver-Child Interaction Behaviors	Caregiver-Child Interaction Behaviors	G
7	CCIS	Caregiver Child Interaction Scale	G
8	CECERS	Chinese Early Childhood Environment Rating Scale	G and M
9	CIP	Caregiver Interaction Profile	G
10	CIS	Arnett Caregiver Interaction Scale	G
11	CKEQRS-TV	Chinese Kindergarten Education Quality Rating Scale - Teacher Version	G
12	CLASS	Classroom Assessment Scoring System	G
13	CLASS Infant*	Classroom Assessment Scoring System Infant Version	G
14	CLASS Toddler	Classroom Assessment Scoring System Toddler Version	G
15	CLASS Pre-K	Classroom Assessment Scoring System Pre-Kindergarten Version	G
16	COEMET	Classroom Observation of Early Mathematics Environment and Teaching	G and M
17	COS	Early Childhood Classroom Observation Scale	G
18	DO-RESI-E-KiGs	Dortmund rating scale to measure language-related interactions - extended version for daycare centers and elementary schools**	G and M
19	EA Scales, Version 3	Emotional Availability Scales, Version 3	G
20	EAS	Emerging Academics Snapshot	G and M
21	ECCOM	Early Childhood Classroom Observation Measure	G and M
22	ECERS	Early Childhood Environment Rating Scale	G
23	ECERS-R	Early Childhood Environment Rating Scale Revised Edition	G and M
24	ECERS-E	Early Childhood Environment Rating Scale Extension	M
25	ECERS-3	Early Childhood Environment Rating Scale Third Edition	G and M
26	ECS	Educator Cognitive Sensitivity Scale	G
27	EmoReg	EmoReg	G

28	EQOS	Educational Quality Observation Scale, Preschool Version	G
29	ESA	Adult Sensitivity Scale**	G
30	Evaluation of the Mentalization of the Significant Caregiver	Evaluation of the Mentalization of the Significant Caregiver	G
31	IMCEIC*	Instrument for Measuring Quality of Early Childhood Education in Colombia	G and M
32	ISSA observation protocol	International Step by Step Association observation protocol	G
33	ITERS	Infant/Toddler Environment Rating Scale	G
34	ITERS-R	Infant/Toddler Environment Rating Scale Revised Edition	G
35	ITERS-3	Infant/Toddler Environment Rating Scale Third Edition	G and M
36	KERS	Kindergarten Environment Rating Scale	G
37	KES-R	Kindergarten Scale Revised Edition**	G and M
38	KES-E	Kindergarten Scale Extension**	M
39	KRIPS-R	Infant/Toddler Scale Revised Edition**	G
40	Language samples of educator-child interactions	Language samples of educator-child interactions	G
41	MELE*	Measure of Early Learning Environments	G and M
42	MLERS	Mediated Learning Rating Scale	G
43	Multi-categorical coding scheme	Multi-categorical coding scheme	G
44	NCKO Sensitivity Scale	NCKO Sensitivity Scale	G
45	OPRS	Outdoor Play Rating Scale	G
46	PICCOLO	Parenting Interactions with Children: Checklist of Observations Linked to Outcomes	G
47	PITC PARS	Program for Infant and Toddler Care Program Assessment Rating Scale	G
48	SBTR	Student Behavior Teacher Response	G
49	SSTEW	Sustained Shared Thinking and Emotional Wellbeing	G
50	TBRS*	Teacher Behavior Rating Scale	M
51	TCIS	Teacher Child Interaction Scale	G
52	TC-SPT	Teacher-Child Structured Play Task	G
53	TILRS (French: GEIPLE)	Teacher Interaction and Language Rating Scale	G
54	TSRS	Teacher Styles Rating Scale	G
55	Watts’s Human Interaction Scale	Watts's Human Interaction Scale	G

Note. *additional hand search; **translated from German into English; G=global interaction quality; M=math-specific interaction quality.

References

Ambady, Nalini, Frank J. Bernieri, and Jennifer A. Richeson. 2000. “Toward a Histology of Social Behavior: Judgmental Accuracy from Thin Slices of the Behavioral Stream.” In *M. P. Zanna (Ed.), Advances in Experimental Social Psychology: Vol. 32*: 201–271. San Diego: Academic Press.

Anders, Yvonne, Christiane Grosse, Hans-Günther Roßbach, Susanne Ebert, and Sabine Weinert. 2013. “Preschool and Primary School Influences on the Development of Children's Early Numeracy Skills between the Ages of 3 and 7 Years in Germany.” *School Effectiveness and School Improvement* 24 (2): 195–211. doi:10.1080/09243453.2012.749794.

Anders, Yvonne, Hans-Günther Roßbach, Sabine Weinert, Susanne Ebert, Susanne Kuger, Simone Lehrl, and Jutta von Maurice. 2012. “Home and Preschool Learning Environments and Their Relations to the Development of Early Numeracy Skills.” *Early Childhood Research Quarterly* 27 (2): 231–44. doi:10.1016/j.ecresq.2011.08.003.

Beller, Emanuel K., Marita Stahnke, Petra Butz, Walter Stahl, and Holger Wessels. 1996. “Two Measures of the Quality of Group Care for Infants and Toddlers.” *European Journal of Psychology of Education* 11 (2): 151–167.

Belsky, Jay, Deborah L. Vandell, Margaret Burchinal, K. A. Clarke-Stewart, Kathleen McCartney, and Margaret T. Owen. 2007. “Are There Long-Term Effects of Early Child Care?” *Child Development* 78 (2): 681–701. doi:10.1111/j.1467-8624.2007.01021.x.

Berk, Laura E., and Adam Winsler. 1995. *Scaffolding Children's Learning: Vygotsky and Early Childhood Education. NAEYC Research into Practice Series. Vol. 7.* Washington, DC: National Association for the Education of Young Children.

Bichay-Awadalla, Krystal, and Rebecca J. Bulotsky-Shearer. 2021. “Examining the Factor Structure of the Classroom Assessment Scoring System Toddler (CLASS-T) in Early Head Start and Subsidized Child Care Classrooms.” *Early Education and Development*, 1–17. doi:10.1080/10409289.2021.1887700.

Bihler, Lilly-Marlen, Alexandru Agache, Katharina Kohl, Jessica A. Willard, and Birgit Leyendecker. 2018. “Factor Analysis of the Classroom Assessment Scoring System Replicates the Three Domain Structure and Reveals No Support for the Bifactor Model in German Preschools.” *Frontiers in Psychology* 9: 1232. doi:10.3389/fpsyg.2018.01232.

Bisceglia, Rossana, Michal Perlman, Diana Schaack, and Jennifer Jenkins. 2009. “Examining the Psychometric Properties of the Infant-Toddler Environment Rating Scale-Revised Edition in a High-Stakes Context.” *Early Childhood Research Quarterly* 24 (2): 121–132. doi:10.1016/j.ecresq.2009.02.001.

Bishop, Crystal D. 2013. “Using the Early Childhood Longitudinal Study Birth Cohort Data to Explore Measurement Invariance for the Early Childhood Environment Rating Scale-revised and the Arnett Caregiver Interaction Scale.” Gainesville, FL: University of Florida.

Bjørnestad, Elisabeth, Martine L. Broekhuizen, Ellen Os, and Anne G. Baustad. 2020. “Interaction Quality in Norwegian ECEC for Toddlers Measured with the Caregiver Interaction Profile (CIP) Scales.” *Scandinavian Journal of Educational Research* 64 (6): 901–20. doi:10.1080/00313831.2019.1639813.

Bleses, Dorthe, Peter Jensen, Pauline L. Slot, and Laura Justice. 2020. “Low-Cost Teacher-Implemented Intervention Improves Toddlers’ Language and Math Skills.” *Early Childhood Research Quarterly* 53: 64–76. doi:10.1016/j.ecresq.2020.03.001.

Boonen, Anton J. H., Meijke E. Kolkman, and Evelyn H. Kroesbergen. 2011. “The Relation between Teachers' Math Talk and the Acquisition of Number Sense within Kindergarten Classrooms.” *Journal of School Psychology* 49 (3): 281–299. doi:10.1016/j.jsp.2011.03.002.

Braak, Dieuwer, Ragnhild Lenes, David J. Purpura, Sara A. Schmitt, and Ingunn Størksen. 2022. “Why do Early Mathematics Skills Predict Later Mathematics and Reading Schievement? The Role of Executive Function.” *Journal of Experimental Child Psychology* 214: 105306. doi:10.1016/j.jecp.2021.105306.

Bronfenbrenner, Urie, and Pamela A. Morris. 1998. "The Ecology of Developmental Processes." In *W. Damon & R. M. Lerner (Eds.), Handbook of Child Psychology: Theoretical Models of Human Development*: 993–1028. Hoboken, NJ: John Wiley & Sons.

Brunsek, Ashley, Michal Perlman, Olesya Falenchuk, Evelyn McMullen, Brooke Fletcher, and Prakesh S. Shah. 2017. “The Relationship between the Early Childhood Environment Rating Scale and Its Revised Form and Child Outcomes: A Systematic Review and Meta-Analysis.” *PloS one* 12 (6): e0178512. doi:10.1371/journal.pone.0178512.

Burchinal, Margaret, Lynne Vernon-Feagans, Virginia Vitiello, and Mark Greenberg. 2014. “Thresholds in the Association between Child Care Quality and Child Outcomes in Rural Preschool Children.” *Early Childhood Research Quarterly* 29 (1): 41–51. doi:10.1016/j.ecresq.2013.09.004.

Burger, Kaspar. 2010. “How Does Early Childhood Care and Education Affect Cognitive Development? An International Review of the Effects of Early Interventions for Children from Different Social Backgrounds.” *Early Childhood Research Quarterly* 25 (2): 140–65. doi:10.1016/j.ecresq.2009.11.001.

Burghardt, Lars, Anja Linberg, Simone Lehrl, and Kira Konrad-Ristau. 2020. “The Relevance of the Early Years Home and Institutional Learning Environments for Early Mathematical Competencies.” *Journal for Educational Research Online* 12 (3): 103–25. doi:10.25656/01:21188.

Colwell, Nicole, Rachel A. Gordon, Ken Fujimoto, Robert Kaestner, and Sanders Korenman. 2013. “New Evidence on the Validity of the Arnett Caregiver Interaction Scale: Results from the Early Childhood Longitudinal Study-Birth Cohort.” *Early Childhood Research Quarterly* 28 (2): 218–33. doi:10.1016/j.ecresq.2012.12.004.

Curby, Timothy W., and Catharine Chavez. 2013. "Examining CLASS Dimensions as Predictors of Pre-K Children's Development of Language, Literacy, and Mathematics." *Dialog* 16 (2): 1–17.

Daucourt, Mia C., Amy R. Napoli, Jamie M. Quinn, Sarah G. Wood, and Sara A. Hart. 2021. "The Home Math Environment and Math Achievement: A Meta-Analysis." *Psychological Bulletin* 147 (6): 565–596. doi:10.1037/bul0000330.

Downer, Jason T., Leslie M. Booren, Olivia K. Lima, Amy E. Luckner, and Robert C. Pianta. 2010. "The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary Reliability and Validity of a System for Observing Preschoolers' Competence in Classroom Interactions." *Early Childhood Research Quarterly* 25 (1): 1–16. doi:10.1016/j.ecresq.2009.08.004.

Early, Diane M., John Sideris, Jennifer Neitzel, Doré R. LaForett, and Chelsea G. Nehler. 2018. "Factor Structure and Validity of the Early Childhood Environment Rating Scale – Third Edition (ECERS-3)." *Early Childhood Research Quarterly* 44: 242–256. doi:10.1016/j.ecresq.2018.04.009.

Farkas, Chamarrita. 2019. "Teachers' Competence Profiles and Their Relation to Language Development in Chilean Children at 12 and 30 Months." *Early Education and Development* 30 (6): 743–764. doi:10.1080/10409289.2019.1597580.

Fassnacht, G. 1995. *Systematische Verhaltensbeobachtung:* Munich: Reinhardt.

Fujimoto, Ken A., Rachel A. Gordon, Fang Peng, and Kerry G. Hofer. 2018. "Examining the Category Functioning of the ECERS-R across Eight Data Sets." *AERA Open* 4 (1): 1–16. doi:10.1177/2332858418758299.

García-Carrión, Rocío, and Lourdes Villardón-Gallego. 2016. "Dialogue and Interaction in Early Childhood Education: A Systematic Review." *REMIE* 6 (1): 51–76. doi:10.17583/remie.2016.1919.

Gordon, Rachel A., Ken Fujimoto, Robert Kaestner, Sanders Korenman, and Kristin Abner. 2013. "An Assessment of the Validity of the ECERS-R with Implications for Measures of Child Care Quality and Relations to Child Development." *Developmental Psychology* 49 (1): 146–160. doi:10.1037/a0027899.

Gordon, Rachel A., Kerry G. Hofer, Ken A. Fujimoto, Nicole Risk, Robert Kaestner, and Sanders Korenman. 2015. "Identifying High-Quality Preschool Programs: New Evidence on the Validity of the Early Childhood Environment Rating Scale–Revised (ECERS-R) in Relation to School Readiness Goals." *Early Education and Development* 26 (8): 1086–1110. doi:10.1080/10409289.2015.1036348.

Grant, Maria J., and Andrew Booth. 2009. "A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies." *Health Information and Libraries Journal* 26 (2): 91–108. doi:10.1111/j.1471-1842.2009.00848.x.

Greve, Werner, and Dirk Wentura. 1997. *Wissenschaftliche Beobachtung: Eine Einführung* (2. Auflage). Weinheim: Beltz.

Hadeed, Julie. 2014. “Reliability and Validity of the Early Childhood Environment Rating Scale, Revised Edition, ECERS-R in Arabic.” *Early Child Development and Care* 184 (6): 819–842. doi:10.1080/03004430.2013.818991.

Halle, Tamara, Jessica E. Vick Whittaker, and Rachel Anderson. 2010. *Quality in Early Childhood Care and Education Settings: A Compendium of Measures.* Washington, DC: Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Hamre, Bridget, Bridget Hatfield, Robert Pianta, and Faiza Jamil. 2014. “Evidence for General and Domain-Specific Elements of Teacher-Child Interactions: Associations with Preschool Children's Development.” *Child Development* 85 (3): 1257–1274. doi:10.1111/cdev.12184.

Hanno, Emily C., Kathryn E. Gonzalez, Rebecca B. Lebowitz, Dana C. McCoy, Armida Lizarraga, and Camilla Korder Fort. 2020. “Structural and Process Quality Features in Peruvian Early Childhood Education Settings.” *Journal of Applied Developmental Psychology* 67: 101105. doi:10.1016/j.appdev.2019.101105.

Hooper, Alison. 2017. “Exploring the Relationship Between Global Quality and Group Engagement in Toddler Child Care Classrooms.” *Journal of Research in Childhood Education* 31 (2): 215–226. doi:10.1080/02568543.2016.1273287.

Hu, Bi Y., Xitao Fan, Chuanhua Gu, and Ning Yang. 2016. “Applicability of the Classroom Assessment Scoring System in Chinese Preschools Based on Psychometric Evidence.” *Early Education and Development* 27 (5): 714–734. doi:10.1080/10409289.2016.1113069.

Hu, Bi Y., Zanmei Song, Shuang Wang, and Jennifer Locasale-Crouch. 2019. “Global Quality Profiles in Chinese Early Care Classrooms: Evidence from the Shandong Province.” *Children and Youth Services Review* 101: 157–164. doi:10.1016/j.childyouth.2019.03.056.

Ishimine, Karin, and Collette Tayler. 2014. “Assessing Quality in Early Childhood Education and Care.” *European Journal of Education* 49 (2): 272–290. doi:10.1111/ejed.12043.

Jamison, Kristen R., Sonia Q. Cabell, Jennifer Locasale-Crouch, Bridget K. Hamre, and Robert C. Pianta. 2014. “CLASS–Infant: An Observational Measure for Assessing Teacher–Infant Interactions in Center-Based Child Care.” *Early Education and Development* 25 (4): 553–572. doi:10.1080/10409289.2013.822239.

Klibanoff, Raquel S., Susan C. Levine, Janellen Huttenlocher, Marina Vasilyeva, and Larry V. Hedges. 2006. “Preschool Children's Mathematical Knowledge: The Effect of Teacher 'Math Talk'.” *Developmental Psychology* 42 (1): 59–69. doi:10.1037/0012-1649.42.1.59.

Kluczniok, Katharina, Simone Lehrl, Susanne Kuger, and Hans-Günther Roßbach. 2013. “Quality of the Home Learning Environment during Preschool Age – Domains and Contextual Conditions.” *European Early Childhood Education Research Journal* 21 (3): 420–438. doi:10.1080/1350293X.2013.814356.

La Paro, Karen M., Bridget K. Hamre, and Robert C. Pianta. 2012. *Classroom Assessment Scoring System (CLASS) Manual, Toddler*. Baltimore, MD: Paul H. Brookes Publishing Company.

Langeloo, Annegien, Mayra Mascareño Lara, Marjolein I. Deunk, Nikolai F. Klitzing, and Jan-Willem Strijbos. 2019. "A Systematic Review of Teacher–Child Interactions with Multilingual Young Children." *Review of Educational Research* 89 (4): 536–568. doi:10.3102/0034654319855619.

LeFevre, Jo-Anne, Lisa Fast, Sheri-Lynn Skwarchuk, Brenda L. Smith-Chant, Jeffrey Bisanz, Deepthi Kamawar, and Marcie Penner-Wilger. 2010. "Pathways to Mathematics: Longitudinal Predictors of Performance." *Child Development* 81 (6): 1753–1767. doi:10.1111/j.1467-8624.2010.01508.x.

Lehrl, Simone, Susanne Ebert, Sabine Blaurock, Hans-Günther Roßbach, and Sabine Weinert. 2020. "Long-Term and Domain-Specific Relations between the Early Years Home Learning Environment and Students' Academic Outcomes in Secondary School." *School Effectiveness and School Improvement* 31 (1): 102–124. doi:10.1080/09243453.2019.1618346.

Lehrl, Simone, Katharina Kluczniok, and Hans-Günther Roßbach. 2016. "Longer-Term Associations of Preschool Education: The Predictive Role of Preschool Quality for the Development of Mathematical Skills through Elementary School." *Early Childhood Research Quarterly* 36: 475–488. doi:10.1016/j.ecresq.2016.01.013.

Lehrl, Simone, Katharina Kluczniok, Hans-Günther Roßbach, and Yvonne Anders. 2017. "Long Term Persistence of Preschool Intervention on Children's Mathematical Development: Results from the German Model Project 'Kindergarten of the Future in Bavaria'." *Global Education Review* 4 (3): 70–87.

Lerkkanen, Marja-Kristiina, Eve Kikas, Eija Pakarinen, Kaili Trossmann, Anna-Maija Poikkeus, Helena Rasku-Puttonen, Martti Siekkinen, and Jari-Erik Nurmi. 2012. "A Validation of the Early Childhood Classroom Observation Measure in Finnish and Estonian Kindergartens." *Early Education and Development* 23 (3): 323–350. doi:10.1080/10409289.2010.527222.

Linberg, Anja, Katharina Kluczniok, Lars Burghardt, and Jan-David Freund. 2017. "Quality of Toddler Childcare – Can It Be Assessed with Questionnaires?" *Early Child Development and Care* 189 (8): 1369–1383. doi:10.1080/03004430.2017.1380636.

Locasale-Crouch, Jennifer, Amanda Williford, Jessica Whittaker, Jamie DeCoster, and Pilar Alamos. 2018. "Does Fidelity of Implementation Account for Changes in Teacher-Child Interactions in a Randomized Controlled Trial of Banking Time?" *Journal of Research on Educational Effectiveness* 11 (1): 35–55. doi:10.1080/19345747.2017.1329365.

Manlove, Elizabeth E., Arcel Vazquez, and Lynne Vernon-Feagans. 2008. "The Quality of Caregiving in Child Care: Relations to Teacher Complexity of Thinking and Perceived Supportiveness of the Work Environment." *Infant and Child Development* 17 (3): 203–222. doi:10.1002/icd.547.

Mashburn, Andrew J., Robert C. Pianta, Bridget K. Hamre, Jason T. Downer, Oscar A. Barbarin, Donna Bryant, Margaret Burchinal, Diane M. Early, and Carollee Howes. 2008. “Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills.” *Child Development* 79 (3): 732–749. doi:10.1111/j.1467-8624.2008.01154.x.

Mathers, Sandra, Rosanna Singler, and Arjette Karemaker. 2012. *Improving Quality in the Early Years: A Comparison of Perspectives and Measures*. Oxford: University of Oxford.

Melhuish, Edward C., Katharina Ereky-Stevens, Konstantinos Petrogiannis, Anamaria Ariescu, Efthymia Penderi, Konstantina Rentzou, Alice Tawell, Pauline L. Slot, Martine L. Broekhuizen, and Paul P. M. Leseman. 2015. “A Review of Research on the Effects of Early Childhood Education and Care (ECEC) upon Child Development. CARE Project." Brussels: European Early Childhood Education and Care Publications.

Möhring, Wenke, Andrew D. Ribner, Robin Segerer, Melissa E. Libertus, Tobias Kahl, Larissa M. Troesch, and Alexander Grob. 2021. “Developmental Trajectories of Children's Spatial Skills: Influencing Variables and Associations with Later Mathematical Thinking.” *Learning and Instruction* 75: 101515. doi:10.1016/j.learninstruc.2021.101515.

Murphy, Nora A. 2005. “Using Thin Slices for Behavioral Coding.” *Journal of Nonverbal Behaviour* 29 (4): 235–246. doi:10.1007/s10919-005-7722-x.

Mutaf-Yıldız, Belde, Delphine Sasanguie, Bert de Smedt, and Bert Reynvoet. 2020. “Probing the Relationship between Home Numeracy and Children's Mathematical Skills: A Systematic Review.” *Frontiers in Psychology* 11: 2074. doi:10.3389/fpsyg.2020.02074.

Pakarinen, Eija, Marja-Kristiina Lerkkanen, Anna-Maija Poikkeus, Noona Kiuru, Martti Siekkinen, Helena Rasku-Puttonen, and Jari-Erik Nurmi. 2010. “A Validation of the Classroom Assessment Scoring System in Finnish Kindergartens.” *Early Education and Development* 21 (1): 95–124. doi:10.1080/10409280902858764.

Peisner-Feinberg, Ellen, M. R. Burchinal, R. M. Clifford, M. L. Culkin, C. Howes, S. L. Kagan, and N. Yazejian. 2001. “The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through Second Grade.” *Child Development* 72 (5): 1534–1553. doi:10.1111/1467-8624.00364.

Peisner-Feinberg, Ellen, Virginia Buysse, Allison Fuligni, Margaret Burchinal, Linda Espinosa, Tamara Halle, and Dina C. Castro. 2014. “Using Early Care and Education Quality Measures with Dual Language Learners: A Review of the Research.” *Early Childhood Research Quarterly* 29 (4): 786–803. doi:10.1016/j.ecresq.2014.04.013.

Pelkowski, Tracy, Everett Herman, Keith Trahan, Dana M. Winters, Cindy Tananis, Mallary I. Swartz, Nancy Bunt, and Chris Rodgick. 2019. “Fostering a 'Head Start' in Math: Observing Teachers in Early Childhood Mathematics Engagement.” *Journal of Early Childhood Teacher Education* 40 (2): 96–119. doi:10.1080/10901027.2018.1522397.

Perlman, Michal, Olesya Falenchuk, Brooke Fletcher, Evelyn McMullen, Joseph Beyene, and Prakesh S. Shah. 2016. “A Systematic Review and Meta-Analysis of a Measure of Staff/Child Interaction Quality (the Classroom Assessment Scoring System) in Early Childhood Education and Care Settings and Child Outcomes.” *PloS One* 11 (12): e0167660. doi:10.1371/journal.pone.0167660.

Perlman, Michal, Gail L. Zellman, and Vi-Nhuan Le. 2004. “Examining the Psychometric Properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R).” *Early Childhood Research Quarterly* 19 (3): 398–412. doi:10.1016/j.ecresq.2004.07.006.

Pianta, Robert C., Karen M. La Paro, and Bridget K. Hamre. 2008. *Classroom Assessment Scoring System (CLASS) Manual, Pre-K*. Baltimore, MD: Paul H. Brookes Publishing Company.

Purpura, David J., Elizabeth Day, Amy R. Napoli, and Sara A. Hart. 2017. “Identifying Domain-General and Domain-Specific Predictors of Low Mathematics Performance: A Classification and Regression Tree Analysis.” *Journal of Numerical Cognition* 3 (2): 365–399. doi:10.5964/jnc.v3i2.53.

Ramani, Geetha B., Meredith L. Rowe, Sarah H. Eason, and Kathryn A. Leech. 2015. “Math Talk during Informal Learning Activities in Head Start Families.” *Cognitive Development* 35: 15–33. doi:10.1016/j.cogdev.2014.11.002.

Rammstedt, Beatrice (Ed.). 2013. *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich: Ergebnisse von PIAAC 2012.* Münster: Waxmann.

Reikerås, Elin, and Tone Salomonsen. 2019. “Weak Mathematical Skills at an Early Age: Persistent or Temporary? Children with Weak Mathematical Skills and Their Development from Toddlers to Preschoolers.” *Early Child Development and Care* 189 (4): 670–682. doi:10.1080/03004430.2017.1337753.

Remsperger, Regina. 2013. “Das Konzept der Sensitiven Responsivität.” *Frühe Bildung* 2 (1): 12–19. doi:10.1026/2191-9186/a000072.

Riedmeier, Magdalena. 2019. “Standardisierte Verfahren zur Erhebung zielkindbezogener Prozessqualität in der Frühpädagogik.” *Frühe Bildung* 8 (3): 144–152. doi:10.1026/2191-9186/a000426.

Salminen, Jenni, Daria Khanolainen, Tuire Koponen, Minna Torppa, and Marja-Kristiina Lerkkanen. 2021. “Development of Numeracy and Literacy Skills in Early Childhood – A Longitudinal Study on the Roles of Home Environment and Familial Risk for Reading and Math Difficulties.” *Frontiers in Education* 6: 725337. doi:10.3389/feduc.2021.725337.

Siraj-Blatchford, Iram, Kathy Sylva, Stella Muttock, Rose Gilden, and Danny Bell. 2002. *Researching Effective Pedagogy in the Early Years (REPEY). Research Report 356*. London: Department for Education and Skills.

Slot, Pauline L., Jan Boom, Josje Verhagen, and Paul P. M. Leseman. 2017. “Measurement Properties of the CLASS Toddler in ECEC in The Netherlands.” *Journal of Applied Developmental Psychology* 48: 79–91. doi:10.1016/j.appdev.2016.11.008.

Sylva, Kathy, Iram Siraj-Blatchford, Brenda Taggart, Pam Sammons, Edward Melhuish, Karen Elliot, and Vasiliki Totsika. 2006. "Capturing Quality in Early Childhood through Environmental Rating Scales." *Early Childhood Research Quarterly* 21 (1): 76–92. doi:10.1016/j.ecresq.2006.01.003.

Sylva, Kathy, Alan Stein, Penelope Leach, Jacqueline Barnes, and Lars-Erik Malmberg. 2011. "Effects of Early Child-Care on Cognition, Language, and Task-Related Behaviours at 18 Months: An English Study." *The British Journal of Developmental Psychology* 29, 18–45. doi:10.1348/026151010X533229.

Thomason, Amy C., and Karen M. La Paro. 2009. "Measuring the Quality of Teacher–Child Interactions in Toddler Child Care." *Early Education and Development* 20 (2): 285–304. doi:10.1080/10409280902773351.

Ulferts, Hannah, Katrin M. Wolf, and Yvonne Anders. 2019. "Impact of Process Quality in Early Childhood Education and Care on Academic Outcomes: Longitudinal Meta-Analysis." *Child Development* 90 (5): 1474–1489. doi:10.1111/cdev.13296.

Van Schaik, Saskia D. M., Paul P. M. Leseman, and Mariëtte de Haan. 2018. "Using a Group-Centered Approach to Observe Interactions in Early Childhood Education." *Child Development* 89 (3): 897–913. doi:10.1111/cdev.12814.

Vandell, Deborah L., Jay Belsky, Margaret Burchinal, Laurence Steinberg, and Nathan Vandergrift. 2010. "Do Effects of Early Child Care Extend to Age 15 Years? Results from the NICHD Study of Early Child Care and Youth Development." *Child Development* 81 (3): 737–756. doi:10.1111/j.1467-8624.2010.01431.x.

Vandell, Deborah L., Margaret Burchinal, and Kim M. Pierce. 2016. "Early Child Care and Adolescent Functioning at the End of High School: Results from the NICHD Study of Early Child Care and Youth Development." *Developmental Psychology* 52 (10): 1634–1645. doi:10.1037/dev0000169.

Vygotsky, Lev S. 1977. "The Development of Higher Psychological Functions." *Soviet Psychology* 15 (3): 60–73.

Zechmeister, Eugene B., Jeanne S. Zechmeister, and John J. Shaughnessy. 2009. *Research Methods in Psychology*. New York, NW: McGraw-Hill Education.

Zimmer, Karin, Jolika Schulte, Minja Dubowy, Jan-Henning Ehm, Susanne Kuger, Jan Lonnemann, Renate Martini, Dominique Rauch, and Marcus Hasselhorn. 2020. "Dokumentenrecherche und -aufbereitung als Qualitätskriterium von Systematic Reviews: Sprachförderung in Kindertageseinrichtungen als Beispiel." *Evidenzbasierte Überprüfung von Sprachförderkonzepten im Elementarbereich*, 51–76. Berlin: Springer.

Appendix

Table 3. Overview of instruments (extended version).

	Instrument	Full name	Studies found	Original author(s)	Targeted domain	(Sub-)scales	Items	Method	Procedure	Age
1	ABC Checklist	Adults Behaviors in Caregiving Checklist	1 (2x EBSCOhost)	Honig and Lally 1973 USA Published	G and M	Version 1: 8 scales (Language facilitation, Social-emotional positive inputs, Social-emotional negative inputs, Piagetian tasks, Caregiving routines with children, Caregiving routines with environment, Physical development, Does nothing) Version 2: 8 scales (Language facilitation, Social-emotional positive inputs, Social-emotional negative inputs, Presentation of Piagetian tasks and opportunities for sensorimotor development, Caregiving routines with children, Caregiving routines with environment, Physical development, Does nothing) Version 2: 9 scales (Facilitates language development, Facilitates development of skills, Facilitates concept development, Social-emotional positive inputs, Social-emotional negative inputs, Caregiving routines with children, Caregiving routines with environment, Qualitative categories, Does nothing)	Version 1: 39 items (Elicits vocalization, Convers with child, Praises/encourages verbally, Offers help or solicitous remarks, Inquires of child or makes requests, Gives information or culture rules, Provides and labels sensory experience, Reads or shows pictures to child, Singst or plays music for child, Smiles at child, Uses raised/loving/reassuring tones, Provides physical/loving contact, Plays social games with child, Eye contact to draw child's attention, Criticizes verbally/scolds/threatens, Forbids/negative demands, Frowns or restrains physically, Punished physically, Isolates child physically, Ignores child when child whos need for attention, Objective permanence, Means and ends, Imitation, Prehension: small-muscle skills, Space, New schemas, Feeds, Diapers or toilets, Dresses or undresses, Washes or cleans child, Prepares child for sleep, Physical shepherding, Eye checks on child's well-being, Prepares food, Tidies up room, Helps other caregivers, Provides kinesthetic stimulation, Provides lang-muscle play) Version 2: 44 items (Converses, Models language, Comments on child's remarks, Praises/encourages, Offers help, Inquires of child or makes request, Gives infromation, Gives culture rules, Labels sensory experiences, Reads or identifies pictures, Sings or plays music with child, Role-plays with child, Promotes child-child play, Gets social games going, Promotes self-help and social responsibility, Helps child recognize his/her own needs, Helps child delay gratification, Promotes persistence/attention span, Small muscle/perceptual motor, Large muscle/kinethesis, Arranges learning of space and time, Arranges learning of seriation/categoriation/polar concepts, Arranges learning of number, Arranges learning of physical causality, Smiles at child, Uses raised/loving or reassuring tones, Provides physical/loving contact, Uses eye contact to draw child's attention, Criticizes verballs/scolds/threatens, Forbids/negative demands, Frowns or restrains physically, Isolates child physically, Ignores child, punishes physically, Gives attention to negative behavior, Diapers/toilets, Gives physical help, Eye-checks on child's well-being, Carries child, Prepares/serves food, Tidies up room, Helps other caregivers, Prepares activites/arranges environment, Encourages creative expression, Matches tempo and/or developmental level of child, Actively engages child's interest in activities, Follows through on request)	Percentages	30 min. observation across different time of day and day of week, conducted live	Version 1: 6-18 months*** Version 2: 18-36 months***
2	Adult Engagement Scale*	Adult Engagement Scale	1 (1x EBSCOhost)	Adapted from Laevers 1994 Belgium Published	G	3 dimensions (Sensitivity, Stimulation, Autonomy)	/	5-point rating scale: Negative behavior (1-2), neutral behavior (3), predominantly positive behaviors (4-5)	Observation, conducted live	0-3 years
3	Adult Interaction Scale	Adult Interaction Scale	1 (1x EBSCOhost)	Rubenstein and Howes 1979 USA Published	G	/	6 items (Adult is not present in the room/not visible to the child, Adult ignores the child, Adult touches the child for routine care but makes no verbal responses to the child, Adult touches the child only for necessary discipline, Adult answers the child's verbal bids but does not elaborate or	7-point rating scale: Absent (0), ignores (1), routine (2), minimal (3), simple	3 hours observation, 4 x 15 min. segments that are coded in	11-30 months***

							use some unnecessary positive physical contact, Adult engages in some positive physical gestures, Adult hugs or holds the child/restates the child's statements/engages the child in conversation)	social (4), elaborated (5), intense (6)	20 sec. intervals, conducted live	
4	AMSS*	Ainsworth's Maternal Sensitivity Scales	1 (1x EBSCOhost)	Ainsworth 1969 USA Published	G	4 scales (Sensitivity/insensitivity to signals, Cooperation/interference, Acceptance/rejection of infant's needs, Accessibility/ignoring)	/	9-point rating scale: Very low (1) to very high (9)	15 min. observation, conducted via video	36-75 months
5	AQI	Assessment for Quality Improvement	1 (1x FIS Bildung)	Perlman et al. 2017 Canada Published	G and M	7 domains (Structure of the day, Activities and experiences planned, Physical environment, Learning areas, Phsyical needs, Health and safety, Interactions)	33 dimensions (Daily and visual schedules, Program plan, Activities and expiriences, Indoor physical environment, Cloakroom space and storage, Displays, Sensory/science/nature, Art, Books, Language and accesories, Physical activities, Clocks and construction, Cognitive and manupulative, Pretend play, Routine care practives, Toileting and diapering routines, Meals and/or snack time, Equipment required for earing/seating and water/refrigeration/minor food preperation, Cots or crobs and bedding, Health and safety, Toys and play equipment washing, Children's hand washing or sanitizing, Transitions, Attencance verification, Positive atmosphere, Supervision of children, Foster children's independence, Supporting the development of self-esteen, Behavior guidance, Supporting the development of communication, Extending children's learning) Infant version: 7–16 items per dimension Toddler version: 7–18 items per dimension	5-point rating scale with stop-scoring approach: Does not meet expectations (1-2), meets expectations (3), exceeds expectations (4-5)	60-90 min. observation, conducted live	Infants version: 0-18 months*** Toddler version: 19-30 months***
6	Caregiver-Child Interaction Behaviors	Caregiver-Child Interaction Behaviors	1 (1x EBSCOhost, 1x Web of Science)	Unpublished	G	Caregiver behavior factors: 4 scales (Autocratic-democratic/rejecting-accepting, Adaptice-attentice, Fostering autonomy, Responsiveness/non-involvement) Child behavior factors: 3 scales (Socially competent/positive affect, Nagetive affect, Cooperative)	Caregiver behavior factors: 12 items (Humiliates, Controlling direction, Autonomy Granting direction, Praises generally, Adapts to child, Need oriented, Visually/verbally attentice, Supports child initiative, Stimulates exploration, Asks whether child wants help, Responsive, Laussez faire) Child behavior factors: 9 items (Acts goal directed/explores, Is responsive, Communicates verbally, Is cheerful, Negative affect, Aggressive, Fearful, Cooperates, Participates in own feeding and toileting)	/	90 x 15 sec. observations on two different days in three natural situations (feeding, diapering/toilettin g, play), conducted live	6-24 months
7	CCIS	Caregiver Child Interaction Scale	1 (1x EBSCOhost)	Carl 2007 USA Unpublished	G	3 domains (Emotional domain, Cognitive/physical domain, Connection with a wider world)	17 items (Tone of voice, acceptance/respect for children, Greeting, Enjoys and appreciated children, Expextations for children, Health and safety, Routines/time spent, Physical attention, Discipline, Language development, Learning opportunities, Involvement with children's activities, Sympolic and literacy interaction, Promotion of prosocial behavior/social emotional learning, Engaging children with special needs, Relationship with families, Cultural competence)	7-point rating scale: Inadequate (1) to adequate (7)	8 hours observation during 5 days, intervalls of 30 min., conducted via video	Infants and toddlers
8	CECERS	Chinese Early Childhood Environment Rating Scale	4 (4x EBSCOhost, 3x Web of Science)	Adapted from ECERS China Unpublished	G and M	8 subscales (Space and furnishings, Personal care routine, Whole-group instruction, Activitites, Guidance and interactions, Curriculum planning and implementation, Parents and staff, Support for inclusion services) *or* 7 subscales (Space and furnishings, Personal care routine, Whole-group instruction, Activitites, Interaction, Curriculum planning and implementation, Parents and staff)	51 items with indicators *or* 53 items with indicators	7-point rating scale with anchors: Inadequate (1), minimal (3), good (5), excellent (7) *or* 9-point rating scale with anchors: Inadequate (1), minimal (3), acceptable (5), good (7), excellent (9)	6 hours observation, teacher interview, conducted live	3-6 years***
9	CIP	Caregiver Interaction Profile	5 (6x EBSCOhost, 3x Web of Science)	Helmerhorst et al. 2014 The Netherlands Published	G	6 scales (Sensitive responsiveness, Respect for autonomy, Structuring and limit setting, Verbal communication, Development stimulation, Fostering positive peer interactions)	/	7-point rating scale: Very low (1), low (2), moderate/low (3), moderate (4), moderate/high (5), high (6), very high (7)	3-4 x 8-10 min. observation, conducted via video	0-4 years***

