



# Qualitative Inhaltsanalyse in naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von Gütekriterien: Ein Review

Maximilian Göhner<sup>1</sup> · Moritz Krell<sup>1</sup>

Eingegangen: 10. Januar 2020 / Angenommen: 29. Februar 2020 / Online publiziert: 16. März 2020  
© Der/die Autor(en) 2020

## Zusammenfassung

Im Rahmen ihrer Entwicklung und Etablierung als empirisch arbeitende Disziplinen findet in den Naturwissenschaftsdidaktiken vermehrt eine Auseinandersetzung mit methodischen und methodologischen Fragen statt. Hierzu gehört unter anderem die Frage danach, was *gute* fachdidaktische Forschung überhaupt ausmacht. Obwohl die qualitative Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung als ein etabliertes Verfahren gilt, unterscheiden sich Vorschläge hinsichtlich anzulegender Gütekriterien und umzusetzender Maßnahmen der Qualitätssicherung. Im Sinne einer exemplarischen Sammlung erprobter Forschungspraxis wird in diesem Beitrag ein Überblick über die in der deutschsprachigen naturwissenschaftsdidaktischen Forschung etablierten Gütekriterien und Maßnahmen zur Qualitätssicherung bei der Umsetzung von Verfahren der qualitativen Inhaltsanalyse gegeben. Hierzu werden 50 in der *Zeitschrift für Didaktik der Naturwissenschaften* publizierte Artikel, die eine qualitative Inhaltsanalyse umsetzen, mit einem deduktiv-induktiv erstellten Kategoriensystem untersucht. Es werden 11 Gütekriterien identifiziert, wobei vorwiegend die klassischen Gütekriterien Validität, Reliabilität und Objektivität in den untersuchten Artikeln beschrieben werden. Von 16 identifizierten Maßnahmen der Qualitätssicherung werden am häufigsten die Modi der Kategorienbildung und die Interrater-Übereinstimmung beschrieben. Es ergibt sich ein signifikant positiver Zusammenhang zwischen dem Publikationsjahr und der Zahl der durchschnittlich beschriebenen Maßnahmen zur Qualitätssicherung. Allerdings werden die einzelnen Maßnahmen der Qualitätssicherung, über alle Artikel hinweg betrachtet, nicht konsistent jeweils einem Gütekriterium zugeordnet. Die Ergebnisse sollen die Diskussion in den Fachdidaktiken anregen und zur weiteren Etablierung und Systematisierung des Verfahrens der qualitativen Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung beitragen.

**Schlüsselwörter** Qualitative Forschung · Qualitative Inhaltsanalyse · Gütekriterien · Objektivität · Reliabilität · Validität

---

✉ Maximilian Göhner  
maximilian.goehner@fu-berlin.de

<sup>1</sup> Didaktik der Biologie, Freie Universität Berlin,  
Schwendenerstraße 1, 14195 Berlin, Deutschland

## Qualitative Content Analysis in Science Education Research Under the Consideration of Quality Criteria: a Review

### Abstract

In the course of developing and establishing the field of science education, methodical and methodological questions have become increasingly important. This includes questioning what poses as *good* qualitative research. Although, qualitative content analysis is considered an established method in science education research, there are different ideas regarding which quality criteria and measures of quality control have to be taken into account. To give an exemplary collection of reliable research practice, the present article reviews the quality criteria and measures of quality control used in science education publications utilizing a qualitative content analysis. 50 publications of the German journal *Zeitschrift für Didaktik der Naturwissenschaften* were investigated with a coding scheme, that was deductively-inductively developed. 11 quality criteria were identified, with validity, reliability and objectivity being the most common. 16 measures of quality control were identified, whereby descriptions of the development process of the categories and interrater-agreement were most commonly described. Also, there is a significant positive correlation between the year of publication and the number of measures of quality control described. However, these measures of quality control are inconsistently assigned to quality criteria, when all publications are taken into account. These results may spark further discussion in the science education community and support utilizing the method of qualitative content analysis more systematically.

**Keywords** Qualitative research · Qualitative content analysis · Quality criteria · Objectivity · Reliability · Validity

### Einleitung

Die Fachdidaktiken haben sich in Deutschland in den letzten Jahrzehnten zu eigenständigen, empirisch arbeitenden wissenschaftlichen Disziplinen entwickelt (Bayrhuber et al. 2011; Leuders 2015). Das Spezifische der Fachdidaktiken liegt hierbei in ihrem Fachbezug, wobei sie methodisch mit benachbarten Disziplinen verbunden sind (z. B. empirische Bildungsforschung, Psychologie). Im Rahmen ihrer Entwicklung und Etablierung als empirisch arbeitende Disziplinen findet in den Naturwissenschaftsdidaktiken vermehrt eine Auseinandersetzung mit methodischen und methodologischen Fragen statt (Leuders 2015); was auch an der Publikation von Lehrbüchern und Sammelbänden zu Methoden in der naturwissenschaftsdidaktischen Forschung erkennbar ist (z. B. Groß et al. 2019; Sandmann und Schmiemann 2016; Krüger et al. 2014; Reinisch et al. in Vorbereitung). Hierzu gehört auch die Frage danach, was *gute* fachdidaktische Forschung überhaupt ausmacht. Aktuelle Reviews zeigen, dass es nicht nur Hinweise auf eine mangelnde Replizierbarkeit von Studien in der sozialwissenschaftlichen und erziehungswissenschaftlichen Forschung gibt (Camerer et al. 2018; Makel und Plucker 2014), sondern auch, dass insbesondere mit Bezug auf qualitative Arbeiten sowohl in den Erziehungswissenschaften (Grunenberg 2004) als auch in den Sozialwissenschaften (Barusch et al. 2011) die praktische Umsetzung qualitätssichernder Maßnahmen oft sehr unterschiedlich ausfällt. Da die hohe Vielfalt qualitativer Methoden (Howe und Eisenhart 1990) und die im Vergleich zu quantitativer Forschung große Bedeutsamkeit von interpretativer Sinnkonstruktion (Patton 2001) die Etablierung von allgemeingültigen Gütekriterien für die qualita-

tive fachdidaktische und erziehungswissenschaftliche Forschung erschweren (vgl. Steinke 2013; Flick 2010), werden methodenspezifisch anzulegende Gütekriterien und Maßnahmen diskutiert (vgl. „methodological awareness“; Seale 2002). Dies zeigt sich auch in einer Forderung des Roundtables der Deutschen Forschungsgemeinschaft (DFG) zur „Qualität qualitativer Forschung in der Erziehungswissenschaft“ (Helsper et al. 2001). Dort wurde die exemplarische Sammlung erprobter Forschungspraxen gefordert, um WissenschaftlerInnen eine Orientierung bei der Umsetzung von qualitätssichernden Maßnahmen in qualitativen Forschungsprojekten zu bieten.

Zu den in der qualitativen sozialwissenschaftlichen und erziehungswissenschaftlichen Forschung etablierten Methoden gehören beispielsweise die verschiedenen Verfahren der qualitativen Inhaltsanalyse (Schreier 2014; Mayring 2019). Auch in der naturwissenschaftsdidaktischen Forschung kommt der qualitativen Inhaltsanalyse eine besondere Bedeutung zu, die sich etwa in spezifisch naturwissenschaftsdidaktischen, methodologischen Beiträgen zeigt (z. B. Gropengießer 2005; Krüger und Riemeier 2014). Die vorliegende Analyse erarbeitet einen Überblick über die in der deutschsprachigen naturwissenschaftsdidaktischen Forschung etablierten Gütekriterien und Maßnahmen zur Qualitätssicherung bei der Umsetzung von Verfahren der qualitativen Inhaltsanalyse. Durch diese Sammlung erprobter Forschungspraxen (Helsper et al. 2001) trägt die Studie dazu bei, Verfahren der qualitativen Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung weiter zu etablieren und mögliche qualitätssichernde Maßnahmen zu systematisieren.

## Qualitative Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung

Unter dem Begriff *qualitative Inhaltsanalyse* werden verschiedene Varianten eines methodischen Verfahrens zusammengefasst, welches auf die „Systematisierung und Interpretation von manifesten und latenten Kommunikationsinhalten unter Verwendung eines Kategoriensystems“ abzielt (Stamann et al. 2016, S. 13). Traditionell wurden qualitativ inhaltsanalytische Verfahren zu den qualitativen Forschungsmethoden gezählt (z. B. Lamnek 2010). Aktuell vertortet Mayring (2019) diese allerdings zwischen dem qualitativen und quantitativen Forschungsparadigma, da einzelne Varianten jeweils spezifische Nähe zu quantitativer oder qualitativer Methodologie aufweisen. Eine generelle Einordnung der qualitativen Inhaltsanalyse als qualitative Methode sei demnach nicht möglich (Mayring 2019). Schreier (2014) unterscheidet zwischen acht verschiedenen Varianten der qualitativen Inhaltsanalyse, wobei zwei Varianten von ihr besonders hervorgehoben werden, die sie als Basisformen der qualitativen Inhaltsanalyse bezeichnet. Die erste Basisform, die inhaltlich-strukturierende Inhaltsanalyse, identifiziert die für eine Fragestellung relevanten Aspekte aus vorliegendem Material (z. B. Texten), indem systematisch und iterativ ein Kategoriensystem entwickelt, erprobt und modifiziert wird (vgl. Kuckartz 2016; Mayring 2015; Schreier 2012). Hierbei zeigen aktuelle Diskussionen jedoch weiteren Abgrenzungsbedarf (Mayring 2019). So unterscheidet sich zum Beispiel auch innerhalb der Verfahren, die von Schreier (2014) zur Basisform der inhaltlich-strukturierenden qualitativen Inhaltsanalyse zugeordnet werden, das Verhältnis von deduktiv und induktiv abgeleiteten Kategorien. Dies kann je nach Forschungsgegenstand variieren (vgl. Kuckartz 2016; Schreier 2012; Mayring 2015), wobei Mayring (2015) die Bedeutung der deduktiven Kategorienbildung für die inhaltlich-strukturierende qualitative Inhaltsanalyse besonders betont. Die zweite Basisform, die qualitative Inhaltsanalyse durch Extraktion (vgl. Gläser und Laudel 2013), zielt auf den Einbezug des gesamten untersuchten Materials unter möglichst hoher ökonomischer Effizienz ab. Deduktiv entwickelte Kategorien dienen als Rahmen für die induktive Verortung aller in den Texten enthaltenen und durch Extraktion zusammengefassten Informationen. Zur weiteren Systematisierung der verschiedenen Varianten schlagen Stamann et al. (2016) vor, neben der Basisform (inhaltlich-strukturierend, extrahierend), der Art der Kategorien (formal, skalierend, inhaltlich) oder des Kategoriensystems (statisch, dynamisch, linear-hierarchisch, netzwerkartig) und den oben genannten Modi der Kategorienbildung (induktiv, deduktiv) auch die weitere Vorgehensweise mit den qualitativ-inhaltsanalytisch gewonnenen Daten (z. B. die Analyse von Kodierhäufigkeiten oder die Entwicklung einer Typologie) als Bezugspunkte zu betrachten.

Diese Vielfalt an Varianten der qualitativen Inhaltsanalyse illustriert, dass unterschiedliche methodologische Positionen zur qualitativen Inhaltsanalyse bestehen, die sich auch in aktuellen Diskussionen und Abgrenzungsversuchen wiederfinden (z. B. Mayring 2019).

