



Vor dem im theoretischen Teil dargelegten Hintergrund wird in der vorliegenden Arbeit die folgende übergeordnete Frage untersucht: Wie gestalten Lehrpersonen und Schulleitenden in elf Schulen des perLen-Projekts (*personalisierte Lernkonzepte in heterogenen Lerngruppen*) ihre unterrichtszentrierte Schulentwicklung in Richtung personalisierten Lernens? Zu deren Klärung werden zwei Studien durchgeführt: eine Längsschnittanalyse qualitativer Daten (Studie 1) und eine transformative Mixed-Methods-Analyse (Studie 2). Nach der Beschreibung des Designs beider Studien (Abschnitt 6.1) folgt die Charakterisierung der elf Schulen umfassenden Stichprobe (Abschnitt 6.2). Anschließend werden die Instrumente der Datenerhebung vorgestellt. Zur Ergänzung werden weitere Instrumente der Datenerhebung des perLen-Projekts aufgeführt, wie etwa Leistungstests oder Unterrichtsvideos. Diese sind zwar für die vorliegende Forschungsarbeit nicht direkt relevant, aber für die Verortung im perLen-Projekt hilfreich (Abschnitt 6.3). Im letzten Kapitel des Methodenteils werden die Verfahren der Datenanalyse erläutert (Abschnitt 6.4).

6.1 Design der Forschungsarbeit: Längsschnittanalyse qualitativer Daten (Studie 1) und transformative Mixed-Methods-Datenanalyse (Studie 2)

Beide Studien gehören zum Typus der fallorientierten Schulentwicklungsforschung und verfolgen das Ziel, die Entwicklungstätigkeiten in elf Schulen mit

Elektronisches Zusatzmaterial Die elektronische Version dieses Kapitels enthält Zusatzmaterial, das berechtigten Benutzern zur Verfügung steht
https://doi.org/10.1007/978-3-658-35070-3_6.

Konzepten personalisierten Lernens zu analysieren, um deren Dynamiken und Prozessverläufe zu verstehen. Unter einem „Fall“ wird eine Schule verstanden, die als soziale Organisation eine reale Handlungsfigur darstellt (Flick, 2019; Häder, 2010; Lamnek, 2010). Studie 1 ist als Längsschnittanalyse qualitativer Daten konzipiert, während Studie 2 auf einer transformativen Mixed-Methods-Datenanalyse beruht.

Studie 1: Längsschnittanalyse von qualitativen Daten

In Studie 1 wird der Prozess der unterrichtszentrierten Schulentwicklung in Richtung personalisierten Lernens als Ganzes, das heißt in seinen verschiedenen Dimensionen in den elf Schulen, untersucht. Hierzu wird im ersten Teil der Entwicklungsprozess jeder Schule im dreijährigen Erhebungszeitraum zwischen 2012 und 2015 (t_1 : Schuljahr 2012/2013, t_2 : Schuljahr 2013/2014 und t_3 : Schuljahr 2014/2015) hinsichtlich folgender Themen beschrieben (siehe Abschnitt 1.3): Rahmenbedingungen, Beginn und Auslöser des Entwicklungsprozesses, organisatorische Struktur des jeweiligen personalisierten Lernkonzepts, Entwicklungstätigkeiten innerhalb des Erhebungszeitraums und langfristige Entwicklungsziele.

Im zweiten Teil von Studie 1 werden die elf Schulen miteinander verglichen mit dem Ziel, Entwicklungsgegenstände zu extrahieren. Die Entwicklungsgegenstände umfassen inhaltlich ähnliche Entwicklungstätigkeiten und sind auf einem Abstraktionsniveau formuliert, das einen Vergleich zwischen den Schulen ermöglicht. Ein Beispiel dafür ist die Erarbeitung von Stoffplänen, in denen die Lerninhalte für alters- und/oder leistungsdurchmischte Lerngruppen neu angeordnet werden. Zusätzlich werden Interviewaussagen in die Analyse aufgenommen, die sich auf den Zeitraum vor der Erhebung beziehen.

Längsschnittanalysen qualitativer Daten sind in der Bildungsforschung ein bisher wenig systematisch diskutiertes Forschungsdesign (Dreier, Leuthold-Wergin & Lüdemann, 2018; Thiersch, 2018). Mit diesem Design können querschnittliche Fallkontrastierungen und Fallvergleiche durchgeführt oder, wie in der vorliegenden Studie umgesetzt, längsschnittliche Veränderungen verschiedenen Fällen nachgezeichnet werden (Dreier et al., 2018). Auf diese Weise wird es möglich, Veränderungen auf Schul- und Unterrichtsebene anhand der in den Interviews beschriebenen Entwicklungstätigkeiten in der unterrichtszentrierten Schulentwicklung in Richtung personalisierten Lernens darzustellen (Berkemeyer, Bos & Gröhlich, 2010a; Flick, 2013; Pfeiffer & Püttmann, 2018). Weil bei jedem der drei Erhebungszeitpunkte alle elf untersuchten Schulen befragt wurden, ist der qualitative Längsschnitt auch als Paneldesign charakterisierbar (Berkemeyer et al., 2010a; Pfeiffer & Püttmann, 2018).

Studie 2: Transformative Mixed-Methods-Datenanalyse

In der zweiten Studie werden die Entwicklungstätigkeiten auf Unterrichtsebene hinsichtlich der Gestaltung von pädagogisch-psychologischen Qualitätsaspekten guten Unterrichts untersucht. Unterricht wird im Kontext des didaktischen Dreiecks und seiner drei didaktischen Teilkulturen verstanden: bildungsinhaltliche Ziel- und Stoffkultur, auf den Lehr- und Lernprozess bezogene Lehr- und Lernkultur sowie kommunikativ-unterstützende Interaktions- und Unterstützungskultur (siehe Abschnitt 4.5.1.2). Die Datengrundlage bilden einerseits die Interviewaussagen zu den drei didaktischen Teilkulturen und andererseits eine Frage aus der Online-Befragung des dritten Erhebungszeitpunkts, in der Lehrpersonen und Schulleitende ihren Stand der Schulentwicklung einschätzten. Diese Daten wurden dem Autor für die Analyse zu Verfügung gestellt. In Studie 2 soll mittels transformativer Mixed-Methods-Datenanalyse (u. a. Kuckartz, 2014) aufgezeigt werden, inwiefern sich die Schulen in ihren Entwicklungstätigkeiten hinsichtlich der pädagogisch-psychologischen Qualitätsaspekte guten Unterrichts unterscheiden und gruppieren lassen (siehe Abschnitt 1.4).

Im ersten Teil von Studie 2 werden die Interviewaussagen zur Unterrichtsentwicklung aus Studie 1) bezüglich der Dimensionen personalisierten Lernens ein zweites Mal ausgewertet. Im zweiten Teil werden inhaltlich ähnliche Aussagen pro Interview zusammengefasst (Summary) und der Oberflächen respektive der Tiefenstruktur des Unterrichts zugeordnet, um auf dieser Grundlage das prozentuale Mengenverhältnis berechnen zu können. Mit dieser Quantifizierung der qualitativen Daten wird im dritten Teil von Studie 2 weitergearbeitet und untersucht, inwiefern sich die Schulen bezüglich des prozentualen Mengenverhältnis von Tiefenmerkmalen sowie dem eingeschätzten Entwicklungsstand unterscheiden und gruppieren lassen. Der Entwicklungsstand stellt einen Indikator für die Erfolgseinschätzung von Rekontextualisierungsprozessen dar.

In der Literatur finden sich zahlreiche Ansätze, die aufzeigen, wie qualitative und quantitative Studienteile gemixt werden können (u. a. Creswell, 2014). Kriterien für die Auswahl eines geeigneten Designs sind erstens die Forschungsfragen und zweitens die Aussagekraft und die Nützlichkeit der verfügbaren Daten hinsichtlich des Erkenntnisinteresses (Creswell, 2014; R. B. Johnson, Onwuegbuzie & Turner, 2007; Plano Clark & Ivankova, 2016). In Studie 2 wird ein transformatives Design eingesetzt, bei dem qualitative Daten mittels strukturierender Inhaltsanalyse quantifiziert und in statistische Analyseverfahren mit quantitativen Datensätzen aufgenommen werden (Caracelli & Greene, 1993; Creswell & Plano Clark, 2018; Kuckartz, 2014, 2017; Onwuegbuzie & Teddlie, 2003).

Die Vorteile eines Mixed-Methods-Ansatzes sind vielfältig. So besteht die Möglichkeit, die komplexe unterrichtszentrierte Schulentwicklung in Richtung

personalisierten Lernens aus verschiedenen methodischen Perspektiven zu durchleuchten (Creswell & Plano Clark, 2018; Kuckartz, 2017; Plano Clark & Ivanova, 2016) und ein umfassendes Bild des Untersuchungsgegenstands zu erhalten (Greco & Völcker, 2018; Kelle, 2014). Des Weiteren wird die Aussagekraft der Ergebnisse erhöht und typische Muster im Datensatz können breiter abgestützt werden als in Forschungsdesigns mit nur qualitativen oder nur quantitativen Methoden (Ortner, 2018). Ein weiterer Vorteil gegenüber rein qualitativen Datensätzen besteht darin, dass durch den Einbezug quantitativer Fragebogendaten eine größere Stichprobe untersucht werden kann. Im perLen-Projekt werden daher nicht nur die Antworten der interviewten Personen, sondern durch den Einbezug von Fragebogendaten auch die Einschätzungen weiterer an den untersuchten Schulen tätigen Lehrpersonen in die Analyse mit aufgenommen.

6.2 Stichprobe

Die Stichprobe des perLen-Projekts umfasst insgesamt 65 Schulen. 53 Schulen gehören der Kernstichprobe und zwölf Schulen der Ergänzungsstichprobe an. Während die Schulen der Kernstichprobe deutlich von tradierten Formen des Unterrichts und der Lernorganisation abweichen, verfügen die zwölf Schulen der Ergänzungsstichprobe über eher traditionelle räumliche und zeitliche Schul- und Unterrichtsstrukturen und implementieren innerhalb dieser Strukturen vermehrt individualisierte und kooperative Lehr- und Lernformen (Stebler et al., 2017). Die in der vorliegenden Arbeit untersuchten elf Schulen (perLen-Fallschulen) gehören zur Kernstichprobe. Diese elf Schulen wurden für die Datenanalyse ausgewählt, weil nur in diesen Schulen Interviews durchgeführt wurden.¹

Die Rekrutierung aller Schulen erfolgte in einem mehrschrittigen Verfahren. Als Erstes formulierte die perLen-Forschungsgruppe Kriterien für Schulen mit Konzepten personalisierten Lernens. Gesucht wurden Schulen, die

sich an einem personalisierten Lehr-Lern-Konzept orientieren und die sich als Schule, Schulabteilung, Kollegium in den letzten Jahren hin zu einer verstärkten Ausrichtung der Unterrichts- und Lernorganisation an den individuellen Bedürfnissen und Voraussetzungen der Schülerinnen und Schüler und dem Aufbau fachlicher und überfachlicher (personaler, sozialer, methodischer) Kompetenzen weiterentwickelt haben. (Reusser, Pauli & Stebler, 2012, S. 1)

¹ Eine umfangreiche Stichprobenbeschreibung des perLen-Projekts ist bei Reusser et al. (2015b) oder Stebler et al. (2017) zu finden.