10	CIS	Arnett Caregiver Interaction Scale	16 (14x EBSCOhost, 4x Web of Science, 4x FIS Bildung)	Arnett 1989 USA Published	G	4 subscales (Positive interactions/sensitivity, Punitiveness/harshness, Permissiveness/authoritarian, Detachment)	26 items (Speaks warmly, Seems critical, Listens attentively, High value on obedience, Semms distant or detached, Enjoys children, Explains rules, Encourages new experiences, does not try to exercise much control, Speaks with irritation, Enthusiastic around activities and efforts, Threatens children, Not involved with children, Plays positive attention, Does not reprimand mishbehavior, Talks to children on level they understand, Punishes without explanation, Firm when necessary, Encourages pro-social behavior, Finds fault easily, Disinterest in children's activities, Prohibits many things, Does not supervise closely, Expects self-control, Talks to children on theis level, Harsh when scolding)	4-point rating scale: Not at all (1), somewhat (2), quite a bit (3), most of the time (4)	45 min.-4 hours observation, conducted live	0-6 years
11	CKEQRS-TV	Chinese Kindergarten Education Quality Rating Scale - Teacher Version	1 (1x Web of Sciences)	Liu 2019 China Published	G	2 scales (Activities, Organization of activities) with 5 subscales (Design/utilization of the physical environment, Interpersonal interactions, Organization of daily routine care, Facilitation/guidance for play, Administration of educational)	17 items with 64 indicators	5-point rating scale: inadequate (1), minimal (2), ordinary (3), good (4), excellent (5)	4 hours observation, teacher interview, review of program materials, conducted live	Kindergarten
12	CLASS	Classroom Assessment Scoring System	6 (5x EBSCOhost, 1x FIS Bildung)	La Paro and Pianta 2003 USA Published	G	3 domains (Emotional support, Classroom organization, Instructional support)	9 dimensions (Positive climate, Negative climate, Teacher sensitivity, Overcontrol, Concept development, Quality of feedback, Behavior management, Productivity, Instructional learning format) with indicators	7-point rating scale: Low (1, 2), medium (3, 4, 5), high (6, 7)	4 x 15-20 min. observation during one day and 4 x 10 min. coding after each observation, conducted live	3-5 years***
13	CLASS Infant*	Classroom Assessment Scoring System Infant Version	1 (1x EBSCOhost)	Hamre et al. 2014 USA Published	G	1 domain (Responsive caregiving)	4 dimensions (Relational climate, Teacher sensitivity, Facilitated exploration, Early language support) with indicators	7-point rating scale: Low (1, 2), medium (3, 4, 5), high (6, 7)	4 x 15-20 min. observation during one day and 4 x 10 min. coding after each observation, conducted live or via video	0-18 months***
14	CLASS Toddler	Classroom Assessment Scoring System Toddler Version	8 (8x EBSCOhost, 2x Web of Science)	La Paro et al. 2012 USA Published	G	2 domains (Emotional and behavioral support, Engaged support for learning)	8 dimensions (Positive climate, Negative climate, Teacher sensitivity, Regard for child perspective, Behavior guidance, Facilitation of learning and development, Quality of feedback, Language modeling) with indicators	7-point rating scale: Low (1, 2), medium (3, 4, 5), high (6, 7)	4 x 15-20 min. observation during one day and 4 x 10 min. coding after each observation, conducted live or via video	15-35 months***
15	CLASS Pre-K	Classroom Assessment Scoring System Pre-Kindergarten Version	33 (32x EBSCOhost, 6x Web of Science, 5x FIS Bildung)	Pianta et al. 2008 USA Published	G	3 domains (Emotional support, Classroom organization, Instructional support)	10 dimensions (Positive climate, Negative climate, Teacher sensitivity, Regard for student perspectives, Behavior management, Productivity, Instructional learning formats, Concept development, Quality of feedback, Language modeling) with indicators	7-point rating scale: Low (1, 2), medium (3, 4, 5), high (6, 7)	4 x 15-20 min. observation during one day and 4 x 10 min. coding after each observation, conducted live	3-5 years***
16	COEMET	Classroom Observation of Early Mathematics Environment and Teaching	2 (1x Web of Sciences, 1x FIS Bildung)	Sarama and Clements 2007 USA Unpublished	G and M	3 scales (Classroom elements, Classroom culture, Specific math activities) with 9 subscales (Environment and interactions, Personal attributes of the teacher, Mathematical focus, Organization/approaches/interactions, Expectations, Eliciting children's solution methods, Supporting children's conceptual understanding, Extending children's math thinking, Formative assessment)	31 items (Number of computers running math activities, Mean number of math activities used, Mean duration of math activities, Percentage teachers stayed in classroom, Interacted with children, Used teachable moments, Percentage time use computers, Math work displayed, Knowledgable about math, Believed math learning enjoyable, Enthusiasm for math ideas, Understanding of topic, Developmentally appropriate, Engaged children's math thinking, Pace appropriate, Management strategies, Actively involved, Percentage time involved, Appropriate strategies, High/realistic ecpectations, Acknowledged effort, Asked children to share ideas, Facilitated children's responses, Encouraged evaluating others, Supported describer's thinking, supported listener's understanding, Gave just enough assistance, Elaborated children's ideas, Encouraged reflection, Listened to children/takes notes, Adapted activities)	5-point rating scale	Half-day observation, conducted live or via video	4-5 years

17	COS	Early Childhood Classroom Observation Scale	1 (1x EBSCOhost)	Unpublished	G	5 subscales (Interactions among staff and children, Cirriculum, Physical environment, Health and safety, Nutrition and food service)	66 items	3-point rating scale: criterion not met (1), criterion partially met (2), criterion fully met (3)	1 hour observation, conducted live	Birth-8th grade
18	DO-RESI-E-KiGs	Dortmund rating scale to measure language-related interactions - extended version for daycare centers and elementary schools**	1 (2x FIS Bildung)	Fried and Briedigkeit 2008 Germany Published	G and M	6 subscales (Organization, Relationships, Adaptive support, Linguistic-cognitive challenges, Mathematical-cognitive challenges, Scientific-cognitive challenges)	27 basic-items, 4 early childcare center-specific items, 9 primary school-specific items	7-point rating scale: Unsatisfying (1) to excellent (7)	For early childcare centers: 3-4 hour observation, for primary schools: 4-6 hours observation, teacher interview, conducted live	3-10 years
19	EA Scales, Version 3	Emotional Availability Scales, Version 3	1 (1x EBSCOhost)	Biringen et al. 1998 USA Published	G	/	4 dimensions for teacher (Sensitivity, Structuring, Non-instructiveness, Non-hostility) 2 dimensions for child (Responsiveness, Involvement)	Global score	30 min. observation, conducted live	10-36 months
20	EAS	Emerging Academics Snapshot	1 (1x EBSCOhost)	Ritchie et al. 2001 USA Published	G and M	3 scales (Activity settings, Preacademic engagement, Teacher-child interactions)	23 items (Basic, Free choice, Individual time, Meals, Small group, Whole group, Esthetics, Fine motor, Gross motor, Letter and sound, Mathematics, Oral language, Prereading, Read to child, Science, Social studies, Writing, Routine, Minimal, Simple, Elaborated, Scaffold, Didactic)	Dichotomous rating scale: Yes or no if the teacher is or is not engaging in the target behaviors during the 20 seconds	6-8 x 20 min. observation, intervalls of 20 sec. observing and 40 sec. recording, conducted live	3-4 years
21	ECCOM	Early Childhood Classroom Observation Measure	2 (2x EBSCOhost)	Stipek and Byler 2004 USA Published	G and M	3 scales (Child-centeres, Teacher-directed, Child-dominated) with 3 subscales (Management, Climate, Instruction)	14 items (Child responsibility, Management, Choice of activities, Discipline strategies, Support for communication skills, Support for interpersonal skills, Student engagement, Individualization of learning activities, Learning standards, Coherence of instructional activities, Teaching concepts, Instructional conversation, Literacy instruction, Math instruction)	5-point rating scale: Practices are seen 0-20% of the time (1) to practices are seen 80-100% of the time (5)	3 hours observation, conducted live	2-7 years
22	ECERS	Early Childhood Environment Rating Scale	14 (12x EBSCOhost, 4x Web of Science, 2x FIS Bildung)	Harms and Clifford 1980 USA Published	G	7 subscales (Furnishing and display, Personal care routines, Language and reasoning experiences, Fine and gross-motor activities, Creative activities, Social development, Adult need)	37 items (Greeting/departing, Meals/snacks, Nap/rest, Diapering/toileting, Personal grooming, Routine care, Learning activities, Relaxation and comfort, Room arrangement, Child-related display, Understand of language, Using language, Using learning concepts, Infromal use of language, Perceptual/fine motor, Supervision of fine motor activities, Space for gross motor, Gross motor equipment, Time for gross activities, Supervision of gross motor activities, Art, Music/movement, Blocks, Sand/water, Dramatic play, Schedule, Supervision of creative activities, Space to be alone, Free play, Group time, Cultural activities, Tone, Provision for expectional children, Sdult personal area, Opportunities for professional growth, Adult meeting area, Provisions for parents) with indicators	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	2-3 hours observation, teacher interview, conducted live	6 months-6 years
23	ECERS-R	Early Childhood Environment Rating Scale Revised Edition	47 (41x EBSCOhost, 11x Web of Science, 16x FIS Bildung)	Harms et al. 1998 USA Published	G and M	7 subscales (Space and furnishing, Personal care routines, Listening and talking, Activities, Interaction, Program structure, Parents and staff)	43 items (Indoor space, Furniture for routine care/play/learning, Furnishings for relaxation and comfort, Room arrangement for play, Space for privacy, Space for gross motor play, Gross motor equipment, Greeting/departing, Meals/snacks, Nap/rest, Toileting/diapering, Health practices, Safety practices, Books/pictures, Encouraging children to communicate, Using language to develop reasoning skills, Informal use of language, Fine motor, Art, Music/movement, Blocks, Sand/water, Dramatic play, Nature/science, Math/number, Use of TV/video and/or computers, Promoting acceptance of diversity, Supervision of gross motor activities, General supervision of children, Discipline, Staff-child interactions, Interaction among children, Schedule, Free play, Group time, Provisions for children with disabilities, Provision for parents, Provision for personal needs and staff, Provision for professional needs and staff, Staff interaction/coopleration, Supervision and evaluation of staff, Opportunities for professional growth) with indicators	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	2-5 hours observation, teacher interview, conducted live	Child care groups in which more than 50% of the children are between the ages of 30-60 months***

24	ECERS-E	Early Childhood Environment Rating Scale Extension	4 (3x EBSCOhost, 1x Web of Science, 3x FIS Bildung)	Sylva et al. 2003 UK Published	M	4 subscales (Literacy, Mathematics, Science and environment, Diversity)	18 items (Environmental print letters and words, Book and literacy areas, Adult reading with the children, Sounds in words, Emergent writing/mark making, Talking and listening, Counting and the application of counting, Reading and writing simple numbers, Mathematical activities: shape/space, Mathematical activities: matching/sorting/comparing, Natural materials, Areas featuring science and science resourcing, Science activities science processes: non-living, Science activities: living processes, Science activities: food preparation, Planning for children's individual needs, Gender equality and awareness, Race equality and awareness) with indicators	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	Full-day observation, teacher interview, conducted live	3-5 years
25	ECERS-3	Early Childhood Environment Rating Scale Third Edition	2 (1x EBSCOhost, 1x FIS Bildung)	Harms et al. 2015 USA Published	G and M	6 subscales (Space and Furnishings, Personal Care Routines, Language and Literacy, Learning Activities, Interaction, Program Structure)	35 items (Indoor space, Room for care/play/learning, Furnishing for care/play/learning, Space for privacy,Cchild-related play, Space for gross motor play, Gross motor quipment, Meals/snacks, Toileting/diapering, Health practices, Safety practices, Helping children expland vocabulary, Encouraging children to use language, Staff use of books with children, Encouriging children's use of books, Becoming familiar with print, Fine motor, Art, Music movement, Blocks, Dramatic play, Nature/science, Math materials and activities, Math in daily events, Understanding written numbers, Promoting acceptance of diversity, Appropriate use of technology, Supervision of gross motor, Individualized teaching and learning, Staff-child interaction, Peer interaction, Discipline, Transition and waiting time, Free play, Whole-group activities for play/learning) with indicators	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	3 hours observation, teacher interview, conducted live	3-4 years
26	ECS	Educator Cognitive Sensitivity Scale	1 (1x EBSCOhost, 1x Web of Science)	Adapted from the Cognitive Sensitivity Scale Prime et al. 2014 Canada Published	G	6 subscales (Mind-reading, Mutuality building, Communicative clarity, Responsivity, Proactiveness, Language promotion)	23 items	5-point rating scale with thin-slice methodology: Not true at all (1) to very true (5)	5 min. observations, conducted live	0-30 months
27	EmoReg	EmoReg	1 (1x EBSCOhost)	Unpublished	G	3 scales (Emotion regulation strategy, Co-regulation by teacher/Self-regulation by child, Emotional coaching by teacher)	/	Event-coding (when challanging emotions appear)	Identifying emotionally challanging situation, coding the emotions, cross-classifying each occurrence (co-regulation by teacher and self-regulation by child), classifying emotion coaching by teacher, conducted via video	47-77 months
28	EQOS	Educational Quality Observation Scale, Preschool Version	4 (3x EBSCOhost, 1x FIS Bildung)	Bourgon and Lavallée 2004 Canada Published	G	4 scales (Physical setting, Programming, Interaction with children, Interaction with parents) with 9 subscales (Space, Material, Planning, Observation, Schedule, Activities, Play values, Intervention, Communication)	Over 100 items	4-point rating scale: Poor (1), minimum (2), good (3), very good (4)	5 hours observation, teacher interview, conducted live	18 months and up***
29	ESA	Adult Sensitivity Scale**	1 (1x EBSCOhost)	Santelices et al. 2012 Chile Published	G	3 subscales (Responsiveness, Playful encouragement, Warm attunement)	19 items	3-point rating scale: Low sensitivity (1) to high sensitivity (3)	Observation, conducted via video	6-36 months***

30	Evaluation of the Mentalizatio n of the Significant Caregiver	Evaluation of the Mentalization of the Significant Caregiver	1 (1x EBSCOhost)	Farkas et al. 2017 Chile Published	G	10 categories (Desires, Cognitions, Emotions, Psychological attributes, States of consciousness, Physiological states, Causal talk, Factual talk, Links with child's life, Physical expressions)	/	3-point rating scale: Low, adequate, high	Observation of a structured situation (storrytelling-situation), conducted via video	Version 1: 0-23 months*** Version 2: 24-48 months***
31	IMCEIC*	Instrument for Measuring Quality of Early Childhood Education in Colombia	1 (1x Web of Sciences)	Adapted from MELE	G and M	3 scales (Safety conditions of the area and the center, Availability of resources and learning materials in the classroom, Quality of teacher-child interactions and domain-specific activities)	45 items	4-point rating scale: Low level (<2), medium level (2-3), hich level (>3)	4 hour observation, conducted live	3-5 years
32	ISSA observation protocol	International Step by Step Association observation protocol	1 (1x Web of Sciences)	Tankersley et al. 2011 Croatia Published	G	5 scales (Interaction in the educational work in the group, Differences and democratic values in the educational work of teachers, Teaching strategies in the educational work, Learning environments in the upbringing and educational practice, Planning educational activities)	46 items	4-point rating scale: Inadequate (1), good start (2), quality practice (3), a step forward (4)	30-45 min. observation, conducted live	2-6 years
33	ITERS	Infant/Toddler Environment Rating Scale	8 (8x EBSCOhost, 2x Web of Science, 3x FIS Bildung)	Harms et al. 1990 USA Published	G	7 subscales (Furnishing and display, Personal care routines, Listening and talking, Activities, Interaction, Programm structure, Adult needs)	35 items (Furnishing for routine care, Use of furnishing for learning activities, Furnishing for relaxation and comfort, Room arrangement, Display for children, Greeting/departing, Meals/snacks, Nap, Diapering/toileting, Personal grooming, Health practive, Health policy, Safety practice, Safety policy, Infromal use of language, Books and pictures, Eye-hand coordination, Active physical play, Art, Music/movement, Blocks, Pretend play, Sand/water play, Cultural awareness, Peer interaction, Caregiver-child interaction, Discipline, Schedule of daily activities, Supervision of daily activities, Staff Cooperation, Provision for exceptional children, Adult personal needs, Opportunities for professional growth, Adult meeting are, Provision for parents) with indicators	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	2-3 hours observation, teacher interview, conducted live	0-30 months***
34	ITERS-R	Infant/Toddler Environment Rating Scale Revised Edition	12 (11x EBSCOhost, 6x Web of Science, 6x FIS Bildung)	Harms et al. 2003 USA Published	G	7 subscales (Space and furnishing, Personal care routines, Listening and talking, Activities, Interaction, Programm structure, Parents and staff)	39 items (Indoor space, Furnishing for routine care and play, Provision for relaxation and comfort, Room arrangement, Display for children, Greeting/departing, Meals/snacks, Nap, Diapering/toileting, Health practices, Safety practices, Helping children understand language, Helping children use language, Using books, Fine motor, Active physical play, Art, Music/movement, Blocks, Damatic play, Sand/water play, Nature/science, Use of TV/video and/or computer, Promoting acceptance of diversity, Supervision of play and learning, Peer interaction, Staff-child interaction, Discipline, Schedule, Free play, Group play activities, Provision for children with disabilities, Provision for parents, Provision for personal needs of staff, Staff interaction and cooperation, Staff continuity, Supervision and evaluation of staff, Opportunities for professional growth) with indicators	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	3-5 hours observation, teacher interview, conducted live	More than 50% are younger than 30 months***
35	ITERS-3	Infant/Toddler Environment Rating Scale Third Edition	1 (1x FIS Bildung)	Harms et al. 2017 USA Published	G and M	6 subscales (Space and furnishing, Personal care routines, Language and books, Activities, Interaction, Program structure)	33 items (Infoor space, Furnishing for care/play/learning, Room arrangement, Display for children, Meals/snacks, Diapering/toileting, Health practices, Safety practices, Talking with children, Encouraging vocabulary development, Responding to children's communication, Encouraging children to communicate, Stadd use of books with children, Encouraging children's use of books, Fine motor, Art, Music/movement, Blocks, Dramatic play, Nature/science, Math/number, Appropriate use of technology, Promoting acceptance of diversity, Gross motor, Supervision of gross motor play, Supervision of play and learning (non-gross motor), Peer interaction, Staff-child interaction, Providing physical warmth/touch,	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	Observation, teacher interview, conducted live	More than 50% are younger than 36 months***

							Guiding children's behavior, Schedule and transitions, Free play, Group play activities) with indicators			
36	KERS	Kindergarten Environment Rating Scale	1 (1x EBSCOhost)	Liu-Yan and Pan-Yuejuan 2008 China Unpublished	G (math-specificity regarding curriculum)	4 subscales (Physical environment, Interactions, Routine care, Curriculum)	25 items (Indoor space, Arrangement of indoor space, Furniture for routine care and learning, Play materials, Child-related display, Space and equipment for gross motor play, Provisions for personal and professional needs of staff, Interactions among teachers, Teacher-child interactions, Interactions among Children, Teacher-parent interactions, Schedule, Discipline, Greeting/departing, Meals, Nap, Toileting, Free play, Health, Society, Science, Mathematics, Language, Art and music, Observation and assessment of young children) with indicators	5-point rating scale with anchors: Inadequate (1), good (3), excellent (5)	Observation, teacher interviews, conducted live or via video	4-5 years
37	KES-R	Kindergarten Scale Revised Edition**	2 (1x EBSCOhost, 3x FIS Bildung)	Tietze et al. 2005 Germany Published	G and M	7 subscales (Space and Furnishings, Personal Care Routines, Language-Reasoning, Activities, Interaction, Program Structure, Parents and Staff)	43 items (Indoor space, Furniture for routine care/play/learning, Furnishings for relaxation and comfort, Room arrangement for play, Space for privacy, Space for gross motor play, Gross motor equipment, Greeting/departing, Meals/snacks, Nap/rest, Toileting/diapering, Health practices, Safety practices, Books/pictures, Encouraging children to communicate, Using language to develop reasoning skills, Informal use of language, Fine motor, Art, Music/movement, Blocks, Sand/water, Dramatic play, Nature/science, Math/number, Use of TV/video and/or computers, Promoting acceptance of diversity, Supervision of gross motor activities, General supervision of children, Discipline, Staff-child interactions, Interaction among children, Schedule, Free play, Group time, Provisions for children with disabilities, Provision for parents, Provision for personal needs and staff, Provision for professional needs and staff, Staff interaction/coopleration, Supervision and evaluation of staff, Opportunities for professional growth) with indicators	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	2-5 hours observation, teacher interview, conducted live	More than 50% are between the ages of 30-60 months***
38	KES-E	Kindergarten Scale Extension**	1 (1x EBSCOhost, 1x FIS Bildung)	Roßbach and Tietze 2010 Germany Published	M	4 subscales (Literacy, Mathematics, Science and environment, Diversity)	18 items (Environmental print letters and words, Book and literacy areas, Adult reading with the children, Sounds in words, Emergent writing/mark making, Talking and listening, Counting and the application of counting, Reading and writing simple numbers, Mathematical activities: shape/space, Mathematical activities: matching/sorting/comparing, Natural materials, Areas featuring science and science resourcing, Science activities science processes: non-living, Science activities: living processes, Science activities: food preparation, Planning for children's individual needs, Gender equality and awareness, Race equality and awareness) with indicators	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	3-4 hour observation, teacher interview, conducted live	3-6 years
39	KRIPS-R	Infant/Toddler Scale Revised Edition**	1 (1x FIS Bildung)	Tietze and Harms 2005 Germany Published	G	7 subscales (Space and furnishing, Personal care routines, Listening and talking, Activities, Interaction, Programm structure, Parents and staff)	41 items (Indoor space, Furnishing for routine care and play, Provision for relaxation and comfort, Room arrangement, Display for children, Greeting/departing, Meals/snacks, Nap, Diapering/toileting, Health practices, Safety practices, Helping children understand language, Helping children use language, Using books, Fine motor, Active physical play, Art, Music/movement, Blocks, Damatic play, Sand/water play, Nature/science, Use of TV/video and/or computer, Promoting acceptance of diversity, Supervision of play and learning, Peer interaction, Staff-child interaction, Discipline, Schedule, Free play, Group play activities, Provision for children with disabilities, Provision for parents, Provision for personal needs of staff, Staff interaction and cooperation, Staff continuity, Supervision and evaluation of staff, Opportunities for professional growth, Acclimatization, Inclusion of the family environment) with indicators	7-point rating scale with anchors and stop-scoring approach: Inadequate (1), minimal (3), good (5), excellent (7), each indicator is scored with yes, no, or not applicable	3-5 hours observation, teacher interview, conducted live	More than 50% are younger than 30 months***
40	Language samples of educator-child interactions	Language samples of educator-child interactions	1 (2x EBSCOhost)	Unpublished	G	2 scales (Guidance, Directives)	11 items (Physical, Cognitive, Language, Socio-emotional, Custodial, Social management, Nurturance, Unproductive, Close-ended, Non response, Other)	/	1 hour observation of structured activities and 1 hour observation unstructured time, conducted live	4 years
41	MELE*	Measure of Early Learning Environments	1 (1x EBSCOhost)	UNESCO et al. 2017 International	G and M	7 scales (Environment and physical setting, Family and community engagement,	25 items	Rating	Observation, teacher interview,	4-6 years

				Published		Personnel, Interactions, Inclusiveness, Pedagogy, Play)			parents interview, director interview	
42	MLERS	Mediated Learning Rating Scale	2 (2x EBSCOhost)	Lidz 1991 USA Published	G	12 subscales (Intentionality, Meaning, Transcendence, Joint regard, Sharing (of experiences), Task regulation, Praise and encouragement, Callange and competence, Psychological differentiation, Contingent Responsivity, Affectice involvement, Change)	/	4-point rating scale: Low (0) to high (3)	10-15 min. observation, conducted via video	18 months- 6 years
43	Multi-categorical coding scheme	Multi-categorical coding scheme	1 (1x EBSCOhost)	Unpublished	G	4 scales (Attention, Sensitivity, Non-sensitivity, Stimulation)	15 items (Attention, Responsiveness, Comforting, Praising, Touching-affective, Non-responsiveness, Ignoring, Intrusiveness, Critique, Touching-aversive, Encouraging, Stimulating, Informing, Giving space, Setting limits)	/	10 min. observation, semi-structured, conducted via video	33-38 months
44	NCKO Sensitivity Scale	NCKO Sensitivity Scale	1 (1x EBSCOhost)	De Kruif et al. 2007 The Netherlands Published	G	1 scale (Sensitivity of teachers responses)	/	7-point rating scale	30 min. observation, intervalls of 10 min., conducted via video	37-64 month
45	OPRS	Outdoor Play Rating Scale	1 (2x EBSCOhost)	Hu and Li 2012 China Unpublished	G	2 subscales (Space and equipment, Activity and organization)	9 items (Space for outdoor play, Outdoor space planning, Types and quantity of equipment, Safety and maintenance, Opportunity and time for outdoor play, Form and content of outdoor avcitivies, Appropriateness of outdoor activities, Supervision of outdoor play)	7-point rating scale with anchors: Inadequate (1), minimum (3), good (5), excellent (7)	Observation, conducted live	3-6 years
46	PICCOLO	Parenting Interactions with Children: Checklist of Observations Linked to Outcomes	1 (1x EBSCOhost)	Roggman et al. 2009 USA Published	G	4 domains (Affection, Responsiveness, Encouragement, Teaching)	29 items (Speaks in a warm tone of voice, Smiles at child, Praises child, Is physically close to child, Uses positive expressions with child, Is engaged in interacting with child, Shows emotional warmth, Pays attention to what child is doing, Changes pace or activity, Is flexible, Follows what child is trying to do, Responds to child's emotions, Looks at child, Replies to child's word and sounds, Waits for child response, Encourages child to handle toys, Supports child in making choices, Supports child in doing things on his/hew own, Verbally encourages child's efforts, Offers suggestions to help child, Shows enthusiasm , Explains reasons, Suggests activities, Repeats or expands, Lables objects or actions, Engages in pretend play, Does acvtivities in a sequence of steps, Talks to child about characteristics of objects, Ask child for inromation)	3-point rating scale: No behavior observed (0) behavior barely observed (1), behavior clearly observed (2)	Observation, conducted via video	10-42 months***
47	PITC PARS	Program for Infant and Toddler Care Program Assessment Rating Scale	1 (1x FIS Bildung)	WestEd USA Published	G	5 subscales (Quality of adult's interactions with children, Family partnerships/cultural responsiveness/inclusive care, Organization of group care, Physical environment, Routines and record keeping)	27 items with 108 indicators	4-point rating scale: Beginning (0-1.7), emerging (1.8-2.7), developing (2.8-3.7), refining (3.8-4), each indicator is scored with observed or not observed	4 hours observation, teacher interview, review of program materials, conducted live	0-36 months***
48	SBTR	Student Behavior Teacher Response	1 (1x EBSCOhost)	Pelham et al. 2008 USA Unpublished	G	Step 1: 2 scales (Teacher acknowledgement, Teacher consequence) Step 2: 1 scale (Post observation inventory)	Step 1: 6 items (No acknowledgment, Appropriate acknowledgment, Inappropriate acknowledgment, No consequence, Appropriate consequence, Inappropriate consequence) Step 2: 5 items (Overall effectiveness of the teacher's use of praise, Commands, Behavioral management strategies, Tone of voice, Overall climate)	Rating in two steps: Step 1: True or not true Step 2: 7-point rating scale: Not at all effective/very negative (1) to very effective/very positive (7)	20 min. observation, 10 min. scoring, conducted live	Kindergarten-6th grade***
49	SSTEW	Sustained Shared Thinking and Emotional Wellbeing	1 (1x EBSCOhost, 1x Web of Sciences)	Siraj et al. 2015 UK Published	G	5 subscales (Bulding trust, Confidence and independence, Social and emotional wellbeing, Supporting and extending language and communication,	14 items	7-point rating scale with anchors: Inadequate (1), basic/minimal (3),	Full-day observation, conducted live	2-5 years***

						Supporting learning and critical thinking, Assessing learning and language)		good (5), excellent (7)		
50	TBRS*	Teacher Behavior Rating Scale	1 (1x EBSCOhost)	Landry et al. 2002 USA Unpublished	M	2 subscales (Literacy, Math) with 10 categories (Oral language quality, Book reading quality, Written expression quality, Print and letter knowledge quality, Math quality, Oral language quantity, Book reading quantity, Written expression quantity, Print and letter knowledge quantity, Math quantity)	20 items	4-point rating scale on quality and quantity: Activity not present (0) to activity high quality/activity happened often or many times (3)	Observation, conducted live	2.5-5 years
51	TCIS	Teacher Child Interaction Scale	1 (1x EBSCOhost)	Farran and Collins 1996 USA Unpublished	G	/	11 items (Physical involvement, Verbal involvement, Responsiveness, Play interaction, Teaching behavior, Control over children's activities, Directives or demands, Relationship among activities, Positive statements, Negative statements, Goal setting)	5-point rating scale on amount, quality and appropriateness	30 min. observation, conducted live	2-6 years
52	TC-SPT	Teacher-Child Structured Play Task	2 (2x EBSCOhost)	Adapted from the Mother-child Interaction Task Egeland and Hiester 1993 USA Published	G	Teacher interactive behavior: 2 scales (Positive teacher interactions, Negative teacher interactions) Child interactive behavior: 2 scales (Child active engagement, Child positive interactions with teacher)	Teacher interactive behavior: 8 items (Ensitive/responsive presence, Positive affect, Teacher confidence, Teacher encourages stimulating environment, Teacher directiveness, Teacher support for child autonomy, Teacher negativity, Affective mutuality/felt security) Child interactive behavior: 11 items (Child enthusiasm, Experience of the session, Affection toward teacher, Reliance on teacher for help, Persistence, Compliance, Child negativity toward teacher, Avoidance, Child's negative emotions, Behavior control, Affective mutuality/felt security)	5-point rating scale	10 min. observation of a structured play task followed by clean-up session, conducted via video	3-4 years
53	TILRS (French: GEIPLE)	Teacher Interaction and Language Rating Scale	1 (2x EBSCOhost)	Girolametto et al. 2000 Canada Published	G	/	11 items (Wait and listen, Follow child's lead, Join in and play, Face to face, Variety of questions, Turn-taking, scan in small groups, Imitate, Variety of labels, Expand, Extend)	7-point rating scale with anchors: Rarely (1), sometimes (3), frequently (5), consistently (7)	15 min. observation, conducted via video	4 years
54	TSRS	Teacher Styles Rating Scale	1 (1x EBSCOhost, 1x Web of Science)	McWilliam et al. 1998 USA Published	G	/	7 interaction behavior items (Redirecting, Introducing, Elaborating, Following, Informing, Acknowledging, Praising) and 13 affective characteristics items	7-point rating scale for interaction behavior with anchors: never (1), occasionally (3), often (5), most of the time (7), 5-point rating scale for affective characteristics	15 min. observation, conducted live	12-36 months
55	Watts's Human Interaction Scale	Watts's Human Interaction Scale	1 (2x EBSCOhost)	Watts and Barnett 1973 USA Published	G	2 subscales (Activities, Techniques of interactions)	9 items (Highly intellectual, Moderately intellectual, Non-intellectual, Social, Didactiv teaching, Facilitation, General information giving, Observation, Restriction) with several items (Verbal/symbolic learning, Perceptual/spatial/fine motor learning, Concrete reasoning, Expressive skills, Executive skills, Exploration/play with household items/toys/nature, Gaining general/routine information, Basic care, Gross motor activity, Unspecific behavior, Positive social/emotional expression, Negative social/emotional expression, Neutral social/emotional expression, Justification, Active participation, Suggestions/commands, Positive reinforcement, Focusing, Assistance, Providing materials, Moving the child, Prohibition, Negative reinforcement, Ignoring, Refusal)	/	45 min. observation, intervalls of 7-15 min., conducted via video	22-26 months

Note. *additional hand search; **translated from German into English; ***as reported in manual, otherwise age of study sample; G=global interaction quality; M=math-specific interaction quality; Number of studies found in brackets include duplicates

Table 4. Reported reliability and validity.