Für die qualitative Inhaltsanalyse als etabliertes Verfahren der erziehungswissenschaftlichen und fachdidaktischen Forschung sind klar definierte Qualitätskriterien erforderlich, die für die Planung und Umsetzung von Forschungsvorhaben, für die Rezeption und Einordnung von Forschungsergebnissen sowie für die externe Begutachtung (z. B. *peer-review*, Forschungsförderung) leitend sein können (Helsper et al. 2016). Vergleichbar mit der Vielfalt qualitativer Forschungsmethoden im Allgemeinen (Howe und Eisenhart 1990), erschwert auch die Vielfalt der qualitativ-inhaltsanalytischen Verfahren im Besonderen die Antwort auf die Frage danach, was eine *gute* qualitative Inhaltsanalyse ausmacht und welche Gütekriterien an qualitativ-inhaltsanalytische Verfahren angelegt werden sollten. So argumentieren einige AutorInnen, dass die qualitative Inhaltsanalyse aus einer quantitativen Tradition hervorgeht und demnach auch klassische Gütekriterien wie Reliabilität und Validität bei deren Umsetzung angelegt werden sollten (vgl. Krippendorff 1980; Schreier 2012; Mayring 2015), wobei Objektivität meist auf Grund des hohen interpretativen Anteils der qualitativen Inhaltsanalyse mit Intersubjektivität gleichgesetzt (Mayring 2015) oder als Gütekriterium für die qualitative Inhaltsanalyse abgelehnt wird (Schreier 2012). Weitere Gütekriterien orientieren sich am Prozess des Verfahrens und können spezifische Maßnahmen zur Qualitätssicherung umfassen. So nennt Mayring (2002) beispielsweise die Verfahrensdokumentation, die argumentative Interpretationsabsicherung, die Regelgeleitetheit, die Nähe zum Gegenstand, die kommunikative Validierung und die Triangulation als Gütekriterien der inhaltlich-strukturierenden qualitativen Inhaltsanalyse. Etwas allgemeiner und orientiert an den zentralen Prinzipien der qualitativen Sozialforschung beschreibt Lamnek (2010) darüber hinaus Offenheit, Kommunikativität, Naturalistizität und Interpretativität als Merkmale einer guten qualitativen Inhaltsanalyse. Insgesamt wird damit deutlich, dass vielfältige Vorschläge zu Gütekriterien und entsprechenden Maßnahmen der Qualitätssicherung vorliegen. Diese Vielfältigkeit wird dem Anspruch klar definierter Qualitätskriterien (Helsper et al. 2001) allerdings nicht gerecht. Beispielsweise weist Mayring (2015) darauf hin, dass die oftmals als Inter-Koderreliabilität bezeichnete Beurteilerübereinstimmung ein Maß für die Auswertungsobjektivität ist, während Schreier (2012) die Beurteilerübereinstimmung entsprechend ihrer Bezeichnung der Reliabilität zuordnet.

## Ziel und Fragestellungen

Vor dem Hintergrund der oben dargestellten Relevanz qualitativ-inhaltsanalytischer Verfahren in der naturwissenschaftsdidaktischen Forschung ist es das Ziel dieser Analyse, einen Überblick über die in der deutschsprachigen naturwissenschaftsdidaktischen Forschung gängige Praxis bei der Beschreibung von Gütekriterien und qualitätssichernden Maßnahmen im Hinblick auf qualitativ-inhaltsanalytische Verfahren zu geben. Die vorliegende Analyse identifiziert häufig genannte Gütekriterien und qualitätssichernde Maßnahmen und stellt damit eine exemplarische Sammlung erprobter Forschungspraxis als Orientierung für Forschungsprojekte dar (Helsper et al. 2001). Ziel ist es ferner, damit die qualitative Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung weiter zu etablieren.

Die folgenden Fragen leiteten die vorliegende Analyse:

1. Welche Gütekriterien werden für die qualitative Inhaltsanalyse in der deutschsprachigen naturwissenschaftsdidaktischen Forschung beschrieben?
2. Welche Maßnahmen der Qualitätssicherung von qualitativen Inhaltsanalysen werden in der deutschsprachigen naturwissenschaftsdidaktischen Forschung beschrieben?
3. Inwieweit werden für die qualitative Inhaltsanalyse in der deutschsprachigen naturwissenschaftsdidaktischen Forschung einzelne Maßnahmen der Qualitätssicherung konsistent bestimmten Gütekriterien zugeordnet?

## Methoden

Um im Sinne eines systematischen Reviews (vgl. Bennett et al. 2005) die gängige Praxis der deutschsprachigen naturwissenschaftsdidaktischen Forschung beschreiben zu können, wurden als Stichprobe alle bis zum 16.10.2018 publizierten Artikel der *Zeitschrift für Didaktik der Naturwissenschaften* mit dem Stichwort „Inhaltsanalyse“ im Fließtext ausgewählt<sup>1</sup>. Als Datengrundlage wurde die *Zeitschrift für Didaktik der Naturwissenschaften* gewählt, da sie als Verbandszeitschrift der Gesellschaft für Didaktik der Chemie und Physik (GDGP) und der Fachsektion Didaktik der Biologie (FDdB) im Verband Biologie, Biowissenschaften und Biomedizin (VBIO) als repräsentativ für deutschsprachige naturwissenschaftsdidaktische Forschung aufgefasst werden kann. Weitere Zeitschriften wurden aus forschungsökonomischen Gründen nicht in die Analyse miteinbezogen. Zudem wurde nicht weitergehend zwischen verschiedenen Varianten der qualitativen Inhaltsanalyse unter-

schieden (Schreier 2014), da in den identifizierten Artikeln größtenteils keine Angabe hierzu gemacht wurde.

58 der 256 insgesamt in diesem Zeitraum publizierten Artikel entsprachen dem Auswahlkriterium. Acht dieser 58 Artikel wurden nach der ersten Sichtung ausgeschlossen, da in ihnen die qualitative Inhaltsanalyse nicht als primäre Methode beschrieben, sondern zum Beispiel nur im Ausblick diskutiert wurde. Die 50 verbliebenen Artikel wurden anschließend qualitativ-inhaltsanalytisch untersucht (siehe separates Literaturverzeichnis; Anhang A). Hierbei wurde entsprechend der inhaltlich-strukturierenden Variante der qualitativen Inhaltsanalyse nach Schreier (2012) verfahren, um beschriebene Gütekriterien (Frage 1) und Maßnahmen zur Qualitätssicherung (Frage 2) in den Artikeln (Analyseeinheit) zu identifizieren, zu konzeptualisieren und die Artikel im Hinblick darauf zu beschreiben (vgl. Schreier 2014).

Ein Kategoriensystem bestehend aus Gütekriterien und Maßnahmen zur Qualitätssicherung wurde deduktiv-induktiv entwickelt. Als theoretische Grundlage für die deduktive Kategorienentwicklung wurden alle Texte herangezogen, die in den Methodenbeschreibungen der 50 analysierten Artikel im Zusammenhang mit der qualitativen Inhaltsanalyse referenziert werden. Das deduktive Vorgehen wurde gewählt, um die Beschreibung der Gütekriterien und Maßnahmen der Qualitätssicherung in der qualitativen Inhaltsanalyse in einzelnen Studien mit methodologischen Empfehlungen kontrastieren zu können. Da oft verschiedene Ausgaben des gleichen Textes referenziert wurden, wurden hier die älteste und die jüngste angegebene Auflage untersucht ( $N=21$ , siehe separates Literaturverzeichnis; Anhang B). Es wurden nur diejenigen Gütekriterien und Maßnahmen in das Kategoriensystem aufgenommen, die von den AutorInnen des jeweiligen Artikels explizit der qualitativen Inhaltsanalyse zugeordnet wurden. Weitere Gütekriterien und Maßnahmen der Qualitätssicherung, etwa die Datenerhebung betreffend, sind nicht Gegenstand der vorliegenden Analyse.

Die induktive Erweiterung des Kategoriensystems erfolgte auf der Basis der Kodierung von 50% der analysierten Artikel. Dabei wurden Abschnitte kodiert, die mindestens einen ganzen Satz umfassen (Kodiereinheit). Zur weiteren Einordnung von Beschreibungen wurde gegebenenfalls die referenzierte methodologische Literatur herangezogen (Kontexteinheit). Die Artikel wurden unabhängig durch zwei Personen ausgewertet (Erstauswertung: Erstautor des Artikels; Zweitauswertung: geschulte studentische Hilfskraft) und es wurde die Interrater-Übereinstimmung geprüft, wobei Cohens Kappa als zufallskorrigiertes Maß der Beurteilerübereinstimmung gewählt wurde (Cohens Kappa = 0,70; „gut“, Wirtz und Caspar 2002). Nicht-übereinstimmungen wurden im Rahmen einer diskursiven Konsensfindung gelöst, dabei wurden gegebenenfalls wei-

<sup>1</sup> Der Zugriff erfolgte über <http://archiv.ipn.uni-kiel.de/zfdn/> und <https://link.springer.com/journal/40573>.

tere Kategorien hinzugefügt, und die bisherigen Kategorien um Definitionen, Kodierhinweise, Stichworte und Ankerbeispiele ergänzt.

Das sich hieraus ergebende Kategoriensystem wurde innerhalb der eigenen Arbeitsgruppe im Sinne eines Peer-Debriefings diskutiert (vgl. Steinke 2013), dabei wurde es insbesondere hinsichtlich seiner Vollständigkeit, der Struktur und den Kodierhinweisen bewertet (vgl. Lisch und Kriz 1978). Anschließend wurde das beschriebene Verfahren auf die übrigen 50 % der Artikel angewendet (Cohens Kappa=0,60; „gut“, Wirtz und Caspar 2002). Für das finale Kategoriensystem (Anhang C) ergaben sich so 27 Kategorien, wobei 11 Kategorien auf die Gütekriterien entfallen und 16 Kategorien Maßnahmen der Qualitätssicherung beschreiben (Tab. 1).

Alle Kodierungen wurden computergestützt mit dem Programm MAXQDA vorgenommen. Zur Beantwortung der Forschungsfragen 1 und 2 wurde jeweils eine Häufigkeitstabelle erstellt, die anzeigt in wie vielen Artikeln die einzelnen Gütekriterien (Frage 1) und qualitätssichernden Maßnahmen (Frage 2) beschrieben werden. Für die Gegenüberstellung von Gütekriterien und qualitätssichernden Maßnahmen (Frage 3) wurde eine Kreuztabelle erstellt, die aufzeigt, wie oft welche Gütekriterien mit welchen qualitätssichernden Maßnahmen zusammenhängend genannt werden.

## Ergebnisse

### Gütekriterien (Frage 1)

Insgesamt beschreiben 50 % der untersuchten Artikel mindestens eines der betrachteten Gütekriterien (Tab. 2). Validität wird in 15 Artikeln (30 %) beschrieben und stellt damit das am häufigsten beschriebene Gütekriterium dar. Hierbei kann Validität als Gütekriterium an das gesamte Verfahren der Auswertung (z. B. Gropengießer 1997), an das Kategoriensystem (z. B. Kühn 2011), an einzelne Kategorien (z. B. Kulgemeyer und Tomczyszyn 2015), an den Entwicklungsprozess der Kategorien (z. B. Mittelsten Scheid und Hößle 2008), an die Untersuchungsergebnisse (z. B. Weitzel und Gropengießer 2009) oder an die Interpretation der Untersuchungsergebnisse angelegt werden (z. B. Kulgemeyer und Tomczyszyn 2015). Zusätzlich werden verschiedene Validitätsaspekte unterschieden, beispielsweise die interne und externe Validität (z. B. Brovelli 2014) oder die Inhaltsvalidität, Kriteriumsvalidität, Handlungsvalidität und prognostische Validität (z. B. Kulgemeyer und Tomczyszyn 2015).

In elf Artikeln (22 %) und somit jeweils ähnlich häufig wie die Validität werden die beiden anderen klassischen Gütekriterien Reliabilität und Objektivität beschrieben. Reliabilität wird hierbei ausschließlich im Zusammen-

**Tab. 1** Übersicht über die im Kategoriensystem enthaltenen Gütekriterien und Maßnahmen der Qualitätssicherung (das vollständige Kategoriensystem ist Anhang C zu entnehmen)

<b>Gütekriterien</b>	
Validität/Gültigkeit	
Reliabilität	
Objektivität/Intersubjektivität	
Transparenz	
Offenheit	
Nähe zum Gegenstand/Authentizität/Naturalizität	
Kommunikativität	
Interpretativität	
Praktische Relevanz	
Praktikabilität/Ökonomie	
Ethische Überlegungen	
<b>Maßnahmen der Qualitätssicherung</b>	
Auswahl/Standardisierung des Settings	
Dokumentation	
Argumentative Interpretationsabsicherung	
Systematisches Vorgehen	
Triangulation	
Kommunikative Validierung	
Theoretisches Sampling	
Formale Aspekte des Kategoriensystems	
Induktive Bildung des Kategoriensystems	
Deduktive Bildung des Kategoriensystems	
Training der BeurteilerInnen	
Diskussion des Kategoriensystems	
Diskussion der Kodierungen	
Konsensbildung	
Prüfung der Intrarater-Übereinstimmung	
Prüfung der Interrater-Übereinstimmung	

**Tab. 2** Nennung der Gütekriterien in den 50 ausgewählten Artikeln der ZfDN, mehrere Kodierungen des gleichen Gütekriteriums werden pro Artikel nur einmal gezählt

Gütekriterien	Artikelanzahl
Validität/Gültigkeit	15
Reliabilität	11
Objektivität/Intersubjektivität	11
Nähe zum Gegenstand/Authentizität/Naturalizität	5
Praktikabilität/Ökonomie	5
Offenheit	3
Praktische Relevanz	3
Transparenz	3
Ethische Überlegungen	2
Kommunikativität	2
Interpretativität	1

hang mit bestimmten qualitätssichernden Maßnahmen beschrieben (vgl. Tab. 4). In den betrachteten Artikeln treten Objektivität und Intersubjektivität nicht gemeinsam auf, beide Begriffe werden jedoch als Gütekriterium für das Kategoriensystem (z. B. Kodierungen anhand von objektiven Textmerkmalen; Starauschek 2006), einzelne Kategorien (z. B. Baur 2018) oder einzelne Kodierungen und davon abgeleitete Interpretationen (z. B. Nehring et al. 2016) verwendet. Schmelzing et al. (2010) nennen darüber hinaus die Auswertungsobjektivität als einen Teilaspekt der Objektivität.

