



# Messung kindlicher Interaktionsqualität in Kindertageseinrichtungen

Eine kritisch-konstruktive Einordnung des Beobachtungsinstruments *Individualized Classroom Assessment Scoring System*

Thilo Schmidt<sup>1</sup>, Katharina Kluczniok<sup>2</sup>, Magdalena Molina Ramirez<sup>3</sup> und Norman Zimmermann<sup>1</sup>

<sup>1</sup>Arbeitsbereich Pädagogik der frühen Kindheit, Institut für Bildung im Kindes- und Jugendalter, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU), Deutschland

<sup>2</sup>Arbeitsbereich Frühkindliche Bildung und Erziehung, Fachbereich Erziehungswissenschaft und Psychologie, Freie Universität Berlin, Deutschland

<sup>3</sup>Abteilung Zentrum für Dauerbeobachtung und Methoden, Deutsches Jugendinstitut München, Deutschland

**Zusammenfassung:** Mit dem *Individualized Classroom Assessment Scoring System* (inCLASS) liegt ein Beobachtungsinstrument vor, das zum Ziel hat, Interaktionen in Kitas auf Kindebene erfassen und in ihrer Qualität bestimmen zu können. Der Beitrag geht den Fragen nach, inwieweit das inCLASS diesem Anspruch nach Maßgabe gängiger Gütekriterien gerecht wird und welche Herausforderungen und Limitationen damit verbunden sind. Die Ausführungen stützen sich auf eigene Forschungserfahrungen mit dem Instrument. Über das inCLASS hinaus werden Bezüge zu weiteren Qualitätsmessverfahren (CLASS und KES) hergestellt. Im Ergebnis zeigt das inCLASS einige Stärken und Schwächen. Stärken liegen unter anderem in der forschungsbasierten Ausrichtung auf die Individualebene von Kindern, was es ermöglicht, Verläufe der kindlichen Interaktionsqualität längsschnittlich nachzuzeichnen. Limitationen und Herausforderungen betreffen insbesondere die Faktorenstruktur und die Konfundierung der Erfassung von kindlichen Interaktionen und Kompetenzen.

**Schlüsselwörter:** inCLASS, Interaktionsqualität von Kindern, Kindertageseinrichtungen, Messgüte

## Measurement of Children's Interaction Quality in Daycare Centers. A Critical-Constructive Study of the Individualized Classroom Assessment Scoring System

**Abstract:** The *Individualized Classroom Assessment Scoring System* (inCLASS) is an observation tool that aims to measure the quality of interactions in daycare centers at the child level. The paper examines whether the inCLASS meets common scientific criteria to fulfill its aim. In addition, it identifies the challenges and limitations of the inCLASS. The study is based on our own research experience with the instrument. Further, we make reference to other quality measurement tools for daycare centers (CLASS and ECERS). The inCLASS shows some strengths and weaknesses. Strengths include the research-based focus on the individual child level, which allows longitudinal tracking of children's interaction quality. Limitations and challenges relate in particular to the factor structure of the inCLASS and the overlapping measurement of children's interactions and competencies.

**Keywords:** inCLASS, interaction quality of children, daycare centers, measurement quality

Der Qualität pädagogischer Interaktionen wird eine zentrale Bedeutung für die frühkindliche Entwicklung zugeschrieben. Interaktionsqualität ist daher ein viel beforschter Gegenstand der frühkindlichen Bildungsforschung (Mayer & Beckh, 2018; Howard et al., 2020; Kluczniok & Schmidt, 2021). Im Blick einschlägig quantitativer Studien sind meistens Interaktionen, die von pädagogischen Fachkräften in Kindertageseinrichtungen (Kitas) ausgehen und auf eine Gruppe von Kindern wirken. Weitaus weniger be-

forscht ist dagegen die Qualität von Interaktionen, die von einzelnen Kindern in Kitas ausgehen und auf andere Kinder und die Fachkräfte gerichtet sind (vgl. Molina Ramirez, 2022). Diese Perspektive einzunehmen, erscheint dann sinnvoll und gewinnbringend, wenn Forschung darauf zielt, herauszufinden, inwieweit es Kindern gelingt, sich mit der sie umgebenden Kita-Umwelt konstruktiv und für ihre Entwicklung gewinnbringend auseinanderzusetzen. Mit dem *Individualized Classroom Assessment Scoring System*

tem (inCLASS) wurde von Downer, Booren, Lima, Luckner und Pianta (2010) ein Beobachtungsinstrument vorgelegt, das den Anspruch erhebt, die Interaktionen einzelner Kinder in Kitas nach Maßgabe gängiger Gütekriterien erfassen und in ihrer Qualität bestimmen zu können. Damit sind, neben der Frage der Übertragbarkeit des US-amerikanischen Instruments auf das deutsche Kita-System, einige weitere Herausforderungen und Limitationen verbunden.

Vor dem Hintergrund des zunehmenden Einsatzes des inCLASS und ähnlicher standardisierter Beobachtungsinstrumente in der frühkindlichen Bildungsforschung geht der Beitrag den Fragen nach, inwieweit das inCLASS geeignet ist, die Interaktionsqualität von Kindern in Kitas objektiv, reliabel und valide zu messen und welche Herausforderungen und Limitationen damit verbunden sind. Beabsichtigt ist eine kritisch-konstruktive methodische und methodologische Einordnung des inCLASS unter Berücksichtigung seiner theoretisch-konzeptionellen Prämissen. Durch Bezüge zu den ähnlichen, weiter verbreiteten Instrumenten CLASS und KES/ECERS ist der Beitrag überdies von exemplarischem Nutzen für die Beurteilung auch dieser standardisierten Beobachtungsinstrumente zur Messung der Qualität von Interaktionen in Kitas – und der mit ihnen erhobenen Daten. Die Ausführungen stützen sich auf eigene Forschungserfahrungen mit dem inCLASS im Projekt *Interaktionsqualität von Kindergartenkindern über die Kindergartenzeit*<sup>1</sup>.