	Reportet reliability (with N)	Reportet validity (with N)
1	**Study 1:** Inter-rater reliability = 84% (N teachers = 2)	/
2	/	/
3	**Study 1:** Double-coding in three centers, inter-rater reliability κ = .84 (N centers = 25, N groups = 63, N teachers = 59, N children = 119)	/
4	**Study 1:** 100% double-coding, inter-rater correlation r = .95 (N teachers = 51, N children = 72)	/
5	**Study 1:** Internal consistency α = .85 (N groups infants = 235) and .89 (N groups toddlers = 386)	**Study 1:** Correlation between Infant version and ITERS-R = .322, p < .05 (N groups infants = 251), correlation between Toddler version and ITERS-R = .402, p < .01 (N groups toddlers = 464)
6	**Study 1:** Inter-rater reliability between six observers r = .99 (N teachers = 72)	**Study 1:** Caregiver behavior factors: EFA: 4-factors explain 71.5% variance, correlation of the 4-factors = .33 - .69, caregiver behavior factors and ECERS subscales were mostly uncorrelated (median correlation r = .23); Child behavior factors: EFA: 3-factors explain 68.2% variance (N teachers = 72)
7	**Study 1:** Internal consistency α = .91 (N teachers = 6)	**Study 1:** Correlation between CCIS and ERS r = .74, r = .75 and r = .75 for each construct (N teachers = 6)
8	**Study 1:** All classrooms except two were rated by a pair, inter-rater reliability 13 items κ > .80, 37 items κ between .60 - .80, 1 item κ = .50, internal consistency for the total score α = .96 (range .82 - .96) (N groups = 176); **Study 2:** All classrooms except two were rated by a pair, inter-rater reliability = 64.2% for exact agreement and 90.7% for agreement within 1-point, estimates range from r = .85 - .94 (M = .89)., internal consistency α = .83 - .93 for the subscale scores, α = .98 for the total composite score (N groups = 178, N children = 1012); **Study 3**: Inter-rater reliability κ = .61 - .88 across items and .83 - .95 across scales, internal consistency α = .89 - .95 (N centers = 193, N groups = 428, N children = 2.110); **Study 4:** 100% double-coding, consistency reliability coefficient of raters = .85 (N groups = 48)	**Study 1:** CFA for 8-factor strucutre: CMIN/DF (1663.6/1052), GFI = 0.72, RMSEA = 0.052, NNFI = .99, CFI = .99, NFI = 0.97 (N groups = 176); **Study 2:** EFA revealed a 2-factor structure (N groups = 178, N children = 1012)
9	**For all studies:** At least 80% agreement within 1-point with a consensus score provided by an expert or meet a criterion of ICC = .70 for each scale before observing in the field; **Study 1:** Inter-rater reliability was computed for 10% of the videos, absolute agreement = 83%, within 1-point agreement = 98% (N teachers = 19); **Study 2:** Inter-rater reliability was computed for 10% of the videos, absolute agreement sensitive responsiveness ICC = .85, respect for autonomy ICC = .90, structuring and limit-setting ICC = .94, verbal communication ICC = .91, developmental stimulation ICC = .91, fostering positive peer interactions ICC = .78 (N centers = 66, N groups = 111, N teachers = 168); **Study 3:** Inter-rater reliability was computed for 10%, r = .85 for sensitive responsiveness, r = .81 for respect for autonomy, r = .76 for verbal communication, r = .80 for developmental stimulation, r = .83 for fostering positive peer interactions (N centers = 47, N groups = 75, N teachers = 145); **Study 4:** Inter-rater reliability was computed for 10%, r = .83 (N centers = 200); **Study 5:** Inter-rater reliability was computed for 16%, average 1- point agreement = 94.8% (N teachers = 72)	**Study 2:** EFA revealed a 1-factor structure, strongest associations were found between ITERS-R scales listening and talking, interactions, and program structure and CIP scales verbal communication and developmental stimulation (N centers = 66, N groups = 111, N teachers = 168); **Study 3:** EFA revealed a 1-factor structure, CIP total score and separate CIP scale scores all correlated significantly and positively with CIS score and with ITERS-R/ECERS-R total score (N centers = 47, N groups = 75, N teachers = 145); **Study 4:** Correlations between CIP scores and ITERS-R/ECERS-R scores were significant and moderate ranging from r = .11 - .49 (N centers = 200); **Study 5:** CIP score α = .84 (N teachers = 72)
10	**Manual:** Observers became reliable at 80% agreement or better for each subscale before data collection; **Study 3:** 10 cases were double-coded, percentage of agreement > .80%, internal consistency supportiveness α = 86, hostility α = .88, detachment α = .92 (N centers = 21, N teachers = 57); **Study 4:** Internal consistency α = .59 - .95 (N groups = 1350); **Study 8:** 10% double-coding, inter-rater reliability = 90% (range 85% - 98%), internal consistency α = .87 (N centers = 47, N groups = 75, N teachers = 145); **Study 10:** Internal consistency sensitivity α = .94, harshness α = 82, detachment α = .83 (N groups = 222, N teachers = 330); **Study 11:** 17% double-coding, inter-rater reliability κ = .78 (range .57 - .96) (N centers = 11, N groups = 24, N teachers = 56); **Study 12:** Internal consistency positive interactions α = .78, punitiveness α = .79, permissiveness α = .79, detachment α = .97 (N centers = 1); **Study 13:** Inter-rater reliability sensitivity r = .95 (range .90 - .98), harshness r = .92 (range .89 - .96), detachment r = .93 (range .90 - .95), permissiveness r = .92 (range .89 - .94), internal consisteny sensitivity subscale α = .85 - .94, involvement subscale α = .39 - .78, acceptance subscale α = .83 - .93 (N groups = 698); **Study 16:** Inter-rater reliability = 85 - 90%, average weighted κ = .80, internal consistency sensitivity subscale α = .90., harshness subscale α = .82 (N centers = 43, N groups = 68)	**Study 4:** CFA 4-factor structure configurial invarianve: RMSEA = .029, CFI = .984, TLI = .981, CFA 4-factor structure strict invariance: RMSEA = .023, CFI = .988, TLI = .989 (N groups = 1350); **Study 7:** Standardized coefficients are highest with CIS and the largest standardized coefficient is seen for ECERS-R factor language-reasoning/interactions with an effect size of .73 (N centers = 1350); **Study 8:** CIS correlated significantly and positively with CIP total score and the separate CIP scale (N centers = 47, N groups = 75, N teachers = 145); **Study 9:** Ratings on CIS were moderately to highly related to scores on TSRS (N teachers = 63); **Study 10:** Correlations between PITC PARS subscale quality of adult's interactions with children and CIS were moderately high and in the expected direction: r = .64, p < .001 with the sensitivity subscale, r = -.62, p < .001 with the harshness subscale, r = -.60, p < .001 with the detachment subscale (N groups = 222, N teachers = 330); **Study 13:** Closest relationship was found between ECERS and CIS sensitivity scale (r = .61-.75), ECERS is moderately correlated with CIS involvement subscale (r = .36 - 53) and acceptance subscale (r = .35 - .59) (N groups = 698)
11	**Study 1:** 100% double-coding, inter-rater reliability r = .82 - .95 in the training phases, internal consistency for dimensions α = .792, α = .748, α = .895, α = .785 and α = .882 (N centers = 90)	**Study 1:** CFA: χ^2/df = 1.772, SRMR = 0.033, CFI = 0.979, AGFI = 0.874, RMSEA = 0.066 (0.043,0.088), IFI = 0.980, TLI = 0.968, GFI = 0.929 (N centers = 90)
12	**Manual:** Compulsory observer certification, 80% agreement within 1-point with the gold standard; **Study 5:** Internal consistency emotional support α = .85, classroom organization α = .92, instructional support α = .80 (N groups = 96, N teachers = 103)	/
13	/	/
14	**Manual:** Compulsory observer certification, 80% agreement within 1-point with the gold standard	**Study 6:** CFA: 3-factor model (χ^2(34) = 82.50, p < 0.001, χ^2/df = 2.43, RMSEA = 0.04, CFI = 0.98, TLI = 0.96, SRMRwithin/between = 0.03/0.04) shows a better model fit then the proposed 2-factor model (N groups = 276, N teachers = 375 teachers); **Study 8:** Good model fit without negative climate (χ^2(13) = 11.40, p = .58; CFI = 1.00; RMSEA = .00) (N groups = 276, N teachers = 375)

15	**Manual:** Compulsory observer certification, 80% agreement within 1-point with the gold standard; **Study 1:** Internal consistency emotional support α = .77, classroom organization α = .80, instructional support α = .79, reliabilities of the dimensions were moderate to high, emotional support r = .28 - .85, classroom organization r = .48 - .68, instructional support = .37 - .65 (N centers = 95, N groups = 177); **Study 2:** Inter-rater reliability κ = .73, with 93% of ratings within 1-point (N groups = 694); **Study 3:** 20% double-coding, 88% for emotional support, 82% for classroom organization, 80% for instructional support (N groups = 53, N children = 304); **Study 4:** 20% double-coding:, 88% for emotional support, 82% for classroom organization, 80% for instructional support (N centers = 8, N groups = 53, N children= 937); **Study 6:** Double-coding at 11 classrooms, inter-rater reliability = 95% (N groups = 119); **Study 7:** Internal consistency instructional domain α = .91, emotional support somain α = .89 (N groups and teachers = 67); **Study 8:** Double-coding at 12 classrooms, inter-rater reliability = 78% - 92%, internal constistency emotional support α = .85, classroom organization α = .84, instructional support α = .78 (N centers = 64, N groups = 128); **Study 9:** Inter-rater reliability emotional support r = .84, classroom organization r = .82, instructional support r = .80, internal consistency emotional support α = .88, classroom organization α = .95, instructional support α = .91 (N groups = 12, N teachers = 57, N children = 1218); **Study 10:** Inter-rater reliability ICC = .82 - .92, internal consistency α = .89 for the overall level, emotional support α = .78, classroom organization α = .84, instructional support α = .92 (N centers = 60, N groups = 180); **Study 11:** Internal reliabilities at T1: emotional support α = .82, classroom organization α = .80, instructional support α = .92 and at T2: emotional support α = .86, classroom organization α = .82, instructional support α = .91 (, N groups = 29N children = 567); **Study 12:** In every classrooms one circle was videotaped that was double-coded afterwards, inter-rater reliability r > .80 (N teachers = 35); **Study 13:** Internal consistency emotional support α = .88, classroom organization α = .88, instructional support α = .88 (N groups = 98); **Study 15:** 100% double-coding, inter-rater reliability r = .80 - .94, internal consistency emotional support α = .93, classroom organization α = .88, instructional support α = .90, test-retest reliability (calculating correlations between two separate days of observation) r = .44 (productivity) - .80 (teacher sensitivity) (N groups and teachers = 49); **Study 16:** 100% double-coding, inter-rater reliability r = .80 - .96, test-retest reliability (calculating correlations between two separate days of observation) r = .55 (instructional learning formats) - .80 (teacher sensitivity), low for two items: r = .44 (productivity), r = .45 (concept development) (N groups and teachers = 49); **Study 17:** 100% double-coding, inter-rater reliability r = .80 - .94, internal consistency emotional support α =.93, classroom organization α = .88, instructional support α = .90, test-retest reliability (calculating correlations between two separate days of observation) r = .44 (productivity) - .80 (teacher sensitivity) (N groups and teachers = 49); **Study 20:** Inter-rater reliability r = .80 - .96 on the first observation day (except for .63 for concept development) and r = .76 - .94 on the second observation day (N groups and teachers = 49, N children = 515); **Study 26:** 15% double-coding, emotional support ICC = .812, classroom organization ICC = .883, instructional support ICC = .902, internal consistency α = .77 - .96 (N groups = 85, N children = 820); **Study 27:** 20.7% double-coding, percent agreement for emotional behavior = 90.3%, classroom organization = 81.2%, instructional support = 94.1% (N teachers = 88); **Study 30:** Inter-rater reliability emotional support ICC = .73, classroom organization ICC = .71, instructional support ICC = .65 (N teachers = 140, N children = 1371); **Study 31:** 20% double-coding, inter-rater reliability emotional support ICC = .82, classroom organization ICC = .76, instructional support ICC = .73, internal consistency emotional support α = .89, classroom organization α = .84, instructional support α =.87 (N teachers = 146, N children = 345)	**Study 1:** CFA: 3-factor structure fits the best: $\chi2$/df = 48.15/29, p < 0.05, CFI = 0.94, TLI = 0.91, RMSEA = 0.06, factor correlations between emotional support and instructional support (r = .31) and between classroom organization and instructional support (r = .20) were low, factor correlation between emotional support and classroom organization (r = .66) was high (N centers = 95, N groups = 177); **Study 6**: ECERS-3 total score was moderately correlated with the domains of CLASS Pre-K (N groups = 95); **Study 8:** 3-factor model did not demonstrate adequate fit: $\chi2(32)$ = 111.85, p < .001; RMSEA = .14; CFI = .90; TLI = .86; SRMR = .09, 3-factor model fit with error correlations: $\chi2(29)$ = 63.78, p < .001; RMSEA = .10; CFI = .96; TLI = .93; SRMR = .05 (N centers = 64, N groups = 128); **Study 10:** 3-factor structure fits the data best with a reasonable fit: $\chi2$ = 161.89, df = 49, p < .001; TLI = .92, CFI = .91, RMSEA = .11 (N centers = 60, N groups = 180); **Study 16:** 3-factor model did not fit the data: $\chi2(32)$ = 112.44, p < .001; CFI = 0.87; TLI = 0.82; RMSEA = 0.23; SRMR = 0.08, 1-factor model did not fit the data: $\chi2(27)$ = 116.27, p = .000; CFI = 0.83; TLI = 0.78; RMSEA = 0.26; SRMR = 0.06, 3-factor model without negative climate and with correlations fits the data well: $\chi2$ (23) = 45.16, p = .004; CFI = 0.96; TLI = 0.94; RMSEA = 0.14; SRMR = 0.04 (N groups and teachers = 49); **Study 27:** Relationship between behavior management effectiveness (SBTR) and group organozation (CLASS Pre-K) was significantly stronger (r = .55, p < .001) than the relationship between behavior management (SBTR) and instructional support (CLASS Pre-K) (r = .31, p = .003), t(78) = 2.82, p = .006, more instances of challenging behavior (SBTR) were related to a lower group organization (CLASS Pre-K) score (r = -.21, p = .027), a higher group organization (CLASS Pre-K) score was related to a higher percent of challenging behaviors acknowledged appropriately by the teacher and the appropriate provision of consequences (SBTR) (r = .38, p = . 001 and r = . 39, p = .001), a lower group organization (CLASS Pre-K) score was related to teachers inappropriate response to behaviors (SBTR) (r = -.21, p = . 004), all correlations between the global SBTR ratings and the behavior management (CLASS Pre-K) were significant, more instances of challanging behavior (SBTR) were associated with a lower behavior management (CLASS Pre-K) score (r = -.24, p = -015), a high behavior management (CLASS Pre-K) score was related to teachers' appropriate acknowledgment of the behavior (r = .36. p < .001) and the appropriate provision of a consequence (SBTR) (r = .42, p < .001), praise (SBTR) was related to the behavior management (CLASS Pre-K) score (r = .37, p < .001) (N teachers = 88)
16	**Study 1:** Inter-rater reliability on 10% of the data, κ = .88, p < .001 (N centers = 18, N children = 179); **Study 2:** 100% double-coding, ICC average measures = .78, internal consistency α = .72 (N teachers = 140, N children = 1371)	/
17	**Study 1:** Internal consistency α = .70, standardized item α = .86. (N centers = 31)	/
18	**Study 1:** Requirement of inter-rater reliability of at least 80 % before data collection, internal consistency for extention dimensions α = .77 (N centers and schools = 53, N teachers = 60)	/
19	**Study 1:** Inter-rater reliability maintained at r > .80 (N teachers = 57)	/
20	/	/
21	**Study 2:** Double-coding at 75 groups, inter-rater reliability r = .74 - .90, p < .001 (N groups and teachers = 83)	**Study 2:** CFA for child-centerted dimension: $\chi2(51)$ = 84.03, p < .01, CFI = 0.95, TLI = 0.93, RMSEA = 0.09, SRMR = 0.05, CFA for teacher-directed dimension: $\chi2(51)$ = 96.31, p < .01, CFI = 0.91, TLI = 0.89, RMSEA = 0.10, SRMR= 0.06, CFA for child-dominated dimension: $\chi2(51)$ = 109.59, p < .01, CFI = 0.85, TLI = 0.81, RMSEA = 0.12, SRMR = 0.08 (N groups and teachers = 83)
22	**Study 4:** Internal consistency personal care and fine and gross motor activities α = .44, furnishings and display for children α = .76, median value α = .57, total ECERS α = .78, social development α =.15 (N groups = 30); **Study 7:** Internal consistency total ECERS α = .96, subscales range α = .82 (social development) - α = .93 (reasoning and language) (N centers = 60, N groups = 120, N children = 526); **Study 8:** Inter-rater reliability = 85 - 94% (M = 90%) (N groups = 9); **Study 9:** One of the values of α below the commonly accepted lower limit of .70 (personal care routines) (N centers = 113); **Study 10:** Inter-rater reliability = 89.5% (N centers = 19); **Study 11:** Inter-rater reliability r = .83 - .98 (M = .94.), internal consistency α = .82 - .96 (N groups = 698);	**Study 4:** Caregiver behavior factors of caregiver-child interaction behaviors and the ECERS subscales were mostly uncorrelated (mean r = .23), EFA: 7-factors accounted for 72% of the variance (N groups = 30); **Study 7:** All items correlated with the total scale with strong association (r ≥ .50, p < .0001), with the exception of items sand/water and space to be alone (r = .39 and .36, p < .0001), correlation between ECERS and CIS r = .51, p < .0001 (N centers= 60, N groups = 120, N children = 526); **Study 11:** Closest relationship was found between ECERS and CIS sensitivity scale (r = .61 - .75), ECERS is moderately correlated with CIS involvement subscale (r = .36 - 53) and acceptance subscale (r = .35 - .59) (N groups = 698)

	Study 13: Inter-rater reliability κ = .74, inter-rater reliability = 85 - 90% (N centers = 43, N groups = 68)	
23	**Study 3:** Internal consistency α = .72 - .94 (N groups = 1350); **Study 13:**15% double-coding, inter-rater reliability = 80% (κ > .60) (N groups and centers = 66); **Study 14:**100% double-coding, κ for mean total = .80, κ for items = .78, κ for indicators = .91, internal consistency α = .95 (N centers = 50); **Study 17:**10% double-coding, inter-rater reliability = 87% (N centers = 47, N groups = 75, N teachers = 145); **Study 18:** 10% double-coding, inter-rater reliability = 89% (N centers and groups = 200); **Study 20:** 20% double-coding, inter-rater reliability *r* = .85 - .99 (N centers and groups = 44); **Study 21:** 20% double-coding, inter-rater reliability *r* = .85 - .99 (N centers and groups = 44); **Study 22:** 100% double-coding, inter-rater reliability *r* = .90 (N centers = 35, N children = 138); **Study 26:** Internal consistency α = .88 (N groups = 222, N teachers = 330); **Study 30:** Internal consistency α = .95 (N groups = 296); **Study 32:** Internal consistency α = .92 (N groups = 10); **Study 33:** Internal consistency α = .90 (N centers = 1); **Study 37:** Inter-rater reliability = 92% (range 86% - 98%) (N groups = 44, N teachers = 72, N children = 636); **Study 40:** Internal consistency α = .70 - .92 (N teachers = 223); **Study 42:** Inter-rater reliability = 86% (range 83% - 87%), internal consistency α = .89 (N centers and groups = 46, N children = 85); **Study 45:** Internal consistency α = .94 for Turkey and α = .90 for the USA (N centers = 20)	**Study 3:** CFA configurial invarianve: inadequate fit of the model, RMSEA = .029 with a 9% confidence interval of .027 to .032., CFI = .946, TLI = .942, CFA structure strict invariance: good fit, RMSEA = .024 with a 95% confidence interval of .022 to .027, CFI = .988, TLI = .989 (N groups = 1350); **Study 10:** 6-factor solution is feasible (NNFI=.990 and RMSEA =.044, without parents and staff), 1-factor solution is not consistent (NNFI=.906 and RMSEA=.133), 3-factor solution was viable (NNFI=.980 and RMSEA=.062), standardized coefficients are highest with CIS and the largest standardized coefficient is seen for ECERS-R factor language-reasoning/interactions with an effect size of .73 (N centers = 1350); **Study 14:** Item correlations between ECERS-R and CIS with *r* = .31 - .58, CFA revealed a 2-factor solution (teaching/interaction and provisions for learning) (N centers = 10); **Study 15:** Moderate positive correlations between ACEI GGA and ECERS-R with *r* = .43 - .70 for subscales from both instruments and *r* = .55 - .70 for total ACEI GGA (N centers = 44); **Study 17:** CIP total score and separate CIP scale scores all correlated significantly and positively with ITERS-R/ECERS-R total score (N centers = 47, N groups = 75, N teachers = 145); **Study 18:** Correlations between CIP scores and ITERS-R/ECERS-R scores were significant and moderate with *r* = .11 - .49 (N centers and groups = 200); **Study 26:** PITC PARS summary rating was found to correlate highly with ECERS-R, *r*(40) = .81, *p* < .001 (N groups = 222, N teachers = 330); **Study 28:** Correlation between ECERS-R and ECERS-3 was positive and modest (*r* = .51, *p* < .001) (N observations = 225); **Study 30:** Average inter-item correlation *r* = .39, item-total correlations *r* = .35 - .76 (*M* =.63), subscale correlations *r* = .48 - .76 (*M* = .62), factor analysis results confirmed that fewer than seven subscales existed, 3-factors were retained by the Kaiser criterion (factor 1: eigenvalue of 13.85, explained 71% of the common variance, factor 2: eigenvalue of 1.93, explained 10% of the common variance, factor 3: eigenvalue of 1.12, explained 6% of the common variance) (N groups = 296)
24	**Study 1:** 100% double-coding, inter-rater reliability *r* = .92 (N centers = 35, N children = 138); **Study 4:** Observers had to meet the following inter-rater reliability standard prior to data collection in field: (1) intra-class correlation exceeding .70 (*M* = .86), (2) correlation exceeding .70 (*M* = .86), (3) mean difference in scores less than .75 (*M* = .43), (4) score agreement within 1-point of at least 80% (*M* = 93%) (N centers = 257, N groups = 323)	/
25	**Study 1:** Inter-rater reliability = 85% within 1-point across all items on two consecutive visits, internal consistency α = .93 (N groups = 1063, N children = 575)	**Study 1:** CFA: 1-factor structure was weak (RMSEA = .081, CFI = .727, Chi-Square(560) = 4429.08, *p* = .0000), 6-factor structure was weak (RMSEA =.104, CFI = .548, Chi-Square(561) = 6979.28, *p* = .0000), 4-factor structure provided the best combination of statistical support and theoretical utility (RMSEA = .046, CFI = .927, Chi-Square (461) = 1486.21, *p* = .0000), ECERS-3 total score were moderately correlated with the domains of CLASS Pre-K (N groups = 1063, for correlation with CLASS Pre-K N groups = 95, N children = 575); **Study 2:** Correlation between ECERS-R and ECERS-3 was positive and modest (*r* = .51, *p* < .001) (N observations = 225)
26	**Study 1:** 10% double-coding, inter-rater reliability *r* = .85., internal consistency α = .96 (N centers = 69, N groups = 135, N teachers = 350)	**Study 1:** ECS was significantly correlated with all but one of the CLASS subscales with *r* = .41 - . 55, ECS scale score was negatively and significantly correlatedwith negative climate (*r* = -.34), all correlations between ECS scale with ITERS-R subscales were significant and positive with *r* = .21 - .40 (N centers = 69, N groups = 135, N teachers = 350)
27	**Study 1:** 25% double-coding, inter-rater reliability identifying challanging situation κ = .91, weighted κ = .76, coding procedure κ = .72 - .87, weighted κ = .95 (N teachers = 9, N children = 28)	/
28	**Study 1:** 10% double-coding, inter-rater reliability = 86% (N teachers = 170, N children = 170); **Study 2:** 10% double-coding, inter-rater reliability = 85.84% (N centers = 122, N children = 179); **Study 3:** 10% double-coding, inter-rater reliability = 91% (N centers = 25, N teacher = 37, N children = 80); **Study 4:** Internal consistency α = .56 - .93 (N centers = 62, N teachers = 181)	**Study 3:** Correlations between EQOS and ECERS-R with *r* = .36 (N centers = 25, N teachers = 37, N children = 80)
29	**Study 1:** Inter-rater reliability κ = .62, internal consistency α = .93 (as reported by original author)	/
30	**Study 1:** 30% double-coding, inter-rater reliability *r* = .83 - .98 (M = .92), κ = .55 - .70 (M = .63), internal consistency α = .70 (N = 99 teachers when children were 12 months, N = 73 teachers when children were 30 months)	/
31	/	/
32	/	/
33	**Study 4:** Internal consistency α = .97 (range .85 - .94) (N groups = 63), **Study 5:** Internal consistency α = .91 (N groups = 10); **Study 8:** Internal consistency α = .69 - .93 (N teachers = 223)	**Study 4:** Item-total correlations were significant (*p* < .0001) with most being greater than *r* = .70 (range .47 - .87.), correlations between items and their corresponding subscale were significant, *r* = .58 - .94 (*p* < .0001), correlations between the subscales and the total scale as well as subscale inter-correlations were strong and significant, with the total scale above *r* = .70 and among the subscales between *r* = .45 - .89, correlation with CIS = .56 (*p* < .0001) (N groups = 63)
34	**For all studies:** Training prior the field observation with inter-rater agreement of at least 80% within 1-point; **Study 1:** Internal consistency α = .91(N groups = 153); **Study 2:** Inter-rater reliability = 88% (range 85 - 96%) (N center = 93, N groups = 206, N teachers = 168); **Study 3:** Internal consistency α = .84 and rho α = .86 for all subscales (N centers = 93, N groups = 206, N children = 2811); **Study 4:** 10% double-coding, inter-rater reliability = 87% (N centers = 47, N groups = 75, N teachers = 145); **Study 5:** 10% double-coding, inter-rater reliability = 89% (N centers = 200); **Study 8:** Internal consistency α = .92 (N groups = 222, N teachers = 330); **Study 11:** Internal consistency α = .87 (N centers = 1); **Study 12:** Inter-rater reliability for the three rounds of program observations = 86%, internal consistency α = .88 - .89 (N centers = 92, N children = 936)	**Study 1:** EFA revealed that the scale measures one global aspect (N groups = 153); **Study 2:** The strongest associations were found between ITERS-R scales listening and talking, interactions, and program structure and CIP scales verbal communication and developmental stimulation (N center = 93, N groups = 206, N teachers = 168); **Study 4:** CIP total score and separate CIP scale scores all correlated significantly and positively with ITERS-R/ECERS-R total score (N centers = 47, N groups = 75, N teachers = 145); **Study 5:** Correlations between CIP scores and ITERS-R/ECERS-R scores were significant and moderate with *r* = .11 - .49 (N centers = 200); **Study 8:** PITC PARS summary rating was found to correlate highly with overall ratings of ITERS-R with *r*(98) = .84, *p* < .001 (N groups = 222, N teachers = 330); **Study 10**: All correlations between ECS scale with ITERS-R subscales were significant and positive with *r* = .21 - .40 (N centers = 69, N groups = 135, N teachers = 350)

35	/	/
36	**Study 1:** Inter-rater reliability κ = .63 - .86 (p < .001) (N centers = 7, N groups = 14)	**Study 1:** Results of the principal factors extraction method with varimax rotation showed 7-factors with eigenvalues over 1, but the scree plot suggested there were 3-factors, items within the subscale interaction are related to curriculum and routine care (N centers = 26, N groups = 50)
37	/	/
38	/	/
39	/	/
40	**Study 1:** 30% double-coding, inter-rater reliability guidance r = .94, directives r = .96, overall κ = .90 (N centers and groups = 44, N teachers = 88)	/
41	/	/
42	**Study 1:** Inter-rater reliability r = .98 (N children = 83); **Study 2:** Inter-rater reliability = 79 - 85% (N teachers = 1)	/
43	**Study 1:** 20% double-coding, inter-rater reliability attention κ = .75, sensitivity κ = .83, stimulation κ = .81 (N groups = 21, N teachers = 37)	/
44	**Study 1:** Inter-rater reliability (mixed ANOVA, absolute agreement) = .88 (range .84 - .91), internal consistency α = .88 (N groups and teachers = 40 , N children = 226)	/
45	**Study 1:** Training prior the field observation with inter-rater agreement of at least 80%, internal consistency α = .91 (N centers = 91, N groups = 174)	**Study 1:** CFA: CMIN/DF = 1.82, RMSEA = 0.07, NFI = 0.95, CFI = 0.97, NFI = 0.97 (N centers = 91, N groups = 174)
46	**Study 1:** Inter-rater reliability total score r = .77, affection r = .80, responsiveness r = .76, encouragement r = .73, teaching r = .69, internal cosnsitency for the total instrument α = .91, affection α = .78, responsiveness α = .75, encouragement α = .77, teaching α = .80 (as reported by original authors)	/
47	**Study 1:** 80% of the subitems across three successive observations have to be in line with the gold standard before observing in the field, internal consistency α = .70 - .90 (N groups = 222, N teachers = 330)	**Study 1:** Correlations between PITC PARS subscale quality of adult's interactions with children and CIS were moderately high and in the expected direction: r = .64, p < .001 with sensitivity subscale, r = -.62, p < .001 with sarshness subscale, r = -.60, p < .001 with detachment subscale, PITC PARS summary rating was found to correlate highly with overall ratings of ERS instruments: ITERS-R, r(98) = .84, p < .001; ECERS-R, r(40) = .81, p < .001; FDCRS, r(80) = .80, p < .001, factor analysis revealed a 3-, not 5-factor structure (α: Factor 1 = .90, Factor 2 = .74, Factor 3 = .80) (N groups = 222, N teachers = 330)
48	**Study 1:** Reliability observation on 19.7%, inter-rater reliability = 71.4 - 92.9%, r = .51 (p < .01) - .82 (p < .05) (N teachers = 88)	**Study 1:** Relationship between behavior management effectiveness (SBTR) and group organozation (CLASS Pre-K) was significantly stronger (r = .55, p < .001) than the relationship between behavior management (SBTR) and instructional support (CLASS Pre-K) (r = .31, p = .003), t(78) = 2.82, p = .006, more instances of challenging behavior (SBTR) were related to a lower group organization (CLASS Pre-K) score (r = -.21, p = .027), a higher group organization (CLASS Pre-K) score was related to a higher percent of challenging behaviors acknowledged appropriately by the teacher and the appropriate provision of consequences (SBTR) (r = .38, p = . 001 and r = . 39, p = .001), a lower group organization (CLASS Pre-K) score was related to teachers inappropriate response to behaviors (SBTR) (r = -.21, p = . 004), all correlations between the global SBTR ratings and the behavior management (CLASS Pre-K) were significant, more instances of challanging behavior (SBTR) were associated with a lower behavior management (CLASS Pre-K) score (r = -.24, p = -015), a high behavior management (CLASS Pre-K) score was related to teachers' appropriate acknowledgment of the behavior (r = .36. p < .001) and the appropriate provision of a consequence (SBTR) (r = .42, p < .001), praise (SBTR) was related to the behavior management (CLASS Pre-K) score (r = .37, p < .001) (N teachers = 88)
49	**Study 1:** Observers had to meet the following inter-rater reliability standard prior to data collection in field: (1) intra-class correlation exceeding .70 (M = .86), (2) correlation exceeding .70 (M = .86), (3) mean difference in scores less than .75 (M = .43), (4) score agreement within -point of at least 80% (M = 93%) (N centers = 257, N groups = 323)	/
50	/	/
51	**Study 1:** Inter-rater reliability = 96% (range 95 - 99%) (N groups = 44 , N teachers = 72, N children = 636)	/
52	**Study 1:** 100% double-coding, positive teacher interactions ICC = .80, negative teacher interactions ICC = .63, internal consistency positive teacher interactions α = .90, negative teacher interactions α = .60 (N centers = 94, N groups = 173, N teachers = 183); **Study 2**: For teacher interactive behaviors, all free play videos were double-coded and 20% of clean-up videos were double-coded, for child interactive behaviors, all free play and clean-up videos were double-coded, inter-rater reliability for positive teacher interactions r = .85 during free play, r = .80 during clean-up, inter-rater reliability for negative teacher interactions r = .82 during free play, r = .63 during clean-up, Inter-rater reliability for child active engagement r = .78 during free play, r = .90 during clean-up, inter-rater reliability for child positive interactions with teacher r =.78 during free play, r = .85 during clean-up, internal consistency for positive teacher behavior α = .91 during free play, α = .90 during clean-up, internal consistency for negative teacher behavior α = .50, internal consistency for child active engagement α = .73 during free play, α = .85 during clean-up, internal consistency for child positive interactions with teacher α = .87 during free play, α = .72 during clean-up (N teachers = 146, N children = 345)	/
53	**Study 1:** 100% double-coding, inter-rater reliability r = .93, (range .64 - 1.00) (N centers = 18, N teachers = 22, N children = 174)	/
54	**Study 1:** Internal consistency for affective characteristics items α = .89 (N teachers = 63)	**Study 1:** Ratings on TSRS were moderately to highly related to scores on CIS (N teachers = 63)
55	**Study 1:** 28 min out of 540 min were double-coded, agreement within ±2% (N children = 6)	/

Note. Only construct and convergent validity are reported; Reliability and validity are reported as reported in the studies.

Table 5. Author(s), year and country of identified studies.

