

Falkenberg, Kathleen

Permanenter Vergleich. Methodologische Überlegungen zu einer an der Grounded-Theory-Methodologie orientierten international vergleichenden Forschung

Tertium comparationis 24 (2018) 1, S. 107-134



Quellenangabe/ Reference:

Falkenberg, Kathleen: Permanenter Vergleich. Methodologische Überlegungen zu einer an der Grounded-Theory-Methodologie orientierten international vergleichenden Forschung - In: Tertium comparationis 24 (2018) 1, S. 107-134 - URN: urn:nbn:de:0111-pedocs-246859 - DOI: 10.25656/01:24685

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-246859>

<https://doi.org/10.25656/01:24685>

in Kooperation mit / in cooperation with:



WAXMANN
www.waxmann.com

<http://www.waxmann.com>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft



Permanenter Vergleich. Methodologische Überlegungen zu einer an der Grounded-Theory-Methodologie orientierten international vergleichenden Forschung

Kathleen Falkenberg

Humboldt-Universität zu Berlin

Abstract

This article discusses the benefits of employing Grounded Theory Methodology (GTM) in international comparative studies from a methodological point of view. After a short introduction to the main characteristics of GTM, the article focuses on three methodological issues concerning GTM and comparative studies: First, the question of case selection and theoretical sampling, second, the parallels of the constant comparative method and comparison as a method, and third, the concepts of tertium comparationis and core category. All three aspects are discussed in relation to a study on Swedish and German teachers' justice beliefs regarding assessment in school. GTM, it is argued, is a tool to face well-known challenges of comparative research, such as methodological nationalism, thereby contributing to ongoing discussions about relating methodological principles of qualitative research to comparative research.

1. Einleitung

Dass der Blick in die Welt und über den eigenen nationalen Tellerrand hinaus eine fruchtbare und inspirierende Perspektivenerweiterung mit sich bringen kann, gehört zunehmend zu den allgemeinen Einsichten erziehungswissenschaftlicher Forschung auch außerhalb der Teildisziplin der International Vergleichenden Erziehungswissenschaft (vgl. Waldow, 2015). Wie eine solche vergleichende Untersuchung angelegt sein kann und wie genau die Vergleichsoperationen vorgenommen werden, gehört hingegen zu den weniger trivialen Fragen, auf die sehr unterschiedliche Antworten gefunden werden können (für einen Überblick über verschiedene Ver-

gleichsmethodologien vgl. u.a. Adick, 2014; Parreira do Amaral, 2015). Vergleichsstudien, die sich auf einige wenige Kontexte konzentrieren und auf qualitativ-rekonstruktive Methoden zurückgreifen, stellen dabei eine vielversprechende Möglichkeit dar, die Komplexität und Vielfalt des interessierenden Phänomens abbilden zu können. Als fall- bzw. kontextsensible Vergleiche (vgl. Steiner-Khamisi, 2010) angelegt, folgen sie dem Diktum „context matters“ (Crossley & Jarvis, 2001) und ermöglichen eine intensive Durchdringung lokaler Spezifiken, von denen in Vergleichsstudien mit sehr vielen Vergleichseinheiten zwangsläufig abstrahiert werden muss.

Als Herausforderung jeglicher vergleichender Forschung stellt sich die grundsätzliche Frage nach dem Einfluss eigener sozio-kultureller Prägungen der Forschenden auf das Verständnis anderer sozialer Wirklichkeiten. Vergleichende Untersuchungen unterliegen dabei immer zweierlei Gefahren: einerseits der Gefahr der Nostrifizierung – also dem Versuch das ‚Andere‘ dem ‚Eigenen‘ anzugleichen – und andererseits der Gefahr einer Altrifizierung – also der Verfremdung des ‚Anderen‘ durch eine Überbetonung von Unterschieden (Parreira do Amaral, 2015, S. 124). Hier setzt auch die Kritik an einem essenzialisierenden Kulturbegriff an, der verschiedene (meist national gefasste) Kontexte als in sich homogene Entitäten konstruiert und diese einander vergleichend gegenüberstellt; der Nationalstaat (und dessen Bildungssystem) wird hier zum „Containerbegriff“ (Wimmer & Glick-Schiller, 2002, S. 307 f.) mit dem sich Unterschiede erklären lassen. Speziell im Rahmen ethnografisch-vergleichender Untersuchungen wird die Frage der Konstruktion und Wahrnehmung des ‚Eigenen‘ bzw. ‚Anderen‘ auch unter der Perspektive der (Selbst-)Positionierung von Forschenden im Forschungsprozess thematisiert und als Insider-Outsider-Perspektive verhandelt (vgl. McNess, Arthur & Crossley, 2015; Milligan, 2016).

Der vorliegende Beitrag diskutiert dagegen einen möglichen Ansatz, wie die Potenziale von kontextsensiblen qualitativ-rekonstruktiven Vergleichen durch die Orientierung an der Grounded-Theory-Methodologie noch stärker genutzt werden können, um neben den Gefahren der Altri- und Nostrifizierung auch einem vor-schnellen Rückgriff auf vermeintlich ‚kulturelle‘ Erklärungsmuster für vorgefundene Unterschiede und damit einem ‚methodologischen Nationalismus‘ (Wimmer & Glick Schiller, 2002; Dale, 2005) vorzubeugen. Anschließend an Überlegungen zur ‚Wahlverwandtschaft von qualitativer Forschung und Kulturvergleich‘ (Hummrich & Rademacher, 2012) verfolgt der Beitrag vordergründig eine methodologische Ausrichtung, basiert jedoch auf Analyseergebnissen aus einer kürzlich abgeschlossenen Studie über die Gerechtigkeitsüberzeugungen schwedischer und deutscher Lehrkräfte in Bezug auf die schulische Leistungsbeurteilung (Falkenberg, 2017).¹ Diese Studie dient dabei der Veranschaulichung der vergleichsmethodologischen

Überlegungen zu einer spezifisch an der Grounded-Theory-Methodologie orientierten vergleichenden Forschung, die sich auf einige wenige makrosoziale Vergleichseinheiten fokussiert.

Dafür werden zunächst einige grundlegende Charakteristika des Forschungsstils der GTM skizziert (Abschnitt 2), um im Weiteren einige zentrale methodologische Grundannahmen der GTM in Bezug zu qualitativen Vergleichsstudien zu setzen (Abschnitt 3). Dabei gehe ich auf die zentrale Frage der Fallauswahl und Wahl der Vergleichseinheiten ein, erläutere die Parallelen der *constant comparative method* der GTM und des ‚Vergleichs als Methode‘ (Schriewer, 1987) und zeige schließlich, inwiefern die zentralen Konzepte der Kernkategorie und des *tertium comparationis* aufeinander bezogen sind. Diese methodologischen Überlegungen werden jeweils anhand von Beispielen aus der oben genannten Studie über Gerechtigkeitsüberzeugungen von Lehrkräften illustriert. Im letzten Abschnitt (4) diskutiere ich die Möglichkeiten und Grenzen einer solchen Vorgehensweise zusammenfassend in Bezug auf allgemeine Herausforderungen vergleichender Forschung.

2. Grundzüge der Grounded-Theory-Methodologie(n)

Mit der Grounded-Theory-Methodologie (GTM) ist ein *spezifischer Forschungsstil* innerhalb des qualitativ-rekonstruktiven Forschungsspektrums gemeint, der auf die Generierung einer gegenstandsverankerten Theorie über ein bisher unbekanntes bzw. ‚unbeforschtes‘ soziales Phänomen abzielt. Die GTM wurde in den 1960er-Jahren als Gegenentwurf einer damals vorrangig hypothetisch-deduktiv geprägten Sozialwissenschaft entwickelt, die von sogenannten *grand theories* dominiert war. Die beiden Gründerväter Anselm Strauss und Barney Glaser formulierten im Anschluss an die Grundideen ihres einzigen gemeinsamen Werks ‚The discovery of Grounded Theory‘ (Glaser & Strauss, 1967) ihr jeweils eigenes Verständnis der Grounded-Theory-Methodologie in den folgenden Jahrzehnten getrennt voneinander immer weiter aus. Mittlerweile existieren durch die Weiterführungen in Zusammenarbeit mit ihren Schülerinnen und Schülern (Strauss & Corbin, 1996) und Forschende der sogenannten *second generation* verschiedene GTM-Varianten (vgl. u.a. Clarke, 2012; Charmaz, 2006; Bryant & Charmaz, 2007). Trotz dieser Ausdifferenzierungen und Weiterentwicklungen können aber einige Grundelemente benannt werden, die allen Varianten gemein sind.

Dies betrifft zum einen den Forschungsprozess selbst, in dem die Phasen von Forschungsplanung, Datenerhebung und -auswertung sowie Theoriegenerierung nicht zeitlich nacheinander ablaufen, sondern in einem *iterativ-zyklischen Forschungsprozess* miteinander verschränkt bzw. in einem ständigen Wechsel vollzogen werden. Die Datengenerierung erfolgt dabei in mehreren Wellen, die jeweils

durch parallel verlaufende Kodier- und Analyseprozesse unterbrochen und neu ausgerichtet wird. Die Datenanalyse beginnt sehr früh im Forschungsprozess, direkt nach der ersten Datenerhebung. Die in dieser Phase entwickelten ersten Konzepte und Ideen dienen in ihrer Vorläufigkeit als Hinweise für die weitere Datenerhebung. Sie stellen den Startpunkt der Analyse dar und werden im Verlauf des Forschungsprozesses immer weiter präzisiert und in die kontinuierliche Hypothesen- und Theoriegenerierung eingespeist. Bereits existierenden Theorien und bestehenden Forschungsergebnissen kommen hierbei die Rolle *sensibilisierender Konzepte* (vgl. Strauss & Corbin, 1996, S. 19–31) zu, die die vorläufige Konstituierung des Untersuchungsgegenstands beeinflussen, im Verlauf des Forschungsprozesses jedoch immer wieder überprüft und gegebenenfalls verworfen sowie durch aus dem Material generierte Konzepte ergänzt werden. Die Forschenden wechseln damit ständig zwischen dem Forschungsfeld bzw. der Datenerhebung im Feld und der theoretischen Reflexion über die beobachteten Phänomene des Erhebungsfelds.²

Zum anderen erfolgt die Datengenerierung nach der Grounded-Theory-Methodologie nicht anhand eines vorab festgelegten Erhebungsplans, sondern sukzessive nach dem Prinzip des *theoretical samplings* (vgl. Glaser & Strauss, 1967). Da es für die Generierung einer Theorie über ein bisher unbekanntes Phänomen kaum Informationen über die Grundgesamtheit, mögliche Merkmalsausprägungen oder die Merkmalsverteilung innerhalb der Grundgesamtheit gibt, können vor Beginn der Studie auch keine Stichprobengrößen festgelegt werden, wie beispielsweise bei hypothesenprüfenden Verfahren üblich. Die Datengenerierung richtet sich allein danach, welche vorläufigen Ergebnisse aus der bisherigen Datenauswertung vorliegen und welche offenen Fragen sich aus dieser Auswertung für die nächste Erhebungsphase ergeben. Glaser und Strauss (1967, S. 49) sprechen in diesem Zusammenhang auch von der *theoretischen Relevanz von Daten* bzw. Vergleichsgruppen. Ziel des theoretischen Samplings ist es, die Theoriegenerierung sukzessive auszubauen, indem systematisch nach Minimal- bzw. Maximalkontrasten gesucht wird. Auf diese Weise werden sukzessive solange neue Daten generiert und in die Analyse einbezogen, bis keine neuen Einsichten und Erweiterungen der Theorie mehr zu erwarten sind, und es lediglich um die Verbesserung der Konsistenz einer Theorie (bzw. der zentralen Kernkategorie) gehen kann. An diesem Punkt spricht man von *theoretischer Sättigung*, die Datenerhebung wird abgeschlossen und die entwickelte gegenstandsverankerte Theorie formuliert.³