Zunächst wird auf das Forschungsparadigma eingegangen, das dem inCLASS zugrunde liegt. Daraufhin wird der Aufbau und der Fokus des inCLASS aufgezeigt. Ausführlicher eingegangen wird im Anschluss auf die Güte des Instruments, insbesondere auf seine Reliabilität und Validität. Der Beitrag schließt mit einem Fazit, das Vorzüge und Grenzen des inCLASS zusammenfasst und bilanziert.

## Paradigmatische Verortung des inCLASS

Das inCLASS reiht sich in bestehende Instrumente zur Erfassung kindlicher Interaktionen in Kitas auf der Ebene einzelner Kinder ein (für eine Übersicht vgl. Riedmeier, 2019). Der Blick richtet sich dabei darauf, wie gut es einzelnen Kindern gelingt, in Interaktion zu ihrer Umwelt in Kitas zu treten und welches Angebot an Interaktionen einzelne Kinder in der Kita konkret erfahren. Interaktionsqualität wird hier als Teilbereich der Prozessqualität verstanden, die im strukturell-prozessualen Qualitätsmodell als „Transmissionsriemen“ gilt. Dieses Rahmenmodell, das zwar in an-

deren Kontexten entwickelt wurde, aber auf das inCLASS angewendet werden kann, geht davon aus, dass Qualität messbar ist und über die Prozessqualität die anderen Qualitätsdimensionen (Struktur- und Orientierungsqualität) im Sinne von Entwicklungsanregungen an die Kinder weitergegeben werden (Kluczniok & Roßbach, 2014). Das inCLASS folgt demnach dem quantitativ-nomothetisch ausgerichteten Forschungsparadigma und formuliert allgemeingültige globale Qualitätsmerkmale (hier: kindliche Interaktionsqualität), die aus Forscherkenntnissen zur Prädiktion sprachlich-kognitiver und sozial-emotionaler Maße der kindlichen Entwicklung abgeleitet sind und daher als wesentlich für die Erfassung kindlicher Interaktionsprozesse betrachtet werden (Downer et al., 2010). Als weitere theoretische Rahmung werden interaktionstheoretische und ökologische Ansätze (Bronfenbrenner & Morris, 1998; Piaget, 1978; Wygotsky, 1987) herangezogen, die die Bedeutung der kindlichen Lernumgebung hervorheben. Damit werden von den Autor\_innen des inCLASS an die Entwicklungspsychologie angelehnte Vorannahmen getroffen, die sich am Kompetenzparadigma orientieren und eine Outcome-Perspektive verfolgen (siehe auch Kärtnner, 2019). Das bedeutet, dass die Kompetenzen des einzelnen Kindes in den Fokus genommen werden, die danach bemessen werden, inwieweit sich das Kind die Verhaltens- und Bedeutungssysteme seiner Lernumwelt angeeignet hat.

Kritiker\_innen sehen bei einem outcome-orientierten Verständnis pädagogischer Qualität die Gefahr, pädagogische Praxis zu verobjektivieren (Bilgi, 2021; Neumann, 2015). Die Fokussierung des Qualitätsbegriffs auf die Entwicklungsunterstützung von Kindern sei demnach nur vermeintlich ein eindeutiger Maßstab und eine „Legitimationsressource“ (Neumann, 2015, S.138) für Evaluationen zu pädagogischer Qualität. Nicht hinreichend berücksichtigt würde dabei, dass Qualität immer abhängig von der jeweiligen Perspektive (z.B. Eltern, Fachkräfte, Träger) und damit interessensgebunden ist. Befürworter\_innen des outcome-orientierten Qualitätsverständnisses setzen dagegen das Wohlbefinden der Kinder und ihre optimalen Entwicklungsmöglichkeiten in den Vordergrund und richten daran den Qualitätsbegriff aus (Tietze et al., 2017).

## Aufbau und Fokus des inCLASS

Mit dem inCLASS werden über standardisierte Beobachtungen die Interaktionen einzelner Kinder in einer Kita hoch-inferent erhoben. Das Qualitäts-Rating umfasst dabei drei übergeordnete Domänen mit jeweils untergeordneten Dimensionen (Downer et al., 2010). Da das In-

<sup>1</sup> Von 2017 bis 2021 gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projekt-Nr. 321014529.

strument bisher nur in englischer Sprache veröffentlicht wurde, haben wir die Domänen und Dimensionen im Folgenden übersetzt:

- *Interaktion mit Fachkräften* mit drei Dimensionen: Beziehungsverhalten gegenüber der Fachkraft, Kommunikation mit der Fachkraft, Konflikthafte Interaktionen mit der Fachkraft
- *Interaktion mit Peers* mit vier Dimensionen: Beziehungsverhalten zu Peers, Kommunikation mit Peers, Führungsverhalten gegenüber Peers, Konflikthafte Interaktionen mit Peers
- *Bewältigung von Herausforderungen/Aufgaben* mit drei Dimensionen: Auseinandersetzung mit Aufgaben, Selbstständigkeit, Verhaltenssteuerung