	Author(s), (year), country
1	**Study 1:** Honig, A. S., Lally, J. R. (1988), USA
2	**Study 1:** Pascal, C., Bertram, T. (1999), UK
3	**Study 1:** Ruprecht, K., Elicker, J., Choi, J. Y. (2016), USA
4	**Study 1:** Oliveira, p. S., Fearon, R. M. P., Belsky, J., Fachada, I., Soares, I. (2015), UK, USA, Portugal
5	**Study 1:** Perlman, M., Brunsek, A., Hepditch, A., Gray, K., Falenchuck, O. (2017), Canada
6	**Study 1:** Beller, E. K., Stahnke, M., Butz, P., Stahl, W., Wessels, H. (1996), Germany
7	**Study 1:** Tanyel, N., Knopf, H. T. (2011), USA
8	**Study 1:** Hu, B. Y., Vong, K., Chen, Y., Li, K. (2015), China; **Study 2:** Li, K., Hu, B. Y., Pan, Y., Qin, J., Fan, X. (2014), China, USA; **Study 3:** Li, K., Zhang, P., Hu, B. Y., Burchinal, M. R., Fan, X., Qin, J. (2019), China, USA; **Study 4:** Liu, X., Hu, B., Huang, J. (2019), China, Thailand
9	**Study 1:** Baustad, A. G., Bjørnestad, E. (2020), Norway; **Study 2:** Bjørnestad, E., Broekhuizen, M. L., Os, E., Baustad, A. G. (2020), Norway, the Netherlands; **Study 3:** Helmerhorst, K. O. W., Riksen-Walraven, J. M., Vermeer, H. J., Fukkink, R. G., Tavecchio, L. W. C. (2014), the Netherlands; **Study 4:** Helmerhorst, K. O. W., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., Tavecchio, L. W. C., Fukkink, R. G. (2015), the Netherlands; **Study 5:** Jilink, L., Fukkink, R., Huijbregts, S. (2018), the Netherlands
10	**Study 1:** Auger, A., Farkas, G., Duncan, G., Burchinal, P., Vandell, D. L., Society for Research on Educational Effectiveness (SREE) (2012), USA; **Study 2:** Bassok, D., Fitzpatrick, M., Greenberg, E., Loeb, S. (2016), USA; **Study 3:** Biringen, Z., Altenhofen, S., Aberle, J., Baker, M., Brosal, A., Bennett, S., Coker, E., Lee, C., Meyer, B., Moorlag, A., Swaim, R. (2012), USA; **Study 4:** Bishop, C. D. (2013), PhD diss., USA; **Study 5:** Campbell, P., Milbourne, S., Silverman, C., Feller, N. (2005), USA; **Study 6:** Connors, M. C., Friedman-Krauss, A. H., Morris, P. A., Page, L. C., Feller, A. (2014), USA; **Study 7:** Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., Abner, K. (2013), USA; **Study 8:** Helmerhorst, K. O. W., Riksen-Walraven, J. M., Vermeer, H. J., Fukkink, R. G., Tavecchio, L. W. C. (2014), the Netherlands; **Study 9:** De Kruif, R. E.L., McWilliam, R. A., Ridley, S. M., Wakely, M. B. (2000), USA; **Study 10:** Mangione, P. L., Kriener-Althen, K., Marcella, J. (2016), USA; **Study 11:** Manlove, E. E., Vazquez, A., Vernon-Feagans, L. (2008), USA; **Study 12:** Rentzou, K. (2017), Cyprus; **Study 13:** Tietze, W., Cryer, D., Bairrao, J., Palacios, J., Wetzel, G. (1996), Germany, USA, Portugal, Spain, Austria; **Study 14:** Torquati, J. C., Raikes, H., Huddleston-Casas, C. A. (2007), USA; **Study 15:** Whitebook, M., Sakai, L., Howes, C. (1997), USA; **Study 16:** Whitebook, M., Sakai, L. M., Howes, C. (2004), USA
11	**Study 1:** Pan, Y.-J., Liu, Y., Yang, Q.-Q., Zheng, X.-L., Wu, X., Song, L.-Q. (2020), China, USA
12	**Study 1:** Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., Howes, C., Barbarin, O. (2006), USA; **Study 2:** Decker-Woodrow, L. (2018), USA; **Study 3:** Downer, J. T., Lopez, M. L., Grimm, K. J., Hamagami, A., Pianta, R. C., Howes, C., National Center for Research on Early Childhood Education (NCRECE), USA; **Study 4:** Howes, C., Fuligni, A. S., Hong, S. S., Huang, Y. D., Lara-Cinisomo, S. (2013), USA; **Study 5:** Jeon, L. Buettner, C. K., Hur, E. (2014), USA; **Study 6:** Wilcox-Herzog, A., McLaren, M., Ward, S., Wong, E. (2013), USA
13	**Study 1:** Research for Action (2018), USA
14	**Study 1:** Henry, A. J. L., Hatfield, B. E., Chandler, K. D. (2021), USA; **Study 2:** Kennedy, A. S., Lees, A. T. (2015), USA; **Study 3:** Pauker, S., Perlman, M., Prime, H., Jenkins, J. (2018), Canada; **Study 4:** Research for Action (2018), USA; **Study 5:** Rodriguez, B. J., Garza, S. (2014), USA; **Study 6:** Slot, P. L., Boom, J. Verhagen, J., Leseman, P. P. M. (2017), the Netherlands; **Study 7:** Slot, P. L., Leseman, P. P. M.; Verhagen, J., Mulder, H. (2015), the Netherlands; **Study 8:** Van Schaik, S. D. M., Leseman, P. P. M., de Haanm, M. (2017), the Netherlands
15	**Study 1:** Bihler, L.-M., Agache, A., Kohl, K., Willard, J. A., Leyendecker, B. (2018), Germany; **Study 2:** Brock, L. L., Curby, T. W. (2014), USA; **Study 3:** Bulotsky-Shearer, R J., Bell, E. R., Carter, T. M., Dietrich, S. L. R. (2014), USA; **Study 4:** Bulotsky-Shearer, R. J., Fernandez, V. A., Bichay-Awadalla, K., Bailey, J., Futterer, J., Qi, C. H. (2020), USA; **Study 5:** Craft-Reiss, B. S. (2012), PhD diss., USA; **Study 6:** Early, D. M., Sideris, J., Neitzel, J., LaForett, D. R., Nehler, C. G. (2018), USA; **Study 7:** Guo, Y., Piasta, S. B., Justice, L. M., Kaderavek, J. N. (2010), USA; **Study 8:** Hanno, E. C., Gonzalez, K. E., Lebowitz, R. B., McCoy, D. C., Lizarraga, A., Korder Fort, C. (2020), USA, Peru; **Study 9:** Hoang, N., Holopainen, L., Siekkinen, M. (2019), Finland; **Study 10:** Hu, B. Y., Fan, X., Gu, C., Yang, N. (2016), China; **Study 11:** Hu, B. Y., Wu, H., Curby, T. W., Wu, Z., Zhang, X. (2018), China, USA; **Study 12:** Jennings, P. A. (2015), USA; **Study 13:** King, E. K., Johnson, A. V., Cassidy, D. J., Wang, Y. C., Lower, J. K., Kintner-Duffy, V. L. (2016), USA; **Study 14:** Lemay, L., Cantin, G., Lemire, J., Bouchard, C. (2018), Canada; **Study 15:** Pakarinen, E., Kiuru, N., Lerkkanen, M.-K., Poikkeus, A.-M., Ahonen, T., Nurmi, J.-E. (2011), Finland; **Study 16:** Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., Nurmi, J.-E. (2010), Finland; **Study 17:** Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Siekkinen, M., Nurmi, J.-E. (2011), Finland; **Study 18:** Perlman, M., Howe, N., Gulyas, C., Falenchuk, O. (2019), Canada; **Study 19:** Research for Action (2018), USA; **Study 20:** Salminen, J., Pakarinen, E., Poikkeus, A.-M., Lerkkanen, M.-K. (2018), Finland; **Study 21:** Sandilos, L. E., Goble, P., Rimm-Kaufman, S. E., Pianta, R. C. (2018), USA; **Study 22:** Stanton-Chapman, T. L., Walker, V. L., Voorhees, M. D., Snell, M. E. (2016), USA; **Study 23:** Tayler, C., Ishimine, K., Cloney, D., Cleveland, G., Thorpe, K. (2013), Australia, Canada; **Study 24:** Tonge, K. L., Jones, R. A., Okely, A. D. (2019), Australia; **Study 25:** Veraksa, A., Bukhalenkova, D., Almazova, O. (2020), Russia; **Study 26:** Vitiello, V. E., Bassok, D., Hamre, B. K., Player, D., Williford, A. P. (2018), USA; **Study 27:** Vujnovic, R. K., Fabiano, G. A., Waschbusch, D. A., Pelham, W. E., Greiner, A., Gera, S., Linke, S., Gormley, M., Buck M. (2014), USA; **Study 28:** Wasik, B. A., Hindman, A. H. (2011), USA; **Study 29:** Whitebook, M., King, E., Philipp, G., Sakai, L. (2016), USA; **Study 30:** Whittaker, J.V., Kinzie, M. B., Vitiello, V., DeCoster, J., Mulcahy, C., Barton E. A. (2020), USA; **Study 31:** Whittaker, J. E. V., Williford, A. P., Carter, L. M., Vitiello, V. E., Hatfield, B. E. (2018), USA; **Study 32:** Williford, A. P., Carter, L. M., Maier, M. F., Hamre, B. K., Cash, A., Pianta, R. C., Downer, J. T. (2017), USA; **Study 33:** Zeng, S. (2017), USA
16	**Study 1:** Bojorque, G., Torbeyns, J., van Nijlen, D., Verschaffel, L. (2018), Belgium; **Study 2:** Whittaker, J.V., Kinzie, M. B., Vitiello, V., DeCoster, J., Mulcahy, C., Barton E. A. (2020), USA
17	**Study 1:** Bartkowiak, E. T. (1996), PhD diss., USA
18	**Study 1:** Isele, P. (2014), PhD diss., Germany
19	**Study 1:** Biringen, Z., Altenhofen, S., Aberle, J., Baker, M., Brosal, A., Bennett, S., Coker, E., Lee, C., Meyer, B., Moorlag, A., Swaim, R. (2012), USA
20	**Study 1:** Howes, C., Fuligni, A. S., Hong, S. S., Huang, Y. D., Lara-Cinisomo, S. (2013), USA
21	**Study 1:** Burns, M. S., Kidd, J. K., Nasser, I., Aier, D. J., Stechuk, R. (2012), USA; **Study 2:** Lerkkanen, M.-K., Kikas, E., Pakarinen, E., Trossmann, K., Poikkeus, A.-M., Rasku-Puttonen, H., Siekkinen, M., Nurmi, J.-E. (2012), Estonia, Finland
22	**Study 1:** Auger, A., Farkas, G., Duncan, G., Burchinal, P., Vandell, D. L., Society for Research on Educational Effectiveness (SREE) (2012), USA; **Study 2:** Bassok, D., Fitzpatrick, M., Greenberg, E., Loeb, S. (2016), USA; **Study 3:** Baştürk, Ramazan; Işikoğlu, Nesrin (2008), Turkey; **Study 4:** Beller, E. K., Stahnke, M., Butz, P., Stahl, W., Wessels, H. (1996), Germany; **Study 5:** Campbell, P., Milbourne, S., Silverman, C., Feller, N. (2005), USA; **Study 6:** Hadeed, J., Sylva, K. (1995), UK; **Study 7:** Herrera, M. O., Mathiesen, M. E., Merino, J. M., Recart, I. (2005), Chile; **Study 8:** McCormick, L., Noonan, M. J., Ogata, V., Heck, R. (2001), USA; **Study 9:** Munton, A. G., Rowland, L., Mooney, A., Lera, M. J. (1997), UK; **Study 10:** Sheridan, S. (2001), Sweden; **Study 11:** Tietze, W., Cryer, D., Bairrao, J., Palacios, J., Wetzel, G. (1996), Germany, USA, Portugal, Spain, Austria; **Study 12:** Whitebook, M., King, E., Philipp, G., Sakai, L. (2016), USA; **Study 13:** Whitebook, M., Sakai, L., Howes, C. (1997), USA; **Study 14:** Whitebook, M., Sakai, L. M., Howes, C. (2004), USA
23	**Study 1:** Amodei, M. L. (2011), PhD diss., USA; **Study 2:** Barandiaran, A., Muela, A., López de Arana, E., Larrea, I., Vitoria, J. R. (2015), Spain; **Study 3:** Bishop, C. D. (2013), PhD diss., USA; **Study 4:** Brinkman, S. A., Hasan, A., Jung, H., Kinnell, A., Nakajima, N., Pradhan, M. (2017), Australia, USA, the Netherlands, South Korea; **Study 5:** Cassidy, D. J., Hestenes, L. L., Hansen, J. K., Hegde, A., Shim, J., Hestenes, S. (2005), USA; **Study 6:** Connors, M. C., Friedman-Krauss, A. H., Morris, P. A., Page, L. C., Feller, A. (2014), USA; **Study 7:** Decker-Woodrow, L. (2018), USA; **Study 8:** Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., Howes, C., Barbarin, O. (2006), USA; **Study 9:** Foster, B. (2017), PhD diss., USA; **Study 10:** Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., Abner, K. (2013), USA; **Study 11:** Grammatikopoulos, V., Gregoriadis, A., Tsigilis, N., Zachopoulou, E. (2018), Greece; **Study 12:** Gregoriadis, A., Tsigilis, N., Grammatikopoulos, V., Kouli, O. (2016), Greece; **Study 13:** Grisham-Brown, J., Cox, M., Gravil, M., Missall, K. (2010), USA; **Study 14:** Hadeed, J. (2014), Bahrain; **Study 15:** Hardin, B. J., Bergen, D., Busio, D. S., Boone, W. (2017), USA; **Study 16:** Hartman, S. C., Warash, B. G., Curtis, R., Day Hirst, J. (2016), USA; **Study 17:** Helmerhorst, K. O. W., Riksen-Walraven, J. M., Vermeer, H. J., Fukkink, R. G., Tavecchio, L. W. C. (2014), the Netherlands; **Study 18:** Helmerhorst, K. O. W., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., Tavecchio, L. W. C., Fukkink, R. G. (2015); **Study 19:** Honeycutt, D. (2008), PhD diss., USA; **Study**

	20: Howe, N., Jacobs, E. (2013), Canada; **Study 21:** Howe, N., Jacobs, E., Vukelich, G., Recchia, H. (2011), Canada; **Study 22:** Ishimine, K., Wilson, R., Evans, D. (2010), Australia; **Study 23:** Jeon, L. Buettner, C. K., Hur, E. (2014), USA; **Study 24:** Jobes, R. D. (2010), PhD diss., USA; **Study 25:** de Kruif, R. E.L., McWilliam, R. A., Ridley, S. M., Wakely, M. B. (2000), USA; **Study 26:** Mangione, P. L., Kriener-Althen, K., Marcella, J. (2016), USA; **Study 27:** Mathers, S., Linskey, F., Seddon, J., Sylva, K. (2007), UK; **Study 28:** Neitzel, J., Early, D., Sideris, J., LaForett, D., Abel, M. B.; Soli, M., Davidson, D. L., Haboush-Deloye, A., Hestenes, L. L., Jenson, D., Johnson, C., Kalas, J., Mamrak, A., Masterson, M. L., Mims, S. U., Oya, P., Philson, B., Showalter, M., Warner-Richter, M., Kortright Wood, J. (2019), USA; **Study 29:** Perlman, M., Howe, N., Gulyas, C., Falenchuk, O. (2019), Canada; **Study 30:** Perlman, M., Zellman, G. L, Le, V.-N. (2004), Canada, USA; **Study 31:** Reinke, S., Peters, L., Castner, D. (2019), USA; **Study 32:** Rentzou, K. (2010), Greece; **Study 33:** Rentzou, K. (2017), Cyprus; **Study 34:** Robertson, N., Yim, B., Paatsch, L. (2020), Australia; **Study 35:** Schilder, D., Smith Leavell, A. (2015), USA; **Study 36:** Setodji, C. M., Schaack, D., Le, V.-N. (2018), USA; **Study 37:** Shim, J., Hestenes, L., Cassidy, D. (2004), USA; **Study 38:** Stanton-Chapman, T. L., Walker, V. L., Voorhees, M. D., Snell, M. E. (2016), USA; **Study 39:** Tayler, C., Ishimine, K., Cloney, D., Cleveland, G., Thorpe, K. (2013), Australia, Canada; **Study 40:** Torquati, J. C., Raikes, H., Huddleston-Casas, C. A. (2007), USA; **Study 41:** Veraksa, A. N., Gavrilova, M. N., Pons, F. (2020), Norway, Russia; **Study 42:** Vermeer, H. J., Groeneveld, M. G., Larrea, I., van IJzendoorn, M. H., Barandiaran, A., Linting, M. (2010), the Netherlands, Spain; **Study 43:** Waldman-Levi, A., Bar-Haim Erez, A. (2015), Israel; **Study 44:** Wilcox-Herzog, A., McLaren, M., Ward, S., Wong, E. (2013), USA; **Study 45:** Yaya-Bryson, D., Scott-Little, C., Akman, B., Cassidy, D. J. (2020), USA, Turkey; **Study 46:** Zan, B. (2005), USA; **Study 47:** Zeng, S. (2017), USA
24	**Study 1:** Ishimine, K., Wilson, R., Evans, D. (2010), Australia; **Study 2:** Mathers, S., Linskey, F., Seddon, J., Sylva, K. (2007), UK; **Study 3:** Neylon, G. (2014), Australia; **Study 4:** Siraj, I. Howard, S. J., Kingston, D., Neilsen-Hewett, C., Melhuish, E. C., de Rosnay, M. (2019), Australia, UK
25	**Study 1:** Early, D. M., Sideris, J., Neitzel, J., LaForett, D. R., Nehler, C. G. (2018), USA; **Study 2:** Neitzel, J., Early, D., Sideris, J., LaForett, D., Abel, M. B.; Soli, M., Davidson, D. L., Haboush-Deloye, A., Hestenes, L. L., Jenson, D., Johnson, C., Kalas, J., Mamrak, A., Masterson, M. L., Mims, S. U., Oya, P., Philson, B., Showalter, M., Warner-Richter, M., Kortright Wood, J. (2019), USA
26	**Study 1:** Pauker, S., Perlman, M., Prime, H., Jenkins, J. (2018), Canada
27	**Study 1:** Silkenbeumer, J. R., Schiller, E.-M., Kärtner J. (2018), Germany
28	**Study 1:** Bigras, N., Lemay, L., Bouchard, C., Eryasa, J. (2017), Canada; **Study 2:** Lehrer, J. S., Lemay, L., Bigras, N. (2015), Canada; **Study 3:** Lemay, L., Bigras, N., Bouchard, C. (2014), Canada; **Study 4:** Lemay, L., Cantin, G., Lemire, J., Bouchard, C. (2018), Canada
29	**Study 1:** Farkas, C. (2019), Chile
30	**Study 1:** Farkas, C. (2019), Chile
31	**Study 1:** Ponguta, L. A., Maldonado-Carreno, C., Kagan, S. L., Yoshikawa, H., Maria Nieto, A., Aragon, C. A., Mateus, A., Rodriguez, A. M., Motta, A., Ramirez Varela, L., Andrea Guerrero, P., Escallon, E. (2019), USA, Colombia
32	**Study 1:** Mićanović, V., Novović, T. (2015), Montenegro
33	**Study 1:** Bassok, D., Fitzpatrick, M., Greenberg, E., Loeb, S. (2016), USA; **Study 2:** Campbell, P., Milbourne, S., Silverman, C., Feller, N. (2005), USA; **Study 3:** Cain, D. W., Rudd, L. C., Saxon, T. F. (2007), USA; **Study 4:** Herrera, M. O., Mathiesen, M. E., Merino, J. M., Recart, I. (2005), Chile; **Study 5:** De Kruif, R. E.L., McWilliam, R. A., Ridley, S. M., Wakely, M. B. (2000), USA; **Study 6:** Rentzou, K. (2010), Greece; **Study 7:** Rudd, L. C., Cain, D. W., Saxon, T. F. (2008), USA; **Study 8:** Torquati, J. C., Raikes, H., Huddleston-Casas, C. A. (2007), USA
34	**Study 1:** Bisceglia, R., Perlman, M., Schaack, D., Jenkins, J. (2009), Canada, USA; **Study 2:** Bjørnestad, E., Broekhuizen, M. L., Os, E., Baustad, A. G. (2020), Norway, the Netherlands; **Study 3:** Bjørnestad, E., Os, E. (2018), Norway; **Study 4:** Helmerhorst, K. O. W., Riksen-Walraven, J. M., Vermeer, H. J., Fukkink, R. G., Tavecchio, L. W. C. (2014), the Netherlands; **Study 5:** Helmerhorst, K. O. W., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., Tavecchio, L. W. C., Fukkink, R. G. (2015), the Netherlands; **Study 6:** Hooper, A., Hallam, R. (2017), USA; **Study 7:** Kleppe, R. (2018), Norway; **Study 8:** Mangione, P. L., Kriener-Althen, K., Marcella, J. (2016), USA; **Study 9:** Mathers, S., Linskey, F., Seddon, J., Sylva, K. (2007), UK; **Study 10:** Pauker, S., Perlman, M., Prime, H., Jenkins, J. (2018), Canada; **Study 11:** Rentzou, K. (2017), Cyprus; **Study 12:** Weinstock, P., Bos, J., Tseng, F., Rosenthal, E., Ortiz, L., Dowsett, C., Huston, A., Bentley, A. (2012), USA
35	**Study 1:** Harms, T., Cryer, D., Clifford, R. M., Yazejian, N. (2017), USA
36	**Study 1:** Liu-Yan, Pan-Yuejuan (2008), China
37	**Study 1:** Bäumer, T., Roßbach, H.-G. (2016), Germany; **Study 2:** Tietze, W., Schuster, K.-M., Grenner, K., Roßbach, H.-G. (2005), Germany
38	**Study 1:** Bäumer, T., Roßbach, H.-G. (2016), Germany
39	**Study 1:** Tietze, W., Harms, T. (2005), Germany
40	**Study 1:** Howe, N., Jacobs, E., Vukelich, G., Recchia, H. (2011), Canada
41	**Study 1:** Anderson, K., Sayre, R., Brookings Institution, Center for Universal Education, United Nations Children's Fund (UNICEF), United Nations Educational, Scientific, and Cultural Organization, World Bank (2016)
42	**Study 1:** Fernández, R. E. E. R. (2019), PhD diss., USA; **Study 2:** Frank, I., Stolarski, E., Scher, A. (2006), Israel
43	**Study 1:** Van Polanen, M., Colonnesi, C., Tavecchio, L. W. C., Blokhuis, S., Fukkink, R. G. (2017), the Netherlands
44	**Study 1:** Barandiaran, A., Muela, A., López de Arana, E., Larrea, I., Vitoria, J. R. (2015), Spain
45	**Study 1:** Hu, B. Y., Li, K., De Marco, A., Chen, Y. (2015), China, USA
46	**Study 1:** Farkas, C. (2019), Chile
47	**Study 1:** Mangione, P. L., Kriener-Althen, K., Marcella, J. (2016), USA
48	**Study 1:** Vujnovic, R. K., Fabiano, G. A., Waschbusch, D. A., Pelham, W. E., Greiner, A., Gera, S., Linke, S., Gormley, M., Buck M. (2014), USA
49	**Study 1:** Siraj, I., Howard, S. J., Kingston, D., Neilsen-Hewett, C., Melhuish, E. C., de Rosnay, M. (2019), Australia, UK
50	**Study 1:** Auger, A., Farkas, G., Duncan, G., Burchinal, P., Vandell, D. L., Society for Research on Educational Effectiveness (SREE) (2012), USA
51	**Study 1:** Shim, J., Hestenes, L., Cassidy, D. (2004), USA
52	**Study 1:** LoCasale-Crouch, J., Williford, A., Whittaker, J., DeCoster, J., Alamos, P. (2018), USA; **Study 2:** Whittaker, J. E. V., Williford, A. P., Carter, L. M., Vitiello, V. E., Hatfield, B. E. (2018), USA
53	**Study 1:** Bouchard, C., Bigras, N., Cantin, G., Coutu, S., Blain-Briere, B., Eryasa, J., Charron, A., Brunson, L. (2010), Canada
54	**Study 1:** De Kruif, R. E.L., McWilliam, R. A., Ridley, S. M., Wakely, M. B. (2000), USA
55	**Study 1:** Biemiller, A., Avis, C., Lindsay, A. (1979), Canada

Note. Detailed references of the identified studies are available from the corresponding author.

Appendix – Study 2

Interaction Quality in German Early Childcare Settings: Investigating the Domains of CLASS Toddler and the Associations with Structural Characteristics

Baron, F., Linberg, A., & Lehrl, S. (2023). Interaction Quality in German Early Childcare Settings: Investigating the Domains of CLASS Toddler and the Associations with Structural Characteristics, *Early Child Development and Care, 193*(13–14), 1485–1502. https://doi.org/10.1080/03004430.2023.2256997

The present study examines the quality and domains of teacher–toddler interactions as well as associations with structural characteristics using data from 95 German early childcare settings. Results of the confirmatory factor analysis supported a two-factor structure of interaction quality assessed by CLASS Toddler (La Paro et al., 2012): emotional and behavioural support (EBS) and engaged support for learning (ESL). The EBS domain showed higher quality ratings ($M = 5.33$, $SD = .59$) than the ESL domain ($M = 3.23$, $SD = .70$). Structural equation modelling was applied to estimate associations between those domains and structural characteristics within classrooms. Structural characteristics predicting interaction quality were teachers' age (for EBS), teachers' education (for ESL) and children's age composition in the classroom (for EBS and ESL). Overall, the two-factor structure of CLASS Toddler could be replicated. For high-quality interactions, teacher and classroom characteristics are crucial but need to be carefully distinguished. Beyond their limitations, these findings have implications that are discussed.

Keywords: early childhood education and care (ECEC); toddler classrooms; teacher–child interactions; interaction quality; structural quality; classroom assessment scoring system CLASS toddler

Introduction

In recent years, a growing number of toddlers have been spending time in childcare settings, both in Germany (children under the age of three in childcare 2011: 25% and 2021: 34%; Autor:innengruppe Bildungsberichterstattung, 2022), and worldwide (children under the age of three in childcare 2010: 25% and 2020: 27%; OECD, 2022). Consequently, an increasing number of toddlers make their experiences a substantial amount of time within teacher–child interactions. Such interactions are described as primary mechanisms of child learning and development in childcare (Hamre & Pianta, 2007). Their quality is associated with concurrent and later (academic) achievements across several age groups (e.g., Belsky et al., 2007; Vandell et al., 2016). Despite the increased usage of early childcare in Germany, little is known about its quality, its quality variation, and possible sources of quality variations, especially in toddler classrooms (exception for German toddler classrooms: NUBBEK, Tietze et al., 2013). In general, high-quality interactions include the promotion of a warm atmosphere, a prompt response to toddlers' signals, physical proximity, back-and-forth interactions, and the provision of diverse opportunities for children to explore and learn (Norris et al., 2015). Although there is congruence in research that different facets of interaction quality can be distinguished (e.g., Thomason & La Paro, 2009), which and how many is not clear across different cultural contexts. Moreover, a huge body of research confirms that teacher–child interactions in general are affected by structural characteristics (Kluczniok & Roßbach, 2014; NICHD, 2002). However, those findings mostly refer to preschool children. As described in the following, according to toddlers' early childhood education and care (ECEC), research is inconsistent (e.g., in terms of how interaction quality occurs, see Slot et al., 2017, as well as in terms of how interaction quality is associated with structural characteristics, see Cadima et al., 2022). This study therefore examines dimensions of interaction quality in German ECEC through defining distinct quality domains in toddlers' childcare and their specific relation to structural characteristics.

Teacher–child interaction quality in classrooms

The bio-ecological theory (Bronfenbrenner & Morris, 1998) states that proximal processes are central mechanism for development and learning and also emphasizes the importance of surroundings (e.g., symbols or objects). The theory relates these processes to organisms and environments in general and describes that they are especially important in early phases. Those processes include (but are not limited to) interactions between a child and the environment or other persons. Empirical evidence underlines the significance of interactions as part of processes as being an important mechanism for children's developmental and learning outcomes (Morrison & Connor, 2002; Rutter &

Maughan, 2002). The teaching through interactions (TTI) framework takes up this theoretical and empirical foundation and focuses primarily on teacher–child interactions as psychological and educational research suggests that this is the core of effective teaching (e.g., Brophy, 1999; Roeser et al., 1999). It distinguishes three major domains of those interactions with preschoolers (Hamre & Pianta, 2007). The emotional domain is seen as the teacher's effort to support the child in its social and emotional functioning in the classroom. High-quality interactions in this domain comprise consistent, predictive, and sensitive interactions, which in turn lead to more self-reliability and self-determination of the child. On this basis, the child can explore their environment and learn (Hamre et al., 2013). The classroom organization domain involves management strategies with which the teacher helps children to regulate their behaviour and attention. Children in a well-organized classroom with clear and consistent routines are able to engage more in meaningful activities and develop good self-regulatory and executive-functioning skills. The teacher's role is to promote children's interactions by, for example, providing activities or materials (Hamre et al., 2013). The instructional support domain includes teachers' interactions promoting higher-order thinking and connecting learning opportunities to children's lives. Central teacher behaviours in this domain are frequent conversations and feedback that expand children's ideas and encourage them to participate in activities, communicate, and keep learning (Hamre et al., 2013).

Depending on the age of the child, these described domains are accentuated and grouped differently. When focusing on infants, responsive caregiving is often seen as one global construct, with warm and meaningful interactions as a foundation for a healthy development and the acquisition of basic skills. Studies could show that this one-domain approach – which is the emotional domain – is applicable for teacher–child interactions in infancy (Norris et al., 2015), although there are also results pointing to more than one facet when considering parent-infant interactions (Linberg, 2018). Concerning teacher–toddler interactions, primarily two domains are considered to be essential for child learning and development: one that addresses emotional and behavioural needs (emotional domain), and one that addresses learning occasions (instructional domain) (Cadima et al., 2022). In this construct, warm and responsive interactions as well as behavioural support are often seen as the basis for learning (Thomason & La Paro, 2009).

The majority of studies support this two-factor structure (Bichay-Awadalla & Bulotsky-Shearer, 2021; Reyhing et al., 2019; Salminen et al., 2021; van Schaik et al., 2018). In contrast, Slot et al. (2017) could not replicate those two domains of interaction quality in toddlerhood. Instead, they revealed a three-factor solution with an emotional, a behavioural, and an instructional domain, which is in line with research on older children (Suchodoletz et al., 2014).

Structural characteristics of quality

Investigating the relationships between structural characteristics and teacher–child interactions can support the identification of indicators that should be prioritized for ECEC improvements. Of major concern throughout the literature are the structural indicators: teacher characteristics such as teacher education and teacher age, as well as classroom characteristics such as group size and children's age composition (e.g., Locasale-Crouch et al., 2007).

With regard to teacher education, results are inconsistent: some studies point to positive associations between the level of teacher education and overall quality in toddler childcare (Burchinal et al., 2002; Goelman et al., 2006), whereas others have not found these relations or even point to negative impacts (van IJzendoorn et al., 1998; Vermeer et al., 2008). For associations with teachers' age, findings are also varying. Whereas Pessanha et al. (2007) revealed a positive effect of being a young teacher, van IJzendoorn et al. (1998) revealed a positive effect of being an older teacher. Surprisingly, the study of van IJzendoorn et al. (1998) showed that fewer years of experience in childcare settings indicated better quality for toddlers. Taking a closer look at different facets of teacher–toddler interaction, Slot et al. (2015) showed that teacher education had a positive (but small) association with emotional quality and that further teacher training at childcare centres had the strongest relation to both emotional and instructional interaction quality. Cadima et al. (2022) compared the effects of several teacher characteristics on teacher–toddler interactions in different countries (Portugal, Poland, Finland, and the Netherlands) and revealed a negative association between teacher education and emotional and instructional support in Portugal. In contrast, teacher education was positively associated with instructional support in the Netherlands, pointing to the fact that there might be country-related differences.

Besides teacher characteristics, classroom components are also essential for high-quality interactions, as they affect whether and how teachers perceive the needs of children in the classroom. Regarding group size and teacher–child ratio, findings mostly agree across all age groups and different interaction domains that quality is higher when teacher–child ratio and group size are smaller (Barros et al., 2016; Gerber et al., 2007; Løkken et al., 2018; Mashburn et al., 2008). This supports the assumption that high quality can be better implemented in smaller classrooms with enough teachers per child, where teachers' attention has to be distributed to fewer children at the same time. However, some studies found a null effect of teacher–child ratio and group size on emotional behavioural and instructional domains of teacher–child interaction quality, probably because it is already highly regulated in those countries (Cadima et al., 2022; Hestenes et al., 2015; Pianta et al., 2005; Slot et al., 2015). Further, responding to children's needs with high-quality interactions can be more difficult when there are many children in the classroom whose care and support requires a high degree of attention. Children with a high need for attention can be, for example, very young children. Since the age composition of children in classrooms differ in each country (because of country-specific regulations), only few findings are available regarding mixed-age classrooms. For classrooms with the age range zero to four, a study from the Netherlands identified a negative effect of the number of children under two years in the classroom on the quality of emotional supportive interactions (Schipper et al., 2007). For older children, a study from Portugal by Cadima et al. (2018) revealed that in mixed-age classrooms the quality of classroom organization declined. In Germany, results point in the same direction as they indicate that the presence or number of very young children in the classroom is negatively associated with emotional support (Sommer & Sechtig, 2016). However, findings on those associations are referring to classrooms with children older than three years, and results concerning toddler and different domains of teacher–toddler interactions are lacking.

In general, results on associations between teacher–toddler interaction quality and structural characteristics are still limited in comparison to interactions with older children. Research has nevertheless identified structural conditions that can affect interaction quality. Yet those findings are not always consistent and, regarding toddler classrooms, findings on associations between structural characteristics and teacher–toddler interactions barely exist or lack evidence (especially in Germany). Hence, this study focuses on a German sample of toddler classrooms.

The German ECEC system

In Germany, the ECEC system does not belong to the public educational system but is part of the social welfare system. It is characterized by a strong socio-pedagogical tradition. Usually, early childcare settings are centre-based and organized in several age-cohorts. The age of children within one classroom often ranges from birth to three years or three to six years, sometimes classrooms are age-mixed with children between one and six years. The recommended average number of children per classroom ranges from 6 to 12 for children under three years and from 14 to 18 for children between three and six years (Viernickel & Fuchs-Rechlin, 2015). Although attending ECEC is voluntary, the attendance rate is moderate to high in Germany (35.0% for children under the age of three years and 92.5% for children between three and six years; BMFSFJ, 2021). German preschool teacher education is highly standardized and practice oriented. Besides attending courses at the vocational school, trainees are required to assist staff in childcare settings to complete a two-year training programme. To become a certified teacher, vocational students have to finish an additional three-year course, resulting in a teacher education lasting five years in total. Even though an academic degree is not required to work in the German ECEC system, several applied, educational, and regular universities have been offering bachelor’s and master’s degrees in early childhood education for the past few years. However, in comparison to other countries, the percentage of teachers working in ECEC with a university degree is rather low in Germany (Bock-Famulla et al., 2021).

The current study

Research has shown that different facets of teacher–child interaction can be distinguished. However, studies have pointed to different conclusions about domains of those interactions, where some support a two-domain solution: emotional and behavioural support (EBS) and engaged support for learning (ESL) (Reyhing et al., 2019; van Schaik et al., 2018), while others support a three-domain solution, where emotional and behavioural interactions are two separate facets (Slot et al., 2017).

Moreover, previous findings indicate that teacher characteristics and classroom components can affect the quality of interactions (e.g., Løkken et al., 2018; Slot et al., 2015). Even though research on the quality of teacher–toddler interactions and its conditions is ongoing, results on associations between interaction quality and structural characteristics in toddler classrooms are still limited in comparison to teacher–child interaction with children aged three years and older (e.g., Hestenes et al., 2015; Locasale-

Crouch et al., 2007; Pianta et al., 2005; Suchodoletz et al., 2014). Besides that, findings are not always consistent as there are varying results from different countries, each with another cultural, historical, and economic context, as well as a different ECEC system (see Cadima et al., 2022 for an overview of quality in toddler classrooms in Europe and Vermeer et al., 2016 for an overview of quality in infant, toddler, and preschool classrooms worldwide). Consequently, it remains unclear which structural characteristics influence teacher–toddler interactions and to what extent.

Therefore, the current study addresses two aims. Its first purpose is to depict teacher–toddler interaction quality in Germany, and to do so established assessments on how to measure those interactions are applied. Its second purpose is to examine associations between domains of teacher–toddler interactions and structural characteristics, which comprise teacher characteristics and classroom components.