Ein weiteres Kernelement des Forschungsprozesses nach der Grounded-Theory-Methodologie ist die *Methode des permanenten Vergleichens* oder *constant comparative method* (Glaser & Strauss, 1967, S. 101–115). Diese umfasst einerseits die bereits geschilderten Vergleichsprozesse bei der Fallauswahl und Integration neuer Daten in die sich entwickelnde Theorie und charakterisiert andererseits den Kodier-

und Auswertungsprozess. Mit Hilfe der *constant comparative method* werden Vergleichsprozeden auf mehreren Ebenen vollzogen: Teile des empirischen Materials werden untereinander, aber auch mit bestehendem (theoretischem) Vorwissen verglichen, bereits vergebene Codes mit neu entstehenden, diese Codes wiederum mit zusammenfassenden Kategorien, usw. Diese iterativen Vergleichsprozesse stellen einen wichtigen Schritt auf dem Weg zur Konzeptualisierung des interessierenden Phänomens dar, indem sukzessive die Eigenschaften der Kategorien herausgearbeitet werden und somit Lücken in der entstehenden Theorie sichtbar werden können, die wiederum Hinweise darauf geben, in welcher Art das theoretische Sampling fortgeführt werden muss, um diese Lücken zu schließen. Auf die konkreten Kodierprozeduren, verschiedene Formen von Codes sowie deren Konsequenzen für vergleichend angelegte Untersuchungen wird in Abschnitt 3.2 näher eingegangen.

Um die zahlreichen Entscheidungen während der Erhebungs- und Auswertungsphasen nachvollziehen zu können, aber auch um sich über die Bedeutung von und die Zusammenhänge zwischen Kategorien klar zu werden, empfehlen Glaser und Strauss (1967, S. 107), das Kodieren regelmäßig zu unterbrechen und die Gedanken in Form von *Memos* schriftlich festzuhalten. *Memos* dienen dabei aber nicht nur der Selbstreflexion der oder des Forschenden, sondern auch der kontinuierlichen Dokumentation des Forschungsprozesses und stellen damit ein Element der Qualitätssicherung von GTM-Studien dar. In ihnen werden die meist impliziten Interpretations- und Vergleichsprozesse expliziert, aus denen später die Theorie entwickelt wird. Neben expliziten Analyse*memos* können auch Planungs- und Erhebung*memos* die nächsten Erhebungsschritte zu klären helfen. Vor allem aber für die schriftliche Ausformulierung einer Theorie und der damit verbundenen Klärung von Begriffen, Zusammenhängen und Beziehungen sind *Memos* ein unverzichtbarer Teil des Forschungsprozesses.

Nach dieser kurzen Einführung in die Grundbegriffe der GTM werden im folgenden Abschnitt drei Aspekte hinsichtlich der methodologischen Parallelen zwischen GTM und Forschung innerhalb der Vergleichenden und Internationalen Erziehungswissenschaft am Beispiel einer Studie zu Gerechtigkeitsüberzeugungen von Lehrkräften näher beleuchtet. Dafür werden zunächst das Studiendesign und die zentralen Ergebnisse dieser Studie skizziert.

3. Methodologische Parallelen zwischen GTM und qualitativen Vergleichsstudien – Am Beispiel einer Untersuchung zu Gerechtigkeitsüberzeugungen von Lehrkräften

Die folgenden vergleichsmethodologischen Überlegungen zur Orientierung an einer Grounded-Theory-Methodologie in qualitativen Vergleichsstudien entstanden im Rahmen einer Untersuchung von Gerechtigkeitsüberzeugungen schwedischer und deutscher Lehrkräfte in Bezug auf die schulische Leistungsbeurteilung (vgl. Falkenberg, 2017). In beiden Ländern gehört die schulische Leistungsbeurteilung zu den zentralen Aufgaben von Lehrkräften, und die von ihnen vergebenen Zensuren und Zeugnisse entscheiden letztlich über den Zugang zu weiterführenden Bildungseinrichtungen und/oder dem Arbeitsmarkt, wodurch Lebenschancen eröffnet, aber auch verwehrt werden können.

Anders als die bisherige Forschung zum Thema gerechter Leistungsbeurteilung, die vor allem auf Fragen der diagnostischen Güte von Lehrerinnen- und Lehrerurteilen einerseits sowie die Frage der Verteilung von Bildungsabschlüssen auf unterschiedliche Gruppen von Schülerinnen und Schülern andererseits fokussierten (vgl. für einen Überblick u.a. Terhart, 2011), widmet sich diese Studie hingegen den Gerechtigkeitsüberzeugungen von Lehrkräften als zentralen Akteuren der Leistungsbeurteilung. Im Anschluss an sozialkonstruktivistische Positionen (Berger & Luckmann, 1977) wird davon ausgegangen, dass ‚Gerechtigkeit‘ ein sozial konstruiertes Phänomen darstellt, das in verschiedenen historischen, soziokulturellen oder nationalen Kontexten unterschiedliche Bedeutungen annehmen kann, weshalb ein Vergleich der Überzeugungen von Lehrkräften aus zwei verschiedenen Bildungssystemen vorgenommen wurde (zur ausführlichen Begründung der Fallauswahl vgl. Abschnitt 3.1). Als hauptsächliche Datengrundlage dienten 44 episodische Interviews (vgl. Flick, 2011) mit Lehrkräften aus Schweden und Nordrhein-Westfalen (NRW), die am Ende der Sekundarstufe I in einem der Kernfächer Mathematik bzw. Landessprache unterrichteten. Ebenfalls in die Analyse einbezogen wurden offizielle Regularien (Gesetze, Handreichungen für Lehrkräfte, Verordnungen) aus dem Zeitraum 2010–2016 und von den Lehrkräften selbst angefertigte Beurteilungsmaterialien. Ziel war es, die komplexen Wechselbeziehungen zwischen institutionellen Rahmungen, informellen Praktiken und berufsbezogenen Überzeugungen (auch *teacher beliefs*, vgl. Reusser, Pauli & Elmer, 2011) zu rekonstruieren. Zentrales Ergebnis dieser Untersuchung sind die Kernkategorie *Gerechte Leistungsbeurteilung als Balanceakt* und die darauf aufbauende gegenstandsverankerte Theorie, deren sukzessive Entwicklung sich mittels zahlreicher Kodier- und Vergleichsprozeden vollzog (vgl. hierzu Abschnitte 3.2 und 3.3).

3.1 Fallauswahl und Vergleichseinheiten

Sowohl in der Vergleichenden Erziehungswissenschaft als auch der GTM kommt der Frage der Fallauswahl eine zentrale Rolle im Forschungsprozess zu, wobei jeweils spezifische Verständnisse davon, was als ‚Fall‘ zu verstehen ist, unterschieden werden können.

Wie bereits im vorhergehenden Abschnitt 2 zu den Grundzügen der GTM erläutert, erfolgt die Auswahl, welcher Fall oder besser, welches Datenmaterial in die Analyse einbezogen wird, in GTM-Studien nach dem Kriterium der theoretischen Relevanz für die zu entwickelnde gegenstandsverankerte Theorie. Die Frage der Fallauswahl in der Vergleichenden Erziehungswissenschaft kann auf verschiedenen Wegen erfolgen, immer in Abhängigkeit davon, welche Art vergleichender Untersuchung vorgenommen werden soll und welches Ziel damit verfolgt wird (vgl. Adick, 2014). Die grundlegende Logik stellt jedoch häufig die Suche nach Ähnlichkeiten und Unterschieden zwischen Fällen dar. Auch die Fallauswahl in qualitativ-vergleichenden Untersuchungen erfolgt auf der Grundlage einer „Gleichheitsunterstellung“ bei gleichzeitiger „Differenzbeobachtung“ (Heintz, 2010, S. 164): einerseits werden solche Fälle ausgewählt, die als ähnlich wahrgenommen werden, um sie in einem sinnvollen Vergleich gegenüberzustellen. Gleichzeitig wird davon ausgegangen, dass sich die Fälle hinsichtlich spezifischer Merkmalsausprägungen unterscheiden und eine Suche nach diesen Differenzen fruchtbare Erkenntnisse für die interessierende Fragestellung liefern wird. Diese Vorannahmen über Ähnlichkeiten und Differenzen können jedoch nur als vorsichtige Versuche verstanden werden, sich der Komplexität der sozialen Wirklichkeiten zu nähern, ihren Wert als geeignete Vergleichseinheiten hinsichtlich der konkreten Fragestellung können sie erst im Laufe der Untersuchung zeigen.

Die Strategie des theoretischen Samplings der Grounded-Theory-Methodologie arbeitet ebenfalls mit Annahmen über Ähnlichkeiten und Unterschiede der in die Analyse einzubeziehenden Daten hinsichtlich des interessierenden Phänomens. Auf der Suche nach minimaler und maximaler Kontrastierung werden ausgehend von der bisherigen Analyse die nächsten einzubeziehenden Daten ausgewählt. Minimale Kontraste dienen dabei der weiteren Dimensionalisierung bereits bestehender Kodes und Kategorien, maximale Kontraste der Abgrenzung von Kodes und Kategorien voneinander. Während des iterativen Forschungsprozesses werden nach und nach explizit Daten gesucht und in die Auswertung einbezogen, die entweder dazu dienen, bereits bestehende Kategorien und Ad-hoc-Hypothesen über deren Verhältnis zueinander zu festigen, oder aber solche, die diese Annahmen hinterfragen und widerlegen könnten. Hier liegt ebenfalls eine Auswahl auf dem Prinzip von Ähnlichkeiten und Unterschieden vor, die der Fallauswahl der vergleichenden Forschung zumindest ähnlich ist.