Die einzelnen Dimensionen sind anhand weiterer Indikatoren und Beobachtungshinweise ausdifferenziert. Pro Kitagruppe können bis zu vier Zielkinder im Wechsel während eines Vormittags (über einen Zeitraum von ca. vier Stunden) anhand der zehn Dimensionen beobachtet werden. Ein Beobachtungsdurchgang gliedert sich in eine zehnminütige Phase der Beobachtung und Protokollierung des kindlichen Verhaltens anhand vordefinierter Indikatoren und in eine fünfminütige, sich anschließende Phase, in der das Niveau der beobachteten Interaktionen auf einer Skala von 1 bis 7 eingeschätzt wird (1 bis 2 = niedriges Niveau; 3 bis 5 = mittleres Niveau; 6 bis 7 = hohes Ni-

veau). Pro Zielkind können an einem Vormittag mehrere Durchgänge durchgeführt werden (z.B. im Mittel 3,5 Durchgänge bei von Suchodoletz et al., 2015). Die Werte der Durchgänge werden anschließend zu einem Gesamtwert gemittelt (Downer et al., 2010).

Im deutschsprachigen Raum liegen unserem Kenntnisstand nach bislang (publizierte) Erfahrungen zum inCLASS aus Forschungsprojekten in Freiburg (von Suchodoletz et al., 2015), Landau/Berlin (Kluczniok & Schmidt, 2021), Tirol (Österreich; Smidt & Embacher, 2021) und der Schweiz (Diebold & Perren, 2022) vor. Die Studien sind mehrheitlich querschnittlich angelegt. Für die nachfolgende Einordnung des inCLASS werden Daten und Erfahrungen aus der Längsschnittstudie *Interaktionsqualität von Kindergartenkindern über die Kindergartenzeit* (Kluczniok & Schmidt, 2021; Molina Ramirez, 2022) herangezogen.

## Güte des Instruments

Im Folgenden wird auf Analysen eingegangen, die die Güte des inCLASS mit Blick auf die klassischen Gütekriterien der Objektivität, Reliabilität und Validität genauer beleuchten. Zur leichteren Orientierung wird zunächst ein knapper Überblick über zentrale deskriptive Projektergebnisse zum Niveau zielkindbezogener Interaktionsqualität gegeben.

**Tabelle 1.** Deskriptive Statistik zur kindlichen Interaktionsqualität (inCLASS)

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>α</i>	<i>Item-Total-Korrelation</i>
<i>Kindliche Interaktionsqualität in Kitas</i> (1 – 2 = niedrig; 3 – 5 = mittel; 6 – 7 = hoch)							
Interaktionen mit Peers MZP1 <sup>a</sup>	160	2.32	.78	1.00	4.90	.84	.65 – .79
Interaktionen mit Peers MZP2	160	2.92	.89	1.10	5.40	.89	.73 – .85
Interaktionen mit Peers MZP3	160	2.95	.99	1.10	6.80	.87	.70 – .80
Interaktionen mit Fachkräften MZP1 <sup>b</sup>	160	2.07	.80	1.00	4.30	.86	.76
Interaktionen mit Fachkräften MZP2	160	1.96	.71	1.00	4.90	.90	.82
Interaktionen mit Fachkräften MZP3	160	1.98	.74	1.00	5.30	.86	.76
Konflikthafte Interaktionen MZP1 <sup>c</sup>	160	1.65	.41	1.00	3.00	.52	.40 – .54
Konflikthafte Interaktionen MZP2	160	1.31	.30	1.00	2.70	.29	.22 – .26
Konflikthafte Interaktionen MZP3	160	1.41	.43	1.00	4.30	.68	.54 – .64
Bewältigung von Herausforderungen/ Aufgaben MZP1 <sup>d</sup>	160	3.62	.79	1.67	6.00	.10	.05
Bewältigung von Herausforderungen/ Aufgaben MZP2	160	3.98	.92	1.88	6.25	.57	.41
Bewältigung von Herausforderungen/ Aufgaben MZP3	160	4.12	.98	1.63	6.38	.59	.43

<sup>a</sup> Beispielim aus der Skala „Interaktionen mit Peers“: Beziehungsverhalten zu Peers.

<sup>b</sup> Beispielim aus der Skala „Interaktionen mit Fachkräften“: Kommunikation mit der Fachkraft.

<sup>c</sup> Beispielim aus der Skala „Konflikthafte Interaktionen“: Konflikthafte Interaktionen mit der Fachkraft.

<sup>d</sup> Beispielim aus der Skala „Bewältigung von Herausforderungen/Aufgaben“: Eigenständigkeit

## Deskriptive Statistik aus der eigenen inCLASS-Studie

Die Studie *Interaktionsqualität von Kindergartenkindern über die Kindergartenzeit* erstreckte sich auf drei Messzeitpunkte über die gesamte Kindergartenzeit der beobachteten Kinder (1. MZP Spät- bis Frühjahr 2017/2018, 2. MZP Spät- bis Frühjahr 2018/2019, 3. MZP Spät- bis Frühjahr 2019/2020). Die Längsschnittstichprobe (Kinder, die zu allen drei MZP beobachtet werden konnten) umfasste 160 Kinder aus 84 Gruppen in 56 Kindergärten der Bundesländer Rheinland-Pfalz und Baden-Württemberg. Die Stichprobe wurde hinsichtlich relevanter Strukturmerkmale (Fachkraft-Kind-Schlüssel, Anteil an Kindern mit nicht-deutscher Familiensprache, Ausbildungsstand des pädagogischen Personals) mit Strukturdaten von Kitas der beiden Bundesländer abgeglichen. Sie zeigt keine nennenswerten Verzerrungen auf. Zum ersten Messzeitpunkt wurden Kinder im ersten Kindergartenjahr im Alter von etwa drei Jahren (gemittelt 42 Monate) einbezogen. Davon wurden pro Kita-Gruppe maximal vier Kinder disproportional zugunsten von Kindern mit Migrationshintergrund ausgewählt. Realisiert werden konnte ein Anteil an Kindern mit Migrationshintergrund (definiert über die in der Familie gesprochene Sprache) von im Mittel 24,5%. Zum dritten Messzeitpunkt waren die Kinder durchschnittlich etwa fünfeneinhalb Jahre alt (66 Monate). Die über die drei Messzeitpunkte auf der Grundlage des amerikanischen Beobachtermanuals ermittelte Interaktionsqualität der Kinder erreichte ein niedriges bis mittleres Niveau (Tabelle 1).<sup>2</sup>

Zur Einordnung der Ergebnisse wären Normstichproben hilfreich. Für das inCLASS, wie auch für andere gängige Beobachtungsinstrumente zur Messung von Interaktionsqualität in Kitas (CLASS, KES bezogen auf Interaktionsmerkmale) liegen allerdings keine Normstichproben vor. Die Zuordnung der beobachteten Interaktionsqualität der Kinder in „niedrig“, „mittel“ und „hoch“ ist beim inCLASS somit nicht statistisch abgesichert. Die Aussagekraft der inCLASS-Werte unterliegt überdies einer Reihe weiterer Herausforderungen und Limitationen, auf die im Folgenden eingegangen wird.

## Objektivität

Bei standardisierten Beobachtungsinstrumenten ist in der Regel von einer hohen Objektivität auszugehen, da den Beobachtungen üblicherweise ein Schema zugrunde liegt, das Beobachtungskategorien vorgibt und möglichst ein-

deutig definiert (vgl. Stangl, 2022). Da das inCLASS auf einem solchen Beobachtungsschema und einem ausführlichen Manual beruht (Downer et al., 2012), liegt es nahe, dem Instrument eine hohe Objektivität zuzuschreiben. Mit der Qualität von Interaktionen ist über das inCLASS allerdings ein komplexes latentes Konstrukt über Ratings hoch inferent einzuschätzen. Das ist anfällig für Rating-Bias (z. B. Milde- oder Strengfeehler) und bedarf daher der Überprüfung der Beobachter\_innen durch systematische Abgleiche ihrer Ratings (z. B. mittels Doppelkodierung und Kalibrierung). Auf diesen Aspekt wird im Abschnitt *Reliabilität* genauer eingegangen.

## Reliabilität

Mit Blick auf die *Reliabilität* können unterschiedliche Maße erfasst und für die Prüfung des inCLASS herangezogen werden. Untersucht wurde im Vorliegenden die *interne Konsistenz* gemessen über Cronbach's  $\alpha$ , die *Interrater-Reliabilität* gemessen durch Cohen's  $\kappa$ , sowie die *Stabilität über die Zeit* gemessen über Pearson  $r$ .

Bezüglich der *internen Konsistenz* der Domänen des inCLASS ergaben sich für die *Interaktionen mit Peers* ( $\alpha = .84-.89$ ), sowie für die *Interaktionen mit Fachkräften* ( $\alpha = .86-.90$ ) zu allen drei Messzeitpunkten hohe Reliabilitätswerte. Die einzelnen Items der beiden Domänen sprechen jeweils für eine gute Trennschärfe (*Interaktionen mit Peers*: Item-Total-Korrelation =  $.65-.85$ ; *Interaktionen mit Fachkräften*: Item-Total-Korrelation =  $.76-.82$ ). Für die Domäne *Konfliktvolle Interaktionen* ( $\alpha = .29-.68$ ), sowie für die *Bewältigung von Herausforderungen/Aufgaben* ( $\alpha = .10-.59$ ) zeigen sich zu zwei Messzeitpunkten niedrige und zu je einem Messzeitpunkt inakzeptable Reliabilitätswerte (Tabelle 1). Dies könnte darauf zurückzuführen sein, dass die beobachteten Kinder sowohl Konflikte als auch Herausforderungen individuell sehr unterschiedlich erleben und damit die Abbildung dieser latenten Konstrukte erschwert wird. Demnach kann auf Basis der erhobenen Daten unserer Studie die interne Konsistenz der von den Autor\_innen des inCLASS (Downer et al., 2010) vorgegebenen Skalenbildung nicht durchgängig bestätigt werden.

Des Weiteren gibt es zunehmend kritische Analysen auch hinsichtlich ähnlicher Instrumente wie dem *Classroom Assessment Scoring System* (CLASS; Pianta, La Paro & Hamre, 2012), der Instrumente aus der Skalenfamilie der *Early Childhood Environment Rating Scale* (ECERS) bzw. der deutschen Adaptionen *Kindergarten-Skala* (KES-RZ; Tietze et al., 2017). Diese Analysen machen auf Einschränkungen in der

<sup>2</sup> Abweichend vom Aufbau des inCLASS (drei Dimensionen) wurden von den Autor\_innen des inCLASS mittels konfirmatorischer Faktorenanalysen vier Faktoren (zusätzlich konfliktvolle Interaktionen) ermittelt (vgl. Molina Ramirez, 2022). Die in Tabelle 1 berichteten Ergebnisse beziehen sich auf diese vier Faktoren.