Die Nähe zum Gegenstand (Mayring 2015), auch Authentizität (vgl. Brovelli et al. 2013) oder Naturalizität genannt (vgl. Lamnek 2010), wird in 10 % der untersuchten Artikel beschrieben. Als Gütekriterium kann die Nähe zum Gegenstand dabei zum einen an die Datenaufnahme angelegt werden, zum Beispiel indem eine besonders authentische Situation geschaffen wird (z. B. Brovelli et al. 2013). Zum anderen wird die Nähe zum Gegenstand im Zusammenhang mit Kategoriensystemen beschrieben, in denen eine „möglichst authentische Erfassung des Bedeutungsgehaltes“ stattfindet (z. B. Kasper und Mikelskis 2008).

Praktikabilität oder ökonomische Überlegungen werden vor allem in Hinsicht auf formale Aspekte oder vorbereitende Schritte der qualitativen Inhaltsanalyse beschrieben. Solche Überlegungen umfassen zum Beispiel die Begründung des Einsatzes von Computerprogrammen (z. B. Reitschert und Höhle 2007), die Reduzierung des Umfangs von eingesetzten Instrumenten zur Datenaufnahme (z. B. Kulgemey-

er und Tomczyszyn 2015) sowie den Stichprobenumfang und die Stichprobenbeschaffenheit (z. B. Uhdn 2016). Beschreibungen der praktischen Relevanz, die im Zusammenhang mit der qualitativen Inhaltsanalyse getroffen werden, umfassen Implikationen für die schulische (Baur 2018) und universitäre Lehre (Kulgemeyer und Tomczyszyn 2015) sowie für die Gestaltung von Prüfungsanforderungen (Florian et al. 2015).

Die weiteren von Lamnek (2010) definierten Gütekriterien Offenheit, Kommunikativität und Interpretativität, die sich an den zentralen Prinzipien der qualitativen Sozialforschung orientieren, werden seltener beschrieben. So begründet Offenheit in drei Artikeln die Auswahl der qualitativen Inhaltsanalyse als Auswertungsmethode (Baalmann et al. 2004; Riemeier et al. 2010; Baur 2018), während Kommunikativität (z. B. Mittelsten Scheid und Höhle 2008) und Interpretativität (Baalmann et al. 2004) als Eigenschaften der qualitativen Inhaltsanalyse hervorgehoben, aber nicht weiter ausgeführt werden.

Ethische Überlegungen zum Einsatz der qualitativen Inhaltsanalyse werden in zwei Artikeln expliziert und heben die Rolle der ProbandInnen als Forschungspartner hervor (z. B. Gropengießer 1997; Uhdn 2016).

## Maßnahmen der Qualitätssicherung (Frage 2)

In 45 Artikeln (90 %) werden eine oder mehrere Maßnahmen der Qualitätssicherung beschrieben, dabei wird eine Zunahme der Anzahl beschriebener Maßnahmen mit

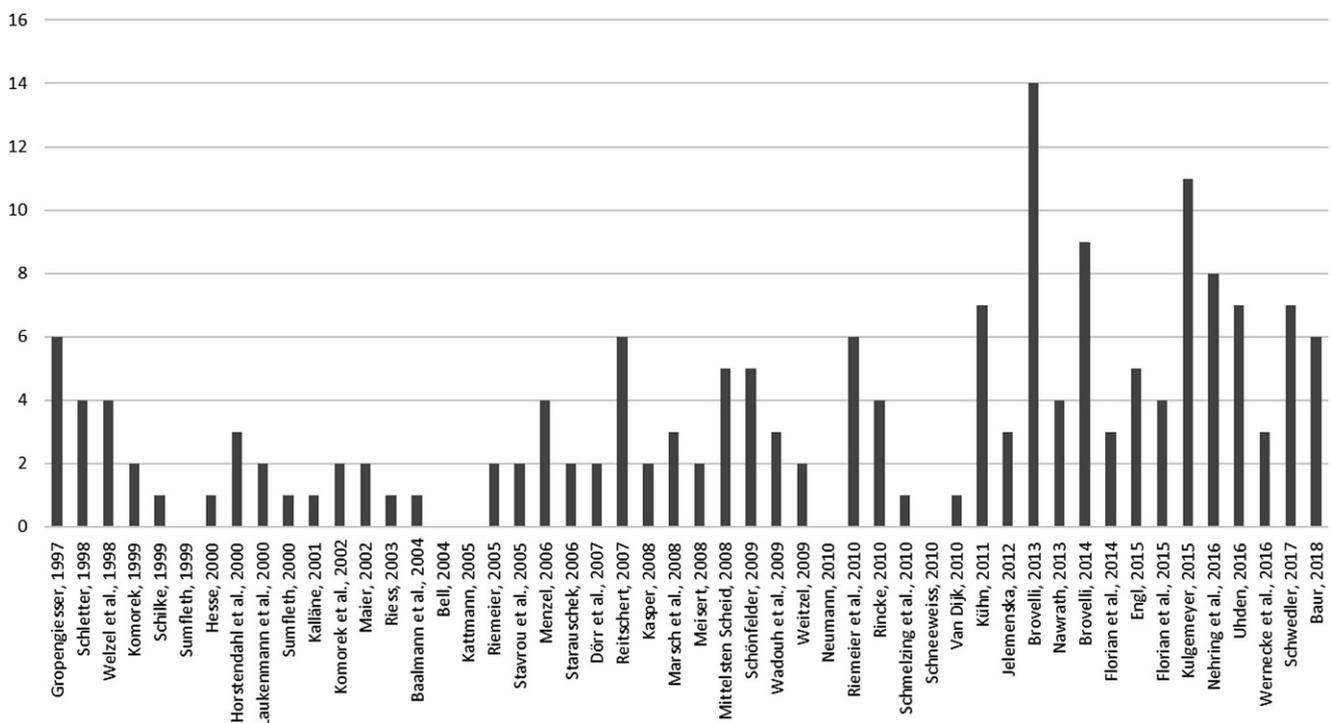


Abb. 1 Anzahl beschriebener Maßnahmen je untersuchtem Artikel

**Tab. 3** Beschreibung qualitätssichernder Maßnahmen in den 50 ausgewählten Artikeln der ZfDN, mehrere Kodierungen der gleichen Maßnahme werden pro Artikel nur einmal gezählt

Maßnahme der Qualitätssicherung	Artikelanzahl
Induktive Bildung des Kategoriensystems	29
Deduktive Bildung des Kategoriensystems	28
Prüfung der Interrater-Übereinstimmung	21
Dokumentation	12
Systematisches Vorgehen	12
Theoretisches Sampling	12
Kommunikative Validierung	10
Triangulation	10
Diskussion der Kodierungen	9
Diskussion des Kategoriensystems	9
Konsensbildung	7
Argumentative Interpretationsabsicherung	5
Formale Aspekte des Kategoriensystems	5
Training der BeurteilerInnen	4
Auswahl/Standardisierung des Settings	1
Prüfung der Intrarater-Übereinstimmung	0

der Zeit deutlich (Abb. 1), wobei vor allem Gropengießer (1997) und Brovelli et al. (2013) relativ umfangreich Maßnahmen entgegen diesem Trend beschreiben. Es ergibt sich ein signifikant positiver Zusammenhang (Pearson Korrelation) zwischen dem Publikationsjahr und der Zahl der durchschnittlich beschriebenen Maßnahmen zur Qualitätssicherung ( $N=50$ ;  $r=0,51$ ;  $p<0,001$ ; großer Effekt).

Die beschriebenen Maßnahmen der Qualitätssicherung (Tab. 3) beziehen sich am häufigsten auf den Modus der Kategorienbildung (vgl. Stamann et al. 2016). In acht Artikeln (16%) wird eine rein induktive Bildung des Kategoriensystems beschrieben. Sieben Artikel (14%) beschreiben eine rein deduktive Bildung des Kategoriensystems. Der Großteil der Artikel (21 Artikel, 42%) kombiniert eine induktive und deduktive Bildung des Kategoriensystems.

Die nächsthäufige Maßnahme der Qualitätssicherung der qualitativen Inhaltsanalyse in den untersuchten Artikeln ist die Prüfung der Interrater-Übereinstimmung (21 Artikel, 42%). Zu den hierbei berechneten Maßen gehören Cohens Kappa (z. B. Engl et al. 2015) sowie die prozentuale Beurteilerübereinstimmung (z. B. Starauschek 2006). Eine Prüfung der Intrarater-Übereinstimmung wird in keinem der untersuchten Artikel beschrieben.

Relativ häufig werden auch die von Mayring (2002) vorgeschlagenen qualitätssichernden Maßnahmen<sup>2</sup> Verfahrensdokumentation (12 Artikel, 24%), argumentative Interpretationsabsicherung (5 Artikel, 10%), systematisches Vorgehen (12 Artikel, 24%), kommunikative Validierung (10 Ar-

tikel, 20%) und Triangulation (10 Artikel, 20%) beschrieben. Zu den expliziten Beschreibungen der Verfahrensdokumentation gehören die Entwicklungsprozesse von Kodierregeln (z. B. Reitschert und Höhle 2007) oder Kodiermanualen (z. B. Marsch et al. 2009) sowie die Begründung der Form der Datenaufnahme (z. B. Riemeier 2005) und deren Limitationen (z. B. Jelemenská 2012). Argumentative Interpretationsabsicherung wird von verschiedenen AutorInnen genannt, aber nicht näher ausgeführt (z. B. Schönfelder und Bögeholz 2009). Systematisches Vorgehen wird entweder als Grundvoraussetzung für die Datenauswertung (z. B. Mittelsten Scheid und Höhle 2008) beschrieben oder begründet Teile des Vorgehens wie die Durchführung einer Pilotstudie (z. B. Mittelsten Scheid und Höhle 2008) oder andere Maßnahmen der Qualitätssicherung wie eine kommunikative Validierung oder die Diskussion der Kodierungen (z. B. Riemeier et al. 2010). Kommunikative Validierungen und Triangulationen werden jeweils in zehn Artikeln (20%) beschrieben. Zur kommunikativen Validierung werden zumeist retrospektiv Interviews mit ProbandInnen genutzt (z. B. Kulgemeyer und Tomczyszyn 2015), zum Teil auch in Verbindung mit Struktur-lege-Techniken (z. B. Jelemenská 2012). Zur Triangulation werden sowohl verschiedene Methoden der Datenerhebung und -auswertung (z. B. Gropengießer 1997) als auch Daten verschiedener Erhebungszeitpunkte genutzt (z. B. Schwedler 2017).

Hinsichtlich diskursiver Maßnahmen der Qualitätssicherung können die interne, zwischen den BeurteilerInnen stattfindende Diskussion der Kodierungen (Kulgemeyer und Tomczyszyn 2015) sowie die Diskussion des Kategoriensystems mit mehr oder weniger externen ExpertInnen (Brovelli et al. 2013) unterschieden werden. Beide Maßnahmen werden in jeweils neun Artikeln (18%) beschrieben. Weniger oft werden dagegen Prozesse der Konsensbildung (7 Artikel, 14%) oder das Training der BeurteilerInnen erläutert (4 Artikel, 8%). Auch formale Aspekte des Kategoriensystems (5 Artikel, 10%) wie beispielsweise eine vorgegebene Länge der Kodiereinheiten, sowie die Standardisierung des Settings (1 Artikel, 2%) werden selten beschrieben.

