

Urteilsgenauigkeit und Beobachtungskompetenz: Untersuchung diagnostischer Kompetenzen frühpädagogischer Fachkräfte

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie (Dr. phil.)
am Fachbereich 5: Erziehungswissenschaften
der Universität Koblenz-Landau

vorgelegt von
Christian Donie
geb. am 04.04.1978 in Saarbrücken

1. Gutachterin: Prof. Dr. Gisela Kammermeyer
2. Gutachterin: Prof. Dr. Anja Wildemann
vorgelegt am: 07.06.2021

„Man sieht oft etwas hundert Mal,
tausend Mal,
ehe man es zum allerersten Mal wirklich sieht.“

(Christian Morgenstern)

Danksagung

Nach Abschluss meiner Dissertationsschrift möchte ich allen herzlich danken, die mich bei der Planung, Durchführung und Fertigstellung dieser Arbeit unterstützt haben.

Mein Dank gilt zunächst Prof. Dr. Gisela Kammermeyer für die Betreuung meiner Doktorarbeit. Vielen Dank für alle hilfreichen Hinweise und Anregungen und dafür, dass Sie zu jeder Zeit bei Fragen ansprechbar waren. Ebenso gilt mein Dank postum Prof. Dr. Susanna Roux, die in der ersten Phase meiner Dissertation eine wertvolle Stütze und Beraterin war. Prof. Dr. Anja Wildemann danke ich für die Übernahme des zweiten Gutachtens und für die Freiräume, die sie mir als Vorgesetzte zur Bearbeitung meiner Dissertation gegeben hat.

Für die Möglichkeit, im Rahmen des DFG-Graduiertenkollegs Unterrichtsprozesse zu promovieren, bin ich ebenfalls sehr dankbar – stellvertretend geht hier mein Dank an die Geschäftsführung, Dr. Heidrun Ludwig, sowie den ehemaligen Leiter des Kollegs, Prof. Dr. Wolfgang Schnotz. Danken möchte ich in diesem Zusammenhang außerdem allen Professorinnen und Professoren, die an den Forschungskolloquien, in denen wir Graduierte unsere Forschungsarbeiten präsentierten und Fragen an das Plenum richten konnten, teilgenommen haben.

Besonders erwähnen möchte ich die Hilfskräfte, Forschungsstudierenden und Mitarbeiterinnen im Projekt *FaBi (Förderung anschlussfähiger Bildungsprozesse im Übergang vom Kindergarten zur Grundschule)*, die indirekt zum Gelingen meiner Arbeit beigetragen haben.

Mein besonderer Dank gilt Prof. Dr. Hanns Petillon, der mich zum Schreiben einer Dissertation überhaupt erst ermutigt hat.

Abschließend möchte ich meiner Frau und meinen Eltern für ihr Vertrauen in mich danken und dafür, dass sie immer für mich da waren bzw. sind – und ganz besonders auch meinem Sohn, der mir immer wieder zeigt, worauf es im Leben wirklich ankommt.

Danke!

Zusammenfassung

Die bisherige Forschung zur Diagnosekompetenz hat trotz langjähriger Forschungstraditionen für die Praxis kaum relevante Befunde erbracht. Dies wurde in der vorliegenden Arbeit als Anlass für den Vorschlag genommen, diagnostische Kompetenz als reine Urteilsgenauigkeit, gemäß der Komponentenanalyse nach Cronbach (1955), zu erweitern. Mit der Einführung der Beobachtungskompetenz als einer weiteren Facette diagnostischer Kompetenz knüpft die vorliegende Studie zudem an der Forschung zu formativem Assessment an.

Die zentralen Fragen der Arbeit beziehen sich darauf, wie die beiden Facetten der diagnostischen Kompetenz Urteilsgenauigkeit und Beobachtungskompetenz von pädagogischen Fachkräften im Elementarbereich ausgeprägt sind, ob diese Facetten zusammenhängen und wie sie durch bereichsspezifische Fortbildungen und explizit durch verschiedene Fortbildungsansätze (Lehrgangs- vs. Lernwegorientierung) verändert werden können. Untersucht wurden insgesamt 103 Erzieherinnen aus 21 Einrichtungen mit eigens für die Arbeit entwickelten Instrumenten (Fragebögen, Filmvignetten).

Zur Beantwortung der ersten Forschungsfrage, wie die bereichsspezifische diagnostische Kompetenz von Erzieherinnen hinsichtlich der Facetten Urteilsgenauigkeit und Beobachtungskompetenz ausgeprägt ist, zeigt sich, dass Erzieherinnen das Niveau ihrer Gruppe im Vergleich der Urteilsgenauigkeits-Komponenten am besten einschätzen können (Niveauelemente), mit Tendenz zur Überschätzung. Auch die Wahrnehmung von Unterschieden zwischen den Kindern ihrer Gruppe gelingt ihnen gut (Personenelemente), dabei neigen sie eher zu einer Überschätzung der Heterogenität. Wesentlich schwerer fällt es ihnen, spezifische Aufgabenschwierigkeiten im mathematischen Bereich korrekt einzuschätzen (Merkmalskomponente) und Unterschiede zwischen den Kindern im Hinblick auf einzelne Aufgaben (Wechselwirkungskomponente) zu beurteilen. Die Erzieherinnen sind jedoch sehr gut in der Lage, die Kinder entsprechend ihrer Leistungen in eine Rangreihenfolge zu bringen, wie es die Korrelationskomponente der Personenelemente zeigt. Die Beobachtungskompetenz von Erzieherinnen zeigt sich in der Anzahl erkannter Fähigkeiten von Kindern während verschiedener Tätigkeiten mit mathematischem Bezug. Dabei wurde der Schwierigkeitsgrad der Beobachtung mittels Expertinneneinschätzung berücksichtigt. Die Erzieherinnen erkennen durchschnittlich die Hälfte der mathematischen Fähigkeiten von Kindern in verschiedenen komple-

nen Alltagssituationen bei mathematischen Tätigkeiten. Die zweite Forschungsfrage, ob es einen Zusammenhang zwischen der Urteilsgenauigkeit und der Beobachtungskompetenz gibt, kann klar beantwortet werden. Erwartungswidrig zeigt die vorliegende Untersuchung, dass die Facetten Urteilsgenauigkeit und Beobachtungskompetenz voneinander unabhängig sind. Damit konnte die Annahme, dass auf der einen Seite eine genaue summative Beurteilung auf einer Vielzahl kompetenter Einzelbeobachtungen beruht, auf der anderen Seite aber häufige gezielte Einzelbeobachtungen in ihrer Summe zu einem genaueren Gesamturteil führen, nicht bestätigt werden. Die dritte Forschungsfrage, ob eine bereichsspezifische Fortbildung von Erzieherinnen zur Verbesserung ihrer Urteilsgenauigkeit führt, muss verneint werden. Dagegen beeinflusst der Faktor Zeit (neun Monate zwischen Prä- und Posttest) die Fähigkeit der Erzieherinnen, Unterschiede zwischen den Kindern signifikant besser wahrzunehmen (Personenkomponente) und spezifische Aufgabenschwierigkeiten signifikant besser einzuschätzen (Merkmalskomponente). Damit verbessert sich auch die globale Urteilsgenauigkeit über die Zeit signifikant. Auf die Einschätzung des Leistungsniveaus der Gruppe (Niveauelemente) sowie von Unterschieden zwischen den Kindern für einzelne Items (Wechselwirkungskomponente) hat jedoch auch der Faktor Zeit keinen signifikanten Einfluss.

Die vierte Forschungsfrage, ob die Unterschiede in der Veränderung der Urteilsgenauigkeit von Erzieherinnen davon abhängig sind, nach welchem Ansatz sie Kinder bereichsspezifisch fördern, lässt sich zusammenfassend verneinen. Die Art und Weise der bereichsspezifischen Förderung spielt keine Rolle für die Variabilität der Urteilsgenauigkeit von Erzieherinnen. Auch die Unterschiedlichkeit in den einzelnen Urteilsgenauigkeitskomponenten lässt sich nicht auf die Art der Fortbildung zurückführen.

Die fünfte Forschungsfrage, ob sich die Beobachtungskompetenz von Erzieherinnen durch eine bereichsspezifische Fortbildung beeinflussen lässt, kann aufgrund einer signifikanten Zunahme der Beobachtungskompetenz nach nur zwei Fortbildungstagen bejaht werden. Damit wird die These gestützt, dass eine Fortbildung zur bereichsspezifischen Förderung für eine gezielte und damit fokussiertere Wahrnehmung früher bereichsspezifischer Kompetenzen bedeutsam ist, selbst wenn sie nicht explizit auf diagnostische Kompetenz, sondern allgemein auf Förderung ausgerichtet ist

Die sechste Forschungsfrage zielt auf Veränderungen der Beobachtungskompetenz von Erzieherinnen in Abhängigkeit von Fortbildungen in verschiedenen bereichsspezifischen Förderansätzen ab. Für eine Zunahme der Beobachtungskompetenz, so zeigt sich,

spielt der bereichsspezifische Ansatz, welcher der Fortbildung zugrunde gelegt ist, keine Rolle.

Die vorliegende Studie ist die erste Arbeit, die sowohl Beobachtungskompetenz (Forschung zu formativem Assessment) als auch Urteilsgenauigkeit (Accuracy-Forschung) anhand derselben Stichprobe untersucht. Trotz des im Ergebnis nicht nachgewiesenen Zusammenhangs zwischen Urteilsgenauigkeit und Beobachtungskompetenz stellt die Studie für die zukünftige Forschung bedeutsame Hinweise für die weitere Untersuchung der Facetten diagnostischer Kompetenz bereit.

Im Hinblick auf den frühpädagogischen Kontext zeigt die vorliegende Arbeit zum einen, dass ein ökologisch valides Treatment (Fortbildung) in einem Umfang von zwei Tagen implizit zu einer Verbesserung diagnostischer Kompetenzen führen kann. Zum anderen wird hier erstmalig im deutschsprachigen Raum die Messung der Urteilsgenauigkeit unter vollumfänglicher Erfüllung eines auf Hoge und Coladarci (1989) beruhenden Anforderungsprofils durchgeführt. Ein weiteres Alleinstellungsmerkmal dieser Studie ist die erstmals mit einer Mindestanzahl von vier Einschätzungen pro Erzieherin unter Berücksichtigung der Kontaktzeiten zwischen Erzieherinnen und Bezugskindern erfolgte Erhebung der Urteilsgenauigkeit von Erzieherinnen.

Inhaltsverzeichnis

Abkürzungsverzeichnis	X
Tabellenverzeichnis	XI
Abbildungsverzeichnis	XII
1. Einleitung	1
2. Diagnostische Kompetenzen von Erzieherinnen	5
2.1 Grundlegendes Verständnis von Diagnostik	5
2.2 Formelle, informelle und semiformelle Diagnostik	7
2.3 Status- und Prozessdiagnostik – summatives und formatives Assessment	8
2.4 Diagnostische Kompetenz als Teil diagnostischer Expertise	9
2.5 Facetten diagnostischer Kompetenz	10
2.6 Theoretische Einbettung der diagnostischen Kompetenz	11
2.6.1 Sozialkonstruktivistische Theorie	12
2.6.2 Modell aus der Lehr-Lernforschung	14
2.6.3 Frühpädagogisches Kompetenzmodell	16
2.6.4 Frühpädagogisches Angebots-Nutzungs-Modell	17
2.6.5 Arbeitsmodell diagnostische Kompetenz	20
3. Urteilsgenauigkeit	27
3.1 Modellierung der Urteilsgenauigkeit	27
3.1.1 Linsenmodell zur Urteilsgenauigkeit	27
3.1.2 Komponentenanalyse nach Cronbach (1955)	29
3.1.3 Komponentenanalyse nach Schrader (1989)	30
3.1.4 Erweiterung nach McElvany et al. (2009)	32
3.2 Berechnung der Urteilsgenauigkeit	33
3.2.1 Urteilsgenauigkeitsmaße nach Cronbach	33
3.2.2 Urteilsgenauigkeitsmaße nach Schrader	36
3.3 Methodische Merkmale zur Messung von Urteilsgenauigkeit	37
3.3.1 Beurteilung: <i>direkt</i> vs. <i>indirekt</i>	38
3.3.2 Urteilspezifität: <i>spezifisch</i> vs. <i>unspezifisch</i>	40
3.3.3 Bezugsnormorientierung: <i>sozial (normbezogen)</i> vs. <i>kriterial (peer-unabhängig)</i>	41
3.3.4 Art der Messung: <i>relativ</i> vs. <i>absolut</i>	41
3.3.5 Analyseeinheit: <i>gepoolt</i> vs. <i>hierarchisch</i>	42
3.3.6 Zusammenfassung	43
3.4 Forschungsstand zur diagnostischen Kompetenz	43
3.4.1 ... als Akkuratheit	45
3.4.2 ... mit expliziter Berücksichtigung eines (bildungswissenschaftlichen) Kompetenzbegriffs	49

3.4.3 ... unter Berücksichtigung weiterer Merkmale.....	50
3.5 Veränderung der Urteilsgenauigkeit.....	55
3.6 Wirkung diagnostischer Kompetenz.....	56
3.7 Zusammenfassung	57
3.8 Folgerungen	60
4. Beobachtungskompetenz	66
4.1 Relevanz von Beobachtung in der Frühpädagogik.....	67
4.2 Begriffsklärung	72
4.3 Beobachtungsfehler	75
4.4 Gütekriterien	78
4.5 Beobachtungskompetenz im Kontext formativen Assessments	80
4.5.1 Relevanz für die Diagnosekompetenzforschung.....	80
4.5.2 Arten formativen Assessments	82
4.5.3 Wirkung	84
4.6 Modelle	85
4.6.1 Erweitertes Linsenmodell zur Beobachtungskompetenz	85
4.6.2 Frühpädagogische Modelle und Modelle der Lehr-Lernforschung	89
4.7 Veränderung der Beobachtungskompetenz	92
4.8 Zusammenfassung	95
4.9 Folgerungen	96
5. Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz	98
6. Forschungsfragen und Hypothesen	102
7. Methodik und Durchführung	106
7.1 Projektrahmen.....	106
7.2 Design	106
7.2.1 Aufbau der Studie	106
7.2.2 Treatment – bereichsspezifischer Fortbildungsansatz	107
7.3 Stichprobe	111
7.4 Variablen und Erhebungsverfahren	114
7.4.1 Urteilsgenauigkeit von Erzieherinnen.....	114
7.4.2 Numerisch-mathematische Fähigkeiten von Vorschulkindern	117
7.4.3 Beobachtungskompetenz von Erzieherinnen	119
7.5 Durchführung.....	128
7.5.1 ... der Stichprobenrekrutierung	128
7.5.2 ... der Fortbildungen.....	129
7.5.3 ... der mathematischen Frühförderung	130
7.5.4 Erhebungszeitpunkte und Begründung	131
7.6 Auswertung und poweranalytische Überlegungen	131
8. Ergebnisse.....	135
8.1 Deskriptive Ergebnisse	135

8.1.1 Urteilsgenauigkeit	135
8.1.2 Beobachtungskompetenz	138
8.2 Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz	140
8.3 Veränderung der Urteilsgenauigkeit	141
8.3.1 Niveauekomponente (Elevation, E)	142
8.3.2 Personenkomponente (Differential Elevation, DE)	144
8.3.3 Merkmalskomponente (Stereotype Accuracy, SA)	146
8.3.4 Wechselwirkungskomponente (Differential Accuracy, DA)	148
8.3.5 Globale Urteilsgenauigkeit (Global Accuracy, ACC)	151
8.3.6 Zusammenfassung	153
8.4 Veränderung der Beobachtungskompetenz durch Fortbildungen	154
9. Zusammenfassung	157
10. Diskussion	160
10.1 Inhaltliche Diskussion der Ergebnisse	160
10.1.1 Einordnung in den Forschungskontext	161
10.1.2 Was bedeutet die Unabhängigkeit der Facetten <i>Urteilsgenauigkeit</i> und <i>Beobachtungskompetenz</i> ?	168
10.1.3 Welche diagnostischen Kompetenzen benötigen Erzieherinnen?	169
10.1.4 Veränderung von <i>Urteilsgenauigkeit</i> und <i>Beobachtungskompetenz</i>	171
10.1.5 Gehören Handlungsentscheidungen zur diagnostischen Kompetenz?	174
10.2 Methodische Diskussion	176
10.2.1 Erfassung von Urteilsgenauigkeit und Beobachtungskompetenz	176
10.2.2 Design	177
10.2.3 Stichprobe	179
10.2.4 Limitationen	180
10.3 Implikationen für Praxis und Forschung	181
10.3.1 Wissenschaftlicher Ertrag der Arbeit	181
10.3.2 Ausblick auf weitere Forschungsansätze	182
10.4 Fazit	184
Literaturverzeichnis	186
Anlagen	213
Curriculum Vitae	234
Eidesstattliche Erklärung	236

Abkürzungsverzeichnis

BEURteiler	Befragung zur Erfassung der Urteilsgenauigkeit von Erzieherinnen
DA	Differential Accuracy (Wechselwirkungskomponente)
DE	Differential Elevation (Personenkomponente)
E	Elevation (Niveauelemente)
FaBi	Projekt „Förderung anschlussfähiger Bildungsprozesse“
FiBB	Filmbasierte Befragung zur Erfassung der Beobachtungskompetenz von Erzieherinnen für die Bereiche Mathematik und Schriftsprache
J	“(…) a Judge (J) ‘predicts’ how another person (O) will respond” (Cronbach, 1955, S. 177)
JFMK	Jugend- und Familienministerkonferenz
JMK	Jugendministerkonferenz
Kita(s)	Kindertagesstätte(n)
KMK	Kultusministerkonferenz
KW	Kalenderwoche
LG	Lehrgangsbasierte Fortbildung
LW	Lernwegsbasierte Fortbildung
M-Vignette	Filmvignette mit mathematischem Schwerpunkt
NeDiKo	Netzwerk zur diagnostischen Kompetenz von Lehrkräften
O	“(…) a Judge (J) ‘predicts’ how another person (O) will respond” (Cronbach, 1955, S. 177)
SA	Stereotype Accuracy (Merkmalskomponente)
x_{oi}	“O’s self-description x_{oi} on a set of items” (Cronbach, 1955, S. 178)
y_{oij}	“set of predictions y_{oij} by J” (Cronbach, 1955, S. 178)

Tabellenverzeichnis

Tabelle 1: Gegenüberstellung der Facetten diagnostischer Kompetenz	11
Tabelle 2: Überblick über die Komponenten diagnostischer Kompetenz nach Schrader (1989) und ihre Bedeutung (Karst, 2017a, S. 22)	32
Tabelle 3: Systematisierung relevanter Urteilsbereiche und Urteilebenen nach McElvany et al. (2009, S. 226)	32
Tabelle 4: Ergebnisse der mathematischen Umformung der Gleichungen für Personen-, Merkmals- und Wechselwirkungskomponente	35
Tabelle 5: Personenbezogene Komponenten nach Schrader (1989)	36
Tabelle 6: Urteilsmaße nach Hoge und Coladarci (1989, S. 300–306)	40
Tabelle 7: Nutzung von Beobachtungsverfahren in Prozent nach Viernickel et al. (2013): mind. für einige wenige Kinder (S. 88), für alle Kinder (S. 89)	68
Tabelle 8: Unterschiede zwischen Gruppen von Lehrpersonen nach Smit und Engeli (2017, S. 296)	74
Tabelle 9: Berufserfahrung und Alter der Erzieherinnen zu t_1	112
Tabelle 10: Verteilung der Erzieherinnen und Bezugsgruppen auf Treatments und Kontrollgruppe	113
Tabelle 11: Verteilung der Bezugsgruppen nach Größe und Treatments	114
Tabelle 12: Split-Half-Reliabilitäten einzelner Komponenten der Erfassung der Urteilsgenauigkeit zu t_1	116
Tabelle 13: Inhaltsbereiche von <i>zahlenstark</i> gemäß Moser und Berweger (2007a)	118
Tabelle 14: Bezüge der FiBB-Vignetten zur Subskala Mathematik der KES-R-E (Tietze & Roßbach, 2010)	121
Tabelle 15: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 3 von 8 (2. Mathematikvignette)	122
Tabelle 16: Auszug aus dem Kategoriensystem Vignette 2 mit Multiplikator (Expertinnengewichtung)	125
Tabelle 17: Intercoderreliabilitäten des qualitativen Anteils der Beobachtungskompetenzerfassung ($N(t_1) = 51$, $N(t_3 \text{ mit Kontrollgruppe}) = 103$)	127
Tabelle 18: Deskriptive Werte zu den Urteilsgenauigkeitskomponenten zu t_1 , Fortbildungsgruppe gesamt und aufgeteilt nach Treatments, Kontrollgruppe und Gesamtgruppe	136
Tabelle 19: t-Test für unabhängige Stichproben (Fortbildungs- und Kontrollgruppe) zu t_1	138
Tabelle 20: Deskriptive Werte zur Beobachtungskompetenz zu t_1	139
Tabelle 21: Korrelationsmatrix zu Urteilsgenauigkeit und Beobachtungskompetenz	141
Tabelle 22: Deskriptive Statistik der Niveauebene (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe	142
Tabelle 23: Deskriptive Statistik der Personenkomponente (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe	144
Tabelle 24: Deskriptive Statistik der Merkmalskomponente (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe	147
Tabelle 25: Deskriptive Statistik der Wechselwirkungskomponente (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe	149
Tabelle 26: Deskriptive Statistik der globalen Urteilsgenauigkeit (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe	151
Tabelle 27: Deskriptive Statistik der Beobachtungskompetenz (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe	155
Tabelle 28: Übersicht der Erfassung von Rahmendaten (Erzieherinnenbefragungen), Domäne Mathematik	214
Tabelle 29: BEUrteilEr: Übersicht der Bereiche und Items aus <i>zahlenstark</i>	215
Tabelle 30: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 1 von 8 (1. Mathematikvignette)	218
Tabelle 31: Kategoriensystem Vignette 1 mit Multiplikator (Expertinnengewichtung)	220
Tabelle 32: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 3 von 8 (2. Mathematikvignette)	222
Tabelle 33: Kategoriensystem Vignette 3 mit Multiplikator (Expertinnengewichtung)	224
Tabelle 34: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 5 von 8 (3. Mathematikvignette)	226
Tabelle 35: Kategoriensystem Vignette 5 mit Multiplikator (Expertinnengewichtung)	228
Tabelle 36: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 7 von 8 (4. Mathematikvignette)	231
Tabelle 37: Kategoriensystem Vignette 7 mit Multiplikator (Expertinnengewichtung)	233

Abbildungsverzeichnis

Abbildung 1: Kompetenzmodell mit Spezifikationen für das Professionswissen von Baumert und Kunter (2011, S. 32)	15
Abbildung 2: Allgemeines Kompetenzmodell nach Fröhlich-Gildhoff et al. (2011, S. 17)	17
Abbildung 3: Frühpädagogisches Angebots-Nutzungs-Modell der Wirkfaktoren akademischer Leistungen in Anlehnung an Helmke und Schrader (2020)	19
Abbildung 4: NeDiKo-Prozessmodell pädagogischer Diagnostik nach Herppich et al. (2017, S. 82)	21
Abbildung 5: NeDiKo-Arbeitsmodell diagnostischer Kompetenz nach Herppich et al. (2017, S. 81)	22
Abbildung 6: Linsenmodell nach Förster und Böhmer (2017, S. 47)	28
Abbildung 7: Theoretische Modellierung diagnostischer Kompetenz nach Schrader (1989, Darstellung in Anlehnung an Karst, 2012, S. 31)	31
Abbildung 8: Strukturmodell zu mathematikbezogenen Kompetenzen pädagogischer Fachkräfte von Hepberger et al. (2017, S. 243)	75
Abbildung 9: Abnahme der Beobachtungsfehler in Abhängigkeit vom Prozessstypus	77
Abbildung 10: Erweitertes Linsenmodell diagnostischer Kompetenz in Anlehnung an Förster und Böhmer (2017), vgl. Abbildung 6	87
Abbildung 11: Modell der kausalen Wirkkette professioneller Kompetenz pädagogischer Fachkräfte nach Seemann et al. (2018, S. 1225)	91
Abbildung 12: Model of professional knowledge and skills for early mathematics education (Gasteiger & Benz, 2018, S. 9)	91
Abbildung 13: Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz	99
Abbildung 14: Forschungsfragen im Überblick	102
Abbildung 15: Evaluationsdesign	107
Abbildung 16: Stichprobenaufteilung	111
Abbildung 17: Beispiel 3 (Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 5)	120
Abbildung 18: Antwort der Erzieherin 003, Vignette 1, Erhebungszeitpunkt 2	122
Abbildung 19: Niveauelemente, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)	143
Abbildung 20: Personenkomponente, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)	145
Abbildung 21: Merkmalskomponente, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)	147
Abbildung 22: Wechselwirkungskomponente, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)	150
Abbildung 23: Globale Urteilsgenauigkeit, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)	152
Abbildung 24: Beobachtungskompetenz, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)	155
Abbildung 25: FiBB-Evaluationsbogen, Deckblatt von Teil 2 mit allgemeinen Hinweisen	216
Abbildung 26: Beispiel 1 von 8 (1. Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 1)	217
Abbildung 27: Beispiel 3 von 8 (2. Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 5)	221
Abbildung 28: Beispiel 5 von 8 (3. Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 9)	225
Abbildung 29: Beispiel 7 von 8 (4. Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 13)	230

1. Einleitung

Zur Gestaltung adaptiver Fördermaßnahmen benötigt eine Lehrperson diagnostische Kompetenz (Praetorius, Karst & Lipowsky, 2012; Schrader & Praetorius, 2018). Dies gilt sowohl für Lehrkräfte in der Schule als auch für das frühpädagogische Fachpersonal¹ in Kitas (vgl. z.B. Bruns, 2014; Dollinger, 2013; Kammermeyer, 2000), da sich die Grundaufgaben von Lehrern an der Schule und von Fachkräften in Kitas nicht wesentlich unterscheiden (Benz et al., 2017). Letztlich geht es darum, Kinder gemäß ihrem augenblicklichen Kompetenzniveau entsprechend ihrer individuellen Lernmöglichkeiten zum fachlich tragfähigen Weiterlernen anzuregen.

Einen traditionellen Zugang zur diagnostischen Kompetenz von Lehrpersonen stellen die Forschungen zur Urteilstgenauigkeit oder Urteilsakkuratheit dar (Überblick vgl. Praetorius & Südkamp, 2017). Der praktische Nutzen dieser Forschung für den Alltag von Bildungseinrichtungen erscheint jedoch gering (vgl. Kap. 2). Zudem wird dieses begrenzte Verständnis von diagnostischer Kompetenz als Genauigkeit von Urteilen von einigen Autoren kritisch hinterfragt (Abs, 2006; Bruder, Klug, Hertel & Schmitz, 2010; Helmke, Hosenfeld & Schrader, 2004; vgl. Praetorius et al., 2012). Aufgrund dieser limitierten Sicht auf diagnostische Kompetenz führt Helmke (2017) das Konzept „Expertise“ ein, „um es von diagnostischer Kompetenz im engeren Sinne abzuheben, die sich meistens lediglich auf die zutreffende Informiertheit bezieht, also auf die Urteilstgenauigkeit (Accuracy) oder Veridikalität (Übereinstimmung von Urteil und Realität)“ (Helmke, 2017, S. 119). Mit diagnostischer Expertise schafft er ein umfassendes Konzept, das sowohl methodisches, prozedurales Wissen als auch konzeptuelles Wissen beinhaltet (Helmke, 2017, S. 119).

Auch in der vorliegenden Arbeit wird die Begrenzung diagnostischer Kompetenz allein auf die Urteilstgenauigkeit als zu eng angesehen. Aus diesem Grund erfolgt eine Erweiterung durch die Einführung der Beobachtungskompetenz als weitere Facette diagnostischer Kompetenz. Dieser erweiterte diagnostische Kompetenzbegriff lässt sich letztlich unter Helmkes Konzept der diagnostischen Expertise verorten.

¹ Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Aufgrund des mehrheitlich weiblichen Personals in vorschulischen Einrichtungen wird der weiblichen Sprachform Vorzug gegeben. Die Begriffe „frühpädagogische Fachkräfte“, „frühpädagogisches Fachpersonal“ und „Erzieherinnen“ werden im Folgenden synonym gebraucht und bezeichnen alle im Elementarbereich tätigen Fachpersonen.

Ziel ist die Untersuchung diagnostischer Kompetenzen unter besonderer Berücksichtigung von formativen (Beobachtungskompetenz) und summativen Anteilen (Urteilsakkuratheit). Zentral ist die Klärung des Zusammenhangs dieser Kompetenzen, weil die übergeordneten Bereiche der Accuracy-Forschung und der Forschung zu formativem Assessment noch weitgehend unverbunden nebeneinander stehen (Südkamp & Praetorius, 2017). Auch erste allgemeinere Modellentwürfe (Herppich et al., 2017) lassen sich nur separat auf beide Forschungszweige anwenden. Daher setzt die vorliegende Untersuchung mit der Entwicklung und Überprüfung eines Prozessmodells zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz an diesem Forschungsdesiderat an. Zudem erfolgt die Untersuchung im Kontext fröhpädagogischer bereichsspezifischer Förderung und erweitert damit die im deutschsprachigen Raum sehr übersichtliche Befundlage zur Accuracy-Forschung und Forschung zu formativem Assessment im vorschulischen Bereich.

Die zentralen Fragestellungen lauten: Wie ist die bereichsspezifische diagnostische Kompetenz von Erzieherinnen hinsichtlich der Facetten Urteilsgenauigkeit und Beobachtungskompetenz ausgeprägt, gibt es einen Zusammenhang zwischen den Facetten und wie verändern sich diese durch bereichsspezifische Fortbildungen?

Aufbau der Arbeit

Grundlegend für die vorliegende Arbeit ist zunächst die Darstellung des allgemeinen pädagogisch(-psychologisch)en Verständnisses von Diagnostik (Kap. 2.1). Die Unterscheidungen in formelle, informelle und semiformelle Diagnostik (Kap. 2.2) sowie in summatives und formatives Assessment (Kap. 2.3) sind für die Formulierung der Arbeitsdefinitionen der Facetten diagnostischer Kompetenz von Erzieherinnen – Urteilsgenauigkeit und Beobachtungskompetenz – relevant, welche beide der diagnostischen Expertise zugeordnet werden können (Kap. 2.4). Eine theoretische Einbettung der diagnostischen Kompetenz (2.6) in Modelle aus der Lehr-Lernforschung und aus der Fröhpädagogik erfolgt vor der Darstellung des sozialkonstruktivistischen Hintergrunds. Eine Gegenüberstellung der zunächst als Facetten diagnostischer Kompetenz angenommenen Urteilsgenauigkeit und Beobachtungskompetenz wird in Kapitel 2.5 aufgezeigt, bevor die Kompetenzen im Einzelnen dargestellt werden.

In Kapitel 3 „Urteilsgenauigkeit“ wird in 3.1 zunächst das Linsenmodell zur Urteilsgenauigkeit von Förster und Böhmer (2017, S. 47) eingeführt. Anschließend werden

sowohl die international bedeutsame Komponentenanalyse nach Cronbach (1955) als auch deren Modifikation nach Schrader (1989) vorgestellt. Beide werden als Voraussetzung für das Verständnis der Erweiterung des Konstrukts „Diagnostische Kompetenz“ in der vorliegenden Studie angesehen und waren zugleich Ausgangspunkt der Überlegungen zur Anlage der vorliegenden Untersuchung. Die Berechnung der Urteilsgenauigkeitskomponenten (3.2) und methodische Merkmale zur Messung von Urteilsgenauigkeit (3.3) folgen zur Vertiefung der für die vorliegende Studie bedeutsamen Komponentenanalyse nach Cronbach (1955). Nach Darstellung des Forschungsstandes (3.4) der nationalen und internationalen Judgment-Accuracy-Forschung unter Einbezug relevanter Studien sowohl aus dem vorschulischen als auch schulischen Bereich folgen Befunde zur Veränderung der Urteilsgenauigkeit (3.5), zur Wirkung von Urteilsgenauigkeit (3.6) und eine Zusammenfassung (3.7). Die Folgerungen (3.8) für die vorliegende Studie sind die Grundlage für die Einführung der Facette Beobachtungskompetenz.

Zur Klärung des Begriffs Beobachtungskompetenz (4.2) wird zunächst die Relevanz von Beobachtung für den frühpädagogischen Bereich erörtert (4.1). Die Beobachtungskompetenz ist einer Vielzahl von Limitationen unterworfen, welche mit den Schwerpunkten Beobachtungsfehler in Kapitel 4.3 und mit dem Schwerpunkt alternative Gütekriterien in Kapitel 4.4 ausgeführt werden.

Anschließend wird die Beobachtung(skompetenz) in der Theorie formativen Assessments verortet (4.5). Diese Verortung in das Linsenmodell zur Urteilsgenauigkeit (3.1.1) führt zu einem erweiterten Linsenmodell zur Beobachtungskompetenz (4.6.1), das bereits die theoretische Grundlage zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz bildet. Die Einordnung der Beobachtungskompetenz in frühpädagogische Modelle und Modelle der Lehr-Lernforschung schließt das Kapitel 4.6. Die Veränderung der Beobachtungskompetenz wird in Kapitel 4.7 beleuchtet. Eine Zusammenfassung des vierten Kapitels folgt in 4.8. Die Folgerungen (4.9) für die vorliegende Studie bilden die Basis für das im nächsten Kapitel folgende Prozessmodell.

Im 5. Kapitel wird der angenommene Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz im „Prozessmodell des bereichsspezifischen Förderprozesses unter Berücksichtigung des Zusammenhangs von Beobachtungskompetenz und Urteilsgenauigkeit“ dargestellt.

In Kapitel 6 werden Forschungsfragen und Hypothesen sowohl zur Urteilsgenauigkeit und Beobachtungskompetenz formuliert als auch zum Zusammenhang beider Facetten diagnostischer Kompetenz.

In Kapitel 7 wird zunächst der Projektrahmen (7.1) erörtert, der sich auf das Design der Studie (7.2) auswirkt. Es folgen eine Stichprobenbeschreibung (7.3), die Schilderung der Variablen und Erhebungsverfahren (7.4) und eine Beschreibung der Durchführung (7.5). Die Darstellung der Auswertungsverfahren und diesbezügliche poweranalytische Überlegungen (7.6) schließen das Kapitel.

Kapitel 8 präsentiert die Ergebnisse. Zunächst werden in Kapitel 8.1 die deskriptiven Ergebnisse zur Urteilsgenauigkeit und Beobachtungskompetenz vorgestellt. Die Ergebnisse zum Zusammenhang zwischen Urteilsgenauigkeit und Beobachtungskompetenz folgen in Kapitel 8.2. Danach werden jeweils die Veränderungen der Urteilsgenauigkeit (8.3) und der Beobachtungskompetenz (8.4) durch ökologisch valide Fortbildungen dargestellt als auch der Einfluss verschiedener bereichsspezifischer Förderansätze auf die jeweilige Veränderung von Urteilsgenauigkeit und Beobachtungskompetenz.

Kapitel 9 gibt eine Zusammenfassung. In Kapitel 10 schließt sich eine inhaltliche (10.1) und eine methodische (10.2) Diskussion an, bevor die Implikationen für Praxis und Forschung (10.3) dargelegt werden. Ein Gesamtfazit (Kap. 10.4) schließt die Arbeit.

2. Diagnostische Kompetenzen von Erzieherinnen

Gegenstand dieses Kapitels ist die Klärung zentraler Begriffe, die zum Verständnis einer Definition diagnostischer Kompetenz von Erzieherinnen mit den Facetten Urteilsgenauigkeit und Beobachtungskompetenz notwendig sind.

Hierzu werden zunächst in Kapitel 2.1 grundlegende Definitionen von Diagnostik aus der Pädagogik und aus der pädagogischen Psychologie vorgestellt. Die in Kapitel 2.2 vorgenommenen Unterscheidungen in formelle, informelle und semiformelle Diagnostik sowie die in Status- und Prozessdiagnostik bzw. summatives und formatives Assessment in Kapitel 2.3 sind für die Klärung der Facetten diagnostischer Kompetenz von Erzieherinnen – Urteilsgenauigkeit und Beobachtungskompetenz – bedeutsam, weil sie sich in diesen Aspekten voneinander unterscheiden. Bezüge zu einem erweiterten Diagnosebegriff werden in Kapitel 2.4 aufgezeigt. Die Definition der Facetten diagnostischer Kompetenz und die Gegenüberstellung zentraler Aspekte folgt in Kapitel 2.5.

Die theoretische Einbettung der diagnostischen Kompetenz in Kapitel 2.6 erfolgt in einem ersten Schritt anhand der sozialkonstruktivistischen Theorie von Wygotski² (Vygotskij, 2002). Die Einordnung in bewährte Kompetenzmodelle aus der Lehr-Lernforschung und der Frühpädagogik schließen sich in einem zweiten Schritt an. Anschließend wird diagnostische Kompetenz in der Komplexität von Lehr-Lernprozessen in einem frühpädagogischen Angebots-Nutzungs-Modell verortet. Den Abschluss des Kapitels bildet ein Arbeitsmodell zur diagnostischen Kompetenz, das theoretische und empirische Ansätze aus teilweise sehr unterschiedlichen Perspektiven verbindet.

2.1 Grundlegendes Verständnis von Diagnostik

Diagnostik wird vor allem mit Medizin und Psychologie oder auch Schule, jedoch kaum mit Frühpädagogik verbunden. Die diagnostische Kompetenz von Erzieherinnen gewinnt im Zuge der Professionalisierung des frühpädagogischen Personals und der damit einhergehenden Akademisierungsdebatte zwar zunehmend an Bedeutung, es gibt jedoch

² Je nach Werk und Übersetzung unterscheiden sich die Schreibweisen, weshalb zur Vereinheitlichung im Folgenden „Wygotski“ und in den Zitationen die jeweilige Version verwendet wird.

unterschiedliche Schwerpunktsetzungen je nachdem, ob diese aus einer eher sozialpädagogischen oder grundschulpädagogischen Orientierung erfolgt.

Während Diagnostik und Leistungsbeurteilung als feste Bestandteile des Lehrerberufs und der Lehrerausbildung etabliert sind, werden im frühpädagogischen Bereich diagnostische Tätigkeiten in erster Linie unter „Beobachten und Dokumentieren“ subsumiert (Geiling & Liebers, 2014, S. 529; Schulz & Cloos, 2013; Viernickel, 2011). Der Begriff Diagnostik ist im vorschulischen Kontext eher negativ konnotiert und wird vorwiegend als schulische Aufgabe wahrgenommen. Vorschulisches Fachpersonal zeigt sich diagnostischen Verfahren gegenüber skeptisch, zum Teil sind auch massive Vorbehalte gegenüber dem Diagnostikbegriff feststellbar (Titz, Ropeter & Hasselhorn, 2018, S. 88). Diagnostik wird nach Geiling und Liebers (2014, S. 529) zuweilen mit Konzepten verknüpft, die elementarpädagogischen Auffassungen von pädagogischem Handeln als Kokonstruktion (Fröhlich-Gildhoff & Strohmmer, 2011) oder als gemeinsames längerfristiges Denken (Sylva, Melhuish, Sammons, Siraj-Blatchford & Taggart, 2004, S. 5–6; Sylva et al., 2004, S. 160) zu widersprechen scheinen. Bei näherer Betrachtung geht es letztlich in allen Fällen um die Optimierung von Lernprozessen – unabhängig davon, ob diese als Betreiben von Diagnostik oder Beobachten und Dokumentieren betitelt werden.

Häufig wird eine Unterscheidung in pädagogische und pädagogisch-psychologische Diagnostik vorgenommen:

- *Pädagogisch-psychologische Diagnostik*

In der pädagogisch-psychologischen Diagnostik geht es vornehmlich um die Beschaffung und Bewertung von Informationen, „die zu einer möglichst akkuraten Einschätzung der aktuellen Ausprägung von Personenmerkmalen (z.B. Fähigkeiten und Einstellungen der Lernenden) oder Merkmalen der Lern- und Entwicklungsumwelt (...) führen und zu einer besseren Erklärung und Prognose in pädagogisch relevanten Problemfeldern beitragen“ (Krapp & Weidenmann, 2006, S. 528). Wilhelm und Kunina-Habenicht (2015, S. 307) betonen zudem die Zielvorstellung, Hilfen zur Lösung praktischer pädagogischer, schulischer oder bildungsbezogener Probleme und Fragestellungen zu liefern.

- *Pädagogische Diagnostik*

Wer pädagogische Diagnostik betreibt, führt Tätigkeiten aus, um für Lernende begründete Entscheidungen zu treffen (Jürgens & Lissmann, 2015, S. 58). Nach Schrader umfasst die pädagogische Diagnostik „alle Bemühungen, für pädagogische Ziele und Entscheidungen relevante Informationen über Personen und ihre Merkmale zu ermitteln, aufzubereiten und verfügbar zu machen“ (Schrader, 2008, S. 169). Dabei kann sie der Zertifizierung dienen; in besonderem Maße aber dient sie der Modifikation, d.h. der erzieherischen Einflussnahme (vgl. ebd.). Dies betonen auch Ingenkamp und Lissmann (2008), die als Kriterium für das Ziel der pädagogischen Diagnostik die Optimierung individueller Lernprozesse einschließen: „Pädagogische Diagnostik umfasst alle diagnostischen Tätigkeiten, durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren“ (Ingenkamp & Lissmann, 2008, S. 13).

Festzustellen ist, dass die Beschaffung und Bewertung von Informationen in beiden Bereichen bedeutsam ist, der Zweck dieser Beschaffung und Bewertung jedoch das zentrale Unterscheidungsmerkmal ist. Erst wenn individuelles Lernen optimiert wird, handelt es sich definitionsgemäß um pädagogische Diagnostik (ebd.).

2.2 Formelle, informelle und semiformelle Diagnostik

Die Pädagogische Psychologie unterscheidet informelle von formeller Diagnostik (Bennett, 1982; Satterly, 1981; Schrader & Helmke, 2001). Während formelle Diagnosen mithilfe wissenschaftlich erprobter Methoden gezielt und systematisch erstellt werden, erfolgt die Gewinnung informeller Diagnosen dagegen eher beiläufig und unsystematisch im Rahmen des alltäglichen erzieherischen Handelns (Schrader, 2006, S. 95). Vielfach handelt es sich um implizite subjektive Urteile, Einschätzungen und Erwartungen (Schrader, ebd.; Hascher, 2008, S. 74). Zur pädagogischen Arbeit von Erzieherinnen gehören vornehmlich informelle Diagnoseleistungen. Formelle Verfahren kommen im Alltag von Erzieherinnen vergleichsweise selten zum Einsatz (Viernickel, Nentwig-Gesemann, Nicolai, Schwarz & Zenker, 2013, S. 88–89).

Für erzieherische Entscheidungen und Maßnahmen sind nach Schrader (2006, S. 99), der sich eher auf den schulischen Kontext bezieht, sowohl formelle als auch informelle

Diagnoseleistungen von Bedeutung und gewinnen zunehmend in ihrem Zusammenspiel an Beachtung. Eine nutzenorientierte Betrachtungsweise haben Cronbach und Gleser bereits 1965 statt der traditionell rein psychometrischen Sichtweise der pädagogisch-psychologischen Diagnostik, im Sinne einer Maximierung der Messgenauigkeit, empfohlen (zitiert nach Schrader, 1997, S. 663). Darauf Bezug nehmend betont Schrader (ebd.), dass auch diagnostische Informationen, die herkömmlichen psychometrischen Kriterien nicht genügen, einen hohen Nutzen haben können, wenn für eine Entscheidung keine anderen oder besseren Informationen zur Verfügung stehen.

Dieses Zusammenspiel bzw. die Tatsache, dass es im erzieherischen Alltag durchaus eine Schnittmenge diagnostischer Handlungen gibt, die weder eindeutig formellen noch informellen Diagnosen zuzuordnen sind, veranlasste Hascher (2008, 2011) zur Erweiterung der bisherigen Unterscheidung formeller und informeller Diagnostik um die semiformelle Diagnostik. Diese bezeichnet die „Gesamtheit aller diagnostischen Tätigkeiten, die nicht den Kriterien der formellen Diagnostik genügen, aber nicht nur zu impliziten Urteilen führen“ (Hascher, 2008, S. 75). „Die Bezeichnung ‚semiformell‘ trifft beispielsweise dann zu, wenn Beobachtungen zwar gezielt, aber nicht mit erprobten Methoden durchgeführt werden; wenn erprobte Methoden nicht gezielt und unsystematisch eingesetzt werden; wenn intuitive Beobachtungen festgehalten werden (...) und explizit in Bewertungen einfließen usw.“ (ebd.). Im frühpädagogischen Bereich ist die semiformelle Diagnostik besonders häufig vertreten.

2.3 Status- und Prozessdiagnostik – summatives und formatives Assessment

Diagnostik lässt sich unterscheiden in die abschließende Feststellung (Statusdiagnostik) von Lernergebnissen, die als summatives Assessment bezeichnet wird, und in die prozessbegleitende Erfassung (Prozessdiagnostik), die als formatives Assessment oder formative Evaluation bezeichnet wird (Schrader, 2008, S. 169); diese Unterscheidung geht auf Scriven (1967, S. 43) zurück.

Mittels Statusdiagnostik werden relativ stabile Personenmerkmale gemessen und mittels Prozessdiagnostik Verläufe und Veränderungen durch wiederholte Messung von sich verändernden Zuständen und modifizierbaren Merkmalen erfasst (Schrader, 2008, S. 169). Diese Unterscheidung wird über den Zweck der Diagnostik definiert (Maier, 2010, S. 294; Winter, 2010, S. 494), wie es die folgenden Begriffspaare verdeutlichen:

Beurteilung des Lernens vs. Beurteilung für (z.B. zielorientiertes) Lernen (*assessment of learning vs. assessment for learning*; Birenbaum et al., 2006) bzw. summative vs. formative Beurteilung (*summative classroom assessment vs. formative classroom assessment*; Black & Wiliam, 2009; Pellegrino, Chudowsky & Glaser, 2001; Stiggins & Chappuis, 2012). Der jeweils erste Begriff bezeichnet die tradierte Annahme, Leistungen der Kinder objektiv messen und für Entscheidungsprozesse, wie z.B. schulische Selektionsentscheidungen, nutzen zu können (Maier, 2010, S. 294). Der jeweils zweite Begriff bzw. die formative Beurteilung hingegen subsumiert Intentionen von Reformansätzen: Leistungsmessung wird als Teil des Lernprozesses betrachtet, der auf die vielperspektivische Erfassung von Leistungen zielt und der Optimierung von Lernprozessen dient (ebd.). Formative Beurteilung zeigt sich unter anderem in der gezielten, auf Kriterien beruhenden Beobachtung und Dokumentation der Beurteilung dieser Beobachtung (Smit & Engeli, 2017, S. 286) und in der Nutzung der gewonnenen Informationen zur individuellen Rückmeldung. Der Zeitpunkt der Interpretation gewonnener Erkenntnisse kann auch zur Unterscheidung von summativem und formativem Assessment herangezogen werden, welcher innerhalb des Lernprozesses liegen kann oder danach, wenn die Informationen für eine abschließende Bewertung genutzt werden (Hattie, 2003, S. 5). Damit wäre die formale Unterscheidung in summative oder formative Leistungsdiagnosen erst dann möglich, wenn klar ist, für welche Entscheidung eine Diagnose genutzt wurde (Harlen, 2005).

Im deutschsprachigen Raum wird statt von formativem Assessment oder formativer Beurteilung auch von ‚förderorientierter Beurteilung‘ gesprochen (z.B. Bürgermeister, 2014; Smit, 2009). In Kapitel 4.5 wird das formative Assessment ausführlich thematisiert.

2.4 Diagnostische Kompetenz als Teil diagnostischer Expertise

Die diagnostische Expertise ist ein von Helmke (2017, S. 119–141) eingeführtes Konzept, das er zur Erweiterung des als Urteilsgenauigkeit oder Veridikalität (Übereinstimmung von Urteil und Realität) inhaltlich sehr eng gefassten Begriffs der diagnostischen Kompetenz verwendet. Sein Expertisebegriff umfasst sowohl methodisches und prozedurales Wissen als auch konzeptuelles Wissen (Helmke, 2017, S. 119). Damit ist sein Konzept wohl geeigneter, den Herausforderungen des pädagogischen Alltags im Hinblick auf diagnostisches Handeln in adaptiven Lehr-Lernsituationen gerecht zu wer-

den als sehr eng gefasste Konstrukte (Urteilsgenauigkeit). Die enge Verzahnung diagnostischer und pädagogischer Anteile im unterrichtlichen Kontext verdeutlicht Praetorius, Hetmanek, Herppich und Ufer (2017, S. 99) zufolge, dass eine konzeptuelle Trennung diagnostischer Anteile adaptiven Unterrichts künstlich ist. Dennoch stellen die Autorinnen heraus, dass eine theoretisch plausible konzeptuelle Trennung kognitiver (diagnostischer) Prozesse von pädagogischen (handlungsentscheidenden) Prozessen wichtig ist, weil eine undifferenzierte, allumfassende adaptive Unterrichtskompetenz weder beschreibbar noch vermitteln- oder erlernbar wäre (ebd.). Dass es auch Handlungsentscheidungen zur Unterrichtsgestaltung als Teil eines diagnostischen Prozesses geben könnte, wird in diesen Überlegungen nicht berücksichtigt. Aus diesem Grund ist die diagnostische Expertise nach Helmke (2017), insbesondere weil sie das prozedurale Wissen als Bestandteil des breit angelegten Konzepts beinhaltet, für die vorliegende Arbeit von besonderer Bedeutung.

2.5 Facetten diagnostischer Kompetenz

Die bisherigen Ausführungen zeigen zwei Facetten diagnostischer Kompetenzen von Lehrenden auf:

Zum einen die der Fähigkeit, im Rahmen einer rückblickenden Statusdiagnostik mittels einer summativen, informellen Diagnostik den Entwicklungsstand von Kindern einzuschätzen. Diese Facette wird im Zusammenhang mit Urteilsgenauigkeit untersucht, die vor allem in der psychologischen Grundlagenforschung zur diagnostischen Kompetenz erforscht wird.

Zum anderen die des Vermögens, im laufenden (Lern-)Prozess mittels einer formativen, semiformalen Diagnostik – durch gezielte Beobachtung kindlicher Aktivitäten – Fähigkeiten von Kindern zu erkennen. Diese zweite Facette wird im Zusammenhang mit Beobachtungskompetenz untersucht, die auf Erkenntnissen aus der anwendungsbezogenen Forschung zu formativem Assessment basiert.

Beide Facetten werden zusammenfassend in Tabelle 1 einander gegenübergestellt und in den Kapiteln 3 und 4 jeweils detailliert ausgeführt, bevor der angenommene Zusammenhang in einem Prozessmodell in Kapitel 5 dargestellt wird.

Tabelle 1: Gegenüberstellung der Facetten diagnostischer Kompetenz

Urteilsgenauigkeit	Beobachtungskompetenz
<ul style="list-style-type: none"> • rückblickend • Statusdiagnostik • summativ • informell • Einschätzung von Lösungsfähigkeiten in Tests • basiert auf Erkenntnissen aus der Grundlagenforschung zur Diagnosekompetenz 	<ul style="list-style-type: none"> • (prozess-)begleitend • Prozessdiagnostik • formativ • semiformell • Erkennen von Fähigkeiten mittels Beobachtung • basiert auf Erkenntnissen aus der anwendungsbezogenen Forschung zu formativem Assessment

2.6 Theoretische Einbettung der diagnostischen Kompetenz

Die diagnostische Kompetenz ist Teil der professionellen Expertise von Lehrpersonen und wird verortet in bewährten Kompetenzmodellen. Da frühpädagogische Fachkräfte auch die Aufgabe haben, Lehr-Lernprozesse anzuregen und zu optimieren, können diese Modelle auch für sie herangezogen werden. Der Stellenwert der diagnostischen Kompetenz im Kontext frühpädagogischer Lehr- und Lernprozesse wird anhand einer Adaption des Angebots-Nutzungs-Modells von Helmke und Schrader (2020) hervorgehoben. Den übergeordneten Rahmen im Hinblick auf adaptive Förderprozesse im Elementarbereich liefert die soziokulturelle bzw. sozialkonstruktivistische Theorie von Wygotski (Vygotsky et al., 1930/1978; Vygotskij, 2002).

Beurteilen stellt in den Standards für die Lehrerbildung der Kultusministerkonferenz (KMK, 2004) einen von vier Kompetenzbereichen neben Unterrichten, Erziehen und Innovieren dar, welcher durch zwei Vermögen näher definiert wird:

- „Lehrerinnen und Lehrer diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt und beraten Lernende und deren Eltern“ (KMK, 2004, S. 11).
- „Lehrerinnen und Lehrer erfassen die Leistungsentwicklung von Schülerinnen und Schülern und beurteilen Lernen und Leistungen auf der Grundlage transparenter Beurteilungsmaßstäbe“ (KMK, 2004, S. 12).

Analog ließe sich für den vorschulischen Bereich formulieren: Frühpädagogisches Fachpersonal beobachtet und dokumentiert Lernvoraussetzungen und Lernprozesse von Kindern, fördert die Kinder gezielt usw. Der Austausch des Verbs ‚diagnostizieren‘ durch die Verben ‚beobachten und dokumentieren‘ lässt diagnostische Kompetenz für Erzieherinnen greifbarer werden. Deutlich wird auch, dass die inhaltlichen Überschneidungen von ‚Diagnostik‘ und ‚Beobachten und Dokumentieren‘ sehr groß bzw. die Unterschiede zwischen beiden marginal sind, wie es für den vorschulischen Bereich z.B. die Bildungs- und Erziehungsempfehlungen für Kindertagestätten in Rheinland-Pfalz (Ministerium für Bildung, 2018) zeigen.

2.6.1 Sozialkonstruktivistische Theorie

Die Bedeutung der diagnostischen Kompetenz zeigt sich insbesondere in der sozialkonstruktivistischen Theorie von Wygotski (Vygotskij, 2002). Hier wird die diagnostische Kompetenz Lehrender, welche in konkreten Fördersituationen zum Tragen kommt, auf der Mikroebene der Interaktion fokussiert, der im Kindergarten eine bedeutsame Rolle zukommt, welche aber auch im schulischen Kontext relevant ist. Seine Theorie wird herangezogen, weil sie letztlich das Fundament der vorausgegangenen Modelle darstellt und die Relevanz diagnostischer Kompetenz für die Interaktion in Lehr-Lernsituationen verdeutlicht. Im frühpädagogischen Bereich hat Wygotskis Lerntheorie (Vygotskij, 2002) im Verlauf der letzten Jahre sowohl national (Textor, 2000) als auch international (Winsler, 2003, S. 156) zunehmend an Bedeutung gewonnen. Siraj-Blatchford (2007, S. 106) zufolge basiert das wissenschaftliche Verständnis von Lernen im gesamten frühpädagogischen Kontext der letzten Jahre auf den Prinzipien des Sozialkonstruktivismus.

In Wygotskis Theorie wird diagnostische Kompetenz als Voraussetzung für eine adaptive Förderung verstanden (Vygotskij, 2002). Eine sachkundige, kompetentere Person gibt den Kindern mit ihrer Unterstützung die Möglichkeit, in die *Zone der nächsten Entwicklung* vorzudringen und dort zu arbeiten. Dabei werden diagnostische Kompetenzen als Voraussetzung bewusster und gezielter Unterstützung, des Scaffoldings gesehen.

Die Verortung der diagnostischen Kompetenz im soziokulturellen Ansatz Wygotskis (Vygotsky, 1930/1978; vgl. hierzu auch Kozulin, Gindis, Ageyev & Miller, 2003; vgl.

auch Woolfolk, 2008, S. 53–60) wird im Folgenden anhand der zwei zentralen Konstrukte ‚Zone der nächsten Entwicklung‘ und ‚Scaffolding‘ dargestellt.

Die Zone der nächsten Entwicklung

Wygotski zufolge wird durch die Diagnose des kindlichen Entwicklungsstandes einerseits nur der bereits herangereifte Teil der kindlichen Entwicklung berücksichtigt, andererseits ein wesentlicher Teil sogenannter „reifender Funktionen“ vernachlässigt (Vygotksij, 2002, S. 326). Diesen Bereich reifender Funktionen bezeichnet er als *Zone der nächsten Entwicklung (Zone of Proximal Development - ZPD)*, die er wie folgt definiert: „Dieser Unterschied im geistigen Alter oder aktuellen Entwicklungsniveau, das durch selbstständig gelöste Aufgaben bestimmt wird, und dem Niveau, das das Kind beim Lösen von Aufgaben zwar nicht selbstständig, aber in Zusammenarbeit erreicht, bestimmt die Zone der nächsten Entwicklung“ (Vygotksij, 2002, S. 327). Wygotski geht davon aus, dass die fähigeren Mitglieder der Kulturgemeinschaft (hier: die diagnostisch fähige Erzieherin) durch Interaktionen mit Kindern deren kognitive Entwicklung fördern (Woolfolk, 2008, S. 63). “The path from object to child and from child to object passes through another person” (Vygotksy, 1978, S. 30). Nahezu alle psychischen Strukturen und kognitiven Fähigkeiten gehen auf soziale Phänomene zurück, die ursprünglich in Interaktionen mit anderen (kompetenteren) Personen auftraten und dann von dem jeweiligen Kind internalisiert werden (Textor, 2000, S. 67). Den Übergang in die *Zone der nächsten Entwicklung* bezeichnet Wygotski als das sensibelste Symptom, um die Entwicklungsdynamik und den Lernerfolg eines Kindes zu kennzeichnen (Vygotksij, 2002, S. 329). Die Feststellung des Entwicklungsniveaus durch Statusdiagnostik reicht nicht aus um Fördersituationen flexibel adaptiv zu gestalten. Eine Erweiterung der Statusdiagnostik um die Wahrnehmung „reifender Funktionen“ (Vygotksij, 2002, S. 326), die sich im Prozess erschließen, erscheint notwendig. Dies bedeutet, dass sich diagnostische Kompetenz hier nicht auf Statusdiagnostik bzw. Urteilsgenauigkeit beschränkt, sondern die Beobachtung in Fördersituationen bzw. die Prozessdiagnostik berücksichtigt.

Scaffolding

Diagnostische Kompetenz spielt auch im Zusammenhang mit Scaffolding eine zentrale Rolle; Wygotksis Arbeiten prägen diesen Begriff. Scaffolding wird zunächst sehr allgemein als Prozess beschrieben, der damit beginnt, das Kind zu Handlungen mit für es erkennbaren Lösungswegen zu animieren; sobald dies erreicht ist, kann der kompetente Partner Unklarheiten für das Kind deuten und seine unterstützende Rolle nur so lange

beibehalten, bis das Kind selbstständig zu arbeiten in der Lage ist (Wood, Bruner & Ross, 1976, S. 96). Das Erkennen und Deuten von Unklarheiten ist also gleichzusetzen mit dem Diagnostizieren. Die genaue Beobachtung und Beurteilung der kindlichen Aktivitäten ist zwingende Voraussetzung, um zu einer Entscheidung für oder gegen die Notwendigkeit von Unterstützung (Rückmeldung im Lernprozess) zu kommen.

Wygotski setzt somit diagnostische Kompetenz voraus, ohne diese explizit zu benennen. Sie zeichnet sich durch genaue Wahrnehmung und Beobachtung kindlicher Fähigkeiten und Fertigkeiten in konkreten Situationen und Interaktionen mit dem Kind aus. Diese Wahrnehmung kindlicher Fähigkeiten in (Förder-)Situationen sind Merkmale einer Beobachtungskompetenz.

Zusammenfassend lässt sich festhalten, dass in Wygotkis Theorie in der „Zone der aktuellen Entwicklung“ die Statusdiagnostik zur Klärung des Entwicklungsstandes von Bedeutung ist und in der *Zone der nächsten Entwicklung* insbesondere die Prozessdiagnostik zur optimalen Gestaltung der adaptiven Förderung eines Kindes Relevanz besitzt.

2.6.2 Modell aus der Lehr-Lernforschung

Um die diagnostische Kompetenz in einem allgemeinen Modell professioneller Handlungskompetenz zu verorten, eignet sich das auf Weinert, Schrader und Helmke (1990) zurückgehende Kompetenzmodell, welches von Baumert und Kunter (vgl. Baumert & Kunter, 2006; Kunter et al., 2011) erweitert wurde. Auch im vorschulischen Kontext wird das Kompetenzmodell zur theoretischen Fundierung der Diagnosegenauigkeit von Erzieherinnen (vgl. Dollinger, 2013, S. 38) bzw. zur Darstellung der professionellen Kompetenzen von pädagogischen Fachkräften im Allgemeinen (vgl. Benz et al., 2017, S. 41) herangezogen.

Im Zentrum des Modells (vgl. Abb. 1) steht das Professionswissen als kognitive Komponente, welche in Zusammenhang steht mit drei weiteren Aspekten professioneller Kompetenz, den motivational-affektiven Komponenten ‚Überzeugungen und Werthaltungen‘, ‚Motivationalen Orientierungen‘ und ‚Selbstregulativen Fähigkeiten‘.

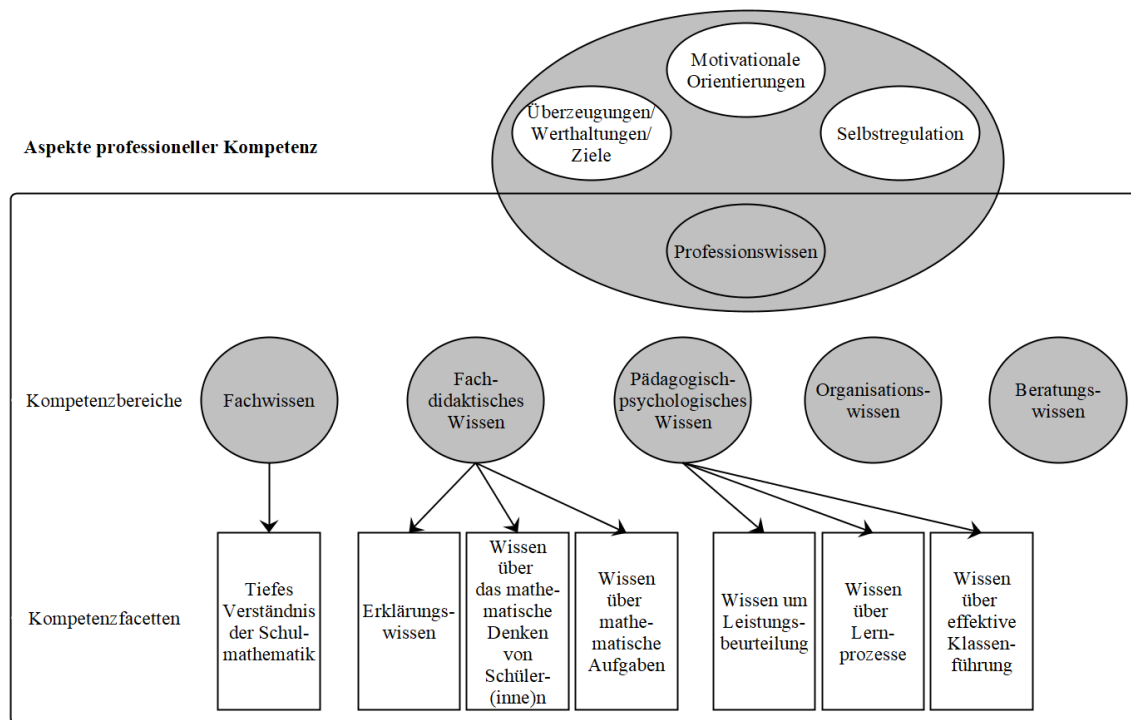


Abbildung 1: Kompetenzmodell mit Spezifikationen für das Professionswissen von Baumert und Kunter (2011, S. 32)

Die Verortung der diagnostischen Kompetenz kann innerhalb des Modells an mehreren Stellen erfolgen. Krauss, Kunter, Brunner und Baumert (2004, S. 43) hingegen definieren diagnostische Kompetenz nur als Facette des Kompetenzbereichs fachdidaktisches Wissen. Bei Baumert und Kunter (2011, S. 37–38) stellt Diagnostik von Schüler(innen)wissen letztlich nur eine von drei Wissensdimensionen fachdidaktischen Wissens dar, welche die Autoren wie folgt beschreiben:

- „- Wissen über das didaktische und diagnostische Potenzial, die kognitiven Anforderungen und impliziten Wissensvoraussetzungen von Aufgaben, ihre didaktische Sequenzierung und die langfristige curriculare Anordnung von Stoffen,
- Wissen über Schülervorstellungen (Fehlkonzeptionen, typische Fehler, Strategien) und Diagnostik von Schülerwissen und Verständnisprozessen,
- Wissen über multiple Repräsentations- und Erklärungsmöglichkeiten.“

Die diagnostischen Fähigkeiten sieht z.B. Dollinger (2013, S. 37) im vorliegenden Kompetenzmodell in den Kompetenzfacetten ‚Wissen über das mathematische Denken von Schüler(inne)n‘, ‚Wissen über mathematische Aufgaben‘ und ‚Wissen um Leistungsbeurteilung‘ eingebettet. Damit wäre das Konstrukt ‚Diagnostische Kompetenz‘ sowohl dem fachdidaktischen als auch dem pädagogisch-psychologischen Wissensbe-

reich zuzuordnen. Dieses Beispiel verdeutlicht, dass das Modell im Hinblick auf die Verortung diagnostischer Kompetenz unterschiedlich interpretiert werden kann.

Weniger die professionellen Handlungskompetenzen als vielmehr die individuellen Wissensstrukturen stehen bei Baumert und Kunter (2006), wie auch schon bei Weinert et al. (1990) im Vordergrund (vgl. auch Karst, 2012, S. 84). Das bedeutet, dass diagnostische Kompetenz isoliert betrachtet wird und nicht im Kontext oder als Bestandteil adaptiver Lehrprozesse.

Im Hinblick auf die Erweiterung des Konstrukts diagnostische Kompetenz, die bis dato zumeist als Urteilsgenauigkeit gefasst wurde, hin zu einem erweiterten diagnostischen Kompetenzbegriff nach z.B. Jürgens und Lissmann (2015, S. 24) oder im Sinne Helmkes diagnostischer Expertise (Helmke, 2017, S. 119), wird als weiterer, insbesondere für den vorschulischen Bereich relevanter Indikator, die Beobachtungskompetenz als bedeutsamer diagnostischer Anteil formativen Assessments angenommen.

2.6.3 Frühpädagogisches Kompetenzmodell

Die diagnostische Kompetenz ist im speziell für den frühpädagogischen Bereich konzipierten und weithin anerkannten sowie breit rezipierten Allgemeinen Kompetenzmodell (vgl. Abb. 2) von Fröhlich-Gildhoff, Nentwig-Gesemann und Pietsch (2011, S. 17) nicht explizit zu finden. Sie ist jedoch implizit in den Begriffen Situationswahrnehmung und -analyse enthalten, welche die Bedeutung der Beobachtung betonen. Die Wahrnehmung und Analyse einer Situation, welche zu Wissen und zu Handlungspotentialen führen, nehmen somit eine zentrale Position ein.

Kern und Ausgangspunkt des Modells ist die Situationswahrnehmung und -analyse, welche im Modell als Element der Differenzierung von Handlungsgrundlagen (Disposition) erscheint und auch unter Berücksichtigung der Handlungsbereitschaft und des Handlungsvollzugs (Performanz) zu einer entsprechenden (Handlungs-)Kompetenz frühpädagogischer Fachkräfte führt (Fröhlich-Gildhoff et al., 2011, S. 17).

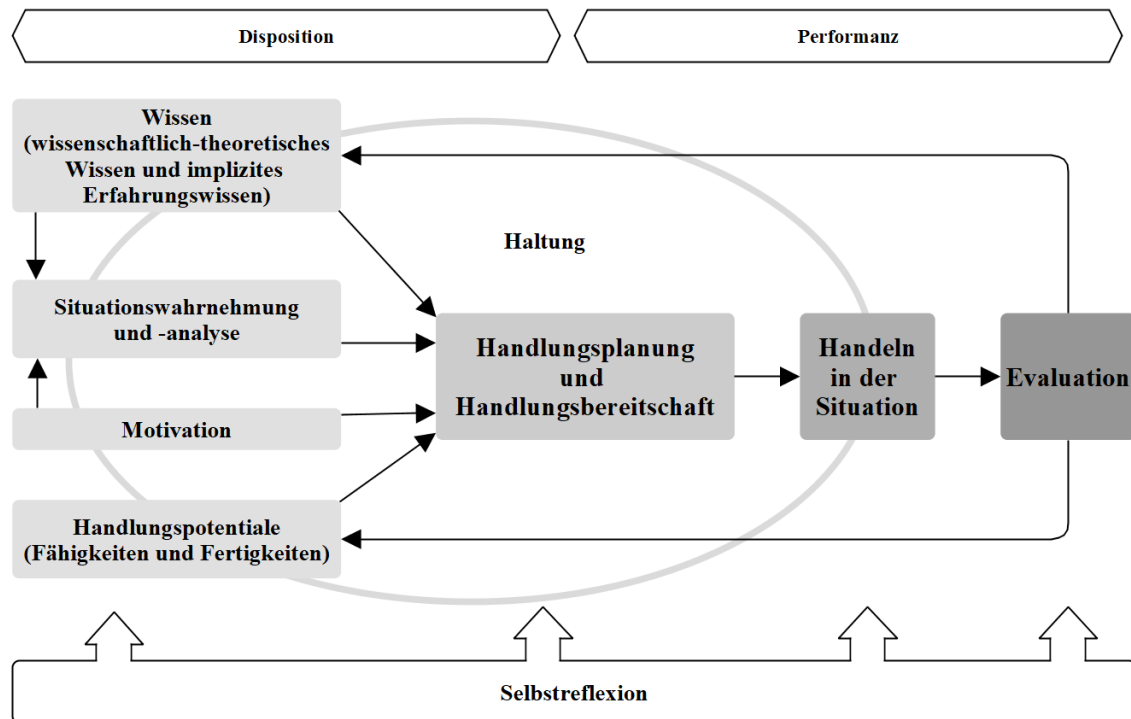


Abbildung 2: Allgemeines Kompetenzmodell nach Fröhlich-Gildhoff et al. (2011, S. 17)

Entscheidend für die Handlungsfähigkeit von frühpädagogischem Fachpersonal ist Fröhlich-Gildhoff et. al. (2011, S. 18) zufolge das wechselseitige Zusammenspiel von explizitem, wissenschaftlich-theoretischem Wissen, implizitem Erfahrungswissen sowie Fähigkeiten und Fertigkeiten, methodischer oder didaktischer Art. Dabei wird die Bereitschaft zum Handeln durch die jeweilige Situationswahrnehmung und Situationsanalyse, die vor allem durch Beobachtung erfolgt, beeinflusst (Röhner & König, 2015, S. 100).

2.6.4 Frühpädagogisches Angebots-Nutzungs-Modell

Wie diagnostische Kompetenz wirkt, kann mithilfe von Angebots-Nutzungs-Modellen aufgezeigt werden. Angebots-Nutzungs-Modelle (Klieme et al., 2006, S. 131; Helmke, 2017, S. 71; Reusser, 2011, S. 21) werden über Fächer, Schultypen und Schulstufen hinweg konzipiert und lassen dementsprechend auch eine Übertragung auf den vorschulischen Kontext zu. Sie wurden bereits in mehreren frühpädagogischen Studien (Kuger & Kluczniok, 2008; Kammermeyer, Roux, Schneider & Stuck, 2011, S. 10; Smidt, 2012, S. 18) insbesondere im Rahmen von Untersuchungen bereichsspezifischer Fördermaßnahmen herangezogen.

Die Tatsache, dass die Kinder optional unterstützende Förderung an deren Lernvoraussetzungen ansetzt (Praetorius, Karst & Lipowsky, 2012, S. 133), gilt für Bildungsprozesse in der Grundschule gleichermaßen wie für die Förderung in vorschulischen Einrichtungen. Die diagnostische Kompetenz ist für diese adaptiven Bildungsprozesse bedeutsam. In einer Vielzahl von Modellen, die sich mit Lehr-Lernprozessen auseinandersetzen, wird der diagnostischen Kompetenz eine zentrale Rolle zugesprochen, weil das Wissen um die Leistungsvoraussetzungen der Kinder als Voraussetzung für Adaptivität und die optimale Nutzung von Lerngelegenheiten (Hardy et al., 2011, S. 820) gilt.

Dabei verdeutlichen Angebots-Nutzungs-Modelle, dass jegliche Form diagnostischer Kompetenz als bedeutsame Voraussetzung für die Gestaltung, Initiierung oder Adaption von Fördersituationen nur ein Einflussfaktor unter vielen ist, der lediglich indirekt zu einer Wirkung auf Ebene der Kinder führen kann.

Das Angebots-Nutzungs-Modell der Wirkfaktoren akademischer Leistung von Helmke (2017, S. 71), das im Folgenden auf den frühpädagogischen Bereich übertragen wird, verdeutlicht, an welcher Stelle die diagnostische Kompetenz von Erzieherinnen im Kontext frühpädagogischer Förderung steht. Es stellt das Unterrichtsangebot und die Nutzung desselben (Lernaktivitäten) in den Mittelpunkt. Helmke betont, dass der eigentliche Motor die individuellen Lernaktivitäten der lernenden Kinder sind. Ausmaß und Qualität dieser Lernaktivitäten bestimmen den Lernerfolg. Zudem handelt es sich „um ein komplexes, multikausales Geschehen, bei dem sich – innerhalb bestimmter Grenzen – unterschiedliche Determinanten des Lernerfolges wechselseitig kompensieren oder substituieren können. Das gesamte Wirkungsmuster ist eingebettet in vielfältige Kontexte“ (Helmke & Schrader, 2020, S. 1).

Dieses Modell bietet sich in seiner Struktur auch für die Darstellung von Lehr-Lernprozessen im vorschulischen Bereich an. Aus diesem Grund wurde das Modell von Helmke (2017, S. 71) adaptiert. Dabei sind Struktur und Kontextmerkmale unverändert geblieben; lediglich wenige Begrifflichkeiten wurden an den frühpädagogischen Bereich angepasst, wie in der folgenden Grafik dargestellt.

Im folgenden Modell wurden Begriffe wie z.B. „Unterricht“ durch „Fördersituation“ oder „außerschulische Lernaktivitäten“ durch „Lernaktivitäten im Freispiel bzw. in offenen Lernphasen“ ersetzt. Die „Gruppenmanagementkompetenz“ wird hierbei als begriffliches Pendant zur „Klassenführungskompetenz“ eingeführt.

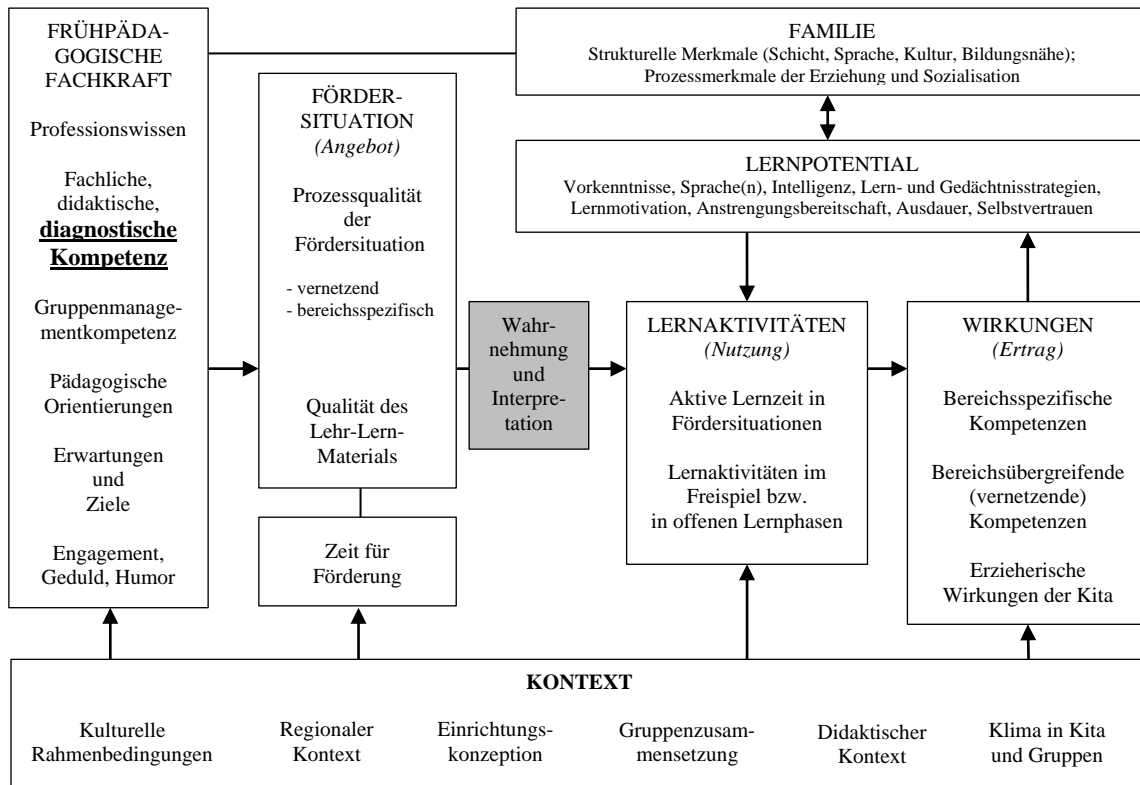


Abbildung 3: Frühpädagogisches Angebots-Nutzungs-Modell der Wirkfaktoren akademischer Leistungen in Anlehnung an Helmke und Schrader (2020)

Am Anfang des Modells steht die frühpädagogische Fachkraft mit ihrer diagnostischen Kompetenz. Es wird deutlich, dass diagnostische Kompetenz nur ein intrapersoneller Faktor von vielen ist. Daneben sind das Professionswissen der Erzieherin, die fachliche und didaktische Kompetenz, ihre pädagogische Orientierung, ihre Erwartungen und Ziele sowie Eigenschaften wie z.B. Engagement, Geduld und Humor von Bedeutung. Je nach Ausprägung dieser Merkmale und Eigenschaften gestaltet die Erzieherin ein Angebot für die Kinder. Dieses kann bereichsspezifisch oder bereichsübergreifend, vernetzend sein. Entscheidend für die Wirkung dieses Angebots ist die Qualität der Fördersituation sowie des Lehr-Lernmaterials, aber auch die Zeit, welche für die Förderung zur Verfügung steht.

Das Angebot wird von den Kindern individuell wahrgenommen und interpretiert und führt bei entsprechender Passung zur Nutzung, d.h. zu Lernaktivitäten. Inwieweit diese Angebote genutzt werden (können), zeigt sich in der aktiven Lernzeit innerhalb der konkreten Fördersituation, aber auch an den Lernaktivitäten, die auf Angeboten basieren und sich dann im Freispiel und in offenen Lernphasen zeigen. Einen großen Einflussfaktor stellt das Lernpotential der Kinder dar. Je nach Vorkenntnissen, sprachlichen

Voraussetzungen, Intelligenz, Lern- und Gedächtnisstrategien, Lernmotivation, Anstrengungsbereitschaft, Ausdauer und Selbstvertrauen der Kinder ist ihnen die Nutzung des Angebots besser oder schlechter möglich. Dieses Lernpotential hängt zu einem großen Teil von den familiären Voraussetzungen (Schicht, Sprache, Kultur, Bildungsnähe) ab, womit zum einen strukturelle Merkmale und zum anderen Prozessmerkmale der Erziehung und Sozialisation gemeint sind.

Je nach Wahrnehmung und Nutzung der Angebote erweist sich die Fördersituation als ertragreicher und wirksamer sowohl was die Entwicklung und den Ausbau bereichsspezifischer, bereichsübergreifender Kompetenzen betrifft als auch in erzieherischer Hinsicht.

Alle Elemente des Modells, welche die Wirkungskette von der frühpädagogischen Fachkraft über die Fördersituation hin zu den Lernaktivitäten und letztlich zu den Wirkungen beschreiben, werden vom Kontext auf Mikro- und Makroebene beeinflusst. Der Kontext umfasst die kulturellen Rahmenbedingungen der Gesellschaft, schließt die Auswirkungen des regionalen Kontextes mit ein, die Konzeption der Einrichtung, die Gruppenzusammensetzung und den didaktischen Kontext sowie das Klima in der Kita wie auch Gruppenklima.

Zusammenfassend muss der Weg von den diagnostischen Kompetenzen zu den Wirkungen als ein weiter und komplexer bezeichnet werden. Er führt vom Diagnostizieren zum einen über die Aktivitäten der Erzieherin (gezielte bereichsspezifische Anregung, Interaktion, Gestaltung von Lernumgebungen etc.) und zum anderen über die Aktivitäten der Kinder (Interesse, Motivation, Beteiligung) zu einem entsprechenden Ertrag einer Lernsituation. Dieser Ertrag bzw. die tatsächliche Wirkung wird von einer Vielzahl Faktoren beeinflusst und ist davon abhängig, inwieweit Kinder Lernchancen und Angebote nutzen (können). Dies gilt sowohl für Lernprozesse im Kindergarten als auch für solche in der Grundschule.

2.6.5 Arbeitsmodell diagnostische Kompetenz

Ein aktuelles Modell, das theoretische und empirische Ansätze aus teilweise sehr unterschiedlichen Perspektiven verbindet, ist das Arbeitsmodell diagnostischer Kompetenz von Herppich et al. (2017, S. 75) des Netzwerks zur diagnostischen Kompetenz von Lehrkräften (NeDiKo). Das Netzwerk stellt dabei „kein Kompetenzmodell im engeren

Sinne“, sondern vielmehr ein Rahmenmodell als ‚Blaupause‘ zur Verfügung, welche die Einordnung bestehender Befunde ermöglicht und weitere integrative Forschung inspirieren soll (ebd.). Das Modell, welches explizit produkt- und prozessorientierte Ansätze verbinden möchte (ebd., S. 76), liefert damit zunächst einen passenden Rahmen zur Verortung diagnostischer Kompetenz.

Dem Modell liegt ein bildungswissenschaftlicher Kompetenzbegriff zugrunde. Kompetenzen werden als kognitive Leistungsdispositionen verstanden, die in unterschiedlichen Situationen wirksam werden und das Potenzial haben, „die Anforderungen der jeweiligen Situationen relativ konsistent und relativ stabil sowie quantifizierbar zu meistern“ (ebd., S. 80). Daraus folgt, dass diagnostische Kompetenz als eine kognitive Leistungsdisposition betrachtet wird, die es Erzieherinnen oder Lehrkräften ermöglicht, „relativ konsistent und relativ stabil sowie quantifizierbar pädagogisch-diagnostische Anforderungen in verschiedenen pädagogischen Handlungssituationen zu meistern“ (ebd.).

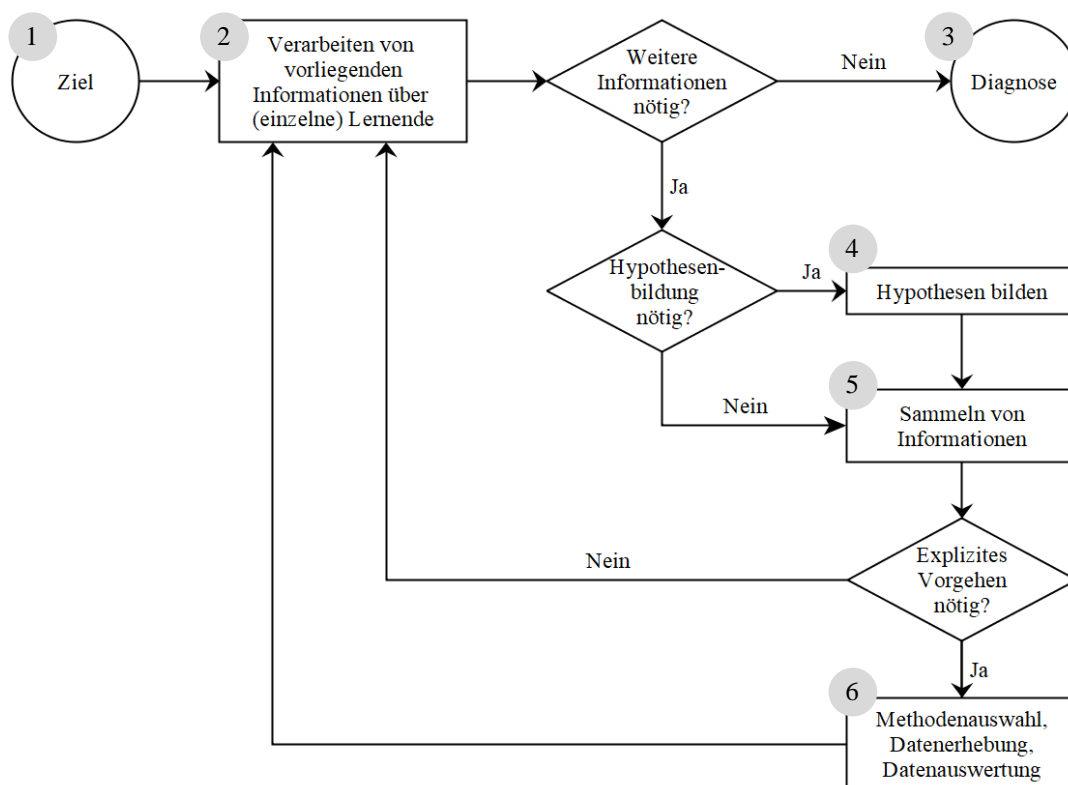


Abbildung 4: NeDiKo-Prozessmodell pädagogischer Diagnostik nach Herppich et al. (2017, S. 82)

Das Modell wird zur besseren Übersicht zweiteilig dargestellt. Zunächst wird das Prozessmodell (Abb. 4) dargestellt, welches einen Fokus auf die quantifizierbare Perfor-

manz der Diagnosekompetenz in der jeweiligen Situation setzt. Danach wird die Verortung des Prozessmodells im übergeordneten Arbeitsmodell (Abb. 5) vorgestellt, welches das Prozessmodell je nach Anzahl der Situationen mehrfach enthalten kann.

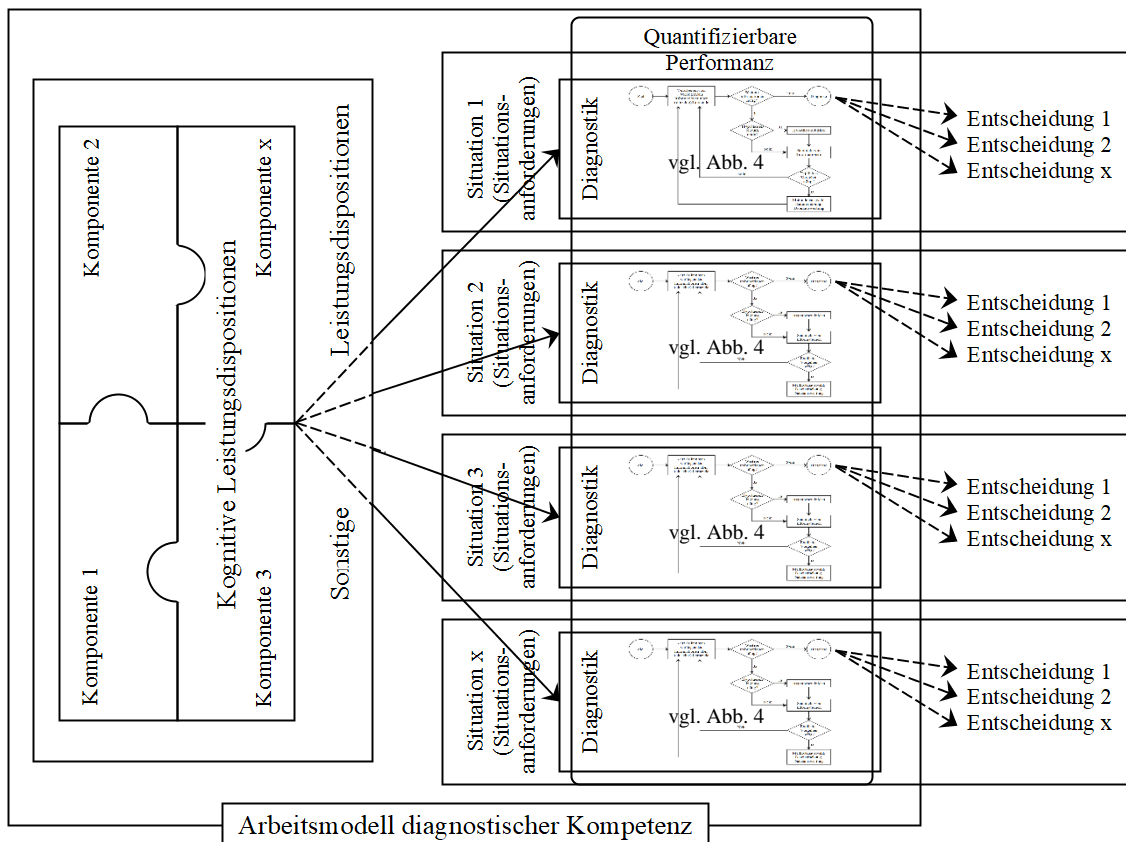


Abbildung 5: NeDiKo-Arbeitsmodell diagnostischer Kompetenz nach Herppich et al. (2017, S. 81)

Das Prozessmodell wird an einem Beispiel aus dem Elementarbereich konkretisiert und anschließend abstrahiert. Zu Beginn des Prozesses (Abb. 4) steht ein diagnostisches Ziel (1): in einer Spielsituation (Brettspiel mit Würfeln) möchte eine Erzieherin z.B. Erkenntnisse über die mathematischen Fähigkeiten zweier Kinder erlangen. Das Verarbeiten von vorliegenden Informationen (2) über die Lernenden bedeutet in diesem Beispiel, dass die Erzieherin anhand der Beobachtung (Kind 1 zählt die Würfelpunkte ab. Kind 2 erkennt die Punktemenge simultan und beginnt direkt zu ziehen.) im (Spiel-)Prozess eine Diagnose bzgl. der Mengenerfassung und Zählfertigkeiten der Kinder erstellt. Sollte es so sein, dass die Erzieherin im Prozess zu keiner abschließenden Diagnose (3) kommt, entwickelt sie entweder eine Hypothese (4) zu den Fähigkeiten der Kinder oder sie sammelt weitere Informationen (5), welche entweder direkt verarbeitet werden können (2) oder ein explizites diagnostisches Vorgehen (6) erfordern. Im Modell sind damit jedoch explizit keine Handlungsentscheidungen gemeint, die Einfluss auf den Prozess

nehmen, sondern das Durchlaufen eines diagnostischen Prozesses im Sinne einer Methodenauswahl, Datenerhebung und Datenauswertung. Die Methodenauswahl bezieht sich dabei auf die Auswahl diagnostischer Verfahren und ebenso wenig auf Handlungsentscheidungen, welche die Aktivitäten der Kinder betreffen. Diese Entscheidungen sind nicht mehr Teil des Arbeitsmodells diagnostischer Kompetenz, wie es Abbildung 5 zeigt.

Handlungsentscheidungen, welche die Spielsituationen zielgerichtet verändern – also den Prozess beeinflussen, um weitere diagnostische Informationen zu generieren, werden im Modell nicht berücksichtigt: Die Erzieherin kann z.B. die ‚diagnostische‘ Entscheidung treffen, sich in das laufende Spiel einzumischen und deutliche Hilfestellungen zu geben. Um das Kind besser diagnostizieren zu können, um zu sehen, auf welcher Stufe der nächsten Entwicklung sich das Kind mit Unterstützung bereits erfolgreich betätigen kann. In diesem Fall verwischen die Grenzen zwischen Diagnostik und Förderung, was die Limitation des wissenschaftstheoretischen Arbeitsmodells aufzeigt.

Das Prozessmodell (Abb. 4) kann im übergeordneten Arbeitsmodell (Abb. 5) mehrfach verortet werden, wobei die Häufigkeit des Vorkommens von der Anzahl der Situationen abhängig ist (Abb. 4 ist als Miniatur in Abb. 5 zu erkennen). Limitation

Bezogen auf das Beispiel zeigt das Arbeitsmodell, dass eine Erzieherin auf Basis ihrer kognitiven Leistungsdisposition, deren Wirksamwerden durch die sonstigen Leistungsdispositionen (z.B. Fachwissen über die Entwicklung des Zählens) gefiltert wird, in verschiedenen pädagogischen Handlungssituationen (z.B. beim Umgang mit Würfeln während eines Brettspiels) pädagogische Diagnostik durchführt (z.B. durch gezielte Beobachtung), an deren Ende jeweils eine Diagnose steht (z.B. Das Kind kann bereits eine Eins-zu-Eins-Zuordnung). Auf Grundlage dieser Diagnose trifft sie eine pädagogische Entscheidung. Diese kann darin bestehen, dass sie dem Kind bestimmte Hilfen, etwa Materialien zur Verfügung stellt oder eine gemeinsame Übung mit dem Kind durchführt, z.B. ein Klatschspiel mit einem Zählvers. Diese Förderung ist im vorliegenden Modell jedoch nicht mehr Teil der Diagnostik und lässt dementsprechend keine direkten Rückschlüsse auf die diagnostische Kompetenz zu.

Hervorzuheben ist die Bedeutung der Situationen im vorliegenden Modell, mit denen die Kontextspezifität von Kompetenzen (Klieme & Leutner, 2006, S. 879) betont wird. Diagnostische Kompetenz ist situationsabhängig, wobei zu berücksichtigen ist, dass die Lernsituationen in der Kita andere sind als in der Schule. In der Kita handelt es sich

weniger um formale Lernsettings wie in der Schule als vielmehr um informelle Lernsettings. Aus diesem Grund unterscheiden sich Kita und Schule im Hinblick auf das Aufsuchen, Finden oder Herstellen diagnostisch geeigneter Situationen. Während Erzieherinnen in der Praxis die für adäquates Diagnostizieren relevanten Situationen zunächst erkennen oder gezielt herstellen müssen, sind diese im schulischen Fachunterricht von vornherein gegeben. Problematisch kann es auch sein, wenn in Spielsituationen aufgrund der unvollständigen Funktionalität des Spiels (Hauser, 2016, S. 20–21) in der Kita Fähigkeiten verborgen bleiben. Dies ist der Fall, wenn z.B. ein sehr emphatischer Sechsjähriger mit einem Dreijährigen spielt und sich an das Niveau des Dreijährigen anpasst, um das Zusammenspiel positiv zu beeinflussen, indem er den Jüngeren auf diese Weise zum Sieg verhilft.

Die Diagnostische Kompetenz als quantifizierbare Performanz wird im NeDiKo-Arbeitsmodell (vgl. Abb. 5) als wiederkehrender schematischer Prozess in verschiedenen Situationen dargestellt. Diese Situationen beeinflussen den Prozess durch ihre Anforderungen genauso, wie die aus der Diagnostik resultierende Entscheidung, die abhängig von der Situation getroffen wird. Deutlich erkennbar wird in diesem Modell auch, dass die Entscheidung nicht mehr Teil der diagnostischen Kompetenz ist.

Je nach Situationsanforderung kommen kognitive und sonstige Leistungsdispositionen in verschiedenen Situationen unterschiedlich zur Geltung, weshalb bei der Erfassung der diagnostischen Kompetenz der Auswahl repräsentativer Situationen eine besondere Bedeutung zukommt (vgl. Praetorius et al., 2017). Schließlich hängt hiervon ab, ob später tatsächlich diagnostische Kompetenz gemessen und gefördert wird und nicht irgendwelche anderen Fertigkeiten, die keine Berufsrelevanz besitzen, wie Herppich et al. (2017, S. 84) herausstellen. Die ökologische Validität der Situationen und damit die Frage nach der Repräsentativität für die tägliche Arbeit z.B. in Kindertagesstätten ist von hoher Bedeutung. Auch die Komplexität der Situation kann eine Rolle spielen, wenn es darum geht, einzelne Kinder oder ganze Gruppen zu diagnostizieren. Diagnostische Kompetenz kommt sowohl bei informeller als auch formeller Diagnostik in Kita und Schule zum Tragen.

Mit der Situation hängt auch die Form der Diagnostik zusammen. Handelt es sich um Alltagsdiagnostik, welche Situationen provoziert, und wird damit eine eben solche herausfordernde Diagnostik betrieben, oder wird der Einsatz eines standardisierten Testinstruments geplant? Je nachdem findet die Diagnostik in unterschiedlichem Ausmaß

kontrolliert oder automatisiert statt. Alle Vorgehensweisen und Verfahren lassen sich im Prozessmodell (Abb. 5) abbilden. „Bei stärker kontrollierter Diagnostik werden auf dem Weg von der Zielstellung zur Diagnose mehr Elemente des Modells durchlaufen als bei stärker automatisierter Diagnostik. Elemente unterhalb der obersten Horizontale [Schritte 1, 2, 3] in der Abbildung [4] zeigen Schritte an, die nicht bei jeder Diagnostik ausgeführt werden.“ (Herppich et al., 2017, S. 82)

Herppich et al. (ebd., S. 81) betonen, dass die diagnostischen Handlungen im diagnostischen Prozess sowie die Diagnose selbst als Performanz messbar, also quantifizierbar sind. Die Autorinnen führen als Beispiele für Komponenten der kognitiven Leistungsdisposition „konzeptuelles Wissen“, „Wissen über diagnostische Methoden, Prozesse und über Ziele“ oder „methodisches Wissen“ an (ebd., S. 83).

Als sonstige Leistungsmerkmale einer Person werden z.B. ihre allgemeinen Fähigkeiten zur Informationsverarbeitung, ihre Motivation, diagnostikrelevante Selbstkonzepte und ihre Überzeugungen verstanden (ebd., S. 83).

Das Potential des zweiteiligen Modells besteht in der Möglichkeit einer Zusammenführung verschiedener Forschungsrichtungen, wie z.B. summativer und formativer Ansätze der Diagnosekompetenzforschung. Es bietet die Möglichkeit, „Forschungsrichtungen einzubeziehen, welche sich mit der Erforschung des diagnostischen Handelns von Lehrkräften befassen, die bisher aber weitgehend unabhängig von der Theoriebildung und Empirie zur diagnostischen Kompetenz untersucht wurden. Ein Beispiel hierfür ist die Forschung zur formativen Diagnostik“ (Herppich et al., 2017, S. 91).

Damit passt das Modell zum Ansatz der vorliegenden Arbeit, welche zur Definition diagnostischer Kompetenz neben der Facette Urteilsgenauigkeit die Facette Beobachtungskompetenz einführt und damit eine Brücke zwischen dem konstruktorientierten Ansatz der Urteilsgenauigkeitsforschung und der prozessorientierten Perspektive auf diagnostische Anteile des formativen Assessments schlägt.

Beide Facetten lassen sich im oben ausgeführten Arbeitsmodell des Netzwerks zur diagnostischen Kompetenz von Lehrkräften (NeDiKo) von Herppich et al. (2017, S. 75) verorten, das vorhandene theoretische und empirische Ansätze aus teilweise sehr unterschiedlichen Perspektiven aufgreift und sie integrierend zusammenführt. Sowohl Urteilsgenauigkeit als auch Beobachtungskompetenz lassen sich auf der obersten Horizontalen verorten (Abb. 5), wobei sich die Ziele unterscheiden. Explizite Ziele werden mit Urteilsgenauigkeit, implizite Ziele mittels Beobachtungskompetenz verfolgt. Ist die Be-

obachtungskompetenz eine gerichtete, erfolgt sie also mit einem speziellen Fokus, verfolgt auch diese explizite Ziele. Im Folgenden werden beide Facetten diagnostischer Kompetenz einzeln dargestellt.

3. Urteilsgenauigkeit

Die wissenschaftliche Auseinandersetzung mit dem Konstrukt Urteilsgenauigkeit hat in der Lehr-Lernforschung bereits eine lange Tradition. Kern der Forschung, Ausgangspunkt und zentrale Methode der vorliegenden Studie ist die international bedeutsame Komponentenanalyse nach Cronbach (1955), ihre Modifizierung nach Schrader (1989) und dessen Erweiterung nach McElvany et al. (2009) (Kap. 3.1). Details zur Berechnung der Komponenten (Kap. 3.2) und die Besonderheiten methodischer Charakteristika der Messung von Urteilsgenauigkeit (Kap. 3.3) sind die Basis für die kritische Auseinandersetzung mit dem Forschungsstand der nationalen und internationalen Judgment-Accuracy-Forschung unter Einbezug relevanter Studien sowohl aus dem vorschulischen als auch schulischen Bereich (Kap. 3.4). Ein besonderes Augenmerk wird auf die Veränderung der Urteilsgenauigkeit (Kap. 3.5) und die Wirkung diagnostischer Kompetenz (Kap. 3.6) gelegt. Dieser folgt eine Zusammenfassung (Kap. 3.7) und die Ableitung von Folgerungen für die weitere Arbeit (Kap. 3.8).

3.1 Modellierung der Urteilsgenauigkeit

Die Fähigkeit eines Urteilers, Personen zutreffend zu beurteilen, definiert Schrader (2006, S. 95) als diagnostische Kompetenz (*diagnostic competence, accuracy of judgment*), weshalb sie die Grundlage für die Genauigkeit diagnostischer Urteile oder Diagnosen ist. Bei der Urteilsgenauigkeit handelt es sich um summatives Assessment bzw. um informelle Statusdiagnostik. Urteilsgenauigkeit bildet den Vergleich zwischen Testergebnis von Kindern und der Einschätzung von Lehrpersonen ab. Je höher die Übereinstimmung von Testergebnis und Einschätzung ist, desto größer ist die Urteilsgenauigkeit.

3.1.1 Linsenmodell zur Urteilsgenauigkeit

Die theoretische Verortung der diagnostischen Kompetenz im Allgemeinen erfolgte bereits in Kapitel 2. Urteilsprozesse, die der Facette Urteilsgenauigkeit zugrunde liegen und deren Verständnis für die Klärung des Begriffs relevant sind, werden zunächst anhand des Linsenmodells nach Förster und Böhmer (2017, S. 47) dargestellt.

Das Linsenmodell ist eine von Brunswick (1956) zur Untersuchung von Wahrnehmung entwickelte allgemeine Darstellung von Urteilsprozessen. Dieses Modell lässt sich in seiner logischen Struktur auf eine Vielzahl von Bereichen und Urteilkriterien anwenden (z.B. Bröder & Hilbig, 2017, S. 637–641). Im Kontext pädagogischer Diagnostik wurde dieses Modell vielfach zur Darstellung subjektiver Wahrnehmung oder impliziter Persönlichkeitstheorien (z.B. Förster & Böhmer, 2017, 47; Helmke, 2015, S. 135; Hoth, 2016, S. 89; Kleber, 1992, S. 131–132) herangezogen. Dem Linsenmodell liegt die Annahme zugrunde, dass ein nicht direkt wahrnehmbares (distales) Merkmal, wie z.B. die mathematische Kompetenz eines Kindes (vgl. Abb. 6), durch wahrnehmbare Verhaltensweisen (proximale Merkmale, Signale) erschlossen werden kann. Dabei sind sowohl die Signalvalidität (Stärke des Zusammenhangs von Signal und Merkmal) als auch die Signalnutzung (Ausmaß der Berücksichtigung einzelner Signale zur Urteilsbildung) von Bedeutung.

Förster und Böhmer (2017) wenden das Linsenmodell auf den Bereich der pädagogisch-psychologischen Diagnostik an und fokussieren das Konstrukt der Urteilsgenauigkeit oder Akkuratheit basierend auf den Vorarbeiten von Back und Nestler (2016), wie es in der folgenden Grafik dargestellt ist.

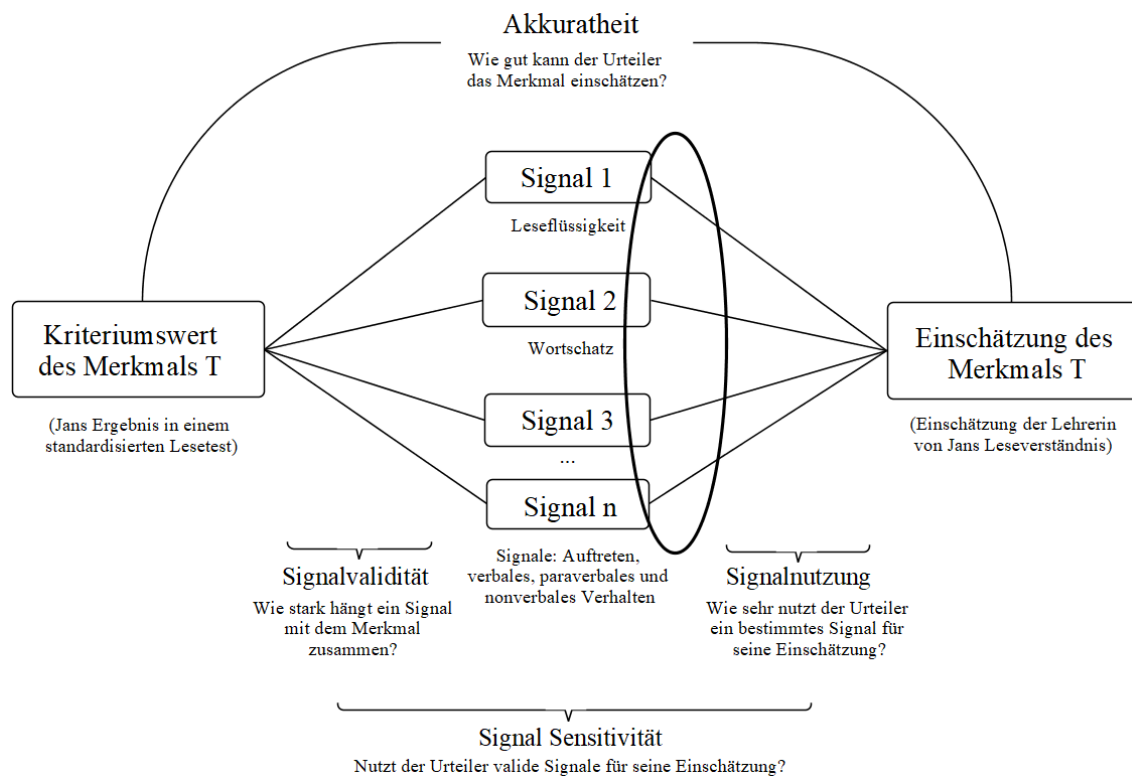


Abbildung 6: Linsenmodell nach Förster und Böhmer (2017, S. 47)

Wie gut eine Lehrkraft ein Schüler(innen)merkmal einschätzen kann, wie hoch die Urteilsgenauigkeit ist, hängt zum einen von der Signalvalidität ab und zum anderen von der Signalnutzung. Unter Signalvalidität wird dabei das Ausmaß verstanden, in dem wahrnehmbare Signale tatsächlich mit dem einzuschätzenden Merkmal zusammenhängen. Signalnutzung meint das Ausmaß der Nutzung valider Signale zur Urteilsbildung; je mehr valide Signale eine Lehrkraft nutzt und entsprechend invalide Signale ignoriert, desto höher ist die Sensitivität (vgl. Förster & Böhmer, 2017, S. 48). In diesem Zusammenhang wird auch von Konsistenz gesprochen, wenn identische Signale auf dieselbe Art und Weise genutzt werden, um verschiedene Schüler(innen) einzuschätzen. Das Linsenmodell bietet Förster und Böhmer (2017) zufolge eine Grundlage zur Untersuchung und Erklärung der Akkuratheit Beurteilender bezüglich bestimmter Zielpersonen und Eigenschaften. Deshalb liefert es die Grundlage für das Verständnis der folgenden Komponentenanalysen.

3.1.2 Komponentenanalyse nach Cronbach (1955)

Wegweisend für die nationale und internationale Judgment-Accuracy-Forschung ist die Komponentenanalyse nach Cronbach (1955). Ursprünglich wurden als Maß für diagnostische Kompetenz globale Differenzmaße verwendet. Die mathematische Beschreibung globaler Differenzmaße lautet wie folgt: „Vom Urteil des Lehrers über die Leistung eines Schülers in einem bestimmten Merkmal (...) wird jeweils die tatsächliche Schülerleistung in diesem Merkmal (...) subtrahiert. Die Differenz wird quadriert, aufsummiert und durch die Anzahl der vom Lehrer abgegeben Urteile geteilt“ (Karst, 2012, S. 41).

An diesem Vorgehen kritisiert Cronbach (1958) zum einen die Vernachlässigung der dyadischen Eigenschaft, also der Tatsache, dass sich das Maß aus Personenurteil und Kriteriumsleistung zusammensetzt. Zum anderen kritisiert er die Globalität eines solchen Maßes, weil darin Urteilstendenzen des Urteilers und Antworttendenzen des Beurteilten konfundiert sind. Das gilt z.B. nicht für den Fall, dass eine Lehrperson ein einziges Kind einschätzt, trifft jedoch dann zu, wenn eine Lehrperson Urteile über mehrere Kinder und mehrere Items abgibt.

Cronbach unterbreitet den Vorschlag, statt globaler Differenzmaße für die Urteilsgenauigkeit diese als Summe aus vier Varianzkomponenten (Cronbach, 1955, S. 178) zu definieren. Es sind dies die Niveauekomponente (*Elevation*), die Personenkomponente

(*Differential Elevation*), die Merkmalskomponente (*Stereotype Accuracy*) sowie die Wechselwirkungskomponente (*Differential Accuracy*):

- Die Niveauelemente (*Elevation*) zeigt, wie exakt eine Lehrperson den mittleren Testwert, das Niveau der Gruppe ihrer Bezugskinder einschätzen kann.
- Die Personenkomponente (*Differential Elevation*) veranschaulicht, wie die Lehrperson die Unterschiedlichkeit der Kinder ihrer Gruppe wahrnimmt.
- Die Merkmalskomponente (*Stereotype Accuracy*) bringt zum Ausdruck, wie sehr die um das generelle Urteilsniveau bereinigten mittleren Merkmalseinschätzungen von den entsprechenden Kriteriumswerten abweichen (Schrader, 1989, S. 76) und beschreibt damit die Fähigkeit der Lehrperson, spezifische Aufgabenschwierigkeiten korrekt einzuschätzen (vgl. Karst, 2012, S. 44).
- Die Wechselwirkungskomponente (*Differential Accuracy*) zeigt die Genauigkeit des Lehrpersonenurteils unter Kontrolle der Haupteffekte von Personen und Merkmalen (Schrader, 1989, S. 78), also die Genauigkeit, mit der die Lehrperson die Unterschiede zwischen den Kindern für jedes Item einschätzt.

3.1.3 Komponentenanalyse nach Schrader (1989)

Eine Vielzahl von Studien im deutschsprachigen Raum gründet auf dem vereinfachten Modell nach Schrader und Helmke (z.B. Artelt, Stanat, Schneider & Schiefele, 2001, S. 99; Spinath, 2005; Thonhauser, 2005, S. 108). Diese ziehen als Indikator für diagnostische Kompetenz von Lehrkräften den Grad an Übereinstimmung von von Urteilen Lehrender und tatsächlicher Ausprägung entsprechender Schüler(innen)merkmale heran (Schrader & Helmke, 1987) und verwenden diagnostische Kompetenz mittels dreier Akkuratheitskomponenten: Niveau, Differenzierungs- und Vergleichskomponente.

Schrader (1989, S. 86) unterscheidet die drei Dimensionen diagnostischer Urteile (personenbezogen, aufgabenbezogen, aufgabenspezifisch), innerhalb derer die einzelnen Komponenten (Niveau-, Differenzierungs-, Rangordnungskomponente) bestimmt werden.

- Die personenbezogene Leistungsschätzung zeigt, wie viele Aufgaben die einzelnen Schüler(innen) der Klasse lösen können, erfasst also die Leistung einer Schülerin oder eines Schülers in einem Merkmal.

- Die aufgabenbezogene Leistungsschätzung veranschaulicht, wie viele Schüler(innen) der Klasse eine Aufgabe (Helmke, Grommelt, Hoppe & Schrader, 1984) lösen können, erfasst also die Lösungshäufigkeit einer Aufgabe pro Klasse und damit die Aufgabenschwierigkeit.
- Die aufgabenspezifische Leistungsschätzung bringt zum Ausdruck, welche Schüler(innen) eine konkrete Aufgabe lösen können, erfasst also die Fähigkeit, Aufgabenprofile von Schülerinnen und Schülern zutreffend einzuschätzen.

Die personen- und die aufgabenbezogene Dimension operationalisiert Schrader (1989, S. 87) in Anlehnung an Cronbach (1955) mittels Niveau-, Differenzierungs- und Rangordnungskomponente. Seine theoretische Modellierung diagnostischer Kompetenz von 1989 lässt sich in Anlehnung an Karst (2012, S. 31) wie folgt grafisch darstellen.

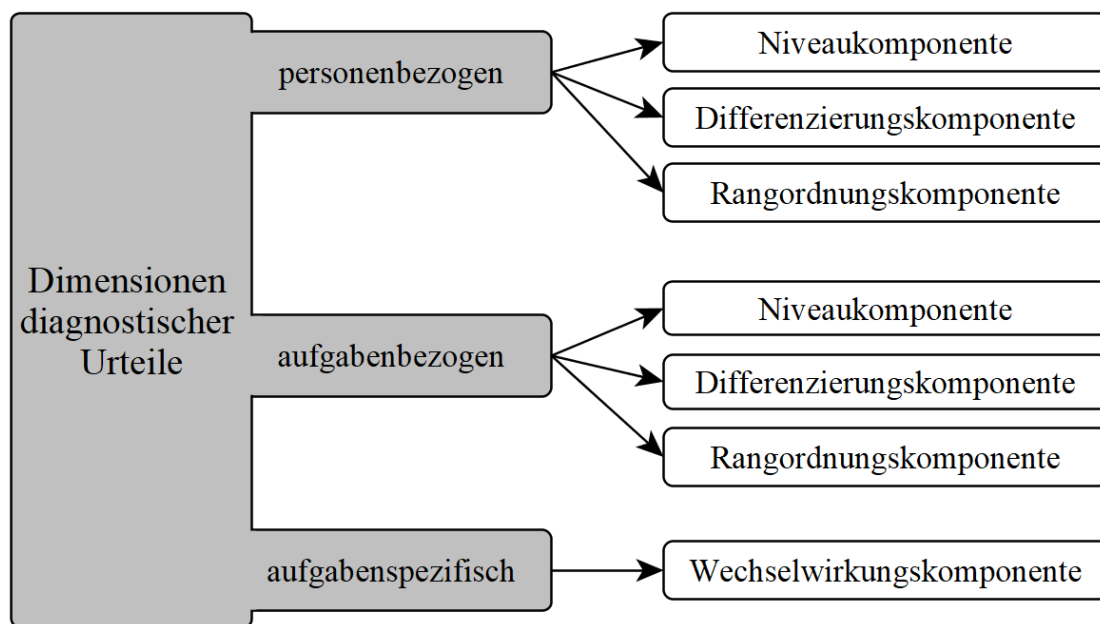


Abbildung 7: Theoretische Modellierung diagnostischer Kompetenz nach Schrader (1989, Darstellung in Anlehnung an Karst, 2012, S. 31)

Die Bedeutung der Komponenten verdeutlicht die folgende tabellarische Übersicht von Karst (2017a, S. 22) (Tab. 2).

Tabelle 2: Überblick über die Komponenten diagnostischer Kompetenz nach Schrader (1989) und ihre Bedeutung (Karst, 2017a, S. 22)

Komponente diagnostischer Kompetenz	Art des Urteils	
	aufgabenbezogen „Wie viele Lernende werden diese Aufgabe lösen?“	personenbezogen „Wie viele Aufgaben werden vom Schüler / von der Schülerin richtig gelöst?“
	Akkuratheit der Einschätzung	
Niveau $\bar{x} - \bar{y}$... der mittleren Aufgabenschwierigkeit	... der mittleren Schülerleistung
Differenzierung σ_y / σ_x	... der Heterogenität von Aufgabenschwierigkeiten	... der Heterogenität von Schülerleistungen
Rangordnung r_{xy}	... der Rangreihe der Aufgabenschwierigkeiten	... der Rangreihe der Schülerleistungen

Anmerkung: \bar{y} = mittleres Lehrerurteil; \bar{x} = mittlere Schülerleistung; σ_y = Streuung des Lehrerurteils; σ_x = Streuung der Schülerleistung; r_{xy} = Produkt-Moment-Korrelation zwischen Lehrerurteil und Schülerleistung.

Die genannten Bestandteile diagnostischer Kompetenz sind nach Schrader (1989, S. 239) weitgehend voneinander unabhängige Bestandteile und kein globales Maß.

3.1.4 Erweiterung nach McElvany et al. (2009)

Eine Erweiterung des Ansatzes von Schrader durch McElvany et al. (2009) mittels einer Systematisierung relevanter Urteilsbereiche und Urteilebenen gemäß Tabelle 3 stellt Karst (2017a) zufolge eine hilfreiche Taxonomie zur Verortung diagnostischer Urteile zur Verfügung und erleichtert damit den Vergleich zwischen gleichen und unterschiedlichen diagnostischen Kompetenzen in ihrer prädiktiven Kraft für die Lernentwicklung Lernender (ebd., S. 24).

Tabelle 3: Systematisierung relevanter Urteilsbereiche und Urteilebenen nach McElvany et al. (2009, S. 226)

	Urteilsbereich	
	Schüler z.B. Kompetenzen	Unterrichtsmaterial z.B. Anforderungen, Schwierigkeit
Urteilebene	individuelle Schüler Klasse übergreifend z.B. Klassenstufe / Schulform	einzelne Aufgaben Gesamttest übergreifend z.B. Themenbereiche

3.2 Berechnung der Urteilsgenauigkeit

Zur Berechnung der Urteilsgenauigkeit existieren zwei bedeutsame Varianten: die Berechnung der Komponenten nach Cronbach (1955) und die auf weniger Komponenten reduzierte Version nach Schrader (1989).

3.2.1 Urteilsgenauigkeitsmaße nach Cronbach

Im Folgenden wird die Beschreibung der Komponenten nach Cronbach (1955) aufgegriffen und mit der Berechnung ergänzt. Zudem werden die Berechnungen explizit auf Erzieherinnen bezogen. Anschließend erfolgt zum Vergleich eine Darstellung der Abwandlungen von Schrader (1989). Die am diagnostischen Prozess beteiligten Personen werden durch die Abkürzungen j (*judge* – hier: Erzieherin) und p (Person – hier: Kind), wobei das Kürzel i für Items steht.

Cronbachs (1955) Niveauelemente (Elevation) zeigt, wie exakt eine Erzieherin den mittleren Testwert bzw. das Niveau der Gruppe ihrer Bezugskinder einschätzen kann. Rechnerisch handelt es sich um die quadrierte Differenz zwischen dem mittleren Urteilswert einer Erzieherin und dem mittleren Testwert der Gruppe ihrer Bezugskinder.

$$E_j^2 = (y_{..j} - x_{..})^2$$

$y_{..j}$ Mittleres Urteil der Erzieherin j (judge)

$x_{..}$ Mittlerer Testwert der Bezugskindergruppe pro Erzieherin

Die Personenkomponente (Differential Elevation) veranschaulicht, wie die Erzieherin die Unterschiedlichkeit der Kinder ihrer Gruppe wahrnimmt. Sie gibt an, wie sehr die um das generelle Urteilsniveau bereinigten mittleren (d.h. über alle Merkmale gemittelten) Einschätzungen jedes Kindes im Durchschnitt von den entsprechenden mittleren Kriteriumswerten abweichen (Schrader, 1989, S. 76). Karst betont, dass bei der Berechnung dieser Komponente die absolute Übereinstimmung der jeweiligen Subtrahenden nicht relevant ist (2012, S. 45). Dies gilt auch für die beiden Minuenden. Hier ist vielmehr die Einschätzung des einzelnen Kindes im Verhältnis zum mittleren Erzieherinnenurteil von Bedeutung (ebd.)

$$DE_j^2 = \frac{1}{N} \sum [(\bar{y}_{p.j} - \bar{y}_{..j})(\bar{x}_p - \bar{x}_{..})]^2$$

N	Anzahl der Kinder
$y_{p.j}$	Mittleres Erzieherinnenurteil für Kind p der Erzieherin j
$y_{..j}$	Mittleres Urteil der Erzieherin j (judge)
x_p	Mittlerer Testwert des Kindes p in der Kindergruppe der Erzieherin j
$x_{..}$	Mittlerer Testwert der Bezugskindergruppe pro Erzieherin

Die Merkmalskomponente (Stereotype Accuracy) bringt zum Ausdruck, wie sehr die um das generelle Urteilsniveau bereinigten mittleren Merkmalseinschätzungen von den entsprechenden Kriteriumswerten abweichen (Schrader, 1989, S. 76) und beschreibt damit die Fähigkeit der Erzieherin, spezifische Aufgabenschwierigkeiten korrekt einzuschätzen (Karst, 2012, S. 44) bzw. die Genauigkeit, mit der verschiedene Merkmale eingeschätzt werden können (Kammermeyer, 2000, S. 92).

$$SA_j^2 = \frac{1}{k} \sum [(\bar{y}_{.ij} - \bar{y}_{..j})(\bar{x}_{.i} - \bar{x}_{..})]^2$$

k	Anzahl der Items
$y_{.ij}$	Mittleres Erzieherinnenurteil für Item i der Erzieherin j
$y_{..j}$	Mittleres Urteil der Erzieherin j (judge)
$x_{.i}$	Mittlerer Testwert der Bezugskindergruppe pro Erzieherin bei Item i
$x_{..}$	Mittlerer Testwert der Bezugskindergruppe pro Erzieherin

Die Wechselwirkungskomponente (Differential Accuracy) zeigt die Genauigkeit des Erzieherinnenurteils unter Kontrolle der Haupteffekte von Personen und Merkmalen (Schrader, 1989, S. 78). Rechnerisch handelt es sich um die Differenz der Residuen von Erzieherinnenurteil und Residuen der Testwerte des Kindes.

$$DA_j^2 = \frac{1}{kN} \sum_p \sum_i (y'_{pij} - x'_{pi})^2$$

k	Anzahl der Items
N	Anzahl der Kinder
y'_{pij}	Residuum des Erzieherinnenurteils von Erzieherin j
x'_{pi}	Residuum der Testwerte der Kinder

Durch mathematische Umformung der Gleichungen für Personen-, Merkmals- und Wechselwirkungskomponente ist auch eine Darstellung möglich, die nur noch die Streuungen der beiden Einzelvariablen sowie die Korrelation zwischen beiden Variablen enthält (Schrader, 1989, S. 78). Diese Umformung stellt zugleich die Berechnung der zugehörigen Korrelationskomponenten dar.

Tabelle 4: Ergebnisse der mathematischen Umformung der Gleichungen für Personen-, Merkmals- und Wechselwirkungskomponente

Personenkomponente Differential Elevation (DE)	$DE_j^2 = \sigma_{y_{p,j}}^2 + \sigma_{x_p}^2 - 2\sigma_{y_{p,j}} \sigma_{x_p} r_{x_p,y_{p,j}}$
Merkmalskomponente Stereotype Accuracy (SA)	$SA_j^2 = \sigma_{y_{i,j}}^2 + \sigma_{x_i}^2 - 2\sigma_{y_{i,j}} \sigma_{x_i} r_{x_i,y_{i,j}}$
Wechselwirkungskomponente Differential Accuracy (DA)	$DA_{ij}^2 = \sigma_{y'_{pij}}^2 + \sigma_{x'_{pi}}^2 - 2\sigma_{y'_{pij}} \sigma_{x'_{pi}} r_{x'_{pi},y'_{pij}}$

$\sigma_{y_{p,j}}^2$	Varianz der mittleren Urteilswerte von Erzieherin j für die Gruppe ihrer Kinder
$\sigma_{x_p}^2$	Varianz der mittleren Testwerte der Bezugskindergruppe pro Erzieherin
$\sigma_{y_{i,j}}^2$	Varianz der mittleren Erzieherinnenurteile der Erzieherin j für die Items
$\sigma_{x_i}^2$	Varianz der Aufgabenschwierigkeiten in einer Gruppe pro Erzieherin
$\sigma_{y'_{pij}}^2$	Varianz des Residuums des Erzieherinnenurteils von Erzieherin j
$\sigma_{x'_{pi}}^2$	Varianz des Residuums der Testwerte der Kinder
$r_{x_p,y_{p,j}}$	DE_r = Korrelationskomponente der Differential Elevation = Korrelation von kindspezifischen Urteilen einer Erzieherin und den über die Items gemittelten Testwerten der Kinder
$r_{x_i,y_{i,j}}$	SA_r = Korrelationskomponente der Stereotype Accuracy = Korrelation von merkmalspezifischen Itemschwierigkeiten und den von den Erzieherinnen geschätzten Itemschwierigkeiten
$r_{x'_{pi},y'_{pij}}$	DA_r = Korrelationskomponente der Differential Accuracy = Korrelation der Residuen von Testwerten der Kinder und Erzieherinnenurteilen

Die Korrelationskomponente der Personenkomponente (Differential Elevation, $DE_r = r_{x_p,y_{p,j}}$) zeigt die Kompetenz der Erzieherin, ihre Bezugskinder entsprechend ihrer Leistungen in eine Rangreihenfolge zu bringen (Karst, 2012, S. 46). Diese Korrelationskomponente stellt das Verhältnis zwischen den bezugskindspezifischen Urteilen einer Erzieherin und den über die Items gemittelten Testwerten der Kinder dar.

In der Korrelationskomponente der Merkmalskomponente (Stereotype Accuracy, $SA_r = r_{x_i,y_{i,j}}$) kommt insbesondere der Aspekt des fachdidaktischen Wissens zum Ausdruck, nämlich die Fähigkeit, Aufgaben nach ihrer Schwierigkeit zu sortieren (vgl. ebd., S. 47). Berechnet wird die Korrelation von merkmalspezifischen Itemschwierigkeiten und den von den Erzieherinnen geschätzten Itemschwierigkeiten.

Die Korrelationskomponente der Wechselwirkungskomponente (Differential Accuracy, $DA_r = r_{x'_{pi},y'_{pij}}$) beinhaltet die korrelierten Residuen, welche als Zusammenhang zwischen den von der Erzieherin eingeschätzten bezugskindspezifischen Aufgabenprofilen und den gemessenen Aufgabenprofilen einzelner Kinder zu interpretieren ist (ebd., S. 47).

Die globale Urteilsgenauigkeit (Accuracy) als Summenwert der Einzelkomponenten Niveau-, Personen-, Merkmals- und Wechselwirkungskomponente bildet die Gesamtheit der Komponenten diagnostischer Kompetenz in einem Wert ab.

$$ACC_j^2 = E_j^2 + DE_j^2 + SA_j^2 + DA_j^2$$

Da es sich bei den Urteilsgenauigkeitskomponenten Cronbachs (1955) um Genauigkeitsmaße handelt, steht eine Ausprägung von null für eine genaue Einschätzung des Kindes durch die Erzieherin. Je ungenauer die Einschätzung einer Erzieherin ist, desto größer sind die Ausprägungen der Komponenten, d.h. desto weiter sind die Werte der Komponenten von null entfernt. Dabei sind theoretisch Werte von null bis unendlich möglich (ebd.). Dies ist ein zentraler Unterschied zu den Merkmalen von Schrader (1989).

3.2.2 Urteilsgenauigkeitsmaße nach Schrader

Schrader (1989, S. 89) beschränkt sich hinsichtlich der aufgabenspezifischen Leistungsschätzung auf die Bildung der Wechselwirkungskomponente aufgrund der Tatsache, dass ihm für die reliable Erfassung dieser aufgabenspezifischen Leistungsschätzung keine hinreichende Zahl von Urteilen vorlag.

Tabelle 5 zeigt beispielhaft an den personenbezogenen Komponenten Schraders (1989) Berechnung der Komponenten.

Tabelle 5: Personenbezogene Komponenten nach Schrader (1989)

Komponente	Formel	Interpretation der Ergebniswerte
Personenbezogene Niveauelemente	$\hat{y} - \hat{x}$	> 0 Überschätzung des Leistungsniveaus < 0 Unterschätzung des Leistungsniveaus = 0 exakte Schätzung des Leistungsniveaus
Personenbezogene Differenzierungskomponente	$\frac{s_y}{s_x}$	> 1 Überschätzung der Leistungsstreuung < 1 Unterschätzung der Leistungsstreuung = 1 exakte Schätzung der Leistungsstreuung
Personenbezogene Rangordnungs- komponente	r_{yx}	Ausmaß zutreffender Einschätzung der Rangordnung

Anmerkungen: \hat{y} = die mittlere Einschätzung für die Stichprobe der Beurteilten (Klasse); \hat{x} = die mittlere Testleistung in der Stichprobe der Beurteilten; s_y = die Streuung der Einschätzungen in der Stichprobe der Beurteilten; s_x = die Streuung der Testleistungen in der Stichprobe der Beurteilten; r_{yx} = Produktmoment-Korrelation zwischen Einschätzungen und Leistungen.

Cronbach und Schrader nutzen unterschiedliche Varianten der Berechnung und betonen damit unterschiedliche Perspektiven auf die Urteile von Lehrenden. So setzt Schrader (1989) z.B. die Berechnung seiner Niveauekomponente nicht zum Quadrat, was den Vorteil hat, dass Über- und Unterschätzungen von Leistungen durch Werte über und unter Null zu erkennen sind. Cronbach (1955) verleiht höheren Fehleinschätzungen wesentlich mehr Gewicht, weil er den Wert der Niveauekomponente mit Zunahme von Fehleinschätzungen exponentiell anwachsen lässt. Problematisch ist, dass die Komponenten beider Autoren in unterschiedlichen Studien unter gleicher Bezeichnung geführt werden, obgleich sich die Aussagen wie dargelegt unterscheiden.

Die Fokussierung einer Vielzahl von Forschungsarbeiten auf die von Schrader (1989, S. 119) als diagnostische Kompetenz im engeren Sinne bzw. von Helmke (2015, S. 134) als Kernstück diagnostischer Kompetenz bezeichnete Rangordnungskomponente wird im Hinblick auf adaptive Förderung (Bruns, 2014, S. 170) kritisch bewertet. Auch Karst (2012, S. 36) sieht dieses Vorgehen als Rückschritt. Zudem steht diese Beschränkung auf nur eine Komponente der diagnostischen Kompetenz im Widerspruch zu Schraders (1989) eigener Forderung nach einer differenzierteren Erfassung der diagnostischen Kompetenz (S. 270). Zwar bildet die Leistungsreihenfolge den ersten Schritt in der adaptiven Planung des Alltags, für eine angemessene Förderung in der Zone der nächsten Entwicklung (vgl. Kap. 2.6.1) ist aber als zweiter Schritt das Setzen individueller Lernziele für das Kind erforderlich (Hattie, 013, S. 235).

3.3 Methodische Merkmale zur Messung von Urteilsgenauigkeit

Die Messung der Urteilsgenauigkeit kann auf unterschiedliche Weise erfolgen. Hoge und Coladarci (1989, S. 300–302) unterscheiden fünf methodische Merkmale: Beurteilung (direkt vs. indirekt), Urteilspezifität (spezifisch vs. unspezifisch), Bezugsnormorientierung (normbezogen vs. peer-unabhängig bzw. sozial vs. kriterial), Art der Messung (relativ vs. absolut) und Analyseinheit (gepoolt vs. hierarchisch). Für die Belastbarkeit der Ergebnisse ist es wichtig, auf welche Art die Urteilsgenauigkeit erhoben wurde, wie die folgenden Gegenüberstellungen unterschiedlicher methodischer Merkmale zeigen.

3.3.1 Beurteilung: *direkt* vs. *indirekt*

In den Untersuchungen mit direkten Urteilen schätzen Lehrer Items ein, welche auch den Schülerinnen und Schülern vorgelegt wurden; in den Untersuchungen mit indirekten Urteilen hingegen schätzen Lehrer ihre Schüler(innen) ein, ohne die zur Leistungsmessung eingesetzten Tests zu kennen. Insgesamt werten Hoge und Coladarci (1989) zehn Studien zu indirekter Beurteilung und sieben Studien zu direkter Beurteilung aus, wobei eine Studie von Wright und Wiese (1988) sowohl direkte als auch indirekte Urteilsformen aufweist und somit doppelt berücksichtigt wird.

Als Beispiel für eine Studie zur diagnostischen Kompetenz mit direkten Urteilen (direct judgments) nennen Hoge und Coladarci (1989, S. 300) die Studie von Helmke und Schrader (1987). Die Ergebnisse dieser Studie beruhen auf Urteilen Lehrender, die anhand der Items ermittelt werden, mit denen auch die Leistungen von Schülerinnen und Schülern erfasst werden. Hierzu benennen Lehrkräfte eine Anzahl von für ihre Schüler(innen) lösbaren Aufgaben. Die Leistung der Schüler(innen) wird mit denselben Aufgaben erhoben, weshalb Urteil und Kriterium direkt vergleichbare Skalen zugrunde gelegt sind.

Als Beispiel für eine indirekte Beurteilung führen Hoge und Coladarci (1989, S. 300) eine Studie von Airasian, Kellaghan, Madaus und Pedulla (1977) an, in der Lehrkräfte auf einer fünfstufigen Skala die Fähigkeitsniveaus ihrer Schüler(innen) in Englisch und Mathematik einschätzen; anschließend werden die Urteile der Lehrkräfte mit den Leistungswerten der Schüler(innen) verglichen, welche mit einem standardisierten Test erhoben wurden. Lehrereinschätzung (Urteil) und Schülerwerte (Kriterium) beziehen sich zwar auf ein gemeinsames Konstrukt, sie werden jedoch unterschiedlich operationalisiert und daher auf unterschiedlichen Skalen erfasst. Der Zusammenhang besteht folglich nur indirekt.

Insgesamt stellen Hoge und Coladarci (1989, S. 303) für direkte und indirekte Urteile eine mittlere Urteil-Kriterium-Korrelation von $r = .66$ fest, die als moderat bezeichnet wird. Eine Aufteilung der Studien nach Urteilsformen (direkt/indirekt) macht deutlich, dass Untersuchungen mit direkten Urteilen ($r = .69$) den Studien mit indirekten Urteilen ($r = .62$) nur knapp überlegen sind (ebd., S. 305).

Studien mit direkten Urteilen zeichnen sich nach Hoge und Coladarci (1989, S. 300) durch einen stärkeren logischen Zusammenhang zwischen Urteil und Kriterium aus.

Südkamp, Kaiser und Möller (2012) schließen mit ihrer mehrebenenanalytisch ausgewerteten Metaanalyse zeitlich an die von Hoge und Coladarci (1989) an. Während Letztere und indirekte Urteilsformen unterscheiden, wählen Südkamp et al. (2012) die Unterscheidung zwischen Urteilen mit Testkenntnis (*informed*) und ohne Testkenntnis (*uninformed*). Sie sehen den Hauptunterschied zwischen direkten und indirekten Urteilen darin, dass Lehrkräfte über den Test oder den Vergleichsstandard, auf dem ihre Beurteilung beruht (ebd., S. 3), entweder informiert oder nicht informiert sind. Ihre Analysen von 75 Studien zum Zusammenhang von Urteilen Lehrender und Leistungen von Schüler(innen) aus den Jahren 1989 bis 2010 legen den Fokus auf theoretisch und methodisch relevante Variablen, die einen Einfluss auf die Urteilsgenauigkeit von Lehrern haben. In ihrer Metaanalyse werden zwei Moderatorvariablen signifikant: Testkenntnis und Übereinstimmung im Spezifitätsgrad von Urteil und Leistung und sie berichten eine mittlere Effektstärke von .63 (Südkamp et al., 2012, S. 8). Damit bestätigen Sie die Metaanalyse von Hoge und Coladarci (1989, S. 303). Zudem finden Südkamp et al. (2012) in Übereinstimmung mit den Ergebnissen von Hoge und Coladarci (1989), Feinberg und Shapiro (2003, 2009) und von Demaray und Elliott (1998) höhere Korrelationen bei direkten Urteilen (meist *informed* bzw. mit Testkenntnis). Auch zeigen sich höhere Korrelationen zwischen schulischen Leistungen der Schüler(innen) und Urteilen von Lehrkräften mit Testkenntnis (mittlere Effektstärke von .76) als Urteile von Lehrkräften ohne Testkenntnis (.61) (Südkamp et al., 2012, S. 10). In 86.8% der von Südkamp et al. (2012) einbezogenen Studien liegen Urteile ohne Testkenntnis vor, in lediglich 13.2% Urteile mit Testkenntnis.

Getrennt nach sprachlicher und mathematischer sowie nach spezifischer (*direct*) und globaler (*indirect*) Einschätzung berechnet Lorenz (2011, S. 52–53) Korrelationen auf Basis der Metaanalysedaten von Hoge und Coladarci (1989), um sowohl den jeweils zugrunde liegenden Leistungsbereichen als auch der Art der abgegebenen Urteile Rechnung zu tragen. Er berichtet die Tendenz, dass eine spezifische Leistungseinschätzung, bei der Lehrer die Leistung in konkreten, ihnen bekannten Aufgaben einschätzen, sowohl für den sprachlichen als auch für den mathematischen Leistungsbereich genauer ausfallen als bei einer globalen Einschätzung, wo nur allgemeine Urteile zur Leistungsfähigkeit abgegeben werden sollten (Lorenz, 2011, S. 53).

Zusammenfassend ist festzuhalten, dass die direkte Erfassung mit Testkenntnis die geeignetere Operationalisierung darstellt, weil Schüler(innen)merkmale und Urteile von Lehrkräften mit gleichen Skalen erhoben werden, damit in einem logischen Zusammen-

hang stehen und direkt vergleichbar sind. Zudem ist die Urteil-Kriterium-Korrelation in Studien mit direkten Leistungseinschätzungen etwas höher.

3.3.2 Urteilspezifität: *spezifisch vs. unspezifisch*

Die Unterscheidung zwischen direkten und indirekten Urteilen hat auch Auswirkungen auf die Urteilspezifität. Hoge und Coladarci (1989, S. 300) stellen fest, dass direkte Maße zur Beurteilung von Lehrkräften definitionsgemäß spezifischer sind, da indirekte Maße nicht ausdrücklich an ein Kriterium im Urteilsprozess gebunden sind, obwohl der Zusammenhang nicht vollkommen konsistent ist.

Tabelle 6: Urteilsmaße nach Hoge und Coladarci (1989, S. 300–306)

Spezifitätsgrad	Bezeichnung der Urteilsmessung	Messmethodische Operationalisierung des Urteils	r
Niedriges Maß an Genauigkeit	[1] Einschätzungen	Einschätzung von Leistungsgruppen, z.B. anhand einer fünfstufigen Skala („niedrigstes ...“ bis „höchstes Fünftel der Klasse“)	.61
	[2] Ranglisten	Bildung einer Schülerrangreihe geschätzter Leistungen, z.B. hinsichtlich des Abschneidens in standardisiertem Test	.76
	[3] Notenäquivalenz	Leistungseinschätzung jedes einzelnen Schülers anhand einer notenähnlichen Skala	.70
	[4] Anzahl korrekter Antworten	Leistungseinschätzung durch Benennen der Gesamtanzahl richtig gelöster Aufgaben	.67
Hohes Maß an Genauigkeit	[5] Benennung korrekter Items	Leistungseinschätzung durch Benennen der einzelnen Items, welche die Schüler zu lösen imstande sind	.70

Die Autoren ordnen die Studien ihrer Metaanalyse anhand der Spezifitätsgrade fünf Urteilsmaßen zu, die in Tabelle 6 dargestellt sind.

Die Analysen zeigen, dass Lehrer insbesondere durch Ranglisten (Rankings) eine hohe Urteil-Kriterium-Korrelation ($r = .76$) erreichen. Einschätzungen (Ratings), die von Hoge und Coladarci (1989, S. 306) als die vorherrschende Form der Lehrendenurteile bezeichnet werden, zeigen die kleinste Urteils-Kriteriums-Korrelation von $r = .61$.

Die Übereinstimmung im Spezifitätsgrad von Urteil und Leistung als signifikante Moderatorvariable identifizieren auch Südkamp et al. (2012, S. 13) in ihrer mehrbenenanalytisch ausgewerteten Metaanalyse. Sie unterscheiden übereinstimmende Spezifi-

tätsgrade von Leistungstests und Urteilsaufgabe (67.6% aller berichteten Effektmaße) von nicht übereinstimmenden Spezifitätsgraden von Urteil und Leistung (32.4% aller berichteten Effektmaße). Größere Effektstärken werden dabei in Studien beobachtet, in denen die Domänenspezifität des Leistungstests und der Ratingaufgabe übereinstimmend war ($r = .67$), als in Studien, in denen dies nicht der Fall war ($r = .54$).

Insgesamt sind Verfahren mit höherer Spezifität (als z.B. bei Einschätzungen) für die Operationalisierung der diagnostischen Kompetenz geeigneter, weil direkte Maße im Gegensatz zu indirekten definitionsgemäß ausdrücklich an ein Kriterium im Urteilsprozess gebunden sind und die Urteil-Kriterium-Korrelation höher ist.

3.3.3 Bezugsnormorientierung: *sozial (normbezogen) vs. kriterial (peer-unabhängig)*

Ein normbezogenes Urteil liegt vor, wenn Lehrkräfte ihre Schüler(innen) beispielsweise anhand einer fünfstufigen Skala einschätzen (Hoge & Coladarci, 1989, S. 302). Diese Gruppierung kann nur durch Vergleiche der Schüler(innen) zustande kommen. Wenn sie hingegen, wie bei Schrader und Helmke (1987), eine Anzahl von für ihre Schüler(innen) lösbaren Aufgaben benennen, liegt ein peerunabhängiges Urteil vor, weil dieses keinen Vergleich der Schüler(innen) erfordert (Hoge & Coladarci, 1989). Die Untersuchung der Urteil-Kriterium-Korrelation zeigt jedoch keine großen Unterschiede zwischen peerunabhängigen Ratings ($r = .68$) und Untersuchungen mit normbezogenen Urteilen ($r = .64$) (ebd., S. 306).

Zur Operationalisierung diagnostischer Kompetenz sind hinsichtlich der Bezugsnormorientierung peer-unabhängige Urteile vorzuziehen, weil diese Form der Messung in unterschiedlichen Gruppen unter Berücksichtigung der kriterialen Bezugsnorm zu vergleichbareren Ergebnissen führt als Messungen, die sich an der sozialen Bezugsnorm orientieren, was die Vergleichbarkeit zwischen unterschiedlichen Gruppen (z.B. Klassen oder Kita-Gruppen) einschränken würde.

3.3.4 Art der Messung: *relativ vs. absolut*

Die Genauigkeit von Urteilen Lehrender wird durch die Urteil-Kriterium-Korrelation ausgedrückt, welche durch zwei Maße entscheidend beeinflusst wird: die Beurteilung der Schüler(innen) durch die Lehrpersonen und die aktuelle Leistung der Schüler(innen) im relevanten standardisierten Test (Hoge & Coladarci, 1989, S. 302). Grundlage hier-

für sind eine Vielzahl von messmethodischen Operationalisierungen der Urteilsgenauigkeit: Einschätzungen, Ranglisten, Notenäquivalenz, Anzahl korrekter Antworten oder Benennung korrekter Antworten (ebd.). Die meisten der von Hoge und Coladarci berücksichtigten Studien ihrer Metaanalyse berichten relative Maße in Form von Korrelationen oder ähnlichen Regressionskoeffizienten als Kennwerte für die Urteilsgenauigkeit.

Als Alternative zu diesen relativen Maßen der Urteilsgenauigkeit liefern Coladarci (1986) und Leinhardt (1983) ein absolutes Maß für die Urteilsgenauigkeit, das Hoge und Coladarci als „a total teacher judgement“ bezeichnen (1989, S. 302). Dieses absolute Maß, das auf einer Trefferquote basiert, die Leinhardt (1983) auch als „hit-rate“ bezeichnet, wird durch die Anzahl der Übereinstimmungen zwischen Urteil und Kriterium auf Itemebene abgebildet.

Diese absoluten Maße sind zur Operationalisierung diagnostischer Kompetenz vorzuziehen, weil somit Urteile im Hinblick auf eine Erfassung diagnostischer Kompetenz Lehrender in verschiedenen Gruppen anhand gleicher Aufgaben, wie z.B. anhand der Items eines standardisierten Tests, gewonnen und somit durch dieselben zugrunde liegenden Skalen ähnliche Voraussetzungen für einen späteren Vergleich geschaffen werden.

3.3.5 Analyseeinheit: *gepoolt vs. hierarchisch*

Die Analyseeinheiten können in gepoolte und hierarchische Analyseeinheiten unterschieden werden (Hoge & Coladarci, 1989). Gepoolte Analyseeinheiten finden sich in Studien, in denen Schüler(innen) und Lehrkräfte zu jeweils einer Gruppe zusammengefasst werden. Somit wird die Gruppenzugehörigkeit ignoriert, was zu einer Über- oder Unterschätzung des Urteil-Kriterium-Zusammenhangs führen kann. (Hoge & Coladarci, 1989, S. 302). In Untersuchungen mit gepoolten Gruppen liegt die Urteil-Kriterium-Korrelation bei $r = .64$ (ebd., S. 306).

Die hierarchische Struktur der Daten wird in vier Studien der Metaanalyse von Hoge und Coladarci (ebd.) berücksichtigt, welche die Urteil-Kriterium-Korrelation klassenweise berechnen. In diesen liegt die Urteil-Kriterium Korrelation bei $r = .70$.

Zur Operationalisierung diagnostischer Kompetenz sind hierarchische Analyseeinheiten vorzuziehen, weil diese zum einen tendenziell höhere Urteil-Kriterium-Korrelationen

aufweisen und zum anderen die Zugehörigkeit zu Gruppen (anfallende Klumpenstichproben) nicht ignorieren.

3.3.6 Zusammenfassung

Insgesamt ergeben sich bei der Messung der Urteilsgenauigkeit gemäß den fünf methodischen Kriterien von Hoge und Coladarci (1989, S. 300–302) klare Anforderungen.

Folgende Merkmale, die sich zur Erfassung der Urteilsgenauigkeit als geeignet erwiesen haben, bilden ein Anforderungsprofil:

- Operationalisierung der Urteilsgenauigkeit mittels direkter Beurteilungen und Testkenntnis (3.3.1)
- Verwendung von Verfahren höchster Spezifität (3.3.2)
- Domänenspezifische Erfassung (3.3.2)
- Berücksichtigung peer-unabhängiger Urteile (3.3.3)
- Ermittlung auf Itemebene (3.3.4)
- Berücksichtigung der hierarchischen Struktur (Klassen / Bezugsgruppen) (3.3.5)

Diese Merkmale werden jedoch in der Forschung zu Urteilsgenauigkeit von Erzieherinnen nicht erfüllt und in den Forschungen im schulischen Kontext nur zum Teil, wie das folgende Kapitel u.a. ausführt.

3.4 Forschungsstand zur diagnostischen Kompetenz

Die Forschung zur diagnostischen Kompetenz bezieht sich vornehmlich auf den schulischen Bereich, die Forschungen zur Diagnosekompetenz von Erzieherinnen bauen auf diesen auf.

Die meisten Studien der Diagnosekompetenzforschung fokussieren kognitive Leistungsmerkmale von Kindern und Jugendlichen. Dabei wird Diagnosekompetenz in der Regel als Urteilsgenauigkeit verstanden. Als Maß dient zumeist die Korrelation zwischen Einschätzung der Rangordnung von Leistungen durch die Lehrkraft und tatsächli-

cher Testleistung. Zwei zentrale Metaanalysen fassen die Erkenntnisse dieses Forschungszweigs bis 2010 zusammen und zeigen wesentliche Tendenzen auf. Dabei handelt es sich um die Analyse *Teacher-Based Judgments of Academic Achievement: A Review of Literature* von Hoge und Coladarci (1989), welche 16 Studien aus den Jahren 1962 bis 1988 berücksichtigt, und die Analyse von Südkamp et al. (2012) mit dem Titel *Accuracy of Teachers' Judgments of Students' Academic Achievement: A Meta-Analysis*, in die empirische Befunde aus 75 Arbeiten der Jahre 1989 bis 2010 eingegangen sind. Eine umfassende Darstellung des aktuellen Forschungsstands zur Erfassung diagnostischer Kompetenz im schulischen Bereich liefert zudem das von der Deutschen Forschungsgemeinschaft (DFG) geförderte wissenschaftliche Netzwerk zur diagnostischen Kompetenz von Lehrkräften (NeDiKo) mit der von Südkamp und Praetorius (2017) herausgegebenen Zusammenfassung der Netzwerkarbeit.

Auf Basis der bisherigen Diagnosekompetenzforschung und theoretischer Überlegungen entwickelte NeDiKo das *Arbeitsmodell für die empirische Erforschung der diagnostischen Kompetenz von Lehrkräften* (Herppich et al., 2017), welches mit dem Ziel konzipiert wurde, als eine integrierende Basis für zukünftige Theoriebildung und empirische Forschung zu fungieren (ebd., S. 75). Dieses bildet den aktuellen Stand der Forschung zu diagnostischer Kompetenz ab (vgl. Kap. 2.6.5 in der vorliegenden Arbeit).

Eine Einteilung der Forschungslinien in folgende vier Herangehensweisen stammt von Karst und Förster (2017):

1. Diagnostische Kompetenz operationalisiert als Akkuratheit von Lehrerurteilen (Karst, 2017a; McElvany et al., 2009; Schrader, 1989, 2006)
2. Diagnostische Kompetenz mit expliziter Berücksichtigung eines (bildungswissenschaftlichen) Kompetenzbegriffs (Karst, 2012, 2017b; Ohle et al., 2017)
3. Diagnostische Kompetenz unter Berücksichtigung weiterer Merkmale (Behrmann & van Ophuysen, 2017; Südkamp et al., 2012, 2017; van Ophuysen & Lintorf, 2013)
4. Diagnostische Kompetenz unter Berücksichtigung des Prozesscharakters der Urteilsbildung (Behrmann & Kaiser, 2017; Böhmer, English & Böhmer, 2017; Förster & Böhmer, 2017; Klug, 2017)

Im Folgenden wird der Forschungsstand anhand dieser Einteilung für die ersten drei Forschungslinien dargelegt. Die Forschung zur diagnostischen Kompetenz unter Berücksichtigung des Prozesscharakters wird in Kapitel 4 Beobachtungskompetenz ausgeführt.

3.4.1 ... als Akkuratheit

In den meisten Arbeiten zur diagnostischen Kompetenz wird diese als Akkuratheit im Sinne einer akkuraten Einschätzung der Leistungen von Schülerinnen und Schülern operationalisiert; diesen Arbeiten liegt die Niveau- und die Differenzierungs- oder auch nur die Rangordnungskomponente zugrunde (Karst & Förster, 2017). Die Komponentenanalyse nach Cronbach (1955) erfuhr in der deutschsprachigen Diagnosekompetenzforschung insbesondere durch die Forschungsarbeit von Schrader und Helmke (Schrader, 1989; Helmke & Schrader, 1987; Schrader & Helmke, 1987) Beachtung und wird seitdem zumeist in einer vereinfachten Form verwendet, bei der eine Urteilerin oder ein Urteiler (z.B. eine Lehrperson) mehrere Personen (z.B. die Schüler(innen) der eigenen Klasse) hinsichtlich eines einzigen Merkmals (statt mehrerer wie bei Cronbach, 1955) beurteilt (Schrader, 2013, S. 157; vgl. auch Karst, 2012). Zumeist sind Studien, welche Cronbachs Komponentenanalyse angewandt oder die von Schrader adaptierten Komponenten Cronbachs verwendet haben, im schulischen Kontext entstanden. Bisherige Befunde zur Akkuratheit von Lehrendenurteilen (Karst, 2012; Schrader, 1989; Spinath, 2005; Südkamp, Möller & Pohlmann, 2008), welche auf Cronbach (1955) zurückgehen, zeigen, dass die Komponenten innerhalb einer Urteilsdimension (aufgabenbezogen, aufgabenspezifisch, personenbezogen), aber auch zwischen den Urteilsdimensionen nur schwach bis gar nicht korrelieren (Karst, 2017a, S. 23).

Neben diesen Studien zur Akkuratheit von Urteilen Lehrender liegen kaum Ergebnisse zur Akkuratheit von Erzieherinnenurteilen vor. Lediglich drei Arbeiten im deutschsprachigen Raum setzen die Komponentenanalyse bzw. einzelne adaptierte Komponenten zur Erfassung der diagnostischen Kompetenz von Fachkräften im vorschulischen Bereich ein. Kammermeyer (2000) untersucht die diagnostische und prognostische Kompetenz von Lehrkräften und Erzieherinnen hinsichtlich der Einschätzung ausgewählter Schulfähigkeitskriterien. Dollinger (2013) greift dieses Thema 13 Jahre später auf und untersucht die Diagnosegenauigkeit von Erzieherinnen und Lehrkräften hinsichtlich der Einschätzung schulrelevanter Kompetenzen in der Übergangsphase. Darüber hinaus untersucht Bruns (2014) einzelne Komponenten, die auf Cronbachs Komponentenana-

lyse zurückgehen, welche sie unter dem Begriff diagnostische Fähigkeiten von frühpädagogischem Fachpersonal zusammenfasst. Diese drei Studien werden im Folgenden ausführlicher dargestellt (Bruns, 2014, S. 132).

In den Studien von Kammermeyer (2000) und Dollinger (2013) basieren die Berechnungen auf Cronbachs Komponentenanalyse. Kammermeyer ließ 18 Erzieherinnen insgesamt 94 Kinder hinsichtlich der Schulfähigkeitskriterien Denkfähigkeit, Feinmotorik, Gedächtnis, Gliederungsfähigkeit und Konzentration anhand des Mannheimer Schuleingangsdagnostikums von Jäger, Beetz, Erler und Habersang-Walther (1994) einschätzen. Ihre Berechnung einer Komponentenanalyse nach Cronbach (1955) zeigt, dass sich globale Urteilsgenauigkeit wie folgt aufteilt: 13% des Summenwertes sind auf die Niveauebene, 23% auf die Merkmals-, 25% auf die Personen- und 39% auf die Wechselwirkungskomponente zurückzuführen (Kammermeyer, 2000, S. 209). Sie schlussfolgert, dass die Differenzen zwischen Vorhersage- und Testwerten gleichermaßen von den untersuchten Merkmalen und von den eingeschätzten Kindern abhängen und das Leistungsniveau ziemlich gut eingeschätzt werden kann (ebd., S. 209–210). Zudem zeigt die Untersuchung äußerst geringe Korrelationen zwischen den einzelnen Komponenten ($r = -.09$ bis $r = .27$; ebd., S. 209) mit Ausnahme der Korrelation zwischen Person- und Wechselwirkungskomponente ($r = .46$). Die Autorin sieht die Korrelationen zwischen den einzelnen Komponenten der globalen Diagnosegenauigkeit als Beleg dafür, dass die Komponenten der Diagnosegenauigkeit unabhängig voneinander sind (ebd., S. 210). Diese Unabhängigkeit oder äußerst geringen Zusammenhänge zeigen sich auch in anderen Studien, welche Zusammenhänge zwischen den Cronbach-Komponenten (Murphy, Garcia, Kerkar, Martin & Balzer, 1982; Sulsky & Balzer, 1988; Uggerslev & Sulsky, 2008) oder zwischen den von Schrader adaptierten Komponenten (Schrader, 1989; Spinath, 2005) berechnen.

Zu anderen Ergebnissen kommt Dollinger, die in ihrer Studie die Diagnosegenauigkeit von Erzieherinnen und Lehrkräften in der Übergangsphase (Dollinger, 2013, S. 21) untersucht. Sie stellt fest, dass Niveau- und Vergleichskomponenten entgegen den Erwartungen miteinander in Beziehung stehen, und geht aufgrund der nur teilweise vorhandenen Zusammenhänge von Niveau- und Vergleichskomponenten mit den jeweiligen Differenzierungskomponenten und -fehlern von zumindest zum Teil unabhängigen Kennwerten der Diagnosegenauigkeit aus (ebd., S. 124). Ihrer Studie zufolge verfügen pädagogische Fachkräfte über keine fachübergreifende Diagnosegenauigkeit, obgleich Ni-

veueinschätzung und Differenzierungsfähigkeit in den einzelnen Bereichen in einem Zusammenhang stehen (ebd., S. 174).

Dollinger stellt fest, dass Erzieherinnen und Lehrkräfte die Kompetenzen der Kinder in den Bereichen Lesen und Mathematik unterschätzen und in der phonologischen Bewusstheit und zum Teil im Wortschatz überschätzen (Dollinger, 2013, S. 174). Lehrkräfte zeigten im Hinblick auf die Rangordnungskomponente mehr Genauigkeit als Erzieherinnen, während letztere die Leistungsstreuung unterschätzten, die wiederum von Lehrkräften relativ exakt eingeschätzt werde (ebd.).

Eine Problematik der Studie liegt in der Belastbarkeit ihrer Ergebnisse, wie die Autorin selbst ausführt (Dollinger, 2013, S. 193). Aufgrund mangelnder Einverständniserklärungen der Eltern zur Teilnahme ihres Kindes an der Studie schätzen manche Erzieherinnen nur ein oder zwei Kinder ein. Zur methodischen Legitimation ihrer mehrerebenenanalytischen Auswertung führt sie vergleichend Berechnungen mit den Erzieherinnen durch, die fünf und mehr Kinder eingeschätzt haben (Dollinger, 2013, S. 193). Die Unterschiede der Ergebnisse bezeichnet sie als geringfügig. Während sich die Rangordnungskomponente in den einzelnen Bereichen denen der Lehrkräfte annähert, werden die Niveauelemente zum Großteil etwas exakter, allerdings bleiben die Über- und Unterschätzungen nach wie vor in allen Bereichen signifikant (ebd.).

Bruns (2014) wählt eine starke Vereinfachung der diagnostischen Kompetenz von Erzieherinnen zur Erfassung der diagnostischen Fähigkeiten von pädagogischen Fachpersonen, basierend auf Helmkes Niveau-, Streuungs- und Rangordnungskomponente (ebd., S. 41). Zwar vergleicht auch sie Testergebnisse von Kindern mit Einschätzungen des frühpädagogischen Fachpersonals, jedoch werden diesen nur sieben Leistungsstufen vorgelegt und keine einzelnen Testaufgaben (ebd., S. 124). Die maximal mögliche Abweichung von sechs Stufen wird durch die Testergebnisse der Kinder limitiert, wobei die Mehrheit der Kinder allerdings zwischen den Niveaustufen drei und fünf liegt. Entsprechend können Bruns zufolge die mittleren Differenzen zwischen der Einschätzung der Fachkraft und dem Testergebnis der Kinder, die eine bis drei Niveaustufen betragen, bereits als relativ hoch betrachtet werden (ebd., S. 161). Auf dieser Basis berechnet sie diagnostische Fähigkeiten auf zwei Weisen: die diagnostische Akkuratheit als Abweichung der Einschätzungen der Fachkräfte von den Testergebnissen der Kinder und die diagnostische Sensitivität als Korrelation zwischen der Rangreihenfolge der Kinder aufgrund der Einschätzung durch die Fachkraft und der Rangreihenfolge der Kinder auf-

grund des Testergebnisses (ebd., S. 132). Es zeigen sich bei den 31 untersuchten Fachkräften keine Zusammenhänge zwischen ihren diagnostischen Fähigkeiten und a) der Angemessenheit der Aktivitäten zu dem von der Fachperson angenommenen Lern- und Entwicklungsniveau der Kinder, b) der Vielfältigkeit des mathematischen Inputs im Alltag und c) der Differenz zwischen Testwerten zweier Zielkinder einer jeweiligen Beobachtung (ebd.). Allerdings zeigt sich ein Zusammenhang zwischen den Abweichungen der Einschätzungen der Fachkräfte von den Testergebnissen und den mittleren Testwerten der Kinder ihrer Gruppe. Die negative Korrelation zeigt Folgendes an: je höher die Abweichung der Einschätzung der Fachkräfte von den Testwerten der Kinder ausfällt, desto niedriger liegen die Punktzahlen der Kinder ihrer Gruppe im Mittel (ebd.); für die diagnostische Sensitivität kann dieser Zusammenhang nicht nachgewiesen werden.

Auch die Untersuchung des Zusammenhangs zwischen den beiden diagnostischen Fähigkeiten und dem Bildungsverständnis (Selbstbildung, Ko-Konstruktion, Direkte Instruktion) bleibt ohne signifikanten Befund (Bruns, 2014, S. 132–133). Trotz des groben 7-stufigen Rasters stellt Bruns fest, dass aufgrund der Fehleinschätzungen der Erzieherinnen 85% der Lernangebote mindestens eine Stufe von den Bedürfnissen der Kinder abweichen. Aus demselben Grund sei die Planung einer angemessenen Förderung nahezu unmöglich, so ihre Folgerung (ebd., S. 169).

Kritisch zu sehen ist das Konstrukt der diagnostischen Kompetenz als eine generelle Fähigkeit. Spinath (2005) befürwortet aufgrund des oben beschriebenen Fehlens von Korrelationen zwischen den Urteilsdimensionen den Verzicht auf die Verwendung des Begriffs der diagnostischen Kompetenz, soweit damit die Fähigkeit zu treffenden Beurteilungen von Personmerkmalen gemeint ist (ebd., S. 94). Einschränkungen bezüglich der Aussagekraft einzelner Komponenten wurden bereits von Schrader aus theoretischen Überlegungen heraus diskutiert und aus methodischer Sicht die Rangordnungskomponenten als biasfreie Maße für die diagnostische Kompetenz im engeren Sinne favorisiert (Schrader, 1989, S. 119). Diese auffallende Fokussierung auf die Rangordnungskomponente ist Karst (2017a) zufolge jedoch problematisch, weil nur ein einziger Aspekt der Urteilsakkuratheit in Betracht gezogen wird, weshalb ihrer Einschätzung nach die Frage nach der Struktur, Generalität und Spezifität diagnostischer Kompetenz theoretisch und empirisch offen bleibt (ebd., S. 24). Zudem stellt sie heraus, dass bei einer Urteil-Kriterium-Korrelation über mehrere Aufgaben und Lernende hinweg Mit-

telwerts- und Streuungsunterschiede außer Acht gelassen werden, was zu einer Überschätzung diagnostischer Kompetenz führen kann (ebd., S. 25).

Insgesamt ist festzuhalten, dass die wenigen Untersuchungen, die sich mit der Urteilsgenauigkeit von Erzieherinnen beschäftigen, methodisch zum Teil nur begrenzt belastbar sind und das Konstrukt diagnostische Kompetenz als generelle Fähigkeit von einzelnen Autoren sehr kritisch gesehen wird.