Nun könnte man einwenden, dass die Entscheidung, welche makrosozialen Vergleichseinheiten (sprich Länder, Weltregionen, Bildungssysteme etc.) in die Untersuchung einbezogen werden, bei der Vergleichenden Erziehungswissenschaft bereits vor der ersten Datenerhebung getroffen wird, und damit dem Grundsatz einer sukzessiven Fallauswahl im Sinne des *theoretical samplings* diametral gegenübersteht. Denn die Entscheidung für spezifische Fälle aufgrund von Vorwissen birgt immer auch die Gefahr, dass bisher unbekannte Zusammenhänge unentdeckt bleiben und nur bereits Bekanntes bestätigt wird. Eine Entscheidung über die in den Vergleich einzubeziehenden makrosozialen, geografischen oder kulturellen Einheiten vor Beginn der Untersuchung, wie in den meisten vergleichenden Untersuchungen, scheint damit unvereinbar mit der Idee der sukzessiven Fallauswahl in der Grounded-Theory-Methodologie. Gegenüber einer standardisierten und quantifizierenden international vergleichenden Forschung, wie z.B. den internationalen Schulleistungsstudien PISA und TIMSS, wäre dieser Einwand sicherlich berechtigt. Hier werden die Gemeinsamkeiten und Unterschiede zwischen den Vergleichseinheiten bereits durch die Operationalisierung der Fragestellung in den Erhebungsinstrumenten gesetzt und nicht erst im Prozess des Vergleichs herausgearbeitet. Qualitativ-vergleichende Forschungen, die rekonstruktiv-sinnverstehend angelegt sind und den Prinzipien der Offenheit und Gegenstandsangemessenheit folgen (vgl. Kruse, 2014), können hingegen sensibler für die kontextgebundenen Eigenheiten sein. Zwar werden auch hier auf einer abstrakten Makroebene die sozialräumlichen, geografischen oder kulturellen Einheiten des Vergleichs vorab bestimmt, die Auswahl der konkreten empirischen Fälle (welche Akteure, in welchem Bereich, welche Institutionen), und damit das empirische Fundament der Theoriebildung, kann dann aber wiederum ebenso sukzessiv und theoriegeleitet wie bei der Grounded-Theory-Methodologie erfolgen (vgl. hierzu auch Schittenhelm, 2012). Ein offener gestalteter Zugang zum Untersuchungsfeld, bei dem bewusst auch Raum gelassen wird für bisher Unbekanntes, kann darüber hinaus auch helfen, die oben beschriebenen – bewussten oder unbewussten – Annahmen über Unterschiedlichkeiten und Gemeinsamkeiten zwischen den Vergleichseinheiten zu relativieren. Je offener Forschende einem Phänomen in unterschiedlichen Kontexten zu begegnen versuchen, desto klarer können auch andere Erklärungsmuster neben das der ‚kulturellen‘ Verschiedenheit treten. Möglicherweise lassen sich dann vorgefundene Unterschiede des interessierenden Phänomens auf andere Merkmale, wie beispielsweise Geschlecht, Bildungsniveau oder soziale Rolle, zurückführen. Dafür ist es allerdings notwendig, die Fallauswahl nicht anhand der Zugehörigkeit zu einem spezifischen Kontext zu treffen, sondern analog zum Vorgehen des *theoretical samplings* nach Fällen zu suchen, die das interessierende Phänomen zu verstehen helfen.

Für die Untersuchung der Gerechtigkeitsüberzeugungen von Lehrkräften in Bezug auf die schulische Leistungsbeurteilung wurde zwischen der Entscheidung für die beiden makrosozialen Vergleichseinheiten Schweden und Deutschland (bzw. Nordrhein-Westfalen) und der konkreten Fallauswahl auf der Individualebene unterschieden. Beide Prozesse werden im Folgenden kurz erläutert.

Konkrete Wahl der Vergleichseinheiten und Fallauswahl in der Beispieluntersuchung

In dieser Arbeit wurde davon ausgegangen, dass ‚Gerechtigkeit‘ ein sozial konstruiertes Phänomen darstellt, dessen Bedeutung in verschiedenen historischen, soziokulturellen oder nationalen Kontexten unterschiedlich sein kann (vgl. Walzer, 1983). Ähnlich verhält es sich mit der Frage nach einer ‚gerechten Leistungsbeurteilung‘ und dem, was Lehrkräfte verschiedener Systeme als solche erachten. Insofern bietet die vergleichende Untersuchung der Gerechtigkeitsüberzeugungen von Lehrkräften aus zwei verschiedenen kulturellen Kontexten mit unterschiedlichen Bildungssystemen eine Möglichkeit, das komplexe Wechselspiel zwischen institutionell eingeschriebenen Gerechtigkeitsnormen, situativ-kontingenten Handlungs-routinen und individuellen Überzeugungen hinsichtlich der schulischen Leistungsbeurteilung vor dem Hintergrund ihrer zugeschriebenen gesellschaftlichen und pädagogischen Funktionen zu analysieren. Dafür wurden einerseits Deutschland, bzw. aufgrund der föderalen Struktur der BRD das Bundesland Nordrhein-Westfalen (NRW), und andererseits Schweden als Vergleichseinheiten ausgewählt. Beide Vergleichskontexte stellen sowohl hinsichtlich der Systemarchitektur der Bildungssysteme (z.B. gegliedertes vs. Gesamtschulsystem) als auch der institutionellen Rahmenbedingungen der schulischen Leistungsbeurteilung und darin eingelassener Gerechtigkeitsnormen einen interessanten Kontrast dar (vgl. Waldow, 2011, 2014). In beiden Ländern tragen hauptsächlich die einzelnen Lehrkräfte die Verantwortung für die schulische Leistungsbeurteilung, gleichzeitig unterscheiden sie sich hinsichtlich der Rolle schulischer Leistungsbeurteilung erheblich: Durch die hochselektive Grundstruktur des gegliederten Schulwesens, die zentrale Rolle von Noten und Zeugnissen für die Versetzung in die nächsthöhere Klassenstufe sowie den Übergang in weiterführende Schulformen, aber auch die zentralen Abschlussprüfungen kommt den Benotungen und daraus resultierenden Bildungsabschlüssen in NRW eine zentrale Rolle bei der Verteilung von Lebenschancen zu. Im schwedischen inklusiven Gesamtschulsystem, in dem Schülerinnen und Schüler erst ab der 6. Klasse überhaupt Noten erhalten⁴ und es weder Sitzenbleiben noch Abschlussprüfungen gibt, liegt der Fokus stärker auf Fragen der Egalität und Gleichwertigkeit von Beurteilungen, die durch ergänzende standardisierte zentrale Tests (*nationella prov*) sichergestellt werden sollen. Beide Kontexte können darüber hinaus ver-

schiedenen Wohlfahrtsstaatsregimen (Esping-Andersen, 1990) und damit in Verbindung stehenden *education regimes* (West & Nikolai, 2013) zugeordnet werden. Zudem finden sich verschiedene Traditionen der Steuerung und Kontrolle des Bildungssystems, die als eher auf die Kontrolle pädagogischer Prozesse (*process control*) und einer an der Überprüfung der Ergebnisse dieser Prozesse (*product control*) orientierten Steuerung charakterisiert werden (Hopmann, 2003). Insgesamt weisen die Bildungs- und Beurteilungssysteme Schwedens und Deutschlands bzw. NRWs einige Gemeinsamkeiten und Unterschiede auf, die eine interessante Ausgangslage für eine vergleichende Untersuchung darstellen. Gleichzeitig wurden diese auf der Grundlage theoretischer Vorannahmen identifizierten Charakteristika lediglich als vorläufig angesehen, denn ob und wenn ja, inwiefern sich die unterschiedlichen institutionellen Rahmenbedingungen auf die Gerechtigkeitsüberzeugungen einzelner Lehrkräfte auswirken, blieb zu jedem Zeitpunkt des Forschungsprojekts eine strikt empirisch zu beantwortende Frage.

Die konkrete Fallauswahl für die Interviews dieser Studie, d.h. welche Lehrkräfte im jeweiligen Kontext befragt wurden, erfolgte über drei Jahre (2012–2015) in mehreren Wellen und auf der Grundlage theoretisch abgeleiteter Samplingkriterien. Dies waren zum einen systembezogene Samplingkriterien, die sich aus der jeweiligen Systemarchitektur der Schul- und Beurteilungssysteme ableiten ließen (z.B. Schulform, Unterricht in einem Kernfach am Ende der Sekundarstufe I), zum anderen personenbezogene Kriterien, die auf der Ebene der individuellen Lehrkräfte angesiedelt sind (z.B. Geschlecht, Berufserfahrung). Für den nordrhein-westfälischen Kontext wurden beispielsweise Unterschiede zwischen Lehrkräften der verschiedenen Sekundarschulformen angenommen, sodass gezielt Gymnasien, Realschulen, Hauptschulen und Gesamtschulen kontaktiert und entsprechende Lehrkräfte der Sekundarstufe I für ein Interview angefragt wurden.⁵ Da im schwedischen Gesamtschulsystem der erste Übergang erst nach der neunjährigen Grundschule stattfindet, wurden zunächst nur Lehrkräfte dieser Schulform ins Sample aufgenommen, die in der Sekundarstufe I (Kl. 7–9) unterrichten. Allerdings zeigte sich in den Interviews mit diesen Lehrkräften eine auffällige Nicht-Thematisierung von Übergangs- und Selektionsproblematiken, sodass (entgegen der ursprünglichen Fokussierung auf Lehrkräfte am Ende der Sekundarstufe I) auch Lehrkräfte der weiterführenden Schulen (*gymnasieskola*, Kl. 10–13) ins Sample aufgenommen wurden. So wurde das ursprünglich aufgrund der Mehrgliedrigkeit des nordrhein-westfälischen Schulsystems aufgenommene Samplingkriterium der Schulform auch auf das Gesamtschulsystem Schwedens ausgeweitet – allerdings erst *nachdem* sich dieses als empirisch interessantes Kriterium herausgestellt hatte.

Auch die Berufserfahrung fand erst im Anschluss an die erste Erhebungsrunde in NRW Eingang in den Samplingprozess, da sich hier Unterschiede in den Erzäh-

lungen von sehr erfahrenen Lehrkräften im Vergleich zu Lehrkräften am Beginn ihrer Berufstätigkeit zeigten. Im nachfolgenden Zyklus wurden dementsprechend Kontrastfälle gesucht, um diesen Hinweisen systematisch nachgehen zu können. Eine weitere Ergänzung der Samplingkriterien in beiden Kontexten stellten schulinterne Funktionen der Lehrkräfte dar, wodurch sukzessive neben ‚einfachen‘ Fachlehrkräften auch Vertrauenslehrerinnen und -lehrer, eine Schulleiterin, pädagogische Leiterinnen und Leiter und Oberstufenkoordinatorinnen und -koordinatoren in das Sample einbezogen wurden. Als weniger bedeutsam stellte sich hingegen die Kategorie Geschlecht heraus, sodass dieses personenbezogene Samplingkriterium im Verlauf des Forschungsprozesses in den Hintergrund trat.

Durch die Anpassung der Samplingkriterien an den jeweiligen Analyse- und Erkenntnisstand entstand so ein kontrastives Sample unterschiedlichster Lehrkräfte, in dem gleichzeitig die schulsystemspezifischen Besonderheiten Berücksichtigung fanden. Durch die auf diese Weise gestufte Fallauswahl – indem zunächst die größeren makrosozialen Entitäten (Schweden und NRW) als Vergleichseinheiten bestimmt wurden, innerhalb derer dann die konkreten Fälle (einzelne Lehrkräfte) identifiziert wurden – konnte dem Anspruch eines theoretisch fundierten Samplings und gleichzeitig den kontextspezifischen Besonderheiten entsprochen werden. Auf den Zusammenhang von Kontext und Einzelfall wird im Abschnitt 3.3 weiter eingegangen.

Daneben wurde die Frage der Fallauswahl auch auf die sukzessive Einbeziehung weiterer Datenquellen – neben den episodischen Interviews mit Lehrkräften – im Forschungsverlauf ausgeweitet. Hierauf gehe ich im folgenden Abschnitt näher ein, in dem auch die konkrete Kodier- und Vergleichsarbeit diskutiert wird.

3.2 *Constant comparative method* und Vergleich als Methode

Vergleichsprozeden finden sowohl in der GTM als auch der vergleichenden Forschung statt, im Folgenden werden daher die grundlegenden Konzepte der *constant comparative method* (GTM) und des ‚Vergleichs als Methode‘ (Schriewer, 1987) aufeinander bezogen diskutiert.

Die grundlegende Operation, die den Kodier- und Analyseprozessen, aber auch der Fallauswahl, nach der GTM zugrunde liegt, ist die *constant comparative method* oder ‚Methode des ständigen Vergleichens‘. Gemeint sind damit die zahlreichen fortlaufenden Vergleichsprozesse, die Forschende zwischen verschiedenen empirischen ‚Vorkommnissen‘ (*incidents*) und daraus abgeleiteten Codes vornehmen, mit dem Ziel, die einer Theorie zugrunde liegenden Kategorien zu verfeinern und zu spezifizieren.⁶ Glaser (2002) weist darauf hin, dass Forschungsarbeit im Sinne der GTM immer *conceptual work* sei, die über eine bloße Beschreibung von Unterschieden und Gemeinsamkeiten hinausgehe (Glaser, 2002, S. 15). Ähnliches

lässt sich auch bei Schriewer (2006) finden, wenn er zwischen einfachen und komplexen Vergleichen unterscheidet. Komplexe Vergleiche zeichnen sich demnach dadurch aus, dass sie nicht die soziokulturellen Phänomene *als solche* miteinander vergleichen, sondern nach den *Beziehungen zwischen verschiedenen Phänomenen* suchen. Erst diese „Relationierung von Relationen“ (Schriewer, 1987, S. 633) ermöglicht den Vergleich auf einem höheren und damit über den Einzelfall hinausweisenden Abstraktionslevel. Auch Glasers Idee der Konzeptionalisierung liegt der Gedanke zugrunde, die Beziehungen zwischen den im empirischen Material identifizierten Konzepten und daraus abgeleiteten Kategorien und deren je spezifische Eigenschaften durch den Vergleich miteinander herauszuarbeiten. Je mehr Konzepte und Kategorien einschließlich ihrer Subkategorien und Ausprägungen durch die Konzeptionalisierung erfasst werden, umso mehr Variation des interessierenden Phänomens kann in die Analyse einbezogen werden.

Ein zweiter Aspekt des Verständnisses von Kodierarbeit als *conceptual work* ist die Hervorbringung von übergreifenden Begriffen, die mehr sind als nur die Zusammenfassung der empirisch beobachtbaren Vorkommnisse. Erst durch die begriffliche Fassung werden die konkreten Ereignisse in theoretische Konzepte überführt, die so über das singuläre Vorkommnis hinausweisen können. Auch hier findet sich eine Parallele zu Schriewers Verständnis des Vergleichs als sozialwissenschaftlicher Methode, wenn er von den Überbrückungsfunktionen des Vergleichs spricht (Schriewer, 2003, S. 21). Erst der Vergleich ermöglicht in seinen Augen die Überbrückung zwischen „raum-zeitlich (geographisch, regional, national, sozial-kulturell) spezifischen Beobachtungen und raum-zeit-enthobenen Hypothesen und Modellen, zwischen den ‚singulären Sätzen‘ der Empirie und den ‚allgemeinen Sätzen‘ der Theorie“ (ebd.). Vergleichende Untersuchungen ermöglichen es von der Ebene der empirischen Einzelheiten zu abstrahieren und auf einer abstrakteren Ebene zu erkennen, wofür diese lokalen Ausprägungen stehen. Um es mit Glasers Worten auszudrücken, Konzeptualisierung ermöglicht es uns zu erkennen „what is happening in the data“ (Glaser, 1978, S. 55).

Vergleichende Kodierarbeit am Beispiel

In der hier beispielhaft herangezogenen Untersuchung zu Gerechtigkeitsüberzeugungen von Lehrkräften in Schweden und Nordrhein-Westfalen erfolgte die Kodierarbeit des zweisprachigen Materials entsprechend der von Strauss (1994) vorgeschlagenen drei Formen des Kodierens – offenes, axiales und selektives Kodieren. Das offene Kodieren meint dabei den ersten offenen Zugang zum Material, bei dem durch das Stellen von generativen Fragen (vgl. Mey & Mruck, 2009, S. 117) und das Anstellen von Vergleichen die Daten ‚aufgebrochen‘ (Strauss & Corbin, 1996, S. 44) werden, um das interessierende Phänomen Benennen und Ka-

tegorisieren zu können. Beim axialen Kodieren werden bisher entwickelte Codes weiter zu Kategorien verdichtet, miteinander in Beziehung gesetzt und verglichen, um weitere Dimensionalisierungen zu ermöglichen. Dies erfolgt auf der Grundlage von sowohl fallinternen Vergleichen ähnlicher Passagen wie auch von fallübergreifenden Vergleichen. Mit dem selektiven Kodieren wird schließlich ein höherer Abstraktionsgrad angestrebt, indem die bisherigen Kategorien am gesamten Materialkorpus geprüft und zu übergreifenden Kategorien verdichtet werden, die dann in Beziehung zu einer oder mehreren Kernkategorie(n) gesetzt werden können. Alle drei Phasen des Kodierens können dabei ineinandergreifen, da die Kodierarbeit mit jedem neugewonnenen Datenmaterial aus jeder neuen Erhebungsrunde weiter vorangetrieben wird.

Dies soll an einem konkreten Beispiel verdeutlicht werden. So wurde in den ersten Interviews mit nordrhein-westfälischen Lehrkräften häufiger erwähnt, dass es für eine gerechte Beurteilung notwendig sei, eine gewisse Transparenz für die Schülerinnen und Schüler herzustellen. In-vivo-Codes (d.h. Codes, die wortwörtlich von den Befragten übernommen wurden) hierfür waren beispielsweise „*alles offenlegen*“ und „*transparent machen*“. Dem zunächst entstandenen Code „*Transparenz der Note*“ wurden später entsprechende Transkriptstellen aus verschiedenen (schwedischen und nordrhein-westfälischen) Interviews zugeordnet. Im fallübergreifenden Vergleich dieser Transkriptstellen wurde deutlich, dass teilweise sehr verschiedene Beurteilungsstrategien mit dem Begriff ‚Transparenz‘ von den Lehrkräften thematisiert wurden: beispielsweise die nachträgliche Offenlegung der Bepunktung einer schriftlichen Klassenarbeit, die Benennung von Beurteilungskriterien bereits zu Beginn einer neuen Unterrichtseinheit oder das Auslegen von Klassenlisten mit den Einzelnoten vor der Zeugnisnotenerstellung. Daraus folgte u.a. die Unterscheidung des Codes „*Transparenz der Note*“ in die Dimensionen „*Transparenz des Verfahrens*“ und „*Transparenz des Ergebnisses*“. Im Zuge der weiteren Analyse wurde zudem deutlich, dass es Unterschiede hinsichtlich der von den Lehrkräften zugeschriebenen Funktion einer transparenten Beurteilung gab – u.a. auch hinsichtlich der Frage *für wen* eine Note transparent sein müsse, um gerecht zu sein. „*Transparenz als Rechenschaftslegung*“ wurde somit für Transkriptstellen vergeben, in denen vor allem die Belegbarkeit einer Beurteilung im Sinne einer Rechenschaftslegung für Dritte (zumeist Eltern oder die Schulleitung) deutlich wurde. Andere Lehrkräfte betonten, dass die Offenlegung der Beurteilungskriterien vor allem der Orientierung von Schülerinnen und Schülern und der Unterstützung ihres Lernprozesses diene, wodurch überhaupt erst eine in ihren Augen gerechte Beurteilung möglich wäre – hier wurde der übergeordnete Code „*Transparenz als Lernunterstützung*“ vergeben. Alle diese Codes flossen später in die übergeordnete Kategorie „*Beurteilungsverständnis*“ ein, die u.a. auch die präferier-

te Bezugsnormorientierung, das Verhältnis zu summativer bzw. formativer Leistungsbeurteilung sowie das Leistungsverständnis enthält. Im Zusammenspiel mit weiteren zentralen Kategorien schälten sich so allmählich vier Muster von Gerechtigkeitsüberzeugungen heraus (vgl. hierzu auch Abschnitt 3.3).

Sukzessives Einbeziehen verschiedener Datenquellen

Für beide Kontexte – Schweden und NRW – wurden bereits in der Vorbereitung der Interviewleitfäden offizielle Regularien (wie Schulgesetze, Verordnungen und Vorschriften zur schulischen Leistungsbeurteilung) analysiert, die im Zuge der Auswertung der Interviewtranskripte sukzessiv um weitere zusätzliche Datensorten ergänzt wurden. Dem Diktum „all is data“ (Glaser, 2004, Abs. 45) folgend wurden so beispielsweise die zentralen Lehrpläne für die schwedische Grund- und weiterführende Schule sowie die Kernlehrpläne für alle Schulformen in NRW in die Analyse einbezogen, ebenso wie die zahlreichen von der schwedischen Schulbehörde *Skolverket* bereitgestellten Handreichungen und Informationsbroschüren für Lehrkräfte, die konkrete Hinweise für die Umsetzung der schulischen Leistungsbeurteilung enthalten. Demgegenüber fanden sich für nordrhein-westfälische Lehrkräfte neben den schulrechtlichen Gesetzen, Erlassen und Verordnungen keinerlei Handreichungen oder Praxishinweise. Dieses Fehlen von Materialien im nordrhein-westfälischen Kontext kann dabei bereits als ein Befund betrachtet werden, der auf eine sehr unterschiedliche institutionelle Rahmung der schulischen Leistungsbeurteilung verweist und nur durch den Vergleich sichtbar wurde.

Die zusätzlichen Materialien lieferten einerseits wichtige Kontextinformationen zu den jeweiligen Schul- und Beurteilungssystemen, gleichzeitig zeigten sich in ihnen bereits institutionell eingelassene Gerechtigkeitsnormen, die für beide Kontexte spezifisch waren. So wird beispielsweise in den schwedischen Materialien der Schulbehörde immer wieder darauf verwiesen, dass nur eine gleichwertige Beurteilung (*likvärdig bedömning*) eine gerechte Beurteilung sein könne. Die ministeriellen Regularien für NRW beschränken sich hingegen eher auf formale Hinweise wie die Zusammensetzung von Zeugnisnoten aus mündlichen und schriftlichen Beurteilungen sowie die Anzahl der zu schreibenden Klassenarbeiten – Details, die im schwedischen Kontext offengelassen werden. Neben diesen offiziellen Materialien wurden zunehmend auch von den Lehrkräften verwendete und zum Teil selbst erstellte Beurteilungsmaterialien – wie Notenlisten, Sitzpläne, Erwartungshorizonte, Punkteraster etc. – in die Analyse einbezogen, da sich in ihnen die individuelle Beurteilungspraxis einzelner Lehrkräfte manifestiert. Durch den GTM-typischen iterativen Forschungsprozess und den permanenten Wechsel zwischen Datengenerierung und -analyse erfolgte so auch eine sukzessive Erweiterung des einbezogenen Datenmaterials, die sich an der theoretischen Relevanz des Materials orientierte

und so kontextsensibel Schritt für Schritt angepasst werden konnte. Der Abgleich der Interviewaussagen mit diesen weiteren Materialien ermöglichte dabei auch eine Rückbindung des subjektiven Sinns individueller Überzeugungen auf überindividuelle Strukturen, also auch eine Verbindung der Mikro- mit der Makroebene.⁷

Zusammenfassend lässt sich als eines der universellen Merkmale der schulischen Leistungsbeurteilung in beiden Kontexten konstatieren, dass die schulische Leistungsbeurteilung als berufliche Kernaufgabe eine hochgradig individuelle und situativ-kontingente Praxis mit erheblichen Deutungs- und Entscheidungsspielräumen für die einzelne Lehrkraft darstellt. Wie diese Spielräume von einzelnen Lehrkräften genutzt werden, wird dabei nur zu einem Teil von den institutionellen Rahmenbedingungen bestimmt. Der regulative Rahmen ist daher nicht als präskriptiv anzusehen, vielmehr variiert die konkrete Ausgestaltung der Leistungsbeurteilung situativ und in Abhängigkeit von den individuellen Gerechtigkeitsüberzeugungen.

Herausforderungen der vergleichenden Analyse

Wie in jedem Forschungsprozess, der sich auf sprachliche Äußerungen und deren Interpretation stützt, waren mit der Analyse der Interviews und weiteren Materialien zahlreiche Verstehensprozesse verknüpft, die sich auf kulturelle, soziale und historisch gewachsene Bedeutungszuschreibungen beziehen und in ihrer Kontextgebundenheit analysiert werden müssen. Für das Verständnis des schwedischen Kontexts war – neben den spezifischen Charakteristika des Bildungssystems im Allgemeinen und den Rahmungen der schulischen Leistungsbeurteilung im Besonderen – auch das Erlernen der Landessprache sowie der Einbezug muttersprachlicher Personen in alle Phasen des Forschungsprozesses eine Grundvoraussetzung. Gleichzeitig sollte nicht darüber hinweggesehen werden, dass auch das vermeintlich ‚bekanntere‘ nordrhein-westfälische Schulsystem sowie die Besonderheiten der schulischen Leistungsbeurteilung eine intensive Auseinandersetzung und Erarbeitung notwendigen Kontextwissens erforderten. Denn trotz der gemeinsamen Sprache kam es auch in den nordrhein-westfälischen Interviews teilweise zu kurzen Irritationen, wenn beispielsweise die Lehrkräfte eine spezifische Fachsprache und mir bis dahin unbekannte Begriffe und Abkürzungen verwendeten. In beiden Kontexten bestand zudem gleichermaßen die Herausforderung, sich den Handlungslogiken und Deutungsmustern einer Berufsgruppe anzunähern, die nicht meinem eigenen beruflichen Erfahrungshorizont als Erziehungswissenschaftlerin entsprechen. McNess et al. (2015) diskutieren dies als „working in that ‚third space‘ in between“ (S. 311) und betonen, dass Forschende „neither complete observers nor complete participants“ seien. Milligan (2016) greift die Idee der Auflösung des Insider-Outsider-Dualismus auf und führt als Erweiterung die Figur des *inbetweeners* ein

und zeigt, wie die Selbst- und Fremdzuschreibung von Positionierungen im Forschungsprozess abwechseln und ineinander übergehen können. Letztlich, so Hella-well (2006), gehe es jedoch weniger um Fragen von „subtly varying shades of ‚insiderism‘ and ‚outsiderism‘. The more important point has to do with empathy, trying to understand the other person, or the other context, rather than closeness or distance“ (S. 489).

Bezogen auf die hier dargestellte GTM-Studie kann die materialnahe, teilweise sehr kleinschrittige Auswertung mithilfe der skizzierten Kodierprozeduren sowie permanente fallinterne und fallübergreifende Vergleiche als Hilfsmittel verstanden werden, um in der Analyse immer wieder zu einer offenen und gegenstandszentrierten Haltung gegenüber dem Material und einem Fokus auf das Phänomen ‚gerechter Leistungsbeurteilung‘ zurückzukehren – anstelle einer Fixierung auf die Gemeinsamkeiten und Unterschiede zwischen Lehrkräften verschiedener Kontexte. Der permanente Wechsel zwischen dem schwedischen und dem nordrhein-westfälischen Kontext sowie zwischen der Analyse von schwedischen und deutschen Interviewtranskripten stellte sich einerseits als herausfordernd dar, andererseits wurde ein vorschnelles Erklären von vorgefundenen Unterschieden durch ‚kulturelle‘ Faktoren so erschwert und die Analyseperspektive von der Deskription hin zur ‚Relationierung von Relationen‘ (Schriewer, 1987, S. 633) erweitert. Durch die fortlaufenden Vergleichsprozesse über den gesamten Materialkorpus hinweg konnte so auch ein ‚gemeinsamer Denkraum‘ (Matthes, 1992, S. 96) geschaffen werden, indem der übergreifende Fokus auf dem Verständnis des Phänomens der Gerechtigkeitsüberzeugungen lag und erst im zweiten Schritt u.a. die rahmenden Kontextbedingungen der beiden Bildungssysteme einbezogen wurden. Durch diese Gegenstandsfokussierung konnte auch einer zu starken Exotisierung des schwedischen Kontexts entgegengewirkt werden, und eine nostrifizierende Überhöhung des vermeintlich bekannten nordrhein-westfälischen Kontexts wurde erschwert. Zentral hierfür war einerseits eine kontinuierliche Dokumentation und Reflexion des Forschungsprozesses sowie der verschiedenen Samplingentscheidungen und Kodierprozeduren in Form von Memos und einem Forschungstagebuch.

Andererseits spielte der systematische Einbezug muttersprachlicher Personen als ‚kulturvertraute Co-Interpreten‘ (Schröer, 2013) in die Datengenerierung, Transkription und Auswertung der schwedischen Interviews eine zentrale Rolle. Die materialnahen Auswertungs- und Kodierprozeduren erfordern eine hohe fremdsprachliche Kompetenz, die auch das Erkennen von Zwischentönen, Ironie oder sprachlichen Bildern ermöglicht sowie Äußerungen der Befragten im jeweils spezifischen Kontext einordnet. Beides kann durch den Austausch mit muttersprachlichen Personen unterstützt werden, ersetzt jedoch nicht die eigene Beherrschung und Auseinandersetzung mit der jeweiligen Landessprache. Im vorliegenden Fall

der Befragung von schwedischen Lehrkräften führte dies zu einer zeitlichen Verzögerung des Feldeintritts, da ich zunächst die schwedische Sprache auf entsprechend gutem Niveau erlernen musste. Gleichzeitig erwies sich der Umstand, als Forschende aus Deutschland das Schwedische mächtig zu sein und damit die Interviews in der Landessprache durchführen zu können, als außerordentlich förderlich für die Teilnahmemotivation und Auskunftsfreude der befragten Lehrkräfte. Der damit verbundene zeitliche Mehraufwand ist allerdings für die Planung und Durchführung einer solchen Untersuchung ebenso zu berücksichtigen, wie die personellen und finanziellen Ressourcen abzuwägen sind.⁸

3.3 Tertium comparationis und Kernkategorie

Zuletzt widme ich mich den methodologischen Parallelen zwischen dem für vergleichende Untersuchungen zentralen Konzept des tertium comparationis und der Kernkategorie in der GTM.

Ein Vergleich zweier Gegenstände, der über die bloße Konstatierung von Ähnlichkeiten und Unterschieden hinausgehen will, bedarf eines den beiden Vergleichsgegenständen gemeinsamen Dritten – des *tertium comparationis*. Dieses ‚Dritte des Vergleichs‘ kann verstanden werden als ein gemeinsamer Maßstab, an dem die beiden Vergleichsobjekte gemessen werden, um zu verhindern, dass lediglich die Merkmale eines der Objekte auf das andere übertragen werden. Insbesondere in international vergleichenden Untersuchungen ist diese Suche nach einem abstrakten Vergleichskriterium ein entscheidendes Element, um über die bloße Bewertung der Unterschiede und Gemeinsamkeiten von Phänomenen in verschiedenen Kontexten auf der Folie der eigenen soziokulturellen Prägungen hinauszugehen. Entscheidend für die Vergleichbarkeit von Phänomenen ist dabei nicht, ob auf den ersten Blick eine ‚Gleichheit‘ oder ‚Ähnlichkeit‘ besteht, sondern ob es gelingt, eine passende übergreifende Kategorie zu finden, anhand derer sich die Merkmale eines Phänomens vergleichen lassen. Denn „Vergleichbarkeit ist kein A-priori-Merkmal bestimmter Sachverhalte, sondern muss immer in Relation zur Fragestellung und zum jeweiligen Kontext festgestellt werden“ (Rippl & Seipel, 2008, S. 66).

Eine interessante Frage hinsichtlich des Charakters des tertium comparationis ist nun, ob es sich hierbei um eine *notwendige Voraussetzung* für vergleichende Untersuchungen handelt, oder ob das tertium comparationis erst das *Ergebnis* einer solchen Forschung sein kann. Wenn man es als dem Vergleich zugrunde liegende Kategorie versteht, wäre es der Ausgangspunkt für vergleichende Untersuchungen. Andererseits kann es aber auch als Zielpunkt des Vergleichsprozesses angesehen werden, als theoretisches Konstrukt, das gerichtete Vergleichsprozesse erst ermöglicht und gleichzeitig ihr Ergebnis darstellt.

In der zweiten Auffassung finden sich Parallelen zum Charakter der Kernkategorie in der Grounded-Theory-Methodologie. Sie stellt die zentrale Kategorie einer Grounded Theory dar, auf der die gesamte gegenstandsbezogene Theorie aufbaut und die durch zahlreiche Vergleichsprozesse im Verlauf des Forschungsprozesses herausgearbeitet wurde. In qualitativ-vergleichenden Forschungen kann das tertium comparationis als gemeinsames Drittes nicht immer von Beginn der Untersuchung an bestimmt werden, es muss – ebenso wie die Kernkategorie der Grounded Theory – erst aus dem Material rekonstruiert und in wiederkehrenden Vergleichsoperationen bestätigt werden (vgl. Kelle, 2015, S. 94). Waterkamp (2006, S. 196) weist darauf hin, dass der gelegentliche Verweis darauf, das tertium comparationis müsse bereits vor Beginn einer vergleichenden Untersuchung festgelegt sein, die Gefahr eines relativ schematischen Forschungsvorgehens birgt. Der wirkliche Verstehensprozess vollziehe sich erst im Laufe des Vergleichs, und die Bestimmung des tertium comparationis als das Gemeinsame zweier unterschiedlicher Vergleichseinheiten könne daher auch „als der eigentliche Zielpunkt des Vergleichs gesehen werden“ (ebd.).

Hier lohnt es sich – wie bereits in Abschnitt 3.1 eingeführt – zwischen den verschiedenen Phasen einer vergleichenden Untersuchung zu unterscheiden: *Zu Beginn eines vergleichenden Forschungsprojekts* steht die grundsätzliche Entscheidung darüber, den Untersuchungsgegenstand zu bestimmen und festzulegen mit Hilfe welcher Vergleichseinheiten die Untersuchung der spezifischen Fragestellung erfolgen soll. Weiß und Nohl (2012) sprechen hier auch von der „implizite[n] Entscheidung für einen Kontext“ (S. 57). Diese Auswahl erfolgt auf der vorläufigen Einschätzung, dass die gewählten Untersuchungseinheiten über Gemeinsamkeiten hinsichtlich der interessierenden Fragestellung verfügen – diese Einschätzung erfolgt zunächst auf der Grundlage persönlichen Vorwissens und bereits bestehender Forschungsergebnisse. Ausgehend von dieser Vergleichsbasis wird dann die möglichst umfassende Erfassung des interessierenden Phänomens angestrebt, um schlussendlich Aussagen darüber treffen zu können, inwiefern sich die Erscheinungsformen des Phänomens in den gewählten Vergleichseinheiten tatsächlich unterscheiden bzw. was die verbindende Gemeinsamkeit ist. Gleichzeitig begründen auch Differenzannahmen die Auswahl der Untersuchungseinheiten, denn „Gemeinsamkeit und Differenz lassen sich nicht unabhängig voneinander bestimmen, sondern müssen gleichermaßen im und durch den Vergleich zweier oder mehrerer empirischer Erscheinungen gewonnen werden“ (Rademacher, 2013, S. 68). Die Auswahl der Vergleichseinheiten ist damit immer als vorläufige Entscheidung zu betrachten, die durch spätere Erkenntnis im Vergleichsprozess auch revidiert werden kann.

Für die *konkrete Fallauswahl und Materialauswertung* in den zu vergleichenden Kontexten sollten die oben genannten Vorannahmen jedoch bewusst ausgeblendet werden bzw. lediglich als sensibilisierende Konzepte in die weitere Forschungsarbeit einbezogen werden (vgl. hierzu auch Liebeskind, 2012, S. 335; Huf, 2015, S. 19), um dem Material mit ausreichender Offenheit gegenüberzutreten zu können. Beide Prozesse, die Wahl der Vergleichseinheiten wie auch die konkrete Fallauswahl, sind von der Fragestellung und dem Untersuchungsgegenstand her bestimmt. Als gedachtes Drittes neben den zwei empirischen Vergleichsgegenständen hat das *tertium comparationis* zwar keinen eigenen empirischen Gehalt, jedoch liegt sein Mehrwert für die vergleichende Untersuchung genau darin: Es fungiert als eine theoretische Konstruktion auf einer höheren Abstraktionsebene als die beiden ursprünglichen Vergleichsgegenstände. Die Besonderheiten und Eigenheiten des Einzelnen verweisen damit auf den Charakter eines allgemeinen, darüber liegenden Prinzips.

Wenn man das *tertium comparationis*, wie eingangs erläutert, nicht nur als die Fallauswahl und Vergleichsdimensionen bestimmende Kategorie versteht, sondern darin auch ein prozessuales Element erkennt, scheinen die Kernkategorie und das *tertium comparationis* auf nicht mehr ganz unterschiedlichen theoretischen Ebenen zu liegen. Die Rolle der Kernkategorie für die Generierung einer neuen Theorie wird von Strauss (1994, S. 66) folgendermaßen beschrieben: „Sie ist relevant und funktioniert. Die meisten anderen Kategorien mit ihren Eigenschaften haben einen Bezug zu ihr, sodass sie in starkem Maße der Qualifikation und der Modifikation unterliegt. Darüber hinaus hat sie ... primär die Funktion, die Theorie zu integrieren, zu verdichten und zu sättigen, sobald die Bezüge herausgearbeitet sind“.

Ähnliches gilt auch für das *tertium comparationis*: Während des Vergleichsprozesses dient es gleichzeitig als Analysebrille und als vorläufiges Ergebnis des Vergleichs. Seine endgültige Form kann es erst durch die fortlaufenden Vergleichsprozesse und die damit verbundenen Änderungen und Verdichtungen annehmen, bis es letztlich als Ergebnis des Vergleichs der Unterschiede und Gemeinsamkeiten formuliert werden kann. Als *tertium comparationis* der hier diskutierten Studie fungierten die Gerechtigkeitsüberzeugungen von Lehrkräften, wobei die vorläufige Annahme unterschiedlicher Überzeugungen zu Beginn der Untersuchung erst im Forschungsverlauf sukzessive erhärtet und die konkreten Ausprägungen im Analyseprozess unterschieden werden konnten. Damit zeichnen sich sowohl das *tertium comparationis* als auch die Kernkategorie durch ihre Vorläufigkeit und Prozesshaftigkeit aus.

Kernkategorie und tertium comparationis am Beispiel

Die für eine von der GTM inspirierte Forschung typische Suche nach der Kernkategorie ermöglicht dabei eine Fokusverschiebung auf das Gemeinsame zwischen verschiedenen Einzelfällen, hier einzelnen Lehrkräften, anstelle einer einseitigen Fokussierung auf vermeintliche Unterschiede. Die Entwicklung der Kernkategorie der hier diskutierten Untersuchung erfolgte über mehrere Zwischenschritte, die an dieser Stelle nur kursorisch dargestellt werden können. So wurde bereits in den ersten drei Interviews deutlich, dass die Erzählungen der Lehrkräfte zwischen zwei Polen eingeordnet werden konnten: einerseits der Selbst- und Fremdpositionierung „alle wollen gerecht sein“, in der sich ein grundsätzliches Streben nach einer gerechten Leistungsbeurteilung ausdrückte, und andererseits der Feststellung der Vieldeutigkeit dessen, was als ‚gerecht‘ verstanden wurde. Zudem wurde deutlich, dass zwischen verschiedenen Beurteilungsformen und -zeitpunkten unterschieden werden muss, die später als Kategorie ‚Beurteilungssituationen‘ im Rahmen kontextueller Bedingungen zusammengefasst wurden.

Im nächsten Schritt kristallisierten sich verschiedene professionelle Spannungsfelder heraus, innerhalb derer die Lehrkräfte ihre eigene Beurteilungspraxis vertorteten. Eines dieser Spannungsfelder wurde beispielsweise zwischen der Orientierung am einzelnen Kind bei der Beurteilung (Einzelfallorientierung) und der Anwendung eines für alle Schülerinnen und Schüler gleichermaßen gültigen Beurteilungsmaßstabs ausgemacht, ein anderes verhandelte die Frage, inwiefern eine gerechte Beurteilung ‚objektiv‘, im Sinne einer unpersönlichen Beurteilung, sein müsse. Der nächste Zwischenschritt bestand darin, die diesen Spannungsfeldern zugrunde liegenden Einflussfaktoren zu kategorisieren, wobei die Unterscheidung zwischen offiziellen Regularien, internen und externen Erwartungen und Normen sowie der individuellen Interpretation dieser erfolgte. ‚Gerechte Leistungsbeurteilung‘ wurde damit zunächst als Ergebnis eines Aushandlungsprozesses zwischen Regularien, Erwartungen und Interpretationen verstanden.

In die finale Kernkategorie ‚*Gerechte Leistungsbeurteilung als Balanceakt*‘ flossen all diese Analyseschritte ein. Die gegenstandsverankerte Theorie lässt sich folgendermaßen zusammenfassen: Ausgehend von der Feststellung, dass die schulische Leistungsbeurteilung eine der Kernaufgaben von Lehrkräften ist und diese Aufgabe durch zum Teil widersprüchliche Anforderungen geprägt ist, ergibt sich ein Handlungsdruck für Lehrkräfte, zu dem sie sich verhalten müssen (*ursächliche Bedingung*). Die institutionellen Rahmenbedingungen der schulischen Leistungsbeurteilung wirken als äußere Erwartungen und bilden zusammen mit den konkreten Beurteilungssituationen den Handlungskontext (*kontextuelle Bedingungen*). Neben den Kontextbedingungen wirken das professionelle Selbstbild der Lehrkräfte, ihr spezifisches Beurteilungsverständnis sowie die Beziehungskonstellationen,

in die sie eingebettet sind und die sie aufbauen, als *intervenierende Bedingungen* auf die Gerechtigkeitsüberzeugungen und Beurteilungsstrategien ein. Aus diesen Bedingungsgefügen ergibt sich die grundlegende Charakterisierung einer gerechten Leistungsbeurteilung als Balanceakt (*Kernkategorie*). Das Bild des Balanceakts bezieht sich dabei auf das aktive Ausgleichen, Aushandeln und Handhabbarmachen der verschiedenen Erwartungen an die Leistungsbeurteilung durch die Lehrkräfte, sprich den Prozess der Herstellung gerechter Leistungsbeurteilung unter Anwendung verschiedener Beurteilungsstrategien. Gleichzeitig wird das Ergebnis dieses Herstellungsprozesses – die als gerecht empfundene Beurteilung – wiederum unter Rückgriff auf unterschiedliche Legitimationsquellen als gerecht oder ungerecht beurteilt, als legitime oder illegitime pädagogische Praxis kategorisiert. Das Ergebnis ‚gerechte Beurteilung‘ ist jedoch flüchtig und muss immer wieder aufs Neue hergestellt, ausgehandelt und erarbeitet werden, weshalb das Bild des Balanceakts sowohl als Prozess- als auch als Zustandsmetapher verstanden werden muss. Dieser Balanceakt wird unter Rückgriff auf verschiedene Gerechtigkeitsüberzeugungen (mathematisch-rechnerisch, prozedural-bürokratisch, diskursiv-interaktiv, kompensatorisch) und durch verschiedene Beurteilungsstrategien von den Lehrkräften bewältigt, die sich auf unterschiedliche Legitimationsquellen stützen und in Abhängigkeit zu den kontextuellen und intervenierenden Bedingungen stehen. Die gewählten Beurteilungsstrategien führen damit eine in den Augen der einzelnen Lehrkraft gerechte(re) Leistungsbeurteilung herbei, ermöglichen die (Wieder-)Herstellung ihrer Handlungsfähigkeit und legitimieren so ihre Beurteilungspraxis (*Konsequenzen*).

Die Kernkategorie *Gerechte Leistungsbeurteilung als Balanceakt* und die weiteren zentralen Kategorien der Theorie wurden dabei nicht durch voneinander getrennte länderspezifische Analysen gewonnen, sondern vielmehr im permanenten Vergleich über den gesamten Materialkorpus hinweg entwickelt. In den vier aus der Analyse gewonnenen Gerechtigkeitsüberzeugungen – mathematisch-rechnerisch, prozedural-bürokratisch, diskursiv-interaktiv und kompensatorisch – zeigen sich die in der Kernkategorie benannten Balanceakte, die Lehrkräfte vollführen, um eine in ihren Augen gerechte Beurteilung zu ermöglichen. Sie sind als Idealtypen im Weber'schen Sinne zu verstehen, als analytische Hilfsmittel, um der empirisch vorfindlichen Vielfalt strukturierend gegenüberzutreten. Gleichzeitig demonstrieren sie den analytischen Mehrwert einer an der GTM orientierten vergleichenden Forschung, da in ihnen deutlich wird, dass die Unterschiede zwischen verschiedenen Lehrkräften hinsichtlich ihrer Gerechtigkeitsüberzeugungen nicht zwangsläufig entlang von Ländergrenzen verlaufen bzw. auf das jeweilige Bildungssystem zurückzuführen sind. Vielmehr finden sich für das gesamte Sample übergreifende Strukturierungskategorien, wie das professionelle Selbstbild, die Be-

ziehungskonstellationen und das Beurteilungsverständnis, die im Wechselspiel mit verschiedenen institutionellen Rahmungen eine spezifische Gerechtigkeitsüberzeugung begünstigen. Teilweise überschneiden sich diese Gerechtigkeitsüberzeugungen, andere hingegen haben keine Berührungspunkte miteinander. Im länderübergreifenden Vergleich zeigt sich, dass alle vier Muster in beiden Ländern auftreten, jedoch in unterschiedlichen Ausprägungen. Während die befragten schwedischen Lehrkräfte beispielsweise nur sehr marginal auf die mathematisch-rechnerische Gerechtigkeitsüberzeugung rekurrieren, spielt diese bei den nordrhein-westfälischen Lehrkräften eine zentrale Rolle. Die kompensatorische Gerechtigkeitsüberzeugung wird demgegenüber in NRW nur unter bestimmten Bedingungen vertreten, im schwedischen Kontext scheint sie dagegen selbstverständlicher zu sein.

Das *tertium comparationis* dieser Untersuchung – Gerechtigkeitsüberzeugungen von Lehrkräften – wurde damit einerseits zu Beginn des Forschungsprozesses als vorläufige Kategorie bestimmt, gleichzeitig schälten sich erst im Verlauf des iterativen Forschungsprozesses und im Zuge der vielzähligen Kodier- und Vergleichsprozeden die genaue Gestalt dieser Gerechtigkeitsüberzeugungen heraus. Dass die rekonstruierten Unterschiede zwischen Gerechtigkeitsüberzeugungen verschiedener Lehrkräfte nicht entlang vorab festgelegter Samplegruppen verlaufen – z.B. deutsche vs. schwedische Lehrkräfte, Lehrkräfte bestimmter Schulformen – sondern vielmehr quer zu diesen vermeintlichen Gruppierungen liegen, zeigt, dass „differences are not constructed on the basis of researcher’s assumptions but on the basis of attributions and beliefs *in the field*“ (Kelle, 2015, S. 94). Inwiefern die rekonstruierten Gerechtigkeitsüberzeugungen auch bei Lehrkräften aus anderen Bildungssystemen als den hier untersuchten zu finden wären, inwiefern die Ergebnisse also auf Lehrkräfte ‚im Allgemeinen‘ generalisierbar und auf andere Kontexte übertragbar wären, ist eine ausnahmslos empirisch zu beantwortende offene Frage, die nur durch weitere Forschung beantwortet werden kann.

4. Zusammenfassende Diskussion und Ausblick

Die Orientierung an Prämissen der Grounded-Theory-Methodologie (GTM) erweist sich im Rahmen qualitativ-rekonstruktiver internationaler Vergleichsstudien auf verschiedenen Ebenen als besonders fruchtbar.

Dies betrifft zunächst die Frage der Vergleichseinheiten und der Fallauswahl, die sowohl bei der vergleichenden Forschung als auch bei der GTM eine zentrale Rolle spielt. Der Forschungslogik der GTM entsprechend wird eine sukzessive Erweiterung der in die Analyse einzubeziehenden Fälle und Daten durch das theoretische Sampling angestrebt. Die Fallauswahl kann dabei als Suchbewegung vom Besonderen zum Allgemeinen verstanden werden. Auf der Suche nach minimalen und

maximalen Kontrasten wird so versucht, das interessierende Phänomen möglichst umfassend zu beleuchten. Da zu Beginn einer Untersuchung meist noch relativ wenig über das Phänomen bekannt ist, ist es jedoch kaum möglich bereits vorab festzulegen, welche und vor allem wie viele Fälle in die Analyse einbezogen werden. Wie geht dies nun mit der Logik einer international vergleichend angelegten Studie einher, bei der die Entscheidung für bestimmte sozialräumliche, geografische oder kulturelle Einheiten zumeist bereits *vor* Beginn der Forschung festgelegt wird? Ich sehe darin keinen grundsätzlichen Widerspruch, solange die Auswahl dieser Einheiten theoretisch begründet erfolgt und die Forschenden sich darüber bewusst sind, dass bereits die Konstitution des Vergleichsgegenstands durch die Forschenden selbst einen Teil der Analyse darstellt. Wie Kelle (2015, S. 94) betont, entspricht diese vorläufige und daher „hypothetical construction of the tertium comparationis“ auch der zumeist induktiven (bzw. abduktiven) Forschungslogik qualitativer Forschung generell, die eine sukzessive Konkretisierung des tertium comparationis im Verlauf der Forschung erfordert, aber auch ermöglicht. Die Auswahl der konkreten empirischen Fälle (also z.B. welche Akteure oder Organisationen, in welchem Bereich des Bildungswesens einbezogen werden), kann dann innerhalb der jeweiligen Vergleichseinheiten entsprechend der Prämissen der GTM sukzessive und theoretisch begründet erfolgen. Ein solcherart offen gestalteter Zugang zum Untersuchungsfeld lässt dabei wiederum Raum für bisher Unbekanntes und ermöglicht die Suche nach ‚funktionalen Äquivalenten‘ des Vergleichs (vgl. Schriewer, 2003).

Die Methode des permanenten Vergleichs der GTM und die darin eingelassene konsequente Gegenstandsorientierung kann dabei ein Weg sein, eine zu starke Fixierung auf kulturelle oder nationale Erklärungsmuster für Unterschiede und Gemeinsamkeiten zu überwinden – ebenso wie die Ignoranz gegenüber nationalen Besonderheiten, die weiterhin bestehen (vgl. Dale, 2005). Indem in der hier dargestellten Studie zu Gerechtigkeitsüberzeugungen von Lehrkräften die VergleichsprozEDUREN permanent auch über die Fälle und Vergleichseinheiten hinweg vollzogen wurden, blieben beispielsweise länderspezifische Unterschiede eine radikal empirisch zu beantwortende Frage, die sich am Material bewähren mussten. Gleichzeitig konnte durch die Analyse unterschiedlichen Datenmaterials (Regulativen, Handreichungen für Lehrkräfte, Beurteilungsmaterialien) eine Rückbindung der individuell rekonstruierten Überzeugungen von Lehrkräften an kontextbezogene überindividuelle Strukturen des jeweiligen Bildungssystems erfolgen. Die hierbei vorgenommenen Differenzierungen erfolgten auch im Bewusstsein des „Zurechnungsproblems“ (Nohl, 2009, S. 96) vergleichender Untersuchungen. Die entwickelte Typologie von Gerechtigkeitsüberzeugungen basiert dabei auf der Analyse zentraler Kategorien, wie dem professionellen Selbstverständnis der Lehr-

kräfte, die als kontextübergreifende strukturierende Elemente für alle im Sample vertretenen Lehrkräfte identifiziert werden konnten und nicht entlang von Ländergrenzen verlaufen. Durch das systematische Offenlegen und Reflektieren der Vergleichsprozeden in Form von Memos und die Zusammenarbeit in einem international zusammengesetzten Forschungsteam konnten zudem eigene Vorannahmen relativiert und hinterfragt werden. Beispielsweise zeigte sich die aus der Literatur ersichtliche Dominanz standardisierter Leistungstests zur Sicherung einer landesweit gleichwertigen und damit gerechten Beurteilung im schwedischen System als weniger handlungsleitend für die befragten Lehrkräfte, die wiederum den kollegialen Austausch oder kleinteilige Dokumentationen als wichtiger für eine in ihren Augen gerechte Leistungsbeurteilung betonten.

Zu guter Letzt wird – im Anschluss an die oben dargestellten Überlegungen von Kelle (2015) – vorgeschlagen, das *tertium comparationis* vergleichender Untersuchungen (analog zur Kernkategorie der GTM) stärker als vorläufiges und prozessuales Element zu betrachten, das sich erst im Verlauf der Untersuchung durch die zahlreichen Vergleichs- und damit auch Abgrenzungsprozesse in seiner finalen Gestalt herauschält (ähnlich hierzu auch Huf, 2015). Das *tertium comparationis* zeichnet sich dann – ebenso wie die Kernkategorie – durch seine Vorläufigkeit und Prozesshaftigkeit aus, die dem Prinzip der Offenheit qualitativen Forschens (vgl. Kruse, 2014) auch in der Analyse gerecht wird.

Insbesondere für explorative vergleichende Untersuchungen zu bisher wenig erforschten Fragestellungen – wie im Fall der hier zitierten Untersuchung von Gerechtigkeitsüberzeugungen von Lehrkräften – können die Prämissen der GTM hilfreiche Orientierungen bei der Wahl der einzubeziehenden Fälle und Datensorten, aber auch der komparativen Analyse geben. Gleichwohl ermöglicht sie eine gegenstandsverankerte Theoriebildung, die wiederum das Verständnis des untersuchten sozialen Phänomens systematisch erweitert und die komplexen Wechselbeziehungen zwischen institutionellen Rahmenbedingungen, lokalen Praktiken und individuellen Überzeugungen sichtbar macht. Andererseits sind der Generalisierbarkeit der Ergebnisse von GTM-Studien Grenzen gesetzt – wie jedoch grundsätzlich bei qualitativen Forschungszugängen.

Ein Allheilmittel gegenüber den eingangs erwähnten Gefahren vergleichender Untersuchungen der Nostri- bzw. Altrifizierung (vgl. Parreira do Amaral, 2015) stellt natürlich auch die GTM nicht dar. Allerdings denke ich, dass sie einige Hilfsmittel bietet, diese Gefahren zu minimieren und eine fruchtbare Erweiterung des methodologisch-methodischen Spektrums qualitativ-vergleichender erziehungswissenschaftlicher Forschung darstellt.

Anmerkungen

1. Diese Studie ist im Rahmen des von der DFG geförderten Emmy-Noether-Projekts ‚Unterschiedliche Welten der Meritokratie? Schulische Leistungsbeurteilung und Verteilungsgerechtigkeit in Deutschland, Schweden und England im Zeitalter der ‚standards-based reform‘‘ entstanden (Laufzeit 2010–2017, Projektleitung Prof. Dr. Florian Waldow, Humboldt-Universität zu Berlin).
2. Für eine Diskussion der Vorteile, aber auch Herausforderungen dieses iterativen Forschungsprozesses der GTM in vergleichend angelegten Untersuchungen vgl. Rupp (2016).
3. Zur Schwierigkeit, den Zeitpunkt der theoretischen Sättigung genau zu bestimmen vgl. Kelle (2003), der diese Frage für ausschließlich forschungspragmatisch beantwortbar hält, also in Abhängigkeit dessen, wie viele Daten innerhalb eines Forschungsprojekts erhoben und ausgewertet werden können. Das Ergebnis eines solchen Forschungsprozesses könne daher auch ‚nur‘ als vorläufig betrachtet werden (Kelle, 2003, S. 244).
4. Bis 2011 erhielten schwedische Grundschülerinnen und Grundschüler erst ab der 8. Klasse Noten. Die Rückmeldung in den unteren Klassen erfolgt durch Elemente formativer Beurteilung, wie Beurteilungsgespräche, Selbsteinschätzungsbögen oder Verbalzeugnisse.
5. Die 2011 eingeführten Sekundarschulen wurden nicht ins Sample aufgenommen, da die angefragten Schulen zum Befragungszeitpunkt zumeist noch in der Gründungsphase waren und aus organisatorischen Gründen eine Teilnahme an der Studie ablehnten.
6. In diesem Zusammenhang wird oft auf das sogenannte Konzept-Indikator-Modell verwiesen, mit dem Glaser (1978) in ‚Theoretical sensitivity‘ die bereits in ‚The discovery of Grounded Theory‘ angelegte Grundidee der Auswertung festhielt (vgl. hierzu auch Mey & Mruck, 2009, S. 109).
7. Hier finden sich auch Parallelen zum Konzept qualitativer Mehrebenenanalysen (vgl. Helsper, Hummrich & Kramer, 2010).
8. Für eine ausführlichere Diskussion der Frage von Mehrsprachigkeit in vergleichenden GTM-Studien vgl. Vogt (2017, S. 42–49), Rupp (2016) sowie für qualitative Interviewforschung allgemein Kruse, Bethmann, Niermann & Schmieder (2012).

Literatur

- Adick, C. (2014). Vergleichen – aber wie? Methodik und Methodologie in der Vergleichenden Erziehungswissenschaft. In C. Freitag (Hrsg.), *Methoden des Vergleichs. Komparatistische Methodologie und Forschungsmethodik in interdisziplinärer Perspektive* (S. 15–38). Opladen: Budrich UniPress.
- Berger, P.L. & Luckmann, T. (1977). *Die gesellschaftliche Konstruktion der Wirklichkeit: Eine Theorie der Wissenssoziologie* (5. Aufl.). Frankfurt a.M.: Fischer.
- Bryant, A. & Charmaz, K. (Hrsg.). (2007). *The SAGE handbook of Grounded Theory*. London: Sage.
- Charmaz, K. (2006). *Constructing Grounded Theory: A practical guide through qualitative analysis*. Thousand Oaks: Sage Publications.
- Clarke, A.E. (2012). *Situationsanalyse: Grounded Theory nach dem Postmodern Turn*. Wiesbaden: Springer VS.
- Crossley, M. & Jarvis, P. (2001). Context matters. *Comparative Education*, 37 (4), 405–408.
- Dale, R. (2005). Globalisation, knowledge economy and comparative education. *Comparative Education*, 41 (2), 117–149.

- Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge: Polity Press.
- Falkenberg, K. (2017). *Gerechte Noten? Eine Grounded Theory-Studie zu Gerechtigkeitsüberzeugungen von Lehrkräften in Bezug auf schulische Leistungsbeurteilung im deutsch-schwedischen Vergleich*. Unveröffentlichte Dissertation, Humboldt-Universität zu Berlin.
- Flick, U. (2011). Das Episodische Interview. In G. Oelerich & H.-U. Otto (Hrsg.), *Empirische Forschung und Soziale Arbeit. Ein Studienbuch* (S. 273–280). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Glaser, B.G. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.
- Glaser, B.G. (2002). Conceptualization: On theory and theorizing using Grounded Theory. *International Journal of Qualitative Methods*, 1 (2), 1–31.
- Glaser, B.G. (2004). Remodeling Grounded Theory. *Forum: Qualitative Sozialforschung*, 5 (2). Verfügbar unter www.qualitative-research.net/index.php/fqs/article/view/607/1315 [24.04.2017].
- Glaser, B.G. & Strauss, A.L. (1967). *The discovery of Grounded Theory: Strategies for qualitative research*. Hawthorne, New York: de Gruyter.
- Heintz, B. (2010). Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs. *Zeitschrift für Soziologie*, 39 (3), 162–181.
- Hellawell, D. (2006). Inside-out: Analysis of the insider-outsider concept as heuristic device to develop reflexivity in students doing qualitative research. *Teaching in Higher Education*, 11 (4), 483–494.
- Helsper, W., Hummrich, M. & Kramer, R.-T. (2010). Qualitative Mehrebenenanalyse. In B. Friebertshäuser, A. Langer & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (3. Aufl.) (S. 119–135). Weinheim: Juventa.
- Hopmann, S.T. (2003). On the evaluation of curriculum reforms. *Journal of Curriculum Studies*, 35 (4), 459–478.
- Huf, C. (2015). Children's agency and teacher's control – Comparative Ethnography in Childhood Institutions. In B. Fritzsche & C. Huf (Eds.), *The benefits, problems, and issues of comparative and cross-cultural research: Ethnographic perspectives* (pp. 15–34). Gloucestershire: E&E Publishing.
- Hummrich, M. & Rademacher, S. (2012). Die Wahlverwandtschaft von qualitativer Forschung und Kulturvergleich und ihre Bedeutung für die Erziehungswissenschaft – strukturtheoretische Überlegungen. *Zeitschrift für Qualitative Forschung*, 13 (1), 39–53.
- Kelle, H. (2015). The problems of comparative ethnographic research and its potential: A meta-perspective. In B. Fritzsche & C. Huf (Eds.), *The benefits, problems, and issues of comparative and cross-cultural research: Ethnographic perspectives* (pp. 93–110). Gloucestershire: E&E Publishing.
- Kelle, U. (2003). Die Entwicklung kausaler Hypothesen in der qualitativen Sozialforschung. Methodologische Überlegungen zu einem häufig vernachlässigten Aspekt qualitativer Theorie- und Typenbildung. *Zentralblatt für Didaktik der Mathematik (ZDM)*, 35 (6), 232–246.
- Kruse, J. (2014). *Qualitative Interviewforschung. Ein integrativer Ansatz*. Weinheim: Beltz Juventa.
- Kruse, J., Bethmann, S. Niermann, D. & Schmieder, C. (2012). (Hrsg.). *Qualitative Interviewforschung in und mit fremden Sprachen. Eine Einführung in Theorie und Praxis*. Weinheim: Beltz Juventa.
- Liebeskind, U. (2012). Komparative Verfahren und Grounded Theory. In K. Schittenhelm (Hrsg.), *Qualitative Bildungs- und Arbeitsmarktforschung. Grundlagen, Perspektiven, Methoden* (S. 325–358). Wiesbaden: Springer Fachmedien.

- Matthes, J. (1992). The Operation Called ‚Vergleichen‘. In J. Matthes (Hrsg.), *Zwischen den Kulturen? Die Sozialwissenschaften vor dem Problem des Kulturvergleichs* (Soziale Welt, Sonderbd. 8) (S. 75–99). Göttingen: Schwartz.
- McNess, E., Arthur, L. & Crossley, M. (2015). ‚Ethnographic dazzle‘ and the construction of the ‚other‘: Revisiting dimensions of insider and outsider research for international and comparative education. *Compare*, 45 (2), 295–316.
- Mey, G. & Mruck, K. (2009). Methodologie und Methodik der Grounded Theory. In W. Kempf & M. Kiefer (Hrsg.), *Forschungsmethoden der Psychologie. Zwischen naturwissenschaftlichem Experiment und sozialwissenschaftlicher Hermeneutik. Bd. 3: Natur und Kultur* (S. 100–152). Berlin: Regener.
- Milligan, L. (2016). Insider-outsider-inbetween? Researcher positioning, participative methods and cross-cultural educational research. *Compare*, 46 (2), 235–250.
- Nohl, A.-M. (2009). Der Mehrebenenvergleich als Weg zum kontextuierten Ländervergleich. Methodologische Überlegungen anhand eines internationalen Projektes zur Migrationsforschung. In S. Hornberg, I. Dirim, G. Lang-Wojtasik & P. Mecheril (Hrsg.), *Beschreiben – Verstehen – Interpretieren. Stand und Perspektiven International und Interkulturell Vergleichender Erziehungswissenschaft in Deutschland* (S. 95–110). Münster: Waxmann.
- Parreira do Amaral, M. (2015). Methodologie und Methode in der International Vergleichenden Erziehungswissenschaft. In M. Parreira do Amaral & K. Amos (Hrsg.), *Internationale und Vergleichende Erziehungswissenschaft. Geschichte, Theorie, Methode und Forschungsfelder* (S. 107–130). Münster: Waxmann.
- Rademacher, S. (2013). Fallvergleich und Kulturvergleich. Methodologische Implikationen einer kulturvergleichenden qualitativen Forschung. In M. Hummrich & S. Rademacher (Hrsg.), *Kulturvergleich in der qualitativen Forschung. Erziehungswissenschaftliche Perspektiven und Analysen* (S. 65–77). Wiesbaden: Springer VS.
- Russer, K., Pauli, C. & Elmer, A. (2011). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 478–495). Münster: Waxmann.
- Rippl, S. & Seipel, C. (2008). *Methoden kulturvergleichender Sozialforschung: Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rupp, C. (2016). Doing qualitative comparative research on teaching: Challenges and benefits of working with Grounded Theory. *Comparative and International Education*, 11 (4), 422–433.
- Schittenhelm, K. (2012). Sampling und die Suche nach fallübergreifender Gültigkeit. Vergleichende Analysen von Statusübergängen zwischen Bildungsabschluss und Arbeitsmarkt. In K. Schittenhelm (Hrsg.), *Qualitative Bildungs- und Arbeitsmarktforschung. Grundlagen, Perspektiven, Methoden* (S. 407–437). Wiesbaden: Springer Fachmedien.
- Schriewer, J. (1987). Vergleich als Methode und Externalisierung auf Welt: Vom Umgang mit Alterität in Reflexionsdisziplinen. In D. Baecker, J. Markowitz, R. Stichweh, H. Tyrell & H. Willke (Hrsg.), *Theorie als Passion. Niklas Luhmann zum 60. Geburtstag* (S. 629–668). Frankfurt a.M.: Suhrkamp.
- Schriewer, J. (2003). Problemdimensionen sozialwissenschaftlicher Komparatistik. In H. Kaelble & J. Schriewer (Hrsg.), *Vergleich und Transfer. Komparatistik in den Sozial-, Geschichts- und Kulturwissenschaften* (S. 11–54). Frankfurt a.M.: Campus.
- Schriewer, J. (2006). Comparative social science: Characteristic problems and changing problem solutions. *Comparative Education*, 42 (3), 299–336.

- Schröder, N. (2013). Zur hermeneutisch-wissenssoziologischen Auslegung des Fremden. Interpretieren mit Unterstützung kulturvertrauter Co-Interpreten. In R. Bettmann & M. Roslon (Hrsg.), *Going the distance. Impulse für die interkulturelle Qualitative Sozialforschung* (S. 61–76). Wiesbaden: Springer VS.
- Steiner-Khamsi, G. (2010). The politics and economics of comparison. *Comparative Education Review*, 54 (3), 323–342.
- Strauss, A.L. (1994). *Grundlagen qualitativer Sozialforschung: Datenanalyse und Theoriebildung in der empirischen soziologischen Forschung*. München: Fink.
- Strauss, A.L. & Corbin, J.M. (1996). *Grounded Theory: Grundlagen Qualitativer Sozialforschung*. Weinheim: Beltz.
- Terhart, E. (2011). Die Beurteilung von Schülern als Aufgabe des Lehrers: Forschungslinien und Forschungsergebnisse. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 699–717). Münster: Waxmann.
- Vogt, B. (2017). *Just assessment in school. A context-sensitive comparative study of pupil's conceptions in Sweden and Germany*. Växjö: Linnaeus University Press.
- Waldow, F. (2011). Juristen oder Testspezialisten? Zur Rolle von Experten bei der Herstellung von Notengerechtigkeit in Deutschland und Schweden. *Zeitschrift für Pädagogik*, 57 (4), 484–496.
- Waldow, F. (2014). Conceptions of justice in examination systems of England, Germany and Sweden: A look at safeguards of fair procedure and possibilities of appeal. *Comparative Education Review*, 58 (2), 322–343.
- Waldow, F. (2015). Coda: Hat sich die Vergleichende Erziehungswissenschaft zu Tode gesiegt? Wozu brauchen wir sie noch? *Tertium Comparationis*, 21 (1), 130–139.
- Walzer, M. (1983). *Spheres of justice: A defense of pluralism and equality*. New York: Basic Books.
- Waterkamp, D. (2006). *Vergleichende Erziehungswissenschaft. Ein Lehrbuch*. Münster: Waxmann.
- Weiß, A. & Nohl, A.M. (2012). Fälle und Kontexte im Mehrebenenvergleich. Ein Vorschlag zur Überwindung des methodologischen Nationalismus in der Migrationsforschung. *Zeitschrift für Qualitative Forschung*, 13 (1/2), 55–75.
- West, A. & Nikolai, R. (2013). Welfare regimes and education regimes: Equality of opportunity and expenditure in the EU (and US). *Journal of Social Policy*, 42 (3), 469–493.
- Wimmer, A. & Glick Schiller, N. (2002). Methodological nationalism and beyond: Nation-state building, migration and the social sciences. *Global Networks*, 2 (4), 301–334.