### Zuordnung von Gütekriterien und Maßnahmen der Qualitätssicherung (Frage 3)

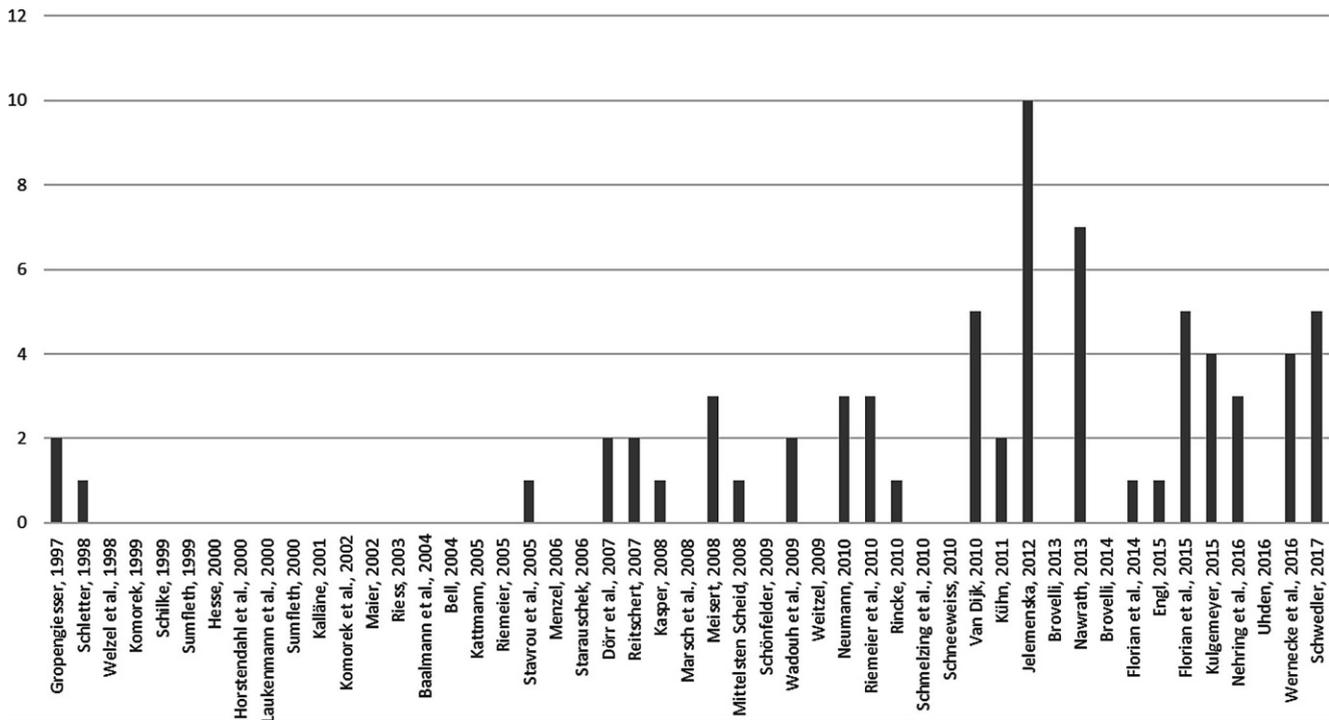
Die 174 insgesamt beschriebenen Maßnahmen der Qualitätssicherung werden zu einem Großteil (74 Maßnahmen, 43%) explizit Gütekriterien zugeordnet, allerdings unterscheiden sich die zugeordneten Gütekriterien in den betrachteten Artikeln. Keine der betrachteten Maßnahmen wird konsistent über mehrere Artikel hinweg einem einzigen Gütekriterium konsistent zugeordnet (Tab. 4).

Die Interrater-Übereinstimmung wird der Validität (in 4 Artikeln), der Reliabilität (in 7 Artikeln) sowie der Ob-

<sup>2</sup> Von Mayring (2002) als Gütekriterien der qualitativen Forschung bezeichnet.

**Tab. 4** Kreuztabelle der zusammenhängend beschriebenen Gütekriterien und qualitätssichernden Maßnahmen. Zeilen (Maßnahmen der Qualitätssicherung) und Spalten (Gütekriterien) sind absteigend nach ihrer Häufigkeit sortiert

	Validität/ Gültigkeit	Reliabilität	Objektivität/ Intersubjektivität	Nähe zum Gegenstand/ Authentizität/ Naturalizität	Trans- parenz	Praktikabilität/ Ökonomie	Offenheit	Praktische Relevanz	Ethische Überlegun- gen	Kommunika- tivität	Interpreta- tivität	Summe
<i>Prüfung der Interrater- Übereinstimmung</i>	4	7	6	-	-	-	-	-	-	-	-	18
<i>Induktive Bildung des Kategoriensystems</i>	3	3	1	2	-	1	1	-	-	-	-	11
<i>Dokumentation</i>	3	1	2	2	2	-	-	-	-	-	-	10
<i>Systematisches Vorgehen</i>	5	-	1	2	-	1	-	-	-	-	-	9
<i>Triangulation</i>	7	1	-	1	-	-	-	-	-	-	-	9
<i>Diskussion der Kodierun- gen</i>	5	3	1	-	-	-	-	-	-	-	-	9
<i>Diskussion des Kategorien- systems</i>	5	-	1	2	-	-	-	-	-	-	-	8
<i>Kommunikative Validierung</i>	4	-	-	1	1	1	-	-	-	-	-	7
<i>Argumentative Interpretati- onsabsicherung</i>	4	-	-	1	-	-	-	-	-	-	-	5
<i>Deduktive Bildung des Kategoriensystems</i>	2	-	-	1	1	-	1	-	-	-	-	5
<i>Theoretisches Sampling</i>	2	-	-	-	-	1	-	-	-	-	-	3
<i>Training der BeurteilerIn- nen</i>	-	2	1	-	-	-	-	-	-	-	-	3
<i>Formale Aspekte des Kate- goriensystems</i>	-	-	1	1	-	-	-	-	-	-	-	2
<i>Konsensbildung</i>	1	-	1	-	-	-	-	-	-	-	-	2
<i>Auswahl/Standardisierung des Settings</i>	-	-	-	-	-	-	-	-	-	-	-	0
<i>Summe</i>	45	17	15	13	4	4	2	0	0	0	0	-



**Abb. 2** Anzahl beschriebener Maßnahmen je untersuchtem Artikel, die einem Gütekriterium zugeordnet wurden

ektivität (in 6 Artikeln) zugeordnet. Dabei findet eine Zuordnung zu mehreren Gütekriterien teilweise auch innerhalb eines Artikels statt, zum Beispiel schreibt Baur (2018): „Fünfundzwanzig Prozent der Daten (Losprinzip) wurden bei der Kategoriensystementwicklung nochmals von einer zweiten Raterin ausgewertet, um die Objektivität und Reliabilität zu bestimmen.“ Auch die induktive Bildung des Kategoriensystems wird der Validität (in 3 Artikeln), der Reliabilität (in 7 Artikeln) und der Nähe zum Gegenstand (in 2 Artikeln), sowie in jeweils einem Artikel der Objektivität, Praktikabilität und Offenheit zugeordnet. Insgesamt werden die meisten beschriebenen Maßnahmen auch dem am häufigsten beschriebenen Gütekriterium der Validität zugeordnet.

Werden nur diejenigen Maßnahmen berücksichtigt, die explizit einem beliebigen Gütekriterium zugeordnet wurden (Abb. 2), ergibt sich ebenso ein signifikant positiver Zusammenhang zwischen dem Publikationsjahr und der Zahl der durchschnittlich beschriebenen und einem Gütekriterium zugeordneten Maßnahmen ( $N=50$ ;  $r=0,44$ ;  $p<0,01$ ; mittlerer Effekt).

## Diskussion

In der qualitativen Forschung werden Gütekriterien nach wie vor umfangreich diskutiert (z.B. Steinke 2013) und auch für die qualitative Inhaltsanalyse existieren vielfälti-

ge theoretische Vorschläge, Gütekriterien und Maßnahmen der Qualitätssicherung anzulegen (Mayring 2015; Schreier 2012; Lamnek 2010). Die vorliegende Untersuchung zeigt für die publizierten Artikel in der *Zeitschrift für Didaktik der Naturwissenschaften*, dass auch in der praktischen Umsetzung im Laufe der letzten Jahre vermehrt Gütekriterien und Maßnahmen der Qualitätssicherung für die qualitative Inhaltsanalyse beschrieben werden (Abb. 1). Dieser positive Trend könnte auf eine umfassendere Qualitätssicherung der qualitativen Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung hindeuten, die mit der zunehmenden Etablierung der Fachdidaktiken als eigenständige, empirisch arbeitende Disziplinen zusammenhängen kann (Bayrhuber et al. 2011; Leuders 2015). Als Gütekriterien werden dabei, entgegen der theoretischen Vielfalt, vorwiegend Validität, Reliabilität und Objektivität angelegt, die sonst klassischerweise der quantitativen Forschung zugeordnet werden (Tab. 2). Die dominierende Position der Gütekriterien Validität, Reliabilität und Objektivität könnte dabei Hinweis auf die zunehmende Etablierung methodologischer Positionen (Mayring 2015; Schreier 2012) oder eine vermehrte Nutzung integrativer Ansätze sein, die quantitative und qualitative Methoden zusammenführen (Kelle 2008). Die Dominanz der Validität als meistgenanntes Gütekriterium passt außerdem zu aktuellen Entwicklungen in der quantitativen Forschung. Hier wird ein „argument-based approach to validation“ vorgeschlagen (Kane 2013), nach dem durch Berücksichtigung verschiedener Evidenzquellen der argu-

mentative Nachweis erbracht werden soll, dass die intendierten Testwertinterpretationen legitim sind. Dabei werden vielfältige Argumente für die Validität der intendierten Interpretation eines Testwertes geprüft. Hierbei kann nicht von der Validität einer Methode oder eines Tests im Allgemeinen gesprochen werden, diese ist vielmehr in Abhängigkeit des gegebenen Untersuchungskontexts darzulegen (Hartig, Frey und Jude 2012).

Betrachtet man die verschiedenen identifizierten Maßnahmen der Qualitätssicherung ergibt sich auf den ersten Blick ein ebenfalls relativ einheitliches Bild: Die Modi der Kategorienbildung und die Interrater-Übereinstimmung stellen klar die am häufigsten genannten Maßnahmen dar (Tab. 3), was sich auch theoretisch wiederfindet, so weist beispielsweise Mayring (2015) der Interrater-Übereinstimmung eine besondere Bedeutung zu. Auf den zweiten Blick zeigen sich zwischen den Artikeln Unterschiede im Hinblick auf den Zusammenhang zwischen den beschriebenen Maßnahmen der Qualitätssicherung und ihrer Zuordnung zu den Gütekriterien. Hier existiert weder theoretisch (vgl. „Objektivität“ Mayring 2015; Schreier 2012) noch praktisch (Baur 2018; Kulgemeyer und Tomczyszyn 2015; Kühn 2011) ein methodologischer Konsens, der als erprobte Forschungspraxis aufgefasst werden kann (vgl. Helsper et al. 2001). Diese Unterschiede könnten in der qualitativen Inhaltsanalyse selbst begründet sein, die durch ihren Variantenreichtum immer wieder Begriffsbestimmungen und Systematisierungsversuchen unterliegt (Ramsenthaler 2013; Schreier 2014; Stamann et al. 2016). Allerdings findet sich hierfür in den untersuchten Artikeln kaum ein Hinweis, da bis auf den Modus der Kategorienbildung selten klar definiert oder benannt wird, welche Variante der qualitativen Inhaltsanalyse dem jeweiligen Forschungsprojekt zugrunde liegt. Weiterhin könnten diese Unterschiede in der Natur der qualitativen Forschungsprojekte liegen, wobei abgewogen werden muss, in welcher Detailliertheit Methoden, Gütekriterien und Maßnahmen der Qualitätssicherung überhaupt konkretisiert werden können, ohne dass die Offenheit und Flexibilität qualitativer Methoden zu Gunsten von Systematik und Transparenz verloren geht (Denzin 2016). Um eine Diskussion über die Zusammenhänge zwischen spezifischen Gütekriterien und Maßnahmen der Qualitätssicherung zu vermeiden, können letztere auch hier analog zur quantitativen Forschung im Sinne eines „argument-based approach to validation“ (Kane 2013) als Argumente für die Validität der abgeleiteten Interpretationen aufgefasst werden. So kann jede Maßnahme der Qualitätssicherung als eine Evidenzquelle für eine valide Interpretation der Untersuchungsergebnisse aufgefasst werden. Dies bedeutet aber auch, dass ein Zusammenspiel mehrerer Maßnahmen der Qualitätssicherung sowie deren umfassende qualitative Beschreibungen an Bedeutung gewinnen. Gerade in der Anzahl der be-