3.4.2 ... mit expliziter Berücksichtigung eines (bildungswissenschaftlichen) Kompetenzbegriffs

Die diagnostische Kompetenz mit expliziter Berücksichtigung eines bildungswissenschaftlichen Kompetenzbegriffs untersucht Karst (2012), indem sie auf Basis der Arbeiten von Cronbach (1955), Schrader (1989) und McElvany et al. (2009) eine perspektivische Erweiterung vornimmt. Karst bezieht in ihr Modell diagnostische Situationen mit ein, die sie zunächst als aufgabenbezogen, personenbezogen und personenspezifisch (Karst, 2012, S. 97) und später als klassenbezogene, schülerglobale und schülerspezifische Situationen bezeichnet (Karst, Schoreit & Lipowsky, 2014, S. 241; Karst, 2017a, S. 26–27). Damit berücksichtigt sie die Kontextspezifität von Kompetenzen (Klieme & Leutner, 2006, S. 879). Innerhalb dieser Kontexte und diagnostischen Situationen werden nach Karst (2012) jeweils verschiedene Teilkompetenzen zur Bewältigung von Anforderungssituationen benötigt. So ist in der aufgabenbezogenen diagnostischen Situation die niveaubezogene diagnostische Kompetenz zur Einschätzung der Eignung von Aufgaben für die gesamte Klasse notwendig, während für die Beurteilung der Eignung möglicher Alternativaufgaben die auf die Heterogenität von Aufgaben bezogene diagnostische Kompetenz benötigt wird (vgl. Karst, 2012, S. 97). Auch Karst kommt hier zu dem Gesamtergebnis, dass es sich bei den Teilkompetenzen nicht um Kompetenzen handelt, die sich einer übergeordneten diagnostischen Kompetenz zuordnen lassen; vielmehr sind Kompetenzen wahrscheinlich selbstständig und voneinander unabhängig (Karst, 2012, S. 212). Damit bestätigt sie bisherige Erkenntnisse von Schrader (1989), Spinath (2005) und Murphy et al. (1982).

Drei Ebenen diagnostischer Kompetenzen (klassenbezogene, schülerglobale und schülerspezifische) identifizieren Karst et al. (2014, S. 237) anhand der Daten des Projekts PERLE (Persönlichkeits- und Lernentwicklung von Grundschulern) (vgl. Greb, Poloczek, Lipowsky & Faust, 2011) und stellen den Einfluss dieser drei Ebenen auf die

Lernentwicklung von Grundschulkindern im Fach Mathematik fest. Karst et al. (2014) zeigen mittels Mehrebenenanalysen auf, dass die klassenbezogene diagnostische Kompetenz die Lernentwicklung positiv vorhersagt, während die diagnostische Kompetenz, bei der die Lehrkraft einzelne Schüler(innen) in Lernbereichen einschätzt (schülerglobale diagnostische Kompetenz und Rangordnungskomponente), die Lernentwicklung negativ prädiziert. Diese teilweise erwartungswidrigen Befunde werfen die Frage auf, ob mit einer Zunahme diagnostischer Kompetenzen eine Verbesserung der Unterrichtsqualität und damit der Lernentwicklung einhergeht. Die schüler(innen)spezifische diagnostische Kompetenz zeigt keine empirische Relevanz.

3.4.3 ... unter Berücksichtigung weiterer Merkmale

Diagnostische Kompetenz wird von einer Vielzahl von Faktoren beeinflusst, die auf verschiedenen Untersuchungsebenen zu finden sind, wie z.B. auf Ebene der Diagnostizierenden (Lehrkräfte oder Erzieherinnen), Diagnostikanden (Schüler(innen) bzw. Kindergartenkinder), der Untersuchungsgegenstände (Testinhalt), der Inhaltsbereiche (Domänen), der Lerngruppe bzw. Klasse etc. Der folgende Überblick an Studien gibt die Vielschichtigkeit der Einflussfaktoren wieder und verdeutlicht die Komplexität der Erfassung.

Ebene der Lehrkräfte und Erzieherinnen

Auf Ebene der Lehrenden werden vor allem die Berufserfahrung und die Urteilssicherheit fokussiert. Unterschiede zwischen den Lehrkräften in der Beurteilung der Schüler(innen)leistung untersuchen Hoge und Coladarci (1989). Sie stellen fest, dass sich Lehrkräfte dahingehend stark unterscheiden, wie akkurat sie die Leistung ihrer Schüler(innen) beurteilen (ebd., S. 306). Hopkins, George und Williams (1985) finden eine Streuung des mittleren r von $r = .44$ bis $r = .88$ fest (ebd., S. 179) und Helmke und Schrader (1987) übereinstimmend mit den meisten Resultaten in der Literatur eine Streuung von $r = .03$ bis $r = .90$ (ebd., S. 94).

Die Berufserfahrung von Lehrpersonen gehört für diagnostische Kompetenz zu den vielfach untersuchten Einflussfaktoren und geht letztlich auf das Expertenparadigma zurück (vgl. z.B. van Ophuysen, 2006, S. 155), wobei das Expertentum dabei häufig als Berufserfahrung oder auch als professionelles Wissen operationalisiert wird. Für dieses

wird ein großer positiver Einfluss auf das diagnostische Handeln vermutet (Ohle et al., 2017, S. 33); Belege hierfür stehen jedoch noch aus.

Ein positiver statistisch signifikanter Zusammenhang der Berufserfahrung mit der diagnostischen Sensitivität bei der vergleichenden Einschätzung von individuellen Schüler(innen)leistungen findet sich bei McElvany et al. (2009, S. 230). Ihnen zufolge können Lehrkräfte mit mehr Berufserfahrung Schüler(innen) demnach zwar besser in eine leistungsbezogene Rangreihe bringen ($r = .21$), es gelingt ihnen jedoch statistisch signifikant weniger gut, die Schwierigkeit der einzelnen Aufgaben vergleichend einzuschätzen ($r = -.23$) (ebd.).

Keine Zusammenhänge von diagnostischer Kompetenz und Berufserfahrung finden sich im Hinblick auf kognitive Fähigkeiten von Schülerinnen und Schülern bei Wild und Rost (1995) oder in Bezug auf Schüler(innen)leistung bei Schrader (1989) sowie Demaray und Elliot (1998).

Helmke und Fend (1981) ermitteln sogar negative Effekte des Berufsalters und stellen fest, dass Lehrkräfte mit maximal fünfjähriger Berufserfahrung über das Selbstvertrauen, die Schulfreude und auch die Konzentrationsfähigkeit der von ihnen unterrichteten Schüler(innen) deutlich besser im Bilde sind als Lehrer mit längerer Berufserfahrung (vgl. Helmke & Fend, 1981, S. 356). Mögliche Erklärungen sehen die Autoren zum einen in der verbesserten Lehrerausbildung, die den Erfordernissen pädagogischer Diagnostik im Schulalltag in höherem Maße gerecht wird, und zum anderen darin dass Schüler(innen) jungen Lehrkräften mehr Vertrauen entgegen bringen oder junge Lehrkräfte engagierter sein könnten als ihre älteren Kollegen (vgl. ebd.).

Aufgrund des uneinheitlichen Forschungsstandes gehen auch Praetorius, Karst, Dickhäuser und Lipowsky (2011) der Frage nach, ob interindividuelle Unterschiede in der diagnostischen Kompetenz mit der Berufserfahrung zusammenhängen. Die Korrelationen von diagnostischer Kompetenz (Rangordnungskomponente) und Berufserfahrung ergeben folgende bereichsspezifischen Unterschiede: Mathematik $r = -.13$, Lesen $r = -.15$, Schreiben $r = .25$, wobei die Zusammenhänge ohne einheitliche Richtung statistisch nicht signifikant sind, ebenso wenig wie die Korrelationen zwischen Berufserfahrung und Niveau- sowie Differenzierungskomponente (ebd., S. 87).

Ob die Dauer der Tätigkeit im Lehrerberuf mit der Genauigkeit von Schwierigkeitseinschätzungen zusammenhängt, untersuchen auch Hoffmann und Böhme (2014). Sie weisen jedoch nur vereinzelt statistisch signifikante Effekte auf die Über- oder Unterschät-

zung der Schwierigkeit einer Aufgabe für Schüler(innen) der eigenen Klasse nach. Dieses Ergebnis erscheint nach Ansicht der Autoren zwar kontraintuitiv, weil Lehrkräfte mit größerer Erfahrung eine höhere Expertise vermuten lassen, stimmt jedoch mit den Befunden anderer Studien überein (Hoffmann & Böhme, 2014, S. 51).

Der Frage, ob auch die Diagnosekompetenz von Erzieherinnen von der Berufserfahrung abhängt, geht Dollinger (2013) nach. Ihre mehrbenenanalytische Auswertung zeigt, dass die Berufserfahrung keinen Einfluss auf die Genauigkeit der Einschätzungen hat (Dollinger, 2013, S. 156). Es findet sich in nur einem Bereich ein signifikanter Einfluss der Berufserfahrung auf die Niveauebene: im Bereich der Mathematik zeigt sich hinsichtlich der Einschätzung der Rangreihe ein Effekt (ebd.). Während eine Fachkraft mit durchschnittlich drei Jahren Berufserfahrung eine Rangordnungskomponente von .14 aufweist, erhöht sich diese für eine Fachkraft mit durchschnittlich 20 Jahren Berufserfahrung auf .38; diese Tatsache, dass mit zunehmender Berufserfahrung sowohl die Einschätzung der wahren Testwerte der Kinder als auch die Differenzierungsfähigkeit der Pädagoginnen etwas exakter werden (ebd.), erklärt Dollinger damit, dass Erzieherinnen mit mehr Berufserfahrung möglicherweise nicht mehr so genau hinsehen, weil sie zu wissen glauben, was Kinder dieses Alters können und auch Veränderungen hinsichtlich einer zum Teil zunehmenden häuslichen Förderung nicht wahrnehmen (ebd., S. 158).

Die Ausbildung wird von der Autorin als weitere mögliche Ursache für die Ergebnisse diskutiert, weil Erzieherinnen zunehmend gezielter auf die Wahrnehmung der kognitiven Kompetenzen hin geschult werden und somit fehlende Erfahrungen mit einem „fokussierteren Blick“ ausgleichen (ebd., S. 159). Unter Berücksichtigung der Tatsache, dass in ihrer Studie nach eigener Angabe „die Einschätzungen in allen untersuchten Bereichen insgesamt nicht sehr exakt sind, (...) scheinen die Differenzen weder in der Erfahrung noch in der Ausbildung begründet zu sein, sondern an der Tatsache zu liegen, dass im Kindergarten eher die sozialen (und zum Teil sprachlichen) Kompetenzbereiche Gegenstand der Beobachtungen sind und weniger die kognitiven“ (ebd.).

Wie bereits angemerkt, ist bei der Bewertung der Ergebnisse jedoch zu berücksichtigen, dass Dollinger (ebd., S. 193) selbst die Belastbarkeit ihrer Studie durch methodische Schwächen eingeschränkt sieht.

Mit der Urteilssicherheit von Lehrkräften identifizieren Praetorius, Berner, Zeinz, Scheunpflug und Dresel (2013) einen weiteren diagnostisch relevanten Faktor. Die Un-

tersuchung zeigt für den mathematischen Bereich einen moderaten Zusammenhang der Urteilssicherheit mit der Urteilsgenauigkeit; im Bereich Deutsch hingegen können keine signifikanten Zusammenhänge aufgedeckt werden (vgl. ebd., S. 64). Zudem zeigen die Ergebnisse, dass sich Mathematiklehrkräfte ihrer Urteilsgenauigkeit teilweise bewusst sind, Deutschlehrkräfte jedoch nicht (vgl. ebd., S. 74).

Ebene der Kinder

Ein häufig untersuchter Einflussfaktor auf der Ebene der Kinder ist die Geschlechtszugehörigkeit. Es existieren vielfache Belege dafür, dass dieser Faktor für die Diagnosegenauigkeit von Lehrkräften keine Rolle spielt (Demaray & Elliot, 1998). Hoge and Butcher (1984, S. 779) finden keine signifikanten Zusammenhänge zwischen Diagnosegenauigkeitsmaßen und Geschlecht. Auch bei Dollinger (2013) korrelieren Geschlechtszugehörigkeit der Kinder und Rangordnungskomponenten der Lehrkräfte nicht miteinander. Erzieherinnen differenzieren dagegen besser zwischen den Mädchen als zwischen den Jungen, wobei Erstere im Vergleich zu den Jungen im Hinblick auf ihre mathematischen Kompetenzen signifikant häufiger unterschätzt werden (Dollinger, 2013, S. 177).

Auch die Intelligenz der Schüler(innen) bzw. deren akademische Leistungen werden als Einflussfaktor auf die diagnostische Kompetenz ihrer Lehrkräfte untersucht. Hoge und Butcher (1984) stellen im Rahmen einer multiplen Regressionsanalyse signifikante Zusammenhänge fest zwischen verbalem IQ-Maß von Schülerinnen und Schülern und der Einschätzung seitens der Lehrenden hinsichtlich a) der von Leistungen in einem Lesetest, b) des Vertrauens in dieses Rating, c) der grundlegenden intellektuellen Fähigkeiten der Schüler(innen) und d) von Motivationsausprägungen von Schülerinnen und Schülern bezüglich der Erledigung von Schulaufgaben (Hoge & Butcher, 1984, S. 778).

Wie exakt Lehrkräfte Leistungen von Schülerinnen und Schülern beurteilen können, wird nach Coladarci möglicherweise von deren akademischen Fähigkeiten beeinflusst (1986, S. 308). Dafür sprechen auch die Befunde von Hoge und Butcher (1984) sowie von Demaray und Elliot (1998).

Zusammenhänge zwischen der Urteilsgenauigkeit und den Leistungen der Schüler(innen) in unterschiedlichen Fächern deckt Coladarci auf (1986, S. 145). Er weist in mehreren Fächern nach, dass Kinder mit hoch eingestuften Fähigkeiten genauer beurteilt werden als Kinder mit gering eingeschätzten Fähigkeiten (ebd.).

Die Tendenz von Lehrkräften, das Leistungsniveau leistungsstarker Schüler(innen) zu überschätzen zeigt sich auch in der Studie von Hoge und Butcher, die ihre Berechnungen unter Einbezug eines Intelligenzquotienten der Schüler(innen) als Prädiktor für Urteilsgenauigkeit anstellen (Hoge & Butcher, 1984, S. 780).

Deutliche Unterschiede in der Einschätzung der Leistung verschiedener Gruppen von Schülerinnen und Schülern finden Demaray und Elliot (1998, S. 17), wobei sie eine Übereinstimmung von Urteilen Lehrender und Testleistungen von Schülerinnen und Schülern von 79% für die Gesamtskala und einheitliche Werte für die Subskalen Lesen, Rechtschreibung und Mathematik ermitteln. Die Urteil-Kriterium-Korrelation zwischen Einschätzung und Testleistung für *lower-achieving students* beträgt $r = .56$, die für *higher-achieving students* $r = .75$ (Demaray & Elliot, 1998, S. 17–18). Die Unterschiede zwischen den beiden Schüler(innen)gruppen hinsichtlich der Übereinstimmung von Einschätzung und Testleistung pro Item des gleichen Instruments sind gering, aber statistisch signifikant. Der durchschnittliche Zusammenhang von Leistung und Urteil für Lehrkräfte und Schüler(innen) mit niedrigeren Leistungswerten beträgt 77% und der für Lehrkräfte und Schüler(innen) mit höheren Leistungswerten 80% (ebd., S. 18).

Ebene der Gruppe bzw. Klasse

Hinsichtlich der Einflussmerkmale auf Gruppenebene, der Frage also, welche Klassenmerkmale mit der Urteilsakkuratheit zusammenhängen, sind die Forschungsergebnisse inkonsistent (vgl. Praetorius et al., 2012, S. 129), was sich z.B. in Bezug auf die Klassengröße zeigt. Keine statistisch signifikanten oder praktisch bedeutsamen Unterschiede in Abhängigkeit von der Klassengröße finden Wild und Rost (1995, S. 83–84) für die Intelligenzdaten oder für die Lehrereinschätzungen. Eine signifikante Korrelation zwischen aufgabenbezogener diagnostischer Kompetenz und der Klassengröße von $r = .37$ ermittelt hingegen Schrader (1989, S. 123). Je größer eine Klasse ist, desto eher können Lehrkräfte die Schwierigkeit von Aufgaben einschätzen. Kleinere Klassen haben Schrader (1989) zufolge den Vorteil, dass die Lehrkraft Einzelne intensiver fördern kann, dass sie weniger Zeit in die Organisation von Lernumgebungen investieren muss und deshalb „quantitativ und qualitativ bessere Beobachtung Gelegenheiten hat“ (ebd., S. 122). Trotz dieser Vorteile zeichnen sich Lehrkräfte kleiner Klassen nicht durch höhere diagnostische Kompetenz aus, sondern eher durch eine geringere, was damit erklärt wird, dass individuelle Einschätzungen in großen Klassen genauer ausfallen könnten, weil hier auf mehr Indikatoren zurückgegriffen werden kann (ebd.).

3.5 Veränderung der Urteilsgenauigkeit

Die Veränderung der Urteilsgenauigkeit wurde als Kontaktdauer zwischen Urteiler und Beurteiltem auf Klassenebene untersucht.

Es ist theoretisch plausibel, dass Lehrende umso mehr Informationen über ihre Bezugskinder oder Schüler(innen) sammeln, je länger sie diese kennen. Daher sollte mit zunehmender Kontaktzeit das Urteil Lehrender über die Fähigkeiten ihrer Kinder genauer ausfallen, weshalb die Dauer des Kontakts zu den Kindern, die eingeschätzt werden, einen weiteren möglichen Einflussfaktor darstellt. Zum Einfluss der Kontaktdauer auf die diagnostische Kompetenz von Erzieherinnen liegen keine Erkenntnisse vor. Die bisherigen Befunde im schulischen Kontext zeigen keine oder nur minimale Zusammenhänge zwischen der Kontaktdauer und der Ausprägung diagnostischer Komponenten.

Die Untersuchung der in der Klasse verbrachten Unterrichtszeit auch von Lehrkräften mit Unterricht in zusätzlichen Fächern zeigte, dass auch die Gelegenheit, durch zusätzlichen Unterricht in anderen Fächern die Schüler(innen) der eigenen Klasse besser kennenzulernen, keinen positiven Einfluss auf das diagnostische Wissen der Lehrkräfte hat (Schrader, 1989, S. 122). Wie die Studie von Wild und Rost zeigt, ermöglicht auch ein längerer Kontakt mit der beurteilten Schulklasse in Kombination mit kleinen Schulklassen keine genauere Beurteilung (1995, S. 86). Die Studie von Lorenz (2011) bestätigt dieses Ergebnis.

Ebenfalls nur nichtsignifikante Effekte der Kontaktdauer fanden Oerke, McElvany, Ohle, Ullrich und Horz (2016, S. 43–44), was darauf hindeutet, dass Lehrkräfte, die ihre Klasse ein Jahr länger unterrichten, weder das Leistungsniveau ihrer Klasse noch die relative Position ausgewählter Schüler(innen) akkurater einschätzen können.

Bei der Untersuchung des Zusammenhangs von Kontaktdauer (mit der jeweiligen Klasse) und der Genauigkeit von Schwierigkeitseinschätzungen wurden dagegen vereinzelt statistisch signifikante Effekte auf die Über- oder Unterschätzung der Schwierigkeit einer Aufgabe für die Schüler(innen) der eigenen Klassen nachgewiesen (vgl. Hoffmann & Böhme, 2014, S. 51).

Zusammenfassend lassen die bisherigen Befunde keinen überzeugenden Schluss auf einen Zusammenhang zwischen der Kontaktdauer und der Ausprägung der Urteilsgenauigkeit zu.

3.6 Wirkung diagnostischer Kompetenz

Wie das frühpädagogische Angebots-Nutzungs-Modell (Kap. 2.6.4) aufzeigt, sind aufgrund der Vielzahl von Einflussfaktoren direkte Zusammenhänge von diagnostischer Kompetenz und dem Ergebnis einer Förderung kaum zu erwarten. Dementsprechend können nur wenige Studien Belege für einen Zusammenhang liefern; zudem sind diese Befunde hinsichtlich der Auswirkung der diagnostischen Kompetenz der Lehrenden auf die Lernentwicklung von Kindern und Jugendlichen uneinheitlich.

Keinen direkten signifikanten Zusammenhang zwischen Lernerfolg der Klassen und der diagnostischen Kompetenz der jeweiligen Lehrer findet Schrader (1989, S. 196). Lediglich in Abhängigkeit von Unterrichtsmerkmalen (Strukturierung, individuelle fachliche Hilfestellung, Aufgabenstellungen) kann er signifikante Zusammenhänge zwischen Diagnosekompetenz des Lehrers und dem Lernerfolg der Klasse nachweisen (ebd., S. 196).

Einen signifikanten Zusammenhang von Urteilsgüte (aufgabenbezogene Urteilsfehler sowie personenbezogene Rangordnungskomponente) und der Unterrichtsqualität belegen Anders, Kunter, Brunner, Krauss und Baumert (2010, S. 190). Hinsichtlich der aufgabenspezifischen Treffer zeigt sich in einer Untersuchung von Karing, Pfost und Artelt (2011) ein signifikanter positiver Zusammenhang mit der Entwicklung der Lesekompetenz der Schüler(innen). Die am stärksten ausgeprägte Lesekompetenzentwicklung stellen Karing et al. (2011, S. 141) fest, wenn eine hohe Trefferquote mit der Gabe von wenigen Strukturierungshilfen im Unterricht einhergeht. Die Erklärung für diesen unerwarteten Befund wird in der unterschiedlichen Leistungsfähigkeit der Schüler(innen) gesehen, weil Leistungsfähigere evtl. weniger auf die Strukturierungsbemühungen der Lehrkräfte angewiesen sind (ebd.).

Einen positiven Zusammenhang der personenbezogenen Rangordnungskomponente von Lehrkräften mit dem Leseverständnis ihrer Schüler(innen) weisen Behrmann und Souvignier (2013, S. 289–290) nach und stellen auch fest, dass eine überdurchschnittliche Häufigkeit individueller Leistungsrückmeldungen einen positiven Moderatoreffekt darstellt.

Im Gegensatz dazu stehen die Befunde von Karst et. al. (2014, S. 244), die erwartungswidrig einen negativen Effekt der personenbezogenen Rangordnungskomponente auf die Lernentwicklung der Schüler(innen) ermitteln.

Großes diagnostisches Wissen kann auf die Effektivität von unterrichtlichem Verhalten einen störenden Einfluss haben, stellt Lingelbach (1995, S. 245–246) fest. Eine Untersuchung der Wechselwirkung zwischen diagnostischem Wissen und der Klassenführung zeigt, dass hohes fallspezifisches Wissen den Einfluss hoher, auf die Gesamtklassebezogener allgemeiner Klassenführungskompetenz mindern kann (Lingelbach, 1995).

Hinsichtlich der Wirksamkeit von Urteilsgenauigkeit sind zudem Unterschiede zwischen untersuchten Domänen bzw. Unterschiede zwischen Schwerpunkten innerhalb von Domänen bedeutsam. Die Tatsache, dass auch Schwerpunkte innerhalb einer Domäne in signifikantem Zusammenhang mit der Urteilsgenauigkeit stehen können, zeigt Coladarci (1986, S. 143). Er stellt fest, dass Urteile Lehrender zu Antworten von Schülerinnen und Schülern bezüglich Items mit mathematischen Berechnungen signifikant genauer (76.52%) sind als solche bezüglich der Items, die sich auf mathematische Begrifflichkeiten beziehen (70.02%). Auch die bereichsspezifische Komplexität beeinflusst die Genauigkeit von Urteilen. Es zeigt sich, dass Lehrkräfte innerhalb einer Domäne unterschiedlich komplexe Konstrukte nicht gleich genau einschätzen. Die Urteile der Lehrer bezüglich der Schülerantworten bei mathematischen Aufgaben waren signifikant genauer (76,52%) als die Urteile bezüglich der Schülerantworten hinsichtlich mathematischer Konzepte (70.02%) (Coladarci, 1986, S. 143).

Die Tatsache, dass sich nur geringe Effekte der Urteilsgenauigkeit auf die Leistungsentwicklung von Kindern und Jugendlichen zeigen, kann Praetorius und Südkamp (2017) zufolge an der Multideterminiertheit der kindlichen Leistungsentwicklung liegen, wie es z.B. die Meta-Metaanalysen von Hattie (2013) verdeutlichen.

3.7 Zusammenfassung

Die Untersuchung diagnostischer Kompetenz als Urteilsgenauigkeit besitzt eine lange Geschichte. In dieser Tradition der Judgment-Accuracy-Forschung wird die Urteilsgenauigkeit als die Fähigkeit verstanden, Personen hinsichtlich bestimmter Merkmale möglichst genau einzuschätzen. Bei dieser Form der Diagnostik handelt es sich um eine rückblickende oder summative Statusdiagnostik.

Im Hinblick auf die Erfassung der Urteilsgenauigkeit ist das messmethodische Vorgehen Cronbachs (1955) mittels der Analyse einzelner Komponenten besonders bedeutsam. Zur Erfassung und Einschätzung des Forschungsstandes sind jedoch auch die Mo-

difizierung der Komponentenanalyse von Schrader (1989) und deren Erweiterung nach McElvany et al. (2009) (Kap. 3.1) relevant.

Weder die Forschungsergebnisse für den schulischen noch für den vorschulischen Bereich zeichnen ein einheitliches Bild, was unter anderem an der sehr unterschiedlichen Qualität der Erfassung diagnostischer Kompetenz liegen könnte. Die Fokussierung oder Reduktion diagnostischer Kompetenz auf die Rangordnungskomponente als das am wenigsten belastete Maß (vgl. Kap. 3.4.1) wird kritisch gesehen. Berücksichtigt werden sollten die gesamten Komponenten der ursprünglichen Messmethodik nach Cronbach (1955). Zudem sollten zukünftige Forschungen zur Urteilsgenauigkeit auf vergleichbaren Qualitätskriterien beruhen. Die sehr unterschiedliche Qualität einzelner Forschungsarbeiten ist insbesondere auf Unterschiede bezüglich messmethodischer Charakteristika zurückzuführen. Daher wird die vollumfängliche Erfüllung des oben dargestellten Anforderungsprofils (Kap. 3.3) als wichtige Grundvoraussetzung für weitere Studien und deren Vergleichbarkeit gesehen. Folgende Merkmale haben sich zur Erfassung der Urteilsgenauigkeit als geeignet erwiesen:

- Operationalisierung der Urteilsgenauigkeit mittels direkter Beurteilungen und Testkenntnis
- Verwendung von Verfahren höchster Spezifität
- Berücksichtigung peer-unabhängiger Urteile
- Ermittlung auf Itemebene
- Berücksichtigung der hierarchischen Struktur (Klassen / Bezugsgruppen)
- Domänenspezifische Erfassung (Bereichsspezifität).

Die langjährige Forschung kann in vier Herangehensweisen gegliedert werden: in diagnostische Kompetenz operationalisiert als Akkuratheit von Lehrkräfte- und Erzieherinnenurteilen (3.4.1), diagnostische Kompetenz mit expliziter Berücksichtigung eines (bildungswissenschaftlichen) Kompetenzbegriffs (3.4.2), diagnostische Kompetenz unter Berücksichtigung weiterer Merkmale (3.4.3) und diagnostische Kompetenz unter Berücksichtigung des Prozesscharakters der Urteilsbildung, welche im Kapitel 4 zur Beobachtungskompetenz ausgeführt wird.

Im Hinblick auf Akkuratheit zeigt sich, dass die Komponenten innerhalb, aber auch zwischen den Urteilsdimensionen höchstens schwach korrelieren, weshalb einzelne Autoren sogar die Verwendung des Begriffs diagnostische Kompetenz, operationalisiert als Fähigkeit zu treffenden Beurteilungen von Personenmerkmalen, für unangemessen hal-

ten. Zur Urteilsgenauigkeit von Erzieherinnen liegen kaum Untersuchungen vor. Zudem sind diese wenigen Studien unter methodischen Gesichtspunkten angreifbar, weil zum einen der Datenerfassung sehr unterschiedliche messmethodische Charakteristika zugrunde liegen und zum anderen teilweise eine sehr reduzierte Sicht auf diagnostische Kompetenz vorliegt, wenn diese einzig durch die Rangordnungskomponente abgebildet wird.

Die Studien zur diagnostischen Kompetenz mit expliziter Berücksichtigung eines (bildungswissenschaftlichen) Kompetenzbegriffs bestätigen die bisherigen Erkenntnisse hinsichtlich der Annahme der Unabhängigkeit einzelner (Teil-)Kompetenzen. Die teilweise erwartungswidrigen Befunde zur diagnostischen Kompetenz auf unterschiedlichen Ebenen (klassenbezogen, schülerglobal und schülerspezifisch) lassen die Frage offen, ob mit Zunahme diagnostischer Kompetenz eine Verbesserung der Unterrichtsqualität und Lernentwicklung einhergeht.

Die Untersuchungen zur diagnostischen Kompetenz unter Berücksichtigung weiterer Merkmale zeigen große Unterschiede zwischen einzelnen Lehrkräften bezüglich der Genauigkeit der Einschätzung von Schüler(innen)leistung, wobei die Berufserfahrung kaum Einfluss auf diagnostische Kompetenz hat. Die Urteilssicherheit von Lehrkräften zeigt sich nur für den mathematischen Bereich als relevanter Einflussfaktor. Auf Ebene der Kinder ist die Geschlechtszugehörigkeit als möglicher Einflussfaktor vielfach untersucht worden, wobei sich Belege dafür finden, dass die Urteile von Lehrkräften davon unberührt bleiben. Im Hinblick auf Erzieherinnen spielt die Geschlechtszugehörigkeit der Kinder im mathematischen Bereich eine Rolle, da Mädchen in diesem Bereich häufiger unterschätzt werden. Eine Replikation dieses einmal erhobenen Befundes steht noch aus. Intelligenz und akademische Leistungsfähigkeit von Kindern sind Einflussfaktoren auf diagnostische Kompetenz. Kinder mit höherer Leistung stehen in Zusammenhang mit genaueren Urteilen ihrer Lehrkräfte. Die Einflussmerkmale auf Gruppenebene sind uneinheitlich. Der Einflussfaktor Bereichsspezifität ist bedeutsam, weil die Validitätskoeffizienten im Bereich Mathematik höher sind als z.B. in den Naturwissenschaften, allerdings spielt innerhalb der Domäne nochmals die Komplexität der Konstrukte eine Rolle, weil weniger komplexe Konstrukte besser eingeschätzt werden können. Der seltener untersuchte Einfluss der Kontaktdauer auf die Urteilsgenauigkeit von Lehrkräften scheint in keinem Zusammenhang mit diagnostischer Kompetenz zu stehen, wobei zum einen vereinzelt statistisch signifikante Effekte der Kontaktdauer auf

die Beurteilung der Aufgabenschwierigkeiten gefunden wurden und zum anderen noch keine Erkenntnisse für Erzieherinnen vorliegen.

Bis dato fehlt der empirische Beweis für einen direkten Nutzen der Urteilsgenauigkeit für das Handeln von Lehrpersonen und letztlich für den Leistungszuwachs der Kinder – auch wenn dieser theoretisch sehr plausibel erscheint. Vereinzelt positive und teilweise indirekte Befunde (Anders et al., 2010; Hosenfeld, 2009; Schrader, 1989) lassen, anders als von Grausam (2018, S. 83) konstatiert, noch keinen allgemeinen Rückschluss auf einen Zusammenhang zwischen Diagnosekompetenz und Lernleistung zu. Zudem sind die Befunde zur Wirkung diagnostischer Kompetenz auf den Lernerfolg von Schülerinnen und Schülern uneinheitlich: Während auf der einen Seite in Abhängigkeit von weiteren Unterrichtsmerkmalen signifikante positive Effekte in einzelnen Domänen zu finden sind, wurden auch erwartungswidrige negative Effekte entdeckt. Insgesamt ist die Forschung zu direkten Effekten der diagnostischen Kompetenz auf die Leistungsentwicklung von Kindern überschaubar. Dies kann daran liegen, dass die kindliche Leistungsentwicklung multideterminiert ist und die Kontrolle aller relevanten Einflussfaktoren zu äußerst komplexen Forschungsdesigns führen würde.

Insgesamt ist festzuhalten, dass zur diagnostischen Kompetenz von Erzieherinnen nur wenige Erkenntnisse vorliegen und vorhandene Arbeiten aufgrund der gewählten Operationalisierungen oder auch im Hinblick auf messmethodische Vorgehensweisen (z.B. geringe Anzahl der Einschätzungen pro Erzieherin) in ihrer Aussagekraft kritisch zu hinterfragen sind. Es ist auch offen, inwiefern internationale Ergebnisse aus vorschulischen Einrichtungen mit einer anderen pädagogisch-didaktischen Tradition und auch mit anderer Ausbildung des Personals im vorschulischen Bereich auf den Kindergarten und die Ausbildung des frühpädagogischen Personals in Deutschland übertragbar sind. Daher lässt sich konstatieren, dass sich die nationale elementarpädagogische Forschung hinsichtlich der Untersuchung diagnostischer Kompetenzen noch in den Anfängen befindet.

3.8 Folgerungen

Auf der einen Seite hat die langjährige Forschung zur diagnostischen Kompetenz als Urteilsgenauigkeit zwar eine Fülle an Erkenntnissen zu Lehrkräften gebracht, es gibt jedoch kaum Erkenntnisse, die für die Gestaltung von Lehr-Lernsituationen und für das

pädagogische Handeln von Lehrpersonen bedeutsam sind. Dies trifft auf den Elementarbereich in besonderer Weise zu.

Auf der anderen Seite stellt die diagnostische Kompetenz theoretisch begründet eine Voraussetzung für die Schaffung adaptiver Lerngelegenheiten dar: Wenn eine Lehrperson den aktuellen Entwicklungsstand ihrer Bezugskinder nicht kennt, kann sie nur schwerlich eine passende Lernumgebung gestalten und die Kinder gezielt in der Zone der nächsten Entwicklungsstufe (Vygotskij, 2002, S. 327) fördern. Daher wird vermutet, dass diagnostische Kompetenz die Voraussetzung für die Planung der Gestaltung adaptiver Lernumgebungen ist.

Wenn es aber um die Optimierung individueller Lernprozesse als das Ziel pädagogischer Diagnostik geht, so ist festzuhalten, dass es mit dem Verständnis diagnostischer Kompetenz als Urteilsgenauigkeit bisher noch nicht gelungen ist, dieses Ziel zu erreichen. Auch ist die Frage danach, wie genau die Kenntnis vom Wissensstand der Bezugskinder oder der Schüler(innen) sein muss, um adaptive Lernumgebungen zu gestalten, noch unbeantwortet: „Ein grundlegendes Problem bisheriger Forschungen zur diagnostischen Kompetenz liegt darin, dass die Bedeutung diagnostischer Kompetenz für die Qualität von Unterrichtsprozessen und die Leistungs- und Persönlichkeitsentwicklung von Schülerinnen und Schülern zwar theoretisch plausibel abgeleitet werden kann, empirisch jedoch kaum untersucht wurde“ (Praetorius & Südkamp, 2017, S. 16).

Es hat den Anschein, als sei die Diagnosekompetenzforschung mit der Fokussierung auf das Konstrukt Urteilsgenauigkeit an ihre Grenzen gelangt. Schrader verwies bereits 1989 darauf, dass in künftigen Untersuchungen versucht werden sollte, „die in der Unterrichtssituation selbst ablaufenden Diagnoseprozesse zum Gegenstand der Analyse zu machen. Nur so kann abgeschätzt werden, ob außerhalb des Unterrichts verwendete Urteilsaufgaben hinreichend valide sind“ (Schrader, 1989., S. 241). Die Notwendigkeit der Erfassung diagnostischer Kompetenz in situativ-authentischen Kontexten erscheint als eine logische Folgerung. Es zeigt sich, dass situierte Ansätze zur Erfassung der Urteilsgenauigkeit, wie z.B. die Erfassung nach Karst (2012) (Kap. 3.4.2), besondere Chancen bieten.

Auf Grundlage des Forschungsstands betonen Praetorius, Hetmanek, Herppich und Ufer (2017, S. 97) die Bedeutung einer repräsentativen Auswahl relevanter diagnostischer Situationen, welche im Bereich der Forschung zu diagnostischer Kompetenz bislang

nicht untersucht wurden. Sie sehen in der Schließung dieser Lücke eine wichtige Herausforderung für zukünftige Studien (vgl. ebd.).

Eine weitere Herausforderung besteht in der Berücksichtigung der Verwobenheit von diagnostischen Urteilen und pädagogischen Handlungen. Diese Verwobenheit ist Schrader (2017) zufolge besonders bei Mikroadaptationen ausgeprägt, „also den während des Unterrichtens stattfindenden kurzfristigen Anpassungsprozessen. Die Verwobenheit zeigt sich zum einen in einer engen Koppelung, zum anderen in der mangelnden Trennbarkeit von diagnostischen und instruktionalen Handlungen“ (ebd., S. 252).

Mit dem Blick auf die instruktionalen Handlungen wird zwar ein wichtiger Schritt unternommen, jedoch bleibt das Konstrukt Urteilsgenauigkeit dabei in seiner Form bestehen, obgleich mehrere Autoren die Frage danach stellen, ob es für den Ertrag bzw. die Wirkung von Unterricht tatsächlich wichtig ist, genaue Urteile vornehmen zu können.

Die Frage danach, inwiefern Urteilsgenauigkeit als alleinige Operationalisierung von diagnostischer Kompetenz für Lehrprozesse relevant ist, wurde wiederholt gestellt (z.B. Abs, 2007; Baumert & Kunter, 2006; Praetorius et al., 2012; Praetorius et al., 2017). Bereits 1989 wies Schrader darauf hin, dass Lehrende mit hohen Urteilsgenauigkeitswerten möglicherweise besser in der Lage seien, Leistungsentwicklungen zu prognostizieren, es aber fraglich bleibe, ob sich durch eine hohe Urteilsgenauigkeit auch die Planungsgüte von Unterricht verbessere (ebd., S. 256). Er benennt sogar einen möglichen gegenteiligen Effekt, der dadurch entstehen könne, dass ein Lehrender mit hoher Urteilsgenauigkeit zwar Vorteile hinsichtlich der Erstellung detaillierter Unterrichtsplanungen haben könnte; bezogen auf die notwendige Flexibilität des unterrichtlichen Handelns jedoch könnte sich dies als nachteilig erweisen (ebd.).

Die unterrichtspraktische Bedeutsamkeit der Urteilsgenauigkeit wird auch von Praetorius et al. (2012, S. 133–135) kritisch hinterfragt, zum einen mit Bezug auf die Vielzahl weiterer Kontext- oder Einflussfaktoren, zum anderen aufgrund der Art und Weise der Urteilsgenauigkeitserfassung, die vielfach weitab der Unterrichtspraxis unter laborähnlichen Bedingungen erhoben wurden.

Die unklare Befundlage der Forschung zur Urteilsgenauigkeit, so lässt sich vorerst festhalten, legt die Vermutung nahe, die diagnostische Kompetenz in dieser Form – also ein Blick in die Tiefe – könnte für die Gestaltung von Lehr-Lernprozessen und die Leistung von Kindern evtl. weniger bedeutsam sein. Möglicherweise wäre weniger die Genauigkeit der Diagnose einzelner Fähigkeiten als vielmehr ein breiter Blick auf die Entwick-

lung im Sinne einer weniger exakten aber bewussten Wahrnehmung vieler Fähigkeiten im laufenden Förderprozess notwendig.

Vor allem im frühpädagogischen Bereich erscheint das Verständnis von diagnostischer Kompetenz als Urteilsgenauigkeit nicht weiterführend. Wird das frühpädagogische allgemeine Kompetenzmodell von Fröhlich-Gildhoff et al. (2011, S. 17) herangezogen, so steht die Situationswahrnehmung und -analyse im Vordergrund, also die Frage danach, welche Leistungen eines Kindes in konkreten Situationen wahrgenommen werden. Dabei spielt die Kontextspezifität von Kompetenzen eine bedeutsame Rolle. Zudem könnte die komplexe Wahrnehmung einer Vielzahl von kindlichen Fähigkeiten und Fertigkeiten bedeutsamer sein als die genaue Einschätzung einer einzelnen Kompetenz, weil entsprechend der Kompetenzen adaptive Fördermaßnahmen gestaltet werden können.

Selbst wenn Wygotskis lerntheoretischer Ansatz herangezogen wird, könnte nicht die Urteilsgenauigkeit das Entscheidende sein, sondern die Wahrnehmung kindlicher Fähigkeiten in konkreten Situationen. Wichtig könnten auch die darauf basierenden Hilfen sein, die eine Erzieherin den Kindern anbietet, um ihnen das Vordringen in die Zone der nächsten Entwicklung und das Arbeiten dort zu ermöglichen.

Die bisherige Urteilsgenauigkeitsforschung offenbart die Notwendigkeit zur Erweiterung des Konstrukts sowie die Bedeutsamkeit eines breiten, offenen Blicks auf das in einzelnen Situationen mögliche Leistungsspektrum des Kindes. Diese neue weitere Facette neben der Urteilsgenauigkeit wird im Folgenden Beobachtungskompetenz genannt und ist der formativen Diagnostik zuzuordnen, welche in der Forschung zur diagnostischer Kompetenz Glogger-Frey und Herppich (2017, S. 45) zufolge noch weitgehend unberücksichtigt ist. Dies wird auf die erheblichen methodischen Herausforderungen zurückgeführt, welche die empirische Untersuchung der informellen sowie der semi-formellen Diagnostik mit sich bringt. Zukünftige Modelle von diagnostischer Kompetenz sollten explizit Überlegungen zur formativen Diagnostik enthalten, so die Forderung von Glogger-Frey und Herppich (ebd.). Zukünftige Modelle von diagnostischer Kompetenz sollten explizit Überlegungen zur formativen Diagnostik enthalten, so die Forderung von Glogger-Frey und Herppich (ebd.). Hierbei sei insbesondere die Berücksichtigung von Prozessmerkmalen pädagogischer Diagnostik wichtig. Die Autorinnen stellen heraus, dass „vor dem Hintergrund der Bedeutung formativen Assessments für das alltägliche unterrichtliche Handeln auch derartige Diagnoseleistungen explizit in einem zukünftigen Rahmenmodell diagnostischer Kompetenz adressiert werden sollten“

(ebd., S. 65). Schrader (2017) schlägt in diesem Zusammenhang vor, die Untersuchung der diagnostischen Kompetenz durch einen Bottom-Up-Ansatz zu ergänzen, bei dem, ausgehend von einer kognitionspsychologischen Analyse des Lehrendenhandelns, identifiziert wird, wo und wie diagnostische Informationen in das Handeln der Lehrperson einfließen und wie insbesondere die während des Unterrichtens ablaufenden Prozesse der Informationsaufnahme, -verarbeitung und -nutzung aussehen (ebd., S. 255). Seiner Meinung nach rekurren bisherige Forschungen den umgekehrten Weg, indem ausgehend von der Existenz und Bedeutsamkeit der diagnostischen Kompetenz die Bedeutung für das unterrichtliche Handeln analysiert wird, was er als Top-Down-Ansatz bezeichnet (ebd.).

Eine weitere Folgerung steht im Zusammenhang mit der zunehmenden Wahrnehmung der hohen Bedeutung lernprozessbezogener Diagnostik (Glogger-Frey & Herppich, 2017; Schütze et al., 2018). Neben Hinweisen zur Bedeutung, welche die Forschung liefert, erscheint es auch in Bezug auf die Vielfalt diagnostischer Aufgaben und Tätigkeiten sinnvoll, die bislang stark auf diagnostische Urteilskompetenz beschränkte Sichtweise zu erweitern (Schrader, 2009, S. 243). Nach Artelt und Gräsel (2009, S. 160) wäre stärker die Frage zu stellen, wie diagnostische Kompetenz der Optimierung von Bildungsprozessen und -ergebnissen dienen kann. Aus diesem Grund sollte eine Konzeptualisierung diagnostischer Kompetenz, die didaktische Relevanz beanspruchen will, stets die didaktische Einbettung von Beurteilungssituationen einbeziehen: „Somit wird Diagnostik von Lehrkräften sinnvoller Weise nicht nur in Analogie zum psychologischen Test zu untersuchen sein, sondern auch auf der Grundlage ihres Potenzials für die Gestaltung von Lernumgebungen“ (Abs, 2007, S. 64). Auch Hoth (2016) hebt die Prozessorientierung hervor, welcher Diagnostik in der Praxis zwangsläufig unterworfen ist: „In den tatsächlich ablaufenden Unterrichtssituationen müssen die Lehrkräfte eher lernprozessorientiert diagnostizieren, um hieran anschließend geeignetes Feedback und den Lernprozess unterstützende Hilfestellungen zu formulieren. Für die Entscheidungen, die eine Lehrkraft während des Unterrichts treffen muss, sind daher eher lernprozessorientierte Diagnosen relevant und weniger das akkurate Beurteilen von Schülerleistungen“ (ebd., S. 91).

Die Beobachtungskompetenz – also die Wahrnehmung von kindlichen Fähigkeiten und Fertigkeiten in konkreten Fördersituationen – spielt auf der Mikroebene von Lehrprozessen eine bedeutsame Rolle. Mit dieser werden jedoch nicht die Mikroadaptationen selbst in den Fokus genommen, sondern die diagnostische Fähigkeit, welche diesen

Adaptionen vorausgeht. Dieser breite Blick auf das Schaffen der Kinder hat in der Frühpädagogik eine lange Tradition. Im folgenden Kapitel wird die Beobachtungskompetenz als weitere Facette der diagnostischen Kompetenz beschrieben.

4. Beobachtungskompetenz

Die Beobachtungskompetenz ist neben der Urteilsgenauigkeit eine weitere Facette diagnostischer Kompetenz. Von Bedeutung ist sie deshalb, weil die Beobachtung eine grundlegende Voraussetzung für die Gestaltung von Lehr-Lernprozessen ist. Die Wahrnehmung wird im Kontext der Erfassung professioneller Kompetenzen von Lehrenden als „Anker zur Erfassung professioneller Kompetenz“ (Seidel, Blomberg & Stürmer, 2010, S. 296) beschrieben. Eine Vielzahl von Studien belegt die Bedeutung der Beobachtung im Vergleich zu anderen diagnostischen Methoden und Verfahren (Schmidt, 2020, S. 68–70). Die Verhaltensbeobachtung gilt als „eine wichtige, wenn nicht die wichtigste förderdiagnostische Methode“ (Breitenbach, 2020, S. 92). In einer Vielzahl von Lernsituationen gewinnen Lehrende durch mündliche Äußerung der Kinder, durch Schriftstücke und auch durch jedes Handeln ihrer Kinder aufschlussreiche Informationen über deren Lernweisen, Entwicklungs- und Leistungsstände (Prenzel, 2017, S. 20). Diese „in Lernsituationen gewonnenen Detailkenntnisse sind in der Regel für die Unterrichtsgestaltung wertvoller als Ergebnisse aus Testsituationen mit normierten Verfahren, die vor allem konzipiert wurden, um in kürzestmöglicher Zeit über Indizien für allgemeine Entwicklungsniveaus zu verfügen“ (ebd.). Dies gilt auch für Lernprozesse im frühpädagogischen Bereich.

Die Beobachtung und damit auch die Beobachtungskompetenz spielen deshalb auch in der Frühpädagogik eine bedeutsame Rolle. Mit der Beobachtungskompetenz wird das in der pädagogisch-psychologischen Bildungsforschung oft als Urteilsgenauigkeit oder Veridikalität sehr eng gefasste Konstrukt *diagnostische Kompetenz* im Sinne Helmkes *diagnostischer Expertise* (Helmke, 2017, S. 219) erweitert. Dieses Konzept führt er ein, um ein breiteres Verständnis des Begriffs diagnostischer Kompetenz hervorzuheben.

Im ersten Schritt wird aus der frühpädagogischen Perspektive die Relevanz von Beobachtung dargestellt (Kap. 4.1). Formen der Beobachtung(skompetenz) werden in Kapitel 4.2 beschrieben. Beobachtungsfehler (Kap. 4.3) und Besonderheiten zu Gütekriterien (Kap. 4.4) werden vor der Einordnung der Beobachtungskompetenz als Teil des formativen Assessments (Kap. 4.5) ausgeführt. Das erweiterte Linsenmodell zur Beobachtungskompetenz wird in Kapitel 4.6.1 vorgestellt, danach wird eine Einordnung in Modelle der Frühpädagogik und Lehr-Lernforschung (Kap. 4.6.2) vorgenommen. Überlegungen zur Veränderung der Beobachtungskompetenz werden in Kapitel 4.7 ausge-

führt. Eine Zusammenfassung (Kap. 4.8) und Folgerungen (Kap. 4.9) schließen das Kapitel 4 ab.

4.1 Relevanz von Beobachtung in der Frühpädagogik

Die Beobachtungskompetenz kann sich auf sehr unterschiedliche Formen von Beobachtung beziehen. Die folgende Übersicht nach Schmidt-Atzert und Amelang (2012, S. 310) zeigt wesentliche domänenunabhängige Aspekte auf, welche die Vielzahl möglicher Varianten von Beobachtung verdeutlicht:

- frei oder systematisch
- direkt oder indirekt (anhand von Aufzeichnungen)
- in natürlicher Umgebung oder in einer Situation, die vom Beobachter geschaffen wurde
- verdeckt oder offen
- wenn offen, dann teilnehmend oder nicht teilnehmend
- Selbst- oder Fremdbeobachtung

Die Beobachtungskompetenz bezieht sich folglich auf nahezu alle aufgezählten Formen von Beobachtung. In erster Linie geht es bei der Beobachtungskompetenz jedoch um eine freie (bzw. semiformelle), direkte, offene und teilnehmende Fremdbeobachtung in natürlicher Umgebung.

Die Beobachtungskompetenz ist deshalb so wichtig, weil die Beobachtung in der Frühpädagogik eine hohe Relevanz hat. Im frühpädagogischen Kontext ist Beobachten traditionell eine wichtige Aufgabe, die in den letzten Jahren weiter an Bedeutung gewonnen hat (Leu, 2013, S. 250). Während Röbe, Huppertz und Füssenich (2010, S. 16) noch feststellten, dass Beobachtung und Dokumentation nicht zum Kerngeschäft der Erzieherinnen gehören, wird bereits drei Jahre später Beobachten als pädagogische Fachaufgabe als in der Praxis angekommen und Teil des Arbeitsalltags von Erzieherinnen bestätigt (Viernickel et al., 2013, S. 91).

Die freie Beobachtung ist das am häufigsten genutzte Beobachtungsverfahren in Kindertagesstätten (Viernickel et al., 2013, S. 888–889). Sie spielt damit eine zentrale Rolle für den Alltag des Kindergartens. Im Zusammenhang mit Beobachtungskompetenz ist deshalb vor allem die freie Form von Beobachtung relevant. Die Studie zeigt, dass freie

Beobachtungen und Kurzzeitbeobachtungen mit und ohne schriftliche Notizen das Mittel der Wahl sind, wenn es um die Beobachtung von wenigen Kindern geht (vgl. Tab. 7).

Tabelle 7: Nutzung von Beobachtungsverfahren in Prozent nach Viernickel et al. (2013): mind. für einige wenige Kinder (S. 88), für alle Kinder (S. 89)

Beobachtungsverfahren	<i>N</i>	Verfahren wird genutzt für einige wenige Kinder	Verfahren wird nicht genutzt	Verfahren wird genutzt für alle Kinder	Verfahren wird nicht genutzt
Freie Beobachtungen/Kurzzeitbeobachtungen mit schriftlichen Notizen	1.199	94,5	5,5	52,5	47,5
Freie Beobachtungen/Kurzzeitbeobachtungen ohne schriftliche Notizen	1.143	90,8	9,2	65,2	34,8
weitere ...					

Die Studie liefert Hinweise darauf, dass die Fachkräfte in der Praxis neben der vorrangig freien Beobachtung eine hohe Affinität zu offenen und an individuellen Entwicklungsverläufen orientierten Beobachtungsverfahren mitbringen (Viernickel et al., 2013, S. 93). Viernickel et al. (2013) stellen fest, dass in der Praxis insbesondere freie Beobachtungen und selbst entwickelte Beobachtungsverfahren eingesetzt werden, dass diese als vollkommen ausreichend für das Erkennen von Entwicklungsdefiziten betrachtet werden und dass nur im Falle eines konkreten Verdachts auf besondere Entwicklungsverzögerungen oder -besonderheiten ein diagnostisches Verfahren eingesetzt wird (ebd., S. 93). Die Autoren stellen dabei die Limitationen heraus, welche durch Validität und Reliabilität sowohl der Beobachtungsergebnisse als auch der Anschlussfähigkeit der Förderstrategien gegeben sein können, wenn Beobachtungskompetenz fehlt.

Es handelt sich jedoch in der Regel nicht um spontane und willkürliche Alltagsbeobachtung, die einer *informellen Diagnostik* (Schrader, 2006, S. 99) zugeordnet würde, sondern um gerichtete (Alltags-)Beobachtungen, welche zur semiformalen Diagnostik (Hascher, 2008, S. 75) gehören. Ob diese Beobachtungen direkt oder anhand von z.B. Filmaufnahmen indirekt erfolgen, ist weniger relevant. Auch ist denkbar, dass Beobachtungskompetenz in natürlicher Umgebung oder in Situationen, die vom Beobachter geschaffen wurden, zum Tragen kommt. Sowohl verdeckte als auch offene Beobachtungen können relevant sein. Wenn offen beobachtet wird, ist sowohl eine teilnehmende als auch eine nicht teilnehmende Form denkbar. Weil es bei der Beobachtungskompetenz vornehmlich um die situierte Beobachtung von Kindern geht, ist die Selbstbeobachtung hier nicht relevant.

Der Beobachtung kindlicher Bildungs- und Entwicklungsprozesse wird eine zentrale Funktion für das pädagogische und fachdidaktische Handeln von Erzieherinnen zugeschrieben (Lonnemann & Hasselhorn, 2018, S. 133), womit die Beobachtungskompetenz eine Schlüsselposition einnimmt. Die Bedeutung gezielter Beobachtungen im Elementarbereich betonen Jugend- und Kultusministerkonferenz (JMK & KMK, 2004) in ihrem gemeinsamen Rahmen der Länder für die frühe Bildung in Kindertageseinrichtungen: Die Kinder sollen im Hinblick darauf beobachtet werden, welche Stärken und Schwächen sie im jeweiligen Bildungsbereich haben und wie sie Anregungen aufnehmen. Systematische Beobachtung und Dokumentation der kindlichen Entwicklungsprozesse sind erforderlich (ebd., S. 5). Schulz und Cloos (2013, S. 788–789) sehen in dieser Forderung die fachliche Herausforderung, die Beobachtung und Dokumentation kindlichen Tuns als Merkmal professioneller Organisationen systematisch in den pädagogischen Alltag einzubinden und die Beobachtungskompetenzen der Mitarbeitenden zu schulen.

Der Beobachtungskompetenz kommt in der Praxis des Kindergartens gerade im Zusammenhang mit bereichsspezifischen Förderprozessen eine wachsende Bedeutung zu, auch wenn sie als Begriff meist umschrieben wird. So stehen Umschreibungen beispielsweise für eine Steigerung der Beobachtungskompetenz wie die ‚Sensibilisierung der Wahrnehmung‘ oder der ‚wahrnehmenden Beobachtung‘, der gemäß Ruppig, Prigge, Pages und Adam (2015, S. 156) ein besonderer Stellenwert zugeschrieben wird. Auch das Konzept der wahrnehmenden Beobachtung nach Schäfer (2010) weist Parallelen zur Beobachtungskompetenz auf – gerade im Hinblick auf die situierte Wahrnehmung kindlichen Tuns.

Beobachten und Dokumentieren werden als zentrale Aufgaben der frühpädagogischen täglichen Praxis gesehen (Roux, 2007, S. 3) und existieren als Begriffspaar in den Bildungsprogrammen aller Bundesländer (Viernickel et al., 2013, S. 6). Schulz und Cloos (2013, S. 788) erläutern Beobachtung und Dokumentation im Hinblick auf den professionellen Anspruch genauer: Sie verstehen Beobachten zunächst als eine alltagsnahe Wahrnehmungstätigkeit, die erst durch systematisierte Verfahren der regelmäßigen Aufzeichnung (Audio-, Bild-, Schriftmedien), Auswertung und Reflexion zur professionellen Methode wird. Der Begriff der Beobachtungskompetenz wird dabei nicht näher ausgeführt. Im Hinblick auf kompetenzorientierte Verfahren ließe sich Beobachtungskompetenz bei Schulz und Cloos (2013) jedoch auch als Expertise im Hinblick auf den Einsatz von Beobachtungsverfahren deuten. Dieser Einsatz ist jedoch für einige Erzie-

herinnen schwierig; sie nehmen die Umsetzung von Beobachtungsverfahren als sehr anspruchsvoll wahr und fühlen sich davon überfordert (Ruppin et al., 2015, S. 147).

Auch ohne direkten Bezug auf die Kompetenz zum Einsatz von Verfahren ist Beobachtung und Dokumentation mit Unsicherheiten auf Seiten der Erzieherinnen verbunden ist (Beher & Walter, 2012, S. 29). Dabei wird das Beobachten und Dokumentieren als Teil eines komplexen Aufgabenprofils aufgrund der Gleichzeitigkeit vieler Anforderungen von zwei Dritteln der Erzieherinnen als Belastung erlebt, wie eine repräsentative Befragung in Nordrhein-Westfalen belegt (Viernickel, Voss & Mauz, 2017, S. 80–81). Zudem wird die Aufgabe des Beobachtens und Dokumentierens von der Hälfte der Erzieherinnen als ‚eher schwierig‘ bis ‚sehr schwierig‘ wahrgenommen (Röbe et al., 2010, S. 16), weshalb sowohl Erzieherinnen als auch Führungskräfte die Schulung dieses Bereichs als zentralen Weiterbildungsbedarf ansehen (Beher & Walter, 2012, S. 32). Dies ist Ausdruck des Bedürfnisses nach einer verbesserten Beobachtungskompetenz. Der Weiterbildungsbedarf von pädagogischen Fachkräften im Bereich Beobachtung und Dokumentation wurde auch von Fachberaterinnen mit 34% am zweithäufigsten benannt, wie eine Studie von Leygraf (2013, S. 47) zeigt.

Im Rahmen der aktuellen Erzieherinnenausbildung, welche König, Kratz, Stadler und Uihlein (2018) zufolge inzwischen als Bildungsberuf konturiert wird, zeigt eine länderübergreifende Auswertung der Lehrpläne, dass Beobachtung in mehreren Lernfeldern des länderübergreifenden Lehrplans verortet ist (ebd. S. 34–35), wobei in der rheinland-pfälzischen Ausbildung von Erzieherinnen „Entwicklungsprozesse beobachten, reflektieren und dokumentieren“ ein eigenes fachrichtungsbezogenes Modul ist (ebd., S. 37; Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur, Rheinland-Pfalz, 2011, S. 18). Innerhalb dieser Aus- und Weiterbildungsbemühungen wird Beobachtung jedoch meist mit als Vorstellung von Beobachtungsverfahren thematisiert. Die freie Beobachtung und die Schulung der Wahrnehmung spielen in diesem Kontext eine nachgeordnete Rolle und dies, obwohl die freie Beobachtung im Kindergartenalltag gemäß einer Studie von Viernickel et al. (2013) am häufigsten vorkommt.

In der frühpädagogischen Praxis spielen dabei vor allem die alltäglichen Beobachtungen eine zentrale Rolle, wie bereits die Begleitstudie zur Umsetzung der Bildungs- und Erziehungsempfehlungen für Kindertagesstätten in Rheinland-Pfalz (Honig, Schreiber & Lang, 2006) zeigt. Auch aktuellere Studien weisen darauf hin, dass Erzieherinnen in ihrer Berufspraxis weitgehend routiniert arbeiten, jedoch nicht explizit auf wissenschaft-

liches oder methodisch-didaktisches Wissen aus ihrer Qualifikationsphase zurückgreifen (Thole, Göbel, Milbradt, Reißmann & Wedtstein, 2015, S. 126).

Zu berücksichtigen ist, dass sich die fröhpädagogische Praxis in den letzten Jahren verändert hat. Durch die Entwicklung der Kindertagesstätten zur ersten Stufe im Bildungssystem (Anders, 2018, S. 186) hat sich der Aufmerksamkeitsfokus von Erzieherinnen von einem weiten Spektrum hin zu einem gezielten Beobachten verändert (vgl. Schulz & Cloos, 2013, S. 789). Im Zentrum des Beobachtungs- und Dokumentationsinteresses stehen Bildungsprozesse und Kompetenzentwicklung genauso wie besondere Begabungen und Risikofaktoren (ebd.). Dabei kommt der Kompetenz zur Beobachtung ein besonderer Stellenwert zu. Knauf (2019, S. 7–8) zeigt mit einer Analyse von in Deutschland verbreiteten pädagogischen Ansätzen, dass die Beobachtung einen Kern der pädagogischen Konzepte bildet, die Bildungsdokumentation hingegen in den meisten Programmen eine nachgeordnete Rolle spielt.

Zusammenfassend ist zu konstatieren, dass ein auf Beobachtung basierendes Verständnis kindlicher Lernwege sowie das Wissen um den aktuellen Entwicklungsstand als Voraussetzungen für adaptive Lehrprozesse gelten (Petermann & Viernickel, 2012, S. 61), weshalb Erzieherinnen Beobachtungskompetenz benötigen. Eine kindorientierte Pädagogik ohne regelmäßige und gezielt eingesetzte Beobachtung ist nicht denkbar; die regelmäßige Beobachtung und insbesondere die Fähigkeit, Beobachtungen für die Schaffung passender Bildungsangebote zu nutzen, stellt ein unverzichtbares Element des pädagogischen Handelns von Erzieherinnen dar (Viernickel & Völkel, 2009, S. 32). Diese Fähigkeit wird in der vorliegenden Studie als Beobachtungskompetenz (vgl. Kap. 2.5) bezeichnet und im nächsten Kapitel näher ausgeführt, der Fokus liegt dabei auf der Kompetenz zur Beobachtung auf der Mikroebene der Interaktion. Beobachtungskompetenz wird in der vorliegenden Arbeit als Voraussetzung für eine hohe Prozessqualität gesehen, v.a. im Hinblick auf qualitativ hochwertige Interaktionen oder bei der Bereitstellung und Gestaltung anregender Lernumgebungen, was letztlich zu einer hohen Bildungsqualität führt (vgl. Anders, 2018).

4.2 Begriffsklärung

Der Begriff Beobachtungskompetenz ist im Kontext der deutschsprachigen einschlägigen Fachliteratur noch nicht etabliert, wie das Ergebnis von lediglich 21 Treffern in der Datenbank FIS-Bildung (Stand 15.02.2020) zeigt.

Die Beobachtungskompetenz ist unter unterschiedlicher Schwerpunktsetzung an verschiedenen Stellen normativ verankert und wird neben einer Vielzahl von Begriffen, die wenig trennscharf sind, verwendet. Beobachtungskompetenz findet sich im gemeinsamen Orientierungsrahmen „Bildung und Erziehung in der Kindheit“ der Kultusminister- und der Jugend- und Familienministerkonferenz (KMK & JFMK, 2010) als Teil der erforderlichen Qualifizierungen von Erzieherinnen. Beobachtungskompetenz wird sowohl von Wahrnehmungs-, Deutungs- und Reflektionskompetenz unterschieden als auch von diagnostischer (Basis-)Kompetenz, die im Kontext von Analysefähigkeit benannt wird (ebd., S. 7). Eine Trennung von Beobachtungs- und diagnostischer Kompetenz übernimmt auch die Autorengruppe Fachschulwesen im „Qualifikationsprofil ‚Frühpädagogik‘ – Fachschule/Fachakademie“ (2011, S. 12): die Begriffe Beobachtungskompetenz und Diagnosekompetenz werden beide aufgeführt. Während die beiden Quellen, die den Begriff der Beobachtungskompetenz normativ verwenden, die Wichtigkeit der Beobachtungskompetenz zeigen, findet sich in diesen jedoch keine Definition oder Operationalisierung von Beobachtungskompetenz, die wissenschaftlichen Ansprüchen genügen würde.

Ein sehr weit gefasstes Verständnis von Beobachtungskompetenz findet sich im frühpädagogischen Kontext bei Wildgruber (2010, S. 14), der Beobachtungskompetenz mit einem Beobachtungsbegriff verbindet, welcher zwar auch direkte Beobachtung meint, jedoch explizit das Handeln der Erzieherinnen im gesamten Beobachtungsprozess einschließt, wie z.B. Durchführung, Aufzeichnung, Auswertung und Reflexion von Beobachtungen oder Ableitung von Zielsetzungen und Planung. Ein sehr breites Verständnis von Beobachtungskompetenz vertritt auch Smit (2009), der folgende Komponenten einer förderorientierten Beurteilung erfasst: „systematisches Beobachten, individuelle Förderung, Diagnostizieren, Rückmeldekultur, Orientierungsgespräch, Selbstbeurteilung und Lernzielorientierung“ (ebd., S. 133; siehe auch Smit & Engeli, 2017). Trotz eines wissenschaftlichen Anspruchs bleiben diese beiden weitgefassten Beobachtungskompetenzbegriffe vage.

Der Begriff Beobachtungskompetenz wird von Oesterhelt (2018, S. 84–85) verwendet, die ihn mit dem Terminus „fokussierter Blick“ umschreibt, welcher sich auf verschiedene Ebenen richten kann: mit dem Fokus der Lehr-Lernsituation auf die Analyse unterrichtlicher Prozesse und mit dem Fokus auf einzelne Lernende auf eine Bestimmung von individuellen Lernausgangslagen oder auf die Lernstandserfassung.

In der Förderpädagogik findet sich der Begriff der pädagogischen Beobachtungskompetenz (Werning, 2007; Werning & Lichtblau, 2012), welche als systemisch-konstruktivistische Beobachtungskompetenz zur Erkundung, Nutzung und Aktivierung von Kompetenzen und Ressourcen gilt (Werning, 2007, S. 39–40). Dabei kommt dem Erkennen vorhandener Potentiale, Fähigkeiten und Ressourcen von Kindern eine besondere Bedeutung hinsichtlich der effektiven Förderung zu (ebd., S. 39; Werning & Lichtblau, 2012, S. 251).

Beobachtungskompetenz kann auch unter dem Begriff einer situationsbezogenen Diagnosekompetenz von Lehrkräften verortet werden, unter welcher, in Anlehnung an die Begrifflichkeit der pädagogischen Diagnostik, diejenigen Kompetenzen verstanden werden, „die notwendig sind, um Diagnoseanforderungen in Unterrichtssituationen mit informellen bis semiformalen Methoden zu bewältigen, zu impliziten Urteilen zu kommen und hierzu geeignete Modifikationsentscheidungen auf der Mikroebene zu treffen“ (Hoth, 2016, S. 99).

Es werden zwei Diagnosetypen identifiziert (ebd., S. 208–209): „Der fachlich-bewertende Diagnosetyp“ nimmt in Unterrichtssituationen besonders fachinhaltliche Aspekte des Unterrichtsgeschehens wahr und interpretiert diese. Bei der Fortführung von Unterrichtssituationen legt dieser erste Typus den Fokus bevorzugt auf fachinhaltliche Aspekte. Der zweite Typus, „der schülerorientiert-handlungsbezogene Typus“, zeichnet sich durch eine allgemeinpädagogische bzw. schülerorientierte Perspektive beim Wahrnehmen und Interpretieren von Unterrichtssituationen aus. Dieser Typus fokussiert besonders die Lernenden, deren Verhalten und Motivation, wobei er auch Aspekte des Classroom-Managements oder unterrichtsmethodische Entscheidungen beachtet. Bei Erzieherinnen dürfte insbesondere der zweite Typus vertreten sein, der fachlich-bewertende Diagnosetyp dagegen eine eher untergeordnete Rolle spielen.

Smit und Engeli (2017, S. 279) untersuchen formative Beurteilung als Teil des Gesamtkonzepts im jahrgangsgemischten Unterricht der Primarstufe, zeigen Einflüsse der formativen Beurteilung auf die Unterrichtsplanung auf und identifizieren zwei Gruppen

von Lehrpersonen bezüglich ihrer formativen Beurteilungskompetenz – eine mit einer traditionellen und eine sehr kleine Gruppe mit einer erweiterten Form (ebd., S. 297), die in Tabelle 8 gegenübergestellt sind.

Tabelle 8: Unterschiede zwischen Gruppen von Lehrpersonen nach Smit und Engeli (2017, S. 296)

	Traditionell	Erweitert
Lektionsplanung basiert eher nicht auf Beobachtungen und Lernkontrollen.	... auf Beobachtungen und Lernkontrollen.
Abklärung des Vorwissens erfolgt nicht (lediglich formative Lernkontrollen in begrenztem Umfang).	... erfolgt.
Beobachtungen werden nicht systematisch durchgeführt und festgehalten.	... gezielt geplant und schriftlich dokumentiert.
Individuelle Rückmeldungen erfolgen ohne entsprechende Lernentwicklungsdiagnostik.	... erfolgen durch Aufzeigen des Lernstandes und der individuellen Entwicklung.
Portfolios kommen nicht zum Einsatz.	... kommen kaum zum Einsatz.
Lerner selbstbeurteilung wird kaum genutzt.	... wird teilweise genutzt.
Elemente formativer Beurteilung kommen nur wenig zum Einsatz.	... kommen zum Teil regelmäßig und andere gelegentlich zum Einsatz.

Beobachtungskompetenz wird außerdem im Kontext der professionellen Wahrnehmung (z.B. Stürmer & Seidel, 2017) als “noticing as teacher expertise” (van Es & Sherin, 2002, 2006) diskutiert. Im Kern geht es hier um das Erkennen (*noticing*) und das wissensbasierte Verarbeiten (*knowledge based reasoning*) des Wahrgenommenen (Seidel et al., 2010; Sherin & van Es, 2009).

Ein der vorliegenden Studie sehr nahes Verständnis von Beobachtungskompetenz findet sich im Strukturmodell zu mathematikbezogenen Kompetenzen von Fachkräften von Hepberger, Lindmeier, Moser Opitz und Heinze (2017, S. 243), das gemäß Abbildung 8 aus Basiswissen, reflexiver und aktionsbezogener Kompetenz besteht.

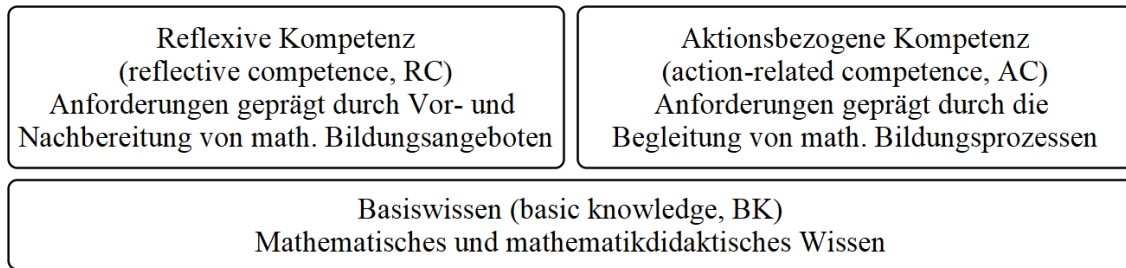


Abbildung 8: Strukturmodell zu mathematikbezogenen Kompetenzen pädagogischer Fachkräfte von Hepberger et al. (2017, S. 243)

Aktionsbezogene Kompetenz wird definiert als „Fähigkeiten und Fertigkeiten sowie die Bereitschaft, um die in der Durchführung von pädagogischen Angeboten mit Bezug zur mathematischen Kompetenzentwicklung auftretenden Anforderungen adaptiv zu bewältigen“ (Hepberger et al., 2017, S. 244). Die Kompetenz wird mittels Videovignetten erhoben, anhand derer die „Fähigkeiten zur spontanen adaptiven Handlung in der pädagogischen Situation in Form von geeigneten Impulsen als Reaktion auf Handlungen, Äußerungen oder Fragen der Kinder“ (ebd.) erfasst werden. Auch im Projekt *WILMA – Wir lernen Mathematik* (Pfennigwerth et al., 2016) wird diese aktionsbezogene Kompetenz anhand von Videosequenzen erhoben, zu denen sich das pädagogische Fachpersonal unter Speedbedingungen direkt und spontan verbal äußert (ebd.). Im Gegensatz zur reflexiven Kompetenz, die mit Bild- und Textvignetten eine andere Situationsspezifität hat (Dunekacke et al., 2018), steht die spontane situierte Diagnostik bei der aktionsbezogenen Kompetenz im Vordergrund (Lindmeier et al., 2020).

Beobachtungskompetenz wird in der vorliegenden Studie als die Kompetenz von Erzieherinnen aufgefasst, im laufenden (Lern-)Prozess mittels einer formativen, semiformalen Diagnostik – durch gezielte Beobachtung kindlicher Aktivitäten – Fähigkeiten von Kindern zu erkennen.

4.3 Beobachtungsfehler

Ein grundlegendes Problem der Beobachtungskompetenz besteht in der Inferenz, weil sich die der Messung zugrunde liegenden Beobachtungen auf einem Kontinuum von ‚niedrig-inferent‘ bis ‚hoch-inferent‘ anordnen lassen (vgl. z.B. Rosenshine, 1970). Von niedriger Inferenz spricht man, wenn Verhaltensweisen in einer Situation unmittelbar wahrgenommen werden können, also direkt beobachtbar sind, von hoher Inferenz, wenn sofort ein allgemeines Urteil abgegeben wird, bzw. wenn globale Einschätzungen auf

Basis von indirekten Beobachtungen getroffen werden müssen (vgl. Lutz, 2007, S. 89). Die Güte der Beobachtungskompetenz ist daher direkt abhängig vom Beobachtungsfokus und -gegenstand. Je nachdem, ob direkt beobachtbare Verhaltensweisen (z.B. das Kind zählt die Würfelbilder laut ab) oder nur indirekt Beobachtbares (z.B. das Kind ist motiviert) im Fokus stehen, liegen unterschiedliche Grade der Inferenz vor, die in unmittelbarem Zusammenhang mit der Güte oder mit der Fehleranfälligkeit von Beobachtungen stehen. Die Limitationen formativer Diagnostik im Hinblick auf Beobachtungsfehler und auf die Güte formativer Assessments werden durch die Anwendung des aus der allgemeinen Psychologie stammenden und auf konzeptueller Ebene sehr breit gefassten Linsenmodells (Brunswik, 1956) besonders deutlich, wie das Linsenmodell zur Urteilsgenauigkeit von Förster und Böhmer (2017) in Kapitel 3.1.1 und die Erweiterung in Kapitel 4.6.1 zeigen. Gerade in diesem erweiterten Modell wird dem Inferenzproblem, das im Kontext der Beobachtungskompetenz eine besondere Rolle spielt, in besonderem Maße Ausdruck verliehen.

Die Beobachtungskompetenz ist darüber hinaus mit einer Vielzahl von Problemen behaftet, weil sie im Kontext semiformeller Diagnostik gemäß Hascher (2008, 2011) nicht den Kriterien der formellen Diagnostik genügt, aber auch nicht nur zu impliziten Urteilen führt (Hascher, 2008, S. 75). Wie viele und welche Beobachtungsfehler auftreten, hängt davon ab, ob es sich um einen reflektierten oder einen unreflektierten Prozess handelt. Diesen beiden Formen der Prozesssteuerung, welche Kleber (1992, S. 104) als einander gegenüberstehende Pole eines Urteilsbildungsprozesses beschreibt, lässt sich die Beobachtungskompetenz nicht zuordnen. Die Beobachtungskompetenz lässt sich weder einem unreflektierten Prozess (informelle Diagnostik) noch einem reflektierten Prozess (formelle Diagnostik) zuschreiben. Sie liegt als semiformelle Diagnostik zwischen den beiden Prozessmerkmalen, wie die folgende Grafik zeigt.

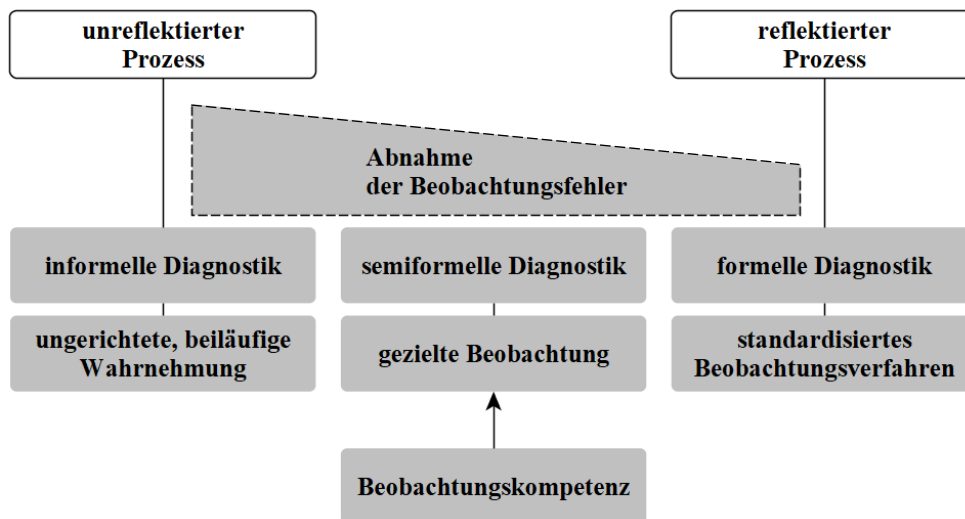


Abbildung 9: Abnahme der Beobachtungsfehler in Abhängigkeit vom Prozessstypus

Im alltäglichen Beobachten – also ohne die Verwendung von standardisierten Beobachtungsverfahren – erschließt sich einer Erzieherin nur das, was sie auch kennt. Umgekehrt heißt dies aber nicht, dass eine Erzieherin, die ein breites Wissen über Stufen oder Ausprägungen z.B. mathematischer Kompetenzen von Kindern besitzt, diese Stufen in der Praxis auch erkennen kann, also zur Verknüpfung ihres theoretischen Wissens mit Beobachtungen fähig ist. Wahrnehmung und Urteilsbildung sind nach Kleber (1992, S. 103) kognitive Grundaktivitäten des Menschen in seiner Auseinandersetzung mit der Umwelt, insbesondere mit der sozialen Umwelt stellen sie eine Grundlage für Handeln im sozialen Kontext und damit auch für diagnostisches Handeln dar.

Verschiedene Phänomene (z.B. Halo-Effekt, Zusammenhangsfehler etc.) können den Blick der Erzieherin auf das Kind verzerren (Wirsing, 2010, S. 463) und sind daher für die Beobachtungskompetenz relevant.

Die bei freien Beurteilungsvorgängen möglichen problematischen Wahrnehmungstendenzen, die zu Urteilsfehlern führen, kategorisiert Lutz (2007, S. 90) wie folgt:

- Wichtiges wird vereinfacht oder gar nicht wahrgenommen (wegen Ermüdung, Ablenkung, Überforderung, ungünstiger Beobachtungssituation).
- Wichtiges wird verzerrt wahrgenommen oder falsch zugeordnet und beurteilt (Tendenz zur Mitte-Fehler, Milde-/Strenge-Fehler, logische Fehler, Halo-Effekt).

- Wichtiges wird durch ‚Übertragung‘ oder Vergleich mit der eigenen Person verfälscht (Kontrast-/Ähnlichkeitsfehler).
- Wichtiges wird wahrgenommen und richtig beurteilt, aber die Beobachtung beeinflusst selbst die weiteren Ereignisse (Erwartungs-/Pygmalion-Effekt; Primacy-Effekt).

Die Entstehung von Beobachtungsfehlern wird dadurch begünstigt, dass eine Erzieherin im Alltag als aktiv Beteiligte an einer pädagogischen Situation diese beobachtet und Rückschlüsse zieht, welche wiederum direkte Folgen auf die pädagogische Situation haben können oder in eine Beurteilung einfließen.

Die Qualität der Beobachtung ist stark abhängig vom Beobachtungsfokus. Was beobachtet eine Erzieherin? Woran erkennt sie bestimmte Fähigkeiten von Kindern? Auf Basis welchen Vorwissens nimmt sie Beobachtungen vor?

4.4 Gütekriterien

Objektivität, Reliabilität und Validität sind für diagnostische Urteile im Sinne psychometrischer Gütekriterien (Helmke, 2015, S. 124–126) bedeutsam, wobei diese im Hinblick auf die Beobachtungskompetenz in besonderer Form berücksichtigt werden. Nach Schmidt (2020) erfüllt die Beobachtung als ein klassisches diagnostisches Verfahren im Kontext formativen Assessments zwar keine Testgütekriterien im engeren Sinne, liefert jedoch wertvolle diagnostische Informationen zu Denk- und Lernprozessen (Schmidt, 2020, S. 47).

Für die der Beobachtungskompetenzmessung zugrunde liegenden Beobachtungen müssen ebenso wie für die lernbegleitende Diagnostik im Allgemeinen, sofern man diese im Kontext und unter Berücksichtigung der speziellen pädagogischen Funktion des formativen Assessments beleuchtet, besondere Anforderungskriterien bzw. alternative Güte-merkmale gesucht werden (Weinert & Schrader, 1986, S. 18).

Wie die Hauptgütekriterien auch im Kontext eines formativen Assessments gesichert werden können, beschreibt Moser Opitz (2006, S. 22) mit einer Vielzahl von Ankerbeispielen: Objektivität, intersubjektive Nachvollziehbarkeit also, und Transparenz bedeutet,

- „dass die Beschreibung der Beobachtungen so dokumentiert ist, dass verschiedene Personen diese nachvollziehen können,
- dass eine zweite Person die Gültigkeit der Beobachtung(en) überprüft, indem sie ein bestimmtes Verhalten in einer bestimmten Situation auch beobachtet,
- dass die Interpretation der Beobachtungen und Beschreibungen offengelegt und begründet wird,
- dass diese Interpretation durch die Einholung von ‚Expertenurteilen‘ von Kolleginnen und Kollegen überprüft wird usw.