Reliabilität und Validität durch Verzerrungen, wie z.B. Rating Bias (Milde- oder Strengefehler) aufmerksam (Leber, Kammermeyer & Roux, 2020; Mayer & Beckh, 2018). Gründe dafür können zum einen unterschiedliche Interpretationen der beobachteten Situation und zum anderen beobachterspezifische Wahrnehmungen sein. Diese Unterschiede zwischen den Einschätzungen werden in der Regel durch Unterschiede in der Tendenz zur Strenge bzw. Milde erklärt (zu Rating-Fehlern siehe Lotz, Gabriel & Lipowsky, 2013). Andere Studien, wie Downer et al. (2012), die die Reliabilität des inCLASS in einem Test-Retest-Design prüften, berichteten über eine mäßige bis hohe Stabilität über zwei Messzeitpunkte (Herbst und Frühjahr), ebenso wie Slot und Bleses (2018), die etwas stärkere Korrelationen zwischen zwei Beobachtungstagen innerhalb von zwei Wochen im Vergleich zu den Korrelationen innerhalb eines Tages (Vormittag und Nachmittag) berichten. Unsere Daten weisen über alle drei Messzeitpunkte hinweg (Jahresabstand bzw. Zweijahresabstand) niedrige bis mittlere Zusammenhänge der einzelnen Skalen auf (zwischen  $r = .15$  und  $r = .29$ ; Ausreißer:  $r = -.05$  Interaktion mit Fachkräften im Zweijahresabstand). Insgesamt sind die Stabilitäten im Jahresabstand leicht höher im Vergleich zum Zweijahresabstand, was darauf hindeutet, dass sich die Interaktionsqualität der Kinder über die Kindergartenzeit verändert. Entwicklungsbedingt ist das erwartbar. Zudem sind Fachkraftwechsel nicht ausgeschlossen, was insbesondere die Fachkraft-Kind-Interaktion verändern kann.

Rating-Fehlern soll in der Regel mit einer umfangreichen Schulung der Erheber\_innen auf Grundlage eines Bewertungsmanuals vorgebeugt werden. Um das inCLASS nach den Vorgaben der Autor\_innen anwenden zu können, haben die Erheber\_innen im Projekt ein zweitägiges Training mit Videosequenzen absolviert. Am Ende der Schulung erfolgte eine Reliabilitätsprüfung der Erheber\_innen, bei der mindestens 80 % Übereinstimmung mit dem von den Autor\_innen vorgegebenen Master-Code nachgewiesen werden musste. Wie bereits in anderen Studien (u.a. Downer et al., 2010) konnte unter den Projektmitarbeitenden eine sehr gute *Interrater-Reliabilität* von  $\kappa = .81$  erreicht werden. Hierfür wurden zum ersten Messzeitpunkt 16 % der Erhebungen doppelt kodiert.

Dennoch bietet die Feststellung der *Interrater-Reliabilität* anhand von Doppelkodierungen nur eine begrenzte Kontrolle. Verzerrungen sind, zumal über einen längeren Zeitraum, nicht auszuschließen. Die Autor\_innen des inCLASS sehen daher, wie auch beim CLASS, nach einem Jahr eine Rezertifizierung mit erneuter Reliabilitätsprüfung vor. Eine Studie mit dem CLASS von Leber und Kolleginnen (2020) zeigt zudem, dass die Datenqualität durch einen mehrstufigen Kalibrierungsprozess (systematischer Abgleich von Ratings auf Basis von Video-Beobachtungen) deutlich verbessert werden kann.

## Validität

Bezüglich des inCLASS haben wir vor allem die Konstruktvalidität (Klarheit des theoretischen Konstrukts) untersucht. Die theoretisch und empirisch hergeleiteten Domänen des inCLASS wurden – wie schon von den Autor\_innen des inCLASS (Downer et al., 2010) – im Rahmen einer konfirmatorischen Faktorenanalyse geprüft. Dabei haben wir drei Faktoren identifiziert (Tabelle 2).

Ein weiterer Faktor *Bewältigung von Herausforderungen/Aufgaben* (vgl. Downer et al., 2010) konnte nicht identifiziert werden. In anderen Studien zeigt sich ein ähnliches Bild (Bohlmann et al., 2019; Slot & Bleses, 2018; von Suchodoletz, Gunzenhauser & Larsen, 2015). Allerdings erscheint es notwendig, die Modellfitwerte der konfirmatorischen Faktorenanalyse zu reflektieren. In unserer Studie ergaben sich folgende Werte:  $\chi^2 (18) = 38,63$ ,  $p = 0.01$ ; RMSEA = 0.09; CFI = 0.96;  $R^2 = 0.98$ . Unter Berücksichtigung des Satorra-Bentler-Schätzers findet sich ein RMSEA-Wert von .09, der gemäß Reinecke (2014) als grenzwertig hoch einzuschätzen ist, auch wenn sich bei Bohlmann et al. (2019) sehr ähnlich hohe Modellfitwerte zeigen, welche von den Autor\_innen als „acceptable to good model fit“ (Bohlmann et al., 2019, S.172) bewertet werden. Aus forschungstheoretischen Arbeiten, die sich mit Modellfitwerten konfirmatorischer Faktorenanalysen mit Blick auf den RMSEA befassen, finden sich widersprüchliche Aussagen: Während der ermittelte RMSEA-Wert nach Byrne (2001) als mittelmäßig einzustufen ist, gilt er gemäß Hennig-Thurau (1998) als noch akzeptabel. Nach Reinecke (2014) ist er dagegen als inakzeptabel einzustufen.

Des Weiteren weist das inCLASS in seinen Dimensionen Unklarheiten auf. Einerseits zielt das Instrument darauf ab, kindliche Interaktionen zu messen. Das wird beispielsweise an folgender Formulierung deutlich: „Positive Engagement with the Teacher measures the degree to which the child is emotionally connected to the teacher [...]“ (Downer et al., 2012, S.23). Andererseits zielt das Rating auf der 7-stufigen Skala häufig darauf ab, die *Quantität*

**Tabelle 2.** Konfirmatorische Faktorenanalyse

Faktor	Items	Ladung
Interaktionen mit Peers	Assertiveness	.71
	Sociability	.84
	communication peers	.90
Interaktionen mit Fachkräften	positive engagement	1.0
	communication teacher	.78
Konflikthafte Interaktionen	behaviour control [umkodiert]	.60
	conflict teacher	.73
	conflict peers	.64

der beobachteten kindlichen Interaktionen zu bewerten, von der aus dann – wie beispielsweise auch bei dem CLASS oder der KES-R – auf Qualität geschlossen wird. Dies zeigt sich an Formulierungen wie „few indications“, „sometimes“ und „consistently“ für niedrige, mittlere und hohe Werte (Downer et al., 2012, S.23).

Unklar ist auch, ob sich die einzelnen Dimensionen nicht gegenseitig ausschließen können, sofern es während eines Beobachtungszyklus zu Bedingungen kommt, unter denen das Kind keine hohe interaktive Eingebundenheit mit Fachkräften, Peers und Anforderungen/Aufgaben zugleich zeigen kann. Durch gemittelte Werte mehrerer Zyklen kann dies allerdings ausgeglichen werden (siehe dazu Molina Ramirez, 2022).

Nicht zuletzt stellt sich die Frage, ob das inCLASS ohne weiteres auf Kitas in Deutschland angewendet werden kann. Zum einen ist die Frühpädagogik in Deutschland durch ein breites Spektrum an teilweise deutlich divergierenden Handlungskonzepten gekennzeichnet (Schmidt, Sauerbrey & Smidt, 2021). Zum anderen sind für die Frühpädagogik in deutschen Kitas große Freispielanteile charakteristisch<sup>3</sup>, in denen sich die pädagogischen Fachkräfte gegenüber den Kindern traditionell eher zurückhalten (Tournier, Wadepohl & Kucharz, 2014). Vor diesem Hintergrund ist zu prüfen, ob zwischen dem inCLASS und spezifischen deutschen Kita-Traditionen eine hinreichende Passung gegeben ist.

## Fazit

Insgesamt betrachtet ergänzt das inCLASS die bisher überschaubare Auswahl an standardisierten Instrumenten zur Messung kindlicher Interaktionen in Kitas. Mit seinem Schwerpunkt auf der Interaktionsqualität von Kindern fokussiert es auf einen bedeutsamen Bereich der Prozessqualität. Erfasst werden können mit dem inCLASS Aspekte kindlicher Interaktionen, die für die spätere Entwicklung und den Schulerfolg von Kindern nachweislich prädiktiv sind. Innovativ für die Kita-Qualitätsforschung ist darüber hinaus die Messung der Interaktionsqualität auf der Individualebene der Kinder. Dies ermöglicht nicht nur Aussagen über die Interaktionsqualität einzelner Kinder, sondern auch, Verläufe der kindlichen Interaktionsqualität längsschnittlich nachzuzeichnen. Das inCLASS ist inzwischen umfangreich getestet und seine Messeigenschaften sind ausführlich dokumentiert (Downer et al., 2010; Slot & Bleses, 2018; Smidt & Embacher, 2021; von Suchodoletz et al., 2015). Das Instrument ermöglicht auf der Grundlage

einer professionellen Schulung und eines klar und ausführlich ausgearbeiteten Manuals an einem Kita-Tag bis zu vier Kinder in ihrer Interaktionsqualität einzuschätzen. Es erfüllt damit wesentliche Aspekte ökonomischer Güte, die den Einsatz auch bei large scale Studien ermöglichen.

Hinsichtlich der Gütekriterien der Reliabilität und Validität kann das Instrument seinem hohen Anspruch jedoch nicht gerecht werden. Die Faktorenstruktur des inCLASS ist suboptimal, was aus forschungstheoretischer Sicht die Frage nach der Auswahl der Domänen und Dimensionen aufwirft. Als problematisch erweist sich zudem die Konfundierung der Erfassung von kindlichen Interaktionen und Kompetenzen. Unklar bleibt dadurch, ob die erhobenen Daten eher etwas über die Interaktionsqualität (Output) oder über die Kompetenzen von Kindern (Outcome) aussagen.

Eine weitere Konfundierung bei der Qualitätsmessung besteht mit dem Kita-Angebot: Inwieweit hat das Kind im Beobachtungszeitraum durch die Gegebenheiten in der Kita (Interaktionsverhalten von Peers und Fachkräften, räumliche und materielle Ausstattung) überhaupt die Möglichkeit, qualitativ zu interagieren? Dies sind Aspekte, die äußerst vielschichtig und vom Interaktionsverhalten des Kindes nicht zu isolieren sind. Mit anderen Worten: Ein ermittelter inCLASS-Wert sagt *zugleich und untrennbar* etwas über die Interaktionsqualität und -quantität von Kindern, ihre Interaktionskompetenzen sowie die Qualität und Quantität des Interaktionsangebotes in der Kita aus. Solche inhaltlichen Überschneidungen bestehen auch bei anderen standardisierten Beobachtungsinstrumenten zur Messung pädagogischer Qualität in Kitas (z.B. CLASS, KES/ECERS). Beim inCLASS sind sie allerdings besonders markant. Sie stellen das inCLASS nicht gänzlich in Frage, verdeutlichen aber, dass die damit erzeugten Zahlenwerte unscharf sind. Fällt ein inCLASS-Wert beispielsweise sehr niedrig aus, ist bei der Interpretation des Wertes zu berücksichtigen, dass dies nicht nur an der gezeigten Interaktionsqualität des Kindes und/oder seinem Entwicklungsstand liegen muss. Der sehr niedrige Wert ist auch als Indikator dafür zu lesen, dass das Interaktionsangebot in der Kita für dieses Kind womöglich unzureichend war. Wichtige Schritte zur weiteren Prüfung des inCLASS stellen systematische Vergleiche mit anderen standardisierten Beobachtungsverfahren auf Zielkindebene und die Erarbeitung von repräsentativen Normstichproben dar.

Die im Abschnitt *Paradigmatische Verortung des inCLASS* angesprochene grundsätzliche Kritik an quantitativ-nomothetisch ausgerichteten und outcome-orientierten Messinstrumenten, die normative Setzungen zur Bewertung der Interaktionsqualität in Kitas enthalten (Bilgi, 2021; Neumann,

<sup>3</sup> Kluczniok und Schmidt (2021) ermittelten auf Basis des vorliegenden Datensatzes einen gemittelten Anteil von rund 62 % Freispiel-Anteilen an Kita-Vormittagen, gefolgt von geplanten (geleiteten) Aktivitäten (14 %), Mahlzeiten (11 %) und Übergangsphasen (7 %).

2015), teilen wir allerdings nicht. Das inCLASS wie auch das CLASS und die KES machen ihre normativen Setzungen in ihren Items, Skalierungen und Manualen u.E. hinreichend transparent. Sie unterliegen Prozessen der Adaption an sich verändernde pädagogische Gegebenheiten und Fachdiskussionen und sind in diesem Sinne kontinuierlich verbesserbar. Die Ausrichtung der genannten Instrumente auf das Wohlbefinden und möglichst gute Entwicklungsmöglichkeiten von Kindern in Kitas erachten wir als plausibel. Vor dem Hintergrund ihrer jeweiligen Begrenzungen (Schwerpunktsetzungen, messtheoretische Mängel) erscheint jedoch eine Pluralität und Weiterentwicklung von Qualitätsmessinstrumenten erforderlich. Zudem sollten aus den mit dem inCLASS erzeugten Messwerten keine pauschalen bildungspolitischen Forderungen abgeleitet werden. Zukünftige Forschung mit dem inCLASS in Deutschland könnte dazu beitragen, das Instrument weiter zu prüfen und ggf. Verbesserungen vorzuschlagen. Dabei sollte auch der Aspekt der Passung zu Spezifika der Frühpädagogik in Deutschland berücksichtigt werden.

## Literatur

- Bilgi, O. (2021). Qualität pädagogisch gedacht: Ein historisch-systematischer Beitrag zum Qualitätsverständnis in der Pädagogik der frühen Kindheit. In O. Bilgi, G. Blaschke-Nacak, J. Durand, T. Schmidt, U. Stenger & C. Stieve (Hrsg.), „Qualität“ revisited, (S. 49 – 67). Weinheim, Basel: Beltz Juventa.
- Bohlmann, N. & Downer, J. T. (2019). Observing children's engagement: Examining factorial validity of the inCLASS across demographic groups. *Journal of Applied Developmental Psychology*, 60, 166 – 176.
- Bronfenbrenner, U. & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, Vol. 1: Theoretical models of human development* (pp. 993 – 1023). New York: John Wiley and Sons.
- Byrne, B. (2001). *Structural Equation Modeling with AMOS, Basic Concepts, Applications, and Programming* (pp. 79 – 88). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Diebold, T. & Perren, S. (2022). Toddlers' peer engagement in Swiss childcare: Contribution of individual and contextual characteristics. *European Journal of Psychology of Education*, 37(3), 627 – 648. <https://doi.org/10.1007/s10212-021-00552-2>
- 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 interaction. *Early Childhood Research Quarterly*, 25(1), 1 – 16.
- Downer, J., Booren, L., Hamre, B., Pianta, R., Williford, A. (2012). *The Individualized Classroom Assessment Scoring System (inCLASS) pre-K coding manual*. University of Virginia, CASTL.
- Howard, S. J., Siraj, I., Melhuish, E. C., Kingston, D., Neilsen-Hewett, C., de Rosnay, M. Duursma, E. & Luu, B. (2020). Measuring interactional quality in pre-school settings: Introduction and validation of the Sustained Shared Thinking and Emotional Wellbeing (SSTEW) scale. *Early Child Development and Care*, 190(7), 1017 – 1030.
- Kärtner, J. (2019). Kind, Kindheit und Entwicklung in der Entwicklungspsychologie. In Dietrich, C., Stenger, U. & Stieve, C. (Hrsg.) *Theoretische Zugänge zur Pädagogik der frühen Kindheit. Eine kritische Vergewisserung*. Weinheim: Beltz Juventa.
- Kluczniok, K. & Roßbach, H.-G. (2014). Conceptions of educational quality for kindergartens. *Zeitschrift für Erziehungswissenschaft*, 17(6), 145 – 158. <https://doi.org/10.1007/s11618-014-0578-2>
- Kluczniok, K. & Schmidt, T. (2021). Zur Bedeutung des pädagogischen Settings für die Interaktionsqualität von Kindern im Kindergarten. *Frühe Bildung*, 10(4), 214 – 223.
- Leber, A., Kammermeyer, G. & Roux, S. (2020). Sicherung der Qualität von Beobachtungen der ErzieherIn-Kind-Interaktion mit dem Classroom Assessment Scoring System. In K. Blatter, K. Groth & M. Hasselhorn (Hrsg.), *Evidenzbasierte Überprüfung von Sprachförderkonzepten im Elementarbereich* (S. 101 – 125). Wiesbaden: Springer VS.
- Lotz, M., Gabriel, K. & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung. *Zeitschrift für Pädagogik*, 59, 357 – 380.
- Hennig-Thurau, T. (1998). *Konsum-Kompetenz – Eine Zielgröße für das Management von Geschäftsbeziehungen*. Frankfurt a. M.: Lang.
- Mayer, D. & Beckh, K. (2018). Erfassung pädagogischer Qualität in Kindertageseinrichtungen. *Frühe Bildung*, 7, 67 – 76. <https://doi.org/10.1026/2191-9186/a000370>
- Molina Ramirez, M. (2022). *Zielkindbezogene Prozessqualität in Kindergärten im Kontext kindspezifischer, familialer und kindergartenbezogener Merkmale* (Dissertation). Universität Bamberg. <https://doi.org/10.20378/irb-53535>
- Neumann, S. (2015). Welche Qualität für wen? Leerstellen einer (immer noch) aktuellen Debatte. *Und Kinder*, 95, 137 – 139.
- Piaget, J. (1978). *Das Weltbild des Kindes*. Stuttgart: Klett-Cotta.
- Pianta, R., La Paro, K. & Hamre, B. (2012). *Classroom Assessment Scoring System (CLASS) manual PRE-K* (8th ed.). Baltimore: Paul H. Brookes Publishing Co.
- Reinecke, J. (2014). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: De Gruyter Oldenbourg.
- Riedmeier (Molina Ramirez), M. (2019). Standardisierte Verfahren zur Erhebung zielkindbezogener Prozessqualität in der Frühpädagogik – ein vergleichender Überblick. *Frühe Bildung*, 8(3), 144 – 152.
- Schmidt, T., Sauerbrey, U. & Smidt, W. (Hrsg.). (2021). *Frühpädagogische Handlungskonzepte. Eine wissenschaftliche Bestandsaufnahme*. Münster: Waxmann utb.
- Slot, P. L. & Bleses, D. (2018). Individual children's interactions with teachers, peers, and tasks: The applicability of the inCLASS Pre-K in Danish preschools. *Learning and Individual Differences*, 61, 68 – 76.
- Smidt, W. & Embacher, E.-M. (2021). Examining the factorial validity of the Individualized Classroom Assessment Scoring System in preschools in Austria. *International Journal of Early Years Education*, 31(3), 675 – 687. <https://doi.org/10.1080/09669760.2021.1893158>
- Stangl, W. (2022). *Standardisierte Beobachtung. Online Lexikon für Psychologie und Pädagogik*. <https://lexikon.stangl.eu/10759/standardisierte-beobachtung>
- Tietze, W., Roßbach, H.-G., Nattefort, R. & Grenner, K. (2017). *Kindergarten-Skala. Revidierte Fassung mit Zusatzmerkmalen (KES-RZ)*. Weimar: das netz.
- Tournier, M., Wadepohl, H. & Kucharz, D. (2014). Analyse des pädagogischen Handelns in der Freispielbegleitung. In D. Kucharz, K. Mackowiak, S. Ziroli, A. Kauertz, E. Rathgeb-Schnierer & M. Dieck (Hrsg.), *Professionelles Handeln im Elementarbereich (PRIMEL). Eine deutsch-schweizerische Videostudie* (S. 99 – 121). Münster: Waxmann.
- von Suchodoletz, A., Gunzenhauser, C. & Larsen, R. A. (2015). Die Beobachtung von Interaktionen im Kindergartenalltag. *Frühe Bildung*, 4(2), 1 – 7. <https://doi.org/10.1026/2191-9186/a000207>
- Wygotsky, L. (1987). *Ausgewählte Schriften. Arbeiten zur psychischen Entwicklung der Persönlichkeit* (Band 2). Köln: Pahl-Rugenstein.

Onlineveröffentlichung: 20.10.2023

#### **Autorenschaften**

Die Mitautor\_innen sind erreichbar unter Katharina Kluczniok, [katharina.kluczniok@fu-berlin.de](mailto:katharina.kluczniok@fu-berlin.de); Magdalena Molina Ramirez, [molina@dji.de](mailto:molina@dji.de); Norman Zimmermann, [norman.zimmermann@rptu.de](mailto:norman.zimmermann@rptu.de)

#### **Förderung**

Open Access-Veröffentlichung ermöglicht durch die Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU).

#### **Thilo Schmidt**

Institut für Bildung im Kindes- und Jugendalter  
Rheinland-Pfälzische Technische Universität  
Kaiserslautern-Landau (RPTU)  
August-Croissant-Straße 5  
76829 Landau i. d. Pfalz  
Deutschland  
[thilo.schmidt@rptu.de](mailto:thilo.schmidt@rptu.de)