schriebenen Maßnahmen der Qualitätssicherung und deren qualitativer Beschreibung unterscheiden sich auch die hier untersuchten Artikel, wobei sich die Unterschiede mitunter aus den spezifischen Projekten heraus ergeben. So werden forschungsökonomisch begründet häufig nur Teile der Daten oder sogar nur das Kodiermanual (z. B. Schmelzing et al. 2010) auf Interrater-Übereinstimmung hin überprüft, obwohl kleine Werte der Beurteilerübereinstimmung auch als Indikator und Entscheidungshilfe für weitere qualitätssichernde Maßnahmen genutzt werden können (z. B. Kulgemeyer und Tomczyszyn 2015). Häufig schwer zu begründen sind darüber hinaus die Grenzwerte der Beurteilerübereinstimmung, die sich im Allgemeinen nach der Komplexität des vorliegenden Datenmaterials beziehungsweise des zu erfassenden Merkmals richten (Wirtz und Caspar 2002). Hier kann eine inhaltliche Begründung (z. B. in Nehring et al. 2016) für Nachvollziehbarkeit und Transparenz sorgen. Insgesamt werden Maßnahmen der Qualitätssicherung, die vor und nach der eigentlichen Kodierung stattfinden, selten und wenig ausführlich beschrieben, hierzu gehören besonders diskursive und daher komplex zu dokumentierende Prozesse wie das Training der BeurteilerInnen, das Vergleichen der Kodierungen und die Bildung eines finalen Konsenses. Hierbei spielt natürlich der zeitliche Aufwand eine große Rolle und sollte für jede qualitätssichernde Maßnahme abgewogen werden (Praktikabilität/Ökonomie). Beispielsweise erfordert die Überprüfung der Intrarater-Übereinstimmung, die in keinem der analysierten Artikel durchgeführt wurde (Tab. 3), einen zusätzlichen Kodierdurchgang. Von der naturwissenschaftsdidaktischen Community als besonders hochwertig eingeschätzte Forschungsprojekte können hier aufzeigen, wie detailliert methodenspezifische Gütekriterien beschrieben sein können, um eine flexible, systematische und transparente Evaluation zu ermöglichen.

Natürlich unterliegen die Ergebnisse des vorliegenden Artikels auch einigen Limitationen. So ist die qualitative Inhaltsanalyse nicht nur Gegenstand, sondern auch Methode der vorliegenden Analyse, was auf einer methodologischen Ebene paradox erscheint: Die qualitative Inhaltsanalyse als Forschungsmethode wird auf Gütekriterien und Maßnahmen der Qualitätssicherung hin untersucht, wobei methodisch selbst eine qualitative Inhaltsanalyse genutzt wird, die wiederum die Gütekriterien und Maßnahmen der Qualitätssicherung berücksichtigen soll. In der praktischen Umsetzung konnten so viele Gütekriterien und Maßnahmen der Qualitätssicherung berücksichtigt und reflektiert werden (vgl. Methode), wobei auch hier aus forschungsökonomischen Gründen auf die Überprüfung der Intrarater-Übereinstimmung verzichtet wurde. Darüber hinaus ergeben sich weitere methodische Einschränkungen. Die *Zeitschrift für Didaktik der Naturwissenschaften* wurde als Datengrundlage ausgewählt, unter der Annahme, dass diese als Ver-

bandszeitschrift der GDGP und FDdB im VBIO eine geeignete Zeitschrift für qualitativ hochwertige und repräsentative naturwissenschaftsdidaktische Forschung im deutschsprachigen Raum darstellt. Zur weiteren Generalisierung der hier beschriebenen Ergebnisse wäre eine Ausweitung auf weitere Zeitschriften nötig; zum Beispiel internationale Journals, Zeitschriften mit einem explizit qualitativen Fokus oder solche, die über die naturwissenschaftsdidaktische Forschung hinausgehen (z. B. *Journal of Research in Science Teaching*, *Zeitschrift für interpretative Schul- und Unterrichtsforschung*). Dabei könnten auch weitere Gütekriterien und Maßnahmen der Qualitätssicherung berücksichtigt werden und die vorliegenden Ergebnisse erweitern, zum Beispiel bezüglich der Datenerhebung. Zusätzlich zu bedenken sind natürlich auch formale und stilistische Einflüsse auf die Beschreibung der Gütekriterien und Maßnahmen der Qualitätssicherung. So ist es denkbar, dass in den jeweiligen Forschungsprojekten durchaus Gütekriterien angelegt und Maßnahmen der Qualitätssicherung durchgeführt worden sind, diese dann im publizierten Artikel aber nicht beschrieben wurden.

## Fazit

Die vorliegende Analyse gibt einen Überblick über die in der deutschsprachigen naturwissenschaftsdidaktischen Forschung etablierten Gütekriterien und Maßnahmen zur Qualitätssicherung bei der Umsetzung von Verfahren der qualitativen Inhaltsanalyse. Dabei können die hier beschriebenen Ergebnisse WissenschaftlerInnen bereits als Ansatzpunkt für Gütekriterien und Maßnahmen der Qualitätssicherung dienen, die an die eigene Forschung angelegt werden können. Zugleich sollen die vorliegenden Ergebnisse aber vor allem zu Diskussionen in den Fachdidaktiken anregen und so zur weiteren Etablierung und Systematisierung des Verfahrens der qualitativen Inhaltsanalyse in der naturwissenschaftsdidaktischen Forschung beitragen. Die Güte qualitativ inhaltsanalytischer Forschung kann im Sinne des „argument-based approach to validation“ (Kane 2013) im Einzelfall und unter Berücksichtigung der intendierten Interpretation von Forschungsergebnissen bewertet werden. Hierdurch zeigt sich die Umsetzung von Maßnahmen zur Qualitätssicherung nicht als Routineverfahren, für das allgemeingültige Regeln vorgeschrieben werden könnten (Hartig et al. 2012). Vielmehr sollte Gütesicherung als ein theoriegeleiteter und komplexer Prozess verstanden werden, bei dem in Abhängigkeit des Forschungskontexts begründet und entschieden werden muss, welche Maßnahmen zur Qualitätssicherung Evidenz für die intendierte Interpretation von Daten liefern können. Neben der Erarbeitung von Gütekriterien für die externe Begutachtung (vgl. Helsper et al. 2016) wird damit einmal mehr die Forderung nach ei-

ner umfassenden und gut zugänglichen Sammlung erprobter Forschungspraxen unterstützt (Helsper et al. 2001), die für WissenschaftlerInnen auch bei der Planung und Umsetzung von Forschungsvorhaben in den Fachdidaktiken leitend sein kann.

**Förderung** Die Autoren danken der Deutschen Forschungsgemeinschaft für die finanzielle Unterstützung des Projekts TypMoL (Projekt-nummer 327507949).

**Funding** Open Access funding provided by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz 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 Artikel 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.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Anhang A

### Verzeichnis der betrachteten ZfDN-Artikel

- Baalman, W., Frerichs, V., Weitzel, H., Gropengießer, H., & Kattmann, U. (2004). Schülervorstellungen zu Prozessen der Anpassung – Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 7–28.
- Baur, A. (2018). Fehler, Fehlkonzepte und spezifische Vorgehensweisen von Schülerinnen und Schülern beim Experimentieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 24.
- Bell, T. (2004). Komplexe Systeme und Strukturprinzipien der Selbstregulation – Konstruktion grafischer Darstellungen, Transfer und systemisches Denken. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 183–204.
- Brovelli, D. (2014). Integrierte naturwissenschaftliche Lehrerbildung – Entwicklung professioneller Kompetenz bei Lehramtsstudierenden. *Zeitschrift für Didaktik der Naturwissenschaften*, 20(1), 21–32.
- Brovelli, D., Bölsterli, K., Rehm, M., & Wilhelm, M. (2013). Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht: Ein Vignettest

- mit authentisch komplexen Unterrichtssituationen und offenem Antwortformat. *Unterrichtswissenschaft*, 41(4), 306–329.
- Engl, L., Schumacher, S., Sitter, K., Größler, M., Niehaus, E., Rasch, R., . . . Risch, B. (2015). Entwicklung eines Messinstrumentes zur Erfassung der Protokollierfähigkeit – initiiert durch Video-Items. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 223–229.
  - Florian, C., Sandmann, A., & Schmiemann, P. (2014). Modellierung kognitiver Anforderungen schriftlicher Abituraufgaben im Fach Biologie. *Zeitschrift für Didaktik der Naturwissenschaften*, 20(1), 175–189.
  - Florian, C., Schmiemann, P., & Sandmann, A. (2015). Aufgaben im Zentralabitur Biologie – eine kategorien-gestützte Analyse charakteristischer Aufgabenmerkmale schriftlicher Abituraufgaben. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 69–86.
  - Gropengießer, H. (1997). Schülervorstellungen zum Sehen. *Zeitschrift für Didaktik der Naturwissenschaften*, 3.
  - Heran-Dörr, E., Wiesner, H., & Kahlert, J. (2007). Schülerorientierung oder Orientierung an Schülervorstellungen? Wie Lehrkräfte vor und nach einer internetunterstützten Fortbildungsmaßnahme über physikbezogenen Sachunterricht denken. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 162–180.
  - Hesse, M. (2000). Erinnerungen an die Schulzeit – Ein Rückblick auf den erlebten Biologieunterricht junger Erwachsener. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 187–201.
  - Horstendahl, M., Fischer, H. E., & Rolf, R. (2000). Konzeptuelle und motivationale Aspekte der Handlungsregulation von Schülerinnen und Schülern im Experimentalunterricht der Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 7–25.
  - Jelemenská, P. (2012). Lehrervorstellungen zum Lehren und Lernen von Evolution – Eine Fallstudie zum fachdidaktischen Coaching. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 231–261.
  - Kalläne, B., & Schilke, K. (2001). Hier und anderswo – Eine Analyse der Lehrpläne für den Sachunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 7(2001), 538.
  - Kasper, L., & Mikelskis, H. (2008). Lernen aus Dialogen und Geschichten im Physikunterricht – Ergebnisse einer Evaluationsstudie zum Thema Erdmagnetismus. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 7–25.
  - Kattmann, U. (2005). Lernen mit anthropomorphen Vorstellungen? – Ergebnisse von Untersuchungen zur Didaktischen Rekonstruktion in der Biologie. *Zeitschrift für Didaktik der Naturwissenschaften*, 11, 165–174.
  - Komorek, M. (1999). Eine Lernprozeßstudie zum deterministischen Chaos. *Zeitschrift für Didaktik der Naturwissenschaften*, 5(3).
  - Komorek, M., Wendorff, L., & Duit, R. (2002). Expertenbefragung zum Bildungswert der nichtlinearen Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 8(1), 33–51.
  - Kühn, S. M. (2011). Weiterentwicklung der Aufgabekultur im naturwissenschaftlichen Unterricht der gymnasialen Oberstufe und im Abitur. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 35–55.
  - Kulgemeyer, C., & Tomczyszyn, E. (2015). Physik erklären – Messung der Erklärensfähigkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 111–126.
  - Laukenmann, M., Bleicher, M., Fuss, S., Gläser-Zikuda, M., Mayring, P., & Rhöneck, C. (2000). Eine Untersuchung zum Einfluss emotionaler Faktoren auf das Lernen im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 139–155.
  - Maier, U. (2002). Eine qualitative Interviewstudie zum Einfluss des Lehrerverhaltens auf Lernemotionen von Schülern im naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 8, 85–102.
  - Marsch, S., Hartwig, C., & Krüger, D. (2009). Lehren und Lernen im Biologieunterricht: Ein Kategoriensystem zur Beurteilung konstruktivistisch orientierter Lernumgebungen. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 109–130.
  - Meisert, A. (2008). Vom Modellwissen zum Modellverständnis – Elemente einer umfassenden Modellkompetenz und deren Fundierung durch lernerseitige Kriterien zur Klassifikation von Modellen. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 243–261.
  - Menzel, S., & Bögeholz, S. (2006). Vorstellungen und Argumentationsstrukturen von SchülerInnen der elften Jahrgangsstufe zur Biodiversität, deren Gefährdung und Erhaltung. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 199–217.
  - Mittelsten Scheid, N., & Höble, C. (2008). Wie Schüler unter Verwendung syllogistischer Elemente argumentieren: Eine empirische Studie zu Niveaus von Argumentation im naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 145–165.
  - Nawrath, D., & Komorek, M. (2013). Kontextorientierung aus Sicht von Physiklehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 233–257.
  - Nehring, A., Stiller, J., Nowak, K. H., Upmeyer zu Belzen, A., & Tiemann, R. (2016). Naturwissenschaftliche Denk- und Arbeitsweisen im Chemieunterricht – eine modellbasierte Videostudie zu Lerngelegenheiten für den Kompetenzbereich der Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 77–96.
  - Neumann, K., Viering, T., & Fischer, H. (2010). Die Entwicklung physikalischer Kompetenz am Beispiel des

- Energiekonzepts. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 285–298.
- Reitschert, K., & Höbke, C. (2007). Wie Schüler ethisch bewerten: Eine qualitative Untersuchung zur Strukturierung und Ausdifferenzierung von Bewertungskompetenz in bioethischen Sachverhalten bei Schülern der Sek. I. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 125–143.
  - Riemeier, T. (2005). Schülervorstellungen von Zellen, Teilung und Wachstum. *Zeitschrift für Didaktik der Naturwissenschaften*, 11(1), 52–72.
  - Riemeier, T., Jankowski, M., Kersten, B., Pach, S., Raabe, I., Sundermeier, S., & Gropengießer, H. (2010). Wo das Blut fließt. Schülervorstellungen zu Blut, Herz und Kreislauf beim Menschen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 77–93.
  - Riess, W. (2003). Die Kluft zwischen Umweltwissen und Umwelthandeln als pädagogische Herausforderung – Entwicklung und Erprobung eines Prozessmodells zum Umwelthandeln in alltäglichen Anforderungssituationen. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 147–159.
  - Rincke, K. (2010). Alltagssprache, Fachsprache und ihre besonderen Bedeutungen für das Lernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16(2010), 235–260.
  - Schilke, K. (1999). Lernvoraussetzungen von Kindern zum Thema Dinosaurier. *Zeitschrift für Didaktik der Naturwissenschaften*, 5(2), 3–14.
  - Schletter, J. C., & Bayrhuber, H. (1998). Lernen und Gedächtnis – Kompartimentalisierung von Schülervorstellungen und wissenschaftlichen Konzepten. *Zeitschrift für Didaktik der Naturwissenschaften*, 4, 3.
  - Schmelzing, S., Wüsten, S., Sandmann, A., & Neuhaus, B. (2010). Fachdidaktisches Wissen und Reflektieren im Querschnitt der Biologielehrerbildung. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 189–207.
  - Schneeweiss, H., & Gropengießer, H. (2010). Schülerkonzepte zu Mikroben. *Zeitschrift für Didaktik der Naturwissenschaften*, 16.
  - Schönfelder, S., & Bögeholz, S. (2009). Bewertungskompetenz in der reflexiven Leitbildarbeit eines Umweltbildungszentrums – Ein Beitrag zur Professionalisierung des pädagogischen Personals. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 247–263.
  - Schwedler, S. (2017). Was überfordert Chemiestudierende zu Studienbeginn? *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 165–179.
  - Starauschek, E. (2006). Der Einfluss von Textkohäsion und gegenständlichen externen piktoralen Repräsentationen auf die Verständlichkeit von Texten zum Physiklernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 127–157.
  - Stavrou, D., Komorek, M., & Duit, R. (2005). Didaktische Rekonstruktion des Zusammenspiels von Zufall und Gesetzmäßigkeit in der nichtlinearen Dynamik. *Zeitschrift für Didaktik der Naturwissenschaften*, 11, 147–164.
  - Sumfleth, E., & Kleine, E. (1999). Analogien im Chemieunterricht – eine Fallstudie am Beispiel des „Balls der einsamen Herzen“. *Zeitschrift für Didaktik der Naturwissenschaften*, 5(3), 39–56.
  - Sumfleth, E., & Telgenbüscher, L. (2000). Zum Einfluss von Bildmerkmalen und Fragen zum Bild beim Chemielernen mit Hilfe von Bildern – Beispiel Massenspektrometrie. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 59–78.
  - Uhdén, O. (2016). Verständnisprobleme von Schülerinnen und Schülern beim Verbinden von Physik und Mathematik. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 13–24.
  - Van Dijk, E. M., & Kattmann, U. (2010). Evolution im Unterricht: Eine Studie über fachdidaktisches Wissen von Lehrerinnen und Lehrern. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 7–21.
  - Wadouh, J., Sandmann, A., & Neuhaus, B. (2009). Vernetzung im Biologieunterricht – deskriptive Befunde einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 69–87.
  - Weitzel, H., & Gropengießer, H. (2009). Vorstellungsentwicklung zur stammesgeschichtlichen Anpassung: Wie man Lernhindernisse verstehen und förderliche Lernangebote machen kann. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 287–305.
  - Welzel, M., Haller, K., Bandiera, M., Hammelev, D., Koumaras, P., Niedderer, H., ... von Aufschneider, S. (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden – Ergebnisse einer europäischen Umfrage. *Zeitschrift für Didaktik der Naturwissenschaften*, 4(1), 29–44.
  - Wernecke, U., Schwanewedel, J., Schütte, K., & Harms, U. (2016). Wie wird Energie im Biologieschulbuch dargestellt? – Entwicklung eines Kategoriensystems und exemplarische Anwendung auf eine Schulbuchreihe. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 215–229.

## Anhang B

### Verzeichnis der referenzierten methodologischen Literatur

- Berelson, B. (1952). *Content analysis in communication research*. New York, NY: Free Press.

- Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. (2. Aufl.) Berlin: Springer.
- Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. *International Journal of Educational Research* 31, 659–671.
- Erickson, F. (1998). *Qualitative research methods for science education*. Dordrecht: Kluwer. 1155–1173.
- Fatke, R. (1997). Fallstudien in der Erziehungswissenschaft. In B. Friebertshäuser & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft*. Weinheim: Juventa. 56–68.
- Flick, U., von Kardorff, E., & Steinke, I. (Hrsg.). (2012). *Qualitative Forschung: Ein Handbuch*. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag.
- Früh, W. (2007). *Inhaltsanalyse: Theorie und Praxis* (6. Aufl.). Konstanz: UVK.
- Gläser, J., & Laudel, G. (2004). *Experteninterviews und qualitative Inhaltsanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gropengießer, H. (2001). *Didaktische Rekonstruktion des „Sehens“: Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung*. Oldenburg: Carl-von-Ossietzky-Universität.
- Gropengießer, H. (2003). *Wie man Vorstellungen der Lerner verstehen kann: Lebenswelten, Denkwelten, Sprechwelten. Beiträge zur Didaktischen Rekonstruktion*. (2. Aufl.). Oldenburg: Didaktisches Zentrum.
- Gropengießer, H. (2005). Qualitative Inhaltsanalyse in der fachdidaktischen Lehr-Lernforschung. In P. Mayring & M. Gläser-Zikuda (Hrsg.), *Die Praxis der Qualitativen Inhaltsanalyse* (2. Aufl.). Weinheim: Beltz. 172–189.
- Haft, H. (1984). Inhaltsanalyse. In H. Haft & H. Kordes (Hrsg.), *Methoden der Erziehungs- und Bildungsforschung*. Stuttgart: Klett-Cotta. 411–418.
- Huber, G. L. (1989). Qualität versus Quantität in der Inhaltsanalyse. In W. Bos & C. Tarnai (Hrsg.), *Angewandte Inhaltsanalyse in Empirischer Pädagogik und Psychologie*. Münster, New York: Waxmann. 32–47.
- Lamnek, S. (1989). *Qualitative Sozialforschung*. München: Psychologie Verlags Union.
- Lamnek, S. (2010). *Qualitative Sozialforschung*. (5. Aufl.). Weinheim: Beltz.
- Lisch, R., & Kriz, J. (1978). *Grundlagen und Modelle der Inhaltsanalyse*. Reinbek: Rowohlt.
- Mayring, P. (1990). *Einführung in die qualitative Sozialforschung*. München: Psychologie Verlags Union.
- Mayring, P. (1990). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Deutscher Studien Verlag.
- Mayring, P. (2010). Qualitative Inhaltsanalyse. In G. Mey & K. Mruck (Eds.), *Handbuch qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag. 601–613.
- Mayring, P., Gläser-Zikuda, M., & Ziegelbauer, S. (2005). Auswertung von Videoaufnahmen mit Hilfe der Qualitativen Inhaltsanalyse – ein Beispiel aus der Unterrichtsforschung. *Medien Pädagogik*.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. (2. Aufl.). Thousand Oaks, CA: SAGE.
- Reinhoffer, B. (2005). Lehrkräfte geben Auskunft über ihren Unterricht. Ein systematisierender Vorschlag zur deduktiven und induktiven Kategorienbildung in der Unterrichtsforschung. In P. Mayring & M. Gläser-Zikuda (Eds.), *Die Praxis der qualitativen Inhaltsanalyse* (2. Aufl.). Weinheim: Beltz Verlag. 123–141.
- Wang, H. A. (1998). Science textbook studies reanalysis: Teachers „friendly“ content analysis methods? Paper presented at the National Association for Research In Science Teaching, San Diego.

# Anhang C

**Tab. C.1** Vollständiges Kategoriensystem

Gütekriterien	Kodierhinweise	Definition	Stichworte	Ankerbeispiel
Validität/ Gültigkeit	Validität/Gültigkeit wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Gültigkeit beinhaltet die Übereinstimmung von Ergebnissen mit dem durch die Untersuchung (Messung) vorgegebenen, theoretisch-begrifflich zu erfassenden.“ (Lisch und Kriz 1978, S. 84) <sup>a</sup>	Validität Gültigkeit Glaubwürdigkeit Validitätsformen (z. B. Augenschein-, Inhalts-, Konstruktvalidität, ...)	„Um eine Aussage zur Güte der Analyse machen zu können, muss abgesichert werden, dass die identifizierten Probleme adäquat analysiert werden, d.h. die Angemessenheit der Bedeutungskonstruktion des Materials muss durch eine Überprüfung verschiedener Expertenurteile validiert werden.“ (Uhlen 2016, S. 17)
Reliabilität	Reliabilität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Zuverlässigkeit beinhaltet die Reproduzierbarkeit von Ergebnissen unter den gleichen intersubjektiven Bedingungen – insbesondere also die Forderung, dass andere Forscher bei Anwendung desselben Erhebungsinstruments in Interaktion mit demselben Untersuchungsgegenstand zu demselben Ergebnis gelangen.“ (Lisch und Kriz 1978, S. 84)	Reliabilität Zuverlässigkeit Verlässlichkeit Trennschärfe	„Fünfundzwanzig Prozent der Daten (Losprinzip) wurden bei der Kategorisierungsystementwicklung nochmals von einer zweiten Katerin ausgewertet, um die Objektivität und Reliabilität zu bestimmen.“ (Baur 2018, S. 6)
Objektivität/ Intersubjektivität	Objektivität/Intersubjektivität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Die Methode soll vom analysierenden Subjekt abgelöst werden, d. h. die Ergebnisse müssen intersubjektiv nachvollziehbar und damit auch reproduzierbar, kommunizierbar und krisisierbar sein.“ (Früh 2007, S. 42)	Objektivität Intersubjektivität Reproduzierbarkeit	„Schließlich wurde die Zuordnung der Antworten zu den von uns formulierten Kategorien durch einen Intersubjektivitätstest innerhalb der Arbeitsgruppe überprüft.“ (Welzel et al. 1998, S. 33)
Transparenz	Transparenz wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Die systematische, offengelegte und damit krisisierbare Vorgehensweise der empirischen Wissenschaft verlangt [die transparente Offenlegung] der Vorstellungen des Forschers, als auch der [s] anvisierte(n) Realitätsausschnitt[en]“ (Früh 2007, S. 20)	Transparenz Offenlegung Dokumentation Nachvollziehbarkeit	„Umstände und Verfahren der Erhebung, der Aufbereitung des Materials und der Auswertung werden nachvollziehbar und detailliert dokumentiert.“ (Gropengießer 1997, S. 76)
Offenheit	Offenheit wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Ggf. wird der Umgang mit unerwarteten Ergebnissen explizit beschrieben	„Qualitative Inhaltsanalyse versucht insoweit diese Offenheit zu praktizieren, als kein vorab entwickeltes inhaltsanalytisches Schema mit Analyseeinheiten, -dimensionen und -kategorien auf die zu untersuchenden Kommunikationsinhalte angelegt wird. Vielmehr versucht sie, den Inhalt selbst sprechen zu lassen und daraus die Analyse zu entfallen.“ (Lamnek 2010, S. 462)	Offenheit Interaktivität	„Dieses Ziel setzt eine subjektbezogene Datenerhebung mit einem offenen und interaktiven Verfahren voraus, wie es mit qualitativen Untersuchungsverfahren vorliegt.“ (Riemer et al. 2010, S. 79)
Nähe zum Gegenstand/ Authentizität/ Naturalisizität	Die Nähe zum Gegenstand wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt	„Das Untersuchungsfeld der qualitativen Sozialforschung „ist die natürliche Welt, die mit naturalistischen Methoden erfasst und beschrieben werden soll“.“ (Lamnek 1989, S. 196)	Nähe zum Gegenstand Authentizität Naturalisizität Berücksichtigung natürlicher Gegebenheiten Vollständigkeit	„In einem zweiten Schritt wurden die Vignetten Lehrkräften und Fachdidaktikern vorgelegt und argumentativ im Blick auf die authentische Abbildung von Unterrichtssituationen validiert.“ (Browelli 2014, S. 27)
Praktikabilität/ Ökonomie	Praktikabilität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Es werden ggf. praktische/ökonomische Vorgaben oder Einschränkungen und ihre Auswirkungen auf die qualitative Inhaltsanalyse beschrieben	„Der pragmatische Sinn jeder Inhaltsanalyse besteht letztlich darin, unter einer bestimmten forschungsleitenden Perspektive Komplexität zu reduzieren. Textmengen werden hinsichtlich theoretisch interessierender Merkmale klassifizierend beschrieben.“ (Früh 2007, S. 44)	Praktikabilität Ökonomie Zeitintensivität Finanzielle Überlegungen	„Die Festlegung der Stichprobengröße ergab sich aus dem zeitintensiven Untersuchungsdesign.“ (Baur 2018, S. 7)
Praktische Relevanz	Die praktische Relevanz wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Die Ergebnisse der qualitativen Inhaltsanalyse werden ggf. als praxisrelevant beschrieben, sodass TeilnehmerInnen oder Öffentlichkeit davon profitieren können	„Even if a study's findings are „valid“ and transferable, we still need to know what the study does for its participants –both researchers and researched– and for its consumers.“ (Miles und Huberman 1994, S. 280)	Praktische Relevanz Berufspraxis Externe Wirkung	„Die dargestellten Fehler, Fehlkonzepte und schülerspezifischen Vorgehensweisen können für die Schulpraxis einen wichtigen Ausgangspunkt darstellen.“ (Baur 2018, S. 11)
Ethische Überlegungen	Ethische Überlegungen werden im Zusammenhang mit der qualitativen Inhaltsanalyse beschrieben. Hierzu gehört bspw. die Werteschatzung der TeilnehmerInnen	„Researchers are obliged ethically to anticipate what will be done in data collection, analysis, and reporting, and to explain to those studied why it will be done that way rather than some other.“ (Erickson 1998, S. 1456)	Ethische Überlegungen Vertrauenswürdigkeit Glaubwürdigkeit Fairness Zumutbarkeit	„Die Technik der interaktiven Tafeln erlaubt es, Tafelvideos synchron zu der Tonaufnahme aufzuzeichnen, so dass die Schüler nicht einer Beobachtungssituation mittels Videokamera ausgesetzt werden mussten. Als Motivation und Gegenleistung für die freiwillige Teilnahme wurde den Schülern ein Gutschein über zwei Nachhilfestunden [...] ausgehändigt, wodurch eine angemessene Gegenleistung geliefert wurde und zudem auch schwächere Schüler zur Teilnahme bewegt werden konnten.“ (Uhlen 2016, S. 16)

Tab. C.1 (Fortsetzung)

Kommunikativität	Kommunikativität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Der Akt der Kommunikation wird als zentral dargestellt und der Güte der qualitativen Inhaltsanalyse zugeordnet	„[...] kommunikative] Verständigung wird bei der Interpretation von Kommunikationsinhalten durch qualitative Inhaltsanalyse besonders wichtig. Um die Realität der jeweiligen Situationen zu erfassen, muss auch die empirische Sozialforschung einen kommunikativen Akt initiieren, der der Alltagskommunikation ähneln.“ (Lamnek 1989, S. 195)	Kommunikativität	„Dieses Ziel setzt eine subjektbezogene Datenerhebung mit einem offenen und interaktiven Verfahren voraus, wie es mit qualitativen Untersuchungs- methoden vorliegt.“ (Riemeier 2010, S. 79)
Interpretativität	Interpretativität wird im Zusammenhang mit der qualitativen Inhaltsanalyse genannt. Es werden Überlegungen oder Methoden dazu geäußert, wie Äußerungen von ProbandInnen interpretiert werden können	„eine wissenschaftlich modifizierte Form des alltagsweltlichen Fremdverstehens [...]“. „[...] aus der naturalistischen, quasi-alltagsweltlichen Untersuchungssituation Handlungsmuster herauszufiltern. Dieser Vorgang besteht aus zwei Phasen: (1) Nachvollzug der alltagsweltlichen Deutungen und Bedeutungs- zuweisungen und (2) typisierende Konstruktion eines Musters.“ (Lamnek 2010, S. 464)	Interpretativität	„Um individuelle Bedeutungsstrukturen zu erheben, ist es notwendig, in einem offenen, interaktiven und problemzentrierten Verfahren interpretativ die Komplexität dieser Bedeutungsstrukturen zu erfassen.“ (Baalman et al. 2004, S. 10)
<b>Maßnahmen der Qualitätssicherung</b>	<b>Kodierhinweise</b> Eine Form der Standardisierung des Settings wird beschrieben und explizit erläutert, welche Relevanz dies für die qualitative Inhaltsanalyse hat. Eine Beschreibung des Settings oder standardisierender Schritte der Datenaufnahme reicht nicht aus	<b>Definition</b> „[man versucht] im subjektiven, inneren Erleben der Befragten vergleichbare Situationen zu erzeugen, indem sich Interviewer, Beobachter usw. individuell auf die untersuchten Personen einstellen. [...] Dahinter steht die Überlegung, dass man unterschiedlichen Probanden Fragestellungen auch unterschiedlich präsentieren muss, um ihnen zu einem vergleichbaren Verständnis der Fragestellung zu verhelfen.“ (Bortz und Döring 1995, S. 302)	<b>Stichworte</b>	<b>Ankerbeispiel</b> „da die Adressaten wegen der Notwendigkeit einer Standardisierung der Erhebungsumgebung zum Zwecke der Testentwicklung vorher geschult werden mussten.“ (Kulgemeyer 2015, S. 124)
Dokumentation	Die Dokumentation der qualitativen Inhaltsanalyse wird auf Metaebene beschrieben und begründet. (z. B. Warum wurde etwas besonders ausführlich/nicht dokumentiert, wie wurden Entscheidungsprozesse dokumentiert?)	„Als eine Möglichkeit dazu erscheint die Forderung, bei Benutzung quantitativer Verfahren die Bedeutungsdimensionen der Daten qualitativ zu erschließen und offenzulegen, bei Benutzung qualitativer Verfahren die Interpretationsprozesse zu systematisieren und zu dokumentieren.“ (Huber 1989, S. 41)	Dokumentation Kodiermanual Datensicherung Schutz der Identität von ProbandInnen	„Die Datenaufnahme erfolgte mithilfe digitaler Videotechnik, um eine eindeutige Sprecheridentifikation zu gewährleisten und über die verbale Sprache hinaus nonverbale Kommunikationselemente wie Gestik und Mimik dokumentieren zu können.“ (Riemeier 2005, S. 46)
Argumentative Interpretationsabsicherung	Es wird beschrieben, dass die Interpretation der Daten einem logisch argumentativen Vorgehen folgt	„Die Interpretationen werden argumentativ abgeleitet. Dies geschieht einerseits in der Explikation, aber auch in den Strukturierungen.“ (Gropengießer 2001, S. 132)	Argumentation Interpretation Regel Koderhinweis	„In einem zweiten Schritt wurden die Vgnetten Lehrkräften und Fachdidaktikern vorgelegt und argumentativ im Blick auf die authentische Abbildung von Unterrichtssituationen validiert.“ (Brovelli 2014, S. 27)
Systematisches Vorgehen	Das systematische Vorgehen der eigenen Studie wird reflektiert (Metaebene) und begründet (Vor-/Nachteile, Gütekriterien)	„Regelteilbarkeit: Das Material wird in Analyseeinheiten zerlegt und schrittweise bearbeitet, einem Ablaufmodell folgend.“ (Gropengießer 2005, S. 10)	Systematisches Vorgehen Pilotstudie mehrere Kodieren regelmäßig gelesenes Vorgehen	„Dazu wurde zunächst ein Teil des Materials gesichtet, zusammengefasst und ein Kategoriensystem induktiv abgeleitet. [...] Es] wurden bereits in der Theorie bekannte Kategorien [...] deduktiv formuliert, dann aber induktiv in größere Kategorien integriert [...]. Anschließend wurde der Rest des Materials regelgeleitet kodiert.“ (Schwedler 2017, S. 170)
Triangulation	Es wird beschrieben, dass die vorliegenden Untersuchungsergebnisse mit anderen Ergebnissen verglichen wurden. Hierbei können andere Methoden auf dieselben Daten oder dieselben Methoden auf andere Daten angewendet werden. Außerdem können Vergleiche mit Untersuchungsergebnissen stattfinden, denen dieselben Annahmen oder Theorien zugrunde liegen	„In der Sozialforschung wird mit dem Begriff <i>Triangulation</i> die Betrachtung eines Forschungsgegenstandes von (mindestens) zwei Punkten bezeichnet.“ (Flick et al. 2012, S. 309)	Triangulation Triangulationsformen (z. B. Datentriangulation, Investigatortriangulation, Theorietriangulation, ...)	„Zum „Erfassen der Lernervorstellungen“ wurden drei Untersuchungen durchgeführt: [...] Erhebung des Stands der Forschung zu empirischen Untersuchungen im Themenbereich Blut und Blutkreislauf und eine anschließende Reinterpretation dieser Ergebnisse, zweitens leitfadenstrukturierte Interviews und drittens Interviewphasen von Vermittlungsexperimenten.“ (Riemeier et al. 2010, S. 79)
Kommunikative Validierung	Es wird beschrieben, dass StudienteilnehmerInnen in die Interpretation miteinbezogen werden (z. B. durch Follow-Up Interviews oder andere Formen der reaktiven Interaktion). Eine Kommunikation zwischen den Auswertenden ist nicht zu kodieren (vgl. Kodierkonferenz)	„Another response to the criticism that traditional qualitative research invites abuses of authority by researchers (including those of self-deception in data identification and analysis) has been for researchers to try to redress the imbalance of power by sharing it more fully with those who are studied. Both participatory action research and practitioner research are attempts to address the power/knowledge issues involved in social research [...]“ (Erickson 1998, S. 1468)	Kommunikative Validierung Interaktives Erfragen Bestätigung von Interpretationen Validierungsinterview	„Die entstandenen Strukturgebilder wurden am nächsten Tag den Kindern zum Zwecke einer kommunikativen Validierung vorgelegt und erläutert. Nachdem die Kinder Korrekturen oder Ergänzungen vor, so wurden diese in der Qualitativen Inhaltsanalyse berücksichtigt.“ (Riess 2003, S. 154)

**Tab. C.1 (Fortsetzung)**

Theoretisches Sampling	Es wird eine gezielte Auswahl von StudienteilnehmerInnen oder Datenquellen beschrieben und begründet, inwieweit diese Auswahl die Qualität der qualitativen Inhaltsanalyse erhöht (z. B. durch die Erhöhung der Wahrscheinlichkeit einer theoretischen Sättigung)	„Die Auswahl der zu untersuchenden Fälle wird hier nicht nach dem Zufallsprinzip, sondern theoretisch gezielte vom Forscher selbst getroffen (theoretisch-systematische Auswahl, bewußte Auswahl, theoretische Stichprobe).“ (Bortz und Döring 1995, S. 310)	Sampling Theoretisches Sampling Datenauswahl	„Um vorunterrichtliche Lernvorstellungen untersuchen zu können, wurde die Stichprobe in allen hier dargestellten Untersuchungen so gewählt, dass die Probanden noch keinen Unterricht zum Thema Blut und Blutkreislauf hatten.“ (Riemeier et al. 2010, S. 79)
Formale Aspekte des Kategoriensystems	Es werden formale Aspekte oder Voraussetzungen des Kategoriensystems beschrieben, hierzu gehören bspw. Trennschärfe, Anzahl oder Beschaffenheit der Kategorien	„[Die] Kodierereinheit legt fest, welches der kleinste Materialbestandteil ist, der ausgewertet werden darf, was der minimale Textteil ist, der einer eine Kategorie fallen kann.“ (Lamnek 1989, S. 213)	Formale Aspekte des Kategoriensystems Analyseeinheit Kodierereinheit Kontexteinheit Kategorienanzahl Kategorienlänge Trennschärfe	„Die Bestimmung der Länge von Events für die Kodierungen wurde über die Struktur von Unterrichtsaktivitäten vorgenommen.“ (Nehring et al. 2016, S. 87)
Induktive Bildung des Kategoriensystems	Es wird beschrieben, dass Kategorien aus den aufgenommenen Daten heraus gebildet werden. Auch eine später im Prozess auftretende datenbasierte Erweiterung oder Ausschärfung des Kategoriensystems wird kodiert	„Induktive Kategorienbildung hat als Grundgedanken, dass die Verfahrenswesen zusammenfassender Inhaltsanalyse genutzt werden, um schrittweise Kategorien aus einem Material zu entwickeln.“ (Flick et al. 2012, S. 472)	Induktiv Datengestützt Ausgehend von Daten	„Aus diesen Antworten haben wir fünf Haupt- und 33 Unterkategorien von Zielen für das Experimentieren gebildet.“ (Weitzel et al. 1998, S. 32)
Deduktive Bildung des Kategoriensystems	Es wird beschrieben, dass Kategorien aus Vorannahmen oder Theorien gebildet werden. Auch eine später im Prozess stattfindende theoriebasierte Umstrukturierung oder Zusammenfassung des Kategoriensystems wird kodiert	„Die aus den theoretischen Vorüberlegungen abgeleiteten Variablen bilden unmittelbar die Grundlage für die Auswertungskategorien. Dadurch wird der Zusammenhang zum existierenden Wissen über den Untersuchungsgegenstand und zu den die Untersuchung strukturierenden theoretischen Vorüberlegungen hergestellt.“ (Gläser und Laudel 2004, S. 198)	Deduktiv Theoriegeleitet Literaturbasiert	„das Datenmaterial wurde nach einem im Vorfeld gebildeten Kategoriensystem strukturiert [...] Die Auswertungskategorien sind aufgrund biologiedidaktischer und fachwissenschaftlicher Überlegungen aufgestellt worden.“ (Hesse 2000, S. 190)
Training der BeurteilerInnen	Es wird beschrieben, dass die BeurteilerInnen geschult werden	„Die Frage ist nun, wann „steht XY sinngemäß da“ und wie weit darf der Codierer die Textinformationen interpretieren? Diese Kompetenz wird bei der Codierschulung geübt und später im Reliabilitätstest überprüft.“ (Früh 2007, S. 52)	Training Schulung	„Zur Sicherung der Qualität des Kategoriensystems wurden verschiedene Maßnahmen durchgeführt, dazu gehörte insbesondere ein intensives, mehrstufiges Training der jeweils zwei fachspezifischen Beurteiler“ (Kühn 2011, S. 42)
Diskussion des Kategoriensystems durch Experten	Es wird beschrieben, dass mehrere Personen das Kategoriensystem (oder mehrere Kategoriensysteme) diskutieren	„Dieses systematische Vorgehen wird ergänzt und abgesichert mit der Kontrolle der Auswertung durch eine andere Person. Zusätzlich begleiten die Mitglieder der Arbeitsgruppe den Prozess kritisch [...]“ (Gropengreber 2001, S. 132)	Diskussion des Kategoriensystems (Experten-)Diskussion	„Mit diesem Ziel wurde die Existenz und die zugehörige analysierende Bezeichnung und Beschreibung der Verständnisprobleme zu mehreren Zeitpunkten der Analyse mit ein oder zwei Physikdidaktikern diskutiert.“ (Uhdén 2016, S. 17)
Diskussion der Kodierungen	Es wird beschrieben, dass mehrere BeurteilerInnen ihre Kodierungen im Einzelfall diskutieren	„Das „klassische“ methodologische Kriterium der Objektivität wird in der Inhaltsanalyse durch Diskursfähigkeit der Ergebnisse gesichert.“ (Haft 1984, S. 418)	Diskussion der Kodierungen (Kodier-)Konferenz Ergebnisdiskussion	„Um eine objektive Vorgehensweise gewährleisten zu können [...] wurde die Auswertung von drei Personen zunächst unabhängig durchgeführt und anschließend gemeinsam diskutiert.“ (Schönfelder 2009, S. 254)
Konsensbildung	Es wird beschrieben, dass die BeurteilerInnen eine gemeinsame Kodierung erarbeiten	„Konsens [...] kann im Verlaufe fachlicher Diskussionen erzielt werden. Eine Konsensbildung in einem heterogenen Forscherteam ist ein stärkeres Indiz für Validität.“ (Bortz und Döring 1995, S. 310)	Konsensbildung Abgleich	„Die Kodierungen und Interpretationen wurden von mindestens zwei Auswertern vorgenommen und Differenzen bis zum Konsens diskutiert.“ (Marsch et al. 2009, S. 116)
Intrater-Übereinstimmung	Es wird beschrieben, dass die Intrater-Übereinstimmung in Form eines gängigen Maßes (bspw. Cohens Kappa) berechnet wurde. Ein Vergleich der Interpretationen allein reicht nicht aus (vgl. Diskussion der Kodierungen)	„Die Intra-Koderreliabilität wird überprüft, indem nach Abschluss der Analyse zumindest Teile des Materials erneut durchgearbeitet werden, ohne auf die zuerst erfolgten Kodierungen zu sehen.“ (Mayring 2010, S. 603)	Intrater-Übereinstimmung Intrater-Reliabilität Zweitkodierung mit zeitlichem Abstand	<i>Kein Beispiel vorhanden</i>
Interrater-Übereinstimmung	Es wird beschrieben, dass die Interrater-Übereinstimmung in Form eines gängigen Maßes (bspw. Cohens Kappa) berechnet wurde. Ein Vergleich der Interpretationen allein reicht nicht aus (vgl. Diskussion der Kodierungen)	„Die Inter-Koderreliabilität [...] wird überprüft, indem zumindest ein Ausschnitt des Materials einem zweiten Kodierer [...] vorgelegt wird.“ (Mayring 2010, S. 604)	Interrater-Übereinstimmung Interrater-Reliabilität Übereinstimmung mehrerer Kodierereinheiten	„die von dem zweiten Rater hinsichtlich der Existenz eines Energiebezugs bewertet wurden. Die Berechnung des Kappa-Koeffizienten als Maß für die Interraterreliabilität ergab $\kappa = 0,726$ .“ (Wernecke et al. 2016)

<sup>a</sup>Die hier aufgeführte Definition von Validität ist abzugrenzen vom im Artikel verwendeten *argument-based approach to validation* nach Kane (2013)

## Literatur

- Baalmann, W., Frerichs, V., Weitzel, H., Gropengießer, H., & Kattmann, U. (2004). Schülervorstellungen zu Prozessen der Anpassung – Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 7–28.
- Barusch, A., Gringeri, C., & George, M. (2011). Rigor in qualitative social work research: a review of strategies used in published articles. *Social Work Research*, 35(1), 11–19.
- Baur, A. (2018). Fehler, Fehlkonzepte und spezifische Vorgehensweisen von Schülerinnen und Schülern beim Experimentieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 24, 115–129.
- Bayrhuber, H., Harms, U., Muszynski, B., Ralle, B., Rothgangel, M., Schön, L., et al. (Hrsg.). (2011). *Empirische Fundierung in den Fachdidaktiken*. Münster: Waxmann.
- Bennett, J., Lubben, F., Hogarth, S., & Campbell, B. (2005). Systematic reviews of research in science education: rigour or rigidity? *International Journal of Science Education*, 27(4), 387–406.
- Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. (2. Aufl.) Berlin: Springer.
- Brovelli, D. (2014). Integrierte naturwissenschaftliche Lehrerbildung – Entwicklung professioneller Kompetenz bei Lehramtsstudierenden. *Zeitschrift für Didaktik der Naturwissenschaften*, 20(1), 21–32.
- Brovelli, D., Bölsterli, K., Rehm, M., & Wilhelm, M. (2013). Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht: Ein Vignettentest mit authentisch komplexen Unterrichtssituationen und offenem Antwortformat. *Unterrichtswissenschaft*, 41(4), 306–329.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2, 637–644.
- Denzin, N. (2016). *The qualitative manifesto: a call to arms*. Taylor and Francis. <https://doi.org/10.4324/9780429449987>.
- Engl, L., Schumacher, S., Sitter, K., Größler, M., Niehaus, E., Rasch, R., et al. (2015). Entwicklung eines Messinstrumentes zur Erfassung der Protokollierfähigkeit – initiiert durch Video-Items. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 223–229.
- Erickson, F. (1998). Qualitative research methods for science education. Dordrecht: Kluwer. 1155–1173.
- Flick, U. (2010). Gütekriterien qualitativer Forschung. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 395–407). Wiesbaden: VS.
- Flick, U., von Kardorff, E., & Steinke, I. (Hrsg.). (2012). *Qualitative Forschung: Ein Handbuch*. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag.
- Florian, C., Schmiemann, P., & Sandmann, A. (2015). Aufgaben im Zentralabitur Biologie – eine kategoriengestützte Analyse charakteristischer Aufgabenmerkmale schriftlicher Abituraufgaben. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 69–86.
- Früh, W. (2007). *Inhaltsanalyse: Theorie und Praxis* (6. Aufl.). Konstanz: UVK.
- Gläser, J., & Laudel, G. (2013). Life with and without coding: two methods for early-stage data analysis in qualitative research aiming at causal explanations. *Forum Qualitative Sozialforschung*, 14. <https://doi.org/10.17169/fqs-14.2.1886>
- Gläser, J., & Laudel, G. (2004). Experteninterviews und qualitative Inhaltsanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gropengießer, H. (1997). Schülervorstellungen zum Sehen. *Zeitschrift für Didaktik der Naturwissenschaften*, 3, 71–87.
- Gropengießer, H. (2001). Didaktische Rekonstruktion des „Sehens“: Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung. Oldenburg: Carl-von-Ossietzky-Universität.
- Gropengießer, H. (2005). Qualitative Inhaltsanalyse in der fachdidaktischen Lehr-Lernforschung. In P. Mayring & M. Gläser-Zikuda (Hrsg.), *Die Praxis der Qualitativen Inhaltsanalyse* (2. Aufl.). Weinheim: Beltz. 172–189.
- Groß, J., Hammann, M., Schmiemann, P., & Zabel, J. (Hrsg.). (2019). *Biologiedidaktische Forschung: Erträge für die Praxis*. Berlin: Springer.
- Grunenberg, H. (2004). Empirische Befunde zur Qualität qualitativer Sozialforschung. Resultate einer Analyse von Zeitschriftenartikeln. In U. Kuckartz, H. Grunenberg & A. Lauterbach (Hrsg.), *Qualitative Datenanalyse: computergestützt: Methodische Hintergründe und Beispiele aus der Forschungspraxis* (S. 65–80). Wiesbaden: VS.
- Haft, H. (1984). Inhaltsanalyse. In H. Haft & H. Kordes (Hrsg.), *Methoden der Erziehungs- und Bildungsforschung*. Stuttgart: Klett-Cotta. 411–418.
- Hartig, J., Frey, A., & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Berlin Heidelberg: Berlin.
- Helsper, W., Herwart-Emden, L., & Terhart, E. (2001). Qualität qualitativer Forschung in der Erziehungswissenschaft. *Zeitschrift für Pädagogik*, 14(2), 251–269.
- Helsper, W., Kelle, H., & Koller, H.-C. (2016). Qualitätskriterien der Begutachtung qualitativer Forschungsvorhaben in der Erziehungswissenschaft. Ergebnisse eines DFG-Roundtable. *Zeitschrift für Pädagogik*, 62(5), 738.
- Hesse, M. (2000). Erinnerungen an die Schulzeit – Ein Rückblick auf den erlebten Biologieunterricht junger Erwachsener. *Zeitschrift für Didaktik der Naturwissenschaften*, 6, 187–201.
- Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: a prolegomenon. *Educational Researcher*, 19(4), 2–9.
- Huber, G. L. (1989). Qualität versus Quantität in der Inhaltsanalyse. In W. Bos & C. Tarnai (Hrsg.), *Angewandte Inhaltsanalyse in Empirischer Pädagogik und Psychologie*. Münster, New York: Waxmann. 32–47.
- Jelemenská, P. (2012). Lehrervorstellungen zum Lehren und Lernen von Evolution – Eine Fