

## Abstract

Assessing the performance of students is a core task of teachers (assessors). Their assessments are supposed to be as objective, reliable, and valid as possible, in order to provide for a reliable feedback on the students' level of skills and competences. Moreover, such assessment is required to reliably certify formal qualifications as a basis for future career paths and additional educational opportunities. In the context of performance assessments during the practical examination of student pre-school teachers, little is known about the diagnostic quality of current assessment instruments. A preliminary study has shown substantial heterogeneity among the assessment instruments used across 18 colleges. Deviations were particularly expressed with regard to the focus of observation, as well as the definition and operationalization of terminology. In addition, the instruments took only limited account of recent scientific findings on how to assess the quality of interactions between the students and the children, although this is considered to be a central variable of influence for the children's educational and developmental processes (Hamre et al., 2012; König, 2011; Reyhing, 2021). Leber et al. (2020) point out the significance of observations in capturing the quality of interactions, as well as the need for extensive training of assessors to reduce rating errors. Against this background, this study develops an instrument for an observation-based performance assessment. This instrument was experimentally tested with 73 assessors by using two video vignettes, to investigate the reliability and validity of the measures proposed by the instrument. In addition, the hypothesis that training increases the psychometric quality of the instrument was tested as follows. One group of participants ( $N = 37$ ) received a brief introduction and five hours of training on the instrument, while the other group ( $N = 36$ ) received only the brief introduction. The groups were compared in terms of their interrater reliability (Krippendorff's alpha). Expert judgment served as a validity check. Sociodemographic data, personality traits, emotional responses during the ratings, as well as a variety of aspects relating the assessors' prior expertise were collected to determine possible influences on their ratings. Overall, the results show significant observer disagreement for both groups. The interrater reliability within the training group was even lower than in the group without additional training. With regard to validity, a strong positive correlation with expert

judgment was found. The results indicate that a standardized observation instrument and a five-hour training for the assessors are insufficient to reliably assess the interaction quality of student pre-school teachers. More extensive training of assessors, and additional calibrations of the assessment instrument appear necessary to increase the quality of performance assessments in this context.

# 1 Forschungsthema und Aufbau der Arbeit

## 1.1 Relevanz des Forschungsthemas

Diese Arbeit beschäftigt sich mit der Erfassung der Interaktionsqualität in der berufspraktischen Fachschulausbildung bei angehenden Erzieher:innen. Hinsichtlich der Bildungs- und Entwicklungsprozesse von Kindern ist die Bedeutung qualitativer Interaktionsprozesse in den letzten Jahren stark in den Vordergrund gerückt (Kluczniok & Schmidt, 2021). Dabei zeigt sich, dass neben den primären Bezugspersonen auch die pädagogischen Fachkräfte familienergänzend das Lernen im Rahmen bindungsähnlicher Beziehungen durch ihre Interaktionen, besonders bei Kindern aus bildungsfernen Kontexten, beeinflussen (Keller et al., 2013; Kuger & Kluczniok, 2009). Interaktionsqualität wird dabei als mehrdimensionales Konstrukt verstanden (Hafen et al., 2014). Untersuchungen im deutschsprachigen Raum weisen darauf hin, dass die Interaktionsqualität in Kindertagesstätten sehr unterschiedlich ist und teilweise sogar ein hohes Entwicklungspotential vor allem im Bereich der Lernunterstützung aufweist (Kuger & Kluczniok, 2009; Wertfein et al., 2015; Wildgruber et al., 2016). Die Fachschulen für Sozialpädagogik tragen im Rahmen der Ausbildung von Erzieher:innen eine wesentliche Verantwortung, deren Interaktionsqualität zu fördern. Während in der fachschulischen Ausbildung u. a. Wissen und Verständnis hinsichtlich der Interaktionsqualität erfasst werden können, dienen die berufspraktische Ausbildung und am Ende die berufspraktische Prüfung dazu, die realen Interaktionsprozesse in authentischen Situationen mit den Kindern zu beobachten und zu bewerten.

Im Bereich der Ausbildung von Erzieher:innen ist über die Güte der verwendeten Beobachtungs- und Bewertungsinstrumente bislang nichts bekannt. Im Rahmen einer unveröffentlichten Vorstudie zu dieser Arbeit wurden 2018 die verwendeten Beobachtungsinstrumente für die berufspraktische Ausbildung von 18 sächsischen Fachschulen für Sozialpädagogik in öffentlicher und privater Trägerschaft ( $N = 63$ ) erhoben. Die Ergebnisse der inhaltlichen Analyse lassen sich in den nachfolgenden Herausforderungen zusammenfassen:

- Der Begriff „Interaktionsqualität“ wird in keinem der Instrumente erwähnt.

- Alle Beobachtungsinstrumente unterscheiden sich hinsichtlich der Auswahl und des Umfangs der zu beobachtenden Aspekte/Merkmale/Kriterien, wenngleich zahlreiche Parallelen erkennbar sind.
- Die zu beobachtenden Aspekte/Merkmale/Kriterien sind teilweise begrifflich sehr abstrakt (z. B. „situationsangemessenes Handeln“) und eröffnen damit eine große Bandbreite an Interpretationsmöglichkeiten, welche individuelle unterschiedliche Blickwinkel, Wahrnehmungen und Einschätzungen begünstigen.

Damit zeigt sich, dass die theoretische Modellierung von Interaktionsqualität unterschiedlich erfolgt.

In einer qualitativen Datenanalyse von Kotzeck (2021) auf Basis ausgewählter Daten aus der vorliegenden Studie, die im Rahmen offener Beobachtungen von Videovignetten eines Fachschülers durch 38 Lehrkräfte erhoben wurden, zeigen sich ebenso ganz unterschiedliche Modellierungen von Interaktionsqualität. Besonders häufig werden Aspekte der Beziehungsgestaltung beobachtet und andere Bereiche, wie beispielsweise eine dialogorientierte Bildungsunterstützung im Rahmen bildungsförderlichen Feedbacks oder kognitiver Aktivierung, vernachlässigt. Gleichzeitig führen diese Beobachtungen zu ganz unterschiedlichen Beurteilungen, was in der Datenanalyse von Kotzeck und Schulze (2022) zu Notenunterschieden von Note zwei bis Note fünf für die gleiche Leistung führt. Die Urteile der Lehrkräfte sind damit nicht objektiv und reliabel. Jäger (2000) verweist ausdrücklich auf die sachlichen Notwendigkeiten sowie die rechtliche Pflicht, eine Transparenz und Chancengleichheit im Rahmen der schulischen Leistungsbeurteilung zu ermöglichen und Willkür zu reduzieren.

Aus dieser Ausgangssituation ergeben sich zwei Fragen:

- Was ist unter Interaktionsqualität angehender Erzieher:innen zu verstehen? (theoretische Modellierung der Interaktionsqualität)
- Wie kann die Interaktionsqualität angehender Erzieher:innen objektiv, valide und reliabel erfasst werden? (empirische Erfassung der Interaktionsqualität)

Da sich die Anforderungen an die Fachschüler:innen im Rahmen der Interaktionen und damit auch die Schwerpunkte zur Erfassung der Interaktionsqualität hinsichtlich der unterschiedlichen Arbeitsfelder (z. B. Krippe, Kindertagesstätte, Jugendwohngruppe) sowie Altersgruppen (z. B. null bis drei Jahre, drei bis sechs Jahre, zwölf bis siebzehn Jahre) zum Teil unterscheiden (Geißler et al., 2022; Sommer & Sechtig, 2016) wird der Fokus in dieser Arbeit auf das Arbeitsfeld Kindertagesstätten und die Altersgruppe von drei bis sechs Jahren gelegt. Dieses Arbeitsfeld und diese Altersgruppe kommen in der berufspraktischen Ausbildung erfahrungsgemäß<sup>1</sup> besonders häufig

---

<sup>1</sup> Die Aussage basiert auf eigenen Erfahrungen mehrjähriger Tätigkeit in der Begleitung von Fachschüler:innen in der berufspraktischen Ausbildung und persönlichen Mitteilungen von Lehrkräften.

vor. Zudem liegen für diese Altersgruppe zahlreiche Studienergebnisse vor, die mit der Erfassung von Interaktionsqualität im Zusammenhang stehen und einen besseren Forschungsanschluss ermöglichen.

## 1.2 Aufbau und Struktur der Arbeit

Ausgehend von den beiden allgemeinen Fragestellungen, die im vorherigen Kapitel formuliert worden, wird im Folgenden zuerst der aktuelle Forschungsstand dargestellt. Dieser bezieht sich auf den Beobachtungsgegenstand (Interaktionsqualität), auf die Art der Datenerfassung (Beobachtungen) sowie die Interpretation der Beobachtungen (Ratings). In diesem Zusammenhang werden auch etablierte Instrumente zur Erfassung der Interaktionsqualität aus dem Bereich der Qualitätsentwicklung in der frühkindlichen Bildung vorgestellt und hinsichtlich ihrer Übertragbarkeit auf die fachschulische Ausbildung eingeschätzt. Auf Grundlage dieser Auseinandersetzung werden abschließend die forschungsleitenden Fragestellungen sowie Hypothesen formuliert.

Daran anschließend erfolgt die Darstellung zur Entwicklung eines Beobachtungsinstrumentes für die fachschulische Ausbildung.

Im dritten Kapitel wird die Studie zur Erprobung des Beobachtungsinstrumentes vorgestellt und detailliert beschrieben. Daran schließen sich in Kapitel vier die Ergebnisdarstellung und in Kapitel fünf die Ergebnisdiskussion an.

Im abschließenden Kapitel der Arbeit werden theoretische sowie praktische Konsequenzen auf Grundlage der Forschungsergebnisse diskutiert und Ausblicke auf weitere Forschungsdesiderate gegeben.