Reliabilität kann angestrebt werden, indem ...

- dieselbe Aufgabe in verschiedenen Situationen oder zu verschiedenen Zeitpunkten gestellt wird,
- mehrere ähnliche Aufgaben vorgelegt werden, die Wirksamkeit von Fördermaßnahmen überprüft wird usw.

Validität kann zu erreichen versucht werden, wenn z.B. ...

- die Herleitung der Beobachtungskategorien und -aufgaben fachlich begründet wird,
- Förderziele anhand von fachlichen, fachdidaktischen, pädagogischen und lern- bzw. entwicklungspsychologischen Kriterien begründet werden können,
- unter Kolleginnen und Kollegen diskutiert wird, ob die ausgewählten Aufgaben auch wirklich den Bereich erfassen, der interessiert (...).“

(Moser Opitz, 2006, S. 22)

Die zuletzt genannte kollegiale Validität spielt in der Frühpädagogik eine besonders große Rolle (Laewen, 2013). Die Chance besteht im gemeinsamen Interpretieren von Beobachtungen. Notwendig hierfür ist der Zugang zu Teamressourcen, insbesondere, weil „wissenschaftlich begründbare Konzepte zur Interpretation der Bedeutung des Verhaltens von Kindern, das über die Beobachtungen erfasst und dokumentiert wird, nicht in ausreichendem Umfang vorliegen“ (Laewen, 2013, S. 100–101).

Die Tatsache, dass es gängige Praxis ist, die Ergebnisse der Beobachtungen mit der unmittelbaren Kollegin zu diskutieren, zeigt eine Befragung von Viernickel et al. (2013, S. 94). Auf Teamressourcen basierende Reflexionen von Beobachtungen helfen im laufenden Förderprozess jedoch nur bedingt weiter. Entscheidend ist zunächst die Kompe-

tenz von Erzieherinnen, im laufenden (Lern-)Prozess bestimmte Fähigkeiten eines Kindes zu erkennen, um an diesen anzuknüpfen.

Diese Form der Validierung in der frühpädagogischen Praxis entspricht wissenschaftlichen Ansprüchen kaum, aber auch im schulischen Kontext kann nicht an jede Diagnostik ein wissenschaftlicher Maßstab angelegt werden. Helmke sieht ebenfalls einen pragmatischen Ansatz im Umgang mit Gütekriterien als gerechtfertigt an (2015, S. 126–127). Bereits Weinert und Schrader (1986, S. 18) stellen fest, dass die Diagnosen in der täglichen praktischen Arbeit „keineswegs besonders genau sein (müssen), wenn sich der Diagnostiker der Ungenauigkeit, Vorläufigkeit und Revisionsbedürftigkeit seiner Urteile bewusst ist“. Diese Entwicklung eines Bewusstseins für Revisionsbedürftigkeit stellt jedoch eine große Herausforderung dar.

Einen gewonnenen Eindruck oder eine Einschätzung im Nachhinein für falsch zu halten und getroffene Urteile zu revidieren, sieht Winter (2015, S. 89) als Herausforderung an und weist darüber hinaus auf die Schwierigkeit der Vermengung einer Vielzahl von Eindrücken bei der Urteilsbildung hin. Dabei besteht prinzipiell „kein Anlass, den Alltags-Diagnosen einer Lehrkraft grundsätzlich zu misstrauen. Schon in der Auswahl normaler Lernaufgaben sind Fördererwartungen aufgrund von vorausgehenden Beobachtungen und Strukturvorstellungen des Lernprozesses enthalten; in ihnen ‚steckt‘ also bereits eine bestimmte Diagnose“ (Lutz, 2007, S. 85).

4.5 Beobachtungskompetenz im Kontext formativen Assessments

4.5.1 Relevanz für die Diagnosekompetenzforschung

Die Forschung zu formativem Assessment, die in den letzten Jahren stark zugenommen hat, ist inhaltlich breit gefächert (zum Überblick siehe Schütze et al., 2018). Aus der Tradition der Diagnosekompetenzforschung finden sich nur wenige belastbare Ansätze bzw. Verknüpfungen: „Trotz ihrer Relevanz blieb die formative, lernprozessorientierte Perspektive in der bisherigen Forschung zu diagnostischer Kompetenz von Lehrkräften weitgehend unberücksichtigt. Dies ist insbesondere im Hinblick auf fehlende empirische Befunde nachvollziehbar, da beispielsweise die Erfassung der Lehrerurteile und der entsprechenden Kriterien bei der informellen „on-the fly“- sowie der semi-formellen „planned-for-interaction“-Diagnostik erhebliche methodische Herausforderungen in

sich birgt“ (Glogger-Frey & Herppich, 2017, S. 45). Eine dieser Herausforderungen ist im Hinblick auf die Beobachtungskompetenz in der Fehleranfälligkeit von Beobachtungen im Allgemeinen zu sehen.

Die bisherige Diagnosekompetenzforschung lässt trotz der erkannten Relevanz die formative, lernprozessorientierte Perspektive weitgehend unberücksichtigt (Glogger-Frey & Herppich, 2017, S. 45). Obwohl der freien Beobachtung eine besondere Relevanz zugesprochen wird, gibt es lediglich in der Forschung zum formativen Assessment entsprechende Ansätze (z.B. Smit & Engeli, 2017) zur Untersuchung derselben oder der damit verbundenen Beobachtungskompetenz (z.B. Dunekacke et al., 2018).

Beobachtungskompetenz kann als Basis formativen Assessments verstanden werden, wenn die drei Phasen des formativen diagnostischen Prozesses nach Prenzel werden: „ein erreichter Kompetenzstand wird erkannt, ein dazu passendes pädagogisches Angebot einschließlich zur Lernausgangslage passender Materialien wird bereitgestellt, Lernen wird praktiziert und das Erreichte wird bilanziert, um die nächste passende Lernaktivität planen und praktizieren zu können“ (Prenzel, 2017, S. 19).

Dementsprechend wird Beobachtungskompetenz aufgrund ihres Prozessbezugs als Facette der diagnostischen Kompetenz zugeordnet, die Voraussetzung für ein erfolgreiches formatives Assessment ist.

Die Relevanz der Beobachtungskompetenz zeigt sich in Studien, in denen insbesondere Personen unter Verwendung informeller diagnostischer Instrumente ein differenziertes Lehren gelingt, wohingegen es Lehrenden nicht gelingt, formelle diagnostische Verfahren im Kontext formativen Assessments adäquat zur Gestaltung adaptiver Lernumgebungen zu nutzen (Schmidt, 2020, S. 256). Aus der Sicht von Grundschullehrkräften geht insbesondere die Nutzung informeller diagnostischer Verfahren, wie z.B. der Beobachtung, mit einer adaptiven Gestaltung von Lernumgebungen einher, während Lehrkräfte Erkenntnisse aus formellen diagnostischen Verfahren weniger gut zur Steigerung der Adaptivität von Lernsituationen von Lehrkräften nutzen können, wie eine Studie von Liebers, Schmidt, Junger und Prenzel (2019, S. 306) zeigt, die damit die Wichtigkeit von Beobachtungskompetenz unterstreicht.

Dabei ist herauszustellen, dass im Kontext formativen Assessments eine Vielzahl verschiedener Formen diagnostischer Kompetenz verortet werden kann, deren Gemeinsamkeit im Zweck der jeweils zugrundeliegenden Diagnostik zu finden ist. Diese Vielzahl an zugrunde liegenden Kompetenzen lässt sich anhand einer Begriffsklärung des

formativen Assessments, das mit sehr unterschiedlicher Schwerpunktsetzung theoretisch gefasst wird, wie folgt darstellen.

4.5.2 Arten formativen Assessments

Glogger-Frey und Herppich (2017) sprechen von unterschiedlichen „Arten formativer Diagnostik“ (ebd., S. 43); Schütze, Souvignier und Hasselhorn (2018) von „Spielarten“ formativer Diagnostik und unterscheiden hierbei (ebd., S. 704) die Spezifität der Assessmentinformation (*monitoring assessment, diagnostic assessment, instructional assessment*) gemäß Wiliam (2010), den Zeitraum, der zwischen Assessment und dessen instruktionaler Nutzung liegt (*asynchronous or synchronous moments of contingency*) gemäß Black & Wiliam (2009) und den Planungs- und Formalitätsgrad des Assessments (*on-the-fly, planned-for-interaction, curriculum-embedded*) gemäß Heritage (2007) und Shavelson et al. (2008). Die Begriffe formative Evaluation, formative Leistungsmessung, formative Beurteilung, formatives Assessment oder auch förderorientierte Beurteilung werden nahezu synonym verwendet und dennoch finden sich perspektivische Unterscheidungen (vgl. Bürgermeister, 2014; Maier, 2010; Schütze et al., 2018; Smit & Engeli, 2017). Formatives Assessment definiert sich über die Funktion (vgl. Scriven, 1967, S. 43), welche in der Optimierung von Lehr-Lernprozessen besteht. Es geht also nicht um die Feststellung von Leistungsständen mit dem Ziel eines Vergleichs oder der Selektion, wie es bei der Statusdiagnostik der Fall ist, sondern um informelle oder semi-formelle Diagnostik. Daher wird die Beobachtungskompetenz als eine der Voraussetzungen für formatives Assessment verstanden. Nach Schmidt (2020, S. 252) geht der Begriff formatives Assessment deshalb weit über die reine Beschreibung von Lern- und Entwicklungsständen hinaus, weil er sowohl diagnostische als auch didaktische Prozesse unter der Maßgabe der Optimierung der Lehr-Lernprozesse einschließt. Deshalb sieht sie eine zu starke Betonung des diagnostischen Anteils und schlägt den neuen Begriff „Formative Leistungsbeurteilung und Unterrichtsgestaltung“ (ebd.) für den deutschen Sprachraum vor. Unterschiedliche Perspektiven auf formatives Assessment werden insbesondere durch die Konzeptualisierung dieses sehr weit gefassten Konstrukts deutlich. Maier (2015) weist darauf hin, „dass das Verständnis von formativer Leistungsdiagnostik von Studie zu Studie in einer gewissen Bandbreite variieren kann“ (ebd., S.121).

Formative Diagnostik ist letztlich ein Sammelbegriff prozessbezogener Ansätze und Methoden, welche das Verstehen (*understanding*) (Black & Wiliam, 2009) und das

Herausfordern kindlicher Denkprozesse (*elicit student thinking*; Briggs, Ruiz-Primo, Furtak, Shepard & Yin, 2012, S. 14) fokussieren. Dabei dient formative Evaluation stets der Adaption von Lehrprozessen (Scriven, 1967, S. 43, 1972, S. 62). Das Modell von Ruiz-Primo und Furtak (2007, S. 78) umfasst einen Dreischritt aus Hervorrufen (*eliciting*), Erkennen (*recognizing*) und Nutzen der Erkenntnisse (*using information*) (ebd.). Diese theoretische Annäherung verdeutlicht, dass die Beobachtungskompetenz auch bezogen auf das Modell von Ruiz-Primo und Furtak (2007) nur einen Ausschnitt der notwendigen diagnostischen Kompetenzen für den zweiten Schritt im Modell – das Erkennen – verdeutlicht.

Die Unterscheidung in *on-the-fly formative assessment*, *planned-for-interaction formative assessment* und *embedded-in-the-curriculum formative assessment* wird von Glogger-Frey und Herppich (2017, S. 44) aufgegriffen und erweitert. Die Autoren wenden die Unterscheidung in formelle, semiformelle und informelle Diagnostik (siehe Kap. 2.2) nicht nur auf Diagnostik im Allgemeinen, sondern auch explizit auf den Bereich formative Diagnostik an (Glogger-Frey & Herppich, 2017).

Smit und Engeli (2017, S. 282) sprechen von verschiedenen „Umsetzungsformen“ formativer Beurteilung, die sich unabhängig vom Fach, in denen sie zum Einsatz kommen, als wirksam erweisen (vgl. auch Black, Harrison, Lee, Marshall & Wiliam, 2004; Greenstein, 2010) und zur Individualisierung und Adaption von Lehrprozessen genutzt werden (z.B. Black & Wiliam, 2009). Dabei gehen die Autoren nicht auf konkrete Formen der Umsetzung ein, sondern verweisen darauf, dass jede Bewertung, wenn sie einen direkten Nutzen für das Lernen der Schüler(innen) im Prozess hat, zum formativen Assessment werden kann (z.B. Black et al., 2004, S. 10). Die konkrete Umsetzung von formativem Assessment ist Gegenstand der Studie von Schmidt (2020), die ihre Ergebnisse und Überlegungen in folgender Definition zusammenfasst: „Die formative Leistungsbeurteilung und Unterrichtsgestaltung dient der Optimierung von Lehr-Lern-Prozessen durch die enge Verzahnung diagnostischen und didaktischen Handelns und den Einsatz lernförderlicher Maßnahmen. (...)“ (ebd., S. 253). Auch hier wird die Relevanz von Diagnostik und damit diagnostischer Kompetenz als Voraussetzung dieser deutlich.

4.5.3 Wirkung

Die Beobachtungskompetenz wird im Zusammenhang mit formativem Assessment empirisch untersucht und lässt sich in Anlehnung an Schmidt (2020) als eine Voraussetzung für gelingendes formatives Assessment auf der Ebene mikroadaptiver Tätigkeiten verorten.

Die folgenden Untersuchungen zeigen die Bedeutung diagnostischer Kompetenzen, die formativem Assessment zugrunde liegen. Hatties Megaanalyse liegen 30 Metaanalysen mit insgesamt 3835 Lernenden zugrunde; es zeigt sich hinsichtlich der Korrelation von formativem Assessment und Schulleistung eine Effektstärke von $d = .90$ (Hattie, 2013, S. 193). Mit Bezug auf die Metaanalyse von Fuchs und Fuchs (1986) stellt Hattie fest, dass formative Evaluationen effektiv sind und zwar unabhängig vom Alter der Lernenden, von der Dauer der Intervention, von der Häufigkeit der Maßnahme und von der Art des besonderen Förderbedarfs (Hattie, 2013, S. 215).

Einen deutlich geringeren Effekt des formativen Assessments mit einer mittleren Effektstärke von $d = .20$ zeigt hingegen die Metaanalyse von Kingston und Nash (2011, S. 33), die zudem die Domäne als einflussreichsten Moderator hervorhebt (ebd., S. 34). Formative Beurteilung führt zu höheren Leistungen als eine summative Rückmeldung mittels Note, stellt Bürgermeister (2014) fest. Den positiven Einfluss formativer Beurteilungsstrategien auf die Leistung zeigt die Metastudie von Black und Wiliam (1998). Robinson, Myran, Strauss und Reed (2014) konnten zeigen, dass die Fortbildung von Lehrpersonen hinsichtlich formativer Beurteilungsstrategien messbare Zuwächse in deren formativer Diagnosekompetenz bewirkt und höhere Schüler(innen)leistungen nach sich zieht. Die Umsetzungsqualität formativer Beurteilung hat einen positiven Einfluss auf die Effektstärke (Yin et al., 2008). Neben der Wirkung formativen Assessments auf die Leistung zeigen sich auch positive Einflüsse auf Aspekte der Selbstkompetenz (Smit, 2009) bzw. auf den Selbstwert und die der Schüler(innen) (Miller & Lavin, 2007).

Die Untersuchung von Schmidt (2020) zeigt zentrale Strategien zur Realisierung einer formativen Leistungsbeurteilung und Unterrichtsgestaltung auf: Dazu zählen die „transparente Kommunikation über Lernziele und Beurteilungskriterien, die Gestaltung kognitiv aktivierender Unterrichtsgespräche, die Aktivierung der Schülerinnen und Schüler zum selbstregulierten und kooperativen Lernen, die regelmäßige Verwendung informeller und formeller diagnostischer Verfahren und die Nutzung diagnostischer Informatio-

nen für konstruktives Feedback und pädagogische Entscheidungen.“ (Schmidt, 2020, S. 253).

Im frühpädagogischen Kontext zeigen die Ergebnisse der Studie *Researching Effective Pedagogy in the Early Years (REPEY)* (Siraj-Blatchford, 2007), in der Befunde aus dem *Effective Provision of Pre-School Education (EPPE) Project* (Sylva et al., 2004) in qualitativen Fallstudien weitergehend untersucht wurden, dass effektiv fördernde Erzieherinnen systematische Beobachtungen der Leistung von Kindern nutzen, um Förderung adaptiv zu gestalten (Siraj-Blatchford, 2007, S. 112–113), weshalb sie dem formativen Assessment zuzuordnen sind. In diesem Zusammenhang hat sich vor allem das länger anhaltende gemeinsame Denken als besonders wirksam erwiesen (ebd.).

Der Frage, wie Fachpersonen beobachten und dokumentieren, also nach der Form der Beobachtung und Dokumentation, geht ein weiterer Zweig der Forschung zu Beobachtungen im vorschulischen Bereich nach. Dieser fokussiert die Untersuchung des Einsatzes standardisierter Beobachtungen derart, dass die Arbeit mit bestimmten Beobachtungsinstrumenten evaluiert (z.B. Gasteiger, 2008, 2010; Walter-Laager et al., 2011) oder auch die Befragung frühpädagogischen Fachpersonals hinsichtlich ihres Beobachtungsverhaltens thematisiert wird (z.B. Bruns, 2014).

Quantitative Daten zu einer situierten Beobachtungskompetenz und einer Untersuchung des Zusammenhangs mit der Urteilsgenauigkeit von frühpädagogischem Personal fehlen.

Die Beobachtungskompetenz kann zusammenfassend als eine Voraussetzung für erfolgreiche Lernprozessbeobachtungen bezeichnet werden, die Teil des sehr weit gefassten diagnostischen Handelns im formativen Assessment ist. Sie ist auf der Mikroebene der Interaktion verortet und lässt sich als im laufenden Prozess einer Lernsituation relevante Kompetenz dem Bereich *on-the-fly formative assessment* (Glogger-Frey & Herppich, 2017, S. 44) zuordnen.

4.6 Modelle

4.6.1 Erweitertes Linsenmodell zur Beobachtungskompetenz

Die Einbindung der Beobachtungskompetenz als weitere Facette diagnostischer Kompetenz erfolgt auf Basis des Linsenmodells von Förster und Böhmer (2017, S. 47), wel-

ches bereits in Kapitel 3.1.1 vorgestellt wurde. Es wird zur Modellierung der Beobachtungskompetenz aufgegriffen und unter Berücksichtigung der Bedingungen des formativen Assessments für die vorliegende Studie entsprechend modifiziert. Die Modifikationen sind in Abbildung 10 rot hervorgehoben.

Das Linsenmodell von Förster und Böhmer (2017, S. 47) verdeutlicht, dass der Mensch sich nur direkt wahrnehmbarer Verhaltensweisen oder äußerlicher Signale bedienen kann, um sich nicht direkt beobachtbare Merkmale wie z.B. das Leseverständnis zu erschließen. Die Autorinnen unterscheiden dabei die Ebene des Kindes (Kriteriumswert) und die Ebene der Lehrperson (Einschätzung des Merkmals); die Linse symbolisiert dabei die Wahrnehmung der Lehrenden.

Dieses Linsenmodell wurde mit dem Ziel erweitert, die Beobachtungskompetenz im Kontext formativen Assessments einzuordnen. Weil die Kompetenz für prozessbezogene formative Diagnostik stets situiert, also an konkrete Situationen gebunden ist, werden die Signale, welche das Kind zu geben in der Lage ist, primär durch die jeweilige Lernsituation limitiert. Aus diesem Grund wurde die Beobachtungskompetenz zudem unter Berücksichtigung der Situiertheit prozessbezogener Diagnosen in das Modell eingearbeitet. Während das ursprüngliche Modell von Förster und Böhmer ausschließlich die Ebenen Kind und Lehrperson berücksichtigt, bezieht das erweiterte Modell (Abb. 10) die Ebene der Situation mit ein. Die Ebenen sind in Abbildung 10 durch gestrichelte Linien hervorgehoben.

Anhand des folgenden Beispiels aus dem frühpädagogischen Bereich wird das erweiterte Linsenmodell erläutert: Ein Kind besitzt eine gewisse mathematische Kompetenz, die nach wissenschaftlichen Gütekriterien mit einem standardisierten Test erfasst werden kann (Kriteriumswert des Merkmals T). In einer Lernsituation, beim Brettspiel mit zwei Würfeln, zeigt das Kind einen Ausschnitt seiner mathematischen Fähigkeiten, weil durch den Gebrauch von zwei Würfeln die Aktivitäten auf einen Zahlenraum bis zwölf begrenzt sind und zur Bewältigung des Spiels nur ein Teil der mathematischen Kompetenz des Kindes gebraucht wird. Das bedeutet eine Einschränkung der sichtbaren Merkmale durch die Situation (Situieretes Anforderungsprofil) und eine Veränderung der Merkmale, wenn sich die Situation verändert. Dieser Limitation der Signale, welche in der Situiertheit prozessbezogener Diagnostik begründet ist, wird Ausdruck verliehen, indem der Linse des erweiterten Linsenmodells (Abb. 10) ein Filter vorgeschaltet wird, der nur für gewisse Signale durchlässig ist (Situationsspezifische Linse). Die Lehrper-

son kann nur erkennen, was die Situation zulässt, was das Kind in einer gewissen Situation zeigt.

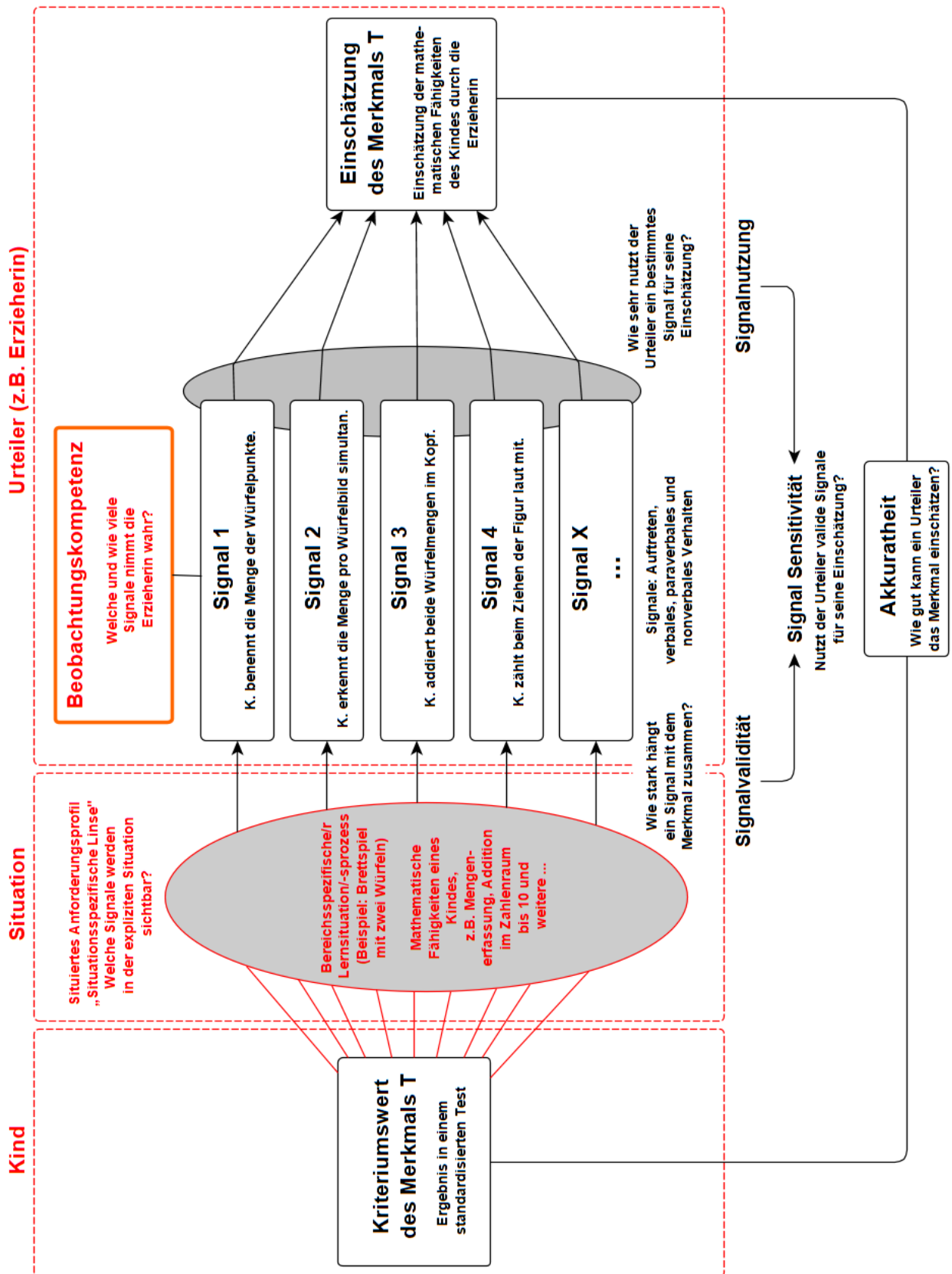


Abbildung 10: Erweitertes Linsenmodell diagnostischer Kompetenz in Anlehnung an Förster und Böhrner (2017), vgl. Abbildung 6

Dem Linsenmodell liegt die Annahme zugrunde, dass ein nicht direkt wahrnehmbares (distales) Merkmal wie z.B. die mathematische Kompetenz eines Kindes (vgl. Abb. 10), durch verschiedene wahrnehmbare Verhaltensweisen (proximale Merkmale, Signale, z.B. lautes Abzählen von Würfelaugen) erschlossen werden kann. Dabei sind sowohl die Signalvalidität als auch die Signalnutzung von Bedeutung. Die Signalvalidität, also das Ausmaß, in dem wahrnehmbare Signale tatsächlich mit dem einzuschätzenden Merkmal zusammenhängen, ist von der Sensitivität der Erzieherin abhängig. Je mehr valide Signale sie nutzt und entsprechend invalide Signale ignoriert, desto höher ist die Sensitivität. Förster und Böhmer (2017) sprechen in diesem Zusammenhang auch von Konsistenz, wenn gleiche Signale auf dieselbe Art und Weise genutzt werden, um verschiedene Kinder einzuschätzen.

Innerhalb des Modells zeigt sich die Beobachtungskompetenz der Erzieherin in der Summe der wahrgenommenen Signale in einer Lernsituation: Wie viele Signale nimmt die Erzieherin wahr? Welche Signale erkennt sie? Dabei bezieht sich die zweite Frage auf den Anspruch der Beobachtung. Werden nur offensichtliche Signale wahrgenommen und damit nur niedrig inferente Einschätzungen aufgrund direkter Beobachtungen vorgenommen (z.B. das Kind zählt laut bis ...) finden auch hoch inferente Einschätzung aufgrund indirekter Beobachtungen statt (z.B. das Kind erfasst die Würfelmenge simultan, weil es direkt die Züge und nicht die Würfelpunkte einzeln laut abzählt)?

Zur deutlichen Kennzeichnung der verschiedenen am Prozess beteiligten Ebenen oder Modellbereiche sind diese in Abbildung 10 beschriftet und mittels gestrichelter Linien eingefasst. Der ergänzte Filter bildet den Bereich Situation/Prozess, der im ursprünglichen Modell nicht existiert. Die Ebenen von Kind und Urteiler (z.B. Erzieherin) existieren bereits im Vorgängermodell (vgl. Kap. 3.1.1), sie werden jedoch nicht durch Beschriftung und Linien hervorgehoben.

Zusammenfassend zeigt das erweiterte Linsenmodell die Verortung der Beobachtungskompetenz im Verlauf des Urteilsprozesses und die formale Berücksichtigung der Situation in formativen Diagnoseprozessen als begrenzenden Filter für das Zustandekommen von Signalen. Die Beobachtungskompetenz wird operationalisiert als die Anzahl der durch die Situation limitierten Signale, die wahrgenommen werden. Eine Teilmenge der wahrgenommenen Signale (Linse auf Ebene der Urteiler) wird herangezogen, um zu einer Einschätzung des Merkmals (mathematische Kompetenz des Kindes) zu gelangen. Anhand dieser Einschätzung kann die Urteilsgenauigkeit bestimmt wer-

den, indem die Einschätzung mit dem objektiv gemessenen Kriteriumswert verglichen wird.

Herauszustellen ist, dass für die Beobachtungskompetenz die Frage nach der Signalnutzung irrelevant ist. Die Beobachtungskompetenz ist nämlich die der Signalnutzung zeitlich vorausgehende Wahrnehmung kindlicher Fähigkeiten und somit der Urteilgenauigkeit zeitlich vorgeschaltet.

4.6.2 Frühpädagogische Modelle und Modelle der Lehr-Lernforschung

Im Gegensatz zur Urteilgenauigkeit, die als Voraussetzung informeller summativer Diagnostik gesehen wird, kann die Beobachtungskompetenz den für semiformelle formative Diagnostik relevanten Kompetenzen Lehrender zugeordnet werden.

Formativ bedeutet, dass die Leistungsmessung als Teil des Lernprozesses verstanden wird und der Optimierung von Lernprozessen von Kindern dient (vgl. Maier, 2010, S. 294) z.B. durch Adaption von Lehr-Lernsituationen. Als semiformell wird eine Diagnostik bezeichnet, die zwar nicht den Kriterien der formellen Diagnostik genügt, aber nicht nur zu impliziten Urteilen führt. Dies trifft auf Beobachtungen zu, die zwar gezielt, aber nicht mit erprobten Methoden durchgeführt werden, oder auch auf intuitive Beobachtungen, die in explizite Bewertungen einfließen (vgl. Hascher, 2008, S. 75). Daher ist auch Beobachtungskompetenz immer situiert, sie kann nicht unabhängig von der Lehr-Lernsituation gesehen werden.

Aus diesem Grund ist sie in bestehende Kompetenzmodelle auch nur teilweise einzuordnen. Im Kompetenzmodell von Baumert und Kunter (2011), das sowohl zur theoretischen Fundierung der Diagnosegenauigkeit von Erzieherinnen (vgl. Dollinger, 2013, S. 38) als auch zur Darstellung der professionellen Kompetenzen von pädagogischen Fachkräften im Allgemeinen (vgl. Benz et al., 2017, S. 41) herangezogen wurde, kann die Beobachtungskompetenz theoretisch an mehreren Stellen verortet werden – im fachdidaktischen wie auch im pädagogisch-psychologischen Wissensbereich (vgl. Kap. 2.6.2). In den Kompetenzfacetten, also dem fachdidaktischen oder auch pädagogisch-psychologischen Wissensbereich oder der Verortung als Facette des Kompetenzbereichs fachdidaktisches Wissen, wird die Diagnostik von Schüler(innen)wissen bzw. diagnostische Kompetenz bei Baumert und Kunter (2011, S. 37–38) jedoch isoliert betrachtet und nicht situiert, wie es bei der Beobachtungskompetenz notwendig wäre. Diese sollte

wie im formativen Assessment zusammen mit den Handlungskompetenzen gedacht werden, obwohl diagnostische Kompetenz, unabhängig von den zugrunde gelegten Operationalisierungen verschiedener Facetten, vielfach einzig als Voraussetzung für das adaptive Gestalten von Unterricht gesehen wird (Südkamp & Praetorius, 2017). Was zunächst wie ein teilweiser Widerspruch erscheint, gewinnt im Kontext der Betrachtung von Beobachtungskompetenz als Teil formativen Assessments und damit einhergehender Modelle an Klarheit.

Im *Allgemeinen Kompetenzmodell* von Fröhlich-Gildhoff et al. (2011, S. 17), das für den frühpädagogischen Bereich konzipiert wurde, lässt sich die Beobachtungskompetenz unter dem Begriff *Situationswahrnehmung und Analyse* verorten. Diese Begriffswahl verdeutlicht die Relevanz des Beobachtungsaspekts, der direkten Einfluss auf die Handlungsplanung und -bereitschaft hat (vgl. Kap. 2.6.3).

Im vorschulischen Bereich finden sich Ansätze zur Erfassung situierter diagnostischer Fähigkeiten in der theoretischen Modellierung professioneller Kompetenzen von pädagogischen Fachkräften von Hepberger et al. (2017). Dieses dreiteilige Strukturmodell, das bereits zur Begriffsklärung herangezogen wurde (Abb. 8, Kap. 4.2) umfasst reflexive Kompetenz (Anforderungen geprägt durch Vor- und Nachbereitung von Bildungsangeboten), aktionsbezogene Kompetenz (Anforderungen geprägt durch die Begleitung von Bildungsprozessen) und das Basiswissen (Mathematisches und mathematikdidaktisches Wissen) (Hepberger et al., 2017, S. 243).

Bereits Knievel, Lindmeier und Heinze (2015) treffen diese Unterscheidung in reflexive und aktionsbezogene Kompetenz, die sie als die zwei Komponenten der fachspezifischen Lehrerkompetenz bezeichnen (ebd., S. 313), wobei sie Ähnlichkeiten zum Ansatz professioneller Wahrnehmung von Sherin und van Es (2009) herausstellen (Knievel et al., 2015, S. 314). Die Beobachtungskompetenz ist der aktionsbezogenen Kompetenz zuzuordnen, weil sie als prozessbegleitende Form diagnostischer Kompetenz nur im Kontext konkreter Lehr-Lernsituationen vorkommt.

Das Modell der kausalen Wirkkette professioneller Kompetenz pädagogischer Fachkräfte nach Seemann et al. (2018, S. 1225) greift diese Unterscheidung in reflexive und aktionsbezogene Kompetenz auf und zeigt, wie die aktionsbezogene Kompetenz, welche in der vorliegenden Arbeit der Beobachtungskompetenz entspricht, über die Lernbegleitung, die entsprechend eine Adaption der Lernumgebung darstellen kann, zu einem Kompetenzzuwachs auf Ebene der Kinder führt. Dieses Modell beinhaltet wie das er-

weiterte Linsenmodell (vgl. Kap. 4.6.1) die drei Ebenen der Lehrenden bzw. der pädagogischen Fachkräfte, der Situation bzw. Lernumgebung sowie der Kinder.

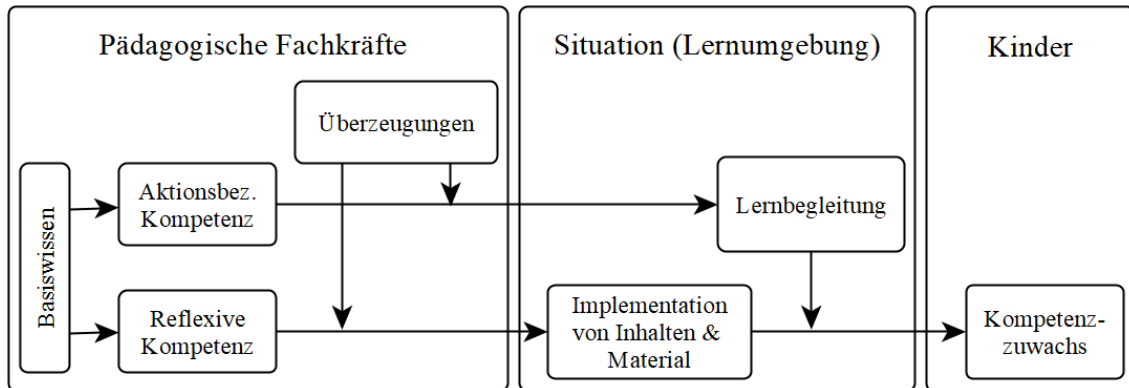


Abbildung 11: Modell der kausalen Wirkkette professioneller Kompetenz pädagogischer Fachkräfte nach Seemann et al. (2018, S. 1225)

Im Modell zu Fachwissen und Fähigkeiten für die frühe mathematische Bildung von Gasteiger und Benz (2018) lässt sich die Beobachtungskompetenz dem Bereich ‚situatives Beobachten und Wahrnehmen‘ zuordnen.

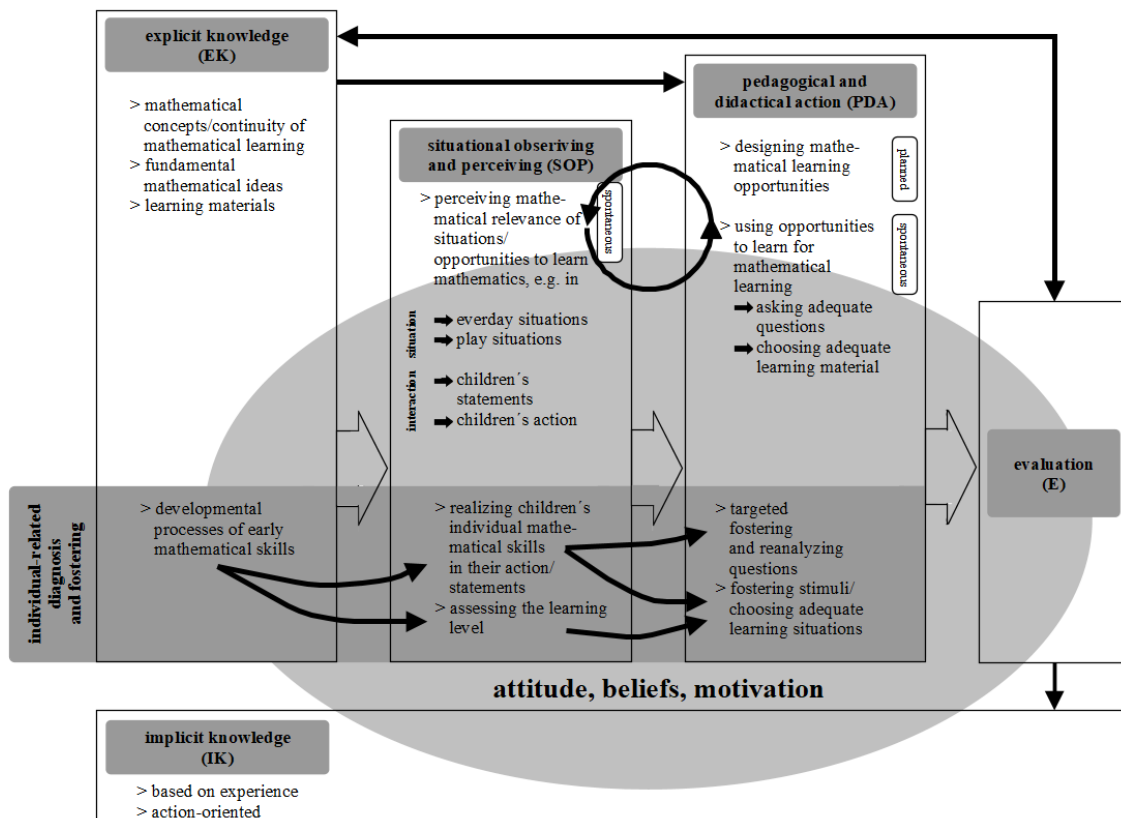


Abbildung 12: Model of professional knowledge and skills for early mathematics education (Gasteiger & Benz, 2018, S. 9)

Hier ist die Wahrnehmung von mathematischen Fördermöglichkeiten durch die Erzieherin verortet und das Erkennen individueller Fähigkeiten bzw. die Einschätzung des Lernlevels. An dieser Schnittstelle *situational observing and perceiving* (Situatives Beobachten und Wahrnehmen) wird zudem das spontane, erfahrungsorientierte, handlungsbasierte Agieren der Erzieherinnen hervorgehoben (Abb. 12).

Angedeutet durch eine Wechselwirkung (kreisrunder Pfeil in Abb. 12) wird der Zusammenhang zwischen dem Bereich ‚pädagogisch-didaktische Handlung‘ und dem Bereich ‚situatives Beobachten und Wahrnehmen‘ sehr deutlich. Diese Verbindung, das Zusammendenken von Wahrnehmung und Handlung bzw. Diagnostik und Lernförderung ist ein zentrales Element des formativen Assessments, dessen Relevanz für die Beobachtungskompetenz bereits erläutert wurde.

4.7 Veränderung der Beobachtungskompetenz

Aus dem beschriebenen erweiterten Linsenmodell (Kap. 4.6.1) können Rückschlüsse auf die Veränderung der einzelnen Facetten diagnostischer Kompetenz gezogen werden. Dabei wird gemäß diesem Modell die Lehr-Lernsituation, welche durch den methodisch-didaktischen Kontext gekennzeichnet sind, als relevantes Einflussmerkmal gesehen. Rückschlüsse auf die Auswirkung dieses Einflussmerkmals werden im Folgenden beispielhaft für den mathematischen Bereich anhand unterschiedlicher bereichsspezifischer Förderansätze theoretisch begründet.

Die Feststellung, dass eine Erzieherin grundsätzlich nur das sehen kann, was sie auch kennt, erscheint trivial, hat jedoch im Hinblick auf die Wirkung von Fachwissen auf die Beobachtungskompetenz bedeutsame Folgen.

Ist einer Erzieherin z.B. das Phänomen der simultanen Mengenerfassung und der Ordinalaspekt gänzlich unbekannt, beobachtet sie vielleicht, dass ein Kind bei einem Brettspiel zwei Würfelbilder schneller abzählt als der Spielpartner, nimmt aber nicht wahr, dass dieses Kind ein Würfelbild simultan erkennt und bereits ab der gesehenen Punktezahln weiterzählt, um auf die gewürfelte Summe zu kommen, während ein Spielpartner z.B. die Punkte beider Würfelbilder noch komplett abzählen muss. Hierfür braucht die Erzieherin entsprechendes Fachwissen, um Fähigkeiten und Fertigkeiten der Kinder gezielt beobachten zu können, also um Signale gemäß dem Linsenmodell (Kap. 4.6.1) als solche überhaupt wahrzunehmen.

Daraus folgt, dass eine Veränderung des Fachwissens z.B. durch den Besuch einer bereichsspezifischen Fortbildung Einfluss auf die Beobachtungskompetenz der Erzieherin haben müsste. Dies gilt nur, wenn die Vermittlung von Fachwissen über frühkindliche mathematische Fähigkeiten und Fertigkeiten Gegenstand der Fortbildung ist. In diesem Zusammenhang können z.B. Fortbildungen zu lehrgangsbasierten Ansätzen von Fortbildungen zu lernwegsorientierten Ansätzen unterschieden werden.

Keine oder nur geringe Veränderungen der Beobachtungskompetenz werden erwartet, wenn z.B. im Rahmen einer Fortbildung zu einem lehrgangsbasierten Programm aufeinanderfolgende Lerneinheiten vorgestellt werden, die nach einem vorgegebenen Schema ablaufen und in deren Mittelpunkt jeweils ein Inhaltsbereich (z.B. eine Zahl) steht. Auch durch die Vorstellung von Stundenbildern und Durchführungshinweisen zu Geschichten, Liedern und Medien, wie z.B. Zahlenpuppen, oder weiteren Materialien kann das Fachwissen über die Entwicklungsschritte mathematischer Fähigkeiten und Fertigkeiten weitgehend unverändert bleiben.

Sind die bereichsspezifischen Entwicklungsverläufe der Kinder jedoch die Basis für Förderung und die Gestaltung adaptiver Lernumgebungen, weil sich die Auswahl der Angebote nach dem Entwicklungsstand der Kinder richtet, wie es in lernwegsorientierten Ansätzen der Fall ist, ist eine Zunahme des Erzieherinnenfachwissens über mathematische Fähigkeiten und Fertigkeiten von Kindern zu erwarten und damit auch eine Zunahme der Beobachtungskompetenz.

Es wurde noch nicht untersucht, ob sich die Entwicklung der Beobachtungskompetenz in verschiedenen (lehrgangsbasierten und lernwegsorientierten) Konzepten so unterscheidet, wie es das folgende Beispiel vermuten lässt. Erzieherinnen, welche eine lehrgangsbasierte Förderung anhand von aufeinanderfolgenden Lerneinheiten durchführen und einem vorgegebenen Schema folgen, ohne Inhalte explizit auf den Lernstand und damit die Bedürfnisse der Kinder abzustimmen, lernen ihre Bezugskinder weniger gut kennen als die Erzieherinnen, welche in der gleichen Kontaktzeit die inhaltliche Planung vom Entwicklungsstand der Kinder abhängig machen, wie es in lernwegsorientierten Ansätzen der Fall ist.

Theoretisch enthalten lernwegsorientierte Ansätze implizit Anteile von Diagnostik, während Förderansätze, die systematisch aufgebaut sind und von der Sache ausgehen, keine implizite Diagnostik enthalten. Deshalb wird auf der einen Seite angenommen, dass zur Umsetzung eines lehrgangsbasierten Ansatzes weniger Beobachtungskompetenz ge-

braucht wird und sich diese entsprechend kaum entwickelt. Die Stundenbilder müssen inhaltlich weder an die Bedürfnisse einzelner Kinder noch an die der Gruppe angepasst werden, weshalb ein solcher Ansatz auch für verschiedene Erzieherinnen mit mittelmäßiger oder geringer Beobachtungskompetenz auf ähnlichem Niveau möglich ist.

Auf der anderen Seite wird angenommen, dass zur Umsetzung eines lernwegsbasierten Ansatzes ein hohes Maß an Beobachtungskompetenz notwendig ist und sich diese dementsprechend im Verlauf der Zeit weiterentwickelt. Die Fördersequenzen und -angebote müssen inhaltlich an die Bedürfnisse einzelner Kinder oder einer bestimmten Gruppe angepasst werden, weshalb die Umsetzung eines solchen Ansatzes stark von der Beobachtungskompetenz der umsetzenden Erzieherinnen abhängt. Im Rahmen eines Förderansatzes, der individualisiert und lernwegsorientiert ausgerichtet ist, müssen Erzieherinnen die Fähigkeiten von Kindern genauer beobachten als in einer lehrgangsorientierten Förderung, die keine Individualdiagnostik voraussetzt.

Im vorschulischen Kontext wird formatives Assessment im Projekt *WILMA – Wir lernen Mathematik* (Pfennigwerth et al., 2016) untersucht. Es werden dort Effekte professioneller Kompetenz von Erzieherinnen analysiert und mittels Interventionen gezielt variiert. Zugrunde gelegt wird dabei nach Seemann et al. (2018, S. 1226) das Modell von Hepberger et al. (2017, S. 243) bestehend aus Basiswissen, reflexiver und aktionsbezogener Kompetenz. In der aktionsbezogenen Kompetenz zeigen sich Bezüge zur Beobachtungskompetenz, weil auch hier Beobachtungen in konkreten Situationen stattfinden. In *WILMA* wird diese anhand von Videosequenzen erhoben, zu denen sich das pädagogische Fachpersonal unter Speedbedingungen direkt und spontan verbal äußert (ebd.). Im Gegensatz zur reflexiven Kompetenz, die mit Bild- und Textvignetten eine andere Situationsspezifität hat (Dunekacke et al., 2018), steht die spontane situierte Diagnostik bei der aktionsbezogenen Kompetenz im Vordergrund.

Erste Ergebnisse zeigen, dass unterschiedliche Trainingsschwerpunkte für Fachkräfte (Interventionsgruppe AC: Fortbildung für aktionsbezogene Kompetenz; Interventionsgruppe RC: Fortbildung für reflexive Kompetenz) im Vergleich zu einer Kontrollgruppe einen Zuwachs der reflexiven und aktionsbezogenen Kompetenz bewirken, es jedoch keinen Unterschied zwischen den beiden Interventionsgruppen gibt (Dunekacke et al., 2018). Eine mögliche Erklärung dafür sehen die Autorinnen in der kurzen Interventionszeit von zehn Stunden.

Zusammenfassend lässt sich festhalten, dass die Art des bereichsspezifischen Förderkonzepts (Lehrgangsbasierung vs. Lernwegsbasierung) von Bedeutung ist und Auswirkungen auf die Beobachtungskompetenz hat. Ein Kennenlernen relevanter Grundlagen bereichsspezifischer Förderung wird als Schlüssel zur Fokussierung bereichsspezifischer Fähigkeiten von Kindern gesehen.

Dabei kann eine Fortbildung direkte Wirkungen (Veränderung des Fachwissens) auf die Beobachtungskompetenz haben; auch hier kommt es auf die Art der Fortbildung (z.B. Lehrgang vs. Lernweg) an. Die Effekte einer Fortbildung auf die Beobachtungskompetenz sind daher direkt messbar – auch mit der Beobachtung von der Erzieherin fremden Kindern.

4.8 Zusammenfassung

Nachdem in Kapitel 3 gezeigt wurde, dass Urteilsgenauigkeit als alleiniges Verständnis von diagnostischer Kompetenz kritisch hinterfragt werden muss (z.B. Abs, 2007; Baumert & Kunter, 2006; Praetorius et al., 2012, 2017), zeigt Kapitel 4 mit der Facette Beobachtungskompetenz eine Erweiterung des diagnostischen Kompetenzbegriffs auf.

Hierzu wurde zunächst die Bedeutung der Beobachtung als Methode dargestellt, der Stellenwert der Beobachtung für die frühpädagogische Praxis und damit die Relevanz der Beobachtungskompetenz.

Es hat sich gezeigt, dass die Beobachtungskompetenz trotz ihrer Bedeutsamkeit bisher kaum untersucht wurde. Dies kann an den methodischen Herausforderungen liegen, die mit der Untersuchung semiformeller Diagnostik und der damit verbundenen Kompetenzen einhergeht: zum einen im Hinblick auf Beobachtungsfehler (Kap. 4.3), zum anderen aber auch auf den besonderen Umgang mit alternativen Gütekriterien (Kap. 4.4).

Bislang wurden lernprozessbezogene Aspekte diagnostischen Lehrerverhaltens kaum thematisiert, wenn es um die Bedeutung der diagnostischen Kompetenz geht (Baumert & Kunter, 2006; Praetorius et al., 2012), obgleich der Bedarf an Studien zu lernprozessbezogenen diagnostischen Kompetenzen immer wieder erkannt und formuliert wurde. Dies gilt auch für den frühpädagogischen Kontext.

Die Erweiterung des Linsenmodells von Förster und Böhmer (2017, S. 47), welches im Zusammenhang mit Urteilsgenauigkeit dargestellt wurde (Abb. 6, Kap. 3.1.1), zeigt,

wie Beobachtungskompetenz im diagnostischen Prozess verortet ist, und dass die Beobachtungssituation einen zentralen Einfluss auf die Wahrnehmung hat (Abb.10, Kap. 4.6.1). Zudem zeigt das erweiterte Linsenmodell, dass die Beobachtungskompetenz der Facette Urteilsgenauigkeit zeitlich vorausgeht.

Eine weitere Einordnung der Beobachtungskompetenz erfolgt durch die Verortung in frühpädagogischen Modellen und Modellen der Lehr-Lernforschung (4.6.2), wobei das Modell der kausalen Wirkkette professioneller Kompetenzen von Seemann et al. (2018, S. 1225) besonders relevant erscheint. Dieses Modell beinhaltet die aktionsbezogene Kompetenz, die in ihrer Operationalisierung mit der Beobachtungskompetenz vergleichbar ist.

Im Kontext der Forschung zu formativem Assessment finden sich Ansätze (z.B. Smit & Engeli, 2017) zur Untersuchung der freien Beobachtung und auch der damit verbundenen Beobachtungskompetenz (z.B. Dunekacke et al., 2018). Einzelne Studien fokussieren Kompetenzen, die Gemeinsamkeiten mit der Beobachtungskompetenz entsprechend der vorliegenden Studie aufweisen. Insbesondere die Studien der Forschungsgruppe WILMA (Pfennigwerth et al., 2016) zeigen mit dem Konstrukt der aktionsbezogenen Kompetenz Bezüge zur Beobachtungskompetenz auf (Lindmeier et al., 2020). Erste Ergebnisse belegen, dass unterschiedliche Trainingsschwerpunkte für Fachkräfte im Vergleich zu einer Kontrollgruppe zwar einen Zuwachs der reflexiven und aktionsbezogenen Kompetenz bewirken, es jedoch keinen Unterschied zwischen verschiedenen Interventionsgruppen gibt (Dunekacke et al., 2018).

4.9 Folgerungen

Während der Forschungsstand zur Beobachtungskompetenz sehr überschaubar ist, liefern bisherige Ansätze zur Erfassung des Konstrukts hilfreiche Modelle. Das Modell der kausalen Wirkkette professioneller Kompetenz pädagogischer Fachkräfte nach Seemann et al. (2018, S.1225) zeigt die Relevanz der Beobachtungskompetenz, welche die Autorengruppe als aktionsbezogene Kompetenz bezeichnet (ebd.), und der reflexiven Kompetenz für die Gestaltung von Lernumgebungen und damit für den Kompetenzzuwachs bei Kindern. Ein gerichteter Zusammenhang zwischen beiden Kompetenzen spielt in diesem Modell keine Rolle.

Die Erweiterung des Linsenmodells von Förster und Böhmer (2017, S. 47) verdeutlicht zum einen die Bedeutung der Situation, in der Beobachtungskompetenz gebraucht wird, und zeigt zum anderen einen möglichen zeitlichen Zusammenhang mit der Urteils-genauigkeit auf. Diese im erweiterten Linsenmodell diagnostischer Kompetenz (Kap. 4.6.1) abgebildete zeitliche Folge von Urteilsgenauigkeit und Beobachtungskompetenz wurde in dieser Form noch nicht untersucht.

Im folgenden Kapitel 5 werden diese Aspekte mit theoretischen Modellen zur Urteils-genauigkeit in einem Prozessmodell zusammengeführt. Dieses Modell ist Grundlage für die Forschungsfragen, die sich in Kapitel 6 anschließen.

5. Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz

Urteilsgenauigkeit und Beobachtungskompetenz als zwei Facetten diagnostischer Kompetenz gehen auf unterschiedliche Forschungstraditionen zurück. Während die Urteilsgenauigkeit in der pädagogisch-psychologischen Forschung im Rahmen einer formalwissenschaftlichen Erfassung eines psychometrischen Konstrukts untersucht wird, ist die Facette Beobachtungskompetenz vor dem Hintergrund praktischer Herausforderungen in Lehr- und Lernsituationen und Ansätzen zur Erfassung von formativem Assessment entstanden. Den Zusammenhang beider Facetten verdeutlicht das Prozessmodell, welches im Folgenden dargestellt wird.

Es wird davon ausgegangen, dass Urteilsgenauigkeit und Beobachtungskompetenz zwei abhängige Facetten der übergeordneten diagnostischen Kompetenz sind. Die Verbindung der beiden Facetten ist aufgrund der Unterschiedlichkeit ihrer Operationalisierungen und ihrer Herkunft nicht spannungsfrei und bildet das aus der psychologischen Diagnostik bekannte Bandbreiten-Genauigkeits-Dilemma (*bandwidth-fidelity trade-off*) (z.B. Cronbach & Gleser, 1965; Hogan & Brent, 1996) ab. Dieses wird in der pädagogischen Diagnostik unter anderem als Häufigkeits-Genauigkeits-Dilemma bezeichnet und kommt z.B. im Zusammenhang mit Kurzzeitbeobachtungen oder Kurzzeitdiagnosen im Vergleich zu aufwendigeren diagnostischen Verfahren zum Tragen (Ingenkamp & Lissmann, 2008, S. 33).

Nach Ingenkamp und Lissmann (ebd., 2008, S. 195) befindet sich die pädagogische Diagnostik in einem diagnostischen Dilemma zwischen Entwicklungsaufwand und Güte der Information; man könne keine umfassende Information von hoher Güte bei minimalem Aufwand erhalten.

Das Bandbreiten-Genauigkeits-Dilemma weist darauf hin, dass vermutlich zwischen der präzisen, reliablen und validen Diagnostik in eng definierten Situationen und der komplexen Diagnostik in nicht standardisierten Alltagssituationen unterschieden werden muss. So findet sich das Bandbreiten-Genauigkeits-Dilemma in der Praxis des Kindergartens z.B. in der Entscheidung einer Erzieherin bezüglich ihres diagnostischen Vorgehens, ob wenige Urteile sehr präzise oder ob viele Urteile eher überblicksartig erfasst werden sollen. Es ist auch möglich, dass Erzieherinnen aufgrund ihrer diagnostischen

Kompetenz nur in der Lage sind, entweder einen präziseren Fokus auf bestimmte bereichsspezifische Fähigkeiten des Kindes zu legen oder nur überblicksartig zu urteilen. Denkbar erscheint auch, dass eine Vielzahl von überblicksartigen Urteilen in konkreten Situationen ein rückblickendes Urteil über einen gewissen Zeitraum zulässt, das aufgrund der Summe von Urteilen entsprechend genau ist. Diese Annahme über den Zusammenhang von Beobachtungskompetenz (1) und Urteilsgenauigkeit (2) liegt dem folgenden Modell (Abb. 13) zugrunde.

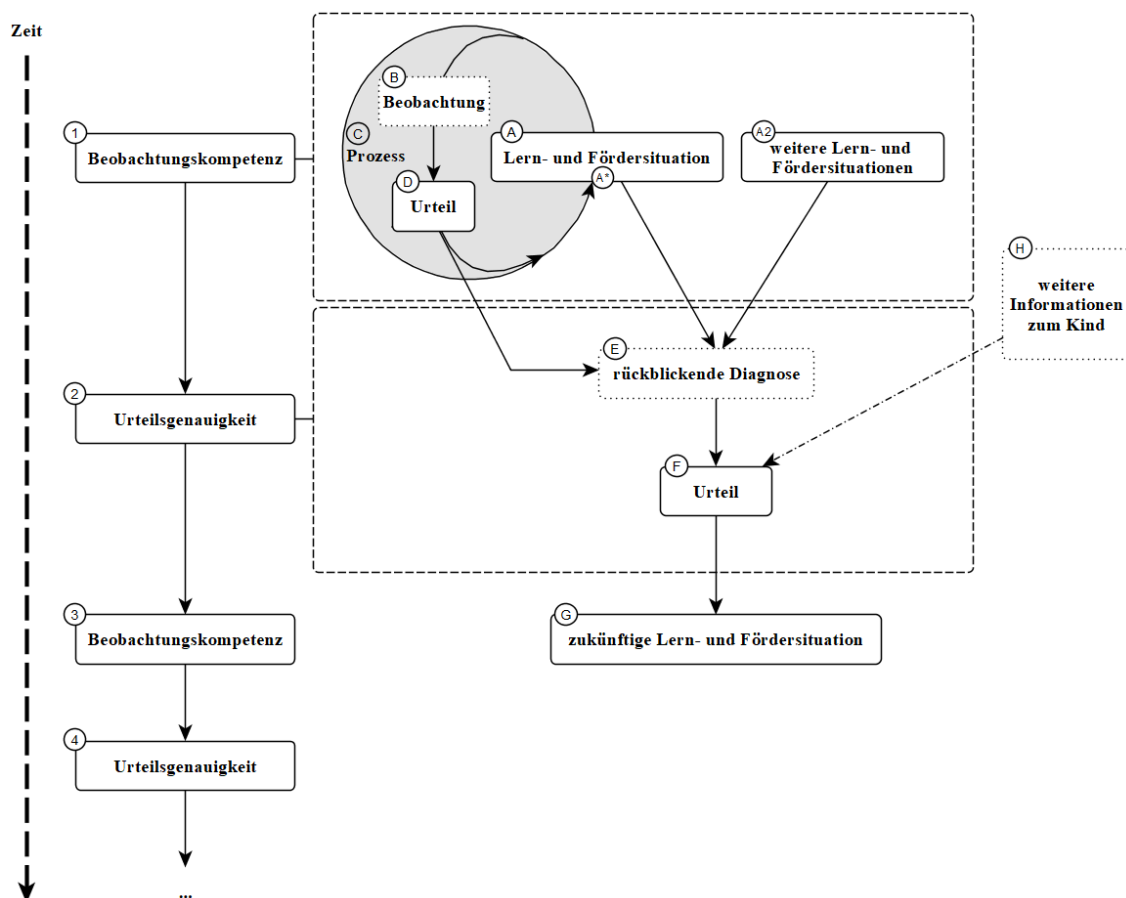


Abbildung 13: Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz

Beide Facetten diagnostischer Kompetenz werden als voneinander abhängig betrachtet. Es wird davon ausgegangen, dass einer Beurteilung eine oder mehrere entsprechend kompetente Einzelbeobachtungen vorausgegangen sein müssen. Folglich wird eine inhaltliche Gemeinsamkeit zwischen den beiden Facetten angenommen, obwohl diese für unterschiedliche Formen von Diagnostik stehen.

Im Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz (Abb. 13) wird der zeitliche Zusammenhang der beiden Facetten erkennbar;

letztlich handelt es sich um eine Wechselwirkung zwischen Beobachtungskompetenz und Urteilsgenauigkeit. Wird die einem Urteil vorgeschaltete Beobachtungskompetenz als ‚Startpunkt‘ angesehen, verläuft der Prozess wie folgt.

Beobachtungskompetenz (1): Die Beobachtungen (B) in konkreten Lern- und Lernsituationen (A, A2, ...) führen im (Lehr-Lern-)prozess (C, grau hinterlegt) zu einem Beobachterurteil (D), das einen Einfluss auf die Fortführung oder Gestaltung der jeweiligen Lernsituation (A*, A2*, ...) haben kann.

Urteilsgenauigkeit (2): Eine rückblickende Diagnose (E) einer Vielzahl von Beobachtungen unterschiedlicher Situationen (A*, A 2, ...) führt zu einem Urteil (F) über den aktuellen Status des Kindes und liefert Hinweise für die Gestaltung zukünftiger Lern- und Fördersituationen (G). Je genauer diese Urteile ausfallen, desto genauer lassen sich zukünftige Beobachtungen (3) durchführen. In dieses Statusurteil (F) fließen weitere Informationen über das Kind (H) ein, die im zeitlichen Verlauf unabhängig von Beobachtungen gewonnen werden.

Elemente dieses Modells finden sich in einer Vielzahl von Prozessmodellen (z.B. Fröhlich-Gildhoff et al., 2011, S. 17), jedoch wird in diesen die Beobachtungskompetenz als solche nicht näher ausgeführt. Zumeist wird Beobachtung als Tätigkeit oder als Informationsaufnahme gesehen, die als Basis der Urteilsgenauigkeit dient. Welche Qualität diese Beobachtung im Prozess hat und wie groß die Ausprägung der Beobachtungskompetenz ist, spielt in diesen Modellen eine untergeordnete Rolle.

Im Prozessmodell des Netzwerks zur diagnostischen Kompetenz von Lehrkräften (Herppich et al., 2017) lassen sich zwar beide Facetten verorten aber nur einzeln und damit unabhängig voneinander betrachten. Der Zusammenhang und der zeitliche Verlauf spielen in diesem Modell keine Rolle. Die einzelnen Prozessequenzen enden dort jeweils mit Handlungsentscheidungen, welche ausdrücklich nicht als Teil des diagnostischen Prozesses verstanden werden (ebd., S. 81). Die Handlungsentscheidung ist in diesem Modell dem Diagnoseprozess nachgeschaltet und nicht mehr Teil des Modells (Abb. 13). Dies ist ein großer Unterschied zum vorliegenden Modell, das die unmittelbare Einflussnahme auf die Fördersituation als Teil der Beobachtungskompetenz einschließt. Das bedeutet, dass von der Gestaltung der Fördersituation auf die Beobachtungskompetenz geschlossen werden kann. So braucht eine Erzieherin z.B. bei der Eingewöhnung eines Kindes, das sie noch nie zuvor gesehen hat, im ersten Schritt Beobachtungskompetenz (Abb. 13 (Start-)Punkt 1). In einer z.B. Spielsituation lernt sie

das Kind kennen, indem sie es beobachtet und immer wieder Impulse setzt oder Materialien anbietet, um ihre Beobachtung zu präzisieren, um detailliertere Informationen zu erhalten – letztlich, um das Kind *besser* kennenzulernen. Daher schließt der Diagnosekreislauf der Beobachtungskompetenz in Abbildung 13 die Einflussnahme auf die Fördersituation ein. Im Gegensatz zum Arbeitsmodell des Netzwerks zur diagnostischen Kompetenz von Lehrkräften (Herppich et al., 2017), das auf alle Diagnostikformen passt und ein allgemeines Modell von Diagnostik darstellt, legt das vorliegende Prozessmodell den Fokus zunächst auf die Mikroebene der Fördersituation. Das NeDiKo-Modell beschreibt das Zustandekommen einer Diagnose im Allgemeinen und zeigt, dass die pädagogische Entscheidung nicht mehr zum diagnostischen Prozess gehört. Weil das hier vorgeschlagene pädagogische Modell (Abb. 13) diagnostisch relevante Handlungsentscheidungen in den diagnostischen Gesamtprozess einschließt, wird es zum einen den Anforderungen von Schütze et al. (2018) gerecht, die betonen, dass formatives Assessment nicht auf ein Merkmal, wie z.B. Leistungsdiagnostik, reduziert werden darf und im Gesamtprozess gesehen werden muss. Zum anderen erfüllt das Modell als ein pädagogisches die Anforderungen Ingenkamps und Lissmanns (2008), welche pädagogische Diagnostik erst dann als solche anerkennen, wenn sie der Optimierung individueller Lernprozesse dient (ebd., S. 13), also den Prozess als immanenten Anteil pädagogischer Diagnostik definiert.

6. Forschungsfragen und Hypothesen

Die Forschungsfragen und Hypothesen beziehen sich auf Fragestellungen, welche die Ausprägung von Urteilsgenauigkeit und Beobachtungskompetenz, die Verbesserung der Facetten und Unterschiede in den Veränderungen zwischen verschiedenen Ansätzen zur bereichsspezifischen Fortbildung fokussieren. Die Fragen und Hypothesen werden getrennt für die beiden Facetten Urteilsgenauigkeit und Beobachtungskompetenz berichtet. Zentral ist die Frage nach dem Zusammenhang beider Facetten; in der folgenden Grafik (Abb. 14), die einen Überblick über die Forschungsfragen gibt, ist sie deshalb mittig platziert.

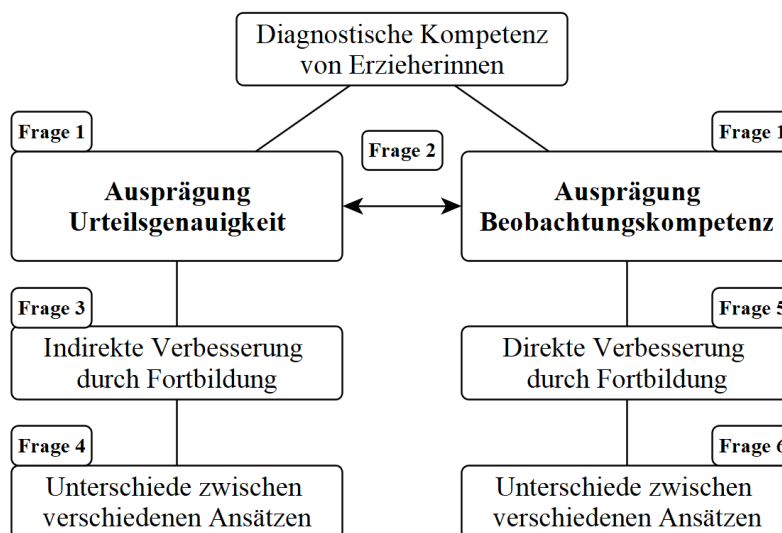


Abbildung 14: Forschungsfragen im Überblick

Zunächst stellt sich die Frage nach den Ausprägungen von Urteilsgenauigkeit und Beobachtungskompetenz (Frage 1).

Forschungsfrage 1: Wie ist die bereichsspezifische diagnostische Kompetenz von Erzieherinnen hinsichtlich der Facetten Urteilsgenauigkeit und Beobachtungskompetenz ausgeprägt?

Die zentrale Fragestellung der vorliegenden Arbeit (Frage 2) bezieht sich auf das vorgeschlagene Prozessmodell zum Zusammenhang von Urteilsgenauigkeit Beobachtungskompetenz und (Kap. 5).

Forschungsfrage 2: Gibt es einen Zusammenhang zwischen Urteilsgenauigkeit und Beobachtungskompetenz?

Hypothese: Es gibt einen Zusammenhang zwischen den Facetten Urteilsgenauigkeit und Beobachtungskompetenz.

Es wird angenommen, dass die Facetten Urteilsgenauigkeit und Beobachtungskompetenz Teile des Konstrukts diagnostische Kompetenz sind, wie es das erweiterte Linsenmodell (Kap. 4.6.1, Abb. 10) und das Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz (Kap. 5, Abb. 13) zeigen. Einem akkuraten Urteil gehen eine Vielzahl von Situationen voraus, die situierte Beobachtungen ermöglichen. In Abhängigkeit davon, wieviel eine Lehrperson in diesen Situationen beobachten kann, gelingen ihr auf Basis dieser Informationen mehr oder weniger genaue Urteile. Der Zusammenhang zeigt sich im Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz und in der Annahme, dass häufige gezielte Beobachtungen zu einer genauen Beurteilung führen. Denkbar wäre jedoch auch, dass Lehrende durch die Anzahl verschiedener Beobachtungen eher oberflächlich und auf der Beobachtungsebene bleiben, weshalb ihnen die Zusammenführung der Beobachtungen in eine summative Diagnose weniger gut gelingt.

Zudem wird die Veränderung von Urteilsgenauigkeit und Beobachtungskompetenz in den Blick genommen. Es stellt sich die Frage, wie durch Fortbildungsmaßnahmen die Beobachtungskompetenz (Frage 5) und die Urteilsgenauigkeit (Frage 3) verändert werden kann und ob sich diese Veränderungen in Abhängigkeit vom Ansatz der Fortbildung unterscheiden (Beobachtungskompetenz: Frage 6, Urteilsgenauigkeit: Frage 4).

Forschungsfrage 3: Unterscheidet sich die Urteilsgenauigkeit von Erzieherinnen mit und ohne Fortbildung zu mathematischer Frühförderung?

Hypothese: Die Urteilsgenauigkeit von Erzieherinnen mit Fortbildung zu mathematischer Frühförderung ist größer.

Es wird eine Zunahme sowohl für einzelne Komponenten der Urteilsgenauigkeit als auch für die globale Urteilsgenauigkeit angenommen, weil Erzieherinnen im Anschluss an eine Fortbildung zur mathematischen Frühförderung (ohne explizites Training der diagnostischen Kompetenz) in vielfältigen Situationen Wissen erwerben, indem sie ihre Bezugskinder bereichsspezifisch fördern. Diese Situationen werden erst nach der Fortbildung hergestellt bzw. erst im Anschluss an die Fortbildung als mathematische Fördersituation von den Erzieherinnen explizit wahrgenommen. Dementsprechend rücken auch die Fähigkeiten der Kinder nach der Fortbildung in den mathematischen Fokus der Erzieherinnen. Sie können somit das Niveau ihrer Gruppe zunehmend besser einschät-

zen (Niveauelemente), ihnen gelingt die Wahrnehmung von Unterschieden zwischen Kindern ihrer Gruppe zunehmend besser (Personenelemente) und sie können sowohl spezifische Aufgabenschwierigkeiten (Merkmalelemente) als auch Unterschiede zwischen den Kindern im Hinblick auf einzelne Aufgaben (Wechselwirkungselemente) zunehmend besser einschätzen.

Forschungsfrage 4: Unterscheidet sich die Veränderung der Urteilsgenauigkeit von Erzieherinnen durch Fortbildungen in verschiedenen mathematischen Förderansätzen?

Hypothese: Die Urteilsgenauigkeit verändert sich durch Fortbildungen zu verschiedenen Ansätzen unterschiedlich.

Es ist anzunehmen, dass im Rahmen eines Förderansatzes, der individualisiert und lernwegsorientiert ausgerichtet ist, Erzieherinnen mehr und differenziertere Urteile treffen müssen. Sie müssen konzeptbedingt adaptiv arbeiten und, je nach Bedürfnissen und Wissenstand der Kinder, in unterschiedlicher Weise im Rahmen offener Lernangebote reagieren. Eine lehrgangsorientierte Förderung setzt dagegen keine Individualdiagnostik voraus. Hier werden den Kindern Angebote mit vorgegebenem systematischem Aufbau unabhängig von unterschiedlichen Leistungsniveaus präsentiert. Dies bedeutet, dass Erzieherinnen, die in einem lernwegsorientierten Förderkonzept fortgebildet werden, stärkere Zunahmen in den einzelnen Urteilsgenauigkeitselementen (Niveau-, Personen-, Merkmale-, Wechselwirkungselemente, globale Urteilsgenauigkeit) zeigen als Erzieherinnen, die eine lehrgangsbasierte Fortbildung erhalten haben. Die Einflüsse auf die Urteilsgenauigkeit können sich jedoch erst nach einer gewissen Zeit zeigen, weil eine Erzieherin ihre Bezugskinder durch eine Fortbildung nicht besser kennenlernt, sondern diese erst nach der Fortbildung gezielter und besser beobachten kann, was in der Folge zu einer höheren Urteilsgenauigkeit führt. Effekte auf die Urteilsgenauigkeit sind daher nur indirekt messbar – nach einer gewissen Kontaktzeit zu eigenen Bezugskindern.

Forschungsfrage 5: Unterscheidet sich die Beobachtungskompetenz von Erzieherinnen mit und ohne Fortbildung zur mathematischen Frühförderung?

Hypothese: Die Beobachtungskompetenz von Erzieherinnen wird durch eine Fortbildung zur mathematischen Frühförderung verbessert.

Es wird eine Zunahme der Beobachtungskompetenz angenommen, weil eine Fortbildung zur mathematischen Frühförderung als Schlüssel für eine bewusste und damit fo-

kussiertere Wahrnehmung früher mathematischer Kompetenzen angenommen wird; Erzieherinnen können nur das wahrnehmen, was sie auch kennen.

Forschungsfrage 6: Unterscheidet sich die Veränderung der Beobachtungskompetenz von Erzieherinnen durch Fortbildungen in verschiedenen mathematischen Förderansätzen?

Hypothese: Die Beobachtungskompetenz verändert sich durch Fortbildungen in verschiedenen Ansätzen unterschiedlich.

Analog zur Urteilsgenauigkeit in Frage 4 wird auch für die Facette Beobachtungskompetenz angenommen, dass im Rahmen einer Fortbildung in mathematischer Frühförderung die Beobachtungskompetenz dann stärker gefördert wird, wenn die Fortbildung auf individuelle Lernwege ausgerichtet ist, weil Erzieherinnen in diesem Fall für die Bedeutung der Individualdiagnostik stärker sensibilisiert werden als im Rahmen einer lehrgangsorientierten Förderung, die keine Individualdiagnostik voraussetzt.

7. Methodik und Durchführung

In Kapitel 7 werden zunächst der Projektrahmen (7.1) und das Design (7.2) der vorliegenden Studie vorgestellt. Kapitel 7.3 thematisiert die Stichprobengewinnung. In Kapitel 7.4 erfolgt die Beschreibung einzelner Variablen und Erhebungsverfahren. Kapitel 7.5 informiert über den Verlauf der Studie, bevor Kapitel 7.6 abschließend die Methodenauswahl pro Forschungsfrage begründet und die jeweils dazugehörigen poweranalytischen Überlegungen präsentiert.

7.1 Projektrahmen

Die vorliegende Arbeit ist im DFG-Graduiertenkolleg im Projekt FaBi (Förderung anschlussfähiger Bildungsprozesse) entstanden (vgl. Donie et al., 2013). Dieses untersucht, inwieweit sich eine bereichsspezifische Förderung im Verlauf des letzten Kindergartenjahres auf mathematische und schriftsprachliche Leistung, Lernfreude und Selbstkonzept von Vorschulkindern auswirkt. Erzieherinnen ohne spezielle Vorkenntnisse erhielten eine Fortbildung zur Förderung mathematischer und schriftsprachlicher Kompetenzen mit zwei unterschiedlichen Förderansätzen. Im Anschluss daran wurden Vorschulkinder in den Bereichen Schriftsprache und Mathematik unter drei Bedingungen gefördert: lehrgangsbasiert, lernwegsbasiert oder nicht speziell bereichsspezifisch.

7.2 Design

7.2.1 Aufbau der Studie

Zur Beantwortung der Forschungsfragen dient ein quasi-experimentelles Drei-Gruppen-Prä-Posttest-Design. Die Randomisierung erfolgte auf Einrichtungsebene, indem die Erzieherinnen den Treatments per Zufall zugeteilt wurden.

Wie in Abbildung 15 dargestellt, wurden die beiden Facetten diagnostischer Kompetenz jeweils zu zwei Messzeitpunkten erhoben: Die Beobachtungskompetenz von Erzieherinnen wurde unmittelbar vor (t_1) und im Abstand weniger Tage nach (t_2) einer zweitägigen Fortbildung zur mathematischen Förderung am Anfang des Kindergartenjahres

2010/11 gemessen. Die Urteilsgenauigkeit wurde zu t_1 und nach neunmonatiger Durchführung von Fördermaßnahmen am Ende des Kindergartenjahres zu t_3 nochmals erhoben. Sowohl die Fortbildungen der Erzieherinnen als auch die darauf basierende Förderung der Vorschulkinder wurden unter zwei Bedingungen (Förderansätze) durchgeführt. Die Kontrollgruppe zur Beobachtungskompetenz wurde nur zu t_2 erhoben, weil von einer Veränderung der Beobachtungskompetenz nach zwei Tagen (Zeitraum der Fortbildung) ohne Intervention nicht auszugehen war. Die Kontrollgruppe zur Urteilsgenauigkeit wurde sowohl zu t_1 als auch zu t_3 untersucht. Im Hinblick auf die Bereichsspezifität erfolgt eine Fokussierung auf den mathematischen Bereich.

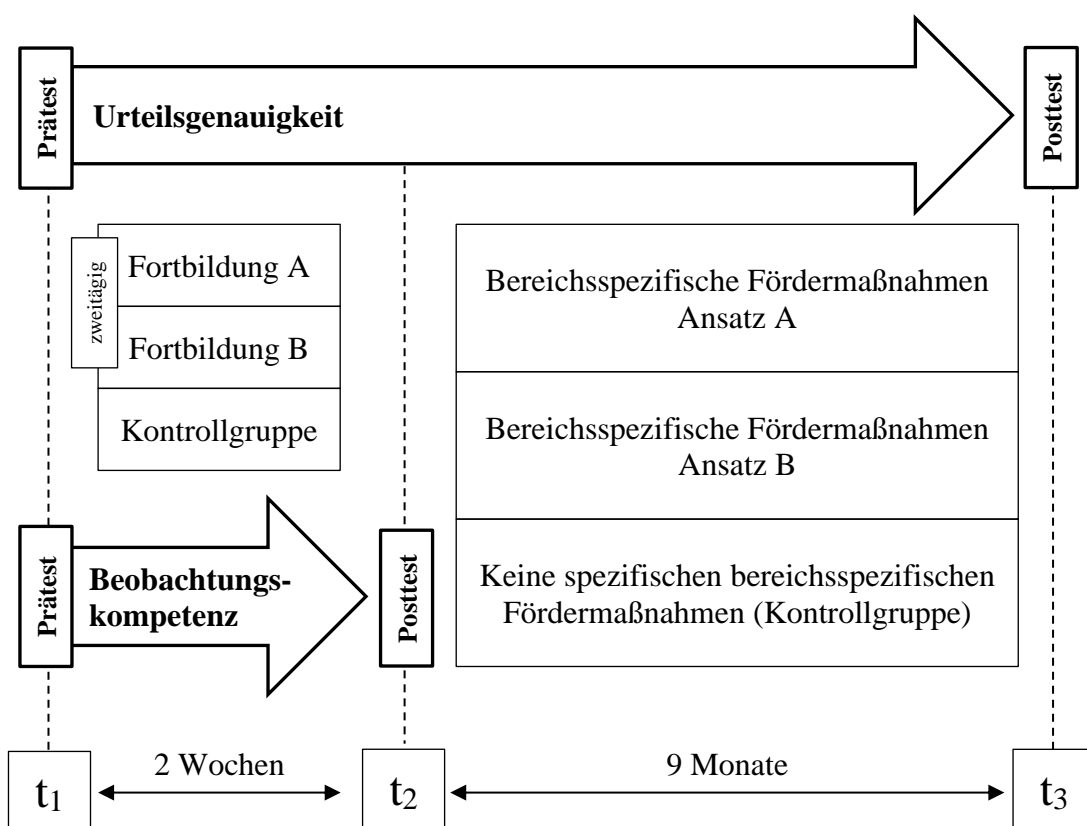


Abbildung 15: Evaluationsdesign

7.2.2 Treatment – bereichsspezifischer Fortbildungsansatz

Fortbildungen zur mathematischen Frühförderung, von denen angenommen wird, dass sie sich auf die diagnostische Kompetenz auswirken, können auf der Basis verschiedener Ansätze durchgeführt werden. Unterschieden wird zwischen Fortbildungen nach einem lernwegorientierten und nach einem lehrgangorientierten Ansatz. Gleichgehalten werden die Dauer der Fortbildung, der Fortbildungsort und die Durchführung durch externe Fortbildnerinnen, die als Expertinnen für den jeweiligen Ansatz bezeichnet

werden können. Diese führten die Förderung ökologisch valide, wie eine reguläre Fortbildung durch, in der es keinen expliziten Fokus auf diagnostische Kompetenz gab.

Treatment 1: Fortbildung zur lehrgangsbasierten Förderung

Kennzeichnend für lehrgangsbasierte Programme ist ihre Konzeption als aufeinander aufbauende und nach einem vorgegebenen Schema ablaufende Lerneinheiten, in deren Mittelpunkt jeweils ein Inhalt(sbereich) steht. Hierzu gibt es detaillierte Hinweise in Form von Stundenbildern, wie diese in Fördereinheiten durchgeführt und wie Medien und Materialien eingesetzt werden können.

Als prototypisches Beispiel für einen pädagogischen Ansatz einer lehrgangsbasierten Förderung wurde das Programm *Komm mit ins Zahlenland* (Friedrich & Galgóczy, 2008) ausgewählt. Das Konzept basiert auf einer Märchenwelt – dem Zahlenland, in dem episodisch zu jeder Zahl von eins bis zehn eine Geschichte (entsprechend zehn Stundenbildern) erzählt wird, die jeweils Ausgangspunkt für Spiele, Aufgaben und Lieder ist. In diesen Geschichten erleben Zahlen mit menschlichen Eigenschaften verschiedene Abenteuer. Feste Bestandteile wie Zahlenweg, Zahlenhäuser und Zahlengärten führen zu ritualisierten Abläufen. Figuren, wie Zahlenkobold und Zahlenfee sind wiederkehrende Elemente mit bestimmten Aufgaben. Der Ansatz versteht sich als ganzheitlich, weil nach Ansicht der Autoren neben den Grundzahlen auch Konzentration und Ausdauer, Gedächtnisleistung, Kreativität, Grundlagen des mathematisch-logischen Denkens und Musikalität gefördert werden (ebd., S. 9).

Ausgewählt wurde dieser Ansatz, nicht nur wegen seines expliziten lehrgangsorientierten Aufbaus, sondern auch, weil er empirisch untersucht wurde. Friedrich und Munz (2004, 2006) evaluierten den Zahlenland-Ansatz in zwei Phasen. In der ersten Phase wurden Kinder mit sprachlichem Entwicklungsrückstand in altershomogenen, in Phase II in altersgemischten Gruppen gefördert; nach elf Wochen mit zehn Fördereinheiten konnten für die mit *Zahlenland* geförderten Kinder im Vergleich zu den nicht geförderten signifikant bessere Effekte nachgewiesen werden. (Friedrich & Munz, 2006). Bei der Bewertung der Ergebnisse ist jedoch zu berücksichtigen, dass die mathematischen Kompetenzen lediglich mit einem informellen Test (Kombination von Aufgaben aus dem *Kieler Einschulungsverfahren* von Fröse, Mölders & Wallrodt (1986) und den *Diagnostischen Einschätzskalen* von Barth (1998) erhoben wurden.

Pauen und Pahnke (2008) verglichen in einer Interventionsstudie im Zweigruppensign ($N = 199$) die Mathematikprogramme *Zahlenland* (Friedrich & Galgóczy, 2008)

und *Mathe 2000* (Wittmann, 2009; Wittmann & Müller, 2009). Nach zehn Wochen Intervention zeigten sich signifikant bessere Effekte in unterschiedlichen Dimensionen mathematischen Denkens, die in beiden Gruppen vergleichbar ausfielen. Die Erkenntnisse aus dieser Studie sind begrenzt, da zum einen eine Kontrollgruppe fehlte und zum anderen die Erzieherinnen beide Programme kennenlernten und anschließend das Treatment frei wählten. Bei der Durchführung kam es nach Angabe von Pauen und Pahnke (ebd., S. 205) zu einer Kombination von Elementen aus beiden Ansätzen, weshalb das Fehlen von Unterschieden nicht verwundert. Hinzu kommt, dass es keine Angaben zur Güte des M-Einstein-Tests gibt, mit dem hier die mathematischen Kompetenzen erhoben wurden. In einer weiteren vergleichbar angelegten Studie der beiden Autorinnen mit Kontrollgruppe (Pauen & Pahnke, 2009) wurden Effekte einer gezielten Förderung für fast alle Grundprinzipien des Zahlenverständnisses gefunden, jedoch nicht für anspruchsvollere Aufgaben.

Im Rahmenprojekt FaBi (Kap. 7.1) zeigte sich eine signifikante Zunahme der numerisch-mathematischen Fähigkeiten von Kindern, die am Zahlenlandansatz teilgenommen hatten, im Vergleich zu einer Kontrollgruppe (Donie et al., 2013, S. 316–317). Ähnliche Ergebnisse zeigte auch eine lernwegsbasierte Förderung, weshalb der Schluss gezogen wurde, im Hinblick auf den Zuwachs der numerisch-mathematischen Fähigkeiten der Kinder sei nur die Förderzeit und nicht die Art der Förderung ausschlaggebend (ebd., S. 321).

Das lehrgangsorientierte Programm *Zahlenland*, welches auch aufgrund der hohen Akzeptanz unter Praktiker(inne)n und wegen der weiten Verbreitung ausgewählt wurde, wird von Fachdidaktiker(inne)n kritisch gesehen (Gasteiger, 2010), weil es vermeintlich verschult sei und Zahlen personifiziert würden. Es wird argumentiert, die Motivation über unnötige ‚Verpackungen‘ sei einer echten Beziehung zur Mathematik abträglich (Wittmann & Deutscher, 2013, S. 214).

Herauszustellen ist, dass im Rahmen der Fortbildung eine Vielzahl von Handlungsabläufen, Medien und Materialien zur mathematischen Frühförderung vorgestellt wurden, die Diagnostik von Lernvoraussetzungen jedoch keine explizite Rolle spielte und diagnostische Kompetenz entsprechend nicht explizit trainiert wurde.

Treatment 2: Fortbildung zur lernwegsbasierten Förderung

Für einen lernwegsorientierten Ansatz ist zum einen kennzeichnend, dass die bereichsspezifische Förderung entwicklungsbasiert in den Alltag integriert ist und zum anderen,

dass es gezielte Angebote sowie Spiel- und Lernumgebungen gibt, bei denen die Erzieherin eine aktive Rolle einnimmt (Kluczniok, Roßbach & Große, 2010). Eingesetzt werden adaptive Lernumgebungen, Lernspiele und themenbezogene Rollenspiele (z.B. Restaurant, Post). Die Auswahl der Angebote richtet sich nach dem Entwicklungsstand der Kinder.

Als lernwegsbasierter Ansatz wurde der im KiDZ-Projekt (Kindergarten der Zukunft, Stiftung Bildungspakt Bayern, 2007) verfolgte handlungsorientierte Ansatz mathematischer Förderung ausgewählt, der nicht nur auf die Kulturtechnik des Rechnens abzielt, sondern vielmehr Mathematik als Kulturgut versteht, welches den Kindern die Welt zu strukturieren und zu begreifen hilft (Steinweg, 2007, S. 139).

Herauszustellen ist die Flexibilität des Konzepts, bei dem Erfahrungsbereiche, wie z.B. Zahlen und Strukturen, Raum und Form oder Zeit und Maße, individuell durch Lernfortschritte makroskopisch miteinander und zu anderen Fächern verknüpft werden, wobei innerhalb eines Bereichs nie nach festen Bauplänen, sondern aktiv vom lernenden Kind Vernetzungen gebildet werden (Steinweg, 2007, S. 141–143).

Die Erzieherinnen erhielten im Rahmen der Fortbildung somit keine festen Handlungsabläufe oder inhaltliche Vorgaben, sondern zu jedem Erfahrungsbereich eine Vielzahl von Anknüpfungspunkten, Aktivitätsvorschlägen und Organisationsideen (z.B. ebd., S. 148–151). Die diagnostische Kompetenz wurde im Rahmen dieser Fortbildung ebenfalls nicht explizit behandelt.

Ausgewählt wurde dieses Konzept wiederum, weil es empirisch untersucht wurde; die mathematischen Kompetenzen von Kindern stiegen in der KiDZ-Gruppe signifikant stärker an als bei Kindern einer Vergleichsgruppe (Roßbach, Sechtig & Freund, 2010, S. 24–26). Auch im Rahmenprojekt FaBi (Kap. 7.1) zeigt sich im Gegensatz zu einer Kontrollgruppe eine signifikante Zunahme der numerisch-mathematischen Fähigkeiten bei den Kindern, die mit dem KiDZ-Ansatz gefördert wurden (Donie et al., 2013, S. 316–317).

Herauszustellen ist jedoch, dass im Rahmen der Fortbildung zur lernwegsbasierten Förderung Diagnostik implizit eine Rolle spielt, weil Inhalte entsprechend der Lernvoraussetzungen angeboten werden sollen. Es wurden aber lediglich Beobachtungsbögen als Möglichkeit zur Dokumentation vorgestellt, während ein explizites Training der diagnostischen Kompetenz ausblieb.

7.3 Stichprobe

Die einzelnen Fragestellungen wurden mit unterschiedlichen Stichproben untersucht. Abbildung 16 zeigt die Aufteilung der Stichproben.

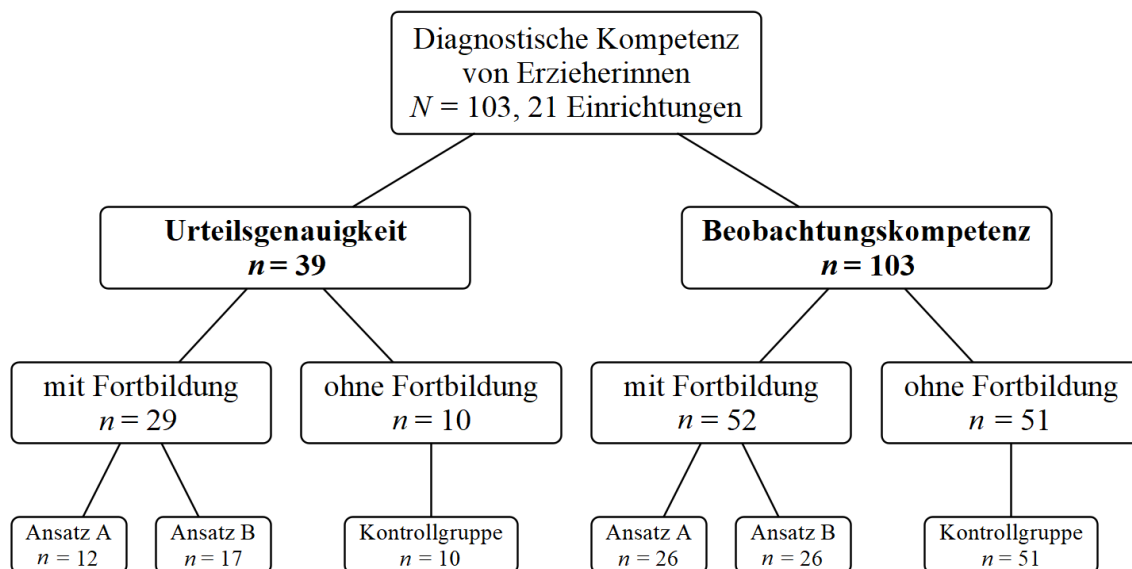


Abbildung 16: Stichprobenaufteilung

Die Gesamtstichprobe umfasst $N = 103$ Erzieherinnen aus 21 Kindergärten in Rheinland-Pfalz. Sowohl die Urteilsgenauigkeit als auch die Beobachtungskompetenz wurde von allen Teilnehmerinnen erhoben.

Zur Verbesserung der Qualität der diagnostischen Kennwerte für die Urteilsgenauigkeit wurde eine Reduktion des Stichprobenumfangs in Kauf genommen. Weil zur Erfassung der Urteilsgenauigkeit Erzieherinnen mehrere Bezugskinder einschätzen sollten, wurden aus der Gesamtstichprobe die Erzieherinnen ausgewählt, die mindestens vier Bezugskinder eingeschätzt hatten und diese entsprechend gut kannten, was anhand der wöchentlichen Kontaktzeit überprüft wurde. Daher besteht die Teilstichprobe Urteilsgenauigkeit abschließend aus 39 Erzieherinnen, die mindestens vier Bezugskinder eingeschätzt und mit diesen eine Kontaktzeit von durchschnittlich 4,9 Stunden zur reichsspezifischen Förderung pro Woche verbracht hatten. Insgesamt schätzten 20 Erzieherinnen jeweils vier und 19 Erzieherinnen jeweils fünf Bezugskinder ein. Die übrigen 64 Erzieherinnen konnten nur drei oder weniger Vorschulkinder einschätzen oder verbrachten mit den Bezugskindern nur sehr wenig Zeit, was z.B. der offenen Struktur der Einrichtungen geschuldet war. Aufgrund dieser pädagogisch-didaktischen Struktur

in den Einrichtungen (z.B. offene und geschlossene Gruppen) schätzten bei der Erfassung der Urteilsgenauigkeit mehrere Erzieherinnen teilweise dieselben Kinder ein, obgleich die Kontaktzeit zu ihnen sehr unterschiedlich sein konnte. Um die Genauigkeit der diagnostischen Kennwerte zu erhöhen, wurden Erzieherinnen nur die von ihnen eingeschätzten Kinder zugeordnet, mit denen sie die meiste Zeit pro Woche verbracht hatten.

Die Teilstichprobe Urteilsgenauigkeit mit Fortbildung besteht somit aus 29 Erzieherinnen, die eine Fortbildung besuchten. Zwölf von ihnen nahmen (41,38%) an einer lernwegsbasierten und 17 Personen (58,62%) an einer lehrgangsbasierten Fortbildung teil. Die übrigen zehn Erzieherinnen der Teilstichprobe haben keine Fortbildung besucht (Kontrollgruppe).

Die Beobachtungskompetenz wurde von der Gesamtstichprobe erhoben. Die Teilstichprobe mit Fortbildung besteht aus 52 Erzieherinnen; davon nahmen jeweils 26 (50%) Personen an einer lernwegsbasierten und einer lehrgangsbasierten Fortbildung teil. Die übrigen 51 Erzieherinnen der Gesamtstichprobe haben keine Fortbildung besucht (Kontrollgruppe).

Die teilnehmenden Erzieherinnen ($N = 103$) aus 21 Kindergärten waren im Durchschnitt 42.88 Jahre alt und verfügten über eine durchschnittliche Berufserfahrung von 17.69 Praxisjahren in der Kindertagesstätte. In Tabelle 9 findet sich eine detailliertere Darstellung der beiden Kennwerte.

Tabelle 9: Berufserfahrung und Alter der Erzieherinnen zu t_1

Berufsjahre	Erzieherinnen Anzahl (Prozent)	Alter $M = 42.88$	Erzieherinnen Anzahl (Prozent)
0–5	16 (15,53)	20–25	9 (8,74)
> 5–10	11 (10,68)	> 25–30	6 (5,83)
> 10–15	17 (16,50)	> 30–35	16 (15,53)
> 15–20	23 (22,33)	> 35–40	8 (7,77)
> 20–25	13 (12,62)	> 40–45	11 (10,68)
> 25–30	6 (5,83)	> 45–50	15 (14,56)
> 30–35	8 (7,77)	> 50–55	24 (23,30)
> 35–40	8 (7,77)	> 55–60	10 (9,71)
> 40	1 (0,97)	fehlend	4 (3,88)
gesamt	103 (100)	gesamt	103 (100)

Aus organisatorischen Gründen war es den eingebundenen Einrichtungen nicht möglich, alle ihre Erzieherinnen an den im Rahmen von FaBi stattgefundenen Fortbildungen teilnehmen zu lassen. Aus diesem Grund wurde die Projektteilnahme an die Bedingung geknüpft, dass die bereichsspezifische Förderung von mindestens einer Person mit Fort-

bildung betreut wird. Infolgedessen haben 52 Erzieherinnen (50,5%) an den bereichsspezifischen Fortbildungen teilgenommen, mit denen die Beobachtungskompetenz erfasst wurde, und 51 Personen (49,5%) nicht. Folglich wurden die Fördermaßnahmen nach verschiedenen Ansätzen zum Teil gemeinsam von Erzieherinnen mit und ohne Fortbildungsbeteiligung durchgeführt, wobei die Verantwortung stets bei der fortgebildeten Person lag. Tabelle 10 zeigt die Anzahl Erzieherinnen pro Einrichtung aufgeteilt nach Treatments. Zudem wird aufgezeigt, wie viele Erzieherinnen innerhalb der Treatments an den Fortbildungen teilgenommen haben und wie viele im Team gemeinsam mit den ausgebildeten Erzieherinnen die Lehrprozesse gestalten. Die Anzahl der Bezugsgruppen gibt an, in wie vielen Gruppen der jeweiligen Einrichtungen Vorschulkinder an der Studie teilnehmen.

Tabelle 10: Verteilung der Erzieherinnen und Bezugsgruppen auf Treatments und Kontrollgruppe

	Erzieherinnen		Bezugsgruppen
	mit Fortbildung	ohne Fortbildung	
Förderansatz A (lehrgnagsorientiert)	3	0	1
	2	0	2
	4	3	3
	3	3	3
	2	5	2
	8	0	2
	4	1	3
Summe:	26 (25,2%)	12 (11,7%)	16
Förderansatz B (lernwegsorientiert)	3	5	3
	6	0	3
	4	3	2
	4	0	3
	5	1	2
	2	2	1
	2	0	1
Summe:	26 (25,2%)	11 (10,7%)	15
Kontrollgruppe		3	1
		2	1
		4	1
		3	3
		5	1
		3	3
		8	4
Summe:		28 (27,2%)	14
Treatment-Summe:	$n = 52$ (50,5%)	$n = 51$ (49,5%)	
	$N = 103$ (100%)		

Tabelle 11 zeigt ergänzend die Anzahl der Einrichtungen aufgeteilt nach Bezugsgruppenanzahl und Treatments.

Tabelle 11: Verteilung der Bezugsgruppen nach Größe und Treatments

	Anzahl der Einrichtungen mit jeweiliger Gruppenanzahl von				Gruppen ($n = 45$)	Einrichtungen ($n = 21$)
	1	2	3	4		
Förderansatz A	1	3	3	/	16	7
Förderansatz B	2	2	3	/	15	7
Kontrollgruppe	4	/	2	1	14	7

7.4 Variablen und Erhebungsverfahren

7.4.1 Urteilsgenauigkeit von Erzieherinnen

Für die Erfassung der Urteilsgenauigkeit wurde das Instrument BEURteilEr (*Befragung zur Erfassung der Urteilsgenauigkeit von Erzieherinnen*) (Donie, 2010a) entwickelt.

Mit diesem schätzen Erzieherinnen mathematische Aufgaben im Hinblick darauf ein, ob den Erzieherinnen bekannte Kinder (Bezugskinder aus der eigenen Einrichtung) imstande sind, bestimmte Aufgaben zu lösen. Die Auswertung erfolgt mittels der Komponentenanalyse von Cronbach (1955). Zur Testkonstruktion wurde das Verfahren *zahlenstark* von Moser und Berweger (2007a, 2007b) ausgewählt, weil diesem vier Aufgabenblöcke (Ordinalzahlaspekt, Kardinalzahlaspekt, Kenntnis von Zahlensymbolen und Rechenzahlaspekt) zugrunde liegen, welche das Konstrukt numerisch-mathematische Fähigkeiten von Kindern abbilden. Damit bietet *zahlenstark* ideale Voraussetzungen für die Anwendung der Komponentenanalyse, die nur dann erfolgen kann, wenn es sich bei den Merkmalen um eine homogene Menge von Merkmalen handelt, denen eine gemeinsame Dimension zugrunde liegt, welche die Bildung von Summenwerten erlaubt (Schrader, 1989, S. 76). Eine Ausführliche Darstellung des Instruments *zahlenstark* von Moser und Berweger (2007a, 2007b) folgt im nächsten Kapitel.

Das Instrument BEURteilEr wurde nach Erkenntnissen zu methodischen Anforderungen zur Erfassung der Urteilsgenauigkeit konstruiert (Hoge & Coladarci, 1989) und entspricht folgendem messmethodischen Anforderungsprofil (vgl. Kap. 3.3):

- Es handelt sich um direkte Urteile, denn den Erzieherinnen wurde pro Bereich von *zahlenstark* jedes zweite Item bzw. jede zweite Itemgruppe zur Beurteilung vorgelegt (vgl. Anhang: Tab. 29: Übersicht der Bereiche und Items aus *zahlenstark*). Dabei wurde pro Bereich ausgelost, ob mit dem ersten oder zweiten Item begonnen werden sollte. Die Erzieherinnen kreuzten die Aufgaben an, welche ein Zielkind ihrer Einschätzung nach zu lösen imstande war. Mit diesem Vorgehen wird gewährleistet, dass direkte Urteile Lehrender mit Kenntnis des zur Leistungsmessung eingesetzten Instruments gefällt werden, was letztlich zu genaueren Urteilen führt (vgl. Hoge & Coladarci, 1989; Südkamp et al., 2012, S. 3).
- Höchste Spezifität wird durch die Leistungseinschätzung mittels Benennung einzelner Items, die ein Bezugskind zu lösen imstande ist, gewährleistet.
- Peer-Unabhängigkeit besteht, weil die Berechnung der Urteilsgenauigkeit ohne die Notwendigkeit eines Vergleichs zwischen Bezugskindern erfolgt.
- Die Ermittlung auf Itemebene erfolgt, weil zur Berechnung der Urteilsgenauigkeit ausschließlich die von der Erzieherin eingeschätzten Items berücksichtigt wurden, die auch vom Zielkind beantwortet wurden. Um eine mögliche Überschätzung zu erfassen, wurden zudem pro Aufgabentyp bzw. Teilbereich (s.o.) drei Items, die vom Zielkind bei steigendem Schwierigkeitsgrad nicht mehr beantwortet wurden, zur Berechnung der Komponenten herangezogen.
- Die hierarchische Struktur wird berücksichtigt, indem die Erzieherinnen nicht zu Gruppen zusammengefasst werden, sondern die Auswertungen pro Erzieherin und den jeweiligen Bezugskindern, also gruppenweise erfolgen.
- Die Domänenspezifität ist durch die Beschränkung auf den mathematischen Bereich gegeben.

Objektivität

Das vorliegende Verfahren zeichnet sich durch einen hohen Grad an Objektivität (vgl. Bühner, 2011, S. 34–35) aus, da die Durchführung standardisiert anhand schriftlicher Instruktionen erfolgt und damit die Einflussnahme eines Testleiters minimiert ist. Auswertungsobjektivität ist durch das standardisierte Vorgehen bei der Auswertung (s.o.) und aufgrund des geschlossenen Frage- und Antwortformats gegeben. Durch dieses

Antwortformat ohne Interpretationsspielraum ist auch die Interpretationsobjektivität gegeben.

Reliabilität

Die Bestimmung der Reliabilität der Diagnosekennwerte des Indikators Urteilsgenauigkeit ist problematisch (vgl. Lintorf et al., 2011; vgl. Spinath, 2005), weil es sich bei den Diagnosekennwerten nicht um einfache Summenscores (oder Mittelwerte), sondern um komplizierte Indizes handelt und hierfür kein testtheoretisches mathematisches Modell existiert (Schrader, 1989, S. 90). Aus diesem Grund wurde in Anlehnung an Schraders Vorgehen (vgl. Gage & Exline, 1953; Gage, Leavitt & Stone, 1955) wie folgt eine Split-Half-Reliabilität berechnet.

Zunächst wurden zwei Gruppen gebildet, indem jedes erste und jedes zweite Item des Instruments BEUrteilEr jeweils einer Gruppe zugeordnet wurde. Anschließend wurden die Diagnosekennwerte für beide Gruppen gebildet und die Reliabilität mittels der Spearman-Brown-Formel berechnet³.

Die durchschnittliche Reliabilität der Urteilsgenauigkeitskomponenten liegt bei .94, wobei nahezu alle Komponenten eine sehr hohe Reliabilität ($> .9$) (Bortz & Döring, 2016, S. 443) aufweisen, wie in Tabelle 12 zusammengestellt. Lediglich die Reliabilität der Niveauelemente liegt bei .87 und erfüllt damit noch den Anspruch eines guten Tests gemäß Bortz und Döring (2009, S. 199).

Tabelle 12: Split-Half-Reliabilitäten einzelner Komponenten der Erfassung der Urteilsgenauigkeit zu t_1

Komponente zu t_1 ($N = 39$)	Split-Half-Reliabilität
Globale Urteilsgenauigkeit (Accuracy)	.97
Niveauelemente (Elevation)	.87
Personenkomponente (Differential Elevation)	.95
Merkmalskomponente (Stereotype Accuracy)	.96
Wechselwirkungskomponente (Differential Accuracy)	.95

Validität

Die Validität des Instruments BEUrteilEr wurde zum einen gesichert, indem Erzieherinneneinschätzungen und Leistungen der Kinder in parallelen Zeiträumen (t_1 , t_3) erhoben wurden, wie es Schrader (1989, S. 57) empfiehlt. Zudem wurde durch die Konstruktion des Instruments basierend auf den gleichen Items, die zur Testung der Kinder eingesetzt wurden, die Validität erhöht. Diese von Coladarci befürwortete Diagnostik

³ Die Berechnungen erfolgten mit SPSS© (Release 20.0.0).

item by item (Coladarci, 1986, S. 142) empfiehlt auch Schrader, um Unterschiede im Merkmalsverständnis zu umgehen (Schrader, 1989, S. 57). Auch Hoge und Butcher (1984) verweisen in diesem Kontext darauf, dass sich das Urteil auf exakt dasjenige Merkmal beziehen sollte, das als Urteilkriterium herangezogen wird. Damit erhöht sich die Validität entsprechend.

7.4.2 Numerisch-mathematische Fähigkeiten von Vorschulkindern

Die numerisch-mathematischen Fähigkeiten von Kindern werden erfasst, um die Urteilsgenauigkeit von Erzieherinnen zu bestimmen. Hierzu werden deren Einschätzungen mit Testwerten von Kindern verglichen. Wie bereits im vorausgegangenen Kapitel dargestellt, basiert das Instrument zur Erfassung der Urteilsgenauigkeit auf dem folgenden Verfahren.

Zur Erfassung der numerisch-mathematischen Fähigkeiten wurde der Subtest *zahlenstark* aus *wortgewandt & zahlenstark* von Moser und Berweger (2007a) eingesetzt. Dieser Test wurde für die Evaluation der Basisstufe in der Schweiz konzipiert und mehrfach erfolgreich eingesetzt, wie z.B. im Projekt PERLE (Corvacho del Toro & Greb, 2007, S. 314). Die Auswahl dieses Verfahrens erfolgte, weil es in der Kindertagesstätte und in der 1. Klasse eingesetzt werden kann, was für das Rahmenprojekt von Bedeutung war (vgl. Kap. 7.1).

Der Test umfasst 120 Einzelaufgaben, die sich auf die vier Bereiche Ordinal-, Kardinal- und Rechenzahlaspekt sowie Kenntnis von Zahlensymbolen verteilen (vgl. Tab. 13). Die Einbindung aller Aufgaben in eine kindgemäße Geschichte und damit Einbettung in Alltagssituationen gewährleistet eine möglichst präzise Abbildung der Fähigkeiten der getesteten Kinder.

Tabelle 13: Inhaltsbereiche von *zahlenstark* gemäß Moser und Berweger (2007a)

Aufgabenblock	Bereiche	Items
Ordinalzahlaspekt	a. Positionen bestimmen	8
	b. Rangreihen mit Bildern bilden	8
	c. Rangreihen mit Zahlen bilden	12
Kardinalzahlaspekt	Mengen bestimmen und vergleichen	24
Kenntnis von Zahlensymbolen	Zahlen benennen	16
Rechenzahlaspekt	a. Rechnen mit Bildern	8
	b. Rechnen mit Zahlen	44
		gesamt 120

Der Test wurde mit einer Eichstichprobe von 1016 Kindern in der Deutschschweiz normiert (Moser & Berweger, 2007b, S. 58). Moser und Bayer (2011, S. 154) betonen den hohen Grad an Standardisierung des Instruments und die damit verbundene Durchführungsobjektivität. Diese ist durch ein Testinstrument mit Aufgaben in Form von Illustrationen, durch eine CD mit den akustischen Aufgabenstellungen und durch eine Testanleitung gegeben, in der sowohl eine abzulesende Anleitung als auch die Angaben zur Durchführung festgehalten sind.

Der Test liefert aufgrund der Konstruktion mittels probabilistischer Testtheorie keine klassischen Testgütekriterien, sondern eine detaillierte Dokumentation der Testgüte in Form einer tabellarischen Aufstellung der Ergebnisse der Itemanalyse sowie der Kalibrierung der Aufgaben (vgl. Moser & Berweger, 2007b, S. 67–69). Schneider, Küspert und Krajewski (2016, S. 81) ziehen in einer Gesamteinschätzung das Fazit, dass das Verfahren im Hinblick auf seine Gütekriterien überzeugt.

Der Grad der Standardisierung wurde in der vorliegenden Studie nochmals erhöht, weil die allgemeinen Abbruchkriterien in der ursprünglichen Form als zu ungenau angesehen wurden. Diese besagen, dass zur Vermeidung einer Überforderung Aufgaben, die sich „als zu schwierig erweisen“ (Moser & Berweger, 2007b, S. 19), übersprungen werden sollen. Wenn ein Kind jeweils drei Wörter, Sätze oder Silben gar nicht oder falsch gelesen hat, werden die restlichen Aufgaben übersprungen (vgl. ebd., S. 20). Dieses Prinzip, das Moser und Berweger auch für einzelne Bereiche des Subtests numerisch-mathematischer Fähigkeiten anwendet (vgl. ebd., S. 25), wurde aus Gründen der Stan-

standardisierung im Rahmen der vorliegenden Studie auf alle Aufgaben übertragen, weil die oben benannte Formulierung „zu schwierig“ einen Ermessensspielraum bietet. Zudem fielen komplette Aufgabenblöcke dann weg, wenn es offensichtlich war, dass ein Kind die entsprechende Aufgabe nicht lösen kann (vgl. Moser & Berweger, 2007b, S. 20).

Auf Basis des Testinstruments *zahlenstark* wurde das Verfahren zur Urteilsgenauigkeitsmessung entwickelt. Diese Entwicklung wird im folgenden Kapitel dargestellt.

7.4.3 Beobachtungskompetenz von Erzieherinnen

Zur Erfassung der Beobachtungskompetenz wurde das selbst entwickelte, auf Video-vignetten basierende Instrument FiBB (*Filmbasierte Befragung zur Erfassung der Beobachtungskompetenz*) eingesetzt (Donie, 2010b). Hierbei handelt es sich um einen Fragebogen zu Filmbeispielen, welche Kinder bei mathematischen Tätigkeiten in alltäglichen

(Spiel-)Situationen in der Kindertagesstätte zeigen. Es handelt sich um folgende Vignetten, anhand derer folgende Beobachtungskriterien erfasst werden:

- Vignette 1 *Sortieren* zeigt ein Vorschulkind, das eine Stiftebox ausgeleert hat, um die Stifte zu sortieren. Dieses Filmbeispiel liefert Beobachtungskategorien zu Sortieren, Mengenunterscheidung und -erfassung, Zahlenkenntnis, Zählen, Zusammen-, Weiterzählen, Rechnen, Vergleichen, Zuordnen, Merken von Ergebnissen.
- Vignette 2 *Brettspiel* zeigt zwei Vorschulkinder, die gemeinsam am Tisch sitzen und ein Brettspiel spielen. Dieses Filmbeispiel liefert Beobachtungskategorien zu (Ab-)Zählen, Zuordnen, Zahlunterschiede erkennen, Erkennen falscher Plätze (Ordinalaspekt), Zusammenzählen mit und ohne Hilfsmittel, Addieren auf Mengengrundlage, Zuordnen, Rücken gewürfelter Zahlen, (Mengen-)Vergleiche, Mengenerlegung, (simultane) Mengenerfassung, Zahlwissen.
- Vignette 3 *Ziffer 6 & Zählen* zeigt drei Vorschulkinder, die gemeinsam am Tisch sitzen und Zahlen schreiben möchten. Das Filmbeispiel liefert Beobachtungskategorien zu Schreiben von Ziffern, (Er-)Kennen und Benennen von Zahlen, Raumlagekenntnisse (z.B. spiegelverkehrtes Zeigen), Vergleichen von Zahlen (z.B. Altersvergleich), Zählen ab 16 (bis 32), Zusammenzählen.

- Vignette 4 *Bauecke* zeigt insgesamt fünf Kinder, die sich in der Bauecke aufhalten. Das Filmbeispiel liefert Beobachtungskategorien zur Einschätzung von Größenverhältnissen, zu räumlichem Vorstellungsvermögen, Vergleichen (wenig-viele, hoch-tief), Erkennen von Größen, mathematische Überlegungen zu Statik, Abmessen, Sortieren, Bauen, Zuordnen, Formen legen.

Die folgende Abbildung zeigt Vignette 2 mit offener Fragestellung.

Im dritten Filmbeispiel sehen Sie zwei Vorschulkinder, die gemeinsam am Tisch sitzen und ein Brettspiel spielen.



Denken Sie bitte an **mathematische Fähigkeiten** von Vorschulkindern, während Sie sich den folgenden kurzen Filmausschnitt anschauen. Sie befinden sich **in der Rolle der Erzieherin und möchten die Kinder mathematisch anregen**.

Fragen zu Filmbeispiel 3

B3.1: Welche **mathematischen Fähigkeiten** haben Sie bei den Kindern beobachten können?

Abbildung 17: Beispiel 3 (Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 5)

Die vier Vignetten wurden aus einer Vielzahl von Aufnahmen mathematischer (Spiel-) Situationen in verschiedenen Kindertagesstätten in Anlehnung an die vier mathematischen Inhaltsbereiche der Kindergarten-Skala-Erweiterung (Roßbach & Tietze, 2010, S. 14–21) ausgewählt. Es wurde auf eine gestufte Anzahl von zu beobachtenden Kindern in den Vignetten geachtet, um eine Zunahme der Komplexität der zu beobachtenden Situationen zu erreichen; die Vignetten sind im Durchschnitt 83,5 Sekunden lang. Tabelle 14 zeigt die Bezüge der Filmbeispiele zu den Inhaltsbereichen, die Dauer der einzelnen Vignetten und die Anzahl der darin vorkommenden Kinder.

Tabelle 14: Bezüge der FiBB-Vignetten zur Subskala Mathematik der KES-R-E (Tietze & Roßbach, 2010)

Vignette	Bereiche gemäß Subskala Mathematik der KES-R-E (Roßbach & Tietze, 2010)	Bezeichnung der Vignette	Dauer [Min.: Sek.]	Anzahl der Kinder
1	1) Zahlen und Zählen 2) mathematische Aktivitäten: Sortieren, Zuordnen und Vergleichen	Stifte	1:59	1
2	Zahlen und Zählen	Brettspiel	1:20	2
3	1) Zahlen und Zählen 2) Lesen und Schreiben einfacher Zahlen	Tisch	1:21	3
4	1) mathematische Aktivitäten: Formen 2) mathematische Aktivitäten: Sortieren, Zuordnen und Vergleichen	Bauecke	0:54	> 3 wechselnd

Entwicklung

Die Pilotierung einer Vorversion des Instruments wurde mit $N = 82$ Erzieherinnen durchgeführt. Änderungen wurden anhand eines Rückmeldebogens teilnehmender Erzieherinnen und anhand des Feedbacks der Untersuchungsleiterinnen vorgenommen. Im Rückmeldebogen wurden die Erzieherinnen sowohl nach schwer verständlichen Fragen oder Begriffen als auch allgemein nach Optimierungsvorschlägen gefragt. Ergänzt wurde z.B. das Bild der videografierten Situation mit einem kurzen einführenden Text und Bezeichnungen für die einzelnen Kinder (K1, K2, ...).

Kategorienbildung (induktiv)

Die folgende qualitative Analyse der mathematischen Vignetten ist als Zwischenschritt zur Gewinnung der Kennwerte für die Beobachtungskompetenz zu sehen.

Die summative Auswertung und vorherige induktive Gewinnung der Kategorien erfolgte mittels qualitativer Inhaltsanalyse nach Mayring (2015) anhand der mit der Endfassung des Instruments gewonnenen Daten. Nach der Transkription der Antworten wurde unter Zuhilfenahme des Programms MAXQDA (Version 10 R240113 © 1995–2011 Udo Kuckartz, Berlin) eine Kategorienbildung gemäß des Prozessmodells induktiver Kategorienbildung von Mayring (ebd., S. 86) durchgeführt. Das Material erwies sich als wenig komplex, da die mathematischen Fähigkeiten von den Befragten zumeist stichwortartig beschrieben wurden (z.B. „K1 erkennt Mengen“ oder „zählt bis 6“), wie es das Beispiel der Erzieherin 003 exemplarisch belegt.

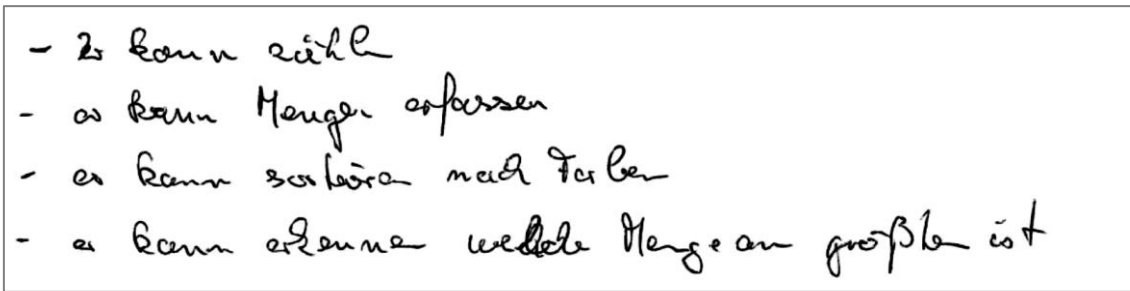


Abbildung 18: Antwort der Erzieherin 003, Vignette 1, Erhebungszeitpunkt 2

Paraphrasen sowie Reduktionsschritte, wie sie von Mayring im Rahmen der MAXQDA-Anwendung empfohlen werden (Mayring, 2015, S. 118), waren deshalb nicht nötig. Das detailreiche Kategoriensystem reichte zur Gewinnung der Kennwerte für die Beobachtungskompetenz aus.

Die induktive Kategorienbildung erfolgte anhand der folgenden drei Kodierrichtlinien:

Kodierrichtlinien

1. Wird ein Code [Hauptkategorie] und werden anschließend mehrere Subcodes [Unterkategorien] identifiziert, so werden nur die Subcodes kodiert.
2. Jede Aussage wird nur einmal kodiert, auch wenn sie in mehrfacher Formulierung vorkommt.
3. In einzelnen Sätzen darf mehrfach kodiert werden.

Weil es sich bei den in den Vignetten dokumentierten bereichsspezifischen Tätigkeiten um keine komplexen mathematischen Handlungen handelte, genügte zur induktiven Kategorienbildung zumeist eine einzige Revision des Materials.

Tabelle 15 zeigt das Ergebnis der induktiven Kategorienbildung der Beispielvignette *Brettspiel* zu einzelnen Phasen des Filmbeispiels. Die Vignettenbeschreibungen und Kategorienbezüge der anderen Vignetten sind im Anhang zu finden.

Tabelle 15: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 3 von 8 (2. Mathematikvignette)

Zeit	Beschreibung des Videos	Kategorien
00:01	Zwei Jungen sitzen an einem Tisch. Vor ihnen liegt ein Brettspiel. K2 fängt an und würfelt mit beiden Würfeln. Er schaut sich die Würfel genau an, sieht in die Kamera und sagt: „Eine Fünf“. K1 schaut ebenfalls auf die Würfel und sagt zeitgleich mit K2 etwas [„Vier“?]. Dann sieht er grinsend zur anderen Seite und sagt schließlich in die Kamera: „Wenn du eine zwei und eine zwei hast, war das vier“.	<ul style="list-style-type: none"> • Zahlwissen <ul style="list-style-type: none"> ○ Zahlen erkennen • Mengenwissen/-erfassung <ul style="list-style-type: none"> ○ simultane Mengenerfassung ○ Mengen erfassen/erkennen ○ (Er)kennen der Würfelaugen/(Er)kennen der Würfelbilder ○ K1 weiß, dass zwei und zwei vier ergibt • Zusammenzählen können (Addieren) <ul style="list-style-type: none"> ○ Zusammenzählen ohne Hilfsmittel ○ Addieren auf Mengenbasis (Fertigkeit/Wissen)

00:17	K2 fällt ihm ins Wort und sagt „drei Figuren? Äh Max, drei Figuren?!“. K1 grinst und antwortet mit „Ja“.	<ul style="list-style-type: none"> • Mengenwissen/-erfassung <ul style="list-style-type: none"> ○ Mengen erfassen/erkennen ○ simultane Mengenerfassung • Vergleichen <ul style="list-style-type: none"> ○ Vergleichen von Mengen ○ Vergleichen von Zahlen • Erkennen der falschen Figuren-Spieler-Zuordnung • Zuordnen
00:25	K2 greift nach der gelben Figur und es erfolgt ein Schnitt. Danach sind nur noch die beiden anderen Figuren im Spiel.	
00:26	K2 rückt fünf Felder mit der roten Figur und zählt dabei laut „1,2,3,4,5“.	<ul style="list-style-type: none"> • Zählen • Abzählen <ul style="list-style-type: none"> ○ K2 Abzählen der Spielfelder • Rücken der gewürfelten Zahl <ul style="list-style-type: none"> ○ Richtiges Rücken der Steine
00:30	K1 beschwert sich: „Hey, ich war doch der Orangene!“. K2: „Du warst der Blaue.“ K1: „Orangene.“ K2: „Blaue.“ K1: „Orangene.“ K2 beendet schließlich die Diskussion: „Egal, sind doch bloß Spielfigur'n!“	
00:41	K1 würfelt mit beiden Würfeln und fragt sich dann: „Fünf oder sechs?“ Er zählt vor sich hin murmelnd die Würfelaugen, indem er jeden Punkt mit dem Finger antippt. Schließlich stellt er fest: „Fünf“. Er schaut verwundert in die Kamera „Immer noch fünf?“. Er indem er die Augen einzeln laut mitzählt, und kommt dann auf die Anzahl sechs. K2 schaut während dessen mit auf die Würfel und sagt ebenfalls „sechs“.	<ul style="list-style-type: none"> • Zählen <ul style="list-style-type: none"> ○ K1 zählt unsicher (mit Fingern) • Abzählen <ul style="list-style-type: none"> ○ K1 zählt die Würfelaugen (noch) ab • Zusammenzählen können (Addieren) <ul style="list-style-type: none"> ○ Zusammenzählen mit Fingern • Ver zählen
00:55	K1 rückt sechs Felder mit der blauen Spielfigur vor und zählt dabei jeden Schritt laut mit. Seine Figur bleibt ein Feld hinter der roten Spielfigur stehen. K2 tippt die letzten beiden Felder mit dem Finger auf dem Brett nach und zählt „fünf, sechs. Ja. Hm“.	<ul style="list-style-type: none"> • Abzählen <ul style="list-style-type: none"> ○ K2 Abzählen der Spielfelder • Rücken der gewürfelten Zahl <ul style="list-style-type: none"> ○ Falsches Rücken der Steine • Nachzählen
01:06	K2 würfelt mit beiden Würfeln. Er schaut auf das Spielfeld und sagt: „Ich hab aber eigentlich fünf gehabt, du musst hier sein normal.“ Er nimmt den blauen Spielstein in die Hand und stellt ihn ein Feld vor den roten. Mit der Aussage „egal“ stellt er den blauen Stein wieder hinter den roten.	<ul style="list-style-type: none"> • Mengenwissen/-erfassung • Zahlunterschiede erkennen <ul style="list-style-type: none"> ○ Erkennen, dass sechs größer fünf ist (Kardinalaspekt) • Erkennen der falsch stehenden Figur (5./6. Platz – Ordinalaspekt) • Korrigieren des falschen Zuges <ul style="list-style-type: none"> ○ Zurücknehmen der Verbesserung
Ende: 01:21		

Auswertung

Im Anschluss an die Bildung der Kategoriensysteme wurde die Häufigkeit der genannten Kategorien pro Erzieherin ausgezählt und mittels MAXQDA ausgewertet. Diese quantitative Analyse qualitativer Daten erfolgte in Anlehnung an das Phasenmodell zum Verhältnis qualitativer und quantitativer Analyse von Mayring (2015, S. 21). Die so gewonnenen einfachen Summenwerte wurden jedoch nicht direkt interpretiert, weil die qualitative Analyse der letzten Phase des Modells („Rückbezug der Ergebnisse auf die Fragestellung“ (ebd.)) im Rahmen einer Pilotierung zeigte, dass eine schlichte Auszählung der wahrgenommenen mathematischen Fähigkeiten der Kinder nicht angemessen ist. Es wurde die Schwierigkeit der einzelnen Erzieherinnenbeobachtungen außer Acht gelassen. Die Beschreibung, dass ein Kind eine Punktemenge simultan erkennt, ist vom Schwierigkeitsgrad höher anzusehen als die Verhaltensbeschreibung, dass ein Kind bis sechs zählt. Um die Schwierigkeit der Antworten zu berücksichtigen, wurde eine Gewichtung der Erzieherinnenbeobachtungen durch einschlägige Expertinnen mit Professur für Mathematik vorgenommen, welche die Vignetten einschätzten und im Anschluss daran die Schwierigkeit der Einschätzung für Erzieherinnen bewerteten.

Den Expertinnen wurden die Fragebögen zusammen mit folgenden Fragen bzw. Arbeitsanweisungen vorgelegt.

1. Welche mathematischen Fähigkeiten sehen Sie im folgenden Beispiel?
2. Bringen Sie bitte die gesehenen Kompetenzen in eine Rangfolge von ‚leicht erkennbar/leicht zu diagnostizieren‘ (oben) bis ‚schwer erkennbar/schwer zu diagnostizieren‘ (unten).
3. Nicht jede Erzieherin differenziert zwischen den einzelnen Kindern. Bringen sie die gefundenen Kompetenzen ohne Bezug zu den einzelnen Kindern in eine Gesamtrangfolge.

Die Sortierung sollte innerhalb der Spalten: Experteneinschätzung von „leicht erkennbar / leicht zu diagnostizieren“ (oben) bis „schwer erkennbar / schwer zu diagnostizieren“ (unten) erfolgen. Die Zahl hinter der genannten Fähigkeit drückt die Stelle in der Gesamtrangfolge der genannten Fähigkeiten aus.

Wenn mehrere Kinder pro Vignette eingeschätzt werden konnten, erfolgte in einem zweiten Schritt die Bildung einer Rangfolge aller Fähigkeiten ohne Bezug zu einzelnen Kindern. Die Expertinnen nummerierten die aufgeführten Fähigkeiten der Kinder pro

Vignette von Eins bis x, wobei Eins ‚einfach zu erkennen‘ bedeutete und ein ansteigender Wert eine Schwierigkeitszunahme für die Erzieherinnen beim Erkennen der entsprechenden Fähigkeit ausdrückte. Anhand dieser Experteneinschätzungen wurden die einzelnen Antworten der Erzieherinnen gewichtet, indem ‚einfach zu erkennende‘ mit eins, ‚schwer zu erkennende Fähigkeiten‘ mit drei und Fähigkeitseinschätzungen mit mittlerem Schwierigkeitsgrad mit zwei multipliziert wurden. Durch diesen Multiplikator wird der Schwierigkeitsgrad der einzelnen Beobachtungen zur Berechnung der Beobachtungskompetenz berücksichtigt.

Die folgende Tabelle zeigt beispielhaft einen Ausschnitt aus dem Kategoriensystem der Vignette *Brettspiel* mit dem jeweiligen Multiplikator.

Tabelle 16: Auszug aus dem Kategoriensystem Vignette 2 mit Multiplikator (Expertinnengewichtung)

Nummer	Vignettenummer \ Kategorie \ Unterkategorie 1 \ Unterkategorie 2 \ ...	Multiplikator
B3.1_01	Nachzählen	1
B3.1_02	Verzählen	1
B3.1_03	Erkennen der falschen Figuren-Spieler-Zuordnung	3
B3.1_04	Zahlunterschiede erkennen	2
B3.1_05	Zahlunterschiede erkennen\Erkennen, dass 6 größer 5 ist (Kardinalaspekt)	2
B3.1_06	Erkennen der falsch stehenden Figur (5./6.Platz – Ordinalaspekt)	3
B3.1_07	Korrigieren des falschen Zuges	1
B3.1_08	Korrigieren des falschen Zuges\Zurücknehmen der Verbesserung	1
B3.1_09	Zählen	1

Die vollständigen Kategoriensysteme mit den durch die Expertinneneinschätzung gewonnenen Multiplikatoren befinden sich im Anhang. Die dort erscheinende Kategorie ‚Sonstige‘ hat den Multiplikator 0 und ging damit nicht in die Berechnung der Beobachtungskompetenz ein, weil in dieser Kategorie fachfremde Beobachtungen, falsche Beobachtungen oder nicht Beobachtbares (Interpretationen) zusammengefasst wurde.

Der Vorteil des geschilderten Verfahrens der qualitativen Inhaltsanalyse nach Mayring (2015) bzw. einer quantitativen Inhaltsanalyse besteht nach Bortz und Döring (2009, S. 149) darin, dass die Ergebnisse aus Häufigkeitsdaten bestehen, die mit entsprechenden inferenzstatistischen Verfahren zu verarbeiten sind und Hypothesentests ermöglichen.

Das Instrument zur Erfassung der Beobachtungskompetenz zeichnet sich durch einen hohen Grad an Objektivität, Reliabilität und Validität aus.

Objektivität

Das Manual und die schriftlichen Testanweisungen gewährleiten die Minimierung der Einflussnahme durch Testleiter und erhöhen damit die Objektivität von Durchführung, Auswertung und Interpretation.

Auswertungsobjektivität ist gegeben, weil das zugrunde gelegte Kategoriensystem ohne Paraphrasen und Reduktionsschritte ausgearbeitet wurde. Dadurch konnten alle transkribierten Beobachtungen direkt und eindeutig einer der detailreichen Kategorien zugeordnet werden. Da die Filmbeispiele wenig komplex sind, bleibt das gesamte Kategoriensystem dennoch sehr übersichtlich. Durch die Eindeutigkeit der Kategorienzuordnung wird der Interpretationsspielraum verringert und damit die Interpretationsobjektivität erhöht.

Die Standardisierung der Beobachtung mittels der Verwendung von Videovignetten sichert die Objektivität wie folgt: Die filmbasierte Form der Beobachtung als ‚nicht-teilnehmendes Verfahren‘ besitzt den Vorteil, dass durch die raum-zeitliche Distanz zwischen Diagnostiker und Diagnostikand Verfälschungen ausgeschlossen werden, welche bei einer teilnehmenden Beobachtung durch die Anwesenheit des Versuchsleiters entstehen, wie zum Beispiel Pygmalion- oder Rosenthal-Effekt (Eid, Gollwitzer & Schmitt, 2010, S. 23). Vorteil der Arbeit mit Filmvignetten ist auch, dass der Blickwinkel des Beobachters standardisiert ist und damit auch Beobachtungsorte, Beobachtungszeiten, Beobachtungsobjekte und Beobachtungseinheiten (Bortz & Döring, 2016, S. 326).

Reliabilität

Die Reliabilität wird durch die Intercoderreliabilität belegt. Sie wurde analysiert, indem die vier Kategoriensysteme der einzelnen Vignetten von jeweils zwei Ratern eingeschätzt und die Übereinstimmung ermittelt wurde, wie es bei inhaltsanalytischen Reliabilitätsbestimmungen üblich ist (vgl. Mayring, 2015, S. 124).

Das Verfahren hat im vorliegenden Fall seine Berechtigung, obwohl die Eignung der Intercoderreliabilität generell als Maß zur Feststellung der Reliabilität kontrovers diskutiert wird (vgl. ebd.), weil es sich bei FiBB um ein detailreiches und zugleich wenig komplexes Kategoriensystem handelt.

Die Berechnung der Intercoderreliabilität erfolgte mittels MAXQDA, wobei die Übereinstimmung der Segmente auf 10% reduziert wurde. Dies erwies sich als notwendig, da in den Kodierrichtlinien keine Angabe zum Umfang der zu markierenden Satzteile

gegeben wurde. So war es möglich, dass Rater 1 den ganzen Satz „das Kind sortiert die Stifte und benennt Mengen ohne lange zu überlegen“ zweimal komplett markiert ($2 * 71 = 142$ Zeichen mit Leerzeichen) und jeweils einer Kategorie zuordnet. Rater 2 hingegen könnte nur „sortiert“ und „benennt Mengen“ markieren und diese Segmente je einer Kategorie (22 Zeichen mit Leerzeichen) zuordnet. Eine Übereinstimmung würde in diesem Fall z.B. bei 15,49% liegen. Da es letztlich nur um das Vorhandensein des entsprechenden Kategoriencodes innerhalb einer Zeile bzw. eines Satzes ging, wurde die Reduzierung der prozentualen Segmentübereinstimmung als unproblematisch gesehen und einer kompletten Neukodierung mit Ergänzung der Kodierrichtlinien auch aus Gründen der Ökonomie vorgezogen.

Wie Tabelle 17 zu entnehmen ist, liegt die Intercooderreliabilität (Mayring, 2015, S. 124) zu Messzeitpunkt t_1 (Prätest) bei 84.54 und zu t_3 (Posttest) bei 83.13. Damit ist die Reliabilitätsanforderung von über .80, die Bortz und Döring (2016, S. 465) als ausreichend betrachten, erfüllt.

Tabelle 17: Intercooderreliabilitäten des qualitativen Anteils der Beobachtungskompetenzerfassung
($N(t_1) = 51$, $N(t_3 \text{ mit Kontrollgruppe}) = 103$)

Vignettennummer	(Titel)	Intercooderreliabilität zu t_1	Intercooderreliabilität zu t_3
1	(Sortieren)	84.71	83.65
2	(Brettspiel)	87.91	83.66
3	(Ziffer 6 und Zählen)	86.10	81.44
4	(Bauecke)	79.44	83.77
1–4		$Mt_1 = 84.54$	$Mt_2 = 83.13$

Validität

Die Sicherung der internen Validität erfolgte durch Experten, die anhand einer Einschätzung der Vignetten auch den diagnostischen Anspruch der Erzieherinnenbeobachtungen einschätzten. Zur Erhöhung der Validität wurde die Auswertung anhand der beschriebenen Kategoriensysteme von zwei Raterinnen durchgeführt und deren Ergebnisse gemittelt.

Das Verfahren ist auch ökologisch valide, weil die Filmvignetten dem natürlichen Alltag der Kindertagesstätte möglichst nahekommen.

7.5 Durchführung

7.5.1 ... der Stichprobenrekrutierung

Zur Stichprobengewinnung wurden im April 2010 die Fachberater(innen) für Kindertageseinrichtungen in Rheinland-Pfalz und die Vertretungen der evangelischen Landeskirchen und katholischen Bistümer, die LIGA der Spitzenverbände der freien Wohlfahrtspflege und die kommunalen Spitzenverbände in Rheinland-Pfalz angeschrieben. Gebeten wurde um die Rückmeldung von Einrichtungen, die sich für die Teilnahme an der Studie interessierten und im laufenden Kindergartenjahr (2009–2010) keinen expliziten Arbeitsschwerpunkt auf den Bereich Mathematik gelegt hatten, z.B. durch den Einsatz bereichsspezifischer Medienpakete oder die Durchführung bestimmter Programme, wie z.B. *Mathe-Kings* (Hoenisch & Niggemeyer, 2004), *Zahlenland* nach Friedrich und Galgóczy (2008) oder Preiß (2009). Bis Ende Mai 2010 wurden 48 Einrichtungen gemeldet, deren Leitungen im Folgenden telefonisch befragt wurden zur Art ihrer Einrichtung (z.B. Teilzeit-, Ganztageseinrichtung), Gruppenstruktur (z.B. geschlossene/offene Gruppen, altersheterogen/-homogen), Konzeption (Situationsansatz, Montessori, Reggio), Anzahl der pädagogischen Fachkräfte (insgesamt und für die Vorschulkinder), Anzahl der Vorschulkinder im Kindergartenjahr 2010/11, bisherigen Erfahrung der Fachkräfte mit mathematischen Förderkonzepten und zu bisher besuchten mathematischen Fortbildungen der Fachkräfte.

Anhand der gewonnenen Informationen wurden 20 Einrichtungen identifiziert, in denen bisher keine bereichsspezifische Förderung durchgeführt worden war und deren Träger oder Leitung sich dazu bereit erklärten, mindestens eine Erzieherin pro Gruppe mit Vorschulkindern für die viertägige Fortbildung freizustellen. Besonders wichtig war darüber hinaus die Akzeptanz einer Zufallszuweisung zu einem lernwegsbasierten *oder* lehrgangsbasierten Förderkonzept *oder* zur Kontrollgruppe, welche eine Fortbildung erst nach Abschluss der Studie erhielt.

Nach Angabe der Fachberater(innen) und der Vertretungen von Kirchen und Spitzenverbänden befanden sich unter den 20 Einrichtungen zwei ‚Brennpunkt‘-Einrichtungen. Dabei handelte es sich um Einrichtungen, welche sich aufgrund der Lage in sozialen Brennpunkten oder eines sehr großen Förderbedarfs der Mehrzahl der Kinder selbst als Brennpunkteinrichtung einstufen. Um eine Gleichverteilung über zwei Treatments und Kontrollgruppe zu ermöglichen, wurde eine weitere ‚Brennpunkt‘-Einrichtung telefo-

nisch akquiriert, wodurch die Gesamtanzahl von 21 Einrichtungen zustande kam. Auch die Brennpunkteinrichtungen wurden den Interventionsgruppen sowie der Kontrollgruppe per Zufall zugewiesen.

Eine Randomisierung der Stichprobe erfolgte auf Einrichtungsebene, die per Zufall (Losverfahren) zwei Treatments und einer Kontrollgruppe zugeordnet wurden. Ein Treatment folgt einem pädagogischen Ansatz lehrgangsbasierter Förderung, das zweite Treatment einem Ansatz lernwegsbasierter Förderung, während die Kontrollgruppe nach einem in Deutschland weit verbreiteten situationsorientierten frühpädagogischen Verständnis (vgl. Donie et al., 2013, S. 310–311) arbeitet. Es wurden 18 der 21 Einrichtungen gemäß ihrer Größe (Anzahl der Gruppen bzw. Kinder) eingeteilt und dann im Losverfahren den Treatments und der Kontrollgruppe zugewiesen. Anschließend wurden die Brennpunkteinrichtungen per Los den Treatments zugeordnet, sodass letztlich sieben Kitas pro Treatment nach dem Zufallsprinzip zugeteilt wurden.

7.5.2 ... der Fortbildungen

Ökologische Validität der Treatments

Die Treatments bestanden aus zwei Fortbildungen zu unterschiedlichen Förderkonzepten. Die ökologische Validität dieser Treatments wurde gesichert, indem die Fortbildungen durch externe Fachleute durchgeführt wurden. Die Dauer von zwei Fortbildungstagen wurde festgelegt, auf die Inhalte jedoch kein Einfluss genommen. Die Qualifizierungen fanden nicht in den teilnehmenden Einrichtungen, sondern in externen Räumlichkeiten statt.

Es gelang zu beiden Förderkonzepten die besten Vertreterinnen des Ansatzes für die Fortbildungen zu gewinnen.

Die Fortbildungen zur lehrgangsbasierten Förderung mittels *Komm mit ins Zahlenland* (Friedrich & Galgóczy, 2008) wurden extern vom Institut für vorschulisches Lernen (Waldkirch/Breisgau) und die Fortbildungen zur lernwegsbasierten Förderung nach dem KiDZ-Ansatz (Stiftung Bildungspakt Bayern, 2007) von einer anerkannten Expertin aus dem KiDZ-Projekt durchgeführt.

Organisation und Implementationskontrolle der Treatments (Fortbildungen)

Um allen pädagogischen Fachkräften eines Kindergartens die Teilnahme an der Fortbildung zu ermöglichen, wurde diese doppelt angeboten, sodass je eine Hälfte des Teams an einer von zwei aufeinander folgenden Veranstaltungen teilnehmen konnte. Mindestanforderung war, dass eine Erzieherin pro Gruppe, in der eine bereichsspezifische Förderung stattfinden sollte, an den Veranstaltungen teilnahm. Zu Zwecken der Dokumentation wurden die Veranstaltungen videografiert, sofern die durchführenden Personen damit einverstanden waren. Alle Einrichtungen erhielten umfangreiche Materialien. In den Fällen, in denen nicht videografiert werden durfte, erfolgte die Implementationskontrolle mittels teilnehmender Beobachtung.

Diese Beobachtung zeigte, dass die Fortbildungen zu beiden Ansätzen sehr handlungsorientiert ausgerichtet waren und ein Schwerpunkt auf konkrete Angebote zur Arbeit mit den Kindern gelegt wurde. Der Inhalt der zugrunde liegenden Förderansätze unterschied sich insbesondere durch formal-didaktische Angebote zur Strukturierung, wobei für den Einsatz im vorstrukturierten Rahmen des lehrgangsbasierten Ansatzes eine Vielzahl von Medien vorgestellt wurde, die auch im lernwegsorientierten Ansatz zur mathematischen Frühförderung empfohlen wurden. Dafür zeigten sich im lernwegsorientierten Ansatz mehr Möglichkeiten zur inhaltlichen Differenzierung.

Kontrollgruppe

Die Erzieherinnen der Kontrollgruppe nahmen an keiner Fortbildung teil, sie führten nach eigenen Aussagen keine gezielte bereichsspezifische Förderung durch, sondern arbeiteten mit dem in Deutschland weit verbreiteten Situationsansatz, in dem alltägliche Situationen und Themen aufgegriffen und Themen gemeinsam mit den Kindern ermittelt werden.

7.5.3 ... der mathematischen Frühförderung

Die mathematische Förderung fand in allen teilnehmenden Einrichtungen einmal pro Woche statt. Gestaltet wurden die Fördermaßnahmen von einer hauptverantwortlichen Erzieherin mit Fortbildung, die von einer zweiten Person unterstützt wurde, welche die Fortbildung nicht besucht haben musste. Die Förderung wurde mit stammgruppenübergreifenden Vorschulgruppen additiv durchgeführt und fand in beiden Treatments entsprechend in separaten Räumlichkeiten statt.

7.5.4 Erhebungszeitpunkte und Begründung

Die Facetten diagnostischer Kompetenz wurden jeweils zu zwei Messzeitpunkten erhoben. Die erste Erhebung der *Urteilsgenauigkeit* fand zu Beginn des Kindergartenjahres 2010/11 statt und die zweite am Ende. Dazwischen lagen die zweitägige Fortbildung und eine ca. neunmonatige Förderphase, in der die Erzieherinnen der Treatmentgruppen bereichsspezifische Fördermaßnahmen durchführten.

Die *Beobachtungskompetenz* der Erzieherinnen wurde unmittelbar vor und im Abstand weniger Tage nach den zweitägigen Fortbildungen zu Beginn des Kindergartenjahres 2010/11 gemessen. Diese Erhebung konnte direkt nach den Fortbildungen durchgeführt werden, weil sie mit Videovignetten erfolgte, bei denen die Erzieherinnen fremde Kinder einschätzen mussten.

Für die Erfassung der *Urteilsgenauigkeit*, die auf Einschätzungen von den Erzieherinnen bekannten Kindern basiert, war eine direkte Erhebung nach den Fortbildungen nicht möglich, weil ein Kennenlernen der Bezugskinder durch die Zusammenarbeit über einen längeren Zeitraum nötig war.

Die Messung der Beobachtungskompetenz erfolgte in der Kontrollgruppe nur einmal in den Kindergärten vor Ort. Es wurde davon ausgegangen, dass es in der Kontrollgruppe zwischen t_1 und t_2 keine Veränderung gab, da dazwischen nur zwei Tage lagen.

7.6 Auswertung und poweranalytische Überlegungen

Im Folgenden werden die Methodenentscheidungen pro Forschungsfrage kurz begründet, poweranalytische Überlegungen zu den jeweiligen Stichproben angestellt und die Voraussetzungen berichtet. Aufgrund des Stellenwerts der Stichprobenqualität wird darauf hingewiesen, dass die zugrunde liegende Gesamtstichprobe aufgrund der Einbettung des Dissertationsvorhabens in das Rahmenprojekt (vgl. 7.1) nicht zu beeinflussen war

Forschungsfrage 1: Wie ist die bereichsspezifische diagnostische Kompetenz von Erzieherinnen hinsichtlich der Facetten Urteilsgenauigkeit und Beobachtungskompetenz ausgeprägt?

Zur Beantwortung der ersten Forschungsfrage nach der Ausprägung der einzelnen Indikatoren diagnostischer Kompetenz wird als Maß der zentralen Tendenz das arithmeti-

sche Mittel, als Dispersionsmaß die Standardabweichung und zudem Mini- und Maximalwert berichtet.

Forschungsfrage 2: Gibt es einen Zusammenhang zwischen Urteilsgenauigkeit und Beobachtungskompetenz?

Berechnet werden die Korrelationen zwischen den Posttestwerten der einzelnen Komponenten der Urteilsgenauigkeit (Niveau-, Merkmals-, Personen-, Wechselwirkungskomponente) einschließlich globaler Urteilsgenauigkeit und den Prätestwerten der Beobachtungskompetenz, welche die Summe der in den Videovignetten erkannten kindlichen Fähigkeiten abbildet, die mittels Expertenurteil unterschiedlich gewichtet wurden.

In einem ersten Schritt erfolgt die Kontrolle der ‚Ausreißer‘. Hierzu geben Bortz und Döring (2016) Folgendes zu bedenken: „Da die Datenaufbereitung einen Eingriff in die Daten darstellt, der prinzipiell auch zu Manipulationszwecken missbraucht werden kann (z.B. hypothesenkonträre Fälle werden als ‚Ausreißer‘ ausgeschlossen), ist das Vorgehen bei der Datenaufbereitung gründlich zu reflektieren und im Ergebnisbericht entsprechend zu begründen und zu dokumentieren“ (ebd., S. 581). Zudem weisen sie darauf hin, dass Korrelationskoeffizienten durch einzelne Ausreißerwerte stark verzerrt werden können (ebd., S. 682). Ausreißerwerte können im vorliegenden Fall ein besonderes Problem darstellen, weil weder Beobachtungskompetenz noch die Komponenten der Urteilsgenauigkeit eine Beschränkung im Sinne z.B. einer Ratingskala aufweisen (vgl. ebd.). Vor allem bei kleinen Stichproben sind Ausreißer, die auf Störungseinflüssen basieren, zu entfernen, da sie zu irreführenden Korrelationen führen (v.a. bei kleinen Stichproben) (Schendera, 2008, S. 33).

Die Ausreißer in der vorliegenden Stichprobe wurden einzeln geprüft. Da die Kontrolle der Ausreißerwerte weder Fehler in der Datenübertragung noch fehlende Werte als Ursache und auch keine mangelnde Plausibilität der Antwortmuster zeigt, also insgesamt keine Störungseinflüsse identifiziert werden konnten, wurden die von SPSS als Ausreißer gekennzeichneten Fälle in der Stichprobe belassen.

Der Frage nach einem möglichen Zusammenhang von Beobachtungskompetenz und den Komponenten der Urteilsgenauigkeit wird mittels einer Pearson-Korrelation (Bortz & Schuster, 2010, S. 153; Bortz & Döring, 2016, S. 660) nachgegangen. Bezüglich der Voraussetzungen konstatieren Bortz und Döring (2016) unter Verweis auf Bortz und Schuster (2010, S. 162): Dieser Korrelationstest setzt eine bivariate Normalverteilung der Merkmale voraus, wobei Bortz und Döring zufolge diese Voraussetzung auf der

einen Seite schwer zu prüfen ist, auf der anderen Seite allerdings dieser Test auf Voraussetzungsverletzungen äußerst robust reagiert (Bortz & Döring, 2016, S. 660). Vor Berechnung der Korrelation wurden die Daten zur Urteilsgenauigkeit gespiegelt, sodass kleine Werte mit einer schlechten und große Werte mit einer guten Urteilsgenauigkeit einhergehen, anschließend wurden die Daten von Beobachtungskompetenz und Urteilsgenauigkeit z- standardisiert.

Die Stichprobenumfangsplanung bzw. das Berechnen der Teststärke a priori erfolgt mittels G*Power (Faul, Erdfelder, Lang & Buchner, 2007). Die Berechnungen basieren auf einem positiven deutlichen Zusammenhang von $\rho = 0.6$ bei einem Signifikanzniveau von 5% und mit einer Power von $> 80\%$. Gemäß G*Power müssen mindestens 19 Personen untersucht werden, weshalb die Stichprobengröße von $N = 39$ als Grundlage der Berechnungen verwendet werden kann.

Zeigt sich primär eine statistisch signifikante Korrelation, werden die Ergebnisse explorativ mittels einfacher Streudiagramme überprüft und – insbesondere im Hinblick auf Scheinkorrelationen – theoretische Überlegungen zur Trendkurve angestellt bzw. überprüft, wie gut sich die vorhandenen Daten an verschiedene Modelle anpassen (vgl. Bortz & Schuster, 2010, S. 199). Gewählt wird das Modell, welches sich den Daten nach dem Kriterium der kleinsten Quadrate am besten anpasst (Bortz & Schuster, 2010, S. 199).

Forschungsfrage 3: Unterscheidet sich die Urteilsgenauigkeit von Erzieherinnen mit einer und ohne eine Fortbildung zur mathematischen Frühförderung, die kein explizites Training diagnostischer Kompetenz beinhaltet?

Forschungsfrage 4: Unterscheidet sich die Veränderung der Urteilsgenauigkeit von Erzieherinnen durch Fortbildungen in verschiedenen bereichsspezifischen Förderansätzen?

Zur Beantwortung der dritten Forschungsfrage nach dem Zusammenhang zwischen Fortbildungsbesuch und etwaiger Urteilsgenauigkeitsveränderungen und zur Beantwortung der vierten Forschungsfrage nach Unterschieden in Abhängigkeit von der Art der Fortbildung werden zweifaktorielle Varianzanalysen mit Messwiederholung mit den Faktoren Treatment und Zeit (t_1 und t_2) (vgl. Bortz & Schuster, 2010, S. 288–290) durchgeführt. Es wird die Interaktion der beiden Faktoren untersucht, die nach Bortz und Döring (2016, S. 840) bei diesem methodischen Vorgehen bedeutsam ist, weil sie über gruppenspezifische Veränderungen informiert.

Der Faktor Treatment ist in den Berechnungen zur Beantwortung der Forschungsfrage 3 zweistufig (Treatmentgesamtgruppe, Kontrollgruppe) und in den Berechnungen zur Forschungsfrage 4 dreistufig (Lehrgang, Lernweg, Kontrollgruppe).

Das Programm G*Power (Faul et al., 2007) erlaubt in der vorliegenden Version noch keine direkte Berechnung der Teststärke für Varianzanalysen mit Messwiederholung auf mehreren Stufen. Die Autoren Rasch, Friese, Hofmann und Naumann (2014a, S. 12) stellen in ihren online publizierten G*Power-Ergänzungen übergangsweise eine alternative Berechnungsart vor, mit der die Rechnung trotzdem mit der aktuellen Version von G*Power durchgeführt werden kann. Die Berechnungen basieren auf einem mittleren Effekt nach Cohen von „ $f = 0.25$ “ (Cohen, 1988, S. 286) und zeigen, dass für den dreistufigen Haupteffekt mind. 20 Personen und für den zweistufigen Haupteffekt mind. 17 Personen untersucht werden müssen, um jeweils einen mittleren Effekt bei einem Signifikanzniveau von 5% und einer angenommenen Korrelation zwischen den Faktorstufen von $r = 0.3$ mit einer Power von $> 80\%$ zu finden (vgl. Rasch et al., 2014a, S. 13–14). Aus diesem Grund ist der Stichprobenumfang der Stichprobe *Urteilsgenauigkeit* ($n = 39$) auch hier als ausreichend zu sehen.

Forschungsfrage 5: Unterscheidet sich die *Beobachtungskompetenz* von Erzieherinnen mit einer und ohne eine Fortbildung zur mathematischen Frühförderung, die kein explizites Training diagnostischer Kompetenz beinhaltet?

Forschungsfrage 6: Unterscheidet sich die Veränderung der *Beobachtungskompetenz* von Erzieherinnen durch Fortbildungen in verschiedenen bereichsspezifischen Förderansätzen?

Für die Forschungsfragen 5 und 6 gelten dieselben Überlegungen wie für die Fragen 3 und 4, weil auch hier eine zweifaktorielle Varianzanalyse mit Messwiederholung eingesetzt wird. Dementsprechend ist die Stichprobengröße ‚Beobachtungskompetenz‘ von 103 Personen (52 mit Fortbildung, 51 ohne Fortbildung) akzeptabel, weil sie die Mindeststichprobengröße überschreitet, die nötig ist, um einen mittleren Effekt bei einem Signifikanzniveau von 5% und einer angenommenen Korrelation zwischen den Faktorstufen von $r = 0.3$ mit einer Power von $> 80\%$ zu finden (vgl. Rasch et al., 2014a, S. 13–14).

8. Ergebnisse

In Kapitel 8 werden zunächst die deskriptiven Statistiken zur Urteilsgenauigkeit und Beobachtungskompetenz berichtet (Kap. 8.1, Forschungsfrage 1). Ob Zusammenhänge zwischen Urteilsgenauigkeit und Beobachtungskompetenz bestehen, zeigt Kapitel 8.2 (Forschungsfrage 2). Kapitel 8.3 legt die Ergebnisse zur Veränderung der Urteilsgenauigkeit durch die Teilnahme an bereichsspezifischen Fortbildungen (Forschungsfrage 3) und etwaige Einflüsse durch verschiedene Fortbildungsansätze (Forschungsfrage 4) dar. In Kapitel 8.4 werden die Ergebnisse zur Veränderung der Beobachtungskompetenz durch die Teilnahme an bereichsspezifischen Fortbildungen (Forschungsfrage 5) und etwaige Einflüsse durch verschiedene Fortbildungsansätze (Forschungsfrage 6) dargestellt.

8.1 Deskriptive Ergebnisse

Zur Beantwortung der Forschungsfrage 1, wie die bereichsspezifische diagnostische Kompetenz von Erzieherinnen hinsichtlich der Facetten Urteilsgenauigkeit und Beobachtungskompetenz ausgeprägt ist, werden im Folgenden zunächst für die Facette Urteilsgenauigkeit die deskriptiven Ergebnisse der Komponentenanalyse nach Cronbach (1955, vgl. Kap. 3.1.2 in dieser Arbeit) berichtet und für die Facette Beobachtungskompetenz die Ergebnisse beschrieben, welche mit der filmbasierten Befragung zur Erfassung der Beobachtungskompetenz (FiBB, vgl. Kap. 7.4.3) erhoben wurden. Die deskriptiven Daten werden jeweils zum einen für die gesamte Teilstichprobe und zum anderen aufgeteilt nach Treatment dargestellt.

8.1.1 Urteilsgenauigkeit

Zur Berechnung der Urteilsgenauigkeit kommt die Komponentenanalyse von Cronbach (1955) zur Anwendung. Bei den ermittelten Genauigkeitsmaßen bedeutet ein Wert von 0 eine perfekte Übereinstimmung zwischen der Einschätzung der Erzieherin und dem Testwert des Zielkinds. Je weiter sich die Werte von 0 entfernen, umso ungenauer ist das Urteil.

Tabelle 18 zeigt die deskriptive Statistik der Urteilsgenauigkeitskomponenten (Niveau-, Personen-, Merkmals- und Wechselwirkungskomponente) sowie der globalen Urteilsgenauigkeit aufgeteilt nach Fortbildungsgruppe, die sich aus den Teilstichproben Lehrgang und Lernweg zusammensetzt, Kontrollgruppe und die Gesamtstichprobe.

Tabelle 18: Deskriptive Werte zu den Urteilsgenauigkeitskomponenten zu t_1 , Fortbildungsgruppe gesamt und aufgeteilt nach Treatments, Kontrollgruppe und Gesamtgruppe

	Fortbildung	Lehrgang	Lernweg	Kontrollgruppe	Gesamt
	FB (LG + LW)	LG	LW	KG	FB + KG
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
	<i>n = 29</i>	<i>n = 12</i>	<i>n = 17</i>	<i>n = 10</i>	<i>N = 39</i>
Niveau- komponente	1.11 (1.51)	1.39 (2.02)	.91 (1.05)	.72 (.61)	1.01 (1.34)
Personen- komponente	1.61 (2.32)	2.43 (3.32)	1.02 (.99)	2.10 (3.40)	1.73 (2.60)
Merkmals- komponente	5.41 (4.71)	7.06 (6.37)	4.25 (2.71)	6.42 (5.54)	5.67 (4.88)
Wechselwirkungs- komponente	5.07 (5.25)	6.51 (7.04)	4.05 (3.38)	6.37 (6.53)	5.40 (5.54)
Globale Urteilsgenauigkeit	12.94 (12.09)	16.78 (16.62)	10.23 (6.85)	15.62 (15.09)	13.81 (12.77)

Die Gesamtstichprobe zeigt, dass es den Erzieherinnen zu t_1 am besten gelingt, das Niveau ihrer Gruppe einzuschätzen (Niveauelement), wobei ihnen auch die Wahrnehmung der Unterschiede unter den Kindern ihrer Gruppe (Personenelement) gut gelingt. Die größeren Mittelwerte in Tabelle 18 zeigen, dass es ihnen wesentlich schwerer fällt, spezifische Aufgabenschwierigkeiten im mathematischen Bereich korrekt einzuschätzen (Merkmalskomponente) und Unterschiede zwischen den Kindern im Hinblick auf einzelne Aufgaben (Wechselwirkungskomponente) einzuschätzen.

Die deskriptive Statistik der Fortbildungsgruppe, die aus den Teilstichproben Lehrgang und Lernweg besteht, weist ebenso wie die Kontrollgruppe ähnliche Größenverhältnisse in den einzelnen Komponenten auf (Tab. 18): Die Mittelwerte zeigen, dass auch innerhalb der Treatments Niveau- und Personenelemente deutlich kleinere und ähnlichere Werte in allen Bedingungen aufweisen als Merkmals- und Wechselwirkungskomponente. Auffällig ist, dass die Erzieherinnen in der Bedingung Lernweg in allen Komponenten – außer der Niveauelemente der Kontrollgruppe – bessere Werte zeigen, was im Summenwert der Komponenten (globale Urteilsgenauigkeit) besonders deutlich wird.

Die globale Urteilsgenauigkeit der Gruppe Lernweg ist 40% größer als die der Gruppe Lehrgang und 35% größer als die der Kontrollgruppe, was sich in den kleineren Werten dieses Genauigkeitsmaßes zeigt.

Die Korrelationskomponente der Personenkomponente, welche die Kompetenz der Erzieherin zeigt, ihre Bezugskinder entsprechend ihrer Leistungen in eine Reihenfolge zu bringen, liegt bei einer mittleren Korrelation (Median) $r = .69$ und einer Spannweite von $r = -.91$ und $r = .96$. Diese Korrelationskomponente der Personenkomponente wird ergänzend berichtet, weil sie oft in internationalen Publikationen als alleiniges Maß verwendet wird (Schrader & Praetorius, 2018, S. 93); sie entspricht der Rangordnungs-komponente nach Schrader (1989). Zudem zeigen Schraders Niveauelemente von $M = .24$ ($SD = .99$) mit einer Spannweite von $r = -2.73$ und $r = 1.8$, dass die mittlere Leistung der Kinder eher überschätzt wird und seine Differenzierungskomponente mit $M = 1.00$ ($SD = .26$) und einer Spannweite von $r = .0$ und $r = 1.6$, dass auch die Heterogenität der Kinderleistungen überschätzt wird.

Prätestunterschiede zwischen Fortbildungs- und Kontrollgruppe

Die Prätestunterschiede werden zwischen Fortbildungs- und Kontrollgruppe auf ihre statistische Signifikanz hin überprüft. Mit Hilfe eines t-Tests für unabhängige Stichproben werden die Unterschiede zwischen Fortbildungs- und Kontrollgruppe zum Zeitpunkt t_1 für die intervallskalierte (vgl. Kap. 3.2) Niveau-, Personen-, Merkmals-, Wechselwirkungskomponente und globale Urteilsgenauigkeit getestet. Zuvor wurde überprüft, dass die Verteilungen der Testwerte in beiden Bedingungen nicht signifikant von einer Normalverteilung abweichen (Kolmogorov-Smirnov-Test; $p > .05$). Eine Annahme auf Normalverteilung zeigt sich nur für die Niveauelemente der Kontrollgruppe (Kolmogorov-Smirnov-Test; $p = .200$). Das Verfahren kann trotz dieser Verletzung der Normalverteilungsannahme für die restlichen Komponenten beibehalten werden, weil der t-Test auf die Verletzung der Normalverteilungsannahme äußerst robust reagiert – auch bei kleinen Stichproben (Rasch, Friese, Hofmann & Naumann, 2014b, S. 44; vgl. Bortz & Schuster, 2010, S. 127). Die Varianzhomogenitätsannahme ist nicht verletzt (Levene-Test; $p_{\text{Niveauelemente}} = .152$; $p_{\text{Personenelemente}} = .158$; $p_{\text{Merkmalskomponente}} = .486$; $p_{\text{Wechselwirkungskomponente}} = .365$; $p_{\text{Globale Urteilsgenauigkeit}} = .359$). Der t-Test für unabhängige Stichproben zeigt gemäß Tabelle 19, dass die beschriebenen Unterschiede zwischen Fortbildungsgruppe und Kontrollgruppe zum Zeitpunkt t_1 statistisch nicht bedeutsam sind.

Tabelle 19: t-Test für unabhängige Stichproben (Fortbildungs- und Kontrollgruppe) zu t_1

Df = 37	<i>t</i>	<i>p</i>
Niveauekomponente	.783	.438
Personenkomponente	-.517	.608
Merkmalskomponente	-.557	.581
Wechselwirkungskomponente	-.637	.528
Globale Urteilsgenauigkeit	-.567	.574

Prätestunterschiede zwischen Treatments (Lehrgang/Lernweg) und Kontrollgruppe

In einem nächsten Schritt werden die Prätestunterschiede zwischen den Treatmentgruppen und der Kontrollgruppe mit einer Varianzanalyse auf ihre statistische Signifikanz hin überprüft. Kontrolliert werden alle Werte zum Zeitpunkt t_1 für Niveau-, Personen-, Merkmals-, Wechselwirkungskomponente und globale Urteilsgenauigkeit. Zuvor wurde sichergestellt, dass die Verteilungen der Testwerte in allen drei Bedingungen nicht signifikant von einer Normalverteilung abweichen (Kolmogorov-Smirnov-Test; $p > .05$). Eine Normalverteilung ist nur für die Niveauekomponente der Kontrollgruppe (Kolmogorov-Smirnov-Test; $p = .200$) gegeben. Auch die Varianzhomogenitätsannahme ist bei der Personen- und Merkmalskomponente sowie bei der globalen Urteilsgenauigkeit verletzt (Levene-Test; $p_{\text{Niveauekomponente}} = .215$; $p_{\text{Personenkomponente}} = .011$; $p_{\text{Merkmalskomponente}} = .021$; $p_{\text{Wechselwirkungskomponente}} = .066$; $p_{\text{Globale Urteilsgenauigkeit}} = .023$). Das Verfahren kann jedoch beibehalten werden, da sich nach Rasch, Friese, Hofmann und Naumann (2014c, S. 31) die Varianzanalyse gegen die Verletzungen der Normalverteilungs- und Varianzhomogenitätsannahme weitgehend robust verhält. Die zweifaktorielle Varianzanalyse zeigt keine signifikanten Unterschiede zwischen den Gruppen (Lehrgang, Lernweg und Kontrollgruppe) zum Zeitpunkt t_1 für Niveau- ($F(2,36) = .760$, $p = .475$), Personen- ($F(2,36) = 1.186$, $p = .317$), Merkmals- ($F(2,36) = 1.355$, $p = .271$), Wechselwirkungskomponente ($F(2,36) = .893$, $p = .418$) und globale Urteilsgenauigkeit ($F(2,36) = 1.095$, $p = .345$).

8.1.2 Beobachtungskompetenz

Wie die bereichsspezifische diagnostische Kompetenz von Erzieherinnen hinsichtlich der Facette Beobachtungskompetenz ausgeprägt ist, wird anhand der Anzahl wahrge-nommener mathematischer Fähigkeiten von Kindern in Alltagssituationen des Kinder-

gartens mittels Filmvignetten erhoben (*Filmbasierte Befragung zur Erfassung der Beobachtungskompetenz* (FiBB) (Donie, 2010b), vgl. Kap. 7.4.3). Es werden die deskriptiven Kennwerte sowohl für die Gesamtstichprobe berichtet als auch für die Interventionsstichprobe der Erzieherinnen, die an einer bereichsspezifischen Fortbildung teilgenommen haben. Diese Fortbildungen folgten zwei unterschiedlichen Ansätzen, einem lehrgangs- und einem lernwegsbasierten (vgl. Kap. 7.2.2).

Tabelle 20 zeigt die deskriptiven Werte zur Beobachtungskompetenz für die Gesamtstichprobe und die Teilstichproben zu Beginn des Kindergartenjahres. Die Gesamtstichprobe setzt sich aus der Fortbildungsgruppe (Lehrgang und Lernweg) der Kontrollgruppe zusammen.

Tabelle 20: Deskriptive Werte zur Beobachtungskompetenz zu t_1

Fortbildung FB (LG + LW)	Lehrgang LG	Lernweg LW	Kontrollgruppe KG	Gesamt FB + KG
<i>M (SD)</i> <i>N = 52</i>	<i>M (SD)</i> <i>N = 26</i>	<i>M (SD)</i> <i>N = 26</i>	<i>M (SD)</i> <i>N = 51</i>	<i>M (SD)</i> <i>N = 103</i>
17.70 (7.87)	20.38 (9.01)	15.02 (5.50)	18.49 (6.98)	18.09 (7.42)

Die Ergebnisse in Tabelle 20 zeigen zu t_1 ähnliche Mittelwerte für Fortbildungs- und Kontrollgruppe, wobei der Unterschied innerhalb der Fortbildungsgruppe zwischen den Treatments Lehrgang ($M = 20.38$) und Lernweg ($M = 15.02$) mit 5.36 Punkten groß erscheint. Die Spannweite liegt zwischen 3 und 37 Punkten, Punkte steht hier für die Anzahl gesehener Fähigkeiten in den Videovignetten, deren Schwierigkeit (Faktoren 1 bis 3 nach Expertinneneinschätzung) berücksichtigt wurde (vgl. Kap. 7.4.3). Die Verteilung der Daten zur Beobachtungskompetenz der Gesamtstichprobe ist normalverteilt (Kolmogorov-Smirnov-Test).

Prätestunterschiede zwischen Fortbildungs- und Kontrollgruppe

Zunächst werden die Prätestunterschiede zwischen Fortbildungs- und Kontrollgruppe auf ihre statistische Signifikanz hin überprüft. Mit Hilfe eines t-Tests für unabhängige Stichproben werden die Unterschiede zwischen den intervallskalierten (vgl. Kap. 7.4.3) Werten der Fortbildungs- und Kontrollgruppe zum Zeitpunkt t_1 getestet. Zuvor wurden die Testwerte in beiden Bedingungen auf Abweichungen von der Normalverteilung überprüft. Es zeigt sich eine Normalverteilung nur für die Kontrollgruppe (Kolmogorov-Smirnov-Test; $p = .200$). Das Verfahren kann wiederum trotz dieser Verletzung

der Normalverteilungsannahme beibehalten werden, weil der t-Test auf die Verletzung der Normalverteilungsannahme äußerst robust reagiert – auch bei kleinen Stichproben (Rasch et al., 2014b, S. 44; vgl. Bortz & Schuster, 2010, S. 127). Die Varianzhomogenitätsannahme ist nicht verletzt (Levene-Test; $p = .620$). Der t-Test für unabhängige Stichproben zeigt, dass der Unterschied zwischen Fortbildungsgruppe ($M = 17.70$, $SD = 7.87$) und Kontrollgruppe ($M = 18.49$, $SD = 6.98$) zum Zeitpunkt t_1 statistisch nicht bedeutsam ist ($t(101) = -.537$, $p = .592$).

Prätestunterschiede zwischen Treatments (Lehrgang/Lernweg) und Kontrollgruppe

In einem nächsten Schritt wurden die Prätestunterschiede zwischen den Treatmentgruppen und der Kontrollgruppe zu t_1 mit einer Varianzanalyse auf ihre statistische Signifikanz hin überprüft. Zuvor wurde sichergestellt, dass die Verteilungen der Testwerte in allen drei Bedingungen nicht signifikant von einer Normalverteilung abweichen (Kolmogorov-Smirnov-Test; $p > .05$). Eine Normalverteilung ist sowohl für die Treatmentgruppe Lernweg (Kolmogorov-Smirnov-Test; $p = .200$) als auch für die Kontrollgruppe (Kolmogorov-Smirnov-Test; $p = .200$) gegeben. Die Varianzhomogenitätsannahme ist verletzt (Levene-Test; $p = .23$). Wegen dieser Verletzung der Voraussetzung für die Treatmentgruppe Lehrgang in zwei von drei Bedingungen wird gemäß Rasch et al. (2014c, S. 106) eine Rangvarianzanalyse (Kruskall-Wallis-Test) durchgeführt. Diese zeigt, dass die Unterschiede in der Beobachtungskompetenz zwischen den Treatmentgruppen Lehrgang ($M = 20.38$, $SD = 9.01$), Lernweg ($M = 15.02$, $SD = 5.50$) und Kontrollgruppe ($M = 18.49$, $SD = 6.98$) zu t_1 statistisch nicht bedeutsam sind (Kruskall-Wallis-Test; $p = .073$).

8.2 Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz

Hinsichtlich der zweiten Forschungsfrage nach dem Zusammenhang zwischen Urteilsgenauigkeit und Beobachtungskompetenz und damit nach der Gültigkeit des postulierten Prozessmodells zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz (Kap. 5) ist festzustellen, dass keine signifikanten Korrelationen (Pearson) zwischen der Beobachtungskompetenz und der globalen Urteilsgenauigkeit oder den einzelnen Urteilsgenauigkeitskomponenten vorliegen, wie Tabelle 21 zeigt. Es handelt sich folglich um voneinander unabhängige Merkmale.

Tabelle 21: Korrelationsmatrix zu Urteilsgenauigkeit und Beobachtungskompetenz

Urteilsgenauigkeit (t ₁)	Beobachtungskompetenz (t ₁)	
	Korrelation nach Pearson	Signifikanz (einseitig)
Niveauelemente	-0.104	0.264
Personenelemente	-0.128	0.219
Merkmalskomponente	-0.133	0.209
Wechselwirkungskomponente	-0.129	0.217
Globale Urteilsgenauigkeit	-0.143	0.192

8.3 Veränderung der Urteilsgenauigkeit

Zur Überprüfung der dritten Forschungsfrage, ob sich die Urteilsgenauigkeit über die Zeit verbessert, indem Erzieherinnen nach der Teilnahme an einer Fortbildung (ohne explizites Training der Urteilsgenauigkeit oder anderer Facetten diagnostischer Kompetenz) Kinder gezielt über einen Zeitraum von neun Monaten bereichsspezifisch fördern, wird pro Komponente eine zweifaktorielle Varianzanalyse mit Messwiederholung durchgeführt. Ob sich etwaige Unterschiede in Hinblick auf die Art der Fortbildung auswirken (Forschungsfrage 4), wird ebenfalls mittels zweifaktorieller Varianzanalyse mit Messwiederholung untersucht.

Voraussetzungen für die zweifaktorielle Varianzanalyse mit Messwiederholung

Es bestehen im Prätest keine statistisch signifikanten Unterschiede zwischen der Fortbildungs- und der Kontrollgruppe (vgl. 8.1.1). Die Unabhängigkeit der Residuen wird wegen der zugrunde gelegten Zufallszuweisung der Förderkonzepte angenommen. Die Sphäritätsannahme ist nicht verletzt, weil der messwiederholte Faktor Zeit zwei Stufen hat (vgl. Janczyk & Pfister, 2015, S. 144). Die Gleichheit der Varianzen der Kennwerte aller Urteilsgenauigkeitskomponenten ist in beiden Gruppen (Fortbildung, Kontrollgruppe) gegeben – sowohl im Prätest (vgl. 8.1.1) als auch im Posttest (Levene-Test; $p_{\text{Niveauelemente}} = .745$; $p_{\text{Personenelemente}} = .282$; $p_{\text{Merkmalskomponente}} = .772$; $p_{\text{Wechselwirkungskomponente}} = .650$; $p_{\text{Globale Urteilsgenauigkeit}} = .831$). Dies ist auch für die Berechnung unter Berücksichtigung der Aufteilung in Treatmentgruppen (Lehrgang, Lernweg) und Kontrollgruppe im Posttest gegeben (Levene-Test; $p_{\text{Niveauelemente}} = .863$; $p_{\text{Personenelemente}} = .074$; $p_{\text{Merkmalskomponente}} = .549$; $p_{\text{Wechselwirkungskomponente}} = .350$; $p_{\text{Globale Urteilsgenauigkeit}} = .328$). Die unkritische Verletzung für drei Komponenten im Prätest wurde in Kapitel 8.1.1 berichtet.

8.3.1 Niveauelemente (Elevation, E)

Cronbachs (1955) Niveauelemente zeigt, wie exakt eine Erzieherin den mittleren Testwert, das Niveau der Gruppe ihrer Bezugskinder einschätzen kann. Rechnerisch handelt es sich um die quadrierte Differenz zwischen dem mittleren Urteilswert einer Erzieherin und dem mittleren Testwert der Gruppe ihrer Bezugskinder.

$$E_j^2 = (y_{..j} - x_{..})^2$$

$y_{..j}$ Mittleres Urteil der Erzieherin j (judge)

$x_{..}$ Mittlerer Testwert der Bezugskindergruppe pro Erzieherin

Ein Wert von 0 beschreibt eine ideale Einschätzung, je weiter sich der Wert von 0 entfernt, desto geringer ist diese Kompetenz ausgeprägt.

In Tabelle 22 finden sich die deskriptiven Ergebnisse, die geringfügige Unterschiede im Prätest und gleiche Werte im Posttest zeigen. Die grafische Darstellung der Kennwerte für Prä- und Posttest zeigt Tabelle 19 für die Gruppen der Erzieherinnen mit Fortbildung und ohne Fortbildung (Kontrollgruppe) sowie für die nach den Treatments Lehrgang und Lernweg differenzierten Gruppen.

Tabelle 22: Deskriptive Statistik der Niveauelemente (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe

	Treatment	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
Prätest	Fortbildung (gesamt)	29	1.11	1.51	.01	7.43
	Lehrgang	12	1.39	2.02	.02	7.43
	Lernweg	17	0.91	1.05	.01	3.24
	ohne Fortbildung (Kontrollgruppe)	10	0.72	0.61	.02	2.10
Posttest	Fortbildung (gesamt)	29	0.63	0.84	.00	3.71
	Lehrgang	12	0.66	0.71	.00	2.18
	Lernweg	17	0.61	0.94	.01	3.71
	ohne Fortbildung (Kontrollgruppe)	10	0.55	0.76	.01	2.40

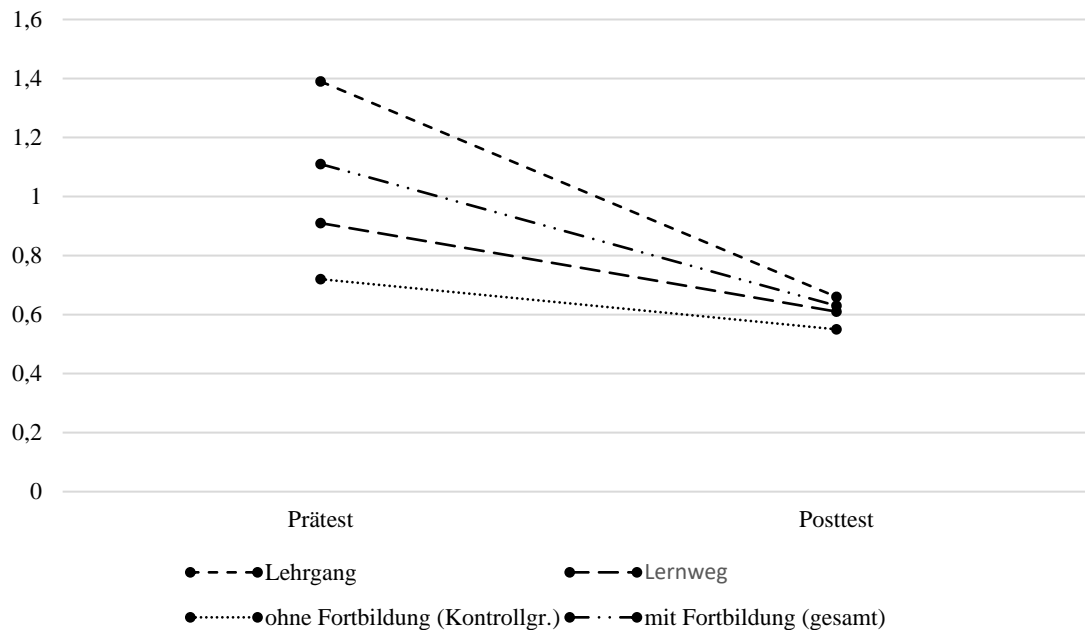


Abbildung 19: Niveauelemente, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)

Ergebnisse zur Varianzanalyse der Gruppen mit und ohne Fortbildung

Die Fähigkeit von Erzieherinnen, das Niveau der Gruppe ihrer Bezugskinder korrekt einzuschätzen, wird weder durch eine Fortbildungsteilnahme noch durch die Zeit oder durch das Zusammenwirken beider Einflussfaktoren verändert. Die zweifaktorielle Varianzanalyse mit Messwiederholung zeigt, dass weder der Einfluss der Zeit auf die Niveauelemente ($F(1) = 1.85, p > .05$) noch der Einfluss der Fortbildungsteilnahme ($F(1) = .48, p > .05$) oder die Wechselwirkungseffekte ($F(1) = .43, p > .05$) signifikant sind. Abbildung 19 zeigt die Kennwerte für die Gruppen mit und ohne Fortbildung für Prä- und Posttest.

Ergebnisse zur Varianzanalyse der Treatmentgruppen (Lehrgang, Lernweg) und der Gruppe ohne Fortbildung (Kontrollgruppe)

Wie deutlich wird, verändert sich die Fähigkeit von Erzieherinnen zur Einschätzung des Gruppenniveaus ihrer Bezugskinder für die Treatmentgruppen weder in Abhängigkeit vom Förderkonzept noch von der Zeit. Auch Effekte, die sich auf einen zeitlichen Einfluss innerhalb der Treatmentgruppen zurückführen lassen, gibt es nicht. Dies zeigt eine zweifaktorielle Varianzanalyse mit Messwiederholung, da keine signifikanten Kennwerte für den Einfluss der Zeit ($F(1) = 3.52, p > .05$), des Treatments ($F(2) = .54, p > .05$) oder Wechselwirkungseffekte ($F(2) = .62, p > .05$) nachzuweisen sind.

Der Post-hoc-Test nach Scheffé zeigt für die Posttestergebnisse keine signifikanten Unterschiede zwischen den Treatments und der Kontrollgruppe ($p > .05$).

8.3.2 Personenkomponente (Differential Elevation, DE)

Die Personenkomponente veranschaulicht, wie die Erzieherin die Unterschiedlichkeit der Kinder ihrer Gruppe wahrnimmt. Sie gibt an, wie sehr die um das generelle Urteilsniveau bereinigten mittleren (d.h. über alle Merkmale gemittelten) Einschätzungen jedes Kindes im Durchschnitt von den entsprechenden mittleren Kriteriumswerten abweichen (Schrader, 1989, S. 76). Wie Karst (2012, S. 45) betont, ist bei der Berechnung dieser Komponente die absolute Übereinstimmung der jeweiligen Subtrahenden irrelevant. Dies gilt auch für die beiden Minuenden. Hier ist vielmehr die Einschätzung des einzelnen Kindes im Verhältnis zum mittleren Erzieherinnenurteil relevant (ebd.).

$$DE_j^2 = \frac{1}{N} \sum [(\bar{y}_{p.j} - \bar{y}_{..j})(\bar{x}_p - \bar{x}_{..})]^2$$

N	Anzahl der Kinder
$y_{p.j}$	Mittleres Erzieherinnenurteil für Kind o der Erzieherin j (judge)
$y_{..j}$	Mittleres Urteil der Erzieherin j
x_p	Mittlerer Testwert des Kindes o in der Kindergruppe der Erzieherin j
$x_{..}$	Mittlerer Testwert der Bezugskindergruppe pro Erzieherin

Ein Wert von 0 beschreibt eine ideale Einschätzung, je weiter sich der Wert von 0 entfernt, desto geringer ist diese Kompetenz ausgeprägt.

Die deskriptive Statistik der Personenkomponente in Tabelle 23 zeigt insbesondere für die Treatmentgruppe Lehrgang einen Kompetenzanstieg, was sich in einer Mittelwertabnahme ausdrückt.

Tabelle 23: Deskriptive Statistik der Personenkomponente (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe

	Treatment	N	M	SD	$Min.$	$Max.$
Prätest	Fortbildung (gesamt)	29	1.61	2.32	.00	10.10
	Lehrgang	12	2.43	3.32	.00	10.10
	Lernweg	17	1.02	0.99	.07	03.71
	ohne Fortbildung (Kontrollgruppe)	10	2.10	3.40	.13	10.16
Posttest	Fortbildung (gesamt)	29	0.64	0.95	.02	04.79
	Lehrgang	12	0.80	1.39	.02	04.79
	Lernweg	17	0.53	0.47	.07	01.62
	ohne Fortbildung (Kontrollgruppe)	10	1.02	1.13	.09	03.15

Die grafische Darstellung der Veränderung zwischen Prä- und Posttest zeigt Abbildung 20 für die Gruppe der Erzieherinnen mit und ohne Fortbildung und die nach den Treatments Lehrgang und Lernweg differenzierten Gruppen.

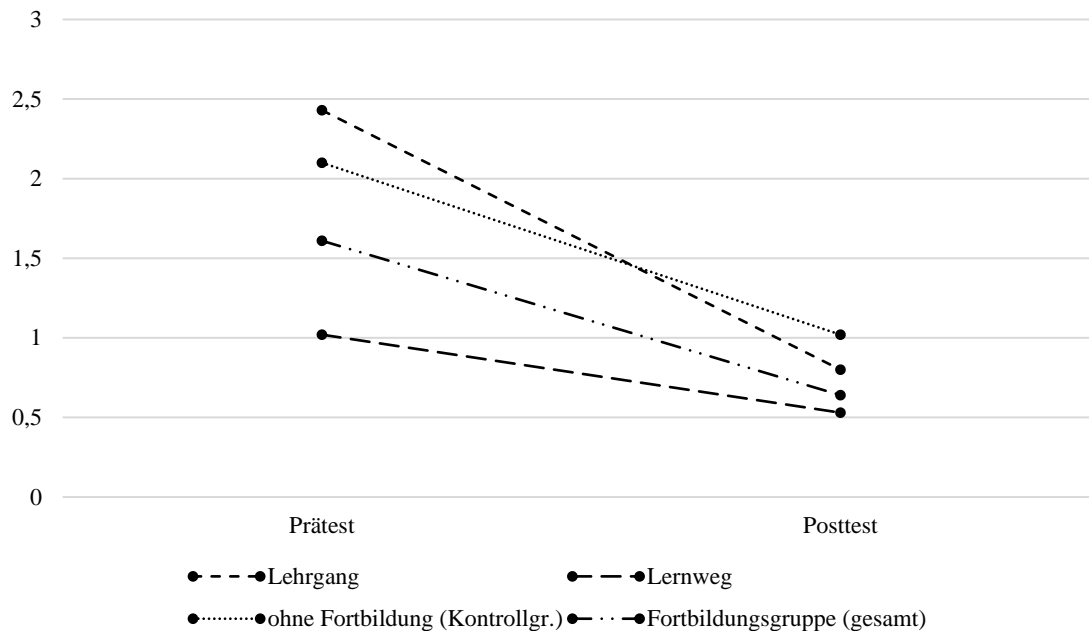


Abbildung 20: Personenkomponente, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)

Ergebnisse zur Varianzanalyse der Gruppen mit und ohne Fortbildung

Die Fähigkeit der Erzieherinnen zu einer zunehmend besseren Wahrnehmung der Unterschiede zwischen den Kindern ihrer jeweiligen Bezugsgruppe wird weder durch eine Fortbildungsteilnahme noch durch die im Treatment verbrachte Zeit verändert, wie die zweifaktorielle Varianzanalyse mit Messwiederholung für den Einfluss der Fortbildungsteilnahme ($F(1) = .48, p > .05$) und die Wechselwirkungseffekte ($F(1) = .71, p > .05$) zeigt. Unabhängig vom Treatment hat der Faktor Zeit jedoch einen marginal signifikanten Einfluss auf die Wahrnehmung der individuellen Leistungen ihrer Bezugskinder ($F(1) = 4.00, p = .053, \eta_p^2 = .098$), wobei die Effektgröße η_p^2 zeigt, dass sich 9,8% der Variabilität der Personenkomponente durch den Faktor Zeit aufklären lassen. Die Effektstärke liegt bei $f = 0.33$, weshalb es sich gemäß Cohen (1988, 284–287) um einen Effekt mittlerer Effektstärke handelt.

Ergebnisse zur Varianzanalyse der Treatmentgruppen (Lehrgang, Lernweg) und der Kontrollgruppe

Die zweifaktorielle Varianzanalyse mit Messwiederholung zeigt, dass auch die Variabilität im Hinblick auf die Fähigkeit der Erzieherinnen zu einer zunehmend besseren Wahrnehmung der Unterschiede zwischen den Kindern ihrer jeweiligen Bezugsgruppe unabhängig von der bereichsspezifischen Förderkonzeption oder von der Wechselwirkung dieses Faktors mit dem der Zeit ist. Es gibt keinen signifikanten Einfluss des Treatments ($F(2) = 1.67, p > .05$) oder der Wechselwirkung zwischen Zeit und Treatment ($F(2) = 0.60, p > .05$) auf die Personenkomponente. Allerdings ist der Einfluss des Faktors Zeit auf die Personenkomponente zwischen Prä- und Posttest signifikant ($F(1) = 5.48, p = .025, \eta_p^2 = .132$), wobei η_p^2 zeigt, dass sich 13% der Variabilität der Personenkomponente durch den Faktor Zeit aufklären lassen. Die Effektstärke liegt bei $f = 0.34$, weshalb es sich gemäß Cohen (1988, 284–287) um einen Effekt mittlerer Effektstärke handelt.

Der Post-hoc-Test nach Scheffé zeigt für die Posttestergebnisse keine signifikanten Unterschiede zwischen den Treatments und der Kontrollgruppe ($p > .05$).

8.3.3 Merkmalskomponente (Stereotype Accuracy, SA)

Die Merkmalskomponente nach Cronbach (1955) bringt zum Ausdruck, wie sehr die um das generelle Urteilsniveau bereinigten mittleren Merkmalseinschätzungen von den entsprechenden Kriteriumswerten abweichen (Schrader, 1989, S. 76) und beschreibt damit die Fähigkeit der Erzieherin, spezifische Aufgabenschwierigkeiten korrekt einzuschätzen, die Genauigkeit also mit der verschiedene Merkmale eingeschätzt werden können.

$$SA_j^2 = \frac{1}{k} \sum [(\bar{y}_{ij} - \bar{y}_{..j})(\bar{x}_{.i} - \bar{x}_{..})]^2$$

k Anzahl der Items

y_{ij} Mittleres Erzieherinnenurteil für Item i der Erzieherin j

$y_{..j}$ Mittleres Urteil der Erzieherin j (judge)

$x_{.i}$ Mittlerer Testwert der Bezugskindergruppe pro Erzieherin bei Item i

$x_{..}$ Mittlerer Testwert der Bezugskindergruppe pro Erzieherin

Ein Wert von 0 beschreibt eine ideale Einschätzung, je weiter sich der Wert von 0 entfernt, desto geringer ist diese Kompetenz ausgeprägt.

Die deskriptive Statistik der Merkmalskomponente in Tabelle 24 zeigt, dass die Erzieherinnen zum Zeitpunkt des Posttests die spezifischen Aufgabenschwierigkeiten in nahezu allen Gruppen gleich gut einschätzen können.

Tabelle 24: Deskriptive Statistik der Merkmalskomponente (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe

	Treatment	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
Prätest	Fortbildung (gesamt)	29	5.41	4.71	0.00	20.92
	Lehrgang	12	7.06	6.37	0.00	20.92
	Lernweg	17	4.25	2.71	1.42	12.70
	ohne Fortbildung (Kontrollgruppe)	10	6.42	5.54	2.19	18.42
Posttest	Fortbildung (gesamt)	29	3.25	2.36	0.80	12.11
	Lehrgang	12	3.48	3.12	0.80	12.11
	Lernweg	17	3.09	1.73	1.00	06.67
	ohne Fortbildung (Kontrollgruppe)	10	3.86	1.98	1.69	07.99

Die grafische Darstellung der Kennwerte für Prä- und Posttest zeigt Abbildung 21 für die Gruppen der Erzieherinnen mit und ohne Fortbildung sowie für die nach den Treatments Lehrgang und Lernweg differenzierten Gruppen.

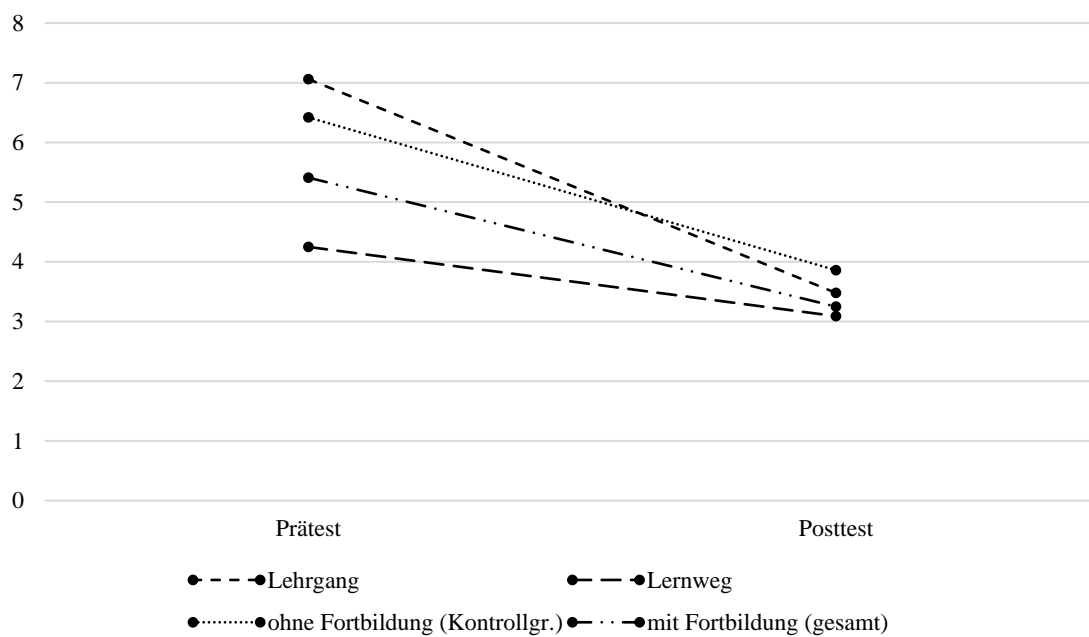


Abbildung 21: Merkmalskomponente, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)

Ergebnisse zur Varianzanalyse der Gruppen mit und ohne Fortbildung

Die Fähigkeit von Erzieherinnen, verschiedene Merkmale genau einzuschätzen (Aufgabenschwierigkeit), wird weder durch eine Fortbildungsteilnahme noch durch die im Treatment verbrachte Zeit verändert, wie es die zweifaktorielle Varianzanalyse mit Messwiederholung für den Einfluss der Fortbildungsteilnahme ($F(1) = .67, p > .05$) und die Wechselwirkungseffekte ($F(1) = .04, p > .05$) zeigt. Die Zeit unabhängig vom Treatment hat jedoch einen marginal signifikanten Einfluss auf die Wahrnehmung der individuellen Leistungen ihrer Bezugskinder. Durch diesen signifikanten Einfluss der Zeit auf die Merkmalskomponente ($F(1) = 5.54, p = .024, \eta_p^2 = .130$) lassen sich 13% der Variabilität der Merkmalskomponente aufklären. Die Effektstärke liegt bei $f = 0.39$, weshalb es sich gemäß Cohen (1988, S. 284–287) um einen Effekt mittlerer Effektstärke handelt.

Ergebnisse zur Varianzanalyse der Treatmentgruppen (Lehrgang, Lernweg) und der Kontrollgruppe

Es zeigt sich, dass die Varianz hinsichtlich der Fähigkeit der Erzieherinnen, spezifische Aufgabenschwierigkeiten korrekt einzuschätzen (vgl. Karst, 2012, S. 44) abhängig vom Faktor Zeit ist, wie die zweifaktorielle Varianzanalyse mit Messwiederholung belegt ($F(1) = 7.44, p = .010, \eta_p^2 = .171$). Wie η_p^2 zeigt, lassen sich 17% der Variabilität der Merkmalskomponente durch den Faktor Zeit aufklären. Die Effektstärke liegt bei $f = 0.45$, weshalb es sich gemäß Cohen (1988, S. 284–287) um einen starken Effekt handelt.

Der Ansatz der bereichsspezifischen Fortbildung oder die im jeweiligen Treatment verbrachte Zeit haben keinen Einfluss auf die Fähigkeit zur Einschätzung von Aufgabenschwierigkeiten, wie es die fehlende Signifikanz der Einflüsse des Treatments ($F(2) = 1.65, p > .05$) und der Wechselwirkungseffekte ($F(2) = 0.72, p > .05$) belegt.

Der Post-hoc-Test nach Scheffé zeigt für die Posttestergebnisse keine signifikanten Unterschiede zwischen den Treatments und der Kontrollgruppe ($p > .05$).

8.3.4 Wechselwirkungskomponente (Differential Accuracy, DA)

Die Wechselwirkungskomponente nach Cronbach (1955) zeigt die Genauigkeit des Erzieherinnenurteils unter Kontrolle der Haupteffekte von Personen und Merkmalen (Schrader, 1989, S. 78). Rechnerisch handelt es sich um die Differenz der Residuen von Erzieherinnenurteil und Residuen der Testwerte des Kindes.

$$DA_j^2 = \frac{1}{kN} \sum_p \sum_i (y'_{pij} - x'_{pi})^2$$

- k** Anzahl der Items
N Anzahl der Kinder
 y'_{pij} Residuum des Erzieherinnenurteils von Erzieherin j
 x'_{pi} Residuum der Testwerte der Kinder

Ein Wert von 0 beschreibt eine ideale Einschätzung, je weiter sich der Wert von 0 entfernt, desto geringer ist diese Kompetenz ausgeprägt.

Die deskriptive Statistik der Wechselwirkungskomponente ist in Tabelle 25 abzulesen. Es zeigt sich, dass die minimale Variabilität zum Zeitpunkt des Prätests bis zum Posttest nahezu verschwindet. Es gelingt den Erzieherinnen zum Zeitpunkt des Posttests in allen Bedingungen nahezu gleich gut, die Unterschiede zwischen den Kindern im Hinblick auf einzelne Aufgaben einzuschätzen.

Tabelle 25: Deskriptive Statistik der Wechselwirkungskomponente (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe

	Treatment	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
Prätest	Fortbildung (gesamt)	29	5.07	5.25	0.00	23.21
	Lehrgang	12	6.51	7.04	0.00	23.21
	Lernweg	17	4.05	3.38	1.13	14.68
	ohne Fortbildung (Kontrollgruppe)	10	6.37	6.53	2.10	19.59
Posttest	Fortbildung (gesamt)	29	3.84	0.84	1.55	6.04
	Lehrgang	12	3.84	1.02	1.55	5.43
	Lernweg	17	3.83	0.73	2.83	6.04
	ohne Fortbildung (Kontrollgruppe)	10	3.99	0.64	3.00	5.00

Die grafische Darstellung der Kennwerte für Prä- und Posttest zeigt Abbildung 22 für die Gruppen der Erzieherinnen mit und ohne Fortbildung sowie für die nach den Treatments Lehrgang und Lernweg aufgeteilten Gruppen.

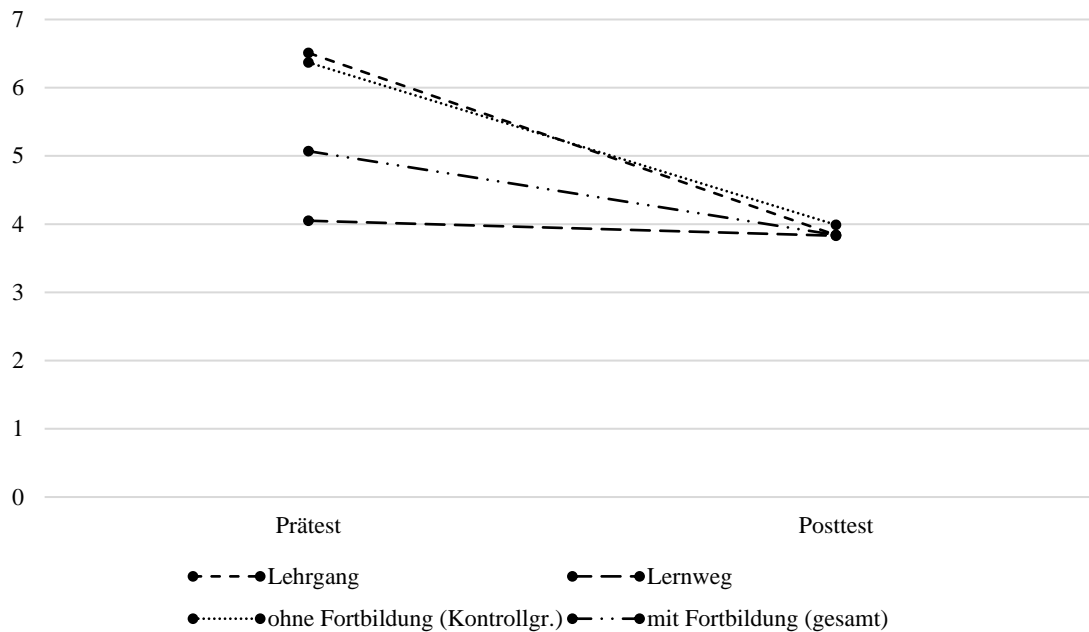


Abbildung 22: Wechselwirkungskomponente, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)

Ergebnisse zur Varianzanalyse der Gruppen mit und ohne Fortbildung

Die Fähigkeit von Erzieherinnen, die Unterschiede zwischen den Kindern im Hinblick auf einzelne Aufgaben einzuschätzen, wird weder durch die Teilnahme an einer Fortbildung noch durch den Faktor Zeit oder durch das Zusammenwirken beider Einflussfaktoren verändert. Die zweifaktorielle Varianzanalyse mit Messwiederholung zeigt, dass weder der Einfluss des Faktors Zeit auf die Wechselwirkungskomponente ($F(1) = 2.93$, $p > .05$) noch der Einfluss der Fortbildungsteilnahme ($F(1) = .52$, $p > .05$) oder die Wechselwirkungseffekte ($F(1) = .30$, $p > .05$) signifikant sind.

Ergebnisse zur Varianzanalyse der Treatmentgruppen (Lehrgang, Lernweg) und der Kontrollgruppe

Es zeigt sich, dass die Variabilität hinsichtlich der Genauigkeit in der Beurteilung von Unterschieden zwischen Kindern im Hinblick auf einzelne Aufgaben unabhängig von der Zeit, von der bereichsspezifischen Förderkonzeption sowie von der in den Treatments verbrachten Zeit ist. Weder der Einfluss der Zeit zwischen Prä- und Posttest ($F(1) = 3.49$, $p > .05$) noch der Einfluss des Treatments ($F(2) = 0.97$, $p > .05$) oder die Wechselwirkungseffekte ($F(2) = 0.791$, $p > .05$) erweisen sich als signifikant.

Der Post-hoc-Test nach Scheffé zeigt für die Posttestergebnisse keine signifikanten Unterschiede zwischen den Treatments und der Kontrollgruppe ($p > .05$).

8.3.5 Globale Urteilsgenauigkeit (Global Accuracy, ACC)

Die globale Urteilsgenauigkeit als Summenwert der Einzelkomponenten Niveau-, Personen-, Merkmals- und Wechselwirkungskomponente bildet die Gesamtheit der Komponenten diagnostischer Kompetenz gemäß Cronbach (1955) wie folgt in einem Wert ab.

$$\text{Globale Urteils-} \\ \text{genauigkeit} = \text{Niveaui} \\ \text{kom-} \\ \text{ponente} + \text{Personenkom-} \\ \text{ponente} + \text{Merkmalskom-} \\ \text{ponente} + \text{Wechsel-} \\ \text{wirkungskom-} \\ \text{ponente}$$

Die deskriptive Statistik der globalen Urteilsgenauigkeit ist in Tabelle 26 abzulesen. Es zeigt sich, dass sich die Mittelwerte nach dem Prätest bis zum Posttest angleichen. Im Durchschnitt aller Teilfähigkeiten, die als unterschiedliche Komponenten in die globale Urteilsgenauigkeit eingehen, gelingt es den Erzieherinnen zum Zeitpunkt des Posttests in allen Bedingungen nahezu gleich gut, die Kinder anhand der Beurteilung ihrer Fähigkeit zum Lösen mathematischer Aufgaben einzuschätzen (vgl. Kap. 7.4.3).

Tabelle 26 zeigt die deskriptive Statistik der globalen Urteilsgenauigkeit.

Tabelle 26: Deskriptive Statistik der globalen Urteilsgenauigkeit (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe

	Treatment	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
Prätest	Fortbildung (gesamt)	29	12.94	12.09	0.00	52.01
	Lehrgang	12	16.78	16.32	0.00	52.01
	Lernweg	17	10.23	06.85	3.04	31.16
	ohne Fortbildung (Kontrollgruppe)	10	15.61	15.09	5.84	48.18
Posttest	Fortbildung (gesamt)	29	8.36	3.71	4.60	23.19
	Lehrgang	12	8.78	5.08	4.60	23.19
	Lernweg	17	8.07	2.47	5.24	15.29
	ohne Fortbildung (Kontrollgruppe)	10	9.43	3.24	5.64	15.95

Die grafische Darstellung der Kennwerte für Prä- und Posttest zeigt Abbildung 23 für die Gruppen der Erzieherinnen mit und ohne Fortbildung sowie für die nach den Treatments *Lehrgang* und *Lernweg* differenzierten Gruppen.

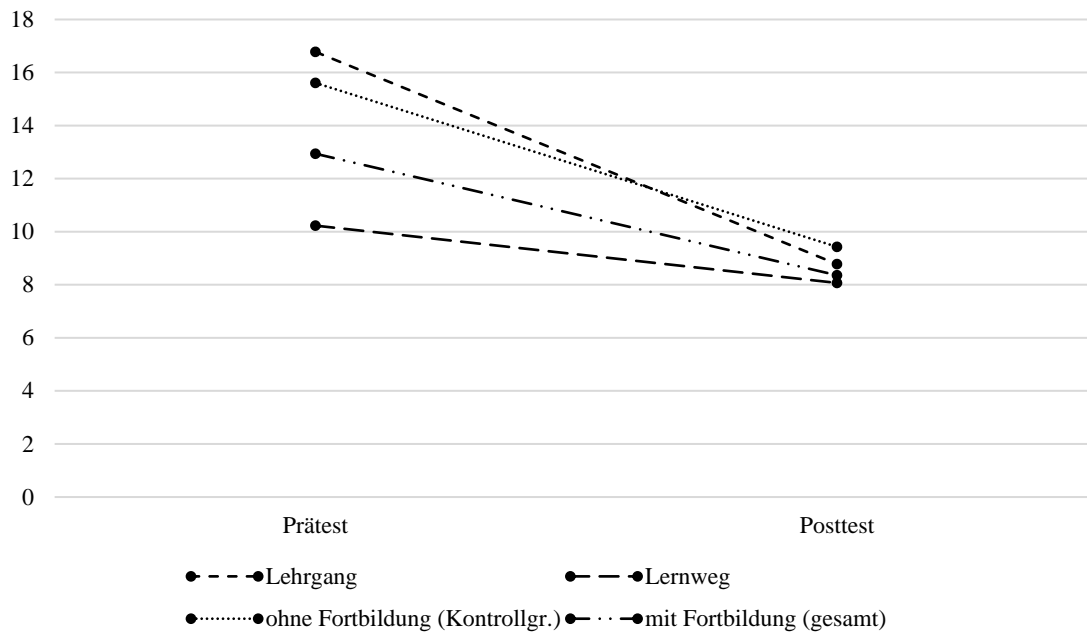


Abbildung 23: Globale Urteilsgenauigkeit, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)

Ergebnisse zur Varianzanalyse der Gruppen mit und ohne Fortbildung

Die globale Urteilsgenauigkeit von Erzieherinnen wird durch den Faktor Zeit beeinflusst ($F(1) = 4.74$, $p = .036$, $\eta_p^2 = .114$), die 11% der Variabilität der globalen Urteilsgenauigkeit aufklärt. Die Effektstärke liegt bei $f = 0.32$ und ist gemäß Cohen (1988, 284–287) als mittelstark zu bezeichnen. Weder die Fortbildungsteilnahme noch die in den Treatments verbrachte Zeit hat einen Einfluss auf die globale Urteilsgenauigkeit, wie die zweifaktorielle Varianzanalyse mit Messwiederholung für den Einfluss der Fortbildungsteilnahme ($F(1) = .59$, $p > .05$) und die Wechselwirkungseffekte ($F(1) = .11$, $p > .05$) zeigt.

Ergebnisse zur Varianzanalyse der Treatmentgruppen (Lehrgang, Lernweg) und der Kontrollgruppe

Die Variabilität des Summenwerts der einzelnen Urteilsgenauigkeitskomponenten ist abhängig vom Faktor Zeit, jedoch nicht von der bereichsspezifischen Förderkonzeption oder vom Zusammenwirken der Faktoren Zeit und Treatment. Wie die Effektstärke η_p^2 zeigt, lassen sich 15% der Variabilität der Urteilsgenauigkeit durch den Faktor Zeit aufklären. Dies belegt eine zweifaktorielle Varianzanalyse mit Messwiederholung, für den Faktor Zeit ($F(1) = 6.12$, $p = .018$, $\eta_p^2 = .145$), den Faktor Treatment ($F(2) = 2.04$, $p > .05$) und die Wechselwirkungseffekte ($F(2) = 1.39$, $p > .05$). Die Effektstärke liegt bei $f = 0.41$ und ist gemäß Cohen (1988, 284–287) als stark zu bezeichnen.

Der Post-hoc-Test nach Scheffé zeigt für die Posttestergebnisse zwischen den Treatments und der Kontrollgruppe keine signifikanten Unterschiede ($p > .05$).

8.3.6 Zusammenfassung

Die Frage, ob sich die Urteilsgenauigkeit der Erzieherinnen in Abhängigkeit vom Besuch einer Fortbildung oder in Abhängigkeit davon verändert, nach welchem Ansatz sie fortgebildet wurden, lässt sich zusammenfassend verneinen. Die Art und Weise der bereichsspezifischen Fortbildung (Treatment) spielt keine Rolle für die Variabilität der Urteilsgenauigkeit von Erzieherinnen. Auch die Unterschiedlichkeit in den einzelnen Urteilsgenauigkeitskomponenten lassen sich nicht auf die Fortbildung oder deren Ansatz (Treatment) zurückführen. Der Haupteffekt *Zeit* hingegen ist für die Summe der Komponenten (Global Accuracy) signifikant, für die Einzelkomponenten gilt dies nur teilweise. Der Faktor *Zeit* hat einen jeweils marginal signifikanten Einfluss auf die Merkmals- und Personenkomponente bei der varianzanalytischen Untersuchung der Gruppen mit und ohne Fortbildung. Der Einfluss des Faktors *Zeit* ist sowohl für Merkmals- als auch Personenkomponente signifikant bei der varianzanalytischen Untersuchung unter Berücksichtigung der Treatments. Für die Niveauelemente sind hinsichtlich der Varianz keine signifikanten Unterschiede zwischen Prä- und Posttest nachweisbar.

Die Effektgrößen (η_p^2) der signifikanten Effekte zeigen, dass der Faktor *Zeit* in der varianzanalytischen Untersuchung der Gruppen mit und ohne Fortbildung 11% der Variabilität der globalen Urteilsgenauigkeit aufklärt, 13% der Varianz der Merkmalskomponente und 9,8% der Varianz der Personenkomponente. Die Effektstärken sind gemäß Cohen (1988, 284–287) sowohl bei der globalen Urteilsgenauigkeit ($f = .32$) als auch bei der Merkmals- ($f = .39$) und der Personenkomponente ($f = .32$) als mittel zu bezeichnen.

Bei der varianzanalytischen Untersuchung unter Berücksichtigung der einzelnen Treatmentgruppen zeigen die Effektgrößen (η_p^2) der signifikanten Effekte zeigen, dass der Faktor *Zeit* 15% der Variabilität der globalen Urteilsgenauigkeit aufklärt, 17% der Varianz der Merkmalskomponente und 13% der Varianz der Personenkomponente. Die Effektstärken sind gemäß Cohen (1988, S. 284–287) bei der globalen Urteilsgenauigkeit ($f = .41$) und der Merkmalskomponente ($f = .45$) als stark und bei der Personenkomponente ($f = .34$) als mittel zu bezeichnen.

Auffallend ist die ähnliche Entwicklung der einzelnen Urteilsgenauigkeitskomponenten, die sich in den grafischen Darstellungen zu den Varianzanalysen zeigt. Es gibt keine signifikanten Unterschiede zwischen den Gruppen zum Zeitpunkt des Posttests.

Entscheidend für die Veränderungen von Merkmals-, Personenkomponente und globaler Urteilsgenauigkeit ist nicht der Besuch einer Fortbildung oder der Typus dieser Fortbildung, sondern einzig der Einfluss der Zeit.

8.4 Veränderung der Beobachtungskompetenz durch Fortbildungen

Zur Überprüfung der fünften Forschungsfrage, ob sich die Beobachtungskompetenz von Erzieherinnen nach ihrer Teilnahme an einer zweitägigen bereichsspezifischen Fortbildung (ohne explizites Training der Beobachtungskompetenz oder anderer Facetten diagnostischer Kompetenz) verbessert, wird eine zweifaktorielle Varianzanalyse mit Messwiederholung durchgeführt. Ob sich etwaige Unterschiede in Hinblick auf die Art der Fortbildung auswirken (Forschungsfrage 6), wird ebenfalls mittels zweifaktorieller Varianzanalysen mit Messwiederholung untersucht.

Voraussetzungen für die zweifaktorielle Varianzanalyse mit Messwiederholung

Es bestehen zu t_1 keine statistisch signifikanten Unterschiede zwischen der Fortbildungs- und Kontrollgruppe (siehe 8.1.2). Die Unabhängigkeit der Residuen wird aufgrund der zugrunde gelegten Zufallszuweisung der Förderkonzepte angenommen. Die Sphärizitätsannahme ist nicht verletzt, weil der messwiederholte Faktor Zeit zwei Stufen hat (vgl. Janczyk & Pfister, 2015, S. 144).

Die Gleichheit der Varianzen der Kennwerte aller Urteilsgenauigkeitskomponenten ist in beiden Gruppen (Fortbildung, Kontrollgruppe) gegeben – sowohl im Prätest (Levene-Test; $p = .620$) als auch im Posttest (Levene-Test; $p = .282$). Die Voraussetzung *Gleichheit der Varianzen* ist für die Aufteilung in Treatmentgruppen (Lehrgang, Lernweg) und Kontrollgruppe zu t_1 nicht gegeben (Levene-Test; $p = .23$). Das Verfahren kann jedoch beibehalten werden, da sich nach Rasch et al. (2014c, S. 31) die Varianzanalyse gegen diese Verletzungen der Voraussetzung weitgehend robust verhält. Im Posttest ist die Gleichheit der Varianzen auch für die Aufteilung in Treatmentgruppen (Lehrgang, Lernweg) und Kontrollgruppe gegeben (Levene-Test; $p = .338$).

Die Beobachtungskompetenz der Kontrollgruppe wurde nur einmal erhoben. Weil zwischen den Messzeitpunkten der Interventionsgruppe wegen der zweitägigen Fortbildung ein so kurzer Zeitraum lag, dass keine Veränderungen zu erwarten waren, wurden die t_1 -Daten der Kontrollgruppe für die Berechnungen zu t_2 erneut eingesetzt.

In Tabelle 27 finden sich die deskriptiven Ergebnisse. Während im Prätest Unterschiede zu sehen sind, die allerdings statistisch nicht signifikant sind (vgl. 8.1.2), zeigen sich im Posttest nahezu gleiche Werte.

Tabelle 27: Deskriptive Statistik der Beobachtungskompetenz (Prä- und Posttest), getrennt nach Fortbildung (gesamt), Treatments (Lehrgang, Lernweg) und Kontrollgruppe

	Treatment	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
Prätest	Fortbildung (gesamt)	52	17.70	7.87	3	37
	Lehrgang	26	20.38	9.01	7	37
	Lernweg	26	15.02	5.50	3	25
	ohne Fortbildung (Kontrollgruppe)	51	18.49	6.98	3	34
Posttest	Fortbildung (gesamt)	52	20.72	6.18	9	35
	Lehrgang	26	21.37	6.67	11	34
	Lernweg	26	20.08	5.72	9	35
	ohne Fortbildung (Kontrollgruppe)	51	18.49	6.98	3	34

Dies zeigt auch die grafische Darstellung der Kennwerte für Prä- und Posttest (Abb. 24) für die Gruppen der Erzieherinnen mit Fortbildung und ohne Fortbildung (Kontrollgruppe) sowie für die nach den Treatments *Lehrgang* und *Lernweg* differenzierten Gruppen.

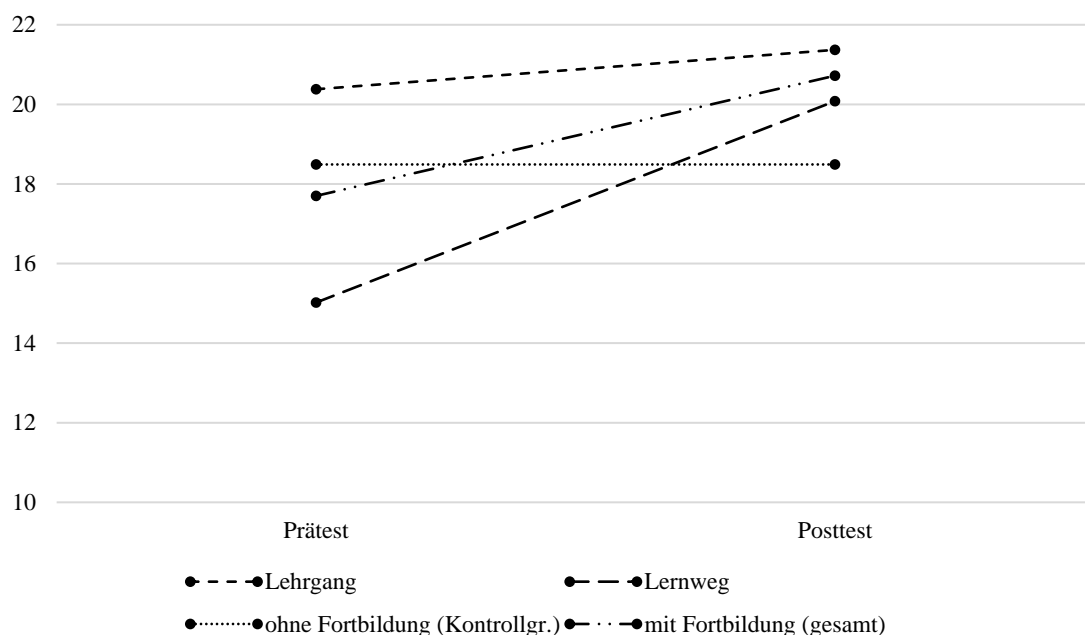


Abbildung 24: Beobachtungskompetenz, absolute Werte (Prä- und Posttest) von Gruppen mit Fortbildung (gesamt, Lehrgang, Lernweg) und ohne Fortbildung (Kontrollgruppe)

Ergebnisse zur Varianzanalyse der Gruppen mit und ohne Fortbildung

Die Kompetenz von Erzieherinnen, mathematische Fähigkeiten von Kindern in Videovignetten zu erkennen, wird durch die Zeit in Abhängigkeit von der Teilnahme an der Fortbildung signifikant beeinflusst, wie der signifikante Interaktionseffekt zwischen Zeit und Gruppen zeigt ($F(1) = 10.76$, $p = .001$, $\eta_p^2 = .096$). Dabei werden 9,6% der Variabilität der Beobachtungskompetenz aufgeklärt. Die Effektstärke liegt bei $f = 0.33$ und ist gemäß Cohen (1988, S. 284–287) als mittel zu bezeichnen.

Ergebnisse zur Varianzanalyse der Treatmentgruppen (Lehrgang, Lernweg) und der Kontrollgruppe

Auch für die Unterscheidung der Veränderung in Abhängigkeit von den einzelnen Treatmentgruppen zeigt sich, dass die Kompetenz von Erzieherinnen, mathematische Fähigkeiten von Kindern in Videovignetten zu erkennen, durch die Zeit in Abhängigkeit von den Gruppen signifikant beeinflusst wird. Dies zeigt der signifikante Interaktionseffekt zwischen Zeit und Gruppen ($F(1) = 11.34$, $p < .001$, $\eta_p^2 = .185$). Dadurch werden 18,5% der Variabilität der Beobachtungskompetenz aufgeklärt. Die Effektstärke liegt bei $f = 0.48$ und ist gemäß Cohen (1988, S. 284–287) als stark zu bezeichnen. Der Post-hoc-Test nach Scheffé zeigt für die Posttestergebnisse keine signifikanten Unterschiede zwischen den Treatments und der Kontrollgruppe ($p > .05$).

9. Zusammenfassung

Die bisherige Diagnosekompetenzforschung hat trotz langjähriger Forschungstraditionen für die Praxis kaum relevante Befunde erbracht (vgl. Kap. 3.4), was für den Elementarbereich in besonderer Weise gilt. Dies wurde in der vorliegenden Arbeit als Anlass für den Vorschlag genommen, die in der Accuracy-Forschung als reine Urteilsgenauigkeit gefasste diagnostische Kompetenz um eine zweite Facette zu erweitern. Die vorliegende Untersuchung knüpft damit an die traditionelle Forschung an, welche auf die Komponentenanalyse Cronbachs (1955) zurückgeht. Mit der Einführung der Beobachtungskompetenz als weitere Facette setzt die vorliegende Studie zudem an der Forschung zu formativem Assessment an.

Es wird analysiert, wie die beiden Facetten der diagnostischen Kompetenz Urteilsgenauigkeit und Beobachtungskompetenz im Elementarbereich ausgeprägt sind, ob diese Facetten gemäß dem neu entwickelten Prozessmodells (Kap. 5) zusammenhängen und wie sie durch bereichsspezifische Fortbildungen und explizit durch verschiedene Fortbildungsansätze (Lehrgangs- vs. Lernwegorientierung) verändert werden können.

Zur Beantwortung der ersten Forschungsfrage, wie die bereichsspezifische diagnostische Kompetenz von Erzieherinnen hinsichtlich der Facetten Urteilsgenauigkeit und Beobachtungskompetenz ausgeprägt ist, wurden verschiedene Instrumente entwickelt: die *Befragung zur Erfassung der Urteilsgenauigkeit von Erzieherinnen* (BEUrteilEr) (Donie, 2010a) und die *Filmbasierte Befragung zur Erfassung der Beobachtungskompetenz* (FiBB) (Donie, 2010b). Hinsichtlich der Ausprägung von Urteilsgenauigkeit und Beobachtungskompetenz (Forschungsfrage 1, Kap. 8.1) zeigt sich, dass Erzieherinnen das Niveau ihrer Gruppe im Vergleich der Urteilsgenauigkeits-Komponenten am besten einschätzen können (Niveauelement), mit Tendenz zur Überschätzung. Auch die Wahrnehmung von Unterschieden zwischen den Kindern ihrer Gruppe gelingt ihnen gut (Personenelement), dabei neigen sie eher zur Überschätzung der Heterogenität. Wesentlich schwerer fällt es ihnen, spezifische Aufgabenschwierigkeiten im mathematischen Bereich korrekt einzuschätzen (Merkmalselement) und Unterschiede zwischen den Kindern im Hinblick auf einzelne Aufgaben (Wechselwirkungselement) einzuschätzen. Die Erzieherinnen sind darüber hinaus problemlos in der Lage, die Kinder entsprechend ihrer Leistungen in eine Rangreihenfolge zu bringen, wie es die Korrelationskomponente der Personenelement zeigt.

Die Beobachtungskompetenz von Erzieherinnen zeigt sich in der Anzahl erkannter Fähigkeiten von Kindern während verschiedener Tätigkeiten mit mathematischem Bezug. Dabei wurde der Schwierigkeitsgrad der Beobachtung mittels Expertinneneinschätzung berücksichtigt. Die Beobachtungskompetenz der Erzieherinnen ist normalverteilt. Die Erzieherinnen erkennen durchschnittlich die Hälfte der mathematischen Fähigkeiten von Kindern in verschiedenen komplexen Alltagssituationen bei mathematischen Tätigkeiten.

Die zweite Forschungsfrage, ob es einen Zusammenhang zwischen der Urteilsgenauigkeit und der Beobachtungskompetenz gibt, kann klar beantwortet werden. Erwartungswidrig zeigt die vorliegende Untersuchung, dass die Facetten Urteilsgenauigkeit und Beobachtungskompetenz voneinander unabhängig sind. Es gibt weder signifikante Korrelationen zwischen der globalen Urteilsgenauigkeit und der Beobachtungskompetenz noch zwischen den einzelnen Urteilsgenauigkeitskomponenten (Niveau-, Personen-, Merkmals- und Wechselwirkungskomponente) und der Beobachtungskompetenz. Damit konnte die auf dem Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz (Kap. 5) basierende Annahme nicht bestätigt werden, dass auf der einen Seite eine genaue summative Beurteilung auf einer Vielzahl kompetenter Einzelbeobachtungen beruht, und dass auf der anderen Seite häufige gezielte Einzelbeobachtungen in ihrer Summe zu einem genaueren Gesamturteil führen.

Die dritte Forschungsfrage, ob eine bereichsspezifische Fortbildung von Erzieherinnen zur Verbesserung ihrer Urteilsgenauigkeit führt, muss verneint werden.

Der Faktor Zeit hingegen (neun Monate zwischen Prä- und Posttest) beeinflusst die Fähigkeit der Erzieherinnen, Unterschiede zwischen den Kindern signifikant besser wahrzunehmen (Personenkomponente) und spezifische Aufgabenschwierigkeiten signifikant besser einzuschätzen (Merkmalskomponente). Damit verbessert sich auch die globale Urteilsgenauigkeit über die Zeit signifikant.

Auf die Einschätzung des Leistungsniveaus der Gruppe (Niveauelemente) sowie von Unterschieden zwischen den Kindern für einzelne Items (Wechselwirkungskomponente) hat auch der Faktor Zeit keinen signifikanten Einfluss.

Die vierte Forschungsfrage, ob die Unterschiede in der Veränderung der Urteilsgenauigkeit von Erzieherinnen davon abhängig sind, nach welchem Ansatz sie Kinder bereichsspezifisch fördern, lässt sich zusammenfassend verneinen. Die Art und Weise der bereichsspezifischen Förderung spielt keine Rolle für die Variabilität der Urteilsgenauigkeit.

igkeit von Erzieherinnen. Auch die Unterschiedlichkeit in den einzelnen Urteilsgenauigkeitskomponenten lässt sich nicht auf die Art der Fortbildung zurückführen.

Die fünfte Forschungsfrage, ob sich die Beobachtungskompetenz von Erzieherinnen durch eine bereichsspezifische Fortbildung von zwei Tagen beeinflussen lässt, kann aufgrund einer signifikanten Zunahme der Beobachtungskompetenz nach nur zwei Tagen bejaht werden. Damit wird die These gestützt, dass eine Fortbildung, auch wenn sie nicht explizit auf diagnostische Kompetenz, sondern allgemein auf Förderung ausgerichtet ist, zur bereichsspezifischen Förderung bedeutsam für eine gezielte und damit fokussiertere Wahrnehmung früher bereichsspezifischer Kompetenzen ist.

Die sechste Forschungsfrage zielt auf Veränderungen der Beobachtungskompetenz von Erzieherinnen in Abhängigkeit von Fortbildungen in verschiedenen bereichsspezifischen Förderansätzen ab. Für eine Zunahme der Beobachtungskompetenz, so zeigt sich, spielt der bereichsspezifische Ansatz, welcher der Fortbildung zugrunde gelegt ist, keine Rolle.

10. Diskussion

10.1 Inhaltliche Diskussion der Ergebnisse

Diagnostische Kompetenzen von Lehrpersonen sind relevant, weil eine adaptive Förderung eine vorausgehende Diagnostik benötigt. Die diagnostische Kompetenz erweist sich nach Schmidt (2020) gerade „im Zusammenhang mit formativem Assessment als besonders relevante Facette adaptiver Lehrerkompetenz“ (ebd. S. 87). Die adaptive Gestaltung von Lernumgebungen gilt nach Häcker (2017, S. 280) innerhalb der Lehr-Lernforschung als das Konzept, das wissenschaftlich am besten fundiert ist, um auf interindividuelle Unterschiede didaktisch angemessen zu reagieren (Helmke & Weinert, 1997, S. 137). Dies gilt auch für die bereichsspezifische Förderung im Kindergarten, die sich gemäß Anders (2018, S. 186) zur ersten Stufe im deutschen Bildungssystem entwickelt hat. Dort findet zwar kein Unterricht im klassischen Sinne statt, jedoch spielen gezielt hergestellte oder zufällige als Lernchance genutzte Situationen gerade im Hinblick auf die bereichsspezifische Förderung eine Rolle. Dabei wird im frühpädagogischen Kontext der freien Beobachtung von Kindern und der damit verbundenen Beobachtungskompetenz der Erzieherinnen ein sehr hoher Stellenwert zugeschrieben (Kap. 4.1).

In der vorliegenden Studie wurde davon ausgegangen, dass die Urteilsgenauigkeit im Hinblick auf die Planung von Förderprozessen und die Beobachtungskompetenz für die Adaption laufender Förderprozesse eine Rolle spielen. Wird die Unterscheidung in Mikro- und Makroadaptionen vorgenommen (Martschinke, 2015), ist die Beobachtungskompetenz besonders relevant für Adaptionen auf der Mikroebene in konkreten Situationen mit Förderpotential und die Urteilsgenauigkeit als summative Diagnostik für Adaptionen, die nicht an konkrete Situationen gebunden sind. Die Urteilsgenauigkeit ist vornehmlich für Adaptionen auf der Makroebene relevant.

10.1.1 Einordnung in den Forschungskontext

Deskriptive Ergebnisse zur Urteilsgenauigkeit

Hinsichtlich der Urteilsgenauigkeit zeigt die vorliegende Untersuchung, dass Erzieherinnen das Niveau ihrer Gruppe im Vergleich der Komponenten am besten einschätzen können (Niveauelement). Auch die Wahrnehmung von Unterschieden zwischen Kindern ihrer Gruppe gelingt ihnen ähnlich gut (Personenelement). Wesentlich schwerer fällt es ihnen, zum einen spezifische Aufgabenschwierigkeiten im mathematischen Bereich korrekt einzuschätzen (Merkmalskomponente) und zum anderen Unterschiede zwischen den Kindern im Hinblick auf einzelne Aufgaben (Wechselwirkungskomponente) zu beurteilen. In der vorliegenden Untersuchung zeigt sich zudem eine durchschnittliche Rangordnungskorrelation von $r = .69$. Diese besagt, dass Erzieherinnen ihre Kinder entsprechend ihrer Leistungen problemlos in eine Rangreihenfolge bringen können.

Die vorliegenden Ergebnisse zur Urteilsgenauigkeit stimmen mit den Ergebnissen der Metaanalyse von Hoge und Coladarci (1989) und den Befunden der daran anschließenden Metaanalyse von Südkamp et al. (2012) überein. Beide Metaanalysen kommen zu dem Ergebnis, dass es Lehrkräften im Durchschnitt gut gelingt, die Rangreihe ihrer Schüler(innen) einzuschätzen. Hoge und Coladarci (ebd.) berichten einen durchschnittlichen Korrelationskoeffizienten von $r = .66$ und Südkamp et al. (ebd.) von $r = .63$. Auch aktuellere Veröffentlichungen zum Thema liefern keine neuen Orientierungswerte zur Urteilsgenauigkeit von Leistungseinschätzungen (Schrader & Praetorius, 2018; Südkamp & Praetorius, 2017).

Im nationalen Kontext lassen sich die Ergebnisse der vorliegenden Untersuchung, welche sich im Hinblick auf die Urteilsgenauigkeit auf die einzelnen Komponenten nach Cronbach (1955) beziehen, nur mit der Studie von Kammermeyer (2000) vergleichen, in der ebenfalls eine vollständige Komponenteanalyse berechnet wird. Mit dieser stimmen die vorliegenden Ergebnisse zum Teil überein. In beiden Studien hat die Niveauelement den geringsten Anteil an der globalen Urteilsgenauigkeit und die Wechselwirkungskomponente den größten. Unterschiede zu Kammermeyer (ebd.) gibt es im Hinblick auf die Bedeutung der Personen- und Merkmalskomponente für die globale Urteilsgenauigkeit. Die Personenelement ist in der vorliegenden Arbeit mit 13% wesentlich geringer und die Merkmalskomponente mit 41% wesentlich stärker ausgeprägt als in der Studie von Kammermeyer (ebd., S. 209), wo beide Komponenten

mit 23% (Merkmalskomponente) und 25% (Personenkomponente) nahezu gleiche Anteile an der globalen Urteilsgenauigkeit besitzen. Somit gelingt den Erzieherinnen in der vorliegenden Untersuchung die Wahrnehmung von Unterschieden zwischen den Kindern wesentlich besser als dies bei Kammermeyer (ebd.) der Fall ist. Spezifische Aufgabenschwierigkeiten zu erkennen, fällt den Erzieherinnen dagegen in der vorliegenden Untersuchung schwerer als den Erzieherinnen in der Studie von Kammermeyer (ebd.).

Eine mögliche Erklärung für diesen Befund ist, dass aufgrund der Bereichsspezifität der vorliegenden Untersuchung das Merkmal (hier: numerisch-mathematische Fähigkeiten) eine größere Rolle spielt als in der Studie von Kammermeyer (ebd.), wo es um Schulfähigkeitskriterien (Denkfähigkeit, Feinmotorik, Gedächtnis, Gliederungsfähigkeit, Konzentration) geht (Kammermeyer, ebd., S. 200).

In der vorliegenden Studie gelingt Erzieherinnen sowohl die Einschätzung des Niveaus der Gruppe als auch die Wahrnehmung von Unterschieden zwischen Kindern vergleichsweise gut. Dies könnte daran liegen, dass Erzieherinnen vor allem die Kinder ihrer Gruppe miteinander vergleichen, sich also vorwiegend an der sozialen Bezugsnorm orientieren. Die Orientierung an fachlichen Kriterien in dem Sinne, dass sich eine Erzieherin fragt, welches Kind bereits simultan Mengen auf Würfelbildern bzw. im Zahlenraum bis sechs erfassen kann, ist hingegen weniger stark ausgeprägt. Diese kriteriale Bezugsnormorientierung ist vermutlich mehr mit Lehr-Lernvorstellungen im schulischen als im vorschulischen Kontext vereinbar. Für Erzieherinnen steht das Kind im Mittelpunkt ihrer Aufmerksamkeit, der Gegenstand oder der Bildungsbereich hingegen ist vermutlich – anders als bei Lehrkräften – im Vergleich dazu eher nachrangig. Wenn fachliche Lernziele weniger bedeutsam sind, ist es nicht verwunderlich, dass Erzieherinnen einzelne Aufgabenschwierigkeiten weniger gut einschätzen können.

Ein Grund dafür kann in der Qualifikation der Erzieherinnen und in ihrem Aufgabenprofil liegen, wie es z.B. in den Bildungs- und Erziehungsempfehlungen für Kindertagesstätten in Rheinland-Pfalz (Ministerium für Bildung, 2018) sehr deutlich wird. In diesen wird lediglich ein „spielerischer Umgang mit mathematischen Inhalten“ (ebd., S. 71) empfohlen – ebenso wie das Einräumen der Möglichkeit zum „Zählen, Messen, Vergleichen“ (ebd., S. 73). Es bleibt bei unverbindlichen Empfehlungen. Verbindliche Standards für den mathematischen Bereich existieren in Deutschland nicht (JMK & KMK, 2004), im Gegensatz zu den USA, wo es spezifische Curricula für Mathematik im Vorschulbereich gibt, die sich vielfach an die Standards des National Council of

Teachers of Mathematics (NTCM) anlehnen (Baroody, Clements & Sarama, 2019; Clements & Sarama, 2014).

Es kann auch an der Ausbildung liegen, dass es Erzieherinnen schwerfällt, Aufgabentypen und -schwierigkeiten einzuschätzen. Die Ausbildung zur Erzieherin an Fachschulen bietet nur wenig Lernanlässe zur Begleitung früher mathematischer Bildungsprozesse (Eichen & Bruns, 2017, S. 67). Lediglich in fünf Bundesländern gibt es Lehrpläne für die Ausbildung von Erzieherinnen mit konkreten Hinweisen zu den Themen und Inhalten des Bereichs Frühe mathematische Bildung (ebd.). Ein „Lernen von der Sache aus“ ist deshalb für viele Erzieherinnen noch ungewohnt (Erath & Rossa, 2017); zudem müssten sie damit teilweise methodisch-didaktische Routinen aufgeben und neue Lehrformen praktizieren (ebd., S. 116), was für viele Erzieherinnen eine Hürde darstellt. Es könnte also an fehlendem Fachwissen liegen, wenn im Hinblick auf fachliche Aspekte weniger exakte Einschätzungen getroffen werden.

Ein weiterer Grund für die vorliegenden Ergebnisse könnten etwaige Vorbehalte von Erzieherinnen dem Fach Mathematik gegenüber sein. Dieses Fach bereitet ihnen oftmals Schwierigkeit, bisweilen entwickeln sie sogar Ängste (Gasteiger, 2010, S. 168), weshalb der Umgang mit diesem Bildungsbereich durch ein geringes Zutrauen in die eigenen Fähigkeiten gekennzeichnet ist. Auch die Einstellungen von Erzieherinnen im Hinblick auf die Relevanz einer frühen mathematischen Förderung könnten eine Hürde sein (ebd.). Möglicherweise fokussieren Erzieherinnen auch deshalb weniger auf den mathematischen Bereich, weil sich ihr Interesse stärker auf andere Bildungsbereiche konzentriert (vor allem auf den sozial-emotionalen und den sprachlichen Bereich).

Zur Einordnung der Ergebnisse können neben der Untersuchung von Kammermeyer (2000) auch die Studien von Dollinger (2013) und Bruns (2014) herangezogen werden, die jedoch beide nur einzelne Komponenten der Komponentenanalyse nach Cronbach (1955) analysiert haben.

Mit den Ergebnissen von Dollinger (2013) ist die vorliegende Untersuchung nur im Hinblick auf die Niveau-, Differenzierungs- und Rangordnungskomponente zu vergleichen. Gemessen an der Niveauebene von Schrader (1989) belegt die vorliegende Studie die Tendenz der Erzieherinnen zur Überschätzung des Gruppen-Niveaus, wobei die Spannweite sehr groß ist. Im Widerspruch dazu stellt Dollinger (2013) für den mathematischen Bereich eine überwiegende Unterschätzung der kindlichen Leistungen fest (ebd., S. 121). Hinsichtlich der Differenzierungskomponente nach Schrader (1989) zeigt

die vorliegende Untersuchung, dass die Heterogenität der Kinderleistungen überschätzt wird. Auch dies steht im Widerspruch zu den Ergebnissen von Dollinger (2013), die hinsichtlich der Differenzierungskomponente eine signifikante Unterschätzung der Leistungsstreuung berichtet (ebd., S. 121). In der Rangordnungskomponente zeigt sich die Kompetenz der Erzieherin, ihre Bezugskinder entsprechend ihrer Leistungen in eine Rangreihenfolge zu bringen. Diese liegt in der aktuellen Untersuchung bei einer mittleren Korrelation von $r = .69$, während Dollinger eine Rangkorrelation von $r = .36$ (ebd.) berichtet. Damit zeigen die Erzieherinnen in der vorliegenden Studie eine größere Kompetenz im Hinblick auf das Wahrnehmen einer Rangreihe. Die fehlenden Übereinstimmungen der Befunde können mit hoher Wahrscheinlichkeit auf methodische Limitationen der Studie von Dollinger (2013) zurückgeführt werden, die in Kapitel 3.4 ausgeführt sind.

Der Rangkorrelation liegt in der Studie von Bruns (2014) bei $r = .763$ (Spannweite $r = -1.00$ und $\rho = 1.00$) und damit etwas höher als in der vorliegenden Studie. Trotz der sehr unterschiedlichen Qualität der Erfassung ist dieser Wert mit dem oben bereits beschriebenen Korrelationskoeffizienten der Personenkomponente in der vorliegenden Studie vergleichbar, welcher bei einer mittleren Korrelation von $r = .69$ mit einer Spannweite von $r = -.91$ und $r = .96$ liegt.

Bei diesem Vergleich ist jedoch zu berücksichtigen, dass die Studie von Bruns (2014) wegen ihrer starken Vereinfachung der diagnostischen Kompetenz von Erzieherinnen basierend auf Helmkes Niveau-, Streuungs- und Rangordnungskomponente (Bruns, 2014, S. 41) nur sehr begrenzt vergleichbar ist mit den vorliegenden Daten. Zwar vergleicht auch sie Testergebnisse von Kindern mit Einschätzungen des frühpädagogischen Fachpersonals, jedoch werden diesen nur sieben Leistungsstufen vorgelegt und keine einzelnen Testaufgaben (ebd., S. 124).

Obwohl in den Studien von Dollinger (2013) und Bruns (2014) die diagnostische Kompetenz sehr unterschiedlich gemessen wurde, kommen beide zu vergleichbaren Rangordnungskomponenten.

Zur Veränderung der Urteilsgenauigkeit durch bereichsspezifische Fortbildungen liegen keine Studien im vorschulischen Bereich vor, in welche die vorliegenden Ergebnisse eingeordnet werden könnten.

Zusammenfassend kann festgehalten werden, dass die Ergebnisse der vorliegenden Untersuchung zur Urteilsgenauigkeit von Erzieherinnen im Bildungsbereich Mathematik

mit nationalen und internationalen Befunden übereinstimmen. Insbesondere die Passung zu den Ergebnissen der einzigen Studie mit Erzieherinnenbeteiligung (Kammermeyer, 2000), welche die Komponenten nach Cronbach (1955) untersucht, ist herauszustellen.

Deskriptive Ergebnisse zur Beobachtungskompetenz

Bezüge zur Beobachtungskompetenz, wie sie in der vorliegenden Studie verstanden wird, finden sich im Kontext der Lehrerbildung. Das Ziel, die Entwicklung der Beobachtungskompetenz von Lehramtsstudierenden zu fördern als „Basis einer fragenentwickelnden und forschenden Auseinandersetzung mit unterrichtspraktischen Situationen“, wird von Oesterhelt (2018, S. 83) verfolgt. Übereinstimmend mit der vorliegenden Studie wählt auch Oesterhelt in Abgrenzung von Alltagsbeobachtungen einen bereichsspezifisch gezielten Fokus. Die Autorin beschreibt im Zusammenhang mit der Ausbildung von Lehramtsstudierenden, dass sich dieser ‚fokussierte Blick‘ z.B. auf eine schüler(innen)bezogene Bestimmung von Lernausgangslagen, auf die Lernstandserfassung also richten kann.

Die Beobachtungskompetenz wurde bisher erst in wenigen Studien untersucht. Eine begriffliche Annäherung findet sich bei Wildgruber im Zusammenhang mit Ergebnissen seiner Untersuchung der Beobachtungspraxis von 32 Erzieherinnen in einem qualitativ-interpretativen Verfahren (Wildgruber, 2010, S. III). Er untersucht neun Einzelfälle und erkennt zwei zentrale Orientierungen, welche sich auf den gesamten diagnostischen Prozess auswirken: Identifiziert werden erstens Erzieherinnen, die sich idealtypisch am ‚Kind‘ orientieren und zweitens Erzieherinnen, die sich an ‚Kriterien‘ orientieren, welche außerhalb des Kindes liegen (ebd., S. 268). Dass seine Fälle jedoch weder eindeutig der einen noch der anderen Orientierung zuzuordnen sind, führt zur Definition einer dritten Orientierung, welche letztlich eine Mischform darstellt und von ihm als ambivalente Orientierung bezeichnet wird (ebd., S. 269). Eine quantitative Verteilung der untersuchten Fälle auf die verschiedenen Orientierungen wird nicht berichtet. Parallelen zur vorliegenden Studie finden sich insbesondere in der Konzeptualisierung von Beobachtungskompetenz, wobei Wildgruber den Ansatz hier als Forschungsdesideratum aufführt. Er sieht die vergleichende Untersuchung von Urteilsgenauigkeit und Beobachtungskompetenz pädagogischer Fachkräfte anhand von Videovignetten als eine Möglichkeit zur Erhöhung der Anschlussfähigkeit an die bisherige Forschung im schulischen Bereich und als einen weiteren Schritt zur Untersuchung der Kompetenzen ausgezeichneter Fachkräfte (ebd., S. 282).

Zur Einordnung der Ergebnisse zur Beobachtungskompetenz können auch Studien zu formativem Assessment herangezogen werden. Hierzu gibt es jedoch im deutschen Sprachraum nur wenige Studien, die sich auf Erzieherinnen beziehen. Wildgruber (2010) zeigt die Forschungslücke auf, die bisher noch nicht geschlossen wurde: „Die Erforschung von Kompetenzen zur Beobachtung und Dokumentation, jenseits von Erfahrungen mit einem einzelnen Beobachtungsverfahren, wurde jedoch bisher nicht fokussiert“ (Wildgruber, 2010, S. 279).

Auch aus der Forschung mit Lehrkräften gibt es nur wenige Studien, in welche die vorliegenden Ergebnisse eingeordnet werden könnten, unabhängig davon, ob diese auf Erzieherinnen übertragbar sind. Auch über die Praxis bezüglich formativen Assessments in deutschen Schulen ist bislang nur wenig bekannt (Schmidt, 2020). Bisherige Befunde lassen zwar auf eine sehr heterogene Praxis schließen (Inckemann, 2008; Racherbäumler, 2009; Solzbacher, 2012), einige Befunde belegen jedoch die Bedeutsamkeit informeller situierter Diagnostik (z.B. prozessbezogene Beobachtung) für die aktuelle Schulpraxis (Eckerth, 2013; Smit & Engeli, 2017; Solzbacher, 2012).

Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz

Theoretische Modelle zur diagnostischen Kompetenz fokussieren entweder die Genauigkeit von Urteilen oder prozessbegleitende Diagnostik (Herppich et al., 2017, S. 89). Ein sowohl auf Urteilsgenauigkeit als auch auf Beobachtungskompetenz anwendbares Modell stellt das *Arbeitsmodell des Netzwerks zur diagnostischen Kompetenz von Lehrkräften* (NeDiKo) (Kap. 2.6.5) von Herppich et al. (ebd., S. 75) dar. Dort lassen sich Urteilsgenauigkeit und Beobachtungskompetenz jedoch nur einzeln verorten, Zusammenhänge und Wechselwirkungen beider Facetten werden im NeDiKo-Modell nicht abgebildet. Um auch diese miteinzubeziehen, wurde in der vorliegenden Arbeit ein Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz (Kap. 5) entwickelt, das im Gegensatz zum NeDiKo-Modell sowohl den zeitlichen Verlauf als auch den Zusammenhang der beiden Facetten berücksichtigt sowie Handlungsentscheidungen ins Modell aufnimmt.

Im neuen Prozessmodell (Kap. 5) wurde gemäß dem erweiterten Linsenmodell (Kap. 4.6.1) angenommen, dass im diagnostischen Prozess im ersten Schritt die Beobachtungskompetenz eine Rolle spielt: Erst aufgrund von mehreren prozessbegleitenden Beobachtungen (Beobachtungskompetenz) ist die Lehrperson nach einer gewissen Zeit

im zweiten Schritt in der Lage, ein summatives Urteil zu fällen; sie kann den Leistungsstand eines Kindes mehr oder weniger genau einschätzen (Urteilsgenauigkeit).

Trotz hoher Plausibilität eines Zusammenhangs von Urteilsgenauigkeit und Beobachtungskompetenz konnte der im neuen Prozessmodell dargestellte Zusammenhang jedoch nicht nachgewiesen werden. Möglicherweise lässt sich das Modell unter den Bedingungen der vorliegenden Untersuchung im vorschulischen Bereich nicht bestätigen, besitzt aber im schulischen Kontext Gültigkeit. Lernzeit und Zuordnung zu Erzieherinnen sind in der Kindertagesstätte weniger eindeutig als Lernzeit und Zuordnung zu Lehrpersonen in der Schule. Über die Erfassung der Kontaktzeit wurde dies in der vorliegenden Studie zwar nach Möglichkeit kontrolliert, jedoch haben Kinder in offenen Einrichtungen zu sehr vielen Erzieherinnen Kontakt und sind weniger an eine Lehrkraft gebunden, als dies in der Grundschule der Fall ist. Ob die Kontaktzeit, die nur ein quantitatives Maß ist und die Qualität des Kontakts in Fördersituationen unberücksichtigt lässt, die beste Zuordnung von Kindern und Erzieherinnen darstellt, muss kritisch hinterfragt werden. Eventuell hat eine Erzieherin, die mit einem Kind weniger Zeit verbracht hat, den größeren Einfluss auf das Kind im Hinblick auf die bereichsspezifische Förderung. Im vorschulischen Bereich gibt es zudem weniger geplante instruktionsorientierte Lernphasen, sondern mehr informelle Lernphasen oder Situationen, die sich im Freispiel ergeben. Diese können diagnostisch relevant sein und von verschiedenen Erzieherinnen unterschiedlich genutzt werden, wie im erweiterten Linsenmodell (4.6.1) dargestellt. Auch das frühpädagogische Angebots-Nutzungs-Modell (Kap. 2.6.4) zeigt, wie vielen Einflüssen Lehrprozesse unterworfen sind und weshalb es sein kann, dass die Erfassung der Urteilsgenauigkeit, die mittels Vergleich von Erzieherinneneinschätzung und Kinderleistung erhoben wurde, in einem so offenen didaktischen Setting wie dem im Kindergarten vorzufindenden fehlerbehaftet sein kann. Auch damit wäre eine fehlende Korrelation mit der mittels Videovignetten erhobenen Beobachtungskompetenz zu erklären.

Dennoch muss vorläufig angenommen werden, dass es sich bei der Urteilsgenauigkeit und der Beobachtungskompetenz um voneinander unabhängige Facetten handelt. Damit kann das in dieser Arbeit vorgeschlagene *Prozessmodell zum Zusammenhang von Urteilsgenauigkeit und Beobachtungskompetenz* in der vorliegenden Form unter den Bedingungen der vorliegenden Untersuchung nicht bestätigt werden. Hierzu sind auch die methodischen Limitationen der Studie zu berücksichtigen (Kap. 10.2.4).

Letztlich stärken die Ergebnisse das NeDiKo-Modell (*Arbeitsmodell des Netzwerks zur diagnostischen Kompetenz von Lehrkräften*) von Herppich et al. (2017, S. 75) und führen dazu, dass Urteilsgenauigkeit und Beobachtungskompetenz als voneinander unabhängig und damit getrennt betrachtet werden müssen.

10.1.2 Was bedeutet die Unabhängigkeit der Facetten *Urteilsgenauigkeit* und *Beobachtungskompetenz*?

Die ursprüngliche Annahme bestand darin, dass die Ausprägung spezifischer Kompetenzen einzelner Kinder durch wiederholte Beobachtungen seitens ihrer Erzieherinnen in Abhängigkeit von der Häufigkeit der Beobachtungen immer genauer beurteilt wird. Die in der vorliegenden Untersuchung herausgestellte Unabhängigkeit der diagnostischen Kompetenzen Urteilsgenauigkeit und Beobachtungskompetenz zeigt jedoch, dass der Blick der Erzieherin für Details hinsichtlich der Leistungen eines Kindes und der Blick auf die Komplexität kindlicher Fähigkeiten in konkreten Fördersituationen voneinander unabhängig sind. Beide diagnostische Kompetenzen zielen jedoch auf eine Optimierung adaptiver Bildungsprozesse ab. Beobachtungskompetenz und Urteilsgenauigkeit können als Kompetenzen oder unabhängige Teilfähigkeiten des übergeordneten Konzepts der diagnostischen Expertise gesehen werden. Diagnostische Expertise ist ein breit gefasstes Konzept, unter das Helmke (2017) eine Vielzahl inhaltlich abgegrenzter Konstrukte zusammenfasst, die letztlich einem gemeinsamen Ziel verpflichtet sind – der Herstellung einer Passung zwischen Lernvoraussetzungen und Anforderungen. Er subsumiert unter diesem Begriff letztlich zahlreiche diagnostische Leistungen (Helmke, 2017, S. 121). Die diagnostische Expertise (ebd., S. 119) als umfassendes Konzept ermöglicht die Zusammenführung der beiden Forschungslinien zu Urteilsgenauigkeit und formativem Assessment.

Bereits der Begriff diagnostische Kompetenz als Urteilsgenauigkeit kann darüber hinaus als Sammelbegriff unabhängiger Teilfähigkeiten verstanden werden. Spinath (2005) zieht nach Untersuchung verschiedener Komponenten der Akkuratheit und aufgrund deren Unabhängigkeit sogar den Schluss, „dass der Begriff der diagnostischen Kompetenz, soweit damit die Fähigkeit zur treffenden Personenbeurteilung gemeint ist, vermieden werden sollte“ (ebd., S.85). Dies hat sich bis dato allerdings nicht durchgesetzt.

Mit der Distanzierung vom Konstrukt diagnostische Kompetenz und der Einordnung unter dem breit gefassten Konzept der diagnostischen Expertise (Helmke, 2017) ist ein

theoretischer Rahmen für die unabhängigen diagnostischen Kompetenzen Urteilsgenauigkeit und Beobachtungskompetenz gegeben.

10.1.3 Welche diagnostischen Kompetenzen benötigen Erzieherinnen?

Angelehnt an Helmkes Begriff der diagnostischen Expertise (2017, S. 119–141) nähert sich die vorliegende Arbeit mit der Erweiterung des Konstrukts diagnostische Kompetenz als reine Urteilsgenauigkeit um die Beobachtungskompetenz einem breiteren diagnostischen Verständnis. Gerade im vorschulischen Bereich und zu Anfang der Grundschulzeit spielt die Beobachtungskompetenz im Vergleich zur Urteilsgenauigkeit die größere Rolle (Kap. 3.7 und 4.8), weil hier weniger die Genauigkeit der Diagnose einzelner Fähigkeiten als vielmehr die Wahrnehmung der momentanen Fähigkeiten zum Setzen weiterer Arbeitsimpulse oder zur Auswahl passender Medien von Bedeutung sein könnte. Bereits 1989 vermutete Schrader, dass Lehrende mit hoher Urteilsgenauigkeit zwar Vorteile hinsichtlich der Erstellung detaillierter Planungen haben könnten, im Hinblick auf die notwendige Flexibilität des unterrichtlichen Handelns jedoch seien Nachteile wahrscheinlich.

Ein weiterer zentraler Aspekt, der im Rahmen der Arbeit an der Erweiterung des Linsenmodells (Kap. 4.6.1) besonders deutlich wurde, ist die Notwendigkeit, der (Förder-)Situation als Element des diagnostischen Prozesses eine bedeutende Rolle zuzusprechen – im Sinne eines “noticing as teacher expertise” (van Es & Sherin, 2002, 2006). Zwar wird mit der Beobachtungskompetenz das Erkennen kindlicher Fähigkeiten im laufenden Prozess kindlicher Förderung in diagnostisch relevanten Situationen erfasst, jedoch ist diese Kompetenz von einer weiteren entscheidenden Fähigkeit abhängig: Die Fähigkeit Lehrender, diagnostisch relevante Situationen zu erkennen, diese gegebenenfalls herauszufordern oder zu planen und entsprechend zu nutzen. Dies ist ein Forschungsdesiderat. Wenn Situationen wie z.B. Kinder beim Brettspiel mit Würfeln nicht als mathematisch relevant erkannt werden, bleiben diese Situationen ungenutzt.

Es existieren geeignetere und weniger geeignete Situationen, um etwaige Fähigkeiten zu beobachten, weil je nach Situation nur ein Teil der kindlichen Fähigkeiten sichtbar wird, wie das erweiterte Linsenmodell (Kap. 4.6.1) zeigt. Die Beachtung der Situation als ein wichtiger Schlüssel zur weiteren Erfassung von diagnostischer Kompetenz betonen auch Praetorius et al. (2017, S. 97), wobei sie die Bedeutung einer repräsentativen Auswahl relevanter und im Bereich der Forschung zu diagnostischer Kompetenz bislang

nicht untersuchter diagnostischer Situationen als bedeutsam erachten. In der vorliegenden Arbeit wird die Erfassung der Fähigkeit zum gezielten Herausfordern von diagnostisch bedeutsamen Situationen als weitere Facette diagnostischer Kompetenz oder diagnostischer Expertise gesehen und hiermit als weiteres Forschungsdesiderat benannt.

Wie die vorliegende Untersuchung zeigt, ist nicht davon auszugehen, dass Lehrpersonen, die in einer Fördersituation eine Vielzahl kindlicher Fähigkeiten gut erfassen können – also eine große Beobachtungskompetenz besitzen – auch genau auf die Ausprägung einzelner kindlicher Fähigkeiten achten können, also über eine große Urteilsgenauigkeit verfügen. Es schließt sich die Frage an, ob eine Erzieherin in der täglichen praktischen Zusammenarbeit mit Kindern eher die Kompetenz zur Einschätzung mehrerer Fähigkeiten benötigt oder eher diejenige zur genauen Einschätzung einzelner Fähigkeiten.

Im Kita-Alltag mit einer Vielzahl von Kindern scheint es für die Erzieherin entweder schwierig oder irrelevant zu sein, Erkenntnisse aus Beobachtungen dauerhaft mit einem Kompetenzniveau einzelner Kinder zu verknüpfen.

Im Hinblick auf die Relevanz der durchgeführten Beobachtungen wird bereichsspezifische Förderung im Vergleich zu anderen parallel ablaufenden Prozessen möglicherweise als weniger wichtig wahrgenommen. Evtl. spielt es für eine Erzieherin auch keine Rolle, wo das Kind ‚genau‘ steht, weil sie in ihren methodisch-didaktischen Fähigkeiten limitiert ist und sich die mathematischen Anteile ihrer bereichsspezifischen Fördermaßnahmen auf die Durchführung einiger Spiele oder das Abarbeiten von Stundenbildern beschränkt (vgl. Kap. 7.2.2), unabhängig vom Kenntnisstand des einzelnen Kindes z.B. mit der gesamten Gruppe der Vorschulkinder.

Auf der einen Seite sind diagnostische Kompetenzen notwendig, um in unbekanntem Situationen mit einer lösungsorientierten Flexibilität adäquat pädagogisch zu handeln, also situativ professionell zu reagieren (Ruppin et al., 2015, S. 141). Deshalb geht eine Vielzahl von theoretischen Ansätzen und Studien von der Auswirkung der diagnostischen Kompetenz auf diejenige zur Gestaltung adaptiver Lernumgebungen (Südkamp & Praetorius, 2017).

Auf der anderen Seite existieren jedoch auch vereinzelte Belege dafür, dass Personen, die weniger genaue Urteile treffen, die Leistung von Kindern erhöhen. Dies wurde allerdings nur in Abhängigkeit von Unterrichtsmerkmalen (Strukturierung, individuelle fachliche Hilfestellung, Aufgabenstellungen) nachgewiesen (Schrader, 1989, S. 196). Es

ist deshalb auch denkbar, dass die diagnostische Kompetenz für Erzieherinnen eher nachrangig und die methodisch-didaktische Kompetenz für die adaptive Gestaltung von Lernumgebungen entscheidend ist. Nachrangig könnte in diesem Zusammenhang bedeuten, dass Erzieherinnen keinen Grund zum Diagnostizieren haben, weil ihr methodisch-didaktisches Repertoire für ein adaptives bereichsspezifisches Förderangebot nicht ausreicht.

Die Limitation der diagnostischen Kompetenz kann folglich in der methodisch-didaktischen Kompetenz der Erzieherin liegen. Das Repertoire an Handlungsentscheidungen kann begrenzt sein, sodass Urteilen und Beobachtungen keine Mikroadaptationen der Lernumgebung folgen. Dies würde die untrennbare Verbindung von Diagnostik und Handlungsentscheidung unterstreichen und verdeutlichen, wie wichtig die Ausbildung von Erzieherinnen im Hinblick auf die Gestaltung adaptiver bereichsspezifischer Fördermaßnahmen ist. Die Limitation der diagnostischen Kompetenz kann aber auch darin begründet sein, dass Erzieherinnen eher den Blick auf die Förderung der sozialen Kompetenzen legen und weniger auf den Sachgegenstand, auf die Förderung einzelner Fachkompetenzen, weil dies im Rahmen der Ausbildung keinen wichtigen Stellenwert hat, wie z.B. die Lernfelder des länderübergreifenden Lehrplans zur Ausbildung von Erzieherinnen (König et al., 2018, S. 34) und deren Ausgestaltung zeigen.

10.1.4 Veränderung von *Urteilsgenauigkeit* und *Beobachtungskompetenz*

Die vorliegende Studie liefert Hinweise auf die Verbesserbarkeit der diagnostischen Kompetenzen Urteilsgenauigkeit und Beobachtungskompetenz durch Fortbildungen. Diese zunächst trivial anmutende Erkenntnis gewinnt an Bedeutung, wenn zum einen beachtet wird, dass die Fortbildungen keinen expliziten Fokus auf diagnostische Kompetenzen legten und zum anderen lediglich zwei Tage dauerten.

Für die Diskussion der Ergebnisse zur Veränderung von Urteilsgenauigkeit und Beobachtungskompetenz sind Unterschiede in der Erfassung der diagnostischen Kompetenzen von Bedeutung, die Auswirkungen auf die Posttestzeitpunkte hatten. Deshalb werden diese primär erläutert: Die Verbesserung der Urteilsgenauigkeit von Erzieherinnen wurde durch ein ökologisch valides Treatment (zweitägige Fortbildung) zur Gestaltung bereichsspezifischer Förderprozesse und die sich daran anschließende Umsetzung des Gelernten im Rahmen der neunmonatigen Förderung von Kindern erreicht. Der Posttest zur Urteilsgenauigkeit erfolgte also nicht direkt nach der Fortbildung, sondern

erst am Ende des Kindergartenjahres. Die Beobachtungskompetenz, welche sich durch die Teilnahme an einer zweitägigen Fortbildung verbesserte, wurde im Anschluss an die Fortbildung innerhalb von zwei Wochen erhoben. Diese unterschiedlichen Zeitpunkte des Posttests sind auf die verschiedenen Erhebungsverfahren zurückzuführen. Weil die Beobachtungskompetenz mittels Videovignetten erhoben wurde, konnte der Prätest direkt nach der Fortbildung erfolgen. Zur Erfassung der Urteilsgenauigkeit war ein Abgleich mit Daten von Bezugskindern aus der Praxis der Erzieherinnen notwendig, weshalb der Prätest nicht direkt nach der Fortbildung stattfinden konnte. Die Erzieherinnen lernen nämlich ihre BezugsKinder durch die Fortbildung nicht besser kennen, sondern müssen dazu eine gewisse Förderzeit mit ihnen verbracht haben. Aus diesem Grund erfolgte der Posttest zur Urteilsgenauigkeit erst am Ende des Projektzeitraums.

Einfluss von Fortbildung zu unterschiedlichen Förderansätzen

Die Verbesserung der diagnostischen Kompetenz hängt nicht von der Art der Förderansätze (Lehrgangs- vs. Lernwegsbasierung) ab.

Die Teilnahme an Fortbildungen mit verschiedenen Förderansätzen und die anschließende Durchführung von mathematischen Fördermaßnahmen über einen Zeitraum von neun Monaten führte zu einer Verbesserung der diagnostischen Kompetenz im Hinblick auf die Urteilsgenauigkeitskomponenten. Die Beobachtungskompetenz verbesserte sich lediglich durch die Teilnahme an der zweitägigen Fortbildung zu beiden Förderansätzen.

Dieses Ergebnis überrascht, da von einer Verbesserung der diagnostischen Kompetenz eher durch die Auseinandersetzung mit den individuellen Lernwegen der Kinder in der Fortbildung als mit der Struktur der Sache bei einer Fortbildung zu lehrgangsorientierter Förderung ausgegangen wurde.

Im vorliegenden Projekt lässt sich diese Unabhängigkeit von der Art der Fortbildung dadurch erklären, dass sich die beiden Ansätze in ihrer konkreten Umsetzung durch die Vertreter der Ansätze als weniger unterschiedlich erwiesen haben als es die Literatur hätte erwarten lassen (Donie et al. 2013); dieser Umstand war nicht zu beeinflussen. Möglicherweise ist auch das ökologisch valide Treatment (zweitägige Fortbildung) eine Ursache für die vom Fortbildungstypus unabhängige Verbesserung der diagnostischen Kompetenz. Die Erzieherinnen wurden zwar mit dem Thema konfrontiert; ihre Ausei-

nersetzung damit war aber nicht derart differenziert, als dass sich minimale Unterschiede in der konzeptuellen Gestaltung der Fortbildung hätten auswirken können.

Gemeinsamkeiten gibt es mit den ersten Befunden zu den Auswirkungen von Fortbildungen zu aktionsbezogenen und reflexiven Kompetenzen aus dem Projekt WILMA (Dunekacke et al., 2018; Pfennigwerth et al., 2016), in dem die Entwicklung beider Kompetenzen sich auch unabhängig von den zwei Fortbildungsschwerpunkten zu den untersuchten Kompetenzen gezeigt hat. Zur Erfassung der aktionsbezogenen Kompetenz wurden Erzieherinnen, ähnlich wie in der vorliegenden Untersuchung, Videovignetten mit Kindern in bereichsspezifischen situativ-authentischen Kontexten vorgespielt. Erste Befunde aus WILMA weisen eine Steigerung der aktionsbezogenen Kompetenz sowohl durch eine diese fördernde Fortbildung als auch durch eine Fortbildung zur Förderung der reflexiven Kompetenz in ähnlichem Umfang nach (Dunekacke et al., 2018; Pfennigwerth et al., 2016; Lindmeier et al., 2020).

Zu berücksichtigen sind jedoch die methodischen Unterschiede im Vergleich zur vorliegenden Studie: Im Projekt WILMA wurden die Äußerungen der Erzieherinnen zur Erfassung der aktionsbezogenen Kompetenz unter Speedbedingungen mittels Audioaufzeichnung festgehalten, was die ökologische Validität erhöhte. Im Vergleich dazu erfolgte im vorliegenden Projekt die Erhebung der Beobachtungskompetenz in schriftlicher Form.

Indem die Erzieherinnen dazu aufgefordert wurden, auf bestimmte bereichsspezifische Fähigkeiten zu achten, konnte ihre kriterienorientierte Sicht der Beobachtungskompetenz herausgefordert werden.

Bestärkt durch Befunde, die zeigen, dass auch ein explizites Training unterschiedlicher professioneller Kompetenzen keine Unterschiede im Kompetenzzuwachs nach sich zieht (Dunekacke et al., 2018), könnte die adaptive Gestaltung bereichsspezifischer Lernumgebungen als Fortbildungsinhalt in den Mittelpunkt der Betrachtungen von Einflüssen zur Verbesserung diagnostischer Kompetenz rücken. Zukünftige Studien müssen zeigen, ob möglicherweise eine indirekte Förderung von diagnostischer Kompetenz zielführender sein kann als eine explizite Fortbildung oder Training.

Dies könnte für die Fort- und Ausbildung von Erzieherinnen bedeuten, dass in der Fortbildung nicht nur der Weg zur Förderung diagnostischer Kompetenzen als Voraussetzung für adaptive Lehrprozesse beschritten werden kann; vielmehr kann es auch sinnvoll sein, über die gezielte Gestaltung von Fördermaßnahmen indirekt diagnostische

Kompetenz zu stärken. Bereits jetzt ist eine Vielzahl von Fortbildungen so gestaltet, dass Erzieherinnen insbesondere im Hinblick auf die Gestaltung von Lernumgebungen und Handlungsoptionen zur Förderung geschult werden und diagnostische Kompetenz nicht explizit trainiert wird. Diesbezüglich hat die vorliegende Studie die positive Auswirkung einer zweitägigen Fortbildung zur bereichsspezifischen Förderung auf die Urteilsgenauigkeit und Beobachtungskompetenz bei Novizen gezeigt, die keine Erfahrung mit bereichsspezifischer Förderung hatten. Gerade im Hinblick auf den kurzen Interventionszeitraum von zwei Tagen sind diese Befunde für die Praxis bedeutsam.

10.1.5 Gehören Handlungsentscheidungen zur diagnostischen Kompetenz?

Ein zentrales Ergebnis der vorliegenden Studie ist, dass durch eine Fortbildung zur Verbesserung der methodisch-didaktischen Kompetenz zur Gestaltung adaptiver Fördermaßnahmen pädagogischer Fachkräfte indirekt und ohne explizite Förderung auch deren diagnostische Kompetenz verbessert wurde.

Eine Verbesserung der methodisch-didaktischen Kompetenz im Rahmen einer Fortbildung, in der es praxisnah um Fördermöglichkeiten, um Methoden, Medien und Materialien geht, sowie der diagnostischen Kompetenz, wenn diese explizit Thema der Fortbildung ist, kann darum erwartet werden. Insofern ist das Ergebnis der Studie, das sich sowohl auf die Urteilsgenauigkeit als auch auf die Beobachtungskompetenz bezieht, durchaus überraschend und wirft die Frage auf, wie Handlungsentscheidungen und diagnostische Kompetenz zusammenhängen, ob zur diagnostischen Kompetenz auch Handlungsentscheidungen gehören.

Die Position einer notwendigen Trennung von diagnostischen Kompetenzen und Handlungsentscheidungen, konkretisiert als Handlungskompetenz bzw. methodisch-didaktische Kompetenzen, um Entscheidungen im Förderprozess zu treffen, vertreten Praetorius et al. (2017, S. 99). Sie beschreiben diagnostische und pädagogische Handlungen zwar als eng verwoben und betonen, dass diese Verschränkung für die adaptive Gestaltung von Lernumgebungen zielführend sein kann (ebd., S. 98–99) und bezeichnen die Trennung deshalb als künstlich, betonen aber die Vorteile einer klaren konzeptuellen Trennung (ebd., S. 96). Dementsprechend werden im NeDiKo-Modell (vgl. Kap. 2.6.5) Diagnose und Förderung bewusst getrennt: einzelne Prozesssequenzen enden jeweils mit Handlungsentscheidungen, welche ausdrücklich nicht als Teil des diagnostischen Prozesses verstanden werden (Herppich et al., 2017, S. 81). Die Handlungsentscheidung

ist in diesem Modell folglich dem Diagnoseprozess nachgeschaltet und nicht mehr Teil des Modells. Damit zeigt sich die Verwobenheit von diagnostischer Kompetenz und Handlungsentscheidungen im Prozessmodell von NeDiKo nur teilweise, obwohl das Modell das Ziel verfolgt, auch prozessbezogene Ansätze abzubilden (ebd. S. 76).

In der Forschung zum formativen Assessment wird jedoch hinsichtlich der Trennung von Diagnostik und Handlungsentscheidungen eine andere Position vertreten. Dort gibt es keine isolierten Untersuchungen zur Urteilsgenauigkeit, weil hier Diagnostik immer im Gesamtzusammenhang von Förderprozessen gesehen werden muss, wie Schütze et al. (2018, S. 701) betonen. Die Kompetenz zu formativem Assessment darf nach ihnen nicht auf ein Merkmal wie z.B. die Leistungsdiagnostik reduziert werden, sondern muss auch Handlungsentscheidungen, wie z.B. das Geben von Feedback, miteinbeziehen.

Zwar kann es auch Handlungsentscheidungen ohne weiteren diagnostischen Zweck geben, z.B. zur Selektion bei der Einteilung von Gruppen oder Schullaufbahnpfehlungen. In diesem Fall handelt es sich gemäß Ingenkamp und Lissmann (2008) jedoch nicht mehr um pädagogische Diagnostik, weil sie diese erst als solche anerkennen, wenn sie der Optimierung individueller Lernprozesse dient (ebd., S. 13). Damit definieren auch Ingenkamp und Lissmann den Prozess als immanenten Anteil pädagogischer Diagnostik.

Auch Schrader (2017, S. 252) betont die Notwendigkeit zur Berücksichtigung der Verwobenheit von diagnostischen Urteilen und pädagogischen Handlungen bei Mikroadaptation und weist auf die enge Koppelung von diagnostischen und instruktionalen Handlungen hin. Diese sind seiner Einschätzung nach nicht trennbar, was dementsprechend auch auf die jeweiligen zugrundeliegenden Kompetenzen zu beziehen ist.

Die von (Praetorius et al., 2017, S. 95) geforderte konzeptuelle und empirische Trennung von diagnostischem und pädagogischem Handeln sehen sowohl Ingenkamp und Lissmann (2008) als auch Schrader (1989) kritisch. Als Anmerkung zur Weiterentwicklung des Konstrukts diagnostische Kompetenz von Lehrkräften weist Schrader (2017) im Hinblick auf eine situierte Diagnostik auch im schulischen Kontext auf die kaum mögliche Trennung von diagnostischen und instruktionalen Handlungen hin. Ein erweitertes Verständnis von Diagnostik, das dem Kontext der Forschungen zum formativen Assessment entstammt und eine Begrenzung auf rein diagnostische Merkmale als zu eng sieht, findet sich sowohl bei Schütze et al. (2018, S. 701) als auch im Konzept der diagnostischen Expertise nach Helmke (2017) (vgl. Kap. 2.4).

Die vorliegende Studie stützt die formative Position, weil eine Fortbildung der methodisch-didaktischen Kompetenz zur Gestaltung adaptiver Fördermaßnahmen indirekt und ohne explizite Förderung die diagnostische Kompetenz – sowohl Urteilsgenauigkeit als auch Beobachtungskompetenz – verbessert hat und damit die Verwobenheit von Diagnostik und Handlungsentscheidungen nahelegt. Zukünftige Studien müssen zeigen, ob Handlungskompetenz und diagnostische Kompetenz möglicherweise untrennbare Bestandteile eines übergeordneten Konstrukts sind.

10.2 Methodische Diskussion

10.2.1 Erfassung von Urteilsgenauigkeit und Beobachtungskompetenz

Die vorliegende Untersuchung ist die erste Arbeit im deutschsprachigen Raum, die zum einen die Messung der Urteilsgenauigkeit unter vollumfänglicher Erfüllung des folgenden Anforderungsprofils in Anlehnung an Hoge und Coladarci (1989) (vgl. Kap. 3.3) durchgeführt hat:

- Operationalisierung der Urteilsgenauigkeit mittels direkter Beurteilungen und Testkenntnis
- Verwendung von Verfahren höchster Spezifität
- Berücksichtigung peer-unabhängiger Urteile
- Ermittlung auf Itemebene
- Berücksichtigung der hierarchischen Struktur (Bezugsgruppen)
- Domänenspezifische Erfassung,

und die zum anderen die Mindestanzahl von vier Einschätzungen pro Erzieherin veranschlagt und die Bezugskindseinschätzungen anhand der größten Kontaktzeiten zugeteilt hat. Positiv zu werten ist die valide Erfassung der Urteilsgenauigkeit, welche sich durch sieben Aspekte numerisch-mathematischer Fähigkeiten mit 120 Items zeigt.

Die Beobachtungskompetenz wurde anhand standardisierter Videovignetten erhoben, die den Erzieherinnen unbekannte Kinder zeigten.

Beide Verfahren weisen besondere Stärken und Herausforderungen auf. Auf der einen Seite könnte es relevant sein, dass Erzieherinnen zur Erfassung der Beobachtungskom-

petenz am Laptop in einer Laborsituation ihnen fremde Kinder hinsichtlich deren mathematischer Fähigkeiten und Fertigkeiten einschätzen. Zur Erfassung der Urteilsgenauigkeit hingegen mussten die Erzieherinnen auf Basis von Erinnerungen und Eindrücken aus ihrem persönlichen pädagogischen Alltag mit ihnen vertrauten Bezugskindern Urteile fällen. Auf der anderen Seite kann es eine Rolle spielen, dass die Beobachtungskompetenz anhand vertrauter Situationen erhoben wurde (z.B. Kinder bei Brettspiel) und die Urteilsgenauigkeit anhand der Einschätzung zu Fragen eines standardisierten Verfahrens, die für die Erzieherinnen weniger praxisnah und damit weniger vertraut sein könnten.

10.2.2 Design

Eine Stärke des vorliegenden Designs ist vor allem die Randomisierung auf Einrichtungsebene (Kap. 7.2).

Als weiterer Vorzug der Studie darf die Implementationskontrolle (Kap. 7.5.2) der zweitägigen Fortbildungen in den Treatments gelten. Diese Durchführungskontrolle der Erzieherinnenfortbildung erfolgte mittels Videografie oder teilnehmender Beobachtung und zeigt eine weniger unterschiedliche Umsetzung der bereichsspezifischen Fortbildungen als die Literatur vermuten ließ. Im Rahmen der lehrgangsbasierten Fortbildung (Kap. 7.2.2) wurden neben dem strukturierten Rahmenkonzept mit verbindlichen Ritualen, Geschichten und Figuren auch die im Rahmen der Fortbildung zur lernwegsbasierten Förderung empfohlenen Lehrmittel und Spiele vorgestellt. Die Fortbildnerin des lehrgangsbasierten Ansatzes begründete die Einbeziehung einer Vielzahl weiterer Medien mit der Weiterentwicklung und der Offenheit des Konzepts *Zahlenland* nach Friedrich und Galgóczy (2008). Damit setze sich das Konzept in der weiterentwickelten hier umgesetzten Variante deutlich von dem ursprünglich ähnlich strukturierten *Zahlenland* nach Preiß (2009) ab, in dem zehn vorgegebene Lerneinheiten inhaltlich und methodisch geschlossener seien. Diese werden aufgrund ihrer Ähnlichkeit in der Literatur vielfach gemeinsam dargestellt (Gasteiger, 2010, S. 79–84).

Dilemma zwischen interner und ökologischer Validität

Als ein generelles Problem quasiexperimenteller Feldforschung gilt das Dilemma zwischen interner und ökologischer Validität, welches in der vorliegenden Studie im Hinblick auf die Auswahl von Fortbildnerinnen zum Tragen kommt. Eigens für die Studie

geschulte Fortbildnerinnen hätten die interne Validität der Studie erhöht, weil unter Gleichhaltung des Einflusses seitens der Fortbildnerin die Unterschiede zwischen den Ansätzen lehrgangs- und lernwegsbasierter Förderung gezielt herausgearbeitet worden wären, so wie sie in der Literatur beschrieben sind. Wie die Implementationskontrolle der Fortbildung jedoch gezeigt hat, wären dadurch Fortbildungen entstanden, die Discrepanzen zur Realität der Fortbildungspraxis aufgewiesen hätten. Durch die Entscheidung für externe und sehr erfahrene Fortbildnerinnen wurde zugunsten der ökologischen Validität entschieden. Um die Beeinträchtigung der internen Validität zu reduzieren, die auf nicht kontrollierte Unterschiede zwischen den Fortbildnerinnen zurückgeführt werden könnte, wurden für die Fortbildungen gezielt die besten Vertreterinnen des Ansatzes gewonnen, die über die größte Fortbildungserfahrung für den jeweiligen Ansatz verfügten.

Auswahl der Messzeitpunkte

Mit der Erfassung der Beobachtungskompetenz direkt vor und nach der Fortbildung untersucht die Studie nur kurzfristige Effekte der Treatments auf die Beobachtungskompetenz.

Die Ergebnisse zur Urteilsgenauigkeit müssen vor dem Hintergrund gesehen werden, dass die erste Erhebung der Urteilsgenauigkeit vor der Fortbildung stattfand und auch vor dem Beginn der Förderphase der Kinder. Zum Zeitpunkt der Ersteinschätzung ihrer Bezugskinder hatten die Erzieherinnen noch keine explizite mathematische Förderung durchgeführt, zum Zeitpunkt des Posttests hatten sie ihre Bezugskinder über den Zeitraum von neun Monaten hinweg einmal wöchentlich im Kontext der mathematischen Frühförderung auf diesem Gebiet kennengelernt. Zudem kann eine neunmonatige Entwicklung der Kinder dazu führen, dass sie nach diesem Zeitraum mathematische Fähigkeiten erworben haben, die für die Erzieherinnen leichter zu erkennen sind.

Die Tatsache, dass zum Messzeitpunkt 1 in allen Komponenten einige Erzieherinnen auch ideale Übereinstimmungen zwischen Testurteil und Kriteriumswert zeigten (siehe Tab. 18), lässt sich mit den bei einigen Kindern im Prätest entwicklungsbedingt noch wesentlich weniger vorhandenen mathematischen Kompetenzen erklären als dies zum Zeitpunkt des Posttests der Fall war. Dementsprechend war die Einschätzung leichter. Die im Posttest höhere durchschnittliche Kompetenz bei gleichzeitiger Abnahme der idealen Übereinstimmungen spricht für den deutlichen Kompetenzzuwachs der Erzieherinnen.

10.2.3 Stichprobe

Die Stichprobenqualität ist im nationalen Vergleich als besonders hoch hervorzuheben.

Eine Stärke der Studie besteht in der Aussagekraft der Stichprobe aufgrund ihrer Größe. In der vorliegenden Untersuchung schätzten 20 Erzieherinnen jeweils vier und 19 Erzieherinnen jeweils fünf Kinder ein, mit denen sie durchschnittlich 4,9 Stunden pro Woche verbrachten. Keine Studie hat bisher die Kontaktzeit zu den Kindern berücksichtigt. Zudem wurde in der vorliegenden Untersuchung ein Minimum von vier Einschätzungen pro Erzieherin eingehalten. In vorausgegangenen Studien variierte die Anzahl der Einschätzungen pro Erzieherin bei Kammermeyer (2000) zwischen drei und zehn Kindern (ebd., S. 200) oder wurde nicht berichtet (Bruns, 2014). Bei Dollinger (2013) schätzten mehrere Erzieherinnen dieselben Kinder ein, wodurch mehr paarweise Einschätzungen vorlagen als tatsächlich Kinder an der Studie teilgenommen hatten (ebd., S.86). Einige Erzieherinnen schätzten zudem nur ein oder zwei Kinder ein (ebd., S. 193).

Die Poweranalysen (vgl. Kap. 7.6) zeigen, dass eine Stichprobengröße von $N = 39$ für die gewählten Auswertungsverfahren ausreichend ist. Die vorliegende Untersuchung hätte eine Gesamtstichprobengröße von $N = 103$ vorweisen können. Zur Erhöhung der Datenqualität wurde auf diese Größe bewusst verzichtet und eine Auswahl von 39 Erzieherinnen anhand der Kontaktzeit zu den Kindern getroffen (vgl. Kap. 7.3). Damit wurde die Urteilsgenauigkeitsuntersuchung (vgl. Kap. 8.3.1) optimiert und zugleich eine Stichprobengröße bewahrt, die den anderen nationalen Untersuchungen mit Erzieherinnenbeteiligung gleichwertig ist ($N = 18$ bei Kammermeyer, 2000, S. 206; $N = 44$ bei Dollinger, 2013, S. 85; $N = 31$ bei Bruns, 2014, S. 125).

Eine Schwäche der vorliegenden Studie ist die Begrenzung der Stichprobe auf Rheinland-Pfalz, weshalb eine Generalisierung der Ergebnisse nicht möglich ist. Zudem kann die Stichprobe nur als „Gelegenheitsstichprobe“ (Bortz & Döring, 2016, S. 305), und aufgrund des Teilnahmeaufrufs über die Fachberater(innen) im Speziellen als „Selbstselektions-Stichprobe“ bezeichnet werden, was die Aussagekraft der Untersuchung im Rahmen der quantitativen Forschung begrenzt (ebd.). Dieser und ähnlicher Limitationen unterliegt jedoch eine Vielzahl quasiexperimenteller Studien im Feld.

Die Ergebnisse müssen dementsprechend zwar vorsichtig interpretiert werden; im Hinblick aber auf die unter Berücksichtigung bedeutsamer Faktoren (Kontaktzeit und Min-

destanzahl an Einschätzungen pro Erzieherin) erfolgte Erzieherinnen-Kind-Zuteilung ist die Stichprobenqualität insgesamt als sehr hoch einzustufen.

10.2.4 Limitationen

Die vorliegende Betrachtung unterliegt Limitationen, die sich vornehmlich auf die Erhebungsinstrumente und die Erhebungszeitpunkte beziehen. Zudem begrenzt die Beschränkung auf den mathematischen Bereich ihre Aussagekraft.

Die Komponentenanalyse nach Schrader (1989) kann kritisiert werden, da jeder Unterschied zwischen Einschätzung der Kinder und Testwert als Fehler interpretiert wird. Dieses problematisieren Schrader und Helmke bereits 1987 (S. 38–39), indem sie darauf hinweisen, dass die Einschätzungen möglicherweise eher an der Kompetenz als an der Performanz orientiert sind, was auf Einschätzungen von sehr jungen Kindern in besonderer Weise zutreffen könnte. Eventuell fokussieren Erzieherinnen auf bestmögliche Leistung in Alltagssituationen und berücksichtigen nicht die für Kinder ungewohnte Testsituationen und deren Auswirkungen auf die kindliche Leistung. Hierzu merkt Kammermeyer an: „Die ungewohnte Testsituation oder andere Situationsvariablen (z.B. nachlassende Motivation, fehlendes Anweisungsverständnis) können bei so jungen Kindern große Bedeutung haben und könnten dafür verantwortlich sein, daß die Kinder ihr Fähigkeitspotential nicht voll zur Entfaltung bringen konnten“ (2000, S.210).

Die Erhebung der Beobachtungskompetenz erfolgte in der Kontrollgruppe nur zum Zeitpunkt des Prätests, weil eine erneute Erhebung nach zwei Tagen Fortbildung den Erzieherinnen der Kontrollgruppe vermeintlich nur schwer erklärbar gewesen wäre und zudem keine Veränderung nach einem so kurzen Zeitraum erwartet wurde. Zur Kontrolle etwaiger Trainingseffekte wäre tatsächlich eine Wiederholung nach zwei Wochen, in denen die zweitägigen Fortbildungen stattfanden, sinnvoll gewesen.

Eine weitere Limitation stellen die unterschiedlichen Posttestzeitpunkte der Beobachtungskompetenz- und Urteilsgenauigkeitsmessung dar. Während die Erfassung der Beobachtungskompetenz direkt nach der Fortbildung erfolgte, fand der Posttest zur Urteilsgenauigkeit neun Monate später statt, weil die Erzieherinnen zunächst mit ihren Bezugskindern arbeiten mussten, um Effekte auf die Urteilsgenauigkeit messen zu können. Eine gleichzeitige Erfassung von Urteilsgenauigkeit und Beobachtungskompetenz

nach neun Monaten hätte auch langfristige Effekte auf die Beobachtungskompetenz zeigen können.

10.3 Implikationen für Praxis und Forschung

10.3.1 Wissenschaftlicher Ertrag der Arbeit

Der vornehmliche Ertrag der vorliegenden Untersuchung besteht in der Einführung der Beobachtungskompetenz als Facette diagnostischer Kompetenz. Zudem zeichnet sich die Arbeit durch ihre Anschlussfähigkeit an die Accuracy-Forschung aus, weil sie als erste Studie den Zusammenhang zwischen Urteilsgenauigkeit und Beobachtungskompetenz fokussiert und damit erstmalig eine Brücke schlägt zwischen den Forschungslinien zur Accuracy-Forschung und den Forschungen zum formativen Assessment.

Die Einbindung der Beobachtungskompetenz als neue Facette diagnostischer Kompetenz erweitert das Linsenmodell von Förster und Böhmer (2017, S. 47) (vgl. Kap. 3.1.1). Zudem berücksichtigt das erweiterte Linsenmodell diagnostischer Kompetenz (Kap. 4.6.1) die Situiertheit prozessbezogener Diagnostik, indem ein Filter eingefügt wird, der die Situation als begrenzenden Einflussfaktor darstellt.

Damit folgt das erweiterte Linsenmodell auf theoretischer Ebene Schraders (1989) Forderung nach Berücksichtigung der Situation als Teil der Analysen zur Untersuchung diagnostischer Kompetenz. Schrader empfiehlt zudem, in künftigen Untersuchungen „die in der Unterrichtssituation selbst ablaufenden Diagnoseprozesse zum Gegenstand der Analyse“ (ebd., S. 241) zu machen, was das vorliegende Modell mit der Beobachtungskompetenz berücksichtigt. Die Aktualität der Forderung und die Notwendigkeit der Berücksichtigung der Situation in Bezug auf die Erfassung diagnostischer Kompetenz zeigen auch Praetorius et al. (2017, S. 97) mit ihrer repräsentativen Auswahl relevanter diagnostischer Situationen für den schulischen Kontext. Das erweiterte Linsenmodell bedenkt damit die Grundsatzfrage, ob man über Situationen hinweg generalisieren könne und zeigt, dass die Situation den Diagnoseprozess entscheidend beeinflusst und aus diesem Grund stets Besonderheiten der jeweiligen Situation berücksichtigt werden müssen. Dieser Beitrag des erweiterten Linsenmodells bleibt theoretisch und wurde empirisch in der vorliegenden Studie nicht überprüft.

Der im Ergebnis nicht nachgewiesene Zusammenhang zwischen Urteilsgenauigkeit und Beobachtungskompetenz stellt für die zukünftige Forschung bedeutsame Hinweise für die weitere Untersuchung der Facetten diagnostischer Kompetenz bereit.

Im Hinblick auf den frühpädagogischen Kontext zeigt die vorliegende Untersuchung zum einen, dass ein ökologisch valides Treatment (Fortbildung) in einem Umfang von zwei Tagen implizit zu einer Verbesserung diagnostischer Kompetenzen führen kann.

Zum anderen wird hier erstmalig im deutschsprachigen Raum die Messung der Urteilsgenauigkeit unter vollumfänglicher Erfüllung eines auf Hoge und Coladarci (1989) beruhenden Anforderungsprofils durchgeführt (vgl. Kap. 3.3). Ein weiteres Alleinstellungsmerkmal dieser Studie ist die erstmals mit einer Mindestanzahl von vier Einschätzungen pro Erzieherin unter Berücksichtigung der Kontaktzeiten zwischen Erzieherinnen und Bezugskindern erfolgte Erhebung der Urteilsgenauigkeit von Erzieherinnen.

Auf der theoretischen Ebene stellt diese Arbeit Überlegungen an zur besonderen Berücksichtigung diagnostisch relevanter Handlungsentscheidungen als Teil des Diagnoseprozesses im NeDiKo-Modell zur diagnostischen Kompetenz (10.1.5). Damit wird die Konzeptualisierung eines breiter gefassten Diagnosekompetenzbegriffs im Sinne der diagnostischen Expertise von Helmke (2017) angebahnt und ein Vorschlag eingebracht, der zu einer besseren Abbildung der prozessbezogenen Diagnostik im Sinne des formativen Assessments im Modell führen könnte.

10.3.2 Ausblick auf weitere Forschungsansätze

In der vorliegenden Studie werden Urteilsgenauigkeit und Beobachtungskompetenz als Kompetenzen oder unabhängige Teilfähigkeiten des übergeordneten Konzepts der diagnostischen Expertise gesehen. Zukünftige Untersuchungen könnten zeigen, wie beide diagnostischen Kompetenzen zur Optimierung adaptiver Bildungsprozesse letztlich auf verschiedene Qualitätsmerkmale bereichsspezifischer Förderung wirken.

Auch die Frage nach der Übertragbarkeit der hier erlangten Erkenntnisse auf bereichsspezifische Fortbildungen in anderen Domänen könnten zukünftige Studien beantworten.

Vielfach wird die Stärkung diagnostischer Kompetenz als Voraussetzung für eine Zunahme der Fähigkeit Lehrender zur Gestaltung adaptiver Lernumgebungen angenommen. Die vorliegende Untersuchung weist darauf hin, dass sich auch Fortbildungen zur

adaptiven Gestaltung von Lernumgebung indirekt auf die diagnostische Kompetenzen auswirken können. Diese Hinweise sollten in weiteren Studien überprüft werden.

Die vorliegende Studie stellt die Bedeutung der Situation für den diagnostischen Prozess heraus. Diese Situationen (Raum, Personen, didaktisches Material, Handlungsverläufe etc.) gezielt zu kontrollieren und zu variieren, ist mittels *Virtual Reality und Augmented Reality* in geradezu beliebiger Form möglich (Buehler & Kohne, 2019). Sowohl für zukünftige Forschungen als auch für die Gestaltung von Fortbildungen bieten digitale Lernumgebungen große Chancen zur situierten Untersuchung diagnostischer Kompetenzen und situierter Förderung adaptiver Lehrkompetenz.

Hinsichtlich der Beobachtungskompetenz müssen zukünftige Studien zum einen den Anteil etwaiger Trainingseffekte am vorliegenden Ergebnis aufzeigen, und zum anderen die langfristigen Effekte von Fortbildungen auf die Beobachtungskompetenz nachweisen. Die Untersuchung weiterer Einflussfaktoren wie z.B. die Berufserfahrung, das Fachwissen oder die didaktischen Fähigkeiten von Erzieherinnen (Erzieherin-Kind-Interaktion) kann weitere bedeutsame Erkenntnisse zur Veränderung der Beobachtungskompetenz generieren.

Als ein weiteres Forschungsanliegen wäre die Untersuchung der Beobachtungskompetenz mittels der Komponentenanalyse Cronbachs (1955) denkbar. Dabei könnten verschiedene, diagnostisch zu analysierende Situationen für die verschiedenen Merkmale innerhalb der Komponentenanalyse stehen. Analog zu verschiedenen Bereichen der numerisch-mathematischen Fähigkeiten könnten Situationen mit Schauspielern oder mittels *Virtual Reality* passgenau als Vignette umgesetzt werden. Damit würde neben der Aufgabenschwierigkeit beispielsweise die Schwierigkeit der Situationsanforderung messbar werden. Dieser differenzierte Blick auf die einzelnen Komponenten der Beobachtungskompetenz wäre ein Zugewinn und die eventuelle Anschlussfähigkeit dieser Forschung an bisherige Untersuchungen zur Urteilsgenauigkeit damit wahrscheinlicher.

10.4 Fazit

Die vorliegende Untersuchung setzt an aktuellen Anforderungen an die Diagnosekompetenzforschung an, welche z.B. Schrader und Praetorius formulieren (2018, S. 96). Sie fordern eine weitere Aufhellung nicht nur der Urteilsqualität, sondern auch des Urteilsprozesses und die Verankerung der diagnostischen Kompetenz in einem Kompetenzmodell (ebd.). Die Qualität der Erfassung der Urteilsgenauigkeit ist eine Stärke der vorliegenden Studie. Auch die Berücksichtigung der prozessorientierten Perspektive ist durch die Untersuchung der Beobachtungskompetenz gegeben, deren Entstehung mit theoretischen Überlegungen verknüpft ist, die letztlich zu Vorschlägen einer Erweiterung bestehender Modelle geführt hat und damit eine stärkere Berücksichtigung von formativem Assessment in die konzeptuellen Überlegungen weiterführender Studien begründet. Die weiteren Forderungen nach einer stärkeren Anbindung informeller Diagnoseleistungen an die formelle Diagnostik und eine Kombination verschiedener Forschungsstränge und Ansätze wie bspw. produkt- und prozessorientierte Ansätze (ebd.) liefert die vorliegende Studie mit der Untersuchung der Urteilsgenauigkeit und damit der Accuracy-Forschung als Ausgangspunkt der Überlegungen, der Begründung einer Erweiterung dieses für die Praxis kaum bedeutsamen Konstrukts, mit der Einführung der Beobachtungskompetenz als diagnostischer Anteil formativen Assessments und schließlich mit der Untersuchung des Zusammenhangs beider Kompetenzen. Die Forderung von Schrader und Praetorius (2018) nach fachübergreifenden und fachspezifischen Sichtweisen erfüllt die vorliegende Untersuchung mit fachübergreifenden theoretischen Überlegungen und Modellen und mit der fachspezifischen Analyse der diagnostischen Kompetenzen in der Domäne Mathematik.

Zusammenfassend lässt sich festhalten, dass die vorliegende Studie die erste Arbeit ist, die sowohl Beobachtungskompetenz (Forschung zu formativem Assessment) als auch Urteilsgenauigkeit (Accuracy-Forschung) anhand derselben Stichprobe untersucht. Während die theoretischen Überlegungen auf den schulischen Kontext weitgehend übertragen werden können, muss auf die eingeschränkte Vergleichbarkeit dieser im vorschulischen Kontext entstandenen Untersuchung hingewiesen werden. Die diagnostische Kompetenz könnte sich in den beiden Bildungssettings (Vorschule und Schule) nicht nur deshalb unterscheiden, weil Lehrkräfte im Gegensatz zu Erzieherinnen eine fachdidaktische Ausbildung erhalten, sondern auch deshalb, weil sich die Kinder im Elementarbereich auf wesentlich niedrigeren Kompetenzstufen befinden als Schulkin-

der. Damit kann sich die Relevanz eines Blicks in die Tiefe (Urteilsgenauigkeit) und eines breiten Blicks auf im Prozess gezeigte Fähigkeiten (Beobachtungskompetenz) in Schule und Elementarbereich unterscheiden, wobei diese diagnostischen Kompetenzen je nach Zweck der Diagnostik beide relevant sind. Für den vorschulischen Bereich im Allgemeinen und für den schulischen Bereich im Hinblick auf die Gestaltung adaptiver Lernumgebungen erscheint die Beobachtungskompetenz als die bedeutsamere Kompetenz.

Literaturverzeichnis

- Abs, H. J. (2006). Zur Bildung diagnostischer Kompetenz in der zweiten Phase der Lehrerbildung. *Zeitschrift für Pädagogik*, 52(51. Beiheft), 217–234.
- Abs, H. J. (2007). Überlegungen zur Modellierung diagnostischer Kompetenz bei Lehrerinnen und Lehrern. In M. Lüders & J. Wissinger (Hrsg.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (S. 63–84). Münster: Waxmann.
- Airasian, P. W., Kellaghan, T., Madaus, G. F. & Pedulla, J. J. (1977). Proportion and direction of teacher rating changes of pupils' progress attributable to standardized test information. *Journal of Educational Psychology*, 69(6), 702–709.
<https://doi.org/10.1037/0022-0663.69.6.702>
- Anders, Y. (2018). Professionalität und Professionalisierung in der frühkindlichen Bildung. *Zeitschrift für Grundschulforschung*, 11(2), 183–197.
- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57(3), 175–193.
- Artelt, C. & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 157–160. <https://doi.org/10.1024/1010-0652.23.34.157>
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–137). Opladen: Leske + Budrich.
- Autorengruppe Fachschulwesen. (2011). *Qualifikationsprofil „Frühpädagogik“ – Fachschule, Fachakademie* (WiFF-Kooperationen, Bd. 1, Stand : Januar 2011). München: DJI. Zugriff am 09.03.2014.
- Back, M. D. & Nestler, S. (2016). Accuracy of judging personality. In J. A. Hall, M. Schmid Mast & T. V. West (Hrsg.), *The social psychology of perceiving others accurately* (S. 98–124). Cambridge, United Kingdom: Cambridge University Press.

- Baroody, A., Clements, D. H. & Sarama, J. (2019). Teaching and learning mathematics in early childhood programs. In C. P. Brown, M. B. McMullen & N. File (Hrsg.), *The Wiley handbook of early childhood care and education* (S. 329–354). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Barth, K. (1998). *Diagnostische Einschätzskalen zur Beurteilung des Entwicklungsstands und der Schulfähigkeit (DES)*. München: Reinhardt.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–35). Münster: Waxmann.
- Behr, K. & Walter, M. (2012). *Qualifikationen und Weiterbildung frühpädagogischer Fachkräfte. Bundesweite Befragung von Einrichtungsleitungen und Fachkräften in Kindertageseinrichtungen: Zehn Fragen – Zehn Antworten. Eine Studie der Weiterbildungsinitiative Frühpädagogische Fachkräfte* (WiFF-Studien, Weiterbildung, Bd. 15). München: Deutsches Jugendinstitut e.V. Verfügbar unter https://www.weiterbildungsinitiative.de/uploads/media/Studie_BeherWalter.pdf
- Behrmann, L. & Kaiser, J. (2017). Das Modell pädagogischer Diagnostik nach Ingenkamp und Lissmann. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 59–63). Münster: Waxmann.
- Behrmann, L. & Ophuysen, S. (2017). Das Vier-Komponenten-Modell der Diagnosequalität. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 38–42). Münster: Waxmann.
- Behrmann, L. & Souvignier, E. (2013). The relation between teachers' diagnostic sensitivity, their instructional activities, and their students' achievement gains in reading. *Zeitschrift Für Pädagogische Psychologie*, 27(4), 283–293.
<https://doi.org/10.1024/1010-0652/a000112>
- Bennett, R. E. (1982). Cautions for the use of informal measures in the educational assessment of exceptional children. *Journal of Learning Disabilities*, 15(6), 337–339.

- Benz, C., Peter-Koop, A. & Grüßing, M. (2015). *Frühe mathematische Bildung. Mathematiklernen der Drei- bis Achtjährigen*. Berlin: Springer Spektrum.
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J. et al. (2006). A learning integrated assessment system. *Educational Research Review*, (1), 61–67.
- Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8–21. <https://doi.org/10.1177/003172170408600105>
- Black, P. & Wiliam, D. (1998). Inside the black box. Raising standards through classroom assessment. *The Phi Delta Kappan*, 80(2), 139–148. Verfügbar unter <http://www.jstor.org/stable/pdf/20439383.pdf>
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Böhmer, M., Englich, B. & Böhmer, I. (2017). Schülerbeurteilungen aus der Perspektive dualer Prozessmodelle der sozialen Urteilsbildung. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 50–54). Münster: Waxmann.
- Bortz, J. & Döring, N. (2009). *Forschungsmethoden und Evaluation. Für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.
- Bortz, J. & Döring, N. (2016). *Forschungsmethoden und Evaluation in der Sozial- und Humanwissenschaft* (5. Aufl.). Heidelberg: Springer.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7., vollst. überarb. u. erw. Aufl.). Berlin: Springer-Verlag.
- Breitenbach, E. (2020). *Diagnostik. Eine Einführung* (Module Erziehungswissenschaft).
- Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L. & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31(4), 13–17. <https://doi.org/10.1111/j.1745-3992.2012.00251.x>
- Bröder, A. & Hilbig, B. E. (2017). Urteilen und Entscheiden. In J. Müsseler & M. Rieger (Hrsg.), *Allgemeine Psychologie* (Springer Lehrbuch, 3. Aufl., S. 619–659). Berlin: Springer.
- Bruder, S., Klug, J., Hertel, S. & Schmitz, B. (2010). Messung, Modellierung und Förderung der Beratungskompetenz und Diagnostischen Kompetenz von Lehrkräften. In

- K. Beck & O. Zlatkin-Troischanskaia (Hrsg.), *Themenheft Lehrerbildung auf dem Prüfstand* (S. 173–193). Landau: Empirische Pädagogik.
- Bruns, J. (2014). *Adaptive Förderung in der elementarpädagogischen Praxis. Eine empirische Studie zum didaktischen Handeln von Erzieherinnen und Erziehern im Bereich Mathematik* (Empirische Studien zur Didaktik der Mathematik, Bd. 21). Münster: Waxmann.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2. Aufl.). Berkeley, CA: University of California Press.
- Buehler, K. & Kohne, A. (2019). Lernen mit Virtual Reality: Chancen und Möglichkeiten der digitalen Aus- und Fortbildung. In M. Groß, M. Müller-Wiegand & D. F. Pinnow (Hrsg.), *Zukunftsfähige Unternehmensführung. Ideen, Konzepte und Praxisbeispiele* (Bd. 12, S. 209–224). Berlin, Heidelberg: Springer Gabler.
https://doi.org/10.1007/978-3-662-59527-5_11
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (Psychologie, 3., aktual. und erw. Aufl.). München: Pearson Studium.
- Bürgermeister, A. (2014). *Leistungsbeurteilung im Mathematikunterricht. Bedingungen und Effekte von Beurteilungspraxis und Beurteilungsgenauigkeit* (Empirische Erziehungswissenschaft, Bd. 45). Münster: Waxmann.
- Clements, D. H. & Sarama, J. (2014). *Learning and Teaching Early Math. The Learning Trajectories Approach* (Second edition). New York: Routledge.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). New York: Psychology Press.
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78(2), 141–146.
- Corvacho del Toro, I. & Greb, K. (2007). Persönlichkeits- und Lernentwicklung von Grundschulkindern. Zur Anlage des Projekts PERLE. In K. Möller, P. Hanke, C. Beinbrech, A. K. Hein, T. Kleickmann & R. Schages (Hrsg.), *Qualität von Grundschulunterricht. Entwickeln, erfassen und bewerten* (Jahrbuch Grundschulforschung, Bd. 11, 1. Aufl., S. 313–316). Wiesbaden: VS Verlag.
- Cronbach, L. J. (1955). Processes affecting scores on “understanding of others” and “assumed similarity”. *Psychological Bulletin*, (52), 177–193.
- Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Hrsg.), *Person perception and interpersonal behavior* (S. 353–379). Stanford: Stanford University Press.

- Cronbach, L. J. & Gleser, G. C. (1965). *Psychological tests and personnel decisions* (2. Aufl.). Urbana and London: University of Illinois Press.
- Demaray, M. K. & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, 13(1), 8–24. <https://doi.org/10.1037/h0088969>
- Dollinger, S. (2013). *Diagnosegenauigkeit von Erzieherinnen und Lehrerinnen. Einschätzung schulrelevanter Kompetenzen in der Übergangsphase*. Wiesbaden: VS Verlag.
- Donie, C. (2010a). *Befragung zur Erfassung der Urteilsgenauigkeit von Erzieherinnen (BEUrteilEr). Einschätzung vier- bis sechsjähriger Vorschulkinder*. Forschungsversion (unveröffentlicht). Landau in der Pfalz: Universität Koblenz-Landau.
- Donie, C. (2010b). *Filmbasierte Befragung zur Erfassung der Beobachtungskompetenz von Erzieherinnen. FiBB*. Forschungsversion (unveröffentlicht). Landau in der Pfalz: Universität Koblenz-Landau.
- Donie, C., Kammermeyer, G. & Roux, S. (2013). Förderung schriftsprachlicher und mathematischer Kompetenzen im Vorschulalter. *Empirische Pädagogik*, 27(3), 304–325.
- Dunekacke, S., Seemann, S., Heinze, A., Leuchter, M., Moser Opitz, E., Vogt, F. et al. (2018). *Frühe mathematische Bildung: Welche professionellen Kompetenzen brauchen pädagogische Fachkräfte und wie lassen sich diese fördern? Poster, vorgestellt beim Leibniz-Forschungsverbund Bildungspotentiale (LERN)*. Berlin: Bildungspolitisches Forum.
- Eckerth, M. (2013). *Formen der Diagnose und Förderung. Eine mehrperspektivische Analyse zur Praxis pädagogischer Fachkräfte in der Grundschule* (Internationale Hochschulschriften, Bd. 593, 1. Aufl.). Münster: Waxmann.
- Eichen, L. & Bruns, J. (2017). Interventionsstudie zur Entwicklung mathematikbezogener Einstellungen frühpädagogischer Fachpersonen. *Frühe Bildung*, 6(2), 67–73. <https://doi.org/10.1026/2191-9186/a000310>
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden. Lehrbuch mit Online-Materialien* (1. Aufl.). Weinheim: Beltz.
- Erath, P. & Rossa, M. (2017). Lernen von der Sache aus!? Über Möglichkeiten und Schwierigkeiten einer dialogisch-instruktiven Didaktik für den Elementarbereich. In P. Erath, F.-M. Konrad & M. Rossa (Hrsg.), *Der Kindergarten als Bildungseinrichtung. Pädagogische, didaktische und methodische Aspekte einer bildungstheoretischen*

- schen Vertiefung der Arbeit in Kindertageseinrichtungen* (S. 97–118). Bad Heilbrunn: Klinkhardt.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 2(39), 157–191.
- Feinberg, A. B. & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, 18(1), 52–65.
<https://doi.org/10.1521/scpq.18.1.52.20876>
- Feinberg, A. B. & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, 102(6), 453–462. <https://doi.org/10.3200/JOER.102.6.453-462>
- Förster, N. & Böhmer, I. (2017). Das Linsenmodell – Grundlagen und exemplare Anwendungen in der pädagogisch-psychologischen Diagnostik. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 46–50). Münster: Waxmann.
- Friedrich, G. & Galgóczy, V. de (2008). *Komm mit ins Zahlenland. Eine spielerische Entdeckungsreise in die Welt der Mathematik* (4. Aufl.). Stuttgart: Urania.
- Friedrich, G. & Munz, H. (2004). Projekt- und Evaluationsbericht „Komm mit ins Zahlenland“. Ein „ganzheitliches“ Frühförderkonzept am Beispiel elementarer Mathematik. Zugriff am 13.07.2013. Verfügbar unter:
<http://www.ifvl.de/material/Projektbericht-Zahlenland.pdf>
- Friedrich, G. & Munz, H. (2006). Förderung schulischer Vorläuferfähigkeiten durch das didaktische Konzept „Komm mit ins Zahlenland“. *Psychologie in Erziehung und Unterricht*, 53(2), 134–146.
- Fröhlich-Gildhoff, K., Nentwig-Gesemann, I. & Pietsch, S. (2011). *Kompetenzorientierung in der Qualifizierung frühpädagogischer Fachkräfte. Eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF)* (WiFF-Expertisen, Bd. 19). München: DJI.
- Fröhlich-Gildhoff, K. & Strohmer, J. (2011). Untersuchungen zum Stand von Beobachtung, Dokumentation und Diagnostik in Kindertageseinrichtungen. In K. Fröhlich-Gildhoff, I. Nentwig-Gesemann & H. R. Leu (Hrsg.), *Forschung in der Frühpädagogik IV. Schwerpunkt: Beobachten, Verstehen, Interpretieren, Diagnostizieren* (Ma-

- terialien zur Frühpädagogik, Bd. 9, S. 37–68). Freiburg: FEL Verlag Forschung Entwicklung Lehre.
- Fuchs, D. & Fuchs, L. S. (1986). Test procedure bias. A meta-analysis of examiner familiarity effects. *Review of Educational Research*, 56(2), 243–262.
<https://doi.org/10.2307/1170377>
- Gage, N. L. & Exline, R. V. (1953). Social perception and effectiveness in discussion groups. *Human Relations*, 6(4), 381–396.
<https://doi.org/10.1177/001872675300600405>
- Gage, N. L., Leavitt, G. S. & Stone, G. C. (1955). Teachers' understanding of their pupils and pupils' ratings of their teachers. *Psychological Monographs: General and Applied*, 69(21), 1–37. <https://doi.org/10.1037/h0093735>
- Gasteiger, H. (2008). Lernanregungen und -dokumentation im Alltag der Kindertagesstätte - ein kompetenzorientierter Förderansatz. In É. Vásárhelyi (Hrsg.), *Beiträge zum Mathematikunterricht 2008. Online. Vorträge auf der 42. Tagung für Didaktik der Mathematik vom 13.3.2008 bis 18.3.2008 in Budapest*. Verfügbar unter <https://eldorado.tu-dortmund.de/bitstream/2003/33668/1/BzMU2008-gesamt.pdf>
- Gasteiger, H. (2010). *Elementare mathematische Bildung im Alltag der Kindertagesstätte. Grundlegung und Evaluation eines kompetenzorientierten Förderansatzes* (Empirische Studien zur Didaktik der Mathematik, Bd. 3). Münster: Waxmann.
- Gasteiger, H. & Benz, C. (2018). Enhancing and analyzing kindergarten teachers' professional knowledge for early mathematics education. *The Journal of Mathematical Behavior*, 51, 109–117.
- Geiling, U. & Liebers, K. (2014). Diagnostik und Assessment im Elementarbereich. In R. Braches-Chyrek, C. Röhner, H. Sünder & M. Hopf (Hrsg.), *Handbuch Frühe Kindheit* (S. 529–537). Opladen: Verlag Barbara Budrich.
- Gibbons, P. (2002). *Scaffolding language, scaffolding learning. Teaching second language learners in the mainstream classroom*. Portsmouth, NH: Heinemann.
- Glogger-Frey, I. & Herppich, S. (2017). Formative Diagnostik als Teilaspekt diagnostischer Kompetenz. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 42–46). Münster: Waxmann.
- Grausam, N. C. (2018). *Diagnosekompetenz von Lehrpersonen als Voraussetzung individueller Förderung im Bereich „Texte schreiben“*. Eine empirische Studie am Bei-

- spiel einer neu eingeführten integrierten Schulform* (Die Gemeinschaftsschule in Baden-Württemberg, Bd. 1). Münster: Waxmann.
- Greb, K., Poloczek, S., Lipowsky, F. & Faust, G. (2011). *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulern“ (PERLE). Teil 1. PERLE-Instrumente: Schüler, Lehrer & Eltern (Messzeitpunkt 1)* (Materialien zur Bildungsforschung, 23/1, 2., überarb. Aufl.). Frankfurt am Main.
- Greenstein, L. (2010). *What teachers really need to know about formative assessment*. Alexandria: ASCD.
- Häcker, T. (2017). Individualisierter Unterricht. In T. Bohl, J. Budde & M. Rieger-Ladich (Hrsg.), *Umgang mit Heterogenität in Schule und Unterricht. Grundlagentheoretische Beiträge, empirische Befunde und didaktische Reflexionen* (UTB Schulpädagogik, Bd. 4755, S. 275–290). Bad Heilbrunn: Klinkhardt.
- Hardy, I., Hertel, S., Kunter, M., Klieme, E., Warwas, J., Büttner, G. et al. (2011). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. *Zeitschrift für Pädagogik*, 57(6), 819–833.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *Curriculum Journal*, 16(2), 207–223.
<https://doi.org/10.1080/09585170500136093>
- Hascher, T. (2008). Diagnostische Kompetenzen im Lehrerberuf. In C. Kraler & M. Schratz (Hrsg.), *Wissen erwerben, Kompetenzen entwickeln. Modelle zur kompetenzorientierten Lehrerbildung* (S. 71–86). Münster: Waxmann.
- Hascher, T. (2011). Diagnostizieren in der Schule. In *PraxisWissen SchulLeitung* (Bd. 34.11, S. 1–9). Unterschleissheim: Luchterhand.
- Hattie, J. (2003). *Formative and summative interpretations of assessment information*. Zugriff am 25.03.2018. Verfügbar unter <https://cdn.auckland.ac.nz/assets/education/hattie/docs/formative-and-summative-assessment-%282003%29.pdf>
- Hattie, J. N. (2013). *Lernen sichtbar machen. Überarbeitete deutschsprachige Ausgabe von „Visible Learning“, besorgt von Wolfgang Beywl und Klaus Zierer* (1. Aufl.). Baltmannsweiler: Schneider Hohengehren.
- Hauser, B. (2016). *Spielen. Frühes Lernen in Familie, Krippe und Kindergarten* (2. Aufl.). Stuttgart: W. Kohlhammer Verlag.

- Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (6., überarb. Aufl.). Seelze: Klett-Kallmeyer.
- Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (7. Aufl.). Seelze: Klett-Kallmeyer.
- Helmke, A. & Fend, H. (1981). Wie gut kennen Eltern ihre Kinder und Lehrer ihre Schüler? Ergebnisse der Konstanzer Untersuchungen zu Bedingungen der Diagnosegenauigkeit bei Eltern und Lehrern. In G. Zimmer (Hrsg.), *Persönlichkeitsentwicklung und Gesundheit im Schulalter. Gefährdungen und Prävention* (S. 341–360). Frankfurt am Main: Campus Verlag.
- Helmke, A., Grommelt, U., Hoppe, R. & Schrader, F.-W. (1984). *Manual für die Durchführung und Auswertung der Unterrichtsbeobachtungsstudie im Rahmen des Projekts „Unterrichtsqualität und Leistungszuwachs“*. München: Max-Planck-Institut für psychologische Forschung.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griesse (Hrsg.), *Schulleitung und Schulentwicklung. Voraussetzungen, Bedingungen, Erfahrungen* (S. 119–144). Baltmannsweiler: Schneider-Verlag Hohengehren.
- Helmke, A. & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91–98.
- Helmke, A. & Schrader, F.-W. (2020). Angebots-Nutzungs-Modell der Wirkfaktoren akademischer Leistungen. In M. A. Wirtz (Hrsg.), *Dorsch – Lexikon der Psychologie*. Zugriff am 09.02.2020. Verfügbar unter <https://portal.hogrefe.com/dorsch/angebots-nutzungs-modell-der-wirkfaktoren-akademischer-leistungen/>
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie. Psychologie des Unterrichts und der Schule* (Themenbereich D, Serie 1, Bd. 3, S. 71–176). Göttingen: Hogrefe.
- Hepberger, B., Lindmeier, A., Moser Opitz, E. & Heinze, A. (2017). „Zähl’ nochmal genauer!“ – Handlungsnahe mathematikbezogene Kompetenzen von pädagogischen Fachkräften erheben. In S. Schuler, C. Streit & G. Wittmann (Hrsg.), *Perspektiven mathematischer Bildung im Übergang vom Kindergarten zur Grundschule* (S. 239–253). Wiesbaden: Springer Spektrum.

- Heritage, M. (2007). Formative Assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140–145.
- Herppich, S., Praetorius, A.-K., Hetmanek, A., Glogger-Frey, I., Ufer, S., Leutner, D. et al. (2017). Ein Arbeitsmodell für die empirische Erforschung der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 75–93). Münster: Waxmann.
- Hoenisch, N. & Niggemeyer, E. (2004). *Mathe-Kings. Junge Kinder fassen Mathematik an*. Weimar: verlag das netz.
- Hoffmann, L. & Böhme, K. (2014). Wie gut können Grundschullehrkräfte die Schwierigkeit von Deutsch- und Mathematikaufgaben beurteilen? Eine Untersuchung zur Genauigkeit aufgabenbezogener Lehrerurteile auf Klassenebene. *Psychologie in Erziehung und Unterricht*, 61(1), 42–55.
- Hogan, J. & Brent, R. W. (1996). Issues and non-issues in the fidelity – bandwidth trade-off. *Journal of Organizational Behavior*, 17(6), 627–637.
- Hoge, R. D. & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology*, 76(5), 777–781.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313.
- Honig, M.-S., Schreiber, N. & Lang, S. (2006). *Begleitstudie zur Umsetzung der „Bildungs- und Erziehungsempfehlungen für Kindertagesstätten in Rheinland-Pfalz“*. Abschlussbericht an das Ministerium für Bildung, Frauen und Jugend. Trier: Universität Trier.
- Hopkins, K. D., George, C. A. & Williams, D. D. (1985). The concurrent validity of standardized achievement tests by content area using teachers' ratings as criteria. *Journal of Educational Measurement*, 22(3), 177–182.
- Hosenfeld, I. (2009). *Teachers' accuracy in judging mathematics tasks and the relationship to achievement gains*. Paper presented at the 13th biennial conference EAR-LI 2009, Amsterdam.
- Hoth, J. (2016). *Situationsbezogene Diagnosekompetenz von Mathematiklehrkräften* (Perspektiven der Mathematikdidaktik). Dissertation. Wiesbaden: Springer Spektrum.

- Inckemann, E. (2008). Förderdiagnostische Kompetenzen von Grundschullehrerinnen im schriftsprachlichen Anfangsunterricht. *Zeitschrift für Grundschulforschung*, (2), 99–115.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik* (6., neu ausgest. Aufl.). Weinheim: Beltz.
- Jäger, R. S., Beetz, E., Erler, R. & Habersang-Walther, R. (1994). *MSD. Mannheimer Schuleingangsdiagnostikum* (4. Aufl.). Weinheim: Beltz.
- Janczyk, M. & Pfister, R. (2015). *Inferenzstatistik verstehen. Von A wie Signifikanztest bis Z wie Konfidenzintervall* (Springer-Lehrbuch, 2., überarb. u. erw. Aufl.). Berlin: Springer Berlin.
- JMK & KMK. (2004). *Gemeinsamer Rahmen der Länder für die frühe Bildung in Kindertageseinrichtungen. Beschluss der Jugendministerkonferenz vom 13./14.05.2004; Beschluss der Kultusministerkonferenz vom 03./04.06.2004*. Zugriff am 05.03.2016. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_06_04-Fruhe-Bildung-Kitas.pdf
- Jürgens, E. & Lissmann, U. (2015). *Pädagogische Diagnostik. Grundlagen und Methoden der Leistungsbeurteilung in der Schule*. Weinheim: Beltz.
- Kammermeyer, G. (2000). *Schulfähigkeit – Kriterien und diagnostische, prognostische Kompetenz von Lehrerinnen, Lehrern und Erziehern* (1. Aufl.). Bad Heilbrunn: Klinkhardt.
- Kammermeyer, G., Roux, S., Schneider, L. & Stuck, A. (2011, Juli). *Was wirkt wie? Evaluation von Sprachfördermaßnahmen. 2. Zwischenbericht*. Landau: Universität.
- Karing, C., Pfof, M. & Artelt, C. (2011). Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? *Journal for Educational Research Online*, 3(2), 119–147.
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern* (Empirische Erziehungswissenschaft, Bd. 35). Münster u.a: Waxmann.
- Karst, K. (2017a). Akkurate Urteile – die Ansätze von Schrader (1989) und McElvany et al. (2009). In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 21–25). Münster: Waxmann.

- Karst, K. (2017b). Diagnostische Kompetenz und unterrichtliche Situationen. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 25–29). Münster: Waxmann.
- Karst, K. & Förster, N. (2017). Ansätze zur Modellierung diagnostischer Kompetenz. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 19–65). Münster: Waxmann.
- Karst, K., Schoreit, E. & Lipowsky, F. (2014). Diagnostische Kompetenzen von Mathematiklehrern und ihr Vorhersagewert für die Lernentwicklung von Grundschulkindern. *Zeitschrift für Pädagogische Psychologie*, 28(4), 237–248.
<https://doi.org/10.1024/1010-0652/a000133>
- Kingston, N. & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37.
<https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Kleber, E. W. (1992). *Diagnostik in pädagogischen Handlungsfeldern. Einführung in Bewertung, Beurteilung, Diagnose und Evaluation* (Grundlagentexte Pädagogik). Weinheim: Juventa.
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Pädagogik*, 52(6), 876–903.
- Klieme, E., Lipowsky, F. & Rakoczy, K. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Münster: Waxmann.
- Kluczniok, K., Roßbach, H.-G. & Große, C. (2010). Fördermöglichkeiten im Kindergarten. Ein Systematisierungsversuch. In A. Diller, H. R. Leu & T. Rauschenbach (Hrsg.), *Wie viel Schule verträgt der Kindergarten? Annäherungen zweier Lernwelten* (Deutsches Jugendinstitut - Fachforum Bildung und Erziehung, Bd. 7, S. 133–152). München: Deutsches Jugendinstitut.
- Klug, J. (2017). Ein Prozessmodell zur Diagnostik und Förderung von selbstreguliertem Lernen. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische

- Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., 54–49). Münster: Waxmann.
- KMK (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 12.06.2014*. Zugriff am 01.10.2018. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf
- KMK & JFMK (2010). *Weiterentwicklung der Aus-, Fort- und Weiterbildung von Erzieherinnen und Erziehern. Gemeinsamer Orientierungsrahmen „Bildung und Erziehung in der Kindheit“*. Beschluss der Kultusministerkonferenz vom 16.09.2010, Beschluss der Jugend- und Familienministerkonferenz vom 14.12.2010. Zugriff am 05.03.2016. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_09_16-Ausbildung-Erzieher-KMK-JFMK.pdf
- Knauf, H. (2019). *Bildungsdokumentation in Kindertageseinrichtungen: Prozessorientierte Verfahren der Dokumentation von Bildung und Entwicklung* (1. Aufl.). Wiesbaden: VS Verlag. Verfügbar unter <https://link.springer.com/content/pdf/10.1007%2F978-3-658-24101-8.pdf>
- Knievel, I., Lindmeier, A. M. & Heinze, A. (2015). Beyond knowledge: Measuring primary teachers' subject-specific competences in and for teaching mathematics with items based on video vignettes. *International Journal of Science and Mathematics Education*, 13(2), 309–329. <https://doi.org/10.1007/s10763-014-9608-z>
- König, A., Kratz, J., Stadler, K. & Uihlein, C. (2018). *Aktuelle Entwicklungen in der Ausbildung von Erzieherinnen und Erziehern an Fachschulen für Sozialpädagogik. Organisationsformen, Zulassungsvoraussetzungen und Curricula – eine Dokumentenanalyse. Eine Studie der Weiterbildungsinitiative Frühpädagogische Fachkräfte* (WiFF-Studien, Ausbildung, Bd. 29). München: Deutsches Jugendinstitut e.V. Verfügbar unter https://www.weiterbildungsinitiative.de/uploads/media/WiFF_Studie_29_Koenig_Kratz_Stadler_Uihlein.pdf
- Kozulin, A., Gindis, B., Ageyev, V. S. & Miller, S. M. (Hrsg.). (2003). *Vygotsky's educational theory in cultural context*. New York: Cambridge University Press.
- Krapp, A. & Weidenmann, B. (Hrsg.). (2006). *Pädagogische Psychologie* (5., vollst. überarb. Aufl.). Weinheim: Beltz PVU.

- Krauss, S., Kunter, M., Brunner, M. & Baumert, J. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsentwicklung* (S. 31–53). Münster: Waxmann.
- Kuger, S. & Kluczniok, K. (2008). Prozessqualität im Kindergarten – Konzept, Umsetzung und Befunde. *Zeitschrift für Erziehungswissenschaft*, 10 (Sonderheft 11), 159–178. https://doi.org/10.1007/978-3-531-91452-7_11
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Laewen, H.-J. (2013). Funktionen der institutionellen Früherziehung: Bildung, Erziehung, Betreuung, Prävention. In L. Fried & S. Roux (Hrsg.), *Handbuch Pädagogik der frühen Kindheit* (S. 96–107). Berlin: Cornelsen.
- Leinhardt, G. (1983). Novice and expert knowledge of individual student's achievement. *Educational Psychologist*, 18(3), 165–179. <https://doi.org/10.1080/00461528309529272>
- Leu, H. R. (2013). Beobachtung in der Praxis. In L. Fried & S. Roux (Hrsg.), *Handbuch Pädagogik der frühen Kindheit* (S. 250–261). Berlin: Cornelsen.
- Leygraf, J. (2013). *Fachberatung in Deutschland. Eine bundesweite Befragung von Fachberaterinnen und Fachberatern für Kindertageseinrichtungen: Zehn Fragen – Zehn Antworten* (WiFF-Studien, Bd. 20). München: Deutsches Jugendinstitut e.V.. Verfügbar unter https://www.weiterbildungsinitiative.de/fileadmin/Redaktion/Publikationen/old_uploads/media/Studie_Leygraf_Fachberatung_web.pdf
- Liebers, K., Schmidt, C., Junger, R. & Prengel, A. (2019). Formatives Assessment in der inklusiven Grundschule im Spannungsfeld von Wissenschaft und Transfer. In C. Donie, F. Foerster, M. Obermayr, A. Deckwerth, G. Kammermeyer, G. Lenske et al. (Hrsg.), *Grundschulpädagogik zwischen Wissenschaft und Transfer* (Jahrbuch Grundschulforschung, Bd. 23, S. 303–312). Wiesbaden: VS Verlag.
- Lindmeier, A., Seemann, S., Kuratli-Geeler, S., Wullschleger, A., Dunekacke, S., Leuchter, M. et al. (2020). Modelling early childhood teachers' mathematics-specific professional competence and its differential growth through professional develop-

- ment – an aspect of structural validity. *Research in Mathematics Education*, 22(2), 168–187. <https://doi.org/10.1080/14794802.2019.1710558>
- Lingelbach, H. (1995). *Unterrichtsexpertise von Grundschullehrkräften* (Schriftenreihe Erziehung – Unterricht – Bildung, Bd. 35). Hamburg: Kovač.
- Lintorf, K., McElvany, N., Rjosk, C., Schroeder, S., Baumert, J., Schnotz, W. et al. (2011). Zuverlässigkeit von diagnostischen Lehrerurteilen – Reliabilität verschiedener Urteilsmaße bei der Einschätzung von Aufgabenschwierigkeiten. Reliability of diagnostic judgements by teachers when assessing task difficulty. *Unterrichtswissenschaft*, 39(2), 102–120.
- Lonnemann, J. & Hasselhorn, M. (2018). Frühe mathematische Bildung. *Frühe Bildung*, 7(3), 129–134. <https://doi.org/10.1026/2191-9186/a000379>
- Lorenz, C. (2011). *Diagnostische Kompetenz von Grundschullehrkräften. Strukturelle Aspekte und Bedingungen*. Bamberg: University of Bamberg Press.
- Lutz, T. (2007). Lern- und Leistungsdiagnostik. In T. Fleischer, N. Grewe, B. Jötten, K. Seifried & B. Sieland (Hrsg.), *Handbuch Schulpsychologie. Psychologie für die Schule* (S. 82–97). Stuttgart: Kohlhammer.
- Maier, U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, 13(2), 293–308. <https://doi.org/10.1007/s11618-010-0124-9>
- Maier, U. (2015). *Leistungsdiagnostik in Schule und Unterricht. Schülerleistungen messen, bewerten und fördern* (Studientexte Bildungswissenschaft). Bad Heilbrunn: Klinkhardt.
- Martschinke, S. (2015). Facetten adaptiven Unterrichts aus der Sicht der Unterrichtsforschung. In K. Liebers, B. Landwehr, A. Marquardt & K. Schlotter (Hrsg.), *Lernprozessbegleitung und adaptives Lernen in der Grundschule. Forschungsbezogene Beiträge* (Jahrbuch Grundschulforschung, Bd. 19, S. 15–32). Wiesbaden: VS Verlag.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (Beltz Pädagogik, 12., aktual. und überarb. Aufl.). Weinheim: Beltz.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W. et al. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 223–235.