Method

Sample

The present study was embedded in a larger research project about mathematical development of children aged two to four years and the impact of interaction quality in early childcare settings (EarlyMath: Mathematical Development and the Impact of Interaction Quality in Early Childcare). A total of 95 childcare centres divided into two cohorts ($n_{\text{cohort2020}} = 50$, $n_{\text{cohort2021}} = 45$) participated in the study. Both cohorts did not differ in terms of participant characteristics. The ECEC settings were located in Bavaria in rural (28.4%), suburban (25.3%), and urban areas (46.3%). The first data collection took place between December 2020 and April 2021 (with temporary closure of all participating childcare centres due to governmental lockdowns because of COVID-19) and the second data collection took place between January and March 2022. One classroom per centre and one teacher per classroom agreed to participate in the study and was observed in classroom settings by trained research assistants. Almost all participating teachers were women (95.5%) and were born in Germany (83.1%). Of all the children, 48.3% were female. Descriptive information on teacher and classroom characteristics as well as interaction quality are shown in Table 1 (for more information, see Table 3 in the Appendix). All teachers and parents of the children gave their written consent to take part in the project.

Measures

Interaction quality

The Classroom Assessment Scoring System Toddler Version (CLASS Toddler; La Paro et al., 2012) was used to assess teacher–toddler interaction quality in classrooms. It is suitable for classrooms with toddlers aged 15 to 36 months, was developed in the US-American context, and was used since 2010 with its final version being released in 2012. Prior to data collection, a licensed CLASS trainer trained ten observers that assessed teacher–toddler interactions in this study during a two-day course. To become certified, all observers took part in reliability testing, meaning they coded five videos and passed reliability when they achieved at least 80% agreement within one point of the master coder on each dimension of CLASS Toddler across all five reliability videos. All observers passed the reliability test demonstrating their ability to collect data in this study.

Classrooms were visited on one regular morning and teacher–toddler interactions were observed across various activities, including morning routines, free play, structured activities, and mealtimes. During the visit, the trained research assistants conducted three live-observation cycles of 15-20 min (as recommended by the CLASS Toddler manual) and recorded one 20 min video cycle where the teacher had to use material provided by the research assistants with up to three children. To ensure comparability, the material (a book and a board game) and the instruction were the same for each teacher. The teachers were asked to use both materials the same way they would usually do. No further instruction was given. This video-recorded cycle was rated afterwards. Even though a video observation is not directly recommended by the CLASS Toddler manual, it states that both unstructured and structured situations should be observed and rated. Since the daily routine is not necessarily similar in every childcare setting and unstructured situations (like free play) occur more frequently during the day than structured activities with fewer children (Nores et al., 2022), the video-recorded cycle was treated as a structured situation, which is comparable across all classrooms.

According to the CLASS Toddler manual, classrooms were observed and scored on a 7-point Likert scale ranging from low (1, 2) to middle (3, 4, 5) to high (6, 7). Teachers' interactions were observed regarding eight dimensions: positive climate (PC), negative climate (NC), teacher sensitivity (TS), regard for child perspectives (RCP), behaviour guidance (BG), facilitation of learning and development (FLD), quality of feedback (QF), and language modelling (LM). The manual includes three to four specific behavioural indicators for each dimension and provides examples that serve as guidelines for scoring. The ratings of the live and video cycles of each teacher were aggregated to one interaction quality rating for each of the eight dimensions. According to the CLASS Toddler manual, those dimensions are distributed across two domains: the EBS domain includes five dimensions (PC, NC, TS, RCP, BG), and the ESL domain comprises three dimensions (FLD, QF, LM). Both domains show high internal consistency, with Cronbach's alpha of .90 for EBS and .83 for ESL across the cycles. The correlation between both domains was high (r = .66, p < .001). 28.4% of the video cycles and 6.3% of the live observations were double coded by two independent coders, with an agreement within one point of 84.3% for EBS and an agreement within one point of 81.6% for ESL.

Structural characteristics

We selected structural characteristics with a sufficient variance to avoid multicollinearity (see Table 3 & 4 in the Appendix), and therefore considered teacher education and teacher age as teacher characteristics (assessed via self-report questionnaire for the observed teachers) as well as group size and children's age composition as classroom characteristics (see Table 1 for descriptive statistics). Teachers' education was asked through (1) no completed education, (2) internship as part of the education, (3) two-year education, (4) five-year education, (5) bachelor's degree, or (6) master's degree. To calculate the teachers' ages, they were asked to indicate their year of birth, which was then subtracted from the current year. To state the group sizes, the average number of children who were present during all four cycles was applied. For the age composition in the classrooms, teachers were asked how many children were cared for at the age ranges 0-12 months, 12-24 months, 24-36 months, and older than 36 months. Children's age composition was computed as a dichotomous variable whether or not all children in the classroom were within a two-year span.

Data analyses

To address the first research question, whether two domains of teacher–toddler interaction quality can be distinguished as proposed by the CLASS Toddler manual, we tested the factor structure with confirmatory factor analysis (CFA). A two-factor structure with the EBS and ESL domain was tested against a one-factor structure and a three-factor structure by comparing the model fit improvement with Chi-squared difference tests as well as with the Akaike information criterion (AIC). The AIC was used as it considers the complexity of different models, each with different number of parameters and degrees of freedom. Lower AIC values indicate a higher quality of the model (Vrieze, 2012). To evaluate the fit of the models, we applied the following parameters: comparative fit index (CFI), Tucker-Lewis reliability index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). A value of CFI and TLI > .95 indicates good fit, while a value of CFI and TLI > .90 indicates adequate fit. RMSEA values <.06 represent close fit, whereas values < .08 represent reasonable fit. For SRMR, values < .08 are considered an adequate fit (Bentler & Bonett, 1980; Hu & Bentler, 1999; Steiger, 1990). The best-fitting model with the lowest AIC was used for further analyses.

To answer the second research question, whether there are significant associations between teacher–toddler interaction quality and structural characteristics of the early childcare settings, structural equation modelling (SEM) was used to analyse the relations of observed and latent variables at the same time. This is an advantage over classic regression analysis but requires an adequate sample size. Although when conducting SEM, a sample size below 200 is considered small, it is also possible to calculate it with around 100 cases (preferably with continuous data; Savalei & Rhemtulla, 2013). Additionally, other studies have shown that SEM works with a sample size below 100 (Barros et al., 2016; van Schaik et al., 2018).

Observation data was completed for all classrooms ($N = 95$) and information from the self-report questionnaire was available with partly missing data for 94.7% of the classrooms ($N = 90$). As we used the full information maximum likelihood (FIML) option to deal with missing data, the complete sample of 95 childcare centres was included in the subsequent analyses. Two statistical packages – lavaan (Rosseel, 2012) and semPlot (Epskamp, 2019) – were used to conduct the analyses and visualizations in R. All analyses were run using the maximum likelihood (ML) estimator. To better illustrate the models, we used a completely standardized solution in the figures where the variables have a variance of one.

Results

Domains of teacher–toddler interaction quality

To investigate whether different domains of teacher–child interaction quality can be distinguished in toddler classrooms (research question 1), we first tested a two-factor structure with the factor (1) EBS with the dimensions positive climate, negative climate, teacher sensitivity, regard for child perspectives, and behaviour guidance, and the factor (2) ESL with the dimensions facilitation of learning and development, quality of feedback, and language modelling. The fit of the model was feasible, although TLI and RMSEA showed a rather unsatisfactory fit statistic – which we address in the Discussion section – namely $\chi^2(19) = 56.51$, $p < .001$, CFI = .91, TLI = .87, RMSEA = .14 (CI90: .10–.19), SRMR = .07. All factor loadings were significant and ranged from –.27 (NC) to .99 (TS) for EBS, and from .58 (QF) to .92 (FLD) for ESL. The dimension NC indicates a floor effect as it has low values (range 1-2) and a low standard deviation (*SD* = .28). Hence, we tested the two-factor model without the dimension NC, which had no substantial effect on the fit statistics in our sample, even though other studies suggest excluding NC for analyses due to a lack of variance (Salminen et al., 2021; van Schaik et al., 2018). For theoretical reasons and as the model fit did not increase, we maintained NC in the subsequent analyses. Nor did any theoretically justified modifications improve the model fit. Figure 1 shows the final completely standardized two-factor model with factor loadings in the overall sample.

Next, we tested the two-factor model of teacher–toddler interaction quality against a one-factor and three-factor model. The one-factor model showed a poor model fit ($\chi^2(20) = 94.18$, $p < .001$, CFI = .83, TLI = .76, RMSEA = .20 (CI90: .16–.24), SRMR = .09), and a significant deterioration of the model compared to the two-factor model ($\chi^2_{diff}(1) = 37.67$, $p < .001$) with a higher AIC of 1337.27 (two-factor model AIC = 1301.60). For the three-factor model, aligned with other studies, we distinguished the domains (1) emotional support, with the dimensions positive climate, teacher sensitivity, and regard for child perspectives, (2) behavioural support, with the dimensions negative climate and behaviour guidance, and (3) engaged support for learning, with the dimensions facilitation of learning and development, quality of feedback, and language modelling (Slot et al., 2017). The three-factor model showed similar but slightly poorer fit statistics than the two-factor model ($\chi^2(17) = 55.06$, $p < .001$, CFI = .91, TLI = .85,

RMSEA = .15 (CI90: .11–.20), SRMR = .07) and did not significantly improve the model fit compared to the two-factor model of teacher–toddler interaction quality ($\chi^2_{diff}(2)$ = 1.45, *p* = .48) but had a slightly higher AIC (1304.15, compared to 1301.60 for the two-factor model). For a comparison of the fit statistics of the tested models, see Table 2.

The descriptive statistics of the two domains of teacher–toddler interaction quality showed overall means in the mid-range: *M* = 5.33 (*SD* = .59) for EBS and *M* = 3.23 (*SD* = .70) for ESL. Table 1 shows the descriptive statistics of the CLASS Toddler dimensions and the two domains.

Association with domains of teacher–toddler interaction quality and structural characteristics

To examine the association between EBS, ESL, and structural characteristics (research question 2), we used structural equation modelling (SEM; see Table 4 in the Appendix for bivariate correlations). The model showed an adequate fit: $\chi^2(43)$ = 81.65, *p* < .001, CFI = .91, TLI = .88, RMSEA = .10 (CI90: .06–.13), SRMR = .06. However, similar to the CFA, the fit statistics of TLI and RMSEA were not in a satisfactory range, which we address in the Discussion section. Less EBS was observed when teachers were older (*B* = –.02, *SE* = .01, *β* = –.26, *p* = .01) and when classrooms were mixed age (*B* = –.35, *SE* = .15, *β* = –.25, *p* = .02). Slightly more ESL was provided by teachers with a higher educational level (*B* = .12, *SE* = .07, *β* = .16, *p* = .10). Moreover, mixed-age classrooms were negatively associated with teachers' ESL (*B* = –.33, *SE* = .16, *β* = –.26, *p* = .04). All of these associations have medium effect sizes (Ellis, 2012). Group size was not associated with either domain of teacher–child interaction quality. For better illustration, a completely standardized solution with the results is shown in the model in Figure 2.

Discussion

In recent years, the importance of high-quality interactions in ECEC for children under the age of three has come into focus and CLASS Toddler has become more frequently used as a measure to capture teacher–child interaction quality (Bichay-Awadalla & Bulotsky-Shearer, 2021; Cadima et al., 2022; Salminen et al., 2021). However, research on different facets of these interactions as well as on associations with structural characteristics such as teacher attributes or conditions within classrooms is still scarce concerning German settings, particularly regarding children under three years. This study therefore aimed to (1) investigate how many domains of teacher–toddler interaction quality can be distinguished, and (2) examine associations between teacher–toddler interaction quality and structural characteristics. Our findings indicate that the quality of teacher–toddler interactions can be differentiated into two facets with an EBS and an ESL domain. Moreover, the findings provide evidence that both teacher and classroom characteristics have an important impact on interaction quality, with different associations for each domain.

Teacher–child interaction quality in German toddler classrooms

Our analyses added evidence that two distinct domains of interaction quality can be differentiated as proposed in the CLASS Toddler manual (La Paro et al., 2012). This strengthens the assumption that both EBS and ESL need to be considered separately as teachers with high levels of emotionally and behaviourally supportive interactions may not stimulate learning opportunities at the same high level, and vice versa. This is particularly interesting since previous findings are not always consistent with those of other studies using CLASS Toddler (Slot et al., 2017). In our study, the two-factor structure of teacher–toddler interactions showed the best model fit compared to a one- or three-factor solution, even though fit statistics were not adequate for all parameters as further discussed in the Limitations and Future Directions section. The EBS domain assesses the extent to which interactions are characterized by a positive emotional climate with teachers that are sensitive toward children's needs and behaviourally supportive, meaning that teachers use behaviour guidance approaches and support positive behaviours of children. The ESL domain measures the extent to which teacher–child interactions support children in their learning and development, for example through back-and-forth exchanges and by stimulating language use.

An interesting finding is that country-related differences in the quality of teacher–toddler interactions are noticeable. On average, interaction quality in our German sample is moderate to high in the EBS and low to moderate in the ESL domain. This is similar to findings from other European countries (Guedes et al., 2020; Reyhing et al., 2019; van Schaik et al., 2018), though it does not reflect the quality level of the ESL domain in some US-American studies. There, interaction quality regarding teachers' engaged support for learning was lower (Bichay-Awadalla & Bulotsky-Shearer, 2021 (M = 2.84); La Paro et al., 2014 (M = 2.83)). In the European countries, the average quality level of ESL was at least above $M = 3.21$ (which is still considered a low medium level; Guedes et al., 2020). Those results may point to differences of cultural contexts concerning teacher practices in ECEC between the USA (where CLASS Toddler was developed), and Germany and other European countries (this is also supported by Vermeer et al., 2016, although they found the process quality in the USA to be higher than in Europe).

Associations with structural characteristics

Our findings confirmed significant relations of the EBS domain with teachers' age and children's age composition as well as (marginally) significant relations of the ESL domain with teachers' education and children's age composition. Regarding the association between teacher–child interaction quality and structural characteristics, it is a notable result that teacher age had a negative association with interactions that support emotions and behaviour of children in such a way that older teachers appear to be less effective in providing this support. Prior findings on the impact of teachers' age are inconsistent as Pessanha et al. (2007) revealed a disadvantage of older teachers for interaction quality ($\beta = -.54$), while van IJzendoorn et al. (1998) showed that older teachers were associated with better quality ($\lambda = .82$). A possible explanation could be that working in toddler classrooms can be exhausting and, in some cases, might even lead to constant stress and fatigue. Indeed, there is empirical evidence that excessive work-related demands, such as too little time for tasks or shortage of assistance, are negatively associated with ECEC quality (Aboagye et al., 2020; Chen et al., 2020). Studies indicate negative associations between the quality of teaching practices and stress (Penttinen et al., 2020), emotional exhaustion (Ansari et al., 2022; Fukkink et al., 2019), and depressive symptoms (Decker-Woodrow, 2018; Gerber et al., 2007; Pianta et al., 2005). Younger teachers may cope better with unfavourable work conditions, as they did not yet have to face stressful work requirements for a prolonged period. In studies where negative impacts of older teachers on interaction quality could not be depicted, it could possibly be that older teachers either had already established successful strategies to reduce their level of work-related stress, or that the outward circumstances were more appropriate to counteract work-related long-term challenges. A lack of supervision or further training, which address coping strategies for those challenges, might add to this phenomenon. This could also be the case in Germany, but no research has yet been carried out on this topic and the meaning of teachers' age in toddler classrooms.

The teachers' level of education indicated a marginally positive impact on learning-related teacher–child interaction quality in our analyses. Teachers with a higher education level showed more ESL toward children, meaning that they provided qualitatively high interactions that offered diverse opportunities for children to participate in activities and explore their surroundings. This is in line with other studies, which also indicated positive relations between the education of teachers and good quality practices (Barros et al., 2016; Cadima et al., 2018; Slot et al., 2015; β = .21). Higher levels of education might lead to a deeper understanding of child development and broader strategies for actively engaging with them in learning occasions. Their advanced education might have helped teachers to identify and provide learning opportunities where they function as provide guidance for children to solve problems and stimulate cognitive development. German teacher education is highly standardized and thus, it must be noted that the association between teacher education and ESL was marginally significant in our model, which is most likely because of the limited variance as 67.0% of the teachers reported having five years of vocational training (which is common in Germany).

Another noteworthy result is that children's age composition in the classroom was the only structural characteristic that was associated with both domains of teacher–toddler interaction quality. Mixed-age classrooms, meaning that there is an age range of more than two years within the classroom, were negatively related to EBS and ESL. Teachers might face more difficulties in meeting the emotional, behavioural, and learning supportive needs of all the children in the classroom when a wide range of younger and older children is present. With consequently wider ranges of different developmental stages, it might be challenging for teachers to perceive children's signals and respond with high-quality interactions to the same extent as when the classroom held children of nearly the same age. Cadima et al. (2018) found similar associations between mixed-age classrooms and teacher–child interaction quality in their study with older children. Furthermore, when considering the effect of the number of very young children in the classroom, findings suggest that they might have a negative impact (Schipper et al., 2007; Sommer & Sechtig, 2016). A German study of Wieduwilt et al. (2023) revealed that a higher number of children under three years to care for particularly affects the emotional domain of interactions (β = –.30) and leads to a decline in language stimulation (β = –.34), which has particular implications for German classrooms: teachers should be better prepared for this and conditions have to be adjusted. In terms of child competencies, prior research has shown contradicting results regarding the impact of mixed-age classrooms on child development (Ansari et al., 2016; Bell et al., 2013; Guo et al., 2014; Yang et al., 2022) and research on toddlers is lacking. Thus, the interplay between children's age compositions in (toddler) classrooms, interaction quality, and children's competencies remains unclear and needs to be further studied. Concerning the impact of group size, it is most likely that it did not show significant associations with teacher–toddler interaction quality in our analysis because it is typically regulated in German classrooms and therefore the variety was rather low in the sample. Our findings emphasize that further studies should not only consider group size and teacher–child ratio but also the specific age composition of children in the classroom since this may lead to a better understanding of how influential structural characteristics can be.

Limitations and future directions

This study has some limitations that require discussion. First, in the CFA the fit indices TLI and RMSEA were, respectively, below and above their suggested thresholds. Studies indicate that especially TLI is sensitive to sample size and the number of indicators in the model. A low sample size and an increased number of indicators lead to declined values of TLI, which conventionally indicates (but actually is not true in all cases) that the model must be worse (Shi et al., 2019). Likewise, a small sample size and small degrees of freedom can lead to higher values of RMSEA, conventionally indicating that the model is misspecified, even when this is not necessarily the case (Chen et al., 2008; Kenny et al., 2015; Lai & Green, 2016; McNeish et al., 2018). Savalei and Rhemtulla (2013) recommend that for complex models, 150 observations would be suitable for a convenient interpretation of model fit statistics, and that models with smaller sample sizes can fall below or exceed thresholds easily. In our study, we analysed 95 observed classrooms and re-analyses with larger German samples would therefore be of interest. In addition, McNeish et al. (2018) pointed to the fact that a model with a RMSEA value of .06 (usually thought to be a good fit) can be a poor model when factor loadings are low, whereas a RMSEA value of .20 (usually thought to be a poor fit) can be an acceptable model when factor loadings are high. In our study, factor loadings were all above .70 – even above .90 (TS = .99, FLD = .92) – except for two factor loadings. NC had a loading of .27, and QF a loading of .58, presumably because of limited variances ($Range_{NC}$ = 1.00-2.00, $Range_{QF}$ = 1.00-4.33). In line with Gorsuch (1983) who stated that factor loadings should be greater than .40, almost all factor loadings accomplished this requirement in our model (except for NC). The final SEM showed good to feasible, although not perfect fit statistics for all parameters as well. In terms of cutoff values, MacCallum et al. (1996) considered values of RMSEA in the range of .08–.10 to indicate at least moderate fit. Referring to that, the RMSEA value in our model can be accepted. Overall, in the current literature, researchers have cautioned against overinterpreting fit indices, and some have even questioned the applicability of universal cutoff values such as for TLI and RMSEA to determine adequate model fit (Fan & Sivo, 2007; Kenny et al., 2015; Markland, 2007; Marsh et al., 2004). Thus, interpretations of model fit statistics should be done carefully.

Second, data collection was conducted during the COVID-19 pandemic. In the first phase, childcare centres in Germany were temporarily closed (therefore, data collection took longer) and in both phases, there were partial restrictions imposed such as reductions in the number of children in the classroom, absences of teachers or children because of infection, wearing of face masks, and overall more challenging situations than usual. Such circumstances and general uncertainties during the pandemic might have led to changes in interactions when compared to situations prior to the pandemic (for which we have no comparison regarding this study sample). However, comparing quality levels of teacher–child interaction with studies in prepandemic times, we could not detect substantial changes in quality. What should be stressed is the possibility that teachers who took part in our study might have been particularly motivated and interested in participating, even more than it is usually the case with voluntarily studies, since participating in our project required more effort than without a pandemic situation. This topic needs more consideration in further analyses.

Finally, the observations took place at a single instance, leading to a fleeting insight into classroom practices. It would be interesting to assess teacher–child interactions at multiple time points since it is not clear whether interaction quality might differ depending on the time of the year. Regarding older children, a study by Buell et al. (2017) indicated that this could indeed be the case. Hence, an interesting future direction would be to examine teacher–toddler interaction using a longitudinal design to understand better how different characteristics can contribute at different time points to high-quality interactions in both domains.

Conclusion

Overall, this study contributes to an understanding of mechanisms between teacher–toddler interaction quality and the structural characteristics of early childcare settings, thus helping to identify indicators that should be targeted for improving the quality of German ECEC. The conclusion can be drawn that there is a need to provide teachers with good strategies to cope with conditions within childcare settings to support children in the best way possible. Gaining a deeper understanding about the contribution of different age ranges within one classroom to high-quality interactions is important, as teachers may find it more challenging to support children in mixed-age classrooms. Providing high quality to toddlers is essential since toddlerhood is a crucial stage of learning and development where children gain competencies that will help them succeed in education and later in life.

Funding

This work was supported by the German Research Foundation under the grants DO 2304/1-1, LE 3245/1-1, LI 3487/1-1, RO 820/18-1, and WE 1478/13-1. The Ethics Committee of the University of Bamberg approved the project.

Disclosure statement

The authors report there are no competing interests to declare.

Table 1. Descriptive statistics of the structural characteristics and interaction quality.

	N	%	*M*	*SD*	*Min.*	*Max.*
Teachers' highest education	88					
No training		1.1				
2-year vocational training		8.0				
5-year vocational training		67.0				
Bachelor		12.5				
Master		11.4				
Teachers' age in years	89		35.6	11.1	19.0	63.0
Group size	95		7.6	2.2	2.8	12.8
Children's age composition (ref.=mixed-age classroom)	90	61.1				
Positive climate	95		5.51	.71	3.00	7.00
Negative climate	95		6.81[1] (1.19)	.28	6.00[1] (1.00)	7.00[1] (2.00)
Teacher sensitivity	95		4.98	.87	3.00	7.00
Regard for child perspectives	95		4.88	.82	3.25	6.50
Behaviour guidance	95		4.48	.92	1.75	6.25
Facilitation of learning and development	95		3.61	.86	1.50	5.75
Quality of feedback	95		2.36	.86	1.00	4.33
Language modelling	95		3.72	.81	2.00	5.75
Emotional and behavioural support	95		5.33	.59	3.45	6.70
Engaged support for learning	95		3.23	.70	1.75	5.08

Note. 1-2=low-quality; 3-5=mid-quality; 6-7=high-quality; [1]=reversed.

Table 2. Fit statistics of the one-factor, two-factor and three-factor model.

	One-factor model	Two-factor model	Three-factor model
Factors	1	2	3
AIC	1337.27	1301.60	1304.15
χ^2	94.18	56.51	55.06
df	20	19	17
p-value	<.001	<.001	<.001
CFI	.83	.91	.91
TLI	.76	.87	.85
RMSEA	.20	.14	.15
90% CI	.16–.24	.10–.19	.11–.20
SRMR	.09	.07	.07

Note. AIC=Akaike information criterion; χ^2=Chi-square; *df*=Degrees of freedom; CFI=Comparative fit index; TLI=Tucker-Lewis index; RMSEA=Root mean square error of approximation; CI=Confidence interval; SRMR=Standardized root mean residual.

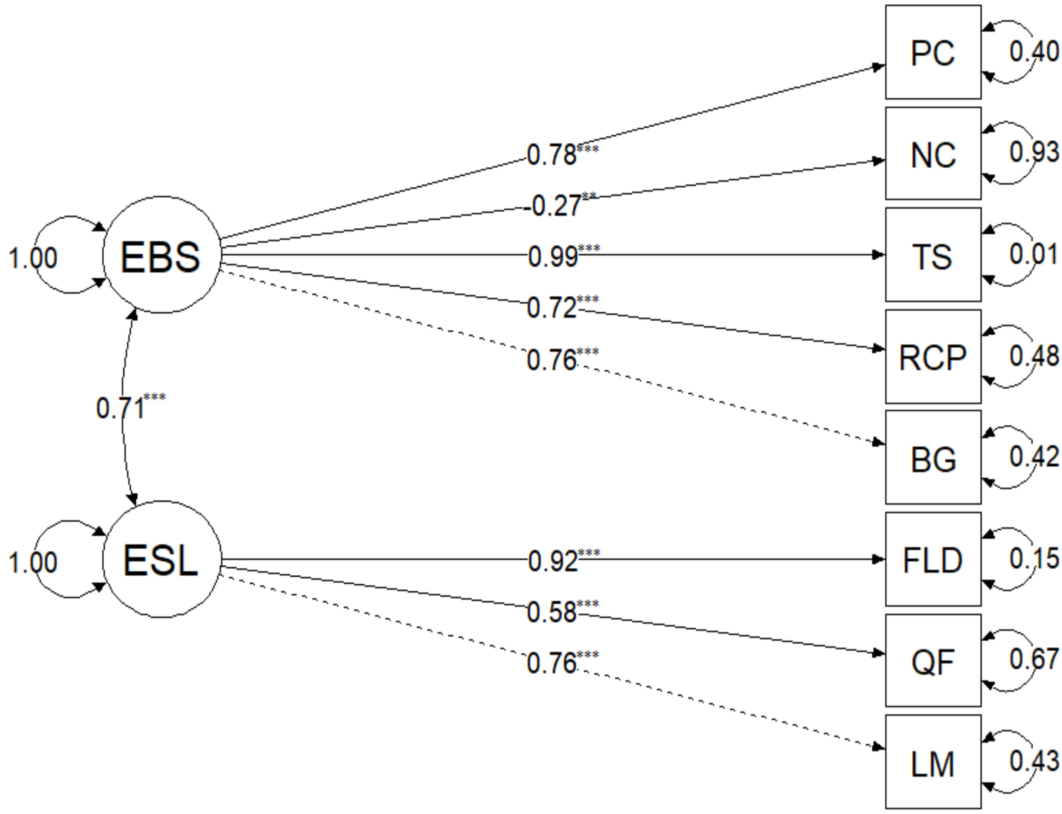


Figure 1. Two-factor model and factor loadings of CLASS Toddler.

Note. EBS=emotional and behavioural support; ESL=engaged support for learning; PC=positive climate; NC=negative climate; TS=teacher sensitivity; RCP=regard for child perspectives; BG=behaviour guidance; FLD=facilitation of learning and development; QF=quality of feedback; LM=language modelling; standardized values are reported; †*p*<.10; **p*<.05; ***p*<.01; ****p*<.001.

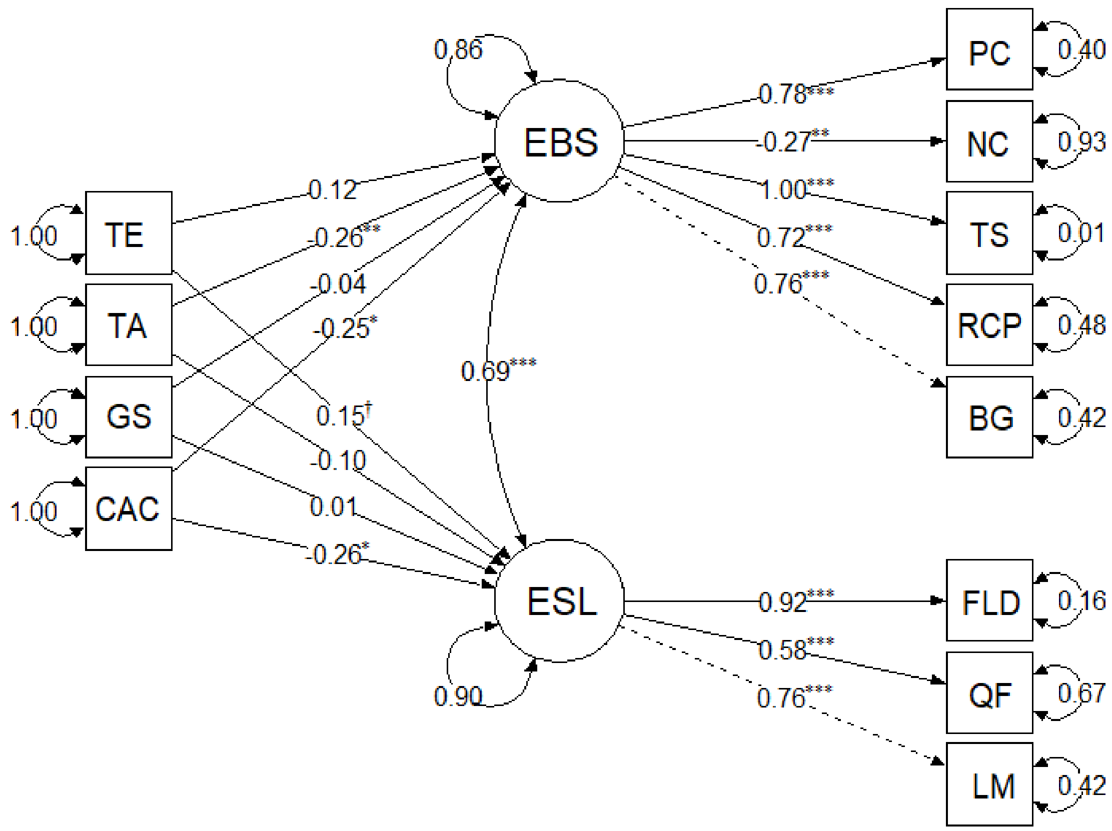


Figure 2. SEM with structural characteristics and factors of interaction quality.

Note. TE=teacher education; TA=teacher age; GS=group size; CAC=children's age composition; EBS=emotional and behavioural support; ESL=engaged support for learning; PC=positive climate; NC=negative climate; TS=teacher sensitivity; RCP=regard for child perspectives; BG=behaviour guidance; FLD=facilitation of learning and development; QF=quality of feedback; LM=language modelling; standardized values are reported; †*p*<.10; **p*<.05; ***p*<.01; ****p*<.001.

References

Aboagye, M. O., Boateng, P., Asare, K., Sekyere, F. O., Antwi, C. O., & Qin, J. (2020). Managing Conflictual Teacher-Child Relationship in Pre-Schools: A Preliminary Test of the Job Resources Buffering-Effect Hypothesis in an Emerging Economy. *Children and Youth Services Review*, *118*, 105468. https://doi.org/10.1016/j.childyouth.2020.105468

Ansari, A., Pianta, R. C., Whittaker, J. V., Vitiello, V. E., & Ruzek, E. A. (2022). Preschool Teachers' Emotional Exhaustion in Relation to Classroom Instruction and Teacher-child Interactions. *Early Education and Development*, *33*(1), 107–120. https://doi.org/10.1080/10409289.2020.1848301

Ansari, A., Purtell, K., & Gershoff, E. (2016). Classroom Age Composition and the School Readiness of 3- and 4-Year-Olds in the Head Start Program. *Psychological Science*, *27*(1), 53–63. https://doi.org/10.1177/0956797615610882

Autor:innengruppe Bildungsberichterstattung. (2022). *Bildung in Deutschland 2022: Ein indikatorengestützter Bericht mit einer Analyse zum Bildungspersonal*. Bielefeld: wbv.

Barros, S., Cadima, J., Bryant, D. M., Coelho, V., Pinto, A. I., Pessanha, M., & Peixoto, C. (2016). Infant Child Care Quality in Portugal: Associations with Structural Characteristics. *Early Childhood Research Quarterly*, *37*, 118–130. https://doi.org/10.1016/j.ecresq.2016.05.003

Bell, E. R., Greenfield, D. B., & Bulotsky-Shearer, R. J. (2013). Classroom Age Composition and Rates of Change in School Readiness for Children Enrolled in Head Start. *Early Childhood Research Quarterly*, *28*(1), 1–10. https://doi.org/10.1016/j.ecresq.2012.06.002

Belsky, J., Vandell, D. L., Burchinal, M. R., Clarke-Stewart, K. A., McCartney, K., & Owen, M. T. (2007). Are there Long-Term Effects of Early Child Care? *Child Development*, *78*(2), 681–701. https://doi.org/10.1111/j.1467-8624.2007.01021.x

Bentler, P. M, & Bonett, D. G. (1980). Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bichay-Awadalla, K., & Bulotsky-Shearer, R. J. (2021). Examining the Factor Structure of the Classroom Assessment Scoring System Toddler (CLASS-T) in Early Head Start and Subsidized Child Care Classrooms. *Early Education and Development*, 1–17. https://doi.org/10.1080/10409289.2021.1887700

BMFSFJ. (2021). *Gute-KiTa-Bericht 2021: Monitoringbericht 2021 des Bundesministeriums für Familie, Senioren, Frauen und Jugend nach § 6 Absatz 2 des Gesetzes zur Weiterentwicklung der Qualität und zur Verbesserung der Teilhabe in Tageseinrichtungen und in der Kindertagespflege (KiTa-Qualitäts- und -Teilhabeverbesserungsgesetz – KiQuTG) für das Berichtsjahr 2020.* Berlin: Bundesministerium für Familie, Senioren, Frauen und Jugend.

Bock-Famulla, K., Girndt, A., Vetter, T., & Kriechel, B. (2021). *Fachkräfte-Radar für KiTa und Grundschule 2021*. Gütersloh: Bertelsmann Stiftung.

Bronfenbrenner, U., & Morris, P. A. (1998). The Ecology of Developmental Processes. In W. Damon, & R. M. Lerner (Eds.), *Handbook of Child Psychology: Theoretical Models of Human Development* (pp. 993–1028). Hoboken, NJ: John Wiley & Sons.

Brophy, J. (1999). Toward a Model of the Value Aspects of Motivation in Education: Developing Appreciation for Particular Learning Domains and Activities. *Educational Psychologist*, *34*(2), 75–85.