Anhand dieser Kriterien wurde eine Ausschreibung veröffentlicht, auf die sich Schulen bewerben konnten. Diese konnten angeben, ob sie als perLen-Fallschule (mit Interviews zur Datenerhebung) oder als Online-Schule (ohne Interviews) teilnehmen wollten. Des Weiteren suchten Projektmitarbeiterinnen und Projektmitarbeiter in ihrem beruflichen Umfeld (u. a. in Pädagogischen Hochschulen) Schulen, die den Kriterien entsprachen. Auch in Bildungsinstitutionen (u. a. Erziehungsdirektionen, Schulinspektorate) wurde nachgefragt, ob ihnen solche Schulen bekannt seien. Insgesamt gingen 24 Bewerbungen für perLen-Fallschulen ein. Aus diesen wurden zwölf Schulen ausgewählt. Dabei wurde darauf geachtet, dass Primar- und Sekundarschulen, öffentliche und private Schulen sowie Schulen aus mehreren Kantonen der Deutschschweiz vertreten waren. Aufgrund dieses mehrstufigen Auswahlverfahrens handelt es sich weder bei den perLen-Schulen ($N = 65$) noch bei der Teilmenge der perLen-Fallschulen ($n = 11^2$) um eine Zufallsstichprobe. Gleichwohl wird davon ausgegangen, dass die perLen-Stichprobe die Vielfalt innovativer Schulen in der deutschsprachigen Schweiz dokumentiert.

Die elf Schulen lassen sich anhand folgender Charakteristika beschreiben:³

- 6 kleine (bis 120 Schülerinnen und Schüler), 2 mittlere (121–220 Schülerinnen und Schüler) und 3 große Schulen (ab 221 Schülerinnen und Schüler);
- 2 private und 6 öffentliche Schulen;
- 2 Primarschulen, 7 Schulen der Sekundarstufe 1 und 2 Gesamtschulen;
- 3 Schulen, die sich in einer Kernstadt befinden, 2 Schulen, die in der Agglomeration angesiedelt sind, und 6 Schulen auf ländlichen Regionen;
- 1 Schule, in der mit der unterrichtszentrierten Schulentwicklung in Richtung personalisierten Lernens im Erhebungszeitraum begonnen wurde. 2 Schulen hatten ihr Unterrichtskonzept 1–3 Jahre, 3 Schulen 4–6 Jahre und 5 Schulen mehr als 7 Jahre vor dem ersten Erhebungszeitpunkt eingeführt.

6.3 Instrumente der Datenerhebung

In der vorliegenden Forschungsarbeit werden die Interviews (Abschnitt 6.3.1) sowie ein Item des Online-Fragebogens t3 (Abschnitt 6.3.2) ausgewertet. Diese

² Eine perLen-Fallschule stieg im ersten Erhebungsjahr aus dem Projekt aus. Aufgrund fehlender Daten im zweiten und dritten Erhebungszeitpunkt wurde Schule C nicht in die Stichprobe aufgenommen.

³ Um einen möglichst hohen Anonymisierungsgrad zu gewährleisten, wird auf eine detaillierte Darstellung, in der die Charakteristika jeder Schule in einer Tabelle aufgeführt sind, verzichtet.

Daten wurden von Mitarbeiterinnen und Mitarbeitern des perLen-Projekts erhoben und dem Autor aufbereitet für die Datenanalyse zur Verfügung gestellt. Für eine bessere Verortung der Forschungsarbeit im perLen-Projekt werden nachfolgend deren vielfältige Erhebungsinstrumente im Überblick aufgeführt (Tabelle 6.1).

Neben der Sammlung von Konzeptdokumentationen und leitfadengestützten, teilstrukturierten Einzel- und Gruppeninterviews mit Schulleitenden und Lehrpersonen wurden Online-Fragebogen von Lehrpersonen und Schulleitenden sowie Schülerinnen und Schülern ausgefüllt. Des Weiteren wurden Tests für Messungen fachlicher und, in einer ausgewählten Teilstichprobe, überfachlicher Kompetenzen sowie videogestützte Unterrichtsbeobachtungen eingesetzt (Reusser, Pauli & Stebler, 2015a; Stebler, 2017).

6.3.1 Qualitative Daten: Teilstrukturierte und leitfadengestützte Einzel- und Gruppeninterviews mit Schulleitenden und Lehrpersonen im Längsschnitt

Pro Schule fanden je ein Einzel- oder Gruppeninterview mit der Schulleitung (t_1) und je drei Gruppeninterviews mit Lehrpersonen (t_1 , t_2 , t_3) statt. Im Anhang (siehe Anhang A im elektronischen Zusatzmaterial) ist exemplarisch der Leitfaden des dritten Erhebungszeitpunkts aufgeführt. Die insgesamt 45 Interviews wurden an den jeweiligen Schulen geführt. Sieben davon waren Einzelinterviews mit Schulleitenden und zwei davon Einzelinterviews mit Lehrpersonen. Bei den anderen 36 Interviews handelte es sich um Gruppeninterviews. Die beiden Einzelinterviews wurden geführt, weil jeweils eine Person zum Interviewzeitpunkt t_2 kurzfristig einen Termin wahrnehmen musste. Bei der Zusammensetzung der Interviewgruppen wurde darauf geachtet, Lehrpersonen zu interviewen, welche diejenigen Klassen/Lerngruppen unterrichteten, die eine Lernlandschaft bildeten und an der Online-Befragung der Schülerinnen und Schüler teilgenommen hatten. Grundsätzlich sollten in den Gruppeninterviews der Lehrpersonen über die drei Erhebungsjahre hinweg dieselben Personen interviewt werden. Tabelle 6.2 zeigt, inwiefern die Gruppen konstant blieben oder sich veränderten.

In drei der elf Schulen waren die Lehrpersonengruppen bei allen drei Interviews (t_1 , t_2 , t_3) gleich zusammengesetzt (in Tabelle 6.2 unterstrichen). Bei sieben Schulen gab es eine konstante Schnittmenge und Abweichungen von einer oder zwei Personen infolge von Mutationen (u. a. Pensionierung, Stellenwechsel) oder Absenzen (u. a. Mutterschaftsurlaub, Weiterbildungen). In Schule B wurde der Entwicklungsprozess im zweiten Erhebungszeitpunkt größtenteils

Tabelle 6.1 Erhebungsinstrumente des perLen-Forschungsprojekts

Erhebungszeitpunkte	Erhebungsinstrumente	Kernstichprobe (53 Schulen)		Ergänzungsstichprobe (12 Schulen)
		perLen Fallschulen (n = 11)	Online-Schulen (n = 42)	
t ₁ 2012/2013	Konzeptdokumentation	x		
	Interviews mit Schulleitenden(gruppen)	x		
	Interviews mit Lehrpersonengruppen	x		
	Online-Befragung mit Lehrpersonen, Schulleitenden und Schülerinnen/Schülern (4./7. Klasse)	x	x	x
	Erfassen fachlicher Kompetenzen von Schülerinnen/Schülern	x	x	
t ₂ 2013/2014	Interviews mit Lehrpersonengruppen	x		
	Online-Befragung mit Lehrpersonen, Schulleitenden und Schülerinnen/Schülern (5./8. Klasse)	x	x	x
	Interviews mit Lehrpersonengruppen	x		

(Fortsetzung)

Tabelle 6.1 (Fortsetzung)

Erhebungszeitpunkte	Erhebungsinstrumente	Kernstichprobe (53 Schulen)		Ergänzungsstichprobe (12 Schulen)
		perLen Fallschulen (n = 11)	Online-Schulen (n = 42)	
	Online-Befragung mit Lehrpersonen, Schulleitenden und Schülerinnen/Schülern (6./9. Klasse)	x	x	x
	Erfassen fachlicher Kompetenzen von Schülerinnen/Schülern	x	x	x
	Erfassen überfachlicher Kompetenzen von Schülerinnen/Schülern	(x)	(x)	(x)
	Videogestützte Unterrichtsbeobachtungen	(x)	(x)	

Anmerkungen: perLen-Fallschulen sind die in dieser Forschungsarbeit untersuchten Schulen; x = alle Schulen der Teilstichprobe; (x) = ausgewählte Schulen der Teilstichprobe; Quelle: Reusser, Pauli & Stibler (2015a) und Stibler (2017).

Tabelle 6.2 Anzahl interviewter Personen und Übereinstimmung der Gruppen über drei Erhebungszeitpunkte (t_1 , t_2 , t_3):

Schule	Anzahl interviewter Personen			
	Schulleitende (t_1)	Lehrpersonen (t_1)	Lehrpersonen (t_2) / Übereinstimmung mit t_1	Lehrpersonen (t_3) / Übereinstimmung mit t_1 / Übereinstimmung mit t_2
Schule A	1	5	5 / 5	4 / 3 / 3
Schule B	1	4	2 / 0	1 / 1 / 0
Schule D	1	<u>4</u>	<u>4 / 4</u>	<u>4 / 4 / 4</u>
Schule E	2	4	4 / 4	3 / 3 / 3
Schule F	1	4	4 / 4	2 / 2 / 2
Schule G	1	2	2 / 2	2 / 1 / 1
Schule H	4	<u>4</u>	<u>4 / 4</u>	<u>4 / 4 / 4</u>
Schule J	1	<u>2</u>	<u>2 / 2</u>	<u>2 / 2 / 2</u>
Schule K	2	3	3 / 2	3 / 2 / 3
Schule L	2	3	3 / 3	3 / 2 / 2
Schule M	1	4	3 / 3	4 / 2 / 2

Anmerkungen: Unterstreichung markiert gleichbleibende Zusammensetzung. Angaben übernommen aus Stebler & Vasarik Staub (2016).

gestoppt, weswegen die Lehrpersonen des ersten Erhebungszeitpunkts im zweiten Erhebungszeitpunkt nicht interviewt werden konnten. Stattdessen wurde je ein Interview mit der Schulleitung und einer Lehrperson durchgeführt, welche die Schule verlassen hatten.

Die Länge der insgesamt 45 Interviews betrug zwischen 35 Minuten und 140 Minuten. Die Interviews wurden auf Tonträgern festgehalten und von der Forschungsgruppe perLen transkribiert. Für die Auswertung lagen die Gespräche daher als schriftliche Textdokumente vor, wobei der gesamte transkribierte Textumfang von ca. 630'000 Wörtern für die vorliegende Forschungsarbeit relevant ist.

Einzel- und Gruppeninterviews mit Expertinnen und Experten

Eine Stärke von Einzelinterviews besteht in der Gewinnung von Informationen zu individuellen Meinungen und Sichtweisen, die nicht durch in der Interviewsituation anwesende andere Personen beeinflusst werden können. Dagegen stehen

in Gruppeninterviews die gemeinsamen Sichtweisen im Zentrum, in der vorliegenden Arbeit beispielsweise das gemeinsame Verständnis von personalisiertem Lernen. Des Weiteren lassen sich angesprochene Widerstände und Meinungsverschiedenheiten, die diskursive Auseinandersetzung damit und deren Aufarbeitung im Interview abbilden (Döring & Bortz, 2016). Des Weiteren lassen sich die Interviews als Experteninterviews charakterisieren. Die Befragten sind Expertinnen und Experten für ihre Aufgaben- und Verantwortungsbereiche in den jeweiligen Konzepten personalisierten Lernens, weshalb sie als „Quelle von Spezialwissen über die zu erforschenden sozialen Sachverhalte“ (Gläser & Laudel, 2010, S. 12) fungieren. Expertinnen und Experten sind sie nicht nur wegen des für den Forschungsgegenstand zentralen Handlungswissens, das heißt wegen des Wissens über die Entwicklungstätigkeiten, sondern auch aufgrund der zentralen Position von Schulleitenden und Lehrpersonen in einer unterrichtszentrierten Schulentwicklung (siehe Abschnitt 4.7.1 und Abschnitt 4.7.2). Denn ohne diese Akteurinnen und Akteure könnten die Konzepte personalisierten Lernens in den Schulen nicht umgesetzt werden (Bogner, Littig & Menz, 2014; Helfferich, 2014; Meusser & Nagel, 2009).

Teilstrukturierte Leitfadeninterviews

In Leitfadeninterviews kommt eine Auflistung von vorgängig definierten Fragen zum Einsatz. Solche thematisch aufeinander abgestimmten Interviewfragen werden auf der Grundlage von theoretischen Vorannahmen und Forschungsfragen konstruiert. Mit diesem Instrument wird das Interviewgespräch gezielt auf die Untersuchungsgegenstände gelenkt, wodurch ein teilweise künstlicher Gesprächsverlauf entstehen kann (Helfferich, 2014). Dieser Künstlichkeit kann in teilstrukturierten Leitfadeninterviews zumindest partiell entgegengewirkt werden, indem beispielsweise die Reihenfolge der Interviewfragen dem Gesprächsverlauf angepasst oder auf für die interviewte Person wichtige und interessante Themen intensiver eingegangen wird. Leitfäden sind in der Regel so konstruiert, dass pro Themenbereich eine kleine Einleitung mit einer oder mehreren Einstiegsfragen formuliert ist. Hinzu kommen Nachfragen, die gestellt werden können, wenn bestimmte Themen durch die Einstiegsfrage nicht oder nur ansatzweise beschrieben wurden (Döring & Bortz, 2016; Helfferich, 2014).

Mithilfe dieses Instruments ist es zudem möglich, längsschnittliche Daten zu Entwicklungsprozessen zu erhalten. So können in einem Interview Aussagen aus vorhergehenden Interviews aufgegriffen und es kann nachgefragt werden, was sich in der Zwischenzeit diesbezüglich allenfalls verändert hat. In der perLen-Studie ermöglichten die Interviews bei ähnlich gestellten Fragen einen Vergleich nicht nur im Längsschnitt, sondern auch zwischen den einzelnen Schulen (Döring

& Bortz, 2016; Pfeiffer & Püttmann, 2018). In Tabelle 6.3 sind die erfragten Themenbereiche aus den Interviews im zeitlichen Verlauf aufgelistet.

Tabelle 6.3 Themenbereiche in den teilstrukturierten Leitfadeninterviews mit

Themenbereiche	Schulleitenden (t ₁)	Lehrpersonen (t ₁)	Lehrpersonen (t ₂)	Lehrpersonen (t ₃)
Vorstellung der Person	x	x		
Pädagogische Leitbegriffe	x			
Beschreibung der Schülerinnen und Schüler	x	x		
Unterricht und Unterrichtsentwicklung	x	x	x	x
Lehrpersonenrolle		x	x	x
Zusammenarbeit mit Eltern		x	x	x
Zusammenarbeit der Lehrpersonen	x	x	x	x
Unterrichtsräume			x	
Fragen zur Schulentwicklung und zu Veränderungen im Allgemeinen	x		x	x
Konzepte personalisierten Lernens	x			x
Ergänzungen	x	x	x	x

Anmerkung: Übernommen aus den Leitfäden (exemplarischer Leitfaden des dritten Erhebungszeitpunkts im Anhang A im elektronischen Zusatzmaterial)

Aus Tabelle 6.3 lässt sich entnehmen, dass für die vorliegende Forschungsarbeit relevante Themen wie etwa „Unterricht und Unterrichtsentwicklung“ zu allen drei Erhebungszeitpunkten erfragt wurden. Somit liegt für die Auswertungen ein Datensatz vor, der es erlaubt, sowohl die Entwicklungstätigkeiten in jeder der elf Schulen längsschnittlich in drei aufeinanderfolgenden Schuljahren zu analysieren als auch die Schulen miteinander zu vergleichen.

6.3.2 Quantitative Daten: Online-Fragebogen für Lehrpersonen und Schulleitende

In Studie 2 der vorliegenden Arbeit werden in der Datenanalyse neben den qualitativen Daten zusätzlich Daten aus der Online-Befragung des dritten Erhebungszeitpunkts verwendet ($n_{t3} = 130$; Rücklauf_{t3} = 53.50 %; Stebler & Mötteli, 2016). Bei einem Item, das 124 Personen beantworteten, konnten sowohl Fachlehrpersonen als auch Klassenlehrpersonen ($n = 97$; 78 %) und Schulleitende mit ($n = 23$; 19 %) und ohne Unterrichtsverantwortung ($n = 4$; 3 %) einschätzen, wie gut ihre Schule in der Schulentwicklung „unterwegs“ sei. Im Durchschnitt waren die befragten Personen 45.77 Jahre alt ($SD = 11.35$) und 56.5 Prozent waren weiblich. Das Antwortformat reichte von „gar nicht gut unterwegs“ (1) bis „hervorragend unterwegs“ (10):

Ist Ihre Schule gut unterwegs? Bitte schätzen Sie die Schulentwicklung auf einer Skala von 1–10 ein, wobei 1 „gar nicht gut unterwegs“ und 10 „hervorragend unterwegs“ bedeutet.

Zu bedenken gilt es diesbezüglich jedoch, dass die Formulierung „unterwegs sein“ in zweifacher Hinsicht gedeutet werden kann: Einerseits kann darunter die Einschätzung der Schulentwicklung in Relation zur Zielerreichung verstanden werden. Gemäß dieser Interpretation würde „gut unterwegs sein“ bedeuten, dass die Entwicklungsziele (fast) erfüllt sind, während die Auslegung „schlecht unterwegs sein“ so aufzufassen wäre, dass die Zielerreichung noch in weiter Ferne liegt. Andererseits kann das Item auf den Prozess bezogen werden, das heißt darauf, inwieweit der Entwicklungsprozess zufriedenstellend voranschreitet, unabhängig davon, ob die Ziele bereits erreicht wurden oder noch nicht erreicht worden sind.

Das Item zum Stand der Schulentwicklung wird für die Beantwortung der Teilfragen 2.5 bis 2.7 (Studie 2/Teil 3) beigezogen (siehe Abschnitt 1.4). Darüber hinaus beinhaltete der Fragebogen aller drei Erhebungszeitpunkte weitere Items, die für den Entwicklungsprozess relevante Gegenstände thematisierten und beispielsweise danach fragten, ob sich durch das jeweilige Konzept personalisierten Lernens die Rolle als Lehrperson verändert habe. Diese Items wurden nicht in die Analyse aufgenommen, weil sie in Bezug auf das Abstraktionsniveau nicht

auf demselben Niveau sind wie die für die Mixed-Methods-Datenanalyse verwendeten quantifizierten Werte der Aussagen zu Tiefenstrukturmerkmalen von Unterricht (siehe Abschnitt 6.4.2.1).

6.4 Verfahren der Datenanalyse

Eine Übersicht über die Datengrundlage und die eingesetzten Verfahren der Datenanalyse ist in Abbildung 6.1 dargestellt.

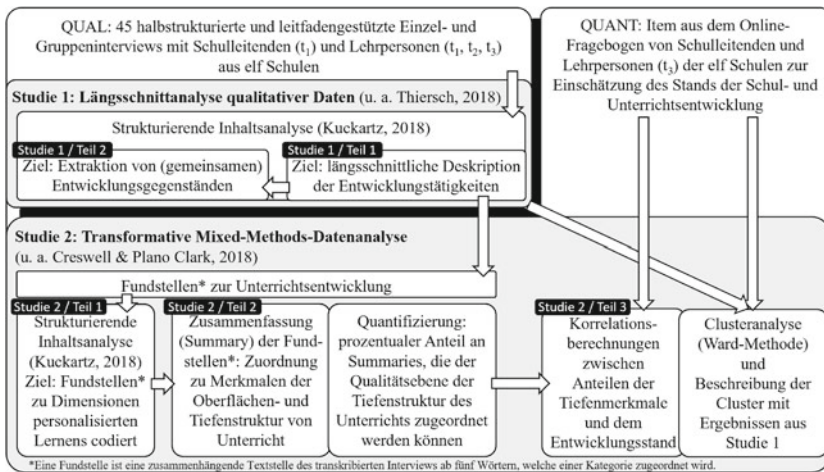


Abbildung 6.1 Überblick über die Datengrundlage und Datenanalyse der zwei Studien (QUALI: qualitative Datengrundlage; QUANT: quantitative Datengrundlage; eigene Darstellung)

In Studie 1 (Längsschnittanalyse qualitativer Daten; siehe Abschnitt 6.4.1) wurden zuerst mit einer strukturierenden Inhaltsanalyse nach Kuckartz (2018) die Entwicklungsprozesse der elf Schulen innerhalb des Erhebungszeitraums rekonstruiert (Studie 1/Teil 1). Darauf aufbauend wurden im zweiten Teil die beschriebenen Entwicklungstätigkeiten vor dem und im Erhebungszeitraum zu Entwicklungsgegenständen abstrahiert und es wurde untersucht, welche der Entwicklungsgegenstände in allen Schulen genannt werden (Studie 1/Teil 2).

In Studie 2 (transformative Mixed-Methods-Datenanalyse; Abschnitt 6.4.2) wurde in einem ersten Schritt die in Studie 1 codierten Fundstellen zur Unterrichtsentwicklung mittels einer strukturierenden Inhaltsanalyse nach Kuckartz (2018) und eines auf den Dimensionen personalisierten Lernens (siehe Abschnitt 3.2, u. a. Stebler et al., 2018) aufbauenden Kategoriensystems ein zweites Mal codiert (Studie 2/Teil 1). Unter einer „Fundstelle“ ist im vorliegenden Zusammenhang eine transkribierte Interviewaussage zu verstehen, die einer Kategorie des Kategoriensystems zugeordnet wird und eine von der vorhergehenden sowie nachfolgenden Fundstelle im Transkript abgrenzbare Sinneinheit bildet. Die Setzung der Grenzen von Fundstellen erfolgte anhand relevanter Inhalte. Was relevant ist und was nicht, wird durch die Forschungsfrage(n), die theoretischen Grundlagen sowie die Kategoriendefinitionen (Memos) bestimmt. Im zweiten Schritt wurden die Fundstellen zusammengefasst. Hierbei wurden die codierten Aussagen zu einer Kategorie für jedes Interview so verdichtet, dass für die betreffende Kategorie relevanten Informationen zusammengefasst vorliegen. Diese Zusammenfassungen werden im Folgenden „Summaries“ genannt (Kuckartz, 2018). Mit der Summariebildung wurden die umfangreichen Textmengen auf überschaubare und für die Forschungsfrage(n) relevanten Inhalte reduziert. Zudem hat dieser Analyseschritt den Vorteil, dass die Interviewführung die Anzahl bei der Auszählung der Summaries weniger verfälscht als bei der Auszählung von Fundstellen. Dies ist besonders im perLen-Projekt von Bedeutung, da verschiedene Personen mit je eigenen Interviewstilen an der Datenerhebung beteiligt waren (siehe Abschnitt 6.4.2.1). In einem dritten Analyseschritt wurden die Summaries zusätzlich im Hinblick darauf differenziert, ob die interviewten Personen Entwicklungstätigkeiten in Bezug auf Merkmale der Oberflächen oder der Tiefenstruktur von Unterricht beschrieben. Anschließend wurden diese Daten quantifiziert, indem pro Schule die Summaries der Oberflächen- und die Summaries der Tiefenstruktur des Unterrichts ausgezählt wurden und das prozentuale Anzahlverhältnis berechnet wurde (Studie 2/Teil 2). Im vierten Schritt wurden die prozentualen Summary-Anteile der Tiefenmerkmale des Unterrichts mit der auf Schulebene aggregierten individuellen Einschätzungen des Entwicklungsstands korreliert und es wurde eine Clusterbildung mit der hierarchischen Clusteranalyse (Ward-Methode) erarbeitet (Studie 2/Teil 3).

6.4.1 Studie 1: Verfahren der Längsschnittanalyse qualitativer Daten

Nachfolgend wird die strukturierende Inhaltsanalyse nach Kuckartz (2018) mit MAXQDA (Abschnitt 6.4.1.1) vorgestellt und es werden Gütekriterien und Verfahren der Qualitätssicherung beschrieben (Abschnitt 6.4.1.2).

6.4.1.1 Strukturierende Inhaltsanalyse (Kuckartz, 2018)

Die im Folgenden dargestellte strukturierende Inhaltsanalyse wird bei Kuckartz (2018) als „inhaltslich strukturierende Inhaltsanalyse“ bezeichnet (Abbildung 6.2).

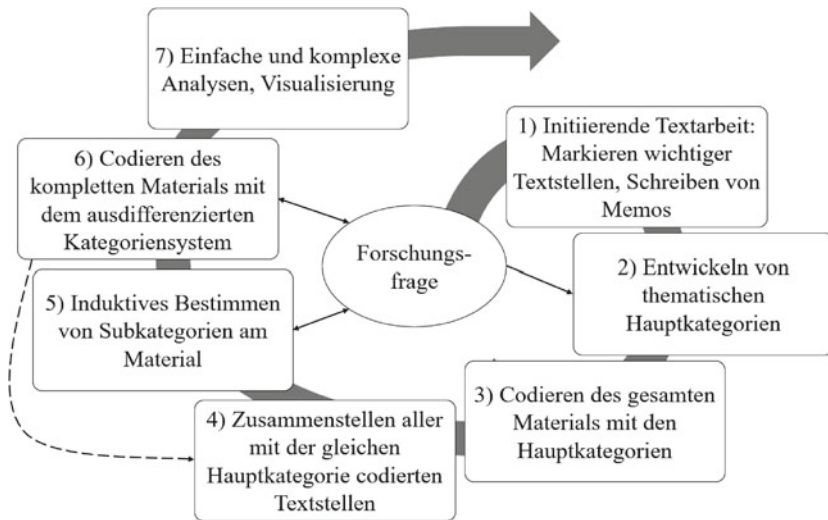


Abbildung 6.2 Ablauf einer strukturierenden Inhaltsanalyse nach Kuckartz (2018)

1) Initiierende Textarbeit

Im ersten Schritt wurden die Transkripte gelesen, wichtige Textpassagen für ein erstes Verständnis unterrichtszentrierter Schulentwicklung in Richtung personalisierten Lernens gesammelt und schulspezifische Besonderheiten im Forschungs- tagebuch festgehalten.

2) Deduktive Entwicklung von thematischen Hauptkategorien

Anschließend wurde deduktiv, aufbauend auf dem uSpL-Modell (siehe Abschnitt 4.2), das in Tabelle 6.4 wiedergegebene Kategoriensystem entwickelt:

Tabelle 6.4 Studie 1: Deduktiv erarbeitetes Kategoriensystem

Kategorien	Theoretischen Grundlagen der Kategorien (Kapitelverweis)
Entwicklungstätigkeiten in den Entwicklungsdimensionen	
Unterrichtsentwicklung in Richtung einer personalisierten Gestaltung von Lehr- und Lernprozessen	Abschnitt 4.5.1
Entwicklung der Ziel- und Stoffkultur	Abschnitt 4.5.1.2
Entwicklung der Lehr- und Lernkultur	Abschnitt 4.5.1.2
Entwicklung der Interaktions- und Unterstützungskultur	Abschnitt 4.5.1.2
Entwicklung organisationaler Strukturen	Abschnitt 4.5.2
Entwicklung der Kooperation zwischen Lehrpersonen	Abschnitt 4.5.3
Professionalisierung der Lehrpersonen und Personalrekrutierung	Abschnitt 4.5.4
Entwicklung eines gemeinschaftlichen Schullebens	Abschnitt 4.5.5
Allgemeine Aussagen zur unterrichtszentrierten Schulentwicklung in Richtung personalisierten Lernens (uSpL)	
Auslöser und Beginn einer uSpL	Abschnitt 4.4
Steuerung einer uSpL (z. B. Instrumente)	Abschnitt 4.6
Stand einer uSpL zu Beginn des Erhebungszeitraums	Abschnitt 4.6
Metaziele einer uSpL	Abschnitt 4.10
Rahmenbedingungen der Schule	
Schulform und Trägerschaft	Abschnitt 4.9
Schulstandort	Abschnitt 4.9
Finanzielle Mittel für die uSpL	Abschnitt 4.9

(Fortsetzung)

Tabelle 6.4 (Fortsetzung)

Kategorien	Theoretischen Grundlagen der Kategorien (Kapitelverweis)
Merkmale der Schülerinnen und Schüler (Anzahl, Heterogenität etc.)	Abschnitt 4.9
Kooperation mit außerschulischen Akteursgruppen	
Schulbehörden	Abschnitt 4.8.1
Eltern	Abschnitt 4.8.2
Beratende	Abschnitt 4.8.3
Kontakte zu anderen Schulen mit Konzepten personalisierten Lernens (z. B. Schulnetzwerk etc.)	Abschnitt 4.8.4

Anmerkung: In dieser Tabelle ist das deduktiv erarbeitete Kategoriensystem dargestellt. Das finale Kategoriensystem ist in Abschnitt 7.1.1 und Anhang D im elektronischen Zusatzmaterial zu finden.

Für jede Kategorie wurde in MAXQDA ein Memo erstellt. Im Memo wurden die Kategorien beschrieben. Zusätzlich wurden Codierregeln definiert (siehe Anhang B im elektronischen Zusatzmaterial) und im Forschungstagebuch festgehalten. Codierregeln sind für regelgeleitetes Codieren und die intersubjektive Nachvollziehbarkeit der Analyse zentral. Ein Beispiel für eine solche Codierregel ist, dass Doppelcodierungen nicht zulässig sind. Das heißt, eine Fundstelle kann nur einer Kategorie zugeordnet werden.

3) Codierung des gesamten Interviewmaterials mit dem deduktiv erstellten Kategoriensystem und Überprüfung der Intercoder-Übereinstimmung

Im dritten Schritt wurden die Transkripte ein weiteres Mal durchgelesen, Fundstellen markiert und einer Kategorie zugeordnet. Fundstellen wurden einer Kategorie aufgrund ihres Sinninhalts zugeordnet. Die Grenzen zwischen den Fundstellen ergaben sich dort, wo ein Sinninhalt endet und ein neuer Sinninhalt beginnt. Passt eine Fundstelle nur teilweise oder gar nicht zu einer Kategorie, so wurde entweder das Memo dieser Kategorien angepasst oder eine neue Kategorie gebildet und deren Definition in Abgrenzung zur ersten formuliert.

Des Weiteren konnten in dieser Analysephase Codierregeln angepasst werden. Zum Beispiel wurde die Codierregel angepasst, dass eine Fundstelle nur einer Kategorie zugeordnet werden darf. Denn es zeigte sich im Codierprozess, dass die einzelnen Textstellen in manchen Fällen so auseinandergenommen wurden, dass bei einzelner Betrachtung der codierten Fundstellen deren Sinninhalt nicht erschließbar war, weil Kontextinformationen, die vor oder nach der

Fundstelle erläutert wurden, fehlen. Darum wurde die Regel so angepasst, dass Fundstellen *einer Hauptkategorie mehreren Unterkategorien* zugeordnet werden können. Die Mehrfachcodierung war jedoch nur erlaubt, wenn sie zu einem besseren Verständnis der Fundstelle beitrug. Sämtliche Änderungen wurden in einem Forschungstagebuch protokolliert.

Nach der ersten Sättigung des Kategoriensystems wurde die Intercoder-Übereinstimmung überprüft. Unter „Sättigung“ wird in diesem Zusammenhang das empirische Durchdringen des Untersuchungsgegenstands mithilfe des Kategoriensystems verstanden (Strübing, Hirschauer, Ayaß, Krähnke & Scheffer, 2018). Eine erste Sättigung war dann erreicht, wenn während des Codierens nur noch selten Veränderungen am Kategoriensystem, an den Memos und an den Codierregeln vorgenommen werden mussten.

Zur Überprüfung der Intercoder-Übereinstimmung wurde in einem ersten Schritt die Gegencodiererin oder der Gegencodierer in die Analyse eingeführt. Das Kategoriensystem und die Codierregeln wurden vorgestellt und das Codieren wurde an einigen Interviewausschnitten trainiert. Im zweiten Schritt wurde ein Interviewtranskript von beiden Personen unabhängig voneinander codiert. In diesem Transkript waren bereits die Grenzen der Fundstellen markiert, aber die Fundstellen waren noch keiner Kategorie zugeordnet worden. Die Grenzen konnten nach wie vor verändert werden (Mayring, 2014). Im dritten und letzten Schritt wurden die Codierungen miteinander verglichen, indem die Grenzen jeder Fundstelle sowie deren Zuordnungen zu den Kategorien überprüft wurde. Ziel war es, eine möglichst hohe Übereinstimmung zu erhalten. Bei Nicht-Übereinstimmung einigten sich die Codierenden in einer konsensuellen Diskussion (Hopf & Schmidt, 1993; Schmidt, 2013) über die Definitionen der Kategorien und die Codierregeln sowie die Zuordnung der Fundstellen.

Die in der vorliegenden Arbeit berichteten Werte der Intercoder-Übereinstimmungen beziehen sich auf prozentuale Werte von Übereinstimmungen, die mit der MAXQDA-Funktion „Codierüberlappung an Segmenten von mindestens 100 %; Auswertung: Segmente beider Dokumente“ ermittelt wurden. Ein absoluter Wert einer Übereinstimmung ergibt sich, wenn dieselbe Fundstelle mit gleichem Anfang und Ende von beiden Codierenden derselben Kategorie zugeordnet wurde (Kuckartz, 2018). Anschließend wurde über die MAXQDA-Funktion „K“ (Cohens Kappa) berechnet. Für die Berechnung des Koeffizienten greift das Analyseprogramm auf die Arbeiten von Brennan und Prediger (1981) zurück. Cohens Kappa ist ein zufallsbereinigter Koeffizient für die Einschätzung der Übereinstimmung von zwei Codierenden. Es wird von der prozentualen Übereinstimmung jener Anteil abgezogen, der sich bei einer zufälligen Zuordnung der Fundstellen ergeben würde.

Die Inter-coder-Übereinstimmung mit den Kategorien der Entwicklungsdimensionen (siehe Anhang C im elektronischen Zusatzmaterial) ist mit 81.20 Prozent (Cohens Kappa = .77) als gut einzuschätzen (Wirtz & Casper, 2002). Einige Nicht-Übereinstimmungen sind entstanden, weil der Begriff des personalisierten Lernens nicht eindeutig definiert werden kann. Wann ist eine beschriebene Entwicklungstätigkeit im Zusammenhang mit personalisiertem Lernen zu sehen und wann nicht? Die zwei Codierenden einigten sich darauf, dass Aussagen codiert werden sollen, die eindeutig im Zusammenhang mit einem Konzept personalisierten Lernens stehen, weil beispielsweise entweder die Interviewfrage darauf abzielt oder die Interviewten konkret ausführen, wie sie Lernangebote schaffen, die vermehrt auf die personalen Voraussetzungen von Schülerinnen und Schülern abzielen. Ausgeschlossen sind hierbei etwa Interviewaussagen zur Einführung der Tagesschule. Nach der Überprüfung der Inter-coder-Übereinstimmung wurde das gesamte Datenmaterial codiert.

4) Überprüfung des Kategoriensystems

Nach dem Codieren des gesamten Datenmaterials wurde das Kategoriensystem abschließend daraufhin überprüft, ob sich die Kategorien inhaltlich eindeutig voneinander abgrenzen lassen.

5) Induktives Bestimmen von Subkategorien am Interviewmaterial

Anschließend folgt eine Arbeitsphase, in der Subkategorien induktiv, das heißt ausgehend vom Datenmaterial, gebildet und definiert wurden.

6) Codierung des kompletten Interviewmaterials mit den ausdifferenzierten Kategorien

Nun wurde mit dem ausdifferenzierten Kategoriensystem von Haupt und Subkategorien das gesamte Interviewmaterial ein zweites Mal codiert und es wurden allfällige inhaltliche Schärfungen der Kategorien vorgenommen sowie für jede Kategorie Ankerbeispiele definiert.

7) Datenanalyse des ersten Teils von Studie 1 (siehe Abschnitt 6.1)

Nachfolgend werden die Analyseschritte erläutert, mit denen die Rahmenbedingungen (Teilfrage 1.1), der Auslöser und der Beginn des Entwicklungsprozesses (Teilfrage 1.2), die organisationale Struktur (Teilfrage 1.3), die Entwicklungstätigkeiten im Erhebungszeitraum (Teilfrage 1.4) und die Metaziele (Teilfrage 1.5) beschrieben werden (Studie 1/Teil 1; siehe Abschnitt 1.3).

Für die Aufbereitung der Ergebnisse wurden zuerst in jedem Interview die codierten Aussagen pro Kategorie regelgeleitet zusammengefasst. Hierbei wurden für die Kategorie relevante Inhalte der jeweiligen Aussagen verdichtet. Falls in mehreren Interviewaussagen dieselben Inhalte beschrieben wurden, wurden diese zusammengenommen. Diese Zusammenfassungen wurden wie bereits festgehalten „Summaries“ genannt. Die Regeln der Zusammenfassung sind im Anhang E im elektronischen Zusatzmaterial dokumentiert. Der Vorteil der Bildung von Summaries besteht darin, dass die sehr umfangreichen Beschreibungen von Entwicklungstätigkeiten nun in kurzer, prägnanter und übersichtlicher Form vorliegen. Beispiele für die Erstellung solcher Summaries sind im Anhang zu finden (siehe Anhang F im elektronischen Zusatzmaterial).

Anschließend wurden die Summaries für jede Schule in einer Tabelle mit vier Spalten dargestellt. Eine Spalte entspricht einem Interview. In einem nächsten Schritt wurden regelgeleitet diejenigen Summaries ausgewählt, in denen thematisch ähnliche Entwicklungstätigkeiten in mindestens zwei Erhebungszeitpunkten beschrieben werden (Kriterien der Summary-Auswahl siehe Anhang G im elektronischen Zusatzmaterial). Ein Beispiel dieser Auswahl ist im Anhang (siehe Anhang H im elektronischen Zusatzmaterial) nachzulesen. Eine Reliabilitätsüberprüfung, die gleich verlief wie die oben unter Punkt 3 beschriebene Überprüfung der Intercoder-Übereinstimmung, zeigt bei 68 Auswahlentscheidungen eine Übereinstimmung und bei 10 Auswahlentscheidungen eine Nicht-Übereinstimmung. Dies ergibt eine als gut zu bewertende prozentuale Übereinstimmung von 87,18 Prozent. Bei Nicht-Übereinstimmungen wurde im konsensuellen Verfahren (Hopf & Schmidt, 1993; Schmidt, 2013) über die Auswahl diskutiert und die Kriterien wurden angepasst. Nicht-Übereinstimmungen sind besonders bei der Auswahl von Summaries mit Zusatzinformationen, die zwar an sich keine Entwicklung beschreiben, aber zum Verständnis der Entwicklungstätigkeiten beitragen, aufgetreten. Die zwei Codierenden einigte sich darauf, dass diese Summaries auch mit aufgenommen werden.

Mit der Auswahl der Summaries lagen nun beschriebene Entwicklungstätigkeiten entlang des Kategoriensystems vor. Mit dieser eher technischen Anordnung der Summaries war es dem Autor jedoch nicht zufriedenstellend möglich, die Schulen hinsichtlich ihrer jeweils eigenen Entwicklungstory zu beschreiben. Die Analysen zeigten, dass sich Tätigkeiten, die verschiedenen Kategorien zugeordnet worden waren, in der Praxis gegenseitig beeinflussen. Deswegen wurden die Summaries aus unterschiedlichen Kategorien bezüglich der von Schulleitenden und Lehrpersonen thematisierten Entwicklungsprojekte (z. B. Erarbeitung qualitätsvoller Lernaufgaben oder Zusammenlegung zweier Lerngruppen) angeordnet.

In einem letzten Schritt wurden auf der Grundlage der ausgewählten Summaries die fünf Forschungsfragen zu den Rahmenbedingungen (Teilfrage 1.1), dem Auslöser und dem Beginn des Entwicklungsprozesses (Teilfrage 1.2), der organisationalen Struktur (Teilfrage 1.3), den Entwicklungstätigkeiten im Erhebungszeitraum (Teilfrage 1.4) und den Metazielen (Teilfrage 1.5) in Fallbeschreibungen beantwortet (Studie 1/Teil 1; siehe Abschnitt 1.3).

8) Datenanalyse des zweiten Teils von Studie 1 (siehe Abschnitt 6.1): Bestimmung von (gemeinsamen) Entwicklungstätigkeiten (vor und im Erhebungszeitraum)

In der letzten Analyse von Studie 1 geht es darum, Entwicklungsgegenstände zu extrahieren und zu untersuchen, welche in allen Schulen bearbeitet werden. Neben Aussagen zum Erhebungszeitraum wurden nun auch Aussagen hinzugenommen, die sich auf die Zeit vor der ersten Erhebung beziehen. Entwicklungsgegenstände sind abstrahierte Entwicklungstätigkeiten, die einen Vergleich zwischen den Schulen zulassen. Zum Beispiel sind das die Erarbeitung von Stoffplänen oder die Erweiterung von Lehr- und Lernformen. Die Entwicklungsgegenstände gründen auf dem induktiv-deduktiv erarbeiteten Kategoriensystem (siehe Abschnitt 7.1.1 und Anhang D im elektronischen Zusatzmaterial).

Die Kategorien wurden dahingehend überprüft, ob sie dem Abstraktionsniveau entsprechen. Wenn dies der Fall war, dann wurde die betreffende Kategorie beibehalten. Wenn dies nicht zutraf, wurde nach inhaltlich sinnvollen Zusammenführungen von Kategorien gesucht. Beispielsweise wurde die Kategorie „Stoffpläne, Lehrpläne und Curricula“ beibehalten und in „Neuanordnung der Lerninhalte für alters- und/oder leistungsdurchmischte Lerngruppe“ umbenannt. Im Gegensatz dazu wurden dem Entwicklungsgegenstand „Anpassung der Lernbegleitung und -förderung an eine heterogene Lerngruppe im offenen Unterricht“ drei Kategorien zugeordnet: (1) „Lernförderung, -begleitung und -coaching der Schülerinnen und Schüler durch Lehrpersonen“, (2) „Gestaltung einer Arbeitsbeziehung zwischen Lehrpersonen und Schülerinnen und Schülern“ und (3) „Pädagogische Konzepte für ein lernförderliches Sozial- und Interaktionsklima“.

Anschließend wurde untersucht, welche der Entwicklungsgegenstände in allen Schulen beschrieben wurden. Das hieraus resultierende Ergebnis sind die gemeinsamen Entwicklungsgegenstände.

6.4.1.2 Gütekriterien und Verfahren der Qualitätssicherung der qualitativen Datenanalyse

Im Diskurs über Gütekriterien und die Qualität qualitativer Inhaltsanalysen argumentieren Kuckartz (2018) und Mayring (2015) für spezifisch angepasste Gütekriterien und Verfahren der Qualitätssicherung. Bei den in der vorliegenden

Untersuchung angewandten Gütekriterien und Verfahren handelt es sich um die folgenden: (1) Regelgeleitetes methodisches Vorgehen, (2) Verfahrensdokumentation für die intersubjektive Nachvollziehbarkeit, (3) Intercoder-Übereinstimmung, (4) konsensuelles Codieren, (5) Diskurs über Vorgehen und Ergebnisse und (6) Beschreibung der Limitation des methodischen Vorgehens.

Regelgeleitetes methodisches Vorgehen

Ein regelgeleitetes methodisches Vorgehen wurde eingehalten, indem die strukturierende Inhaltsanalyse gemäß den Analysenschritten von Kuckartz (2018) durchgeführt wurde.

Verfahrensdokumentation für eine intersubjektive Nachvollziehbarkeit

Grundlage für intersubjektive Nachvollziehbarkeit ist die Dokumentation des methodischen Vorgehens (u. a. Häder, 2010; Steinke, 1999, 2007, 2013). Diese wurde in der vorliegenden Forschungsarbeit mit zwei Dokumenttypen sichergestellt. Erstens erfolgte eine computergestützte Auswertung der transkribierten Interviews mit MAXQDA, weshalb die Zuordnungen der Fundstellen zu den Kategorien einsehbar sind. Zweitens gibt es für jede Studie ein detailliertes Forschungstagebuch, in dem die Auswertungen, die Codierregeln, die Regeln für die Erstellung und für die Auswahl von Summaries sowie Änderungen im Kategoriensystem kleinschrittig dokumentiert und begründet werden.

Intercoder-Übereinstimmung

Drittens wurde mit der Intercoder-Übereinstimmung ein Gütekriterium für die Reliabilität der Codierungen von Fundstellen eingesetzt. Der prozentuale Wert der Übereinstimmung und der Cohens-Kappa-Wert sind Masse für die inhaltliche Trennschärfe der Kategorien sowie für die Exaktheit der Codierregeln.

Konsensuelles Codieren

Das konsensuelle Codieren (Hopf & Schmidt, 1993; Schmidt, 2013) als weiteres Qualitätsmaß qualitativer Inhaltsanalyse kam beim Gegencodieren zum Einsatz, wenn eine Nicht-Übereinstimmung vorliegt. Hier bedurfte es einer Diskussion der Zuteilung, der Definition der Kategorien sowie der Codierregeln. Ziel war es, im Konsens eine einheitliche Zuordnung zu erreichen (Kuckartz, 2018; Mayring, 2015). Konsensuelles Codieren gilt gemeinhin als Merkmal für eine valide Interpretation der Ergebnisse: „Nur wenn intersubjektiver Konsens zwischen Auswertern besteht, kann eine Interpretation als gültig (valide) und wissenschaftlich abgesichert angesehen werden“ (Bortz & Döring, 2006, S. 327).

Expertenvalidierung des methodischen Vorgehens und der Ergebnisse

Ein weiteres Verfahren ist die Expertenvalidierung (Flick, 2014). Damit ist das Präsentieren und Diskutieren des methodischen Vorgehens und der Ergebnisse in der Wissenschaftsgemeinschaft gemeint. Auch die gemeinsame Dateninterpretation in Forschungsgruppen kann als Expertenvalidierung angesehen werden (Steinke, 2007, 2013). Seit 2014 wurden Teilaspekte der vorliegenden Untersuchung bereits in fünf Kolloquien zweier Lehrstühle des Instituts für Erziehungswissenschaft der Universität Zürich sowie in einem Kolloquium und einer Fachgruppensitzung der Pädagogischen Hochschule Zürich, in 13 Sitzungen mit der perLen-Forschungsgruppe sowie an elf Tagungen oder Kongressen in Form von Postern oder Fachvorträgen zur Diskussion gestellt.

Limitation des methodischen Vorgehens

Ein weiteres Qualitätskriterium für qualitative Inhaltsanalyse ist die Beschreibung der Limitation des methodischen Vorgehens (Janssen, Stamann, Krug & Negele, 2017). Sowohl die Interviews als auch die Online-Fragebogen erfassen Selbstbeschreibungen über die Entwicklung. Es ist anhand der Daten nicht rekonstruierbar, ob die befragten Personen dies auch tatsächlich so umsetzen, wie sie es berichten. Hierfür sind teilnehmende Beobachtungen oder Videostudien nötig. Des Weiteren wurde in den Interviews mit Schulleitenden ersichtlich, dass diese häufig einen Sprachduktus verwenden, mittels dessen sie ihr Unterrichts- und Schulkonzept bestmöglich „verkaufen“ wollen. Eine tiefgehende, kritische Reflexion der Schul- und Unterrichtsentwicklung war hingegen selten zu erkennen. Es scheint daher fraglich zu sein, inwiefern die Schulleitenden von ihrer persönlichen Sicht auf die unterrichtszentrierte Schulentwicklung in Richtung personalisierten Lernens in ihrer Schule berichteten oder ob sie eine „geschönte“ Darstellung vornahmen. In dieser Hinsicht waren die Gruppeninterviews mit den Lehrpersonen ergiebiger, weil die Lehrpersonen kritischer über ihre Entwicklungstätigkeiten berichteten. Des Weiteren ist die Generalisierbarkeit der Ergebnisse nur begrenzt möglich, weil erstens die Stichprobe selektiv gezogen wurde und zweitens die Stichprobe mit elf Schulen relativ klein ist.

6.4.2 Studie 2: Verfahren der transformativen Mixed-Methods-Datenanalyse

Eine transformative Mixed-Methods-Datenanalyse hat zum Ziel, entweder qualitative Daten durch Quantifizierung in quantitative Daten zu transformieren oder

quantitative Daten so aufzubereiten, dass sie für eine qualitative Datenauswertung nutzbar werden (Caracelli & Greene, 1993; Creswell & Plano Clark, 2018; Kuckartz, 2014, 2017; Onwuegbuzie & Teddlie, 2003; Prommer, 2018). In der zweiten Studie wird die erste Variante genutzt: Zu Beginn wurden mittels strukturierender Inhaltsanalyse (Kuckartz, 2018) Interviewaussagen zur Unterrichtsentwicklung anhand der Dimensionen personalisierten Lernens (siehe Abschnitt 3.2, u. a. Stebler et al., 2018). erneut codiert. Nach der Zusammenfassung der Fundstellen zu Summaries wurden diese den Merkmalen der Oberflächen- und Tiefenstruktur von Unterricht (siehe Abschnitt 4.5.1.3) zugeordnet. In einem nächsten Analyseschritt wurde das prozentuale Anzahlverhältnis zwischen den Summaries der Oberflächen- und Tiefenstruktur ermittelt (Abschnitt 6.4.2.1). Auf der Grundlage dieser Prozentwerte und der Werte zur Einschätzung des Entwicklungsstands aus dem Online-Fragebogen (t_3) wurden im letzten Schritt Zusammenhangsanalysen und eine Clusteranalyse durchgeführt (Abschnitt 6.4.2.2). Die diesen Analysen zugrunde liegenden methodischen Qualitätsstandards werden im letzten Unterkapitel beschrieben (Abschnitt 6.4.2.3).

6.4.2.1 Quantifizierung qualitativer Daten

Für eine Quantifizierung qualitativer Daten eignet sich eine qualitative Inhaltsanalyse deshalb, weil sich Interviewaussagen durch Auszählung der Kategorien, Fundstellen oder Summaries in Zahlenwerten erfassen lassen (Grunenberg & Kuckartz, 2013; Kuckartz, 2014, 2017; Mayring, 2012). Eine notwendige Voraussetzung für dieses Vorgehen besteht jedoch darin, dass in den Leitfadeninterviews, sowohl im Querschnitt als auch im Längsschnitt, dieselben Fragen gestellt werden (Morse & Maddox, 2014). Die im perLen-Projekt eingesetzten teilstrukturierten Leitfadeninterviews erfüllen diese Bedingung (siehe Abschnitt 6.3.1). Eine Überprüfung der Intercoder-Übereinstimmung während der Codierung gilt als eine weitere Voraussetzung für eine Quantifizierung (Mayring, 2012; Morse & Maddox, 2014).

Strukturierende Inhaltsanalyse

Die Fundstellen zur Unterrichtsentwicklung aus der Datenanalyse von Studie 1 (siehe Abschnitt 6.4.1) wurden mit einem weiteren, deduktiv erstellten Kategoriensystem erneut mit einer strukturierenden Inhaltsanalyse (Kuckartz, 2018) codiert. Dieses Kategoriensystem basiert auf fünf Dimensionen personalisierten Lernens:

- Unterrichtsangebote an Schülerinnen und Schüler und Lerngruppen sowie deren personale Voraussetzungen anpassen,

- personale und soziale Kompetenzen aufbauen und ganzheitliche Förderung der Schülerinnen und Schüler anstreben,
- selbstgesteuertes Lernen auf eigenen Wegen ermöglichen,
- kompetenzorientiertes Lernen zur persönlichen Sache machen und
- als Lehrperson und Lerngemeinschaft bildend und unterstützend wirken (siehe Abschnitt 3.2, u. a. Stebler et al., 2018).

Nach dem Definieren der Kategorien und der Codierregeln (siehe Anhang I im elektronischen Zusatzmaterial) wurden die ersten Interviews codiert, Kategorien induktiv erweitert und Codierregeln angepasst, bis das Kategoriensystem gesättigt war (Strübing et al., 2018). Während dieses Prozesses zeigte sich, dass Textstellen, in denen das Nutzungsverhalten der Schülerinnen und Schüler erläutert wird, nicht eindeutig codiert werden konnten. Dieses Problem wurde unter anderem in folgender Aussage manifest, in der eine Lehrperson ihr didaktisches Handeln zwischen dem alten und dem neuen Unterrichtsmodell mit vermehrt offenen Unterrichtsphasen, zum Beispiel in einer Lernlandschaft, vergleicht: „Die Schüler[innen] kommen [...] dich mehr fragen. Also gerade die schwächeren Schüler[innen]. [...] Im [...] alten [Unterrichtsmodell] [...] haben [sie] nie so viele Fragen gestellt oder dich um Hilfe gebeten wie jetzt im [offenen Unterricht]“ (LP_Schule-H-t2). Die Lehrperson beschreibt das Nutzungsverhalten der Schülerinnen und Schüler, welche auf die Lehrperson zugehen und Fragen stellen sowie um Hilfe bitten. Diese Aussage deutet darauf hin, dass die Schülerinnen und Schüler Verantwortung für ihren Lernprozess übernehmen. Gemäß dieser Interpretation würde sie in die vierte Dimension „Kompetenzorientiertes Lernen zur persönlichen Sache machen“ passen. Jedoch ist diese Dimension, wie die anderen vier Dimensionen auch, stärker aus der Perspektive des didaktischen Handelns der Lehrperson formuliert. Im Kontext eines Angebot-Nutzungs-Modells (siehe Abschnitt 4.3, Reusser & Pauli, 2010b) wird das Nutzungshandeln der Schülerinnen und Schüler für eine inhaltlich eindeutige Zuordnung teilweise zu wenig berücksichtigt. Dies rührt daher, dass das Lernangebot und die Lernnutzung in einem wechselseitigen, ko-konstruktiven Verhältnis zueinander stehen: Nutzungshandeln kann erst auftreten, wenn ein Lernangebot vorhanden ist. Ebenso bedingt ein Lernangebot, dass dieses genutzt wird.

Aufgrund dessen wurden die Dimensionen personalisierten Lernens ergänzt: Wenn Schulleitende oder Lehrpersonen explizit beschreiben, dass sie ein Lernangebot gestalten und entwickeln, dann wurde diese Aussagen entweder in die erste („Unterrichtsangebote an Schülerinnen und Schüler und Lerngruppen sowie deren personale Voraussetzungen anpassen“) oder fünfte Dimension („Als Lehrperson und Lerngemeinschaft bildend und unterstützend wirken“) codiert. Dagegen

wurde Nutzungsverhalten in die dritte („Selbstgesteuertes Lernen auf eigenen Wegen ermöglichen“) oder vierte Dimension („Kompetenzorientiertes Lernen zur persönlichen Sache machen“) zugeordnet. Die zweite Dimension „Personale und soziale Kompetenzen aufbauen und ganzheitliche Förderung der Schülerinnen und Schüler“ passt in keinen der beiden Bereiche, weil hier stärker die multikriterialen Bildungswirkungen (siehe Abschnitt 4.10) beschrieben werden. In dieser Kategorie wurden Aussagen codiert, in denen die interviewten Personen den Kompetenzerwerb der Schülerinnen und Schüler darstellen.

Nach dieser inhaltlichen Feinjustierung der Dimensionen für diese Analyse folgte der nächste Arbeitsschritt: die Prüfung der Inter-coder-Übereinstimmung. Sie verlief gleich wie in der strukturierenden Inhaltsanalyse von Studie 1 (siehe Abschnitt 6.4.1.1). Es ergab sich ein zufriedenstellender Wert von 77,27 Prozent Übereinstimmung (Cohens Kappa = .74). Auch hier wurden im konsensuellen Verfahren (Hopf & Schmidt, 1993; Schmidt, 2013) Nicht-Übereinstimmungen diskutiert. Nicht-Übereinstimmungen gab es beispielsweise bei Fundstellen, in denen die Interviewten ein Lernangebot und die sich daraus ergebenden Nutzungsmöglichkeiten beschreiben: „Ich denke, diese Freiheit [...], die [die Schülerinnen und Schüler] nun haben, ist zum einen Teil, dass sie ihre Arbeiten im [offenen Unterricht] gestalten können: die Reihenfolge, Länge, die Tiefe“ (LP_Schule-L_t1). Die Lehrperson erläutert, dass sie mit ihren Kolleginnen und Kollegen ein Lernangebot erstelle, in dessen Rahmen die Schülerinnen und Schüler die Reihenfolge, zeitliche Dauer und die Verarbeitungstiefe von Lernaufgaben selbst bestimmen können. Allerdings wird nicht beschrieben, dass diese Möglichkeiten von den Schülerinnen und Schülern auch tatsächlich genutzt werden, das heißt, dass sie im offenen Unterricht eigenverantwortlich die Lernaufgaben auswählen und bearbeiten. In der Diskussion erfolgte die Einigung dahingehend, dass diese Fundstelle zu den Dimensionen des Lernangebots codiert werden sollte, weil Ausführungen zu einem konkreten Nutzungsverhalten nicht vorzufinden sind.

In den nächsten zwei Arbeitsschritten wurde das restliche Datenmaterial codiert und die codierten Interviewaussagen wurden pro Interview und pro Kategorie, das heißt pro Dimension personalisierten Lernens, zu Summaries zusammengefasst (Kuckartz, 2018). Diese Verdichtung wurde anhand der Kategorieninhalte⁴ vorgenommen. Das heißt, wenn eine oder mehrere Textstelle(n) eines

⁴ Kategorieninhalte fungieren dabei als inhaltliche Entscheidungskriterien, die im Memo dokumentiert werden und auf deren Grundlage Textstellen der jeweiligen Kategorie zugeordnet werden. Dies ist bei der ersten Dimension („Unterrichtsangebot an Schülerinnen und Schüler und Lerngruppen sowie deren personale Voraussetzungen anpassen“) beispielsweise „Differenzierung von Lerninhalten hinsichtlich von Alters- und Leistungsniveaus“.

Interviews einem solchen Inhalt entsprach(en), wurden sie bezüglich dieses Inhaltes verdichtet. Während dieser Zusammenfassung wurde parallel dazu ein weiterer Analyseschritt durchgeführt, indem bei der Summariebildung gleichzeitig theoriegeleitet unterschieden wurde, ob die interviewten Personen die Entwicklung von Merkmalen der Oberflächen- oder Tiefenstruktur (siehe Abschnitt 4.5.1.3) von Unterricht berichtet.⁵ Zum Beispiel sind ein Merkmal der Oberflächenstruktur von vermehrt personalisierten Lehr- und Lernprozessen die vielfältigen Lehr- und Lernformen: „Ich persönlich lege viel Wert darauf, dass ich über eine Methodenvielfalt im Unterricht verfüge und es nicht nur Frontalunterricht ist, sondern alles Mögliche, x-verschiedene Formen, die es gibt. Ich probiere auch, viele spielerische Sachen einzubauen“ (LP_Schule-G_t1). Diese Fundstelle wurde wie folgt zusammengefasst: „Die Lehrpersonen setzen vielfältige Lehr- und Lernmethoden ein.“ Ein Merkmal der Tiefenstruktur von Unterricht hingegen sind kognitiv aktivierende und motivationsfördernde Lernumgebungen. Hierfür sind Lernaufgaben, die tiefgreifende Verstehensprozesse evozieren, bedeutsam:

Das [entwickeln] wir im Moment immer wieder, dass wir Aufgaben stellen, [bei denen] nicht nur etwas angekreuzt [wird], sondern dass der Auseinandersetzungsgrad höher liegt und die Verarbeitungstiefe grösser ist. Da sind wir einfach unterwegs. [...] Wir haben aber auch sicher noch viel zu lernen, es kann immer noch besser werden. (LP_Schule-G_t1)

Die Lehrperson bezieht sich hierbei auf Aufgabensammlungen. Dabei handelt es sich um eine Ansammlung von thematisch aufeinander abgestimmten Lernaufgaben, die von den Schülerinnen und Schülern größtenteils selbstständig und eigenverantwortlich im offenen Unterricht bearbeitet werden. Die Fundstelle wurde wie folgt zusammengefasst: „Die Lehrpersonen überarbeiten die Aufgabensammlungen in Richtung einer verstärkten kognitiven Auseinandersetzung mit den Lerninhalten.“

Quantifizierung der Summaries

Auf der Grundlage der Zusammenfassung der Interviewaussagen und der Unterteilung dieser Summaries in Merkmale der Oberflächen- und Tiefenstruktur von

⁵ Die Fundstellen wurden so zusammengefasst, dass sie entweder Merkmale der Oberflächen- oder der Tiefenstruktur von Unterricht enthielten. Dadurch entfiel eine Kontrolle der Intercoder-Übereinstimmung im Sinne einer Überprüfung der Zuordnung respektive Codierung von Summaries zur Oberflächen- oder Tiefenstruktur, weil die Kategorisierung bereits vom Autor vorgenommen worden war. Eine inhaltliche und sehr aufwendige Überprüfung in Bezug darauf, ob eine zweite Person die Interviewaussagen ähnlich zusammengefasst hätte wie der Autor, konnte im Rahmen der verfügbaren Ressourcen nicht geleistet werden.

Unterricht (siehe Abschnitt 4.5.1.3) wurde es durch Auszählungen möglich, absolute Häufigkeiten zu erhalten und relative Häufigkeiten zu berechnen. Wie bereits festgehalten wurden Summaries quantifiziert und nicht, wie in der Literatur vorgeschlagen (u. a. Kuckartz, 2014), Fundstellen, weil es in Summaries weniger zu Verzerrungen der Häufigkeiten aufgrund des Interviewverlaufs kommt. Die Häufigkeiten von Fundstellen können durch besondere Ereignisse im Interview (z. B. Störungen; Personen kommen während des Interviews hinzu oder verlassen dieses) und/oder den Interviewstil unterschiedlich ausgeprägt sein. So können Fragebereiche des Leitfadens in den Antworten je nach Interviewstil mehr oder weniger eingehend thematisiert werden und es kann weniger oder häufiger zwischen den Fragebereichen hin- und hergesprungen werden. Zum Beispiel weisen Interviews, die streng der Chronologie des Leitfadens entsprechen, tendenziell weniger Fundstellen auf als ähnlich lange Interviews, in denen zwischen Themenbereichen hin- und hergesprungen wird. Letztere lassen sich durch Merkmale von problemzentrierten Interviews (u. a. Witzel, 2000) charakterisieren, in denen auf einige Bereiche intensiv eingegangen wird, was teilweise zu Abschweifungen führen kann. Damit kann die Fundstellenanzahl auch ein Indikator dafür sein, wie ein Interview verlaufen ist.

Solche durch Fundstellenhäufigkeiten potenziell generierte Verzerrungen sind in großen, längsschnittlichen Forschungsprojekten wie zum Beispiel im qualitativen Teil des perLen-Forschungsprojekts wahrscheinlich, weil über mehrere Jahre viele und wechselnde Projektmitarbeitende und Hilfskräfte an der Datenerhebung mitwirken. Mit einer Schulung in der Interviewführung, der Besprechung des Leitfadens, einer einheitlichen Verfahrensdokumentation und einem Einblick in die qualitative Inhaltsanalyse lassen sich interindividuelle Unterschiede in der Interviewführung zwar reduzieren, was im perLen-Projekt auch erfolgt ist (Vasarik Staub et al., 2019). Allerdings ist eine starke Standardisierung qualitativer Datenerhebungen nicht möglich, weil die Interviews im sozialen Raum stattfinden, in dem Personen unbewusst und bewusst miteinander interagieren und jedes Interview wiederum für sich einzigartig ist (u. a. Flick, 2019).

Da sich während der Datenanalyse gezeigt hatte, dass die Streuung der absoluten Häufigkeiten relativ groß ist, wurde mit relativen Häufigkeiten anstatt mit absoluten Häufigkeiten gerechnet. So liegen beispielsweise zu Schule D insgesamt 66 Summaries vor, zu Schule M mit 123 hingegen fast doppelt so viele (Tabelle 7.7 in Abschnitt 7.4.2). Diese absoluten Unterschiede können zwar dahingehend gedeutet werden, dass in einer Schule mit mehr Summaries vielfältigere Aspekte der Oberflächen- und Tiefenstruktur von Unterricht beschrieben werden, jedoch hängt diese Anzahl auch mit den Rahmenbedingungen des Interviews zusammen, zum Beispiel mit der Länge, der Anzahl interviewter Personen, dem

störungsfreien Verlauf etc. Mit prozentualen Häufigkeiten können solche „störenden“ Einflüsse zwar nicht ausgeschlossen, aber zumindest reduziert werden, während sich die Vergleichbarkeit der Schulen erhöht.

6.4.2.2 Quantitative Datenanalyse

Während die Gesprächsanteile, die den Merkmalen der Tiefenstruktur zugeordnet werden können, im eben beschriebene Analyseschritt für jede Schule ermittelt wurden, müssen die Individualwerte des Entwicklungsstands auf Schulebene aggregiert werden. Zur Prüfung der Eignung der Daten für die Mittelung der Kennwerte pro Schule wurde auf die Intraklassenkorrelation (ICC) zurückgegriffen: ICC(1) und ICC(2). Diese Verfahren werden in der Mehrebenenanalyse eingesetzt. Anschließend wurde für die Untersuchung des Zusammenhangs zwischen den zwei Variablen das nichtparametrische Verfahren der Rangkorrelation nach Spearman angewendet. Als letztes Verfahren der quantitativen Datenanalyse kam die hierarchische Clusteranalyse (Ward-Methode) zum Einsatz.

Intraklassenkorrelation ICC(1) und ICC(2)

Die Intraklassenkorrelation gibt Hinweise darauf, wie geeignet die Daten für die Mittelung auf Schulebene sind. In der Mehrebenenanalyse wird zwischen ICC(1) und ICC(2) unterschieden. Der ICC(1)-Wert gibt an, wieviel Varianz der gesamten Stichprobe durch die Gruppenzugehörigkeit aufgeklärt werden kann. In der vorliegenden Arbeit interessiert, inwiefern sich die 124 Personen, die ihren Entwicklungsstand einschätzten, aufgrund der Schulzugehörigkeit unterscheiden lassen. Die ICC(1) wurde mit Mplus (Version 8.4) berechnet. Mit dem Syntaxbefehl „*type = twolevel*“ wurde der Wert direkt ausgegeben. In Mplus wird die ICC(1) aus der Varianzen innerhalb der Gruppen (*within*; W) und der Varianz zwischen den Gruppen (*between*; B) berechnet: $B/(W + B)$ (Kleinke, Schlüter & Christ, 2017; Muthén, 1991). Der Wertebereich reicht von 0 bis 1. Je höher der aufgeklärte Varianzanteil ist, desto besser eignen sich die Daten. Bei Werten unter .01 wird von einer Aggregation abgeraten (Bliese, 1998).

Für die Berechnung der ICC(2) wurde die Formel von Bliese (2000) verwendet:

$ICC(2) = (k \cdot ICC(1)) / (1 + (k-1) \cdot ICC(1))$. „K“ steht für die durchschnittliche Gruppengröße. Cicchetti und Sparrow (1981) und von Fleiss (1981) interpretieren die Güte der ICC(2)-Werte wie folgt: Kleinere Werte als .4 sind schlecht, Werte zwischen .4 und .59 sind durchschnittlich, Werte zwischen .6 und .74 sind gut und Werte über .75 sind sehr gut. Teils strengere Grenzwerte bestimmen Koo und Li (2016). Sie beurteilen kleinere Werte als .5 als schlecht, Werte zwischen .5 und

.75 als durchschnittlich, Werte zwischen .75 und .9 als gut beurteilt und Werte über .9 als exzellent.

Rangkorrelation: Spearmans Rho

In dieser Arbeit wurde das nichtparametrische Verfahren „Spearmans Rho“ eingesetzt. Dieses ist weniger voraussetzungsreich als beispielsweise die Bravais-Pearson-Korrelation und eignet sich daher besser für kleinere Stichproben. Zum Beispiel reagiert das Verfahren robust auf Ausreißer. Ausreißer können besonders in kleinen Stichproben den Korrelationskoeffizient stark beeinflussen (Bortz & Schuster, 2010; Field, 2018).

Die Rangkorrelation untersucht lineare und ungerichtete Zusammenhänge und sie wurde mit dem Programm IBM SPSS Statistics 26 gerechnet. Die Daten werden zunächst in einer Rangreihenfolge angeordnet und mit den Rangnummern wird anschließend der eigentliche Test durchgeführt. Daher werden die absoluten Abstände zwischen den Werten nicht berücksichtigt. Zwei Variablen können linear miteinander positiv respektive negativ kovariieren. Eine positive Korrelation gibt an, dass ein Wert einer Variable höhere oder tiefere Ausprägungen aufweist, wenn der Wert der anderen Variable höher oder tiefer wird. Bei negativen Korrelationen gehen höhere Werte in einer Variable mit tieferen Werten in der anderen einher und umgekehrt. Welche Variable am Anfang einer Wirkungskette steht, kann mit dem Verfahren nicht ermittelt werden. Der Korrelationskoeffizient, welcher die Stärke des Zusammenhangs angibt, liegt zwischen -1 und + 1. Ist der Wert kleiner als 0, liegt ein negativer Zusammenhang vor. Bei größeren Werten als 0 ist der Zusammenhang positiv. Bei $r = 0$ besteht kein Zusammenhang. Des Weiteren muss geprüft werden, ob sich der Korrelationskoeffizient signifikant von 0 unterscheidet. Hier werden drei Signifikanzniveaus unterschieden: $.1 < p < .05$ signifikant, $.001 < p < .01$ sehr signifikant und $p < .001$ hoch signifikant (Bortz & Schuster, 2010; Field, 2018; Hirsig, 2007a).

Hierarchische Clusteranalyse mit Ward-Methode

Die Clusteranalyse ist ein heuristisches Verfahren, das es erlaubt, systematische Klassifizierungen in Daten zu erkennen. In der vorliegenden Forschungsarbeit wurde die Clusteranalyse als exploratives Verfahren eingesetzt. Materialgestützt wurden mit dem Programm IBM SPSS Statistics 26 Strukturen in den Daten erkundet und plausible Clusterlösungen gebildet. Darauf aufbauend wurden Hypothesen formuliert, die jedoch in Folgestudien zu überprüfen sind.

Mit elf Schulen ist die Stichprobe in dieser Arbeit klein. Gerade agglomerative hierarchische Verfahren wie die Ward-Methode, die zwischen Merkmalen und Systemen unterscheiden, eignen sich für solch kleine Stichproben (Bacher, 1996;

Wiedenbeck & Züll, 2010). Die Merkmale beziehen sich auf die folgenden zwei Variablen: den prozentualen Summary-Anteil der Tiefenstruktur von Unterricht und den eingeschätzten Entwicklungsstand. Die Systeme sind die elf untersuchten Schulen.

In der Clusteranalyse werden diese Merkmale der einzelnen Systeme so in Clustern gruppiert, dass die Systeme innerhalb eines Clusters möglichst ähnlich sind und die Unterschiede zwischen den Clustern möglichst groß ausfallen. Es gibt zahlreiche Verfahren der Clusteranalyse. In dieser Forschungsarbeit wurde das Verfahren der hierarchischen Clusteranalyse mit der Ward-Methode eingesetzt. Im ersten Schritt der hierarchischen Clusteranalyse wurde jedes System, das heißt jede Schule, als ein Cluster definiert. Für jedes Cluster-Paar wurde im zweiten Schritt ein quadriertes euklidisches Distanzmaß auf der Grundlage der zwei Variablen ermittelt. Die Cluster, welche im Distanzmaß minimal voneinander abweichen, wurden zu einem neuen Cluster fusioniert. Dieses Vorgehen wurde so lange wiederholt, bis nur noch ein Cluster übrig blieb.

In einem nächsten Schritt wurde ein Kriterium, das Summe-Abweichungs-Quadrat (SAQ), eingesetzt, das vorgibt, welche Clusterlösung die beste ist. Der SAQ-Wert ist vor der ersten Fusion 0. Bei jeder Fusionsstufe wird für alle möglichen Gruppierungen der Cluster der Betrag der SAQ-Werte (Δ SAQ) berechnet. Auf der Fusionsstufe, die vor derjenigen liegt, auf der der Δ SAQ-Wert am stärksten steigt, kann die Clusteranalyse abgebrochen werden. Ist die Clusterlösung zudem inhaltlich begründbar, kann sie angenommen werden (Bortz & Schuster, 2010; Hirsig, 2007b).

Um die Güte der Clusterlösung zu überprüfen, wurde die Silhouetten-Analyse von Rousseeuw (1987) gerechnet. Sie ist ein Maß für Clusterhomogenität, das heißt, sie gibt an, wie gut sich die untersuchten Schulen einem Cluster zuordnen lassen. Hierzu wurde zuerst das Silhouetten-Maß für jede Schule mit den euklidischen Distanzmassen berechnet. Das Silhouetten-Maß eines Falls errechnet sich aus der mittleren Distanz zu allen anderen Schulen desselben Clusters (A) und der mittleren Distanz zu allen anderen Schulen des nächstliegenden Clusters (B):

$(B-A) / \max(A,B)$. Es sind Werte von -1 bis 1 möglich (König, 2018). Je höher die Werte, desto besser die Clusterstruktur. Nach Kaufman und Rousseeuw (1990) ist keine Struktur vorhanden, wenn das Silhouetten-Maß unter .25 liegt. Eine schwache Struktur ist zwischen .25 und .5, eine mittlere Struktur zwischen .51 und .75 sowie eine starke Struktur ab .751 erkennbar. Die Silhouetten-Masse können pro Cluster und für die gesamte Clusterlösung zu Silhouetten-Koeffizienten gemittelt werden (König, 2018).

6.4.2.3 Gütekriterien und Verfahren der Qualitätssicherung in der Mixed-Methods-Datenanalyse

In der Mixed-Methods-Forschung besteht weitgehend der Konsens, dass die Qualitätskriterien und Verfahren der Qualitätssicherung sowohl der qualitativen als auch der quantitativen Forschung im jeweiligen Forschungsstrang gelten (u. a. Onwuegbuzie & Johnson, 2006; Plano Clark & Ivankova, 2016). In Studie 2 wurden im qualitativen Teil die Verfahren der Qualitätssicherung eingesetzt (siehe Abschnitt 6.4.1.2). In der quantitativen Datenanalyse wurden angemessene und robuste statistische Tests verwendet sowie die Kriterien der Objektivität, Validität, Reliabilität und Replizierbarkeit eingehalten (Creswell & Plano Clark, 2018).

Darüber hinaus bestehen in der Mixed-Methods-Forschung Bestrebungen, eigene Gütemaßstäbe für das Zusammenführen qualitativer und quantitativer Datenanalysen zu definieren. Aufgrund der zahlreichen Kombinationsmöglichkeiten mit unterschiedlicher Gewichtung qualitativer und quantitativer Studienteile (hierzu u. a. Kuckartz, 2014) ist es bisher nicht gelungen, einheitliche Gütekriterien zu definieren (O’Cathain, 2010). Stattdessen haben sich vermehrt Verfahren der Qualitätssicherung etabliert. Dabei handelt es um eher allgemein gehaltene Beschreibungen, die darlegen, wie eine Mixed-Methods-Analyse durchgeführt und dokumentiert werden sollte, um als qualitativvoll zu gelten. So ist beispielsweise relevant, ob sich mit einem solchen Zusammenführen die Forschungsfrage beantworten lässt und dies mit der theoretischen Grundlage vereinbar ist. Des Weiteren sollten die eingesetzten Analyseverfahren und das Mixen zu dem vorhandenen Datensatz passen und es muss schlüssig argumentiert werden können, warum das entsprechende Vorgehen gewählt wurde und welchen Mehrwert dieses mit sich bringt (Bryman, Becker & Sempik, 2008; Creswell & Plano Clark, 2018; Plano Clark & Ivankova, 2016). Einige Autorinnen und Autoren fassen solche Verfahren aufbauend auf Literatur-Reviews in Qualitätsrahmen (*quality frameworks for mixed methods research*) zusammen. Beispielsweise arbeitete O’Cathain (2010) die folgenden acht Maßnahmen heraus:

- Qualität der Planung (*planning quality*): Der Plan der Datenanalyse ist durchführbar, einsehbar sowie verständlich und knüpft an methodischen und theoretischen Erkenntnissen an.
- Qualität des Designs (*design quality*): Das Design der Datenanalyse wird detailliert beschrieben, ist für das Forschungsziel zielführend und umfasst Methoden, die sich gegenseitig ergänzen.
- Qualität der Daten (*data quality*): Die Datenanalyse ist der Datenqualität angemessen und wird methodisch streng durchgeführt.

- Strenge der Interpretation (*interpretive rigor*): In der Diskussion und der Interpretation der Ergebnisse wird besonders auf konsistente und inkonsistente Ergebnisse zwischen den qualitativen und quantitativen Datenanalysen eingegangen.
- Übertragbarkeit der Schlussfolgerungen (*inference transferability*): Des Weiteren wird diskutiert, inwiefern die Ergebnisse auf andere Situationen, Gruppen und Zeitpunkte übertragbar sind.
- Qualitätsbericht (*reporting quality*): Es wird aufgezeigt, inwiefern der Mix aus qualitativen und quantitativen Datenanalysen einen Mehrwert für die gesamte Studie darstellt.
- Verwendbarkeit für Metastudien und Reviews (*synthesizability*): Die Studie hält Qualitätskriterien und Verfahren der Qualitätssicherung ihres Forschungsbereichs und der gewählten Methoden ein, sodass sie für Reviews oder Metaanalysen genutzt werden kann.
- Nützlichkeit (*utility*): Die Studie erbringt einen Nutzen für die Zielgruppe.

Limitation des methodischen Vorgehens

Die Limitationen der qualitativen Datenanalyse dieser Studie sind dieselben wie bereits in Studie 1 beschrieben: Es wurden Beschreibungen der Interviewten analysiert und nicht die tatsächlichen, direkt beobachtbaren Entwicklungsprozesse. Ebenso ist der inhaltliche Wert der Interviews mit Schulleitenden etwas geringer einzuschätzen, weil im Vergleich mit den Lehrpersoneninterviews kritische Reflexionen der Entwicklungsprozesse häufig ausblieben. Die Frage der Generalisierbarkeit stellt sich angesichts der selektiven und kleinen Stichprobe nicht nur bei der qualitativen Datenanalyse, sondern besonders bei quantitativen Auswertungsverfahren. So können Tendenzen aufgezeigt werden, jedoch keine generalisierbaren Schlussfolgerungen gezogen werden.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