- Miller, D. & Lavin, F. (2007). 'But now I feel I want to give it a try': formative assessment, self-esteem and a sense of competence. *Curriculum Journal*, 18(1), 3–25.
<https://doi.org/10.1080/09585170701292109>
- Ministerium für Bildung, Rheinland-Pfalz (2018). *Bildungs- und Erziehungsempfehlungen für Kindertagesstätten in Rheinland-Pfalz plus Qualitätsempfehlungen* (Frühe Kindheit, 4. Aufl.). Berlin: Cornelsen.
- Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur, Rheinland-Pfalz. (2011). *Lehrplan für die Fachschule Sozialwesen*. Speyer: Pädagogisches Landesinstitut.
- Moser, U. & Bayer, N. (2011). „Wortgewandt & zahlenstark“: Ein Diagnostikum zur Erfassung sprachlicher und mathematischer Vorläuferfertigkeiten in der Schuleingangsstufe. In M. Hasselhorn & W. Schneider (Hrsg.), *Frühprognose schulischer Kompetenzen* (Tests und Trends, N.F., 9, Jahrbuch der pädagogisch-psychologischen Diagnostik, S. 147–166). Göttingen: Hogrefe.
- Moser, U. & Berweger, S. (2007a). *wortgewandt & zahlenstark. Lern- und Entwicklungsstand bei 4- bis 6- Jährigen (Testinstrumente)*. Rorschach: Kantonaler Lehrmittelverlag St. Gallen.
- Moser, U. & Berweger, S. (2007b). *Wortgewandt & zahlenstark. Lern- und Entwicklungsstand bei 4- bis 6-Jährigen (Testhandbuch)*. Rorschach: Kantonaler Lehrmittelverlag St. Gallen.
- Moser Opitz, E. (2006). Förderdiagnostik: Entstehung – Ziele – Leitideen – Beispiele. In M. Grüßing & A. Peter-Koop (Hrsg.), *Die Entwicklung mathematischen Denkens in Kindergarten und Grundschule. [Beobachten – Fördern – Dokumentieren]* (3. Aufl., S. 10–28). Offenburg: Mildenerberger.
- Murphy, K. R., Garcia, M., Kerkar, S., Martin, C. & Balzer, W. K. (1982). Relationship between observational accuracy and accuracy in evaluating performance. *Journal of Applied Psychology*, 67(3), 320–325. <https://doi.org/10.1037/0021-9010.67.3.320>
- Oerke, B., McElvany, N., Ohle, A., Ullrich, M. & Horz, H. (2016). Verbessert sich die diagnostische Urteilsgenauigkeit von Lehrkräften bei längerem Kontakt mit der Klasse? *Psychologie in Erziehung und Unterricht*, 63(1), 34–47.
- Oesterhelt, V. (2018). Förderung der Beobachtungskompetenz in der Sportlehrerbildung. In N. Ukley & B. Gröben (Hrsg.), *Forschendes Lernen im Praxissemester. Begründungen, Befunde und Beispiele aus dem Fach Sport* (Bildung und Sport, Bd. 13, 1. Aufl., S. 81–100). Wiesbaden: VS Verlag.

- Ohle, A., McElvany, N., Horz, H. & Ullrich, M. (2017). Aspekte diagnostischer Kompetenz im Projekt „BiTe“. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 29–33). Münster: Waxmann.
- Pauen, S. & Pahnke, J. (2008). Mathematische Kompetenzen im Kindergarten. Evaluation der Effekte einer Kurzzeitintervention. *Empirische Pädagogik*, 22(2), 193–208.
- Pellegrino, J. W., Chudowsky, N. & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: The National Academies Press.
- Petermann, F. & Viernickel, S. (2012). Schwerpunkt: Entwicklungsbeobachtung und Entwicklungsdokumentation. *Frühe Bildung*, 1(2), 61–63.
<https://doi.org/10.1026/2191-9186/a000029>
- Pfennigwerth, S., Dunekacke, S., Heinze, A., Kuratli, S., Leuchter, M., Lindmeier, A. et al. (2016). Effekte fachspezifischer Erzieherinnenkompetenz auf den Kompetenzzuwachs 4-6-jähriger Kinder. In Gesellschaft für Didaktik der Mathematik (Hrsg.), *Beiträge zum Mathematikunterricht 2016. Vorträge auf der 50. Tagung für Didaktik der Mathematik vom 07.03.2016 bis 11.03.2016 in Heidelberg* (S. 1515–1516). Münster: WTM.
- Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A. & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging self-concepts of students. *The Journal of Educational Research*, 106(1), 64–76.
<https://doi.org/10.1080/00220671.2012.667010>
- Praetorius, A.-K., Hetmanek, A., Herppich, S. & Ufer, S. (2017). Herausforderungen bei der empirischen Erforschung diagnostischer Kompetenz. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 95–149). Münster: Waxmann.
- Praetorius, A.-K., Karst, K., Dickhäuser, O. & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, 58(2), 81–91.
- Praetorius, A.-K., Karst, K. & Lipowsky, F. (2012). Diagnostische Kompetenz von Lehrkräften: Aktueller Forschungsstand, unterrichtspraktische Umsetzbarkeit und Bedeutung für den Unterricht. In R. Lazarides & A. Ittel (Hrsg.), *Differenzierung im*

- mathematisch-naturwissenschaftlichen Unterricht. Implikationen für Theorie und Praxis* (S. 115–146). Bad Heilbrunn: Klinkhardt.
- Praetorius, A.-K. & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 13–18). Münster: Waxmann.
- Preiß, G. (2009). *Leitfaden Zahlenland 2* (3., leicht überarb. Aufl.). Kirchzarten: Verlag Zahlenland Prof. Preiß. Verlag Zahlenland Prof. Preiß.
- Prenzel, A. (2017). Individualisierung in der „Caring Community“ – Zur inklusiven Verbesserung von Lernleistungen. In A. Textor, S. Grüter, I. Schiermeyer-Reichl & B. Streese (Hrsg.), *Leistung inklusive? Inklusion in der Leistungsgesellschaft* (S. 13–27). Bad Heilbrunn: Klinkhardt.
- Racherbäumer, K. (2009). *Hochbegabte Schulanfängerinnen. Eine explorative Längsschnittstudie zum Übergang hochbegabter Kinder vom Kindergarten in die flexible Schuleingangsstufe NRW* (Internationale Hochschulschriften, Bd. 533). Münster: Waxmann.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2014a). *G*Power-Ergänzungen. Kapitel 7: Varianzanalyse mit Messwiederholung*. Zugriff am 08.02.2016. Verfügbar unter https://lehrbuch3.s3.amazonaws.com/files/asset/555a28ee319eb100030001b2/Kapitel_7_GPower_Ergaenzungen_A4.pdf
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2014b). *Quantitative Methoden 1. Einführung in die Statistik* (4. Aufl.). Berlin: Springer.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2014c). *Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Reusser, K. (2011). Von der Unterrichtsforschung zur Unterrichtsentwicklung – Probleme, Strategien, Werkzeuge. In W. Einsiedler (Hrsg.), *Unterrichtsentwicklung und Didaktische Entwicklungsforschung* (S. 11–40). Bad Heilbrunn: Klinkhardt.
- Röbe, E., Huppertz, N. & Füssenich, I. (2010). *WiBeOr. Wissenschaftliche Begleitung und Evaluation zur Implementierung des Orientierungsplans für Erziehung und Bildung in baden-württembergischen Kindergärten. Abschlussbericht*. Zugriff am 11.09.2019. Verfügbar unter <https://www.km-bw.de/site/pbs-bw->

- new/get/documents/KULTUS.Dachmandant/KULTUS/Projekte/kindergaerten-bw/Oplan/KM-KIGA_wissenschaftliche_Begleitung_Abschlussbericht.pdf
- Robinson, J., Myran, S., Strauss, R. & Reed, W. (2014). The impact of an alternative professional development model on teacher practices in formative assessment and student learning. *Teacher Development*, 18(2), 141–162.
<https://doi.org/10.1080/13664530.2014.900516>
- Röhner, C. & König, K. (2015). Professionalisierung im frühpädagogischen Bereich am Beispiel des Projekts „Frühe Chancen – Schwerpunkt-Kitas Sprache und Integration“. In D. Blömer, M. Lichtblau, A.-K. Jüttner, K. Koch, M. Krüger & Werning, R. (Hrsg.), *Perspektiven auf inklusive Bildung. Gemeinsam anders lehren und lernen* (Jahrbuch Grundschulforschung, Bd. 18, S. 99–106). Wiesbaden: VS Verlag.
- Rosenshine, B. (1970). Evaluation of classroom instruction. *Review of Educational Research*, 40(2), 279–300. <https://doi.org/10.3102/00346543040002279>
- Roßbach, H.-G., Sechtig, J. & Freund, U. (2010). *Empirische Evaluation des Modellversuchs „Kindergarten der Zukunft in Bayern – KiDZ“*. *Ergebnisse der Kindergartenphase* (Schriften aus der Fakultät Humanwissenschaften der Otto-Friedrich-Universität Bamberg; 7) Bamberg: University of Bamberg Press. Zugriff am 20.08.2013. Verfügbar unter <http://www.opus-bayern.de/uni-bamberg/volltexte/2010/289/pdf/KiDZopus1Ae.pdf>.
- Roßbach, H.-G. & Tietze, W. (2010). *Kindergarten-Skala-Erweiterung (KES-R-E)*. *Grundlage: Forschungsversion Juni 2007. Mit Korrekturen vom 31. März 2010*, Berlin: PädQUIS.
- Roux, S. (2007). Beobachten und Dokumentieren als Aspekte frühpädagogischer Professionalität. In S. Roux (Hrsg.), *Beobachten und Dokumentieren im Kindergarten. PISA und die Folgen* (VEP-Aktuell, Bd. 11, S. 3–13). Landau: Verlag Empirische Pädagogik.
- Ruiz-Primo, M. A. & Furtak, E. M. (2007). Exploring teachers’ informal formative assessment practices and students’ understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84.
<https://doi.org/10.1002/tea.20163>
- Ruppig, I., Prigge, J., Pages, S. & Adam, A. (2015). Professionalisierung der pädagogischen Fachkräfte in Kindertagesstätten durch Beobachtung und Dokumentation. Möglichkeiten und Grenzen von Fort- und Weiterbildungen. In I. Ruppig (Hrsg.), *Professionalisierung in Kindertagesstätten* (S. 140–162). Weinheim: Beltz Juventa.

- Satterly, D. (1981). *Assessment in schools*. Oxford: Blackwell.
- Schäfer, G. E. (2010). Frühkindliche Bildungsprozesse in ethnographischer Perspektive. Zur Begründung und konzeptionellen Ausgestaltung einer pädagogischen Ethnographie der frühen Kindheit. In G. E. Schäfer & R. Staege (Hrsg.), *Frühkindliche Lernprozesse verstehen. Ethnographische und phänomenologische Beiträge zur Bildungsforschung* (S. 69–90). Weinheim: Beltz Juventa.
- Schendera, C. F. G. (2008). *Regressionsanalyse mit SPSS*. München: Oldenbourg Wissenschaftsverlag.
- Schmidt, C. A. (2020). *Formatives Assessment in der Grundschule. Konzept, Einschätzungen der Lehrkräfte und Zusammenhänge*. Wiesbaden: VS Verlag. Verfügbar unter <https://link.springer.com/content/pdf/10.1007%2F978-3-658-26921-0.pdf>
- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik* (5. Aufl.). Berlin: Springer Berlin Heidelberg.
- Schneider, W., Küspert, P. & Krajewski, K. (2016). *Die Entwicklung mathematischer Kompetenzen* (UTB, Bd. 3899, 2., aktual. u. erw. Aufl.). Paderborn: Ferdinand Schöningh.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts* (Europäische Hochschulschriften: Reihe 6, Psychologie, Bd. 289). Frankfurt am Main: Lang.
- Schrader, F.-W. (1997). Lern- und Leistungsdiagnostik im Unterricht. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie. Psychologie des Unterrichts und der Schule* (Themenbereich D, Serie 1, Bd. 3, S. 659–699). Göttingen: Hogrefe.
- Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (3., überarb. u. erw., S. 95–100). Weinheim: Beltz, PVU.
- Schrader, F.-W. (2008). Diagnoseleistungen und diagnostische Kompetenzen von Lehrkräften. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der pädagogischen Psychologie* (S. 168–177). Göttingen: Hogrefe.
- Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 237–245. <https://doi.org/10.1024/1010-0652.23.34.237>
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerbildung*, 31(2), 154–165.

- Schrader, F.-W. (2017). Diagnostische Kompetenz von Lehrkräften – Anmerkungen zur Weiterentwicklung des Konstrukts. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 247–255). Münster: Waxmann.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27–52.
- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45–58). Weinheim: Beltz.
- Schrader, F.-W. & Praetorius, A.-K. (2018). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost, J. R. Sparfeldt & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5., überarb. u. erw. Aufl., S. 92–98). Weinheim: Beltz.
- Schulz, M. & Cloos, P. (2013). Beobachtung und Dokumentation von Bildungsprozessen. In M. Stamm & D. Edelmann (Hrsg.), *Handbuch Frühkindliche Bildungsforschung* (S. 787–800). Wiesbaden: VS Verlag. https://doi.org/10.1007/978-3-531-19066-2_54
- Schütze, B., Souvignier, E. & Hasselhorn, M. (2018). Stichwort – Formatives Assessment. *Zeitschrift für Erziehungswissenschaft*, 21(4), 697–715. <https://doi.org/10.1007/s11618-018-0838-7>
- Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagné & M. Scriven (Hrsg.), *Perspectives of curriculum evaluation* (S. 39–83). Chicago: Rand McNally.
- Scriven, M. (1972). Die Methodologie der Evaluation. In C. Wulf (Hrsg.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München: R. Piper & Co. Verlag. Verfügbar unter https://www.pedocs.de/volltexte/2009/1423/pdf/Scriven_Methodologie_EvaluationD_A.pdf
- Seemann, S., Dunekacke, S., Lindmeier, A., Heinze, A., Leuchter, M., Moser Opitz, E. et al. (2018). Anforderungsbezogene Modellierung und Erfassung domänenspezifischer professioneller Kompetenz frühpädagogischer Fachkräfte. In U. Kortenkamp & A. Kuzle (Hrsg.), *Beiträge zum Mathematikunterricht 2017. 51. Jahrestagung der Gesellschaft für Didaktik der Mathematik* (S. 1225–1228). Münster: WTM. Verfüg-

- bar unter <https://eldorado.tu-dortmund.de/bitstream/2003/36696/1/BzMU-2017-SEEMANN-2.pdf>
- Seidel, T., Blomberg, G. & Stürmer, K. (2010). „Observer“ – Validierung eines video-basierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. Zeitschrift für Pädagogik*, 56. Beiheft, S. 296–306 [Themenheft]. Weinheim: Beltz.
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A. et al. (2008). On the impact of curriculum-embedded formative assessment on learning. A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21(4), 295–314.
<https://doi.org/10.1080/08957340802347647>
- Sherin, M. G. & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60(1), 20–37.
- Siraj-Blatchford, I. (2007). Effektive Bildungsprozesse: Lehren in der frühen Kindheit. In F. Becker-Stoll & M. R. Textor (Hrsg.), *Die Erzieherin-Kind-Beziehung. Zentrum von Bildung und Erziehung* (S. 97–114). Berlin: Cornelsen Scriptor.
- Smidt, W. (2012). *Zielkindbezogene pädagogische Qualität im Kindergarten. Eine empirisch-quantitative Studie* (Internationale Hochschulschriften, Bd. 564). Münster, New York: Waxmann.
- Smit, R. (2009). *Die formative Beurteilung und ihr Nutzen für die Entwicklung von Lernkompetenz. Eine empirische Studie in der Sekundarstufe I* (Schul- und Unterrichtsforschung, Bd. 10). Baltmannsweiler: Schneider- Verlag Hohengehren.
- Smit, R. & Engeli, E. (2017). Formative Beurteilung im jahrgangsübergreifenden Unterricht. *Zeitschrift für Erziehungswissenschaft*, 20(2), 279–303.
- Solzbacher, C. (2012). *Jedem Kind gerecht werden? Sichtweisen und Erfahrungen von Grundschullehrkräften* (Praxiswissen Unterricht). Köln: Link.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1–2), 85–95. <https://doi.org/10.1024/1010-0652.19.12.85>
- Steinweg, S. (2007). Mit Kindern Mathematik erleben. Aktivitäten und Organisationsideen sowie Beobachtungsvorschläge zur mathematischen Bildung der Drei- bis Sechsjährigen. In Stiftung Bildungspakt Bayern (Hrsg.), *Das KIDZ Handbuch*.

- Grundlagen, Konzepte und Praxisbeispiele aus dem Modellversuch „KIDZ – Kindergarten der Zukunft in Bayern“* (S. 136–203). Köln: Wolters Kluwer.
- Stiftung Bildungspakt Bayern (Hrsg.). (2007). *Das KIDZ Handbuch. Grundlagen, Konzepte und Praxisbeispiele aus dem Modellversuch „KIDZ – Kindergarten der Zukunft in Bayern“*. Köln: Wolters Kluwer.
- Stiggins, R. J. & Chappuis, J. (2012). *An introduction to student-involved assessment for learning* (6. Aufl.). Boston: Pearson.
- Stürmer, K. & Seidel, T. (2017). A standardized approach for measuring teachers' professional vision: The observer research tool. In E. O. Schack, M. H. Fisher & J. A. Wilhelm (Hrsg.), *Teacher noticing. Bridging and broadening perspectives, contexts, and frameworks* (Research in mathematics education, Bd. 336, S. 359–380). Cham: Springer. https://doi.org/10.1007/978-3-319-46753-5_21
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>
- Südkamp, A., Kaiser, J. & Möller, J. (2017). Ein heuristisches Modell der Akkuratheit diagnostischer Urteile von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl., S. 33–38). Münster: Waxmann.
- Südkamp, A., Möller, J. & Pohlmann, B. (2008). Der Simulierte Klassenraum. Ein Instrument zur Untersuchung von diagnostischer Kompetenz. In E.-M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 87–97). Münster: Waxmann.
- Südkamp, A. & Praetorius, A.-K. (Hrsg.). (2017). *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 94, 1. Aufl.). Münster: Waxmann.
- Sulsky, L. M. & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology*, 73(3), 497–506. <https://doi.org/10.1037/0021-9010.73.3.497>
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., Taggart, B. & Elliot, K. (2004). The effective provision of pre-school education project – Zu den Auswirkungen vorschulischer Einrichtungen in England. In G. Faust, M. Götz, H. Hacker & H.-

- G. Roßbach (Hrsg.), *Anschlussfähige Bildungsprozesse im Elementar- und Primarbereich* (S. 154–167). Bad Heilbrunn: Klinkhardt.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. & Taggart, B. (2004). *The Effective Provision of Pre-School Education (EPPE) Project: Findings from Pre-school to end of Key Stage 1*. Zugriff am 28.02.2016. Verfügbar unter <http://www.dotwaidecentre.org.au/pdf/EPPE.pdf>
- Textor, M. R. (2000). Lew Wygotski. In W. E. Fthenakis & M. R. Textor (Hrsg.), *Pädagogische Ansätze im Kindergarten. Online Version* (Jahrbuch der Frühpädagogik und Kindheitsforschung, Bd. 3, S. 67–78). Weinheim: Beltz.
- Thole, W., Göbel, S., Milbradt, B., Reißmann, M. & Wedtstein, M. (2015). Thematisierungsweisen pädagogischer Praxis in Kindertageseinrichtungen. In A. König (Hrsg.), *Forschungsperspektiven auf Professionalisierung in der Frühpädagogik. Empirische Befunde der AWiFF-Förderlinie* (Perspektive Frühe Bildung, Bd. 2, S. 124–143). Weinheim: Beltz Juventa.
- Thonhauser, J. (2005). Diagnostische Kompetenzen von Lehrerinnen und Lehrern und die aktuelle Diskussion über Standards – Ein Beitrag über Österreich. In H. Döbert (Hrsg.), *Leistungsmessungen und Innovationsstrategien in Schulsystemen. Ein internationaler Vergleich* (Studien zur international und interkulturell vergleichenden Erziehungswissenschaft, Bd. 6, S. 103–124). Münster: Waxmann.
- Tietze, W. & Roßbach, H.-G. (2010). *Kindergarten-Skala-Zusatzmerkmale (KES-Z). Forschungsversion 2008.09.16A ÜA 20100422*, Berlin: PädQUIS.
- Titz, C., Ropeter, A. & Hasselhorn, M. (2018). Ausgangslagen erfassen und Veränderungen dokumentieren: Zum Mehrwert von Diagnostik. In C. Titz, S. Geyer, A. Ropeter, H. Wagner, S. Weber, M. Hasselhorn et al. (Hrsg.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln* (Bildung durch Sprache und Schrift, Bd. 1, S. 87–100). Stuttgart: Kohlhammer.
- Uggerslev, K. L. & Sulsky, L. M. (2008). Using frame-of-reference training to understand the implications of rater idiosyncrasy for rating accuracy. *The Journal of Applied Psychology*, 93(3), 711–719. <https://doi.org/10.1037/0021-9010.93.3.711>
- Van Es, E. & Sherin, M. G. (2002). Learning to notice: Scaffolding new Teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571–596.
- Van Es, E. & Sherin, M. G. (2006). How different video club designs support teachers in “learning to notice”. *Journal of Computing in Teacher Education*, 22(4), 125–135.

- Van Ophuysen, S. (2006). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahnpfehlung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38(4), 154–161.
<https://doi.org/10.1026/0049-8637.38.4.154>
- Van Ophuysen, S. & Lintorf, K. (2013). Pädagogische Diagnostik im Schulalltag. In S.-I. Beutel, W. Bos & R. Porsch (Hrsg.), *Lernen in Vielfalt. Chance und Herausforderung für Schul- und Unterrichtsentwicklung* (S. 55–76). Münster: Waxmann.
- Viernickel, S. (2011). Beobachtungs- und Dokumentationsverfahren. Einige Reflexionen über ihren Beitrag zur frühpädagogischen Professionalisierung. In P. Cloos & M. Schulz (Hrsg.), *Kindliches Tun beobachten und dokumentieren. Perspektiven auf die Bildungsbegleitung in Kindertageseinrichtungen* (Kindheitspädagogische Beiträge, S. 202–220). Weinheim: Beltz Juventa.
- Viernickel, S., Nentwig-Gesemann, I., Nicolai, K., Schwarz, S. & Zenker, L. (2013). *Schlüssel zu guter Bildung, Erziehung und Betreuung. Bildungsaufgaben, Zeitkontingente und strukturelle Rahmenbedingungen in Kindertageseinrichtungen. Forschungsbericht* (1. Aufl.). Berlin: Der Paritätische Gesamtverband.
- Viernickel, S. & Völkel, P. (2009). *Beobachten und Dokumentieren im pädagogischen Alltag* (1. Aufl. d. vollst. überarb. u. erw. Neuausg.; 5. Gesamtaufl.). Freiburg: Herder.
- Viernickel, S., Voss, A. & Mauz, E. (2017). *Arbeitsplatz Kita. Belastungen erkennen, Gesundheit fördern* (1. Aufl.). Weinheim: Beltz Juventa.
- Vygotskij, L. S. (2002). *Denken und Sprechen. Psychologische Untersuchungen* (Orig.-Ausg.). Weinheim: Beltz.
- Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press (Original erschienen 1930).
- Walter-Laager, C., Pfiffner, M. R. & Schwarz, J. (2011). *Beobachten und Dokumentieren in der Elementarpädagogik. Erste Resultate aus dem internationalen Forschungsprogramm KiDiT (KiDiT)*. Universität Oldenburg: Didaktisches Zentrum.
- Weinert, F. E. & Schrader, F.-W. (1986). Diagnose des Lehrers als Diagnostiker. In H. Petillon, J. W. L. Wagner & B. Wolf (Hrsg.), *Schülergerechte Diagnose. Theoretische und empirische Beiträge zur Pädagogischen Diagnostik. Festschrift zum 60. Geburtstag von Karlheinz Ingenkamp* (S. 11–29). Weinheim: Beltz.

- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1990). Educational expertise: Closing the gap between educational research and classroom practice. *School Psychology International*, 11(3), 163–180. <https://doi.org/10.1177/0143034390113002>
- Werning, R. (2007). Pädagogische Beobachtungskompetenz statt Diagnostik. Perspektiven für die Begabungsförderung aus Sicht der Förderpädagogik. In G. Guttenberger & B. Husmann (Hrsg.), *Begabt für Religion. Religiöse Bildung und Begabungsförderung* (S. 31–45). Göttingen: Vandenhoeck & Ruprecht.
- Werning, R. & Lichtblau, M. (2012). Sonderpädagogische Diagnostik. In R. Werning, R. Balgo, W. Palmowski & M. Sassenroth (Hrsg.), *Sonderpädagogik. Lernen, Verhalten, Sprache, Bewegung und Wahrnehmung* (2., aktual. Aufl., S. 229–259). München: Oldenbourg.
- Wild, K.-P. & Rost, D. H. (1995). Klassengröße und Genauigkeit von Schülerbeurteilungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 27(1), 78–90.
- Wildgruber, A. (2010). *Kompetenzen von Erzieherinnen im Prozess der Beobachtung kindlicher Bildung und Entwicklung* (Münchner Beiträge zur Bildungsforschung, Bd. 20). München: Herbert Utz Verlag.
- Wilhelm, O. & Kunina-Habenicht, O. (2015). Pädagogisch-psychologische Diagnostik. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Springer-Lehrbuch, 2., vollst. überarb. u. aktual. Aufl., S. 305–328). Berlin: Springer.
- Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Hrsg.), *Handbook of formative assessment* (S. 18–40). New York, NY: Routledge.
- Winsler, A. (2003). Observations of children's task activities and social interactions in relation to teacher perceptions in a child-centered preschool: Are we leaving too much to chance? *Early Education & Development*, 14(2), 155–178.
- Winter, F. (2010). Diagnostische Kompetenzen unmittelbar im Dienst des Lernens. *Schule NRW: Amtsblatt des Ministeriums für Schule und Weiterbildung*, 62, 494–496.
- Winter, F. (2015). *Lerndialog statt Noten. Neue Formen der Leistungsbeurteilung* (Pädagogik praxis). Weinheim: Beltz.
- Wirsing, K. (2010). Wahrnehmung. In R. Pousset (Hrsg.), *Handwörterbuch für Erzieherinnen und Erzieher* (2. Aufl., S. 462–464). Berlin: Cornelsen Scriptor.

- Wittmann, E. C. (2009). Das mathe 2000-Frühförderprogramm. In S. Pauen & V. Herber (Hrsg.), *Vom Kleinsein zum Einstein* (1. Aufl., S. 54–66). Berlin: Cornelsen Scriptor.
- Wittmann, E. C. & Deutscher, T. (2013). Mathematische Bildung. In L. Fried & S. Roux (Hrsg.), *Handbuch Pädagogik der frühen Kindheit* (S. 210–216). Berlin: Cornelsen.
- Wittmann, E. C. & Müller, G. N. (2009). *Das Zahlenbuch. Handbuch zum Frühförderprogramm* (1. Aufl.). Stuttgart: Klett.
- Wood, D., Bruner, J. S. & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100.
<https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Woolfolk, A. (2008). *Pädagogische Psychologie* (Psychologie, 10. Aufl. [1. dt. Ausgabe, Übersetzung der 10. amerikanischen Original-Ausgabe]). München: Pearson Studium.
- Wright, D. & Wiese, M. J. (1988). Teacher judgment in student evaluation: A comparison of grading methods. *The Journal of Educational Research*, 82(2), 10–14.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M. et al. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21(4), 335–359. <https://doi.org/10.1080/08957340802347845>

Anlagen

Tabelle 28: Übersicht der Erfassung von Rahmendaten (Erzieherinnenbefragungen), Domäne Mathematik

Bereich	Frage /Aufgabe	Testname, Erhebungszeitpunkt (Aufgabentypus)
Gruppenstruktur (zu t_1)	In welcher Gruppenstruktur arbeiten Sie mit den Vorschulkindern? (geschlossene, offene oder Mischform?) Bitte erläutern Sie kurz:	BeUrteilEr t_1 (offen)
Arbeitszeit mit Gruppe (zu t_1)	Wie viele Stunden nutzen Sie pro Woche für die Arbeit mit den Vorschulkindern?	BeUrteilEr t_1 (offen)
Arbeitszeit mit Zielkind und Gründe für Abweichung (zu t_1)	Beschäftigen Sie sich mit dem genannten Kind öfter oder seltener als durchschnittlich üblich? Bitte erläutern Sie kurz die Gründe.	BeUrteilEr t_1 (offen)
Gruppenstruktur (zu t_2)	Die schriftsprachliche und mathematische Förderung dieses Kindes findet in meiner Gruppe statt. ... findet unter meiner Leitung in einer separaten Vorschulgruppe statt. ... findet unter Leitung einer Kollegin in einer separaten Vorschulgruppe statt. ... findet in folgender Gruppenstruktur statt (Bitte erläutern Sie ggf. kurz)	BeUrteilEr t_2 (Multiple Choice und offen)
Arbeitszeit mit Gruppe (zu t_2)	Wie viele Stunden nutzen Sie pro Woche für die mathematische Arbeit mit den Vorschulkindern?	BeUrteilEr t_2 (offen)
Förderkooperation (zu t_2)	Haben Sie sich bezüglich der mathematischen Förderung dieses Kindes mit KollegInnen beraten? ... nie ... ca. einmal pro Monat ... ca. einmal pro Woche ... mehrmals pro Woche	BeUrteilEr t_2 (Multiple Choice)
Arbeitszeit mit Zielkind und Gründe für Abweichung (zu t_2)	Beschäftigen Sie sich mit dem genannten Kind öfter oder seltener als durchschnittlich üblich? Nein; Ja (Wenn ja, erläutern Sie bitte kurz die Gründe)	BeUrteilEr t_2 (dichotom und offen)
Art der individuelle Förderung (zu t_2)	Haben Sie das genannte Kind individuell mathematisch gefördert? Nein; Ja (Wenn ja, auf welche Art und Weise? Bitte erläutern Sie ggf. kurz)	BeUrteilEr t_2 (dichotom/offen)
Geburtsjahr	Geburtsjahr	FiBB zu t_1 & t_2 (offen)
Eigene Kinder	Anzahl eigener Kinder	FiBB zu t_1 & t_2 (offen)
Praxisjahre	Praxisjahre im Kindergarten	FiBB zu t_1 & t_2 (offen)
Abschlüsse	Berufsabschluss/-abschlüsse	FiBB zu t_1 & t_2 (offen)
Fortbildungen	Welche Fortbildungen haben Sie bisher zu den Themen „Mathematik“ oder „Schriftsprache“ (Lesen und Schreiben) besucht?	FiBB zu t_1 & t_2 (offen)
Didaktisches Wissen	Kennen Sie Spiele/Materialien/Aktivitäten, um die Entwicklung von Mathematik im Kindergarten anzuregen? Ja; Nein (Wenn ja, welche?)	FiBB zu t_1 & t_2 (dichotom/offen)
Didaktisches Wissen	Welches ist Ihrer Meinung nach das beste Vorgehen/Material zur mathematischen Förderung im Kindergarten?	FiBB zu t_1 & t_2 (offen)
Selbstwirksamkeitserwartung	Welche Bedeutung haben Sie Ihrer Meinung nach für die mathematische Entwicklung der Kinder? Bitte kreuzen Sie das am ehesten zutreffende Kästchen an.	FiBB zu t_1 & t_2 (6-stufige Likertskala von „eine große Bedeutung“ bis „eine kleine Bedeutung“)
Medieneinsatz	Haben Sie in Ihren Gruppen bewusst Anregungen zur Auseinandersetzung mit Zahlen und mathematischen Tätigkeiten eingerichtet?	FiBB zu t_1 & t_2 (dichotom/offen)
Erwartungen	Sollte ein Vorschulkind am Ende der Kita-Zeit bestimmte mathematische Fähigkeiten besitzen? Ja; Nein (Wenn ja, welche?/Wenn nein, warum?)	FiBB zu t_1 & t_2 (dichotom/offen)
Alltagsförderung	Thematisieren Sie Zahlen oder mathematische Tätigkeiten im Kita-Alltag? Ja; Nein (Wenn ja, wie?)	FiBB zu t_1 & t_2 (dichotom/offen)
Förderkonzept	Bieten Sie den Kindern ihrer Gruppe ein spezielles Mathematik-Förderkonzept an? Ja; Nein (Wenn ja, welche/welches?)	FiBB zu t_1 & t_2 (dichotom/offen)
Fachwissen über mathematische Entwicklung	Die Kinder durchlaufen in ihrer mathematischen Entwicklung Phasen. Die folgenden Beispiele stellen Kinder auf jeweils einer anderen Entwicklungsstufe dar. Bringen Sie die Beispiele durch Nummerieren in die Ihrer Meinung nach richtige Rangreihenfolge.	FiBB zu t_1 & t_2 (Zuordnungsaufgabe)

Tabelle 29: BEURteiler: Übersicht der Bereiche und Items aus *zahlenstark*

Bereiche	Anzahl	Items aus <i>zahlenstark</i>
1. Positionen bestimmen	4	1_1, 1_3, 1_5, 1_7
2. Mengen bestimmen und vergleichen	16	2.1_a-c, 2.3_a-c, 2.4_a-c, 2.6_a-c, 2.7, 2.9, 2.10, 2.12
3. Zahlen benennen	8	3.1_1, 3.1_3, 3.2_1, 3.2_3, 3.2_5, 3.3_1, 3.3_3, 3.3_5,
4. Rechnen mit Bildern	4	4.1, 4.3, 4.5, 4.7
5. Rangreihen mit Bildern	8	5.1_a, 5.1_b, 5.2_a, 5.2_b, 5.3_a, 5.3_b, 5.4_a, 5.4_b
6. Rangreihen mit Zahlen	8	6.1_2, 6.1_4, 6.2_2, 6.2_4, 6.3, 6.4, 6.5, 6.6
7. Rechnen mit Zahlen	22	7.1_1, 7.1_3, 7.2_1, 7.2_3, 7.3_1, 7.3_3, 7.4_1, 7.4_3, 7.5_1, 7.5_3, 7.6_1, 7.6_3, 7.7_1, 7.7_3, 7.8_1, 7.8_3, 7.9_1, 7.9_3, 7.10_1, 7.10_3, 7.11_1, 7.11_3,
Itemsumme pro Instrument	70	

FiBB-Evaluationsbogen (Auszüge, Mathematischer Bereich)



FaBi
Förderung
anschlussfähiger
Bildungsprozesse

Erzieherin-Code:



UNIVERSITÄT
KOBLENZ · LANDAU

Institut für Bildung im
Kindes- und Jugendalter 

Christian Donie
Tel.: 06341-280-33251
Mail: donie@uni-landau.de

**„Schriftsprache und Mathematik
im Kindergarten“**

Teil 2
Filmbasierter Fragebogen für Erzieherinnen

Allgemeine Hinweise

Dieser filmbasierte Fragebogen zu „Schriftsprache und Mathematik im Kindergarten“ enthält offene Fragen und Fragen, die Sie durch Ankreuzen beantworten sollen.

Bitte beantworten Sie die offenen Fragen **stichwortartig**. Falls der Platz zum Beantworten der Fragen nicht ausreicht, nutzen Sie bitte die **Rückseite des Blattes**. Notieren Sie bitte **erst die Nummer der Frage** und dann die Fortsetzung Ihrer Antwort.

Die zutreffenden Antworten können Sie wie folgt **ankreuzen**: oder .
Möchten Sie eine Antwort korrigieren, **kreuzen** Sie bitte **die falsche Antwort ein** und kreuzen die Richtige an .

Abbildung 25: FiBB-Evaluationsbogen, Deckblatt von Teil 2 mit allgemeinen Hinweisen

Beispiel 1 von 8

Im ersten Filmbeispiel sehen Sie ein Vorschulkind, das eine Stiftebox ausgeleert hat, um die Stifte zu sortieren.



Denken Sie bitte an **mathematische Fähigkeiten** von Vorschulkindern, während Sie sich den folgenden kurzen Filmausschnitt anschauen. Sie befinden sich **in der Rolle der Erzieherin und möchten das Kind mathematisch anregen**.

Fragen zu Filmbeispiel 1

B1.1: Welche **mathematischen Fähigkeiten** haben Sie bei dem Kind beobachten können?

(Zur Fortsetzung ggf. Rückseite benutzen.)

B1.2: **Was würden Sie tun**, um das Kind in der eben gezeigten Situation oder im direkten Anschluss daran mathematisch zu fördern?

(Zur Fortsetzung ggf. Rückseite benutzen.)

[1]

Abbildung 26: Beispiel 1 von 8 (1. Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 1)

Tabelle 30: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 1 von 8 (1. Mathematikvignette)

Zeit	Beschreibung des Videos	Kategorien (Unterkategorien eingerückt)
00:01	Ein Junge sitzt am Tisch und breitet Buntstifte vor sich aus.	
00:06	Mit den Worten „Erst ´mal mach ich ´mal Dunkelblau.“ nimmt er die dunkelblauen Holzstifte und legt die sechs Stifte auf die Seite, wobei er bei jedem Stift noch einmal „Dunkelblau“ sagt. Er zeigt auf einen der Stifte und sagt „eins“, als wolle er beginnen zu zählen, schiebt die Stifte dann aber zur Seite.	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Ab-/Nach-Zählen
00:21	Er nimmt die hellblauen Stifte gezielt aus der Menge, hält sie in der Hand und sagt: „Hellblau ist vier“. Die dunkelblauen Stifte zählt er, indem er die Stifte mit dem Finger antippt und laut von eins bis sechs abzählt.	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> ○ Simultane Mengenerfassung • Ab-/Nach-Zählen • Beherrschen (Kennen) von Zahlen
00:36	Die hellblauen Stifte legt er auf einen Haufen neben die dunkelblauen.	
00:38	Er nimmt einen goldenen Stift, schaut nochmal über die übrigen unsortierten Stifte und stellt dann fest, dass es davon nur einen gibt. Er legt ihn beiseite.	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Beherrschen (Kennen) von Zahlen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> ○ Simultane Mengenerfassung ○ nur einer
00:43	Er hat vier Stifte in der Hand und sagt „Nur vier Hautfarben.“	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Beherrschen (Kennen) von Zahlen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> ○ Simultane Mengenerfassung • Beherrschen (Kennen) von Zahlen
00:52	Beim Einsammeln der vier roten Stifte zählt er laut mit.	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Ab-/Nach-Zählen • Beherrschen (Kennen) von Zahlen
1:00	Ebenso bei den drei lila Stiften, die er als rosa bezeichnet.	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Ab-/Nach-Zählen • Beherrschen (Kennen) von Zahlen
1:06	Die beiden schwarzen Stifte legt er mit folgenden Worten beiseite: „Schwarz sind da nur zwei Stück.“	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> ○ Simultane Mengenerfassung • Beherrschen (Kennen) von Zahlen
1:11	Mit den Worten: „Braun. Ein Stück.“ legt er den einzigen braunen Stift beiseite und fügt dann hinzu: „Genauso wie die Gold.“	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> ○ Simultane Mengenerfassung • Beherrschen (Kennen) von Zahlen • Verwenden von Vergleichswörtern/Vergleichen <ul style="list-style-type: none"> ○ genauso viel/gleichviel
1:20	„Die Orange: Ein Stück.“ Er legt die drei einzelnen Stifte nah zueinander (aber nicht auf einen Haufen).	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> ○ Simultane Mengenerfassung

		<ul style="list-style-type: none"> • Beherrschen (Kennen) von Zahlen • Sortieren nach Mengen
1:26	Die Anzahl der drei grünen Stifte erkennt er ebenfalls ohne laut zu zählen.	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> ○ Simultane Mengenerfassung • Beherrschen (Kennen) von Zahlen
1:36	Danach legt er den einzigen hellgrünen Stift beiseite: „Von Hellgrün ein Stück.“	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> ○ Simultane Mengenerfassung • Beherrschen (Kennen) von Zahlen
1:40	Er nimmt einen silberfarbenen Stift, der wie ein Bleistift aussieht, legt ihn zu den schwarzen und sagt, dass schwarz jetzt drei Stück sind und er den Stift vorhin übersehen hat.	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Zuordnen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> ○ Simultane Mengenerfassung • Beherrschen (Kennen) von Zahlen • Ab-/Nach-Zählen <ul style="list-style-type: none"> ○ Zusammenzählen (Addieren)
1:46	Die verbliebenen fünf gelben Stifte zählt er.	<ul style="list-style-type: none"> • Sortieren nach Farben • Sortieren nach Mengen • Ab-/Nach-Zählen • Beherrschen (Kennen) von Zahlen
1:52	Er lässt den Blick über die nach Farben sortierten Stifte schweifen und stellt fest, dass Dunkelblau am meisten ist.	<ul style="list-style-type: none"> • Verwenden von Vergleichswörtern/Vergleichen • Mengenerfassung/Mengen erkennen <ul style="list-style-type: none"> Erkennen der größten Menge <ul style="list-style-type: none"> ○ Mengen erkennen\viel – wenig ○ Simultane Mengenerfassung • Mengenunterscheidung • Verwenden von Vergleichswörtern/Vergleichen <ul style="list-style-type: none"> ○ mehr als/größer als
Ende: 01:59		<ul style="list-style-type: none"> • Merken (Behalten) von Zahlen (Ergebnissen)

Tabelle 31: Kategoriensystem Vignette 1 mit Multiplikator (Expertinnengewichtung)

Nummer	Vignettensnummer \ Kategorie \ Unterkategorie 1 \ Unterkategorie 2 \ ...	Multiplikator
B1.1_01	Stifte sortieren	1
B1.1_02	Stifte sortieren \ Sortieren nach Mengen	1
B1.1_03	Stifte sortieren \ Sortieren nach Farben	1
B1.1_04	Mengenunterscheidung	2
B1.1_05	Mengenerfassung/Mengen erkennen	2
B1.1_06	Mengenerfassung/Mengen erkennen \ viel – wenig	2
B1.1_07	Mengenerfassung/Mengen erkennen \ Erkennen der größten Menge	2
B1.1_08	Mengenerfassung/Mengen erkennen \ Simultane Mengenerfassung	3
B1.1_09	Ab-/Nach-Zählen	1
B1.1_10	Ab-/Nach-Zählen \ Zusammenzählen (Addieren)/Rechnen	1
B1.1_11	Weiter zählen als bis vier	1
B1.1_12	Beherrschen (Kennen) von Zahlen	1
B1.1_13	Verwenden von Vergleichswörtern/Vergleichen	3
B1.1_14	Verwenden von Vergleichswörtern/Vergleichen \ am wenigsten/nur einer	3
B1.1_15	Verwenden von Vergleichswörtern/Vergleichen \ weniger als/kleiner als	3
B1.1_16	Verwenden von Vergleichswörtern/Vergleichen \ mehr als/größer als	3
B1.1_17	Verwenden von Vergleichswörtern/Vergleichen \ genauso viel/gleichviel	3
B1.1_18	Zuordnen	2
B1.1_19	Merken (Behalten) von Zahlen (Ergebnissen)	2
	Sonstige	0
Aus-	Sonstige \ Benennen/Erkennen von Farben	0
schluss		
B1.1_01	Sonstige \ Beschreibung/Interpretation einer Handlung	0
bis	Sonstige \ Arbeitsweise	0
B1.1_05	Sonstige \ Vermutungen/Behauptungen	0

Beispiel 3 von 8

Im dritten Filmbeispiel sehen Sie zwei Vorschulkinder, die gemeinsam am Tisch sitzen und ein Brettspiel spielen.



Denken Sie bitte an **mathematische Fähigkeiten** von Vorschulkindern, während Sie sich den folgenden kurzen Filmausschnitt anschauen. Sie befinden sich **in der Rolle der Erzieherin und möchten die Kinder mathematisch anregen**.

Fragen zu Filmbeispiel 3

B3.1: Welche **mathematischen Fähigkeiten** haben Sie bei den Kindern beobachten können?

(Zur Fortsetzung ggf. Rückseite benutzen.)

B3.2: **Was würden Sie tun**, um das Kind/die Kinder in der eben gezeigten Situation oder im direkten Anschluss daran mathematisch zu fördern?

(Zur Fortsetzung ggf. Rückseite benutzen.)

[5]

Abbildung 27: Beispiel 3 von 8 (2. Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 5)

Tabelle 32: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 3 von 8 (2. Mathematikvignette)

Zeit	Beschreibung des Videos	Kategorien
00:01	Zwei Jungen sitzen an einem Tisch. Vor ihnen liegt ein Brettspiel. K2 fängt an und würfelt mit beiden Würfeln. Er schaut sich die Würfel genau an, sieht in die Kamera und sagt: „Eine Fünf“. K1 schaut ebenfalls auf die Würfel und sagt zeitgleich mit K2 etwas [„Vier“?]. Dann sieht er grinsend zur anderen Seite und sagt schließlich in die Kamera: „Wenn du eine zwei und eine zwei hast, war das vier“.	<ul style="list-style-type: none"> • Zahlwissen <ul style="list-style-type: none"> ○ Zahlen erkennen • Mengenwissen/-erfassung <ul style="list-style-type: none"> ○ simultane Mengenerfassung ○ Mengen erfassen/erkennen ○ (Er)kennen der Würfelaugen/(Er)kennen der Würfelbilder ○ K1 weiß, dass zwei und zwei vier ergibt • Zusammenzählen können (Addieren) <ul style="list-style-type: none"> ○ Zusammenzählen ohne Hilfsmittel ○ Addieren auf Mengenbasis (Fertigkeit/Wissen)
00:17	K2 fällt ihm ins Wort und sagt „drei Figuren? Äh Max, drei Figuren?!“. K1 grinst und antwortet mit „Ja“.	<ul style="list-style-type: none"> • Mengenwissen/-erfassung <ul style="list-style-type: none"> ○ Mengen erfassen/erkennen ○ simultane Mengenerfassung • Vergleichen <ul style="list-style-type: none"> ○ Vergleichen von Mengen ○ Vergleichen von Zahlen • Erkennen der falschen Figuren-Spieler-Zuordnung • Zuordnen
00:25	K2 greift nach der gelben Figur und es erfolgt ein Schnitt. Danach sind nur noch die beiden anderen Figuren im Spiel.	
00:26	K2 rückt fünf Felder mit der roten Figur und zählt dabei laut „1,2,3,4,5“.	<ul style="list-style-type: none"> • Zählen • Abzählen <ul style="list-style-type: none"> ○ K2 Abzählen der Spielfelder • Rücken der gewürfelten Zahl <ul style="list-style-type: none"> ○ Richtiges Rücken der Steine
00:30	K1 beschwert sich: „Hey, ich war doch der Orangene!“. K2: „Du warst der Blaue.“ K1: „Orangene.“ K2: „Blaue.“ K1: „Orangene.“ K2 beendet schließlich die Diskussion: „Egal, sind doch bloß Spielfigur'n!“	
00:41	K1 würfelt mit beiden Würfeln und fragt sich dann: „Fünf oder sechs?“ Er zählt vor sich hin murmelnd die Würfelaugen, indem er jeden Punkt mit dem Finger antippt. Schließlich stellt er fest: „Fünf“. Er schaut verwundert in die Kamera „Immer noch fünf?“. Er überprüft nochmal die Augenzahl, indem er die Augen einzeln laut mitzählt, und kommt dann auf die Anzahl sechs. K2 schaut während dessen mit auf die Würfel und sagt ebenfalls „sechs“.	<ul style="list-style-type: none"> • Zählen <ul style="list-style-type: none"> ○ K1 zählt unsicher (mit Fingern) • Abzählen <ul style="list-style-type: none"> ○ K1 zählt die Würfelaugen (noch) ab • Zusammenzählen können (Addieren) <ul style="list-style-type: none"> ○ Zusammenzählen mit Fingern • Verzählen • Nachzählen
00:55	K1 rückt sechs Felder mit der blauen Spielfigur vor und zählt dabei jeden Schritt laut mit. Seine Figur bleibt ein Feld hinter der roten Spielfigur stehen. K2 tippt die letzten beiden Felder mit	<ul style="list-style-type: none"> • Abzählen <ul style="list-style-type: none"> ○ K2 Abzählen der Spielfelder • Rücken der gewürfelten Zahl <ul style="list-style-type: none"> ○ Falsches Rücken der Steine • Nachzählen

	dem Finger auf dem Brett nach und zählt „fünf, sechs. Ja. Hm“.	
01: 06 Ende: 01:21	K2 würfelt mit beiden Würfeln. Er schaut auf das Spielfeld und sagt: „Ich hab aber eigentlich fünf gehabt, du musst hier sein normal.“ Er nimmt den blauen Spielstein in die Hand und stellt ihn ein Feld vor den roten. Mit der Aussage „egal“ stellt er den blauen Stein wieder hinter den roten.	<ul style="list-style-type: none">• Mengenwissen/-erfassung• Zahlunterschiede erkennen<ul style="list-style-type: none">○ Erkennen, dass sechs größer fünf ist (Kardinalaspekt)• Erkennen der falsch stehenden Figur (5./6. Platz – Ordinalaspekt)• Korrigieren des falschen Zuges<ul style="list-style-type: none">○ Zurücknehmen der Verbesserung

Tabelle 33: Kategoriensystem Vignette 3 mit Multiplikator (Expertinnengewichtung)

Nummer	Vignettenummer \ Kategorie \ Unterkategorie 1 \ Unterkategorie 2 \ ...	Multiplikator
B3.1_01	Nachzählen	1
B3.1_02	Verzählen	1
B3.1_03	Erkennen der falschen Figuren-Spieler-Zuordnung	3
B3.1_04	Zahlunterschiede erkennen	2
B3.1_05	Zahlunterschiede erkennen \ Erkennen, dass sechs größer fünf ist (Kardinalaspekt)	2
B3.1_06	Erkennen der falsch stehenden Figur (5./6.Platz – Ordinalaspekt)	3
B3.1_07	Korrigieren des falschen Zuges	1
B3.1_08	Korrigieren des falschen Zuges \ Zurücknehmen der Verbesserung	1
B3.1_09	Zählen	1
B3.1_10	Zählen \ K1 zählt unsicher (mit Fingern)	1
B3.1_11	Abzählen	1
B3.1_12	Abzählen \ (Ab-)zählen der Würfelaugen	1
B3.1_13	Abzählen \ (Ab-)zählen der Spielfelder	1
B3.1_14	Zusammenzählen können (Addieren)	1
B3.1_15	Zusammenzählen können (Addieren) \ Zusammenzählen ohne Hilfsmittel (Kopfrechnen)	1
B3.1_16	Zusammenzählen können (Addieren) \ Zusammenzählen mit Fingern	1
B3.1_17	Zusammenzählen können (Addieren) \ Addieren auf Mengebasis (Fertigkeit/Wissen)	1
B3.1_18	Zuordnen	1
B3.1_19	Rücken der gewürfelten Zahl	1
B3.1_20	Rücken der gewürfelten Zahl \ falsches Rücken der Steine	1
B3.1_21	Rücken der gewürfelten Zahl \ richtiges Rücken der Steine	1
B3.1_22	Vergleichen \ Vergleichen von Mengen	2
B3.1_23	Mengenwissen/-erfassung \ Mengenerlegung	1
B3.1_24	Mengenwissen/-erfassung \ Eins-zu-Eins-Zuordnung	2
B3.1_25	Mengenwissen/-erfassung \ (Er-)kennen der Würfelaugen/(Er-)kennen der Würfelbilder	2
B3.1_26	Mengenwissen/-erfassung \ K1 weiß, dass zwei und zwei vier ergibt	1
B3.1_27	Mengenwissen/-erfassung \ Mengen erfassen/erkennen	1
B3.1_28	Mengenwissen/-erfassung \ simultane Mengenerfassung	3
B3.1_29	Zahlwissen	1
B3.1_30	Zahlwissen \ Zahlen erkennen	1
Ausschluss	Sonstige \ Beschreibung/Interpretation einer Handlung	0
B3.1_01 bis	Sonstige \ Arbeitsweise	0
B3.1_03	Sonstige \ FALSCHER Beobachtung oder sprachliches Unvermögen	0

Beispiel 5 von 8

Im fünften Filmbeispiel sehen Sie drei Vorschulkinder, die gemeinsam am Tisch sitzen und Zahlen schreiben möchten.



Denken Sie bitte an **mathematische Fähigkeiten** von Vorschulkindern, während Sie sich den folgenden kurzen Filmausschnitt anschauen. Sie befinden sich **in der Rolle der Erzieherin und möchten die Kinder mathematisch anregen**.

Fragen zu Filmbeispiel 5

B5.1: Welche **mathematischen Fähigkeiten** haben Sie bei dem Kind/den Kindern beobachten können?

(Zur Fortsetzung ggf. Rückseite benutzen.)

B5.2: **Was würden Sie tun**, um das Kind/die Kinder in der eben gezeigten Situation oder im direkten Anschluss daran mathematisch zu fördern?

(Zur Fortsetzung ggf. Rückseite benutzen.)

[9]

Abbildung 28: Beispiel 5 von 8 (3. Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 9)

Tabelle 34: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 5 von 8 (3. Mathematikvignette)

Zeit	Beschreibung des Videos	Kategorien (Unterkategorien eingerückt)
00:01	Drei Mädchen sitzen an einem Tisch, auf dem eine Kiste mit Buntstiften steht. Sie haben jeweils ein leeres Blatt vor sich liegen.	
00:03	K1 sagt, dass sie eine riesen Sechs malen will und tut so, als würde sie eine Sechs auf ihr Blatt schreiben, die so groß ist, wie das Blatt selbst. Sie berührt aber mit dem Stift nicht das Blatt.	<ul style="list-style-type: none"> • Üben (der Zahlen) • (Er-)Kennen/Benennen von Zahlen <ul style="list-style-type: none"> ○ K1 kennt die Zahlen/das Zahlenbild der Sechs • Schreiben der Zahl/en <ul style="list-style-type: none"> ○ K1 kann die Zahl Sechs schreiben • Malen der Zahl <ul style="list-style-type: none"> ○ sie versuchen die Zahl Sechs zu malen
00:06	K3 fragt: „Wie geht eine Sechs?“ K2 sagt „Pst“, legt den Finger an den Mund und zeigt in die Richtung, aus der das Telefonklingeln kommt. K1 sagt: „Soo rum. Guck, ich zeig dir’s.“ und tut noch einmal so, als würde sie eine große Sechs schreiben. Dabei sagt sie noch einmal „Riesen-Sechs“.	<ul style="list-style-type: none"> • K1 zeigt/erklärt (K3), wie die Ziffer Sechs zu schreiben ist • (Er-)kennen/Benennen von Zahlen • Schreiben der Zahl/en <ul style="list-style-type: none"> ○ K1 kann die Zahl Sechs schreiben ○ K3 konnte die Sechs ohne Hilfe nicht schreiben • Malen der Zahl <ul style="list-style-type: none"> ○ sie versuchen, die Zahl Sechs zu malen
00:13	K3 versucht, es genau so nachzumachen (ebenfalls, ohne wirklich zu schreiben), malt die Sechs aber spiegelverkehrt. K1 zeigt es ihr nochmal auf dem Blatt von K3, die daraufhin die Sechs in der Luft über dem Blatt richtig herum schreibt.	<ul style="list-style-type: none"> • K1 zeigt/erklärt (K3), wie die Ziffer Sechs zu schreiben ist • (Er-)kennen/Benennen von Zahlen • Schreiben der Zahl/en <ul style="list-style-type: none"> ○ K1 kann die Zahl Sechs schreiben ○ K3 konnte die Sechs ohne Hilfe nicht schreiben ○ Schreiben der Zahl/en \ K2+3 brauchen beim Schreiben der Zahlen noch Hilfe • Malen der Zahl <ul style="list-style-type: none"> ○ sie versuchen, die Zahl Sechs zu malen • Raumlagekenntnisse (z.B. spiegelverkehrtes Zeigen) <ul style="list-style-type: none"> ○ Erkennen/Wahrnehmen der richtigen Raumlage ○ K3 wollte die Sechs erst spiegelverkehrt schreiben ○ K1 erkennt (spiegelverkehrt), dass K3 Sechs falsch schreiben wollte
00:23	K3 malt die Sechs richtig auf ihr Blatt und lächelt beim Betrachten des Ergebnisses. Sie schaut vom Blatt auf und sagt etwas – anscheinend zu einer Person, die sich nicht im Bild befindet. Schließlich schaut sie, was die anderen Mädchen gerade schreiben.	<ul style="list-style-type: none"> • Üben (der Zahlen) [ohne weiteres Verb] • Schreiben der Zahl/en <ul style="list-style-type: none"> ○ K3 konnte die Sechs ohne Hilfe nicht schreiben • Malen der Zahl <ul style="list-style-type: none"> ○ sie versuchen, die Zahl Sechs zu malen • Raumlagekenntnisse (z.B. spiegelverkehrtes zeigen) <ul style="list-style-type: none"> ○ Erkennen/Wahrnehmen der richtigen Raumlage
00:30	K1 malt auch die Sechs auf ihr Blatt. Die anderen beiden sehen ihr dabei zu	<ul style="list-style-type: none"> • Üben (der Zahlen) [ohne weiteres Verb] • (Er-)Kennen/Benennen von Zahlen <ul style="list-style-type: none"> ○ K1 kennt die Zahlen/das Zahlenbild der Sechs • Schreiben der Zahl/en <ul style="list-style-type: none"> ○ K1 kann die Zahl(en) Sechs schreiben • Malen der Zahl <ul style="list-style-type: none"> ○ sie versuchen, die Zahl Sechs zu malen
00:40	K2 sagt, sie wisse nicht, welche Zahl sie malen solle. K1 meint, sie solle auch eine	<ul style="list-style-type: none"> • Üben (der Zahlen) [ohne weiteres Verb] • (Er-)Kennen/Benennen von Zahlen <ul style="list-style-type: none"> ○ Erfassen von Zahlen

	<p>Sechs malen, da sie ja auch sechs Jahre alt sei. K3 stimmt zu: „Ja, jeder ist sechs Jahre alt von uns.“ Aus dem Hintergrund kommt der Einwand, K1 sei bereits sieben Jahre alt ist, woraufhin K3 erstaunt schaut und K1 sagt: „Egaaal!“.</p>	<ul style="list-style-type: none"> ○ Benennen von Zahlen • Alter und Zahl <ul style="list-style-type: none"> ○ Alter und Zahl in Verbindung bringen ○ Kennen ihr Alter • Vergleichen von Zahlen <ul style="list-style-type: none"> ○ Erkennen von mehr und weniger (z.B. Altersvergleich)
00:46	<p>Es erfolgt ein Schnitt. Eines der Mädchen sagt „Sechzehn“. Ein anderes sagt „Siebzehn“. K2 beschwert sich daraufhin und schaut dann auf ihr Blatt. K3 beginnt die Zahlen ab 16 laut zu zählen, K1 zählt ab 18 mit. K1 zählt dabei flüssig und spricht deutlich. K3 nuschelt immer wieder einmal, als wäre sie nicht ganz sicher, welche Zahl als nächstes kommt. Außerdem steigt sie bei Zahl 28 aus. K1 zählt bis 33. An dieser Stelle endet das Video.</p>	<ul style="list-style-type: none"> • Üben (der Zahlen) [ohne weiteres Verb] • (Er-)Kennen/Benennen von Zahlen <ul style="list-style-type: none"> ○ K2 kennt wenig Zahlen ○ Erfassen von Zahlen • Schreiben der Zahl/en <ul style="list-style-type: none"> ○ K2 schreibt eine Zahl (sehr klein) auf ○ K2 konnte die Zahlen nicht (so gut) schreiben • Malen der Zahl <ul style="list-style-type: none"> ○ K2 ist unsicher beim Malen • Zählen <ul style="list-style-type: none"> ○ Zählen ab 16 (bis 32) ○ K1 und K3 zählen zusammen ○ K3 kann noch nicht so weit zählen wie K1 ○ K2 konnte kaum zählen ○ K2 konnte kaum zählen <ul style="list-style-type: none"> ▪ K2 zieht sich beim Zählen zurück ○ K1 ist fit im Zählen ○ Zählen über 10 hinaus ○ Zählen bis 20 (K3 zählt sicher bis 20) ○ Zählen über 20 hinaus <ul style="list-style-type: none"> ▪ K3 zählt bis ca. 24 → (ungenau) ○ zählen bis 30 <ul style="list-style-type: none"> ▪ K1 (und K3 zählen) zählt sicher bis 28/29 ▪ K3 zählt bis 29/30 durch Nachsprechen von K1 ▪ K3 steigt beim Zählen bei 30 aus ○ Zählen über 30 hinaus <ul style="list-style-type: none"> ▪ K1 zählt weiter als 30
Ende 01:21		

Tabelle 35: Kategoriensystem Vignette 5 mit Multiplikator (Expertinnengewichtung)

Nummer	Kategorie \ Unterkategorie 1 \ Unterkategorie 2 \ ...	Multiplikator
B5.1_01	Üben (der Zahlen) [ohne weiteres Verb]	1
B5.1_02	K1 zeigt/erklärt (K3), wie die Ziffer Sechs zu schreiben ist	2
B5.1_03	(Er-)Kennen/Benennen von Zahlen \ K1 kennt die Zahlen/das Zahlenbild der Sechs	1
B5.1_04	(Er-)Kennen/Benennen von Zahlen \ K2 kann die Zahlen Sechs und Sieben nicht richtig auseinanderhalten	1
B5.1_05	(Er-)Kennen/Benennen von Zahlen \ K2 kennt wenig Zahlen	1
B5.1_06	(Er-)Kennen/Benennen von Zahlen \ Erfassen von Zahlen	1
B5.1_07	(Er-)Kennen/Benennen von Zahlen \ (Er-)kennen/Benennen von Zahlen	1
B5.1_08	Schreiben der Zahl/en	1
B5.1_09	Schreiben der Zahl/en \ K1 kann die Zahl Sechs schreiben	1
B5.1_10	Schreiben der Zahl/en \ K2 schreibt eine Zahl (sehr klein) auf	1
B5.1_11	Schreiben der Zahl/en \ K2 konnte die Zahlen nicht (so gut) schreiben	1
B5.1_12	Schreiben der Zahl/en \ K2 weiß nicht, wie sie die Zahl Sieben schreiben soll	1
B5.1_13	Schreiben der Zahl/en \ K2+3 brauchen beim Schreiben der Zahlen noch Hilfe	1
B5.1_14	Schreiben der Zahl/en \ K3 konnte die Sechs ohne Hilfe nicht schreiben	1
B5.1_15	Malen der Zahl	1
B5.1_16	Malen der Zahl \ K2 ist unsicher beim Malen	1
B5.1_17	Malen der Zahl \ sie versuchen, die Zahl Sechs zu malen	1
B5.1_18	Erfassung der Struktur des Dezimalsystems	3
B5.1_19	Raumlagekenntnisse (z.B. spiegelverkehrtes Zeigen)	3
B5.1_20	Raumlagekenntnisse (z.B. spiegelverkehrtes Zeigen) \ Erkennen/Wahrnehmen der richtigen Raumlage	3
B5.1_21	Raumlagekenntnisse (z.B. spiegelverkehrtes Zeigen) \ K3 wollte die Sechs erst spiegelverkehrt schreiben	3
B5.1_22	Raumlagekenntnisse (z.B. spiegelverkehrtes Zeigen) \ K1 erkennt (spiegelverkehrt), dass K3 Sechs falsch schreiben wollte	2
B5.1_23	Vergleichen von Zahlen	2
B5.1_24	Vergleichen von Zahlen \ Erkennen von mehr und weniger (z.B. Altersvergleich)	2
B5.1_25	Alter und Zahl \ Alter und Zahl in Verbindung bringen	2
B5.1_26	Alter und Zahl \ Erfassen von Mengen	2
B5.1_27	Alter und Zahl \ Kennen ihr Alter	2
B5.1_28	Zählen ab 16 (bis 32)	3
B5.1_29	Zählen	1
B5.1_30	Zählen \ K1 und K3 zählen zusammen	1
B5.1_31	Zählen \ K3 kann noch nicht so weit zählen wie K1	1
B5.1_32	Zählen \ K2 konnte kaum zählen	1
B5.1_33	Zählen \ K2 konnte kaum zählen \ K2 zieht sich beim Zählen zurück	1
B5.1_34	Zählen \ K1 ist fit im Zählen	1
B5.1_35	Zählen \ Zählen über zehn hinaus	1
B5.1_36	Zählen \ Zählen bis 20	1
B5.1_37	Zählen \ Zählen über 20 hinaus	1
B5.1_38	Zählen \ Zählen über 20 hinaus \ K3 zählt bis ca. 24	1
B5.1_39	Zählen \ Zählen bis 30	1

B5.1_40	Zählen \ Zählen bis 30 \ K1 zählt bis 30	1
B5.1_41	Zählen \ Zählen bis 30 \ K1 (und K3 zählen) zählt sicher bis 28/29	1
B5.1_42	Zählen \ Zählen bis 30 \ K3 zählt bis 29/30 durch Nachsprechen von K1	3
B5.1_43	Zählen \ Zählen bis 30 \ K3 steigt beim Zählen bei 30 aus	3
B5.1_44	Zählen \ Zählen über 30 hinaus	1
B5.1_45	Zählen \ Zählen über 30 hinaus \ K1 zählt weiter als 30	1
	Sonstige \ K2 zeigt im Filmausschnitt KEINE mathematischen Fähigkeiten	0
	Sonstige \ FALSCHER Antworten	0
	Sonstige \ Beschreibung/Interpretation einer Handlung	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Verhalten \ K2 noch unsicher/etwas zurückhaltend	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Verhalten \ K1 und K3 schon sehr sicher	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Verhalten \ K1 ist die „Dominanteste“/übernimmt Führungsrolle	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Verhalten \ ein Kind war zurückhaltender	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Verhalten \ sind unsicher	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Orientieren \ K3 orientiert sich stark an K1	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Orientieren \ K2 orientiert sich an den anderen/schaut zu	0
Ausschluss		
B5.1_01 bis	Sonstige \ Beschreibung/Interpretation einer Handlung \ Interesse \ zeigen Interesse	0
B5.1_21	Sonstige \ Beschreibung/Interpretation einer Handlung \ Interesse \ K2 hat kein so großes Interesse an Zahlen/am Zählen	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Stifthaltung	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ helfen sich gegenseitig	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Proportionen zum Blatt gewählt (riesige Sechs)	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Proportionen zum Blatt gewählt (riesige Sechs) \ Begrenzung (Blatt) wird beachtet	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ motivieren sich gegenseitig	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ K2 kann sich nicht entscheiden, welche Zahl es nimmt	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Kinder beschäftigen sich mit ihrem Alter	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Beobachten einander	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Sprechen miteinander	0

Beispiel 7 von 8

Im siebten Filmbeispiel sehen Sie Kinder, die sich während des Freispiels in einer „Bau-Ecke“ aufhalten.



Denken Sie bitte an **mathematische Fähigkeiten** von Vorschulkindern, während Sie sich den folgenden kurzen Filmausschnitt anschauen. Sie befinden sich **in der Rolle der Erzieherin und möchten die Kinder mathematisch anregen**.

Fragen zu Filmbeispiel 7

B7.1: Welche **mathematischen Fähigkeiten** haben Sie bei dem Kind/den Kindern beobachten können?

(Zur Fortsetzung ggf. Rückseite benutzen.)

B7.2: **Was würden Sie tun**, um das Kind/die Kinder in der eben gezeigten Situation oder im direkten Anschluss daran mathematisch zu fördern?

(Zur Fortsetzung ggf. Rückseite benutzen.)

[13]

Abbildung 29: Beispiel 7 von 8 (4. Mathematikvignette, FiBB-Evaluationsbogen, Teil 2, S. 13)

Tabelle 36: Vignettenbeschreibung und Kategorienbezüge zu Beispiel 7 von 8 (4. Mathematikvignette)

Zeit	Beschreibung des Videos	Kategorien
00:01	Fünf Kinder sitzen in der Bauecke und beschäftigen sich mit Holzklötzen. Vor K3 und K4 liegt ein Kreis aus Quadern flach auf dem Boden. K1 baut an etwas, das aussieht wie ein Haus, aber noch ohne Dach: Die Wände bestehen aus senkrecht stehenden Quadern, auf denen oben Quader flach aufliegen. In der Mitte ist ein Loch. Vor K5 steht ein Turm aus bunten Klötzen, die sehr verschachtelt aufgebaut sind. K2 holt gerade neue Bausteine.	<ul style="list-style-type: none"> • Mathematische Überlegungen/Statik • Vorstellungen umsetzen können/Vorgehen kennen • Bauen <ul style="list-style-type: none"> ○ Aneinanderreihen ○ Gebilde bauen können ○ Stapeln • Räumliches Vorstellungsvermögen haben\3D <ul style="list-style-type: none"> ○ 3D-Wahrnehmung <ul style="list-style-type: none"> ▪ Bauen in 3D ▪ noch nicht in 3D bauen • Bauwerke aus Steinen/Formen entstehen lassen • Formen <ul style="list-style-type: none"> ○ Formen legen ○ Formen bauen
00:03	K2 zu K1: „Kannst du mir helfen?“ K1 bejaht und sagt dann etwas über einen ICE, das schlecht zu verstehen ist.	<ul style="list-style-type: none"> • Mathematische Überlegungen/Statik • Vorstellungen umsetzen können/Vorgehen kennen • Bauen <ul style="list-style-type: none"> ○ Aneinanderreihen ○ Gebilde bauen können ○ Stapeln • Räumliches Vorstellungsvermögen haben\3D <ul style="list-style-type: none"> ○ Bauen in 3D ○ noch nicht in 3D bauen • Bauwerke aus Steinen/Formen entstehen lassen • Formen <ul style="list-style-type: none"> ○ Formen legen ○ Formen bauen
00:12	K2 schaut auf das Gebaute von K1 und meint: „Ja, so sieht doch kein Zug aus!“ K2: „Jaaa. Ich weiß. Ich mach auch keinen Zug, ich mach nur ein Haus.“ Es macht sein Gebautes kaputt und fängt nochmal an.	<ul style="list-style-type: none"> • Mathematische Überlegungen/Statik • Vorstellungen umsetzen können/Vorgehen kennen • Bauen <ul style="list-style-type: none"> ○ Aneinanderreihen ○ Gebilde bauen können ○ Stapeln ○ Verändern ○ Bauanleitung geben können ○ planen können, was man baut • Räumliches Vorstellungsvermögen haben\3D <ul style="list-style-type: none"> ○ 3D-Wahrnehmung <ul style="list-style-type: none"> ▪ Bauen in 3D ▪ noch nicht in 3D bauen • Bauwerke aus Steinen/Formen entstehen lassen • Formen <ul style="list-style-type: none"> ○ Formen legen ○ Formen bauen ○ Formen kennen • Größenverhältnisse einschätzen können/Raumerfahrung <ul style="list-style-type: none"> ○ Gebautes mit Echtem vergleichen können
00:23	K2: „Nee, so sieht doch auch kein Haus aus!“	<ul style="list-style-type: none"> • Größenverhältnisse einschätzen können/Raumerfahrung <ul style="list-style-type: none"> ○ Gebautes mit Echtem vergleichen können • Bauen <ul style="list-style-type: none"> ○ Bauanleitung geben können • Formen <ul style="list-style-type: none"> ○ Formen kennen
00:28	K4 mischt sich ein und meint, dass sein Haus viel größer sei.	<ul style="list-style-type: none"> • Größenverhältnisse einschätzen können/Raumerfahrung • Mengen/Größen noch nicht kennen

	<p>Dabei steht es auf, streckt die Arme ganz weit nach oben. K3 sagt: „Bis zum Himmel.“ und K5: „Joo, ganz hoch!“</p> <p>In der Zeit baut K3 an dem Kreis weiter. Es legt eine zweite Reihe Quader auf die erste, wobei es die Lücken der ersten Reihe mit den Steinen überbrückt.</p>	<ul style="list-style-type: none"> • Räumliches Vorstellungsvermögen haben <ul style="list-style-type: none"> ○ Vorstellung von Größen haben ○ 3D-Wahrnehmung <ul style="list-style-type: none"> ▪ bauen in 3D • Unterscheidung/Vergleich <ul style="list-style-type: none"> ○ Größenunterscheidung/Unterscheidung größer – kleiner ○ Unterscheidung hoch – tief • Formen <ul style="list-style-type: none"> ○ Formen legen ○ Formen bauen • Bauen <ul style="list-style-type: none"> ○ Gebilde bauen können • Bauwerke aus Steinen/Formen entstehen lassen • Mathematische Überlegungen/Statik
00:35	<p>Zeitsprung: K2 hat die Bauecke verlassen. K1 baut immer noch an seinem Haus. Die zweite Runde des Steinkreises von K3 ist fast fertig. K4: „Ja, und ich auch.“ K3: „Ja und Andreas auch.“ K1: „Ne, nie im Leben.“ K5: „Un net Alisa. Weisd was?“ Daraufhin verzieht K3 sein Gesicht und streckt die Zunge raus: „Nur Chayenne, die ist lieb.“ K5: „Ja, und Carlos.“</p>	<ul style="list-style-type: none"> • Formen <ul style="list-style-type: none"> ○ Formen legen ○ Formen bauen • Bauen <ul style="list-style-type: none"> ○ Gebilde bauen können • Bauwerke aus Steinen/Formen entstehen lassen • Mathematische Überlegungen/Statik • Räumliches Vorstellungsvermögen haben\3D <ul style="list-style-type: none"> ○ 3D-Wahrnehmung <ul style="list-style-type: none"> ▪ Bauen in 3D
Ende 00:54		

Tabelle 37: Kategoriensystem Vignette 7 mit Multiplikator (Expertinnengewichtung)

Nummer	Kategorie \ Unterkategorie 1 \ Unterkategorie 2 \ ...	Multiplikator
B7.1_01	Größenverhältnisse einschätzen können/Raumerfahrung	2
B7.1_02	Größenverhältnisse einschätzen können/Raumerfahrung \ Gebautes mit Echem vergleichen können	2
B7.1_03	Räumliches Vorstellungsvermögen haben	1
B7.1_04	Räumliches Vorstellungsvermögen haben \ 3D \ 3D-Wahrnehmung/bauen in 3D	1
B7.1_05	Räumliches Vorstellungsvermögen haben \ 3D \ noch nicht in 3D bauen	1
B7.1_06	Räumliches Vorstellungsvermögen haben \ Vorstellung von Größen haben	1
B7.1_07	Unterscheidung/Vergleich	2
B7.1_08	Unterscheidung/Vergleich \ Unterscheidung wenig – viele	2
B7.1_09	Unterscheidung/Vergleich \ Unterscheidung hoch – tief	2
B7.1_10	Unterscheidung/Vergleich \ Größenunterscheidung/Unterscheidung größer – kleiner	2
B7.1_11	Mengen/Größen noch nicht kennen	2
B7.1_12	Mathematische Überlegungen/Statik	1
B7.1_13	abmessen können	1
B7.1_14	Sortieren können	1
B7.1_15	Vorstellungen umsetzen können/Vorgehen kennen	3
B7.1_16	Mathematisches Verständnis/Zählen	1
B7.1_17	Bauen	1
B7.1_18	Bauen \ Bauanleitung geben können	1
B7.1_19	Bauen \ Verändern	1
B7.1_20	Bauen \ planen können, was man baut	1
B7.1_21	Bauen \ aneinander reihen	1
B7.1_22	Bauen \ Gebilde bauen können	1
B7.1_23	Bauen \ stapeln	1
B7.1_24	Bauwerke aus Steinen/Formen entstehen lassen	1
B7.1_25	Mengen abschätzen können	1
B7.1_26	Zuordnen können	1
B7.1_27	Formen	1
B7.1_28	Formen \ Formen legen	1
B7.1_29	Formen \ Formen bauen	1
B7.1_30	Formen \ Formen kennen	1
	Sonstige \ Keine mathematischen Fähigkeiten zu sehen	1
	Sonstige \ Falsche Antworten	0
	Sonstige \ Beschreibung/Interpretation einer Handlung	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ sich gegenseitig austauschen	0
Ausschluss	Sonstige \ Beschreibung/Interpretation einer Handlung \ Dinge ausprobieren	0
B7.1_01 bis	Sonstige \ Beschreibung/Interpretation einer Handlung \ Wütend werden/ Bauwerke umstoßen/kritisiert werden	0
B7.1_10	Sonstige \ Beschreibung/Interpretation einer Handlung \ Konzentration	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Auge Hand Koordination	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ Unterschiedliche Bausteine werden benutzt	0
	Sonstige \ Beschreibung/Interpretation einer Handlung \ sich gegenseitig helfen/gemeinsam bauen	0

Curriculum Vitae

Persönliche Angaben

Name Christian Donie
Jahrgang 1978

Schulbildung / Berufsausbildungen / Studium

- 2006 2. Staatsexamen für das Lehramt an Grund- und Hauptschulen am Studienseminar in Rohrbach
- 2005 Eignungsfeststellungsverfahren zur Promotion im Bereich der Erziehungswissenschaften an der Universität Koblenz-Landau, Campus Landau unter der Betreuung von Prof. Dr. Hanns Petillon und PD Dr. Herbert Günther
- 2003 1. Staatsexamen der Grund- und Hauptschulpädagogik mit dem zweiten Fach Musik an der Universität Koblenz-Landau, Campus Landau
- 2003 Fachdienstausbildung zum „Notfallnachsorgehelfer“ beim DRK Landesverband Rheinland-Pfalz in Bellheim
- 1999 Lehre zum Rettungsassistenten in der Rettungsdienstakademie Saarbrücken
- 1997 Ausbildung zum Rettungssanitäter beim Deutschen Roten Kreuz Pirmasens
- 1997 Allgemeine Hochschulreife, Robert-Schuman-Gymnasium, Saarlouis

Berufstätigkeit

ab 2021 Stellvertretender Schulleiter der Pestalozzischule, Grundschule Landau Mitte

ab 2020 Grundschullehrer an der Pestalozzischule, Grundschule Landau Mitte

- 2012-2020 Wissenschaftlicher Mitarbeiter im Arbeitsbereich Grundschulpädagogik des Instituts für Bildung im Kindes- und Jugendalter, Arbeitsbereich Grundschulpädagogik der Universität Koblenz-Landau, Campus Landau
- 2012-2014 Lehrbeauftragter der Pädagogischen Hochschule Karlsruhe im BA-Studiengang „Sprachförderung und Bewegungserziehung“
- 2009-2012 Doktorand im DFG-Graduiertenkolleg Unterrichtsprozesse an der Universität Koblenz-Landau im Projekt Förderung anschlussfähiger Bildungsprozesse (FaBi)

-
- 2006-2009 Grundschullehrer an der Grundschule Wollmesheimer Höhe, ab 2007 an der Pestalozzi-Grundschule in Landau
- 2004-2009 Lehrbeauftragter des Instituts für Grundschulpädagogik, das zwischenzeitlich umbenannt wurde in Institut für Bildung im Kindes- und Jugendalter der Universität Koblenz-Landau, Campus Landau
- 1998-2010 Chorleiter (Kirchenchor Steinrausch - Bistum Trier, Sing & Swing Rohrbach e.V., Coeurchen Landau e.V., Frauenchor der evangelischen Gemeinde Offenbach), freischaffende Tätigkeit als Gesang- und Klavierlehrer (u.a. an der Kreismusikschule Saarlouis)
- 1999-2005 Hauptamtliche Arbeit als Rettungsassistent in der Rettungswache Landau (DRK Kreisverband Südliche Weinstraße) und nebenberufliche Arbeit in der ärztlichen Notfalldienstzentrale Landau
- 1997-1998 Zivildienst in der DRK-Lehrrettungswache Saarlouis

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbst angefertigt und alle benutzten Quellen und Hilfsmittel in der Arbeit angegeben habe.

Ich versichere, dass ich die Arbeit weder vollständig noch in Teilen als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe. Die gleiche oder eine ähnliche Abhandlung habe ich bei keiner anderen Hochschule als Dissertation eingereicht.

Landau, den 07.06.2021