Buell, M., Han, M., & Vukelich, C. (2017). Factors Affecting Variance in Classroom Assessment Scoring System Scores: Season, Context, and Classroom Composition. *Early Child Development and Care*, *187*(11), 1635–1648. https://doi.org/10.1080/03004430.2016.1178245

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver Training and Classroom Quality in Child Care Centers. *Applied Developmental Science*, *6*(1), 2–11. https://doi.org/10.1207/S1532480XADS0601_01

Cadima, J., Aguiar, C., & Barata, M. (2018). Process Quality in Portuguese Preschool Classrooms Serving Children At-Risk of Poverty and Social Exclusion and Children with Disabilities. *Early Childhood Research Quarterly*, *45*, 93–105. https://doi.org/10.1016/j.ecresq.2018.06.007

Cadima, J., Aguiar, C., Guedes, C., Wysłowska, O., Salminen, J., Slot, P. L., Barata, M. C., & Lerkkanen, M.-K. (2022). Process Quality in Toddler Classrooms in Four European Countries. *Early Education and Development*, *34*(7), 1565–1589. https://doi.org/10.1080/10409289.2022.2139548

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models. *Sociological Methods & Research*, *36*(4), 462–494. https://doi.org/10.1177/0049124108314720

Chen, S., Phillips, B., & Izci, B. (2020). Teacher–Child Relational Conflict in Head Start – Exploring the Roles of Child Behaviour, Teacher Stress, and Bias, and Classroom Environment. *Early Child Development and Care*, *190*(8), 1174–1186. https://doi.org/10.1080/03004430.2018.1524378

Decker-Woodrow, L. (2018). The Relationship between Internal Teacher Profiles and the Quality of Teacher-Child Interactions in Prekindergarten. *Journal of Research in Childhood Education*, *32*(1), 32–51. https://doi.org/10.1080/02568543.2017.1393030

Ellis, P. (2012). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

Epskamp, S. (2019). *semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output: R package semPlot version 1.1.6*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/semPlot/index.html

Fan, X., & Sivo, S. A. (2007). Sensitivity of Fit Indices to Model Misspecification and Model Types. *Multivariate Behavioral Research*, *42*(3), 509–529. https://doi.org/10.1080/00273170701382864

Fukkink, R., Helmerhorst, K., Gevers Deynoot-Schaub, M., & Sluiter, R. (2019). Training Interaction Skills of Pre-service ECEC Teachers: Moving from In-Service to Pre-Service Professional Development. *Early Childhood Education Journal*, *47*(4), 497–507. https://doi.org/10.1007/s10643-019-00928-6

Gerber, E. B., Whitebook, M., & Weinstein, R. S. (2007). At the Heart of Child Care: Predictors of Teacher Sensitivity in Center-Based Child Care. *Early Childhood Research Quarterly*, *22*(3), 327–346.

Goelman, H., Forer, B., Kershaw, P., Doherty, G., Lero, D., & LaGrange, A. (2006). Towards a Predictive Model of Quality in Canadian Child Care Centers. *Early Childhood Research Quarterly*, *21*(3), 280–295. https://doi.org/10.1016/j.ecresq.2006.07.005

Gorsuch, R. L. (1983). *Factor Analysis*. Hillsdale, NJ: Erlbaum Associates.

Guedes, C., Cadima, J., Aguiar, T., Aguiar, C., & Barata, C. (2020). Activity Settings in Toddler Classrooms and Quality of Group and Individual Interactions. *Journal of Applied Developmental Psychology*, *67*, 101100. https://doi.org/10.1016/j.appdev.2019.101100

Guo, Y., Tompkins, V., Justice, L., & Petscher, Y. (2014). Classroom Age Composition and Vocabulary Development among At-Risk Preschoolers. *Early Education and Development*, *25*(7), 1016–1034. https://doi.org/10.1080/10409289.2014.893759

Hamre, B. K., & Pianta, R. C. (2007). Learning Opportunities in Preschool and Early Elementary Classrooms. In R. C. Pianta, M. J. Cox, & K. Snow (Eds.), *School Readiness and the Transition to Kindergarten in the Era of Accountability* (pp. 49–84). Baltimore, MD: Paul H. Brookes Publishing Company.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., & Rivers, S. E. (2013). Teaching through Interactions: Testing a Developmental Framework of Teacher Effectiveness in over 4,000 Classrooms. *The Elementary School Journal*, *113*(4), 461–487.

Hestenes, L. L., Kintner-Duffy, V., Wang, Y. C., La Paro, K. M., Mims, S. U., Crosby, D., Scott-Little, C., & Cassidy, D. J. (2015). Comparisons among Quality Measures in Child Care Settings: Understanding the Use of Multiple Measures in North Carolina's QRIS and Their Links to Social-Emotional Development in Preschool Children. *Early Childhood Research Quarterly*, *30*, 199–214. https://doi.org/10.1016/j.ecresq.2014.06.003

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The Performance of RMSEA in Models with Small Degrees of Freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kluczniok, K., & Roßbach, H.-G. (2014). Conceptions of Educational Quality for Kindergartens. *Zeitschrift für Erziehungswissenschaft*, *17*, 145–158. https://doi.org/10.1007/s11618-014-0578-2

La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2012). *Classroom Assessment Scoring System (CLASS) Manual, Toddler*. Baltimore, MD: Paul H. Brookes Publishing Company.

La Paro, K. M., Williamson, A. C., & Hatfield, B. (2014). Assessing Quality in Toddler Classrooms Using the CLASS-Toddler and the ITERS-R. *Early Education and Development*, *25*(6), 875–893. https://doi.org/10.1080/10409289.2014.883586

Lai, K., & Green, S. B. (2016). The Problem with Having Two Watches: Assessment of Fit when RMSEA and CFI Disagree. *Multivariate Behavioral Research*, *51*(2–3), 220–239. https://doi.org/10.1080/00273171.2015.1134306

Linberg, A. (2018). *Interaktion zwischen Mutter und Kind: Dimensionen, Bedingungen und Effekte*. Münster: Waxmann.

Locasale-Crouch, J., Konold, T., Pianta, R. C., Howes, C., Burchinal, M. R., Bryant, D. M., Clifford, R. M., Early, D. M., & Barbarin, O. (2007). Observed Classroom Quality Profiles in State-Funded Pre-Kindergarten Programs and Associations with Teacher, Program, and Classroom Characteristics. *Early Childhood Research Quarterly*, *22*(1), 3–17. https://doi.org/10.1016/j.ecresq.2006.05.001

Løkken, I. M., Bjørnestad, E., Broekhuizen, M. L., & Moser, T. (2018). The Relationship between Structural Factors and Interaction Quality in Norwegian ECEC for Toddlers. *International Journal of Child Care and Education Policy*, *12*(9), 1–15. https://doi.org/10.1186/s40723-018-0048-z

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Markland, D. (2007). The Golden Rule is that there are No Golden Rules: A Commentary on Paul Barrett’s Recommendations for Reporting Model Fit in Structural Equation Modelling. *Personality and Individual Differences*, *42*(5), 851–858. https://doi.org/10.1016/j.paid.2006.09.023

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D. M., Burchinal, M. R., Early, D. M., & Howes, C. (2008). Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills. *Child Development*, *79*(3), 732–749. https://doi.org/10.1111/j.1467-8624.2008.01154.x

McNeish, D., An, J., & Hancock, G. R. (2018). The Thorny Relation between Measurement Quality and Fit Index Cutoffs in Latent Variable Models. *Journal of Personality Assessment*, *100*(1), 43–52. https://doi.org/10.1080/00223891.2017.1281286

Morrison, F. J., & Connor, C. M. (2002). Understanding Schooling Effects on Early Literacy: A Working Research Strategy. *Journal of School Psychology*, *40*(6), 493–500.

NICHD Early Child Care Research Network (2002). CHILD-CARE STRUCTURE → PROCESS → OUTCOME: Direct and Indirect Effects of Child-Care Quality on Young Children's Development. *Psychological Science, 13*(3), 199–206. https://doi.org/10.5194/gi-2016-11-RC2

Nores, M., Friedman-Krauss, A., & Figueras-Daniel, A. (2022). Activity Settings, Content, and Pedagogical Strategies in Preschool Classrooms: Do these Influence the Interactions We Observe? *Early Childhood Research Quarterly*, *58*, 264–277. https://doi.org/10.1016/j.ecresq.2021.09.011

Norris, D. J., Horm, D., & McMullen, M. B. (2015). Teacher Interactions with Infants and Toddlers. *YC Young Children*, *70*(5), 84–91.

OECD. (2022). *Education at a Glance 2022: OECD Indicators*. Paris: Organization for Economic Co-operation and Development. https://doi.org/10.1787/3197152b-en

Penttinen, V., Pakarinen, E., Suchodoletz, A. v., & Lerkkanen, M.-K. (2020). Relations between Kindergarten Teachers' Occupational Well-being and the Quality of Teacher-child Interactions. *Early Education and Development*, *31*(7), 994–1010. https://doi.org/10.1080/10409289.2020.1785265

Pessanha, M., Aguiar, C., & Bairrão, J. (2007). Influence of Structural Features on Portuguese Toddler Child Care Quality. *Early Childhood Research Quarterly*, *22*(2), 204–214.

Pianta, R. C., Howes, C., Burchinal, M. R., Bryant, D. M., Clifford, R. M., Early, D. M., & Barbarin, O. (2005). Features of Pre-Kindergarten Programs, Classrooms, and Teachers: Do They Predict Observed Classroom Quality and Child-Teacher Interactions? *Applied Developmental Science*, *9*(3), 144–159. https://doi.org/10.1207/s1532480xads0903_2

Reyhing, Y., Frei, D., Burkhardt Bossi, C., & Perren, S. (2019). Die Bedeutung situativer Charakteristiken und struktureller Rahmenbedingungen für die Qualität der unterstützenden Fachkraft-Kind-Interaktion in Kindertagesstätten. *Zeitschrift Für Pädagogische Psychologie*, *33*(1), 33–47. https://doi.org/10.1024/1010-0652/a000233

Roeser, R. W., Eccles, J. S., & Freedman-Doan, C. (1999). Academic Functioning and Mental Health in Adolescence: Patterns, Progressions, and Routes from Childhood. *Journal of Adolescent Research*, *14*(2), 135–174.

Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rutter, M., & Maughan, B. (2002). School Effectiveness Findings 1979–2002. *Journal of School Psychology*, *40*(6), 451–475.

Salminen, J., Guedes, C., Lerkkanen, M.-K., Pakarinen, E., & Cadima, J. (2021). Teacher–Child Interaction Quality and Children's Self-Regulation in Toddler Classrooms in Finland and Portugal. *Infant and Child Development*, *30*(3), 1–23. https://doi.org/10.1002/icd.2222

Savalei, V., & Rhemtulla, M. (2013). The Performance of Robust Test Statistics with Categorical Data. *The British Journal of Mathematical and Statistical Psychology*, *66*(2), 201–223. https://doi.org/10.1111/j.2044-8317.2012.02049.x

Schipper, E. J. de, Riksen-Walraven, J. M., & Geurts, S. A. (2007). Multiple Determinants of Caregiver Behavior in Child Care Centers. *Early Childhood Research Quarterly*, *22*(3), 312–326. https://doi.org/10.1016/j.ecresq.2007.04.004

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the Model Size Effect on SEM Fit Indices. *Educational and Psychological Measurement*, *79*(2), 310–334. https://doi.org/10.1177/0013164418783530

Slot, P. L., Boom, J., Verhagen, J., & Leseman, P. P. M. (2017). Measurement Properties of the CLASS Toddler in ECEC in the Netherlands. *Journal of Applied Developmental Psychology*, *48*, 79–91. https://doi.org/10.1016/j.appdev.2016.11.008

Slot, P. L., Leseman, P. P. M., Verhagen, J., & Mulder, H. (2015). Associations between Structural Quality Aspects and Process Quality in Dutch Early Childhood Education and Care Settings. *Early Childhood Research Quarterly*, *33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

Sommer, A., & Sechtig, J. (2016). Sozio-emotionale Interaktionsqualität vor dem Hintergrund einer erweiterten Altersmischung im Kindergarten. *Frühe Bildung*, *5*(1), 13–21. https://doi.org/10.1026/2191-9186/a000240

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, *25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Suchodoletz, A. v., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A Typical Morning in Preschool: Observations of Teacher-Child Interactions in German Preschools. *Early Childhood Research Quarterly*, *29*(4), 509–519. https://doi.org/10.1016/j.ecresq.2014.05.010

Thomason, A. C., & La Paro, K. M. (2009). Measuring the Quality of Teacher-Child Interactions in Toddler Child Care. *Early Education and Development*, *20*(2), 285–304. https://doi.org/10.1080/10409280902773351

Tietze, W., Becker-Stoll, F., Bensel, J., Eckhardt, A. G., Haug-Schnabel, G., Kalicki, B., Keller, H., & Leyendecker, B. (Ed.). (2013). *Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit (NUBBEK)*. Weimar/Berlin: das netz.

Van IJzendoorn, M. H., Tavecchio, L. W. C., Stams, G.-J. J., Verhoeven, M. J. E., & Reiling, E. J. (1998). Quality of Center Day Care and Attunement between Parents and Caregivers: Center Day Care in Cross-National Perspective. *The Journal of Genetic Psychology*, *159*(4), 437–454. https://doi.org/10.1080/00221329809596163

Van Schaik, S. D. M., Leseman, P. P. M., & Haan, M. de (2018). Using a Group-Centered Approach to Observe Interactions in Early Childhood Education. *Child Development*, *89*(3), 897–913. https://doi.org/10.1111/cdev.12814

Vandell, D. L., Burchinal, M., R. & Pierce, K. M. (2016). Early Child Care and Adolescent Functioning at the End of High School: Results from the NICHD Study of Early Child Care and Youth Development. *Developmental Psychology*, *52*(10), 1634–1645. https://doi.org/10.1037/dev0000169

Vermeer, H. J., van IJzendoorn, M. H., Cárcamo, R. A., & Harrison, L. J. (2016). Quality of Child Care Using the Environment Rating Scales: A Meta-Analysis of International Studies. *International Journal of Early Childhood*, *48*(1), 33–60. https://doi.org/10.1007/s13158-015-0154-9

Vermeer, H. J., van IJzendoorn, M. H., Kruif, R. E. L. de, Fukkink, R. G., Tavecchio, L. W. C., Riksen-Walraven, J. M., & van Zeijl, J. (2008). Child Care in the Netherlands: Trends in Quality over the Years 1995-2005. *The Journal of Genetic Psychology*, *169*(4), 360–385. https://doi.org/10.3200/GNTP.169.4.360-385

Viernickel, S., & Fuchs-Rechlin, K. (2015). Expertise Fachkraft-Kind-Relation und Gruppengröße in Kindertageseinrichtungen. In S. Viernickel, K. Fuchs-Rechlin, P. Strehmel, C. Preissing, J. Bensel, & G. Haug-Schnabel (Eds.), *Grundlagen, Analysen, Berechnungsmodell* (pp. 11–130). Freiburg im Breisgau: Herder.

Vrieze, S. I. (2012). Model Selection and Psychological Theory: A Discussion of the Differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Wieduwilt, N., Lehrl, S., & Anders, Y. (2023). Preschool Teachers' Language-Related Pedagogical Beliefs and Their Relation to Observed Classroom Quality. *Early Childhood Research Quarterly*, *62*, 175–185. https://doi.org/10.1016/j.ecresq.2022.08.001

Yang, Q., Bartholomew, C. P., Ansari, A., & Purtell, K. M. (2022). Classroom Age Composition and Preschoolers' Language and Literacy Gains: The Role of Classroom Engagement. *Early Childhood Research Quarterly*, *60*, 49–58. https://doi.org/10.1016/j.ecresq.2022.01.001

Appendix

Table 3. Descriptive statistics of the study sample.

	N	%	*M*	*SD*	*Min.*	*Max.*
Teacher–child ratio	95		1 : 3.5	1.2	1 : 0.9	1 : 7.2
Teachers' professional experience in years						
With children under 3 years	83		6.8	4.5	0.3	19.0
In total	87		10.1	7.9	1.5	38.0
Teachers' training (ref.=received further training during the past year)	83	50.6				
Teachers' sex (ref.=female)	89	95.5				
Teachers' country of birth (ref.=Germany)	89	83.1				
Children's age in months	90					
0-12 months		2.1				
12-24 months		31.9				
24-36 months		50.7				
36 months and older		15.3				
Children's sex (ref.=female)	90	48.3				
Multilingual children	89	30.7				

Table 4. Correlations of the domains and dimension of CLASS Toddler and structural characteristics.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Emotional and behavioural support																		
2. Engaged support for learning	.66***																	
3. Positive climate	.81***	.54***																
4. Negative climate	-.34**	-.32**	-.28**															
5. Teacher sensitivity	.94***	.67***	.77***	-.27**														
6. Regard for child perspectives	.83***	.50***	.54***	-.19†	.72***													
7. Behaviour guidance	.85***	.54***	.52***	-.15	.76***	.63***												
8. Facilitation of learn. and dev.	.62***	.90***	.44***	-.24*	.64***	.51***	.53***											
9. Quality of feedback	.49***	.77***	.39***	-.28**	.51***	.23*	.51***	.53***										
10. Language modelling	.54***	.82***	.52***	-.28**	.53***	.50***	.31**	.71***	.36***									
11. Teacher education	.08	.10	.07	.01	.05	.09	.08	.14	.08	.02								
12. Teacher training	.09	.19†	.10	-.23*	.11	-.05	.08	.23*	.14	.08	.30**							
13. Prof. experience (under 3y old)	-.05	-.01	.02	.11	-.05	-.19†	.09	-.04	.14	-.15	.16	.22†						
14. Prof. experience (in total)	-.07	.00	.02	.08	-.10	-.10	-.04	-.06	.06	.01	.03	.11	.69***					
15. Teacher age	-.25*	-.09	-.07	.11	-.25*	-.27*	-.24*	-.06	-.07	-.09	.13	.21	.58***	.74***				
16. Group size	-.04	.02	.06	-.00	-.09	-.05	-.06	-.06	.02	.10	.06	-.19	-.05	.10	.12			
17. Teacher–child ratio	-.04	.07	.08	-.07	-.08	-.04	-.09	.06	.08	.02	.07	.01	-.03	-.04	.05	.60***		
18. Children's age composition	-.21†	-.24*	-.22*	.08	-.26*	-.18†	-.06	-.21*	-.09	-.29**	.05	-.08	.15	-.05	.02	.13	.20†	
19. Multilingual children	.04	-.05	-.02	-.05	.01	.10	.02	.05	-.07	-.11	.21†	.05	-.03	-.02	-.08	.04	.07	-.04

Note. † p<.10; * p<.05; ** p<.01; *** p<.001.

Appendix – Study 3

Improving Global and Math-Specific Teacher–Toddler Interactions through an Intervention for Early Childcare Teachers: The Role of Activity Settings

Baron, F., Linberg, A., Dornheim, D., & Lehrl, S. (under review). Improving Global and Math-Specific Teacher–Toddler Interactions through an Intervention for Early Childcare Teachers: The Role of Activity Settings, *Early Childhood Research Quarterly*.

Preprint available: https://doi.org/10.31219/osf.io/6rzu9

This study is part of the 'EarlyMath' project, in which two teacher trainings, identical in terms of method and scope, but different in terms of focus and content, were developed to enhance global and math-specific teacher–toddler interactions in early childhood education and care (ECEC). We aimed to answer two questions: (1) Do trainings improve global and math-specific interactions to the same extent and (2) does this improvement vary with activity setting (free play vs. structured activity)? Teachers from 95 toddler classrooms were randomly assigned to three groups: math-specific intervention (MIG), global intervention (GIG), and control (CG). The training was structured through nine modules emphasizing practical aspects, which included video analyses and role-plays. The results showed that the teacher training improved math-specific interactions in the MIG during free play, but not during structured activities. Global interaction quality did not significantly improve in both intervention groups compared to the control group. This study highlights the potential of teacher training to enhance interactions in ECEC by emphasizing the specific role of activity settings.

Keywords: early childhood education and care (ECEC); intervention; teacher training; toddler classrooms; teacher–child interactions; math-specific interactions

Introduction

How to raise quality in early childhood education and care (ECEC) is one of the central questions in research for decades (e.g., Cassidy et al., 1995; Egert et al., 2020; Werner et al., 2016). Research and meta-analytic evidence have shown that high-quality interactions between teachers and children are essential for promoting positive outcomes, including increased language skills, enhanced self-development, and better academic achievement (Melhuish et al., 2015; Vandell et al., 2016). Interactions with children under the age of three years have been linked to positive developmental outcomes, not only in the short term, but also later in life (Bratsch-Hines et al., 2020; Choi et al., 2019; Dearing et al., 2009). However, studies across different countries have revealed a mid to rather low quality of teacher–toddler interactions, especially in terms of supporting children's learning opportunities in their early age (La Paro et al., 2014; Slot et al., 2017). Besides the quality of global interactions, domain-specific interactions do not always seem to receive high-quality stimulation in toddler classrooms as well, especially concerning math-specific interactions (Early et al., 2010; Frede et al., 2007). Thus, it is of interest to identify efficient ways of enhancing the quality of global and domain-specific teacher–toddler interactions. Here, research indicates that a promising approach is providing further training for teachers (Buøen et al., 2021; Moreno et al., 2015; Weinstock et al., 2012). However, it remains unclear to which extent these trainings improve different types of interactions, namely global and domain-specific interactions, particularly concerning very young children. Unfolding the potential of teacher training also depends on the activities in which those different teacher–child interactions occur (Guedes et al., 2020; Slot et al., 2015; Werner et al., 2018). Depending on the activity setting (e.g., free play vs. structured activity), teachers may find it easier or more challenging to apply certain facets of their interactions (e.g., global or domain-specific interactions), raising the question of whether trainings have varying effects in different settings. Therefore, this study investigates to which extent teacher–toddler interactions can be improved through teacher trainings that have similar methods and scopes, but differ in focus and content, while taking different activity settings where these interactions take place into account.

Global and math-specific interactions

Teacher–child interactions are often described as being global or domain-specific (Purpura et al., 2017; Sylva et al., 2006). Global interactions are activities and interactions that encompass and foster multiple competencies and skills at the same time, rather than targeting a specific area (Siraj-Blatchford et al., 2002). These global interactions encompass emotional, behavioral, and general learning-related strategies, such as creating a warm and supportive atmosphere, providing back-and-forth exchange, and scaffolding a child through an activity (Berk & Winsler, 1995). Vygotsky's theory of the zone of proximal development underlines the importance of global interactions for children's development (Vygotsky, 1977). However, studies point out that different domains of interactions, such as emotional or learning-related aspects, can be qualitatively divergent and that a teacher might not be assumed to support each domain equally well. A study by Cadima et al. (2022) examined global interaction quality in toddler classrooms across four European countries (Portugal, Poland, Finland, and the Netherlands) and showed that in all countries, the support of emotions and behavior was higher than the learning-related support. This also occurred in US-American and German toddler classrooms: interactions that were assigned to emotions and behavior were mostly in the mid to high range (with limited variance), while more limited opportunities aroused for learning-related interactions (USA: Bichay-Awadalla & Bulotsky-Shearer, 2021; La Paro et al., 2014; Germany: Bücklein et al., 2017). The differences in the quality level of interactions raise concerns, as global interactions are overall, but also solemnly associated with child outcomes (Mortensen & Barnett, 2015; Salminen et al., 2021). Therefore, it is particularly important to foster high-quality interactions in all domains of global interactions, as they are crucial for shaping different areas of child development (Bleses et al., 2020; Choi et al., 2019).

Children in childcare experience not only global, but also domain-specific interactions that promote specific competencies, such as mathematical skills (Clements & Sarama, 2011; Kinzie et al., 2014). Given the emerging research, highlighting the importance of early mathematics at school and later in life (Braak et al., 2022; Duncan et al., 2023; Hooper et al., 2010), it is essential to examine how children can develop their math skills through interactions. Math-specific interactions can address different mathematical aspects, from numerical terms and the number system (e.g., counting: considered a narrow understanding of mathematics) to spatial, temporal, and quantitative content (e.g., tall, long, more: considered a broad understanding of mathematics; Ramani et al., 2015). In ECEC settings, math-specific interactions mostly happen informal in everyday life, helping children to develop early mathematical competencies ('everyday mathematics'; Ginsburg et al., 2008). Even though findings show that math-specific interactions, such as engaging children in mathematical problem-solving and encouraging mathematical thinking and reasoning, are positively associated with children's math competencies (Anders et al., 2013; Lehrl et al., 2016), little time is spent on such interactions. A study of Early et al. (2010) showed that during the day in childcare, children were exposed to mathematical activities only for 8% of classroom time, compared to 17% for language and literacy. Moreover, teachers interacted with children in a narrow range of math-specific content, mostly with a limited focus on counting small numbers (Frede et al., 2007). Studies indicated that this might be the case since teachers are poorly trained for accounting math-specificity and are therefore not able to take advantage of math-specific learning opportunities, feel insecure with mathematical content, or think it is not important for young children (Ginsburg et al., 2008; Pelkowski et al., 2019). Teacher–child interactions that seem promising for development and learning refer to the provision of wide-ranging math experiences (e.g., size, space, shapes, patterns; Ginsburg et al., 2008) and the use of mathematical language (Klibanoff et al., 2006). However, most of this evidence refers to US-American preschool classrooms (exception for Germany: Lehrl et al., 2017; Petersen & Gerken, 2018). Although there is growing interest in math-specific support for toddlers in current research (Burghardt et al., 2020; Reikerås & Salomonsen, 2019), little is known about math-specific interactions and how to endorse them at toddlerhood.

The role of activity settings

Looking at interactions during different activities offers insights into contextual factors, interfering with teacher–child interaction quality as teachers might find it easier or more challenging to show different dimensions of their interactions depending on the activity setting. Interactions take place in various situations in daily ECEC, which can be mainly assigned to two broad activity settings: free play, which is child-initiated and not necessarily led by teachers, and structured situations, which are planned, facilitated, and characterized by a high degree of teacher guidance (Goble & Pianta, 2017). During free play, children explore and engage with their environment in a self-directed manner and with limited directive intervention from teachers, making it an important component for children's development and learning (Fuligni et al., 2012; Goble & Pianta, 2017). Interactions during free play demand a high degree of teachers' flexibility to appropriately stimulate and support children's emotions, their behavior, as well as general learning opportunities and math-specific learning occasions. There are findings, showing that a high level of teacher–child interaction quality and profound math-specific activities in classrooms can be found in classrooms with a high degree of free play (Cabell et al., 2013;

Chien et al., 2010). Structured activities are an essential component as well: typically, they are designed to promote specific objectives or skills and require teachers to be more directive in their interactions with children, for example through explicit instructions (Dalli et al., 2011). In structured situations, topics that are oriented towards children's learning or a specific developmental domain can be prepared in advance with a clear objective as primary focus. This might make it easier for teachers to concentrated on their interactions, especially in terms of domain-specific (including math-specific) interactions. Likewise to findings regarding free play, there are studies that report higher interaction quality during structured activities (e.g., shared book reading; Booren et al., 2012; Vitiello et al., 2012; Wildgruber et al., 2016). However, different studies come to different conclusions whether there are higher levels of interaction quality during free play or during structured activities (Nores et al., 2022).

To the best of our knowledge, to date there are only three studies investigating the meaning of different activity settings explicitly for interactions in toddler classrooms. One study showed that teachers' emotional and behavioral support was significantly higher in free play then in early academic situations (that are structured), while it was the other way around for engaged support for learning: teachers provided slightly more instructional support during early academics than during free play (Guedes et al., 2020). This was partly confirmed by another study that revealed educational activities, which are usually structured, to be associated with both higher emotional and behavioral as well as higher learning support. Comparing teachers engaged support for learning across different activity settings showed that interactions during educational activities were higher than during free play. Here, free play did not contribute to higher emotional and behavioral support (Slot et al., 2015). The third study addressed the meaning of activity settings for interactions and development (not including math skills) and demonstrated that an intervention to promote teacher–toddler interactions and relationships was effective in structured activities, but not in free play (Werner et al., 2018). This leads to a diverse picture on the role of different activity settings for enhancing high-quality global and math-specific interactions, although these findings confirm that teacher–child interactions might intervene with activity settings.

Improving interactions in toddler classrooms

Research has demonstrated the benefits of teacher trainings in promoting high-quality global interactions with preschool-aged children (e.g., ‘TEACH’, Cassidy et al., 1995; ‘TCIT’, Lyon et al., 2009; ‘MyTeachingPartner’, Pianta et al., 2008; ‘Head Start’, Resnick & Zill, 1999). Meta-analyses have exposed that such programs can be efficient: Werner et al. (2016) found moderate effects on global teacher–child interactions, on overall childcare quality, on the caregiver level, and on the child level. Fukkink and Lont (2007) found a positive impact of trainings targeting teacher–child interactions on teachers’ competencies. Even though these findings show that teacher trainings seem to have an impact on global interaction quality, it is challenging to draw consistent conclusions about the effective elements of such trainings, given the variety in focus, design, and methodology. Recent (meta-)analyses have revealed that coaching strategies with an appropriate training intensity, individual feedback, as well as direct application of theoretical input (e.g., through in-house trainings, through role-play, through video-feedback) are promising and effective approaches (Egert et al., 2020; Egert et al., 2018; Kennedy & Lees, 2016). Most of the evidence about trainings that foster global interactions in toddler classrooms comes from the USA, but the picture is diverse: some studies indicated positive effects on emotional and behavioral support (‘SCCC’, Biringen et al., 2012; ‘REACH’, Conners-Burrow et al., 2017), some on engaged support for learning (‘Expanding Quality for Infants and Toddlers’, Moreno et al., 2015), some on both domains (‘Responsive Early Childhood Program’, Landry et al., 2014), and some did not find any effects at all (‘PITC’, Weinstock et al., 2012). Studies conducted in the Netherlands gave insights into European classrooms, but show similar findings: one study revealed that a training with six sessions (each 1.5 hours across several weeks) improved the quality of teacher–toddler interactions in terms of sensitive responsiveness (‘VIPP-CC’, Werner et al., 2018), while another study reported no noteworthy difference before and after a training on global teacher–toddler interactions (‘CIP Training’, Helmerhorst et al., 2017). In Norway, the ‘Thrive by Three’ intervention had positive effects on both domains, emotional and behavioral support and engaged support for learning (Buøen et al., 2021). Hence, results regarding teacher trainings in toddler classrooms are mixed.

In order to improve math-specific interactions, mathematical learning programs and everyday-related trainings to increase math stimulations of teachers have been developed in recent years, mostly for preschool classrooms shortly before children’s school enrollment (e.g., Gasteiger, 2012; Krajewski et al., 2008; Preiß, 2006). Two of the most popular programs are ‘Building Blocks’ (Clements & Sarama, 2007) and ‘Big Math for Little Kids’ (Greenes et al., 2004), both of which have been shown to be effective

(Bojorque et al., 2018; Lewis Presser et al., 2015). In preschools, trainings with a significant focus on specific instructional content seemed to be promising, as teachers with math training exhibited better math-specific interactions than teachers without training (‘MyTeachingPartner-Math/Science’, Kinzie et al., 2014; Whittaker et al., 2020). Effective approaches to train teachers how to apply high-quality math-specific interactions in everyday situations (rather than focusing on specific instructional content) conducted in Germany are ‘Kindergarten of the Future’ (Lehrl et al., 2017) and ‘Pyramid’ (Kammermeyer et al., 2016). What these programs across different countries have in common is that they consist of practical exercises with professional support, feedback, and reflections, with a great emphasis on child centricity. However, they are mostly designed to be applied in preschools.

Activities, recommended to provide good opportunities for promoting not only global, but also specific areas (e.g., mathematics) are shared book reading (Noble et al., 2019; Røe-Indregård et al., 2024; Whitehurst et al., 1994) and playing (board) games (Gasteiger, 2012; Røe-Indregård et al., 2024; Siegler & Ramani, 2009). For toddlers, trainings that target math-specific interactions and activities have not been implemented and evaluated that often: the intervention ‘Play and Learn’ conducted by Bleses et al. (2020) revealed a positive training effect on different competencies of toddlers, also on math skills. This intervention provided teachers in toddler classrooms with strategies and tools to be more explicit and intentional in daily interactions and activities, both globally and in terms of specific content (such as math vocabulary and numeracy skills). However, how different activity settings contributes to this remains unclear, although present findings suggest that both free play and structured activities could be assets for effective teacher–toddler interactions (Nores et al., 2022).

The current study

A growing body of research has examined the importance of high-quality teacher–child interactions and how they can be improved through teacher training (Egert et al., 2018). These interactions can have different facets, including global interactions, such as emotional and behavioral as well as instructional support, and domain-specific interactions, such as stimulating math-related content. Some, but not many studies, have considered different types of activity settings within ECEC in their consideration of different interactions (Guedes et al., 2020; Slot et al., 2015; Werner et al., 2018). These findings mostly refer to children shortly before school entry and research targeting toddlers is comparatively sparse, especially regarding math-specific interactions. Only few studies have explored this area and the meaning of different activity setting in toddler classrooms so far. Therefore, in this study we developed teacher trainings, which aimed to improve global and math-specific teacher–toddler interactions, and we examined the following research questions:

(1) To which extent do teacher trainings improve global and math-specific teacher–toddler interactions?
(2) Does the improvement vary with activity setting (free play vs. structured activity)?

Method

Participants and randomization

This study was embedded in the larger research project ‘EarlyMath’ (Lehrl et al., forthcoming), which examines the mathematical development of children aged two to four years and investigates the impact of interaction quality in early childcare settings using a quasi-experimental design. At baseline, ninety-five childcare centers with one teacher from each center divided into two cohorts ($n_{\text{cohort2020}}$ = 50, $n_{\text{cohort2021}}$ = 45) participated in the study. The centers were in Bavarian rural (28.4%), suburban (25.3%), and urban areas (46.3%). Teachers were assigned randomly to the experimental group 1 (math-specific intervention, n_{posttest} = 30), the experimental group 2 (global intervention, n_{posttest} = 29), or the control group (business as usual, n_{posttest} = 30). The cluster randomization was conducted at the childcare center level, with matching based on the location of the center, its size, and the percentage of multilingual children as well as socially disadvantaged children in each center. No statistically significant group differences occurred at baseline, as shown in Table 1 with descriptive information on teachers and classroom characteristics by condition. However, in the control group, the teachers had slightly more professional experience in ECEC. Global and math-specific interactions were assessed in pre- and posttest. Data collection of the pretest took place between December 2020 and April 2021 and between January and March 2022. The intervention was conducted in March/April 2021 and 2022. Data collection of the posttest took place between April and June 2021 and between May and August 2022. There were

COVID-19 related restrictions during all data collection periods, but we found no evidence in our data that these restrictions biased our measurement in such a way that they could distort intervention effects. Teachers' dropout rate was 6.3%, see Figure 1 for the flowchart of participants and for the measurement points.

Intervention

Two trainings were designed and implemented, aimed at improving the quality of interactions between teachers and toddlers. Teachers in the experimental condition received a two-day training on topics related to child development and learning as well as on interaction quality, and they were either assigned to the math-specific intervention group (MIG), or the global intervention group (GIG). Depending on their allocation, the training consists of math-specific and global interaction quality topics for the MIG, or on global interaction quality topics for the GIG. Even though the focus and content of the trainings differed, the method and scope were the same for each group (see Lehrl et al., forthcoming for a detailed description). Due to COVID-19, all meetings were conducted virtually. As previous research suggests that a crucial component of effective training is practical application (direct skill training; Egert et al., 2017), interactive components were integrated despite the COVID-19 related shift to a virtual setting.

In general, the trainings were structured through nine modules, each emphasizing everyday-related interactions. Every module consisted of a short theoretical input, followed by video analyses of teacher–toddler interactions using observation sheets that were adapted to the quality domains of CLASS Toddler (La Paro et al., 2012). To support the practical transfer, role-plays were conducted in small groups, which has been shown to be particularly effective (e.g., Fröhlich-Gildhoff et al., 2011). Due to the virtual setting, the role-plays were carried out in such way that small groups were moved to so-called breakout rooms, where they each took on a role ('teacher', 'child', or 'observer'). Worksheets were used to describe the situation in which they would find themselves and which methods they should apply. After a role play of approximately five minutes, the 'observer' gave feedback to the 'teacher'. Throughout the role play, each person took on every role. Furthermore, standardized boxes with books, (board) games, and materials were sent to each participating teacher prior to training. The box was then used for demonstration and practical components during the training days. All books and games are particularly, but not exclusively, suitable for stimulating mathematical content and can also address other topics.

Experimental groups with teacher training

The length of the training and the structure of each module was the same for MIG and GIG. A special focus was placed on everyday situations and free play in both groups. After refreshing basics on development, learning, and shared book reading in early childhood in module 1, global interactions to support children in the socio-emotional dimension were discussed in module 2, while interactions to support the cognitive-learning dimension were discussed in module 3. Strategies to improve teachers' interactions was introduced through the method of shared book reading, and this method was illustrated in depth using various books in module 4 and 5 (Ennemoser et al., 2013; Whitehurst et al., 1994). Best-practice examples were jointly analyzed on video with observation sheets. Using the shared reading technique with the help of books from the box, teachers developed 'prompts' – short verbal requests that stimulate children's thinking by using mental verbs such as to believe, to think, to suspect, or to estimate – in group work and individually. These self-developed prompts were then practiced in role-plays with feedback from group members. In module 6, sensitive and stimulating interactions were transferred from books to (board) games. Referring back to previously acquired knowledge, the application of prompts was observed in a video, discussed, and practiced using (board) games from the box in role-plays (Hauser et al., 2014; Siegler & Ramani, 2009). In module 7, the techniques were further transferred to everyday situations, such as mealtimes, changing diapers, or playing outside. During this module, teachers worked on developing prompts for their individual pedagogical practice. Finally, in module 8 there was time to develop explicit plans of what the teachers wanted to implement the coming days and weeks as well as for any remaining questions and comments.

Modules 2, 3, and 8 were the same for both intervention groups. However, content in modules 1, 4, 5, 6, and 7 differed for MIG and GIG. For the MIG, the content and development of prompts was always math related. For four mathematical areas, relevant concepts and specific math-vocabulary were introduced and discussed (Benz et al., 2015; Fthenakis, 2009): (1) size and measurement, (2) quantity, numbers, and digits, (3) space, shape, and time, (4) patterns, sequences, sorting, and classifying. Based on best-practice examples and material from the box, it was demonstrated and practiced how to playfully explore mathematical content related to those four areas. Observation sheets that contained these areas were used for this purpose. The practical component was mainly achieved through using mathematical expressions – the ‘math talk’ (Boonen et al., 2011; Klibanoff et al., 2006; Purpura & Logan, 2015; Ramani et al., 2015) – meaning that teachers learned how to integrate mathematical vocabulary in their language use (in combination with strategies that stimulate mental processes in general). This included quantitative and spatial-temporal vocabulary (e.g., more, over, before), numbers (e.g., one, first), and verbal expressions related to numerical operations (e.g., adding). In the training, teachers in the MIG practiced how everyday situations can be recognized and used as a learning occasion to address mathematical concepts (e.g., counting elephants in the book or steps in the game, discovering patterns while building blocks or stringing beads).

In contrast, the GIG focused on four different content areas with mathematics as one of them to get a global point of view: (1) environment and nature, (2) movement, (3) social relationships, and (4) mathematics. Different from the training of the MIG, math-specific content and vocabulary was not addressed explicitly and prompts for activating mental processes did not necessarily refer to this area. Teachers in the GIG worked with the same videos and discussed and practiced the same emotional and cognitive stimulating techniques to keep the structure of the training as parallel as possible, but with a global point of view by using examples from different topics. Consequently, their observation sheets, which they used during the training days, differed from those in the MIG and did not include the four mathematical areas mentioned above.

Implementation

After the first training day (modules 1–4), teachers had one week of implementation (first phase) in their toddler classroom to apply what has been discussed and learned so that the experiences could be integrated into the second training day (modules 5–8). The second implementation phase lasted eight weeks after the intervention. Teachers were instructed to use materials from the box and the techniques learned at least three times a week. In addition, they received weekly reminders to conduct self-observations and self-reflections using a web-based observation sheet, also used during the training (Fröhlich-Gildhoff et al., 2011; Meyer, 2018). Moreover, four weeks after the training, teachers in

MIG and GIG were invited to participate to module 9, a two-hour debriefing module in which a video was analyzed with the global or math-specific observation sheet, questions were answered, and experiences could be shared.

Control group

For the teachers in the control group, a 90-minute virtual information meeting was conducted that contained basic information on the project itself and a short theoretical introduction to the socio-emotional and the cognitive-learning dimensions. This meeting did not include any interactive elements, such as discussions, the use of observation sheets, or role-plays. Although half of the control group (n = 15) received the box prior to the posttest (without instructions or additional materials), the groups were merged for this study as they both did not obtain any training.

Measures

Global interactions

The Classroom Assessment Scoring System Toddler Version (CLASS Toddler; La Paro et al., 2012) was used to assess interaction quality in classrooms at pre- and posttest. Prior to data collection, ten research assistants were trained by a licensed CLASS trainer and passed the certification by conducting a reliability test (at least 80% agreement within one point of the master coder on each dimension of CLASS Toddler across all five reliability videos). All research assistants were reliable for data collection and were blind to the conditions. To observe teacher–toddler interactions, classrooms were visited on one regular morning during various activities, including free play and structured activities. During the visit, the research assistant conducted three live-observation cycles of 15–20 minutes and recorded one 20-minute video cycle that was rated afterwards. To make the video-observation as comparable as possible, the teacher had to use material provided by the research assistant with up to three children. The material (two books and one board game, which were included in the box, but not necessarily addressed in the training with the intervention groups) and the instruction were the same for each teacher: teachers were asked to use the materials as usually and no further instructions were given.

As described in the CLASS Toddler manual, teacher–toddler interactions were scored on a 7-point Likert scale ranging from low (1, 2), middle (3, 4, 5) to high (6, 7).

Eight dimensions were assessed: positive climate (PC), negative climate (NC), teacher sensitivity (TS), regard for child perspectives (RCP), behavior guidance (BG), facilitation of learning and development (FLD), quality of feedback (QF), and language modelling (LM). The manual includes three to four specific behavioral indicators for each dimension and provides examples that serve as guidelines for the rating. The scores of the live and video cycles were aggregated to one interaction quality score for each of the eight dimensions. According to the CLASS Toddler manual and empirical findings (Baron et al., 2023; Bichay-Awadalla & Bulotsky-Shearer, 2021; Cadima et al., 2022; Salminen et al., 2021), these dimensions are distributed across two domains: the emotional and behavioral support (EBS) domain comprises five dimensions (PC, NC, TS, RCP, BG; $M(SD)_{pre}$ = 5.3 (0.6), $M(SD)_{post}$ = 5.5 (0.5)), and the engaged support for learning (ESL) domain three dimensions (FLD, QF, LM; $M(SD)_{pre}$ = 3.2 (0.7), $M(SD)_{post}$ = 3.4 (0.8), see Table 2 for pre- and posttest ratings divided by groups). Both domains show high internal consistency across the cycles, with Cronbach's alpha of .90 for EBS and .83 for ESL in the pretest and .87 for EBS and .84 for ESL in the posttest.

Math-specific interactions

The frequency of math-specific interactions was assessed during the four-cycle observation in addition to the CLASS Toddler observation. When research assistants observed math-specific activities in each 15–20-minute observation, they had to indicate whether the teacher addressed mathematical content (1) one time, (2) two to three times, or (3) four times and more. Those math-specific interactions were assigned to six mathematical categories: (1) quantity/numbers/digits, (2) size/measurement, (3) space/shape, (4) time, (5) patterns/sequences, and (6) sorting/ classifying (Benz et al., 2015; Fthenakis, 2009). A sum score for each cycle was calculated by adding the occurrence of observed math-specific interactions of each category. If no interaction related to mathematical content was observed, teachers' math-specific interaction received a zero. The sum score in the pretest ranged from 0.5 to 9.8 ($M(SD)_{pre}$ = 3.7 (1.9)), and in the posttest from 0.0 to 9.5 ($M(SD)_{post}$ = 3.7 (1.8)). Quantity/numbers/digits was the category that occurred most often in both pre- and posttest ($M(SD)_{pre}$ = 1.2 (0.7), $M(SD)_{post}$ = 1.3 (0.6), see Table 2 for pre- and posttest ratings divided by groups).

Activity settings and structural characteristics

During the four observation cycles, research assistants further indicated the activity settings of each cycle. As proposed by the CLASS Toddler manual, four different activity setting were rated: (1) free choice/interest area, (2) transition, (3) routine, and (4) group time. While observing, they assessed if a certain activity setting lasted 1-25%, 26-50%, 51-75%, or 76-100% of the time in each 15–20-minute cycle. The subsequent analyses were performed as follows: for free play, observations were considered where free choice persisted more than half of the time, and for structured situations, observations were considered where group time persisted more than half of the time. The cycle with the video observation was assessed as being structured group time. Of all 661 cycles in the pre- and posttest where an activity setting was indicated, free play was observed in 283 cycles (42.8%) and structured activities in 363 cycles (54.9%) over half of the time of the observation cycle (> 7.5 minutes).

Structural characteristics of the toddler classrooms and teachers were assessed during the observations and through questionnaires for the participating teachers, see Table 1 for descriptive statistics.

Data analyses

To answer our research questions, we conducted several repeated measures analyses of variance (RMANOVA) in R with the package ‘rstatix’ (Kassambara, 2023). We utilized partial eta-squared (η_p^2) as an indicator of effect size, a common metric in RMANOVA. Following Cohen (1977), the benchmarks for categorizing the effects are: $\eta_p^2 \sim .01$ = small, $\eta_p^2 \sim .06$ = medium, and $\eta_p^2 \sim .14$ = large. For significant group*time effects, post-hoc tests with Bonferroni adjustment were performed by calculating multiple pairwise comparisons (with the package ‘rstatix’ as well). To visualize intervention effects, we used the package ‘gglot2’ in R (Wickham et al., 2023). We collected completed data for all observations ($N = 89$). However, since not each activity settings were observed for at least half of the time (> 7.5 minutes) in at least one of the four cycles, few cases – depending on the activity setting and the measurement point – could not be included in the respective analysis: for six cases in the pretest and for five cases in the posttest, the required amount of time in free play could not be observed. Regarding structured situations, this applied to zero cases in the pretest and to five cases in the posttest. All cases were missing completely at random (Little’s MCAR test: $p = .63$), and therefore we deleted them listwise to perform complete case analyses.

As teachers' professional experience at baseline was marginally significantly different between the groups in favor of the control group (see Table 1), we performed analyses of covariance (ANCOVA) prior to the main analyses to see if it could have an impact on intervention effects. The ANCOVA with the quality and frequency of global and math-specific teacher–toddler interactions at pretest as independent variable, teachers' professional experience as control variable, and the quality and frequency of global and math-specific interactions at posttest as dependent variable showed – considering the different groups – no other results than the RMANOVA without control variables, which is why we only report the results of the RMANOVA in the following.

Results

Intervention effects on global interactions

Regarding global teacher–toddler interactions during free play, findings revealed no significant time, group, or time*group effects for neither emotional and behavioral support (EBS), nor engaged support for learning (ESL).

During structured activities, (marginally) significant changes over time with small effect sizes can be noticed for both domains: EBS: $F(1, 81) = 4.17$, $p = .04$, $\eta_p^2 = .02$; ESL: $F(1, 81) = 3.24$, $p = .08$, $\eta_p^2 = .01$. In structured situations, all teachers in intervention and control groups showed qualitatively higher global interactions with toddlers at posttest than at pretest. Neither a significant group effect, nor a time*group effect occurred regarding EBS and ESL in structured activities. Table 2 shows the descriptive statistics of the pre- and posttest ratings of teacher–toddler interactions divided by activity settings and groups. Table 3 shows the results of the RMANOVA, including the significant as well as the tested (but not significant) main and interaction effects of global interactions in both activity settings.

Intervention effects on math-specific interactions

Regarding the frequency of math-specific teacher–toddler interactions, there was a significant growth over time in free play with a medium effect size ($F(1, 75) = 13.12$, $p = .001$, $\eta_p^2 = .07$). At posttest, all teachers showed more math-specific interactions during free play than at pretest. The group effect was not significant. However, a marginally significant small time*group effect can be noticed in free play ($F(2, 75) = 2.82$, $p = .07$, $\eta_p^2 = .03$). The post-hoc test with Bonferroni adjustment revealed a significant effect only for the math-specific intervention group (MIG; $p < .001$): teacher in the MIG stimulated children more often in terms of math-specific content than teachers in the other groups (global intervention group and control group), see Figure 2 for the visualized time*group effect of math-specific interactions in free play.

During structured activities, findings showed no significant time, group, or time*group effects for math-specific teacher–toddler interactions. See Table 2 for pre- and posttest descriptive statistics for each group and activity setting and Table 3 for all results of the RMANOVA for effects on math-specific interactions decided by activity settings.

Discussion

The importance of teacher–child interaction quality for child development and learning is highlighted by much research (e.g., Bratsch-Hines et al., 2020; Melhuish et al., 2015). How to improve those interactions is an important target but was mainly examined for preschool classrooms in the past years. Recently, more attention has been put on the quality of interactions in toddler classrooms. However, thus far not many studies have investigated effective ways to improve these interactions. Furthermore, few studies addressing this topic have indicated that teacher–toddler interactions and their improvement might be dependent on the setting in which those interactions take place (Guedes et al., 2020; Slot et al., 2015; Werner et al., 2018). Therefore, this study describes a global and math-specific teacher training and investigates its effectiveness by answering two questions: (1) To which extent do teacher trainings improve global and math-specific teacher–toddler interactions? (2) Does the improvement vary with activity setting (free play vs. structured activity)?

Our findings provide evidence that the frequency of math-specific interactions improved through teacher training, but that this improvement could only be identified during free play and not during structured activities. Our study indicates that teachers in the math-specific intervention group (MIG) were more often able to stimulate mathematics during free play after the training. We could not find any study with a similar result, but identified findings that might explain this phenomenon: Cabell et al. (2013) and Nores et al. (2022) found that children engaged in mathematics more often during free play than during structured activities, making it potentially easier for the teacher to tie on children's math-specific impulse during free play. Moreover, the teacher training was primarily designed to enhance interactions in everyday-related situations, which also includes free play. In the MIG, there was a strong emphasis on linking mathematical prompts with everyday situations in ECEC to rehearse 'everyday mathematics' (Ginsburg et al., 2008). Therefore, these learned strategies could have worked particularly well for the MIG in free play. This is a notable result, especially in German early childcare settings, where free play takes up a large part of the time (Bücklein et al., 2017). However, there are also hints that teachers' engagement in mathematical thinking can be improved through trainings (at least in preschool classrooms), independent of a certain activity setting (Whittaker et al., 2015). We could not confirm this finding in our study within toddler classrooms.

No significant intervention effect could be noticed concerning teachers' emotional and behavioral support (EBS) or engaged support for learning (ESL) in neither of the activity settings. There might be several reasons for this. First, global teacher–toddler interaction quality, especially EBS, was already relatively high at the beginning and therefore did not have much potential for improvement. Looking at the quality level, this potential would have been there for ESL. However, interaction quality did not improve (unless in both domains in structured activities over time). It may be more difficult for teachers to change global, possibly long-established interaction strategies, than to change specific interactions in a targeted way. Also, it might be the case that the teacher training had an effect only on those who initially demonstrated a low level of quality: this assumption is supported by a finding from Bleses et al. (2020). In our study, there was low variance, making this difficult to verify. Second, it might not be possible to identify intervention effects based on the observational assessment we used. CLASS Toddler is a measurement that assesses the quality throughout the entire classroom. Even though the research assistants focused on the teacher who participated in the training, (minor) changes at teacher-level might not have been recognized clear enough. Also, CLASS – which was developed in the USA – has been criticized for not taking cultural differences into account (Ishimine & Tayler, 2014; Pastori & Pagani, 2017). That could cause issues, since the importance of different facets of interactions might be weighted differently, for example, in terms of giving instructions. Nonetheless, we decided for CLASS Toddler as it is the most systematic measure of interaction quality so far.

Limitations and perspectives

Some limitations of this study need to be discussed and they give further perspectives. First, methodological issues need to be considered when interpreting our results. With regard to the control group, some teachers might have had other expectations regarding math-specific content because of the project name ‘EarlyMath’. Indeed, some teachers asked for more information on the math topic and were somewhat confused when we did not comply. We could imagine that these teachers informed themselves more deeply on their own. Additionally, as our project was voluntarily, it could be the case that teachers, who were already interested in math (and in improving themselves in general), took part in this study. Furthermore, half of the control group got a box with books and (board) games, which gave them the opportunity to practice with this material. Perhaps this new material gave them too much stimulus to try new global and math-specific interactions, despite not having a teacher training. The evidence of our study might have been stronger without those limitations. Consequently, it would be of interest to conduct the intervention with fewer references to the topic and incentives for the control group. Moreover, math-specific interactions were assessed as frequency of math-related stimulus towards the children. It would be of high interest to elaborate more on teachers’ math talk quality as this seems to be an important mechanism (Boonen et al., 2011; Klibanoff et al., 2006).

Second, there are limitations about the teacher training itself. Although research suggests in-service training and individual video feedback to be most effective (Egert et al., 2017), it was not possible to offer these aspects to the participating teachers in the teacher training. In-service visits were not possible due to the COVID-19 pandemic and no staff was available for individual video feedback. It would be interesting to evaluate how the teachers evolve their abilities with this additional support. Furthermore, the virtual setting might have had an impact on the implementation of the teacher training, but as we have no comparison and as research did not elaborate on the differences between in person and virtual meetings concerning study fidelity so far, we cannot draw conclusions about this. Although our virtual teacher training was overall well accomplished (74.3% of teachers reported that the methodological implementation in the digital format was very good, 20.0% rated it as good; 74.3% found the training could be followed very well despite its digital format, while 25.7% said they could follow it well

(n = 35)), it would be valuable to conduct the study in a non-pandemic situation (with slight modifications) again.

Finally, there are approaches that appear to be promising, but could not be considered. It would have been interesting to investigate how the teacher training might have changed teachers' beliefs, both in general and concerning mathematics, as there are suggestions that attitudinal changes may precede behavioral changes and therefore, improvement in interaction quality might need a longer period of time (Fukkink & Lont, 2007; Werner et al., 2018). Another approach would be to look exclusively at the improvement of teachers who started with a very low quality at baseline, because they probably gain the greatest increase (Bleses et al., 2020). However, we have a limited variance of interaction quality in our sample.

Implications

Despite the limitations and further perspectives, we can draw some implications from this study. A short teacher training – as we conducted – seems to be helpful for improving the frequency of math-specific interactions in a certain activity setting, namely free play. However, the quality does not increase in all domains and across all activity settings equally well. Our training was primarily focused on everyday situations, which is why the improvement in interactions might be more noticeable during free play – also because the quality and frequency of global and math-specific interactions was already relatively high in structured situations. However, our training did not explicitly address different activity settings. Actually, most teacher trainings jointly consider all activity settings at once with little attention to the differences across settings, which may be a reason why some interventions are not as effective in changing teachers' interactions as expected (Dickinson et al., 2011). Therefore, for teacher training and its effectiveness, it is important to consider the settings in which global and math-specific interactions take place and adapt them accordingly to these different settings.

Conclusion

Even though global and math-specific interactions in ECEC are important for child development and learning (Bratsch-Hines et al., 2020; Choi et al., 2019; Dearing et al., 2009), the potential of teachers' high-quality interactions seem to be unrealized: this is the case especially regarding learning support and math stimulation towards toddlers

(Burghardt et al., 2020; La Paro et al., 2014). Therefore, it is crucial to offer training and support that is needed to improve the quality of such interactions as providing teachers with appropriate trainings can play a critical role in supporting children's outcomes. This study showed that it is possible to increase teachers' interactions in math-specific stimulation through a short teacher training, but not in global interaction quality. This improvement was only detected during a specific activity setting (free play), which highlights that different training approaches are needed to enhance interaction quality in all activity settings. Research needs to continue to examine different dimensions of interactions across several activity settings. Identifying specific strategies to promote interactions during certain settings contributes to the field of early childhood education and care as it assures that best practices are transferred into practice.

Funding

This work was supported by the German Research Foundation under the grants DO 2304/1-1, LE 3245/1-1, LI 3487/1-1, RO 820/18-1, and WE 1478/13-1. The Ethics Committee of the University of Bamberg approved the project.

Disclosure statement

The authors report there are no competing interests to declare.

Table 1. Descriptive statistics of the final study sample's characteristics and ratings of teacher–toddler interactions divided by groups.

		MIG (n = 30)		GIG (n = 29)		CG (n = 30)	
	N	%	*M (SD)*	%	*M (SD)*	%	*M (SD)*
Teachers' highest education	85						
No training		3.4		0.0		0.0	
2-year voc. training		10.3		10.7		3.6	
5-year voc. training		58.7		67.9		71.4	
Bachelor		13.8		10.7		14.3	
Master		13.8		10.7		10.7	
Teachers' age (years)	85		34.8 (11.0)		33.4 (11.6)		36.8 (9.9)
Teachers' professional experience (years in total)	83		8.8 (6.9)		8.1 (7.0)		12.4† (9.2)
Group size	89		7.6 (1.9)		7.5 (1.8)		7.7 (2.7)
Teacher–child ratio	89		1 : 3.4 (1.1)		1 : 3.7 (1.2)		1 : 3.4 (1.2)
Children's age composition (ref.=mixed-age classroom)	86	65.5		65.5		53.6	

Note. MIG=math-specific intervention group; GIG=global intervention group; CG=control group; the group differences are not statistically significant (p<.05); †marginal significant group difference (p=.10).

Table 2. Pretest and posttest ratings of teacher–toddler interactions divided by activity settings and groups.

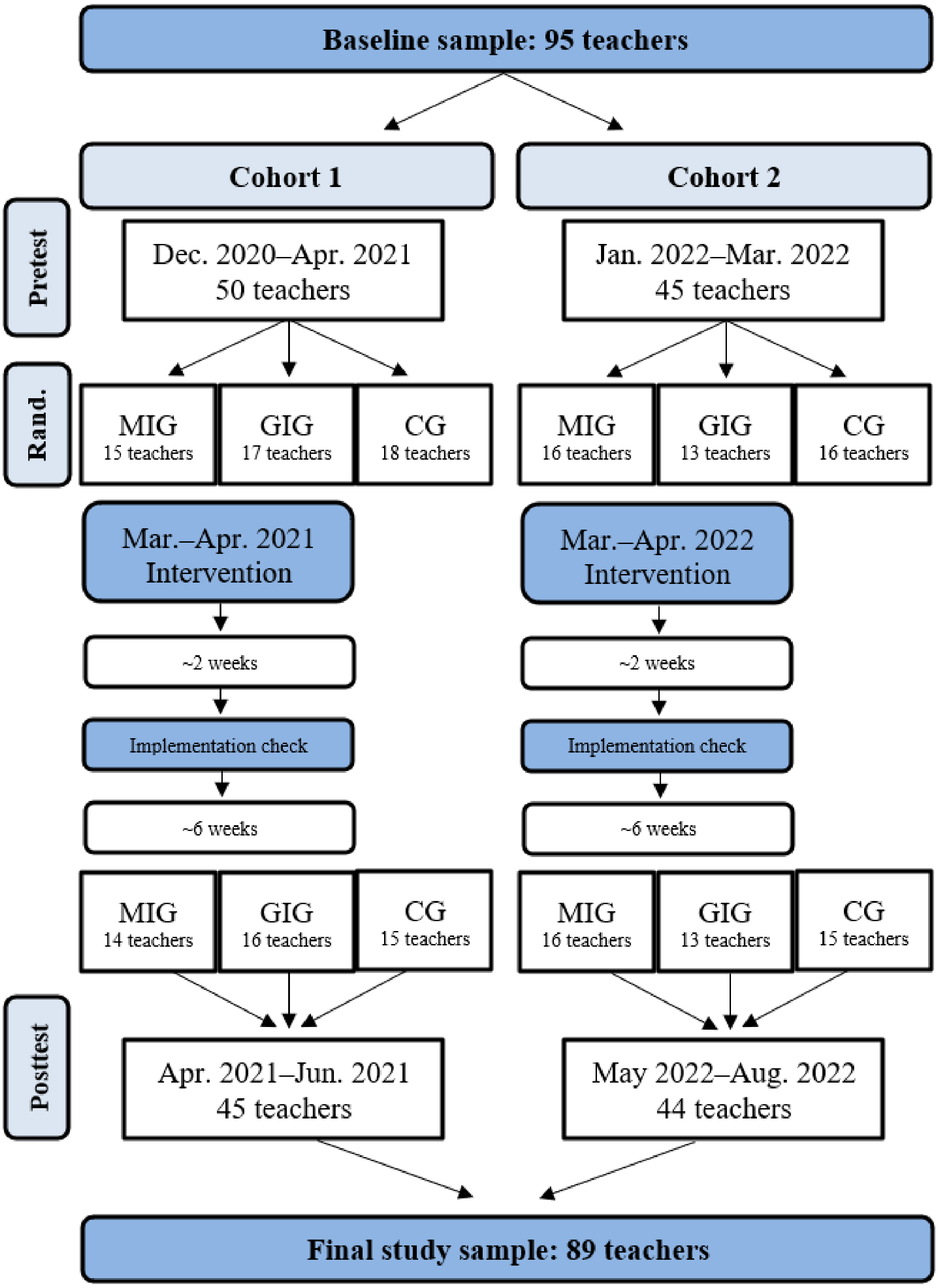
	MIG (*n* = 30)		GIG (*n* = 29)		CG (*n* = 30)	
	$M\ (SD)_{pre}$	$M\ (SD)_{post}$	$M\ (SD)_{pre}$	$M\ (SD)_{post}$	$M\ (SD)_{pre}$	$M\ (SD)_{post}$
Free play						
Emotional and behavioral support	5.2 (0.6)	5.1 (0.7)	5.3 (0.8)	5.4 (0.5)	5.4 (0.7)	5.3 (0.7)
Engaged support for learning	3.0 (0.8)	3.3 (1.0)	2.9 (1.0)	3.3 (1.1)	3.1 (0.7)	3.2 (0.9)
Frequency of math-specific interactions	1.8 (2.3)	4.8 (3.2)	2.7 (2.5)	3.0 (3.2)	2.4 (2.1)	3.2 (2.7)
Structured activities						
Emotional and behavioral support	5.2 (0.8)	5.4 (0.5)	5.5 (0.5)	5.6 (0.7)	5.3 (0.7)	5.5 (0.5)
Engaged support for learning	3.5 (0.8)	3.8 (1.1)	3.7 (0.9)	3.8 (1.1)	3.6 (0.8)	3.8 (0.8)
Frequency of math-specific interactions	5.5 (2.6)	5.3 (2.9)	6.1 (2.9)	5.3 (2.6)	6.3 (3.4)	5.6 (3.2)

Note. MIG=math-specific intervention group; GIG=global intervention group; CG=control group; for emotional and behavioral support and engaged support for learning: 1-2=low-quality; 3-5=mid-quality; 6-7=high-quality; for frequency of math-specific interactions: 0=occurs zero times; 1=occurs one time, 2=occurs two to three times, 3=occurs four times and more; a sum score of math-specific interactions in each cycle was calculated.

Table 3. Results of the repeated measures analyses of variance (RMANOVA).

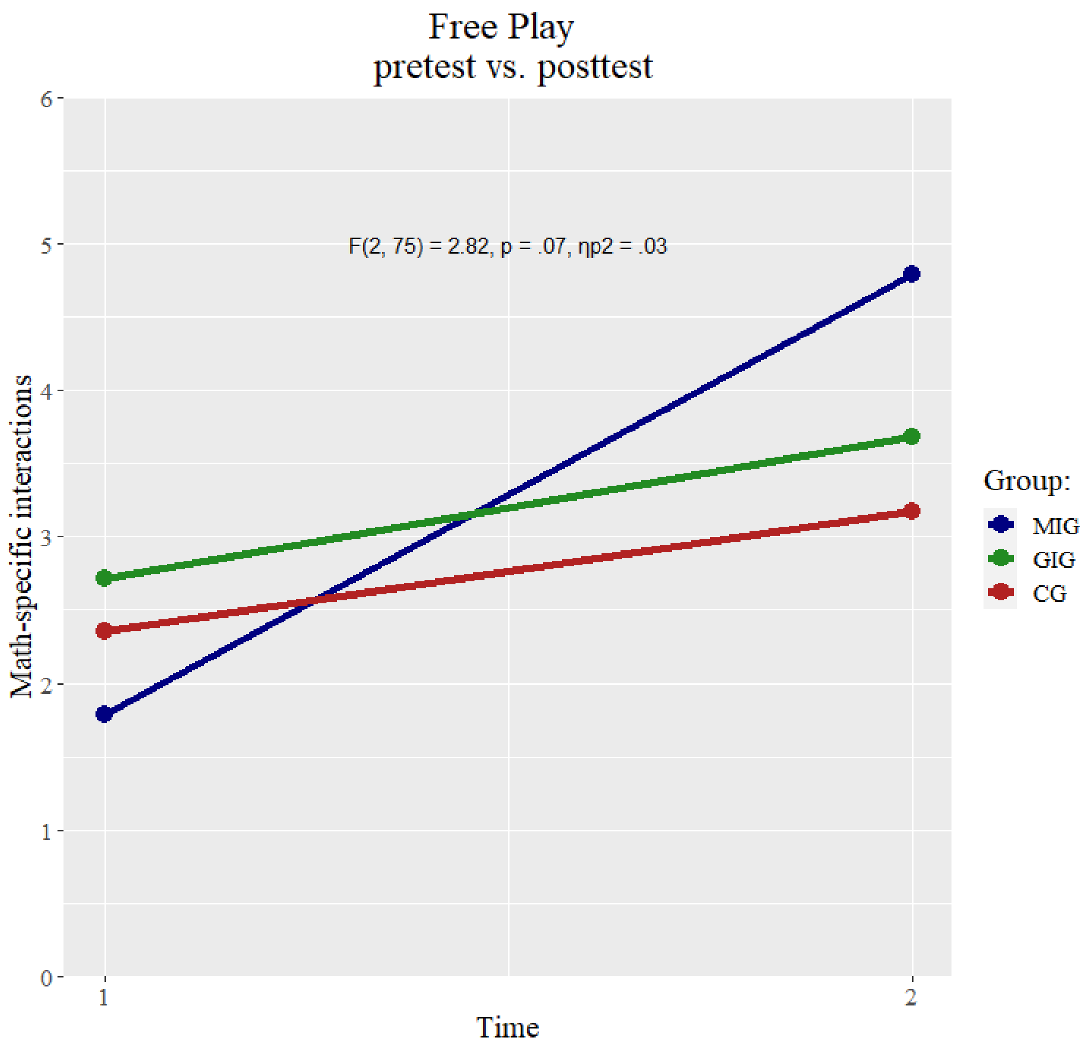
	F	df_n, df_d	p	η_p^2
Free play				
EBS: Time	0.05	1, 75	.83	.00
EBS: Group	0.91	2, 75	.41	.01
EBS: Time*Group	0.75	2, 75	.48	.01
ESL: Time	2.21	1, 75	.14	.01
ESL: Group	0.28	2, 75	.76	.01
ESL: Time*Group	0.42	2, 75	.66	.00
Math: Time	13.12	1, 75	.001***	.07
Math: Group	0.53	2, 75	.59	.01
Math: Time*Group	2.82	2, 75	.07†	.03
Structured activities				
EBS: Time	4.17	1, 81	.04*	.02
EBS: Group	1.18	2, 81	.31	.02
EBS: Time*Group	0.49	2, 81	.61	.01
ESL: Time	3.24	1, 81	.08†	.01
ESL: Group	0.17	2, 81	.84	.00
ESL: Time*Group	0.56	2, 81	.57	.00
Math: Time	1.80	1, 81	.18	.01
Math: Group	0.38	2, 81	.69	.01
Math: Time*Group	0.42	2, 81	.66	.00

Note. EBS=Emotional and behavioral support; ESL=Engaged support for learning; Math=Frequency of math-specific interaction; df_n=Degrees of freedom in the numerator; df_n=Degrees of freedom in the denominator; η_p^2=Partial eta-squared; †p<.10; *p<.05; **p<.01; ***p<.001.



Note. MIG=math-specific intervention group; GIG=global intervention group; CG=control group; Rand.=randomization.

Figure 1. Flowchart of participants and measurement points.



Note. MIG=math-specific intervention group; GIG=global intervention group; CG=control group; for frequency of math-specific interactions: 0=occurs zero times; 1=occurs one time, 2=occurs two to three times, 3=occurs four times and more; a sum score of the math-specific interactions in each cycle was calculated.

Figure 2. Pretest and posttest ratings of the frequency of math-specific interactions during free play and structured activities divided by groups.

References

Anders, Y., Grosse, C., Roßbach, H.-G., Ebert, S., & Weinert, S. (2013). Preschool and Primary School Influences on the Development of Children's Early Numeracy Skills between the Ages of 3 and 7 Years in Germany. *School Effectiveness and School Improvement*, *24*(2), 195–211. https://doi.org/10.1080/09243453.2012.749794

Baron, F., Linberg, A., & Lehrl, S. (2023). Interaction Quality in German Early Childcare Settings: Investigating the Domains of CLASS Toddler and the Associations with Structural Characteristics. *Early Child Development and Care, 193*(13–14), 1485–1502. https://doi.org/10.1080/03004430.2023.2256997

Benz, C., Peter-Koop, A., & Grüßing, M. (Eds.). (2015). *Frühe mathematische Bildung: Mathematiklernen der Drei- bis Achtjährigen*. *Mathematik Primarstufe und Sekundarstufe I + II Didaktik der Mathematik*. Springer Spektrum. https://doi.org/10.1007/978-3-8274-2633-8

Berk, L. E., & Winsler, A. (1995). *Scaffolding Children's Learning: Vygotsky and Early Childhood Education. NAEYC Research into Practice Series. Volume 7*. Washington, DC: National Association for the Education of Young Children.

Bichay-Awadalla, K., & Bulotsky-Shearer, R. J. (2021). Examining the Factor Structure of the Classroom Assessment Scoring System Toddler (CLASS-T) in Early Head Start and Subsidized Child Care Classrooms. *Early Education and Development*, 1–17. https://doi.org/10.1080/10409289.2021.1887700

Biringen, Z., Altenhofen, S., Aberle, J., Baker, M., Brosal, A., Bennett, S., Coker, E., Lee, C., Meyer, B., Moorlag, A., & Swaim, R. (2012). Emotional Availability, Attachment, and Intervention in Center-Based Child Care for Infants and Toddlers. *Development and Psychopathology*, *24*(1), 23–34. https://doi.org/10.1017/S0954579411000630

Bleses, D., Jensen, P., Slot, P. L., & Justice, L. (2020). Low-Cost Teacher-Implemented Intervention Improves Toddlers' Language and Math Skills. *Early Childhood Research Quarterly*, *53,* 64–76. https://doi.org/10.1016/j.ecresq.2020.03.001

Bojorque, G., Torbeyns, J., van Hoof, J., van Nijlen, D., & Verschaffel, L. (2018). Effectiveness of the Building Blocks Program for Enhancing Ecuadorian Kindergartners' Numerical Competencies. *Early Childhood Research Quarterly*, *44*, 231–241. https://doi.org/10.1016/j.ecresq.2017.12.009

Boonen, A. J. H., Kolkman, M. E., & Kroesbergen, E. H. (2011). The Relation between Teachers' Math Talk and the Acquisition of Number Sense within Kindergarten Classrooms. *Journal of School Psychology*, *49*(3), 281–299. https://doi.org/10.1016/j.jsp.2011.03.002

Booren, L. M., Downer, J. T., & Vitiello, V. E. (2012). Observations of Children's Interactions with Teachers, Peers, and Tasks across Preschool Classroom Activity Settings. *Early Education and Development*, *23*(4), 517–538. https://doi.org/10.1080/10409289.2010.548767

Braak, D., Lenes, R., Purpura, D. J., Schmitt, S. A., & Størksen, I. (2022). Why do Early Mathematics Skills Predict Later Mathematics and Reading Achievement? The Role of Executive Function. *Journal of Experimental Child Psychology*, *214*, 105306. https://doi.org/10.1016/j.jecp.2021.105306

Bratsch-Hines, M. E., Carr, R., Zgourou, E., Vernon-Feagans, L., & Willoughby, M. (2020). Infant and Toddler Child-Care Quality and Stability in Relation to Proximal and Distal Academic and Social Outcomes. *Child Development*, *91*(6), 1854–1864. https://doi.org/10.1111/cdev.13389

Bücklein, C., Hoffer, R., & Strohmer, J. (2017). Measuring the Quality of Interactions in Early Childhood Education and Care Settings for Toddlers – An Exploratory Comparison of the Observation Tools GInA and CLASS Toddler. In H. Wadepohl, K. Mackowiak, K. Froehlich-Gildhoff, & D. Weltzien (Eds.), *Psychologie in Bildung und Erziehung. Interaktionsgestaltung in Familie und Kindertagesbetreuung* (pp. 83–114). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-10276-0_4

Buøen, E. S., Lekhal, R., Lydersen, S., Berg-Nielsen, T. S., & Drugli, M. B. (2021). Promoting the Quality of Teacher-Toddler Interactions: A Randomized Controlled Trial of "Thrive by Three" In-Service Professional Development in 187 Norwegian Toddler Classrooms. *Frontiers in Psychology*, *12*, 778777. https://doi.org/10.3389/fpsyg.2021.778777

Burghardt, L., Linberg, A., Lehrl, S., & Konrad-Ristau, K. (2020). The Relevance of the Early Years Home and Institutional Learning Environments for Early Mathematical Competencies. *Journal for Educational Research Online*, *12*(3), 103–125. https://doi.org/10.25656/01:21188

Cabell, S. Q., DeCoster, J., Locasale-Crouch, J., Hamre, B. K., & Pianta, R. C. (2013). Variation in the Effectiveness of Instructional Interactions across Preschool Classroom Settings and Learning Activities. *Early Childhood Research Quarterly*, *28*(4), 820–830. https://doi.org/10.1016/j.ecresq.2013.07.007

Cadima, J., Aguiar, C., Guedes, C., Wysłowska, O., Salminen, J., Slot, P. L., Barata, M. C., & Lerkkanen, M.-K. (2022). Process Quality in Toddler Classrooms in Four European Countries. *Early Education and Development*, *34*(7), 1565–1589. https://doi.org/10.1080/10409289.2022.2139548

Cassidy, D. I., Buell, M. I., Pugh-Hoese, S., & Russell, S. (1995). The Effect of Education on Child Care Teachers' Beliefs and Classroom Quality: Year One Evaluation of the TEACH Early Childhood Associate Degree Scholarship Program. *Early Childhood Research Quarterly*, *10*(2), 171–183. https://doi.org/10.1016/0885-2006(95)90002-0

Chien, N. C., Howes, C., Burchinal, M., Pianta, R. C., Ritchie, S., Bryant, D. M., Clifford, R. M., Early, D. M., & Barbarin, O. A. (2010). Children's Classroom Engagement and School Readiness Gains in Prekindergarten. *Child Development*, *81*(5), 1534–1549. https://doi.org/10.1111/j.1467-8624.2010.01490.x

Choi, J. Y., Horm, D., Jeon, S., & Ryu, D. (2019). Do Stability of Care and Teacher–Child Interaction Quality Predict Child Outcomes in Early Head Start? *Early Education and Development*, *30*(3), 337–356. https://doi.org/10.1080/10409289.2018.1546096

Clements, D. H., & Sarama, J. (2007). Effects of a Preschool Mathematics Curriculum: Summative Research on the Building Blocks Project. *Journal for Research in Mathematics Education*, *38*(2), 136–163. https://doi.org/10.2307/30034954

Clements, D. H., & Sarama, J. (2011). Early Childhood Mathematics Intervention. *Science*, *333*(6045), 968–970. https://doi.org/10.1126/science.1204537

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (Revised Edition). Hillsdale, NJ: Erlbaum Associates.

Conners-Burrow, N. A., Patrick, T., Kyzer, A., & McKelvey, L. (2017). A Preliminary Evaluation of REACH: Training Early Childhood Teachers to Support Children's Social and Emotional Development. *Early Childhood Education Journal*, *45*(2), 187–199. https://doi.org/10.1007/s10643-016-0781-2

Dalli, C., White, E. J., Rockel, J., Duhn, I., Buchanan, E., Davidson, S., & Wang, B. (2011). *Quality Early Childhood Education for Under-Two-Year-Olds: What Should it Look Like? A Literature Review*. New Zealand: Ministry of Education.

Dearing, E., McCartney, K., & Taylor, B. A. (2009). Does Higher Quality Early Child Care Promote Low-Income Children's Math and Reading Achievement in Middle Childhood? *Child Development*, *80*(5), 1329–1349. https://doi.org/10.1111/j.1467-8624.2009.01336.x

Dickinson, D. K., Freiberg, J. B., & Barnes, E. M. (2011). Why are so Few Interventions Really Effective? A Call for Fine-Grained Research Methodology. *Handbook of Early Literacy Research*, *3*, 337–357.

Duncan, R. J., Nordgren, I., Schmitt, S. A., & Vandell, D. L. (2023). Additive and Synergistic Relations of Early Mother–Child and Caregiver–Child Interactions and Socioeconomic Outcomes in Adulthood. *Infant and Child Development*, 1–24. https://doi.org/10.1002/icd.2406

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D.-M. C., Crawford, G. M., Frome, P. M., Clifford, R. M., Burchinal, M., Howes, C., Bryant, D. M., & Pianta, R. C. (2010). How do Pre-Kindergarteners Spend Their Time? Gender, Ethnicity, and Income as Predictors of Experiences in Pre-Kindergarten Classrooms. *Early Childhood Research Quarterly*, *25*(2), 177–193. https://doi.org/10.1016/j.ecresq.2009.10.003

Egert, F., Dederer, V., & Fukkink, R. G. (2020). The Impact of In-Service Professional Development on the Quality of Teacher-Child Interactions in Early Education and Care: A Meta-Analysis. *Educational Research Review*, *29*, 100309. https://doi.org/10.1016/j.edurev.2019.100309

Egert, F., Eckhardt, A. G., & Fukkink, R. G. (2017). Zentrale Wirkmechanismen von Weiterbildungen zur Qualitätssteigerung in Kindertageseinrichtungen. *Frühe Bildung*, *6*(2), 58–66. https://doi.org/10.1026/2191-9186/a000309

Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of In-Service Professional Development Programs for Early Childhood Teachers on Quality Ratings and Child Outcomes: A Meta-Analysis. *Review of Educational Research*, *88*(3), 401–433. https://doi.org/10.3102/0034654317751918

Ennemoser, M., Kuhl, J., & Pepouna, S. (2013). Evaluation des Dialogischen Lesens zur Sprachförderung bei Kindern mit Migrationshintergrund. *Zeitschrift für Pädagogische Psychologie*, *27*(4), 229–239. https://doi.org/10.1024/1010-0652/a000109

Frede, E., Jung, K., Barnett, W. S., Lamy, C. E., & Figueras, A. (2007). *The Abbott Preschool Program Longitudinal Effects Study (APPLES).* New Brunswick, NJ: National Institute for Early Education Research.

Fröhlich-Gildhoff, K., Nentwig-Gesemann, I., & Pietsch, S. (2011). *Kompetenzorientierung in der Qualifizierung frühpädagogischer Fachkräfte. Eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF).* München: Deutsches Jugendinstitut.

Fthenakis, W. E. (Ed.). (2009). *Natur-Wissen schaffen: Band 2. Frühe mathematische Bildung.* Köln: Bildungsverlag EINS.

Fukkink, R. G., & Lont, A. (2007). Does Training Matter? A Meta-Analysis and Review of Caregiver Training Studies. *Early Childhood Research Quarterly*, *22*(3), 294–311. https://doi.org/10.1016/j.ecresq.2007.04.005

Fuligni, A. S., Howes, C., Huang, Y., Hong, S. S., & Lara-Cinisomo, S. (2012). Activity Settings and Daily Routines in Preschool Classrooms: Diverse Experiences in Early Learning Settings for Low-Income Children. *Early Childhood Research Quarterly*, *27*(2), 198–209. https://doi.org/10.1016/j.ecresq.2011.10.001

Gasteiger, H. (2012). Fostering Early Mathematical Competencies in Natural Learning Situations. Foundation and Challenges of a Competence-Oriented Concept of Mathematics Education in Kindergarten. *Journal Für Mathematik-Didaktik*, *33*(2), 181–201. https://doi.org/10.1007/s13138-012-0042-x

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). *Mathematics Education for Young Children: What It Is and How to Promote It. Social Policy Report. Volume 22, Number 1.* Ann Arbor, MI: Society for Research in Child Development.

Goble, P., & Pianta, R. C. (2017). Teacher–Child Interactions in Free Choice and Teacher-Directed Activity Settings: Prediction to School Readiness. *Early Education and Development*, *28*(8), 1035–1051. https://doi.org/10.1080/10409289.2017.1322449

Greenes, C., Ginsburg, H. P., & Balfanz, R. (2004). Big Math for Little Kids. *Early Childhood Research Quarterly*, *19*(1), 159–166. https://doi.org/10.1016/j.ecresq.2004.01.010

Guedes, C., Cadima, J., Aguiar, T., Aguiar, C., & Barata, C. (2020). Activity Settings in Toddler Classrooms and Quality of Group and Individual Interactions. *Journal of Applied Developmental Psychology*, *67*, 101100. https://doi.org/10.1016/j.appdev.2019.101100

Hauser, B., Vogt, F., Stebler, R., & Rechsteiner, K. (2014). Förderung früher mathematischer Kompetenzen. *Frühe Bildung*, *3*(3), 139–145. https://doi.org/10.1026/2191-9186/a000144

Helmerhorst, K. O. W., Riksen-Walraven, J. M. A., Fukkink, R. G., Tavecchio, L. W. C., & Gevers Deynoot-Schaub, M. J. J. M. (2017). Effects of the Caregiver Interaction Profile Training on Caregiver-Child Interactions in Dutch Child Care Centers: A Randomized Controlled Trial. *Child & Youth Care Forum*, *46*(3), 413–436. https://doi.org/10.1007/s10566-016-9383-9

Hooper, S. R., Roberts, J., Sideris, J., Burchinal, M., & Zeisel, S. (2010). Longitudinal Predictors of Reading and Math Trajectories through Middle School for African American versus Caucasian Students across Two Samples. *Developmental Psychology*, *46*(5), 1018–1029. https://doi.org/10.1037/a0018877

Ishimine, K., & Tayler, C. (2014). Assessing Quality in Early Childhood Education and Care. *European Journal of Education*, *49*(2), 272–290. https://doi.org/10.1111/ejed.12043

Kammermeyer, G., Stuck, A., & Roux, S. (2016). Promotion of Literacy and Numeracy in Pyramid Classrooms in Germany. *Early Child Development and Care*, *186*(1), 153–172. https://doi.org/10.1080/03004430.2015.1052424

Kassambara, A. (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests: R package rstatix version 0.7.2*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/rstatix/index.html

Kennedy, A. S., & Lees, A. T. (2016). Preparing Undergraduate Pre-service Teachers Through Direct and Video-Based Performance Feedback and Tiered Supports in Early Head Start. *Early Childhood Education Journal*, *44*(4), 369–379. https://doi.org/10.1007/s10643-015-0725-2

Kinzie, M. B., Whittaker, J. V., Williford, A. P., DeCoster, J., McGuire, P., Lee, Y., & Kilday, C. R. (2014). MyTeachingPartner-Math/Science Pre-Kindergarten Curricula and Teacher Supports: Associations with Children's Mathematics and Science Learning. *Early Childhood Research Quarterly*, *29*(4), 586–599. https://doi.org/10.1016/j.ecresq.2014.06.007

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool Children's Mathematical Knowledge: The Effect of Teacher "Math Talk". *Developmental Psychology*, *42*(1), 59–69. https://doi.org/10.1037/0012-1649.42.1.59

Krajewski, K., Nieding, G., & Schneider, W. (2008). Kurz- und langfristige Effekte mathematischer Frühförderung im Kindergarten durch das Programm „Mengen, zählen, Zahlen". *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *40*(3), 135–146. https://doi.org/10.1026/0049-8637.40.3.135

La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2012). *Classroom Assessment Scoring System (CLASS) Manual, Toddler*. Baltimore, MD: Paul H. Brookes Publishing Company.

La Paro, K. M., Williamson, A. C., & Hatfield, B. (2014). Assessing Quality in Toddler Classrooms Using the CLASS-Toddler and the ITERS-R. *Early Education and Development*, *25*(6), 875–893. https://doi.org/10.1080/10409289.2014.883586

Landry, S. H., Zucker, T. A., Taylor, H. B., Swank, P. R., Williams, J. M., Assel, M., Crawford, A., Huang, W., Clancy-Menchetti, J., Lonigan, C. J., Phillips, B. M., Eisenberg, N., Spinrad, T. L., Villiers, J. de, Villiers, P. de, Barnes, M., Starkey, P., & Klein, A. (2014). Enhancing Early Child Care Quality and Learning for Toddlers at Risk: The Responsive Early Childhood Program. *Developmental Psychology*, *50*(2), 526–541. https://doi.org/10.1037/a0033494

Lehrl, S., Linberg, A., Dornheim, D., Baron, F., Besser, N., Klemm, J., Weinert, S., Roßbach, H.-G. (forthcoming). *Early Mathematical Development and the Importance of Interaction Quality in Daycare Centers. A Methodological Report on the Conception and Implementation of the EarlyMath Project.* Unpublished.

Lehrl, S., Kluczniok, K., & Roßbach, H.-G. (2016). Longer-Term Associations of Preschool Education: The Predictive Role of Preschool Quality for the Development of Mathematical Skills through Elementary School. *Early Childhood Research Quarterly*, *36*, 475–488. https://doi.org/10.1016/j.ecresq.2016.01.013

Lehrl, S., Kluczniok, K., Roßbach, H.-G., & Anders, Y. (2017). Long Term Persistence of Preschool Intervention on Children's Mathematical Development: Results from the German Model Project" Kindergarten of the Future in Bavaria". *Global Education Review*, *4*(3), 70–87.

Lewis Presser, A., Clements, M., Ginsburg, H., & Ertle, B. (2015). Big Math for Little Kids: The Effectiveness of a Preschool and Kindergarten Mathematics Curriculum. *Early Education and Development*, *26*(3), 399–426. https://doi.org/10.1080/10409289.2015.994451

Lyon, A. R., Gershenson, R. A., Farahmand, F. K., Thaxter, P. J., Behling, S., & Budd, K. S. (2009). Effectiveness of Teacher–Child Interaction Training (TCIT) in a Preschool Setting. *Behavior Modification*, *33*(6), 855–884. https://doi.org/10.1177/0145445509344215

Melhuish, E. C., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., Tawell, A., Slot, P. L., Broekhuizen, M., & Leseman, P. P. M. (2015). *A Review of Research on the Effects of Early Childhood Education and Care (ECEC) upon Child Development. CARE Project.* Brussels: European Early Childhood Education and Care Publications.

Meyer, A. (2018). *Kompetenzorientierte Weiterbildungen didaktisch-methodisch planen, durchführen und evaluieren: Ein Leitfaden für frühpädagogische Arbeitsfelder*. München: Deutsches Jugendinstitut.

Moreno, A. J., Green, S., & Koehn, J. (2015). The Effectiveness of Coursework and Onsite Coaching at Improving the Quality of Care in Infant–Toddler Settings. *Early Education and Development*, *26*(1), 66–88. https://doi.org/10.1080/10409289.2014.941260

Mortensen, J. A., & Barnett, M. A. (2015). Teacher–Child Interactions in Infant/Toddler Child Care and Socioemotional Development. *Early Education and Development*, *26*(2), 209–229. https://doi.org/10.1080/10409289.2015.985878

Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C., Gobet, F., & Pine, J. (2019). The Impact of Shared Book Reading on Children's Language Skills: A Meta-Analysis. *Educational Research Review*, *28*, 100290. https://doi.org/10.1016/j.edurev.2019.100290

Nores, M., Friedman-Krauss, A., & Figueras-Daniel, A. (2022). Activity Settings, Content, and Pedagogical Strategies in Preschool Classrooms: Do these Influence the Interactions We Observe? *Early Childhood Research Quarterly*, *58*, 264–277. https://doi.org/10.1016/j.ecresq.2021.09.011

Pastori, G., & Pagani, V. (2017). Is Validation always Valid? Cross-Cultural Complexities of Standard-Based Instruments Migrating out of Their Context. *European Early Childhood Education Research Journal*, *25*(5), 682–697. https://doi.org/10.1080/1350293X.2017.1356545

Pelkowski, T., Herman, E., Trahan, K., Winters, D. M., Tananis, C., Swartz, M. I., Bunt, N., & Rodgick, C. (2019). Fostering a "Head Start" in Math: Observing Teachers in Early Childhood Mathematics Engagement. *Journal of Early Childhood Teacher Education*, *40*(2), 96–119. https://doi.org/10.1080/10901027.2018.1522397

Petersen, L. A., & Gerken, A.-L. (2018). *NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 1 for Four-Year Old Children. Survey Paper No. 45.* Bamberg: Leibniz Institute for Educational Trajectories. https://doi.org/10.5157/NEPS:SP45:1.0

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of Web-Mediated Professional Development Resources on Teacher-Child Interactions in Pre-Kindergarten Classrooms. *Early Childhood Research Quarterly*, *23*(4), 431–451. https://doi.org/10.1016/j.ecresq.2008.02.001

Preiß, G. (2006). *Guten Morgen, liebe Zahlen. Eine Einführung in die „Entdeckungen im Zahlenland"*. Kirchzarten: Preiß.

Purpura, D. J., Day, E., Napoli, A. R., & Hart, S. A. (2017). Identifying Domain-General and Domain-Specific Predictors of Low Mathematics Performance: A Classification and Regression Tree Analysis. *Journal of Numerical Cognition*, *3*(2), 365–399. https://doi.org/10.5964/jnc.v3i2.53

Purpura, D. J., & Logan, J. A. R. (2015). The Nonlinear Relations of the Approximate Number System and Mathematical Language to Early Mathematics Development. *Developmental Psychology*, *51*(12), 1717–1724. https://doi.org/10.1037/dev0000055

Ramani, G. B., Rowe, M. L., Eason, S. H., & Leech, K. A. (2015). Math Talk during Informal Learning Activities in Head Start Families. *Cognitive Development*, *35*, 15–33. https://doi.org/10.1016/j.cogdev.2014.11.002

Reikerås, E., & Salomonsen, T. (2019). Weak Mathematical Skills at an Early Age: Persistent or Temporary? Children with Weak Mathematical Skills and Their Development from Toddlers to Preschoolers. *Early Child Development and Care*, *189*(4), 670–682. https://doi.org/10.1080/03004430.2017.1337753

Resnick, G., & Zill, N. (1999). *Is Head Start Providing High-Quality Educational Services? "Unpacking" Classroom Processes*. Rockville, MD: Westat.

Røe-Indregård, H., Brinchmann, E. I., Rydland, V., Rowe, M. L., Hagtvet, B. E., & Zambrana, I. M. (2024). Teacher–Child Interactions during Toy Play and Book Sharing. *Early Education and Development, 35*(2), 234–249. https://doi.org/10.1080/10409289.2022.2142028

Salminen, J., Guedes, C., Lerkkanen, M.-K., Pakarinen, E., & Cadima, J. (2021). Teacher–Child Interaction Quality and Children's Self-Regulation in Toddler Classrooms in Finland and Portugal. *Infant and Child Development*, *30*(3), 1–23. https://doi.org/10.1002/icd.2222

Siegler, R. S., & Ramani, G. B. (2009). Playing Linear Number Board Games – But Not Circular Ones – Improves Low-Income Preschoolers' Numerical Understanding. *Journal of Educational Psychology*, *101*(3), 545–560. https://doi.org/10.1037/a0014239

Siraj-Blatchford, I., Sylva, K., Muttock, S., Gilden, R., & Bell, D. (2002). *Researching Effective Pedagogy in the Early Years (REPEY). Research Report 356.* London: Department for Education and Skills.

Slot, P. L., Boom, J., Verhagen, J., & Leseman, P. P. M. (2017). Measurement Properties of the CLASS Toddler in ECEC in The Netherlands. *Journal of Applied Developmental Psychology*, *48*, 79–91. https://doi.org/10.1016/j.appdev.2016.11.008

Slot, P. L., Leseman, P. P. M., Verhagen, J., & Mulder, H. (2015). Associations between Structural Quality Aspects and Process Quality in Dutch Early Childhood Education and Care Settings. *Early Childhood Research Quarterly*, *33*, 64–76. https://doi.org/10.1016/j.ecresq.2015.06.001

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing Quality in Early Childhood through Environmental Rating Scales. *Early Childhood Research Quarterly*, *21*(1), 76–92. https://doi.org/10.1016/j.ecresq.2006.01.003

Vandell, D. L., Burchinal, M., & Pierce, K. M. (2016). Early Child Care and Adolescent Functioning at the End of High School: Results from the NICHD Study of Early Child Care and Youth Development. *Developmental Psychology*, *52*(10), 1634–1645. https://doi.org/10.1037/dev0000169

Vitiello, V. E., Booren, L. M., Downer, J. T., & Williford, A. (2012). Variation in Children's Classroom Engagement throughout a Day in Preschool: Relations to Classroom and Child Factors. *Early Childhood Research Quarterly*, *27*(2), 210–220. https://doi.org/10.1016/j.ecresq.2011.08.005

Vygotsky, L. S. (1977). The Development of Higher Psychological Functions. *Soviet Psychology*, *15*(3), 60–73.

Weinstock, P., Bos, J., Tseng, F., Rosenthal, E., Ortiz, L., Dowsett, C., Huston, A., & Bentley, A. (2012). *Evaluation of Program for Infant/Toddler Care (PITC): An On-Site Training of Caregivers. Final Report*. Washington, DC: National Center for Education Evaluation and Regional Assistance.

Werner, C. D., Linting, M., Vermeer, H. J., & van IJzendoorn, M. H. (2016). Do Intervention Programs in Child Care Promote the Quality of Caregiver-Child Interactions? A Meta-Analysis of Randomized Controlled Trials. *Prevention Science: The Official Journal of the Society for Prevention Research*, *17*(2), 259–273. https://doi.org/10.1007/s11121-015-0602-7

Werner, C. D., Vermeer, H. J., Linting, M., & van IJzendoorn, M. H. (2018). Video-Feedback Intervention in Center-Based Child Care: A Randomized Controlled Trial. *Early Childhood Research Quarterly*, *42*, 93–104. https://doi.org/10.1016/j.ecresq.2017.07.005

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A Picture Book Reading Intervention in Day Care and Home for Children from Low-Income Families. *Developmental Psychology*, *30*(5), 679–689. https://doi.org/10.1037/0012-1649.30.5.679

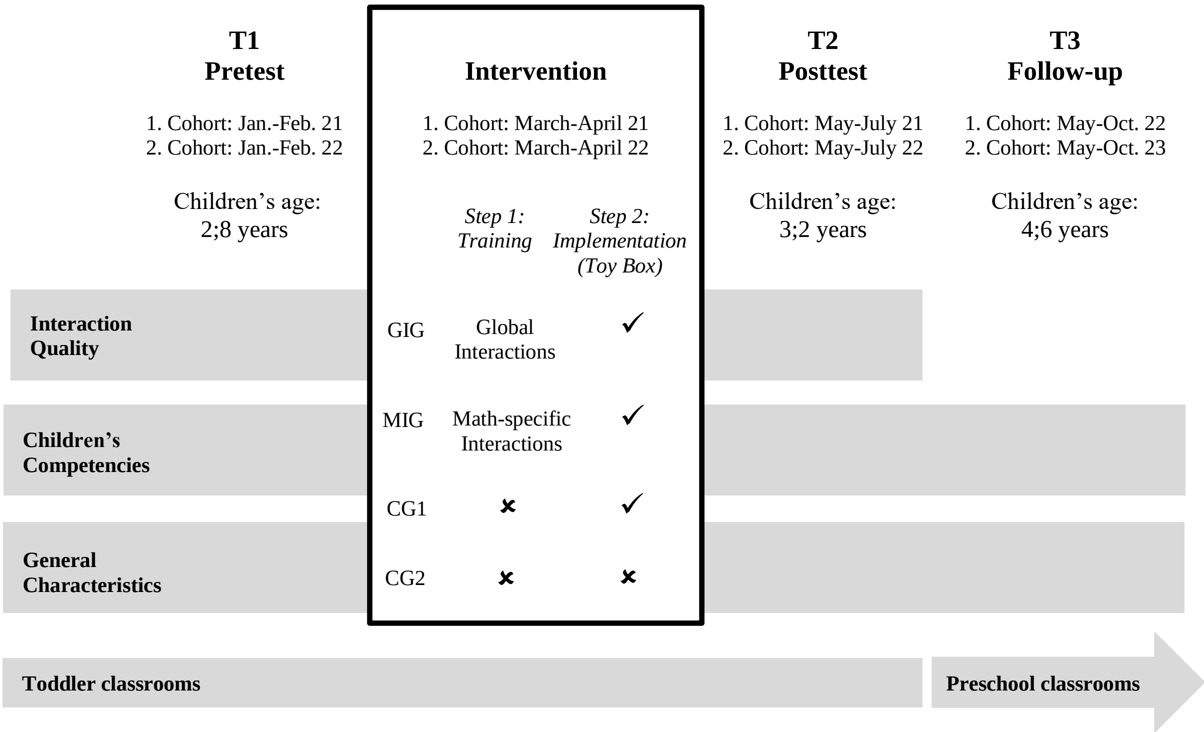
Whittaker, J. V., Kinzie, M. B., Vitiello, V., DeCoster, J., Mulcahy, C., & Barton, E. A. (2020). Impacts of an Early Childhood Mathematics and Science Intervention on Teaching Practices and Child Outcomes. *Journal of Research on Educational Effectiveness*, *13*(2), 177–212. https://doi.org/10.1080/19345747.2019.1710884

Whittaker, J. V., Kinzie, M. B., Williford, A., & DeCoster, J. (2015). Effects of MyTeachingPartner–Math/Science on Teacher–Child Interactions in Prekindergarten Classrooms. *Early Education and Development*, *27*(1), 110–127. https://doi.org/10.1080/10409289.2015.1047711

Wickham, H., Chang, W., Henry, L., Lin Pedersen, T., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2023). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics: R package ggplot2 version 3.4.4*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/ggplot2/index.html

Wildgruber, A., Wertfein, M., Wirts, C., Kammermeier, M., & Danay, E. (2016). Situative Unterschiede der Interaktionsqualität im Verlauf des Kindergartenalltags. *Frühe Bildung*, *5*(4), 206–213. https://doi.org/10.1026/2191-9186/a000283

Further Appendices



Note. T=time of assessment; MIG=math-specific intervention group; GIG=global intervention group; CG=control group.

Figure A1. EarlyMath study design.

Table A1. Content overview of the teacher trainings.

Content of Teacher Trainings	Same Content	Different Content
First Training Day (6 hours, excl. breaks)		
Module 1 Learning and development in early childhood • GIG: Shaping interactions in different content areas: Nature; Mathematics; Movement; Social relationships • MIG: Shaping interactions in content areas of mathematics: Size and measurement; Space, shape, and time; Quantity, numbers, and counting; Patterns and sequences		**X**
Module 2 Theoretical introduction to shared book reading I: Support well-being and positive emotions • Positive climate • Sensitivity • Focus on the child • Behavioral support	**X**	
Module 3 Theoretical introduction to shared book reading II: Support thinking and language • Stimulating thinking and learning processes • Giving good feedback • Stimulating and expanding language	**X**	
Module 4 Shaping (math-specific) interactions during shared book reading I • Provide effective interactions based on a picture book (‘Gemeinsam sind wir riesengroß’) • Focus of GIG and MIG depending on the content areas introduced in module 1		**X**
1 week of implementation in the toddler classrooms		
Second Training Day (6 hours, excl. breaks)		
Module 5 Shaping (math-specific) interactions during shared book reading II • Brief recap of the first training day • Provide effective interactions based on a picture book (‘Mittagsschmaus im Hasenhaus’) • Focus of GIG and MIG depending on the content areas introduced in module 1, and support of thinking and language		**X**
Module 6 Shaping (math-specific) interactions during (board) games • Provide effective interactions based on a board games (‚Bärenhunger‘; ‚Auf, auf, kleiner Pinguin‘; ‚Würfelwürmchen‘; ‚Einkaufen‘) • Focus of GIG and MIG depending on the content areas introduced in module 1		**X**
Module 7 Shaping (math-specific) interactions during everyday situations • Provide effective interactions in everyday situations in childcare (examples: mealtime, excursion, cleaning up, changing diapers) • Focus of GIG and MIG depending on the content areas introduced in module 1, and support of thinking and language, and support of well-being and positive emotions		**X**
Module 8 Reflection • Discussion of further learning opportunities (newsletters, pocket cars), general feedback	**X**	
4 weeks of implementation in the toddler classrooms		
Debriefing (2 hours, part of implementation)		
Module 9 Exchange of experiences • Repetition of content and methods		**X**
4 weeks of implementation in the toddler classrooms *incl. three content area related newsletters, weekly reminders for self-observation and self-reflection*		

Note. MIG=math-specific intervention group; GIG=global intervention group; The methods were the same for GIG and MIG and followed the same procedure for modules 2 to 7: (1) theoretical framing and topic-related examples, (2) video observation with topic-related observation sheets, (3) active practice of the topic-related examples through role-plays, group work, and individual work; Based on Lehrl et al., 2024.

Acknowledgments

I want to express my appreciation for everyone who helped me with this thesis and who shaped me as a researcher. Especially, I would like to acknowledge my PhD supervisor **Prof. Dr. Susanne Kuger**, who supported me throughout this time. I am grateful for the supportive mentoring, the constructive guidance, and the knowledge I have gained from you. You are certainly a role model in academia, and I look forward to staying in touch.

Further, I want to acknowledge **Prof. Dr. Simone Lehrl** for being an excellent mentor for my thesis and in the EarlyMath project. Thank you very much for all your valuable thoughts. Also, special thanks to **Dr. Anja Linberg**, who was an amazing colleague at the German Youth Institute. Thank you for all the time you invested in my progress and in my scientific success. Your open ear definitely had a huge impact on where I am right now. Many thanks to all other colleagues at the German Youth Institute as well.

Moreover, thanks to **Dr. Dorothea Dornheim** for lots of helpful discussions. Particularly, I want to thank **Nadine Besser** and **Johanna Klemm**, who were the best colleagues I could have imagined. Thank you for this great, productive, collaborative, and supportive working atmosphere – and thanks to everyone else I was lucky enough to work with in the EarlyMath project. Also, thanks to my new colleagues at the ICER in Bern, for giving me the time and space I needed to accomplish my goal and to complete this PhD.

I am very grateful to my **friends (Gloria, Vanessa, Annika, Sina, Anna, Larissa, Julia, Elisa, Shannen, Anja, Anja…)** and to my **family (Lukas, Jorge, Gerda, Viktor, Alexandra, Sonja, Andreas…)** who have supported me in many ways. It is always a great pleasure to be around you. Thanks for contributing to my work-life balance.

Many thanks to my parents **Ulrike & Johannes** for the constant support and encouragement. Your effort guided me many years.

Specially, I want to thank my wonderful fiancé **Laurentz**, who has accompanied me closely the whole time and motivated me to stay on track. Thank you for making everything so much more enjoyable! $(x^2+y^2-1)^3 - x^2y^3 = 0$

Finally, I want to thank **everyone** who inspired me, or who has managed to distract me with something completely different at times. Thank you all for paving this way.

List of Publications

Published publications (double-blind peer-reviewed)

Baron, Franka*, Linberg, Anja, & Kuger, Susanne (2023). Assessing Global and Math-Specific Teacher–Child Interaction Quality in Early Childcare Settings: A Systematic Literature Review of Instruments Used in Research, *European Early Childhood Education Research Journal, 31*(4), 640–659. https://doi.org/10.1080/1350293X.2022.2154817**

Baron, Franka*, Linberg, Anja, & Lehrl, Simone (2023). Interaction Quality in German Early Childcare Settings: Investigating the Domains of CLASS Toddler and the Associations with Structural Characteristics, *Early Child Development and Care, 193*(13–14), 1485–1502. https://doi.org/10.1080/03004430.2023.2256997**

Vinokic, Konstantin*, **Baron, Franka**, Kunter, Mareike, Linberg, Anja, Begrich, Lukas, & Kuger, Susanne (2024). Using the Thin Slices Technique to Assess Interactional Quality in Early Childhood Education and Care Settings, *Frontiers in Education, 9*, 1368503. https://doi.org/10.3389/feduc.2024.1368503

Articles submitted for publication (double-blind peer-reviewed)

Baron, Franka*, Linberg, Anja, Dornheim, Dorothea, & Lehrl, Simone (under review). Improving Global and Math-Specific Teacher–Toddler Interactions through an Intervention for Early Childcare Teachers: The Role of Activity Settings, *Early Childhood Research Quarterly*.**

Other publications (not double-blind peer-reviewed)

Baron, Franka, Kuger, Susanne, Riedel, Birgit, & Schacht, Diana D. (2023). Es braucht viele Stimmen, um die frühe Bildung weiterzuentwickeln. Was gute frühkindliche Bildung ausmacht, wird unterschiedlich definiert, erhoben und von vielen Akteur:innen beeinflusst. In Deutsches Jugendinstitut (Hrsg.), *Frühe Bildung weiterentwickeln. Wie es um die Qualität der Kindertagesbetreuung in Deutschland steht und welche positiven Beispiele es aus anderen Ländern gibt* (S. 4–10), München: DJI Impulse.

*first authorship
**included in the present dissertation „Dimensions, Conditions, and Improvement of Teacher–Toddler Interaction Quality in Early Childhood Education and Care: Teachers' Role in the Complex Interplay of Process–Person–Context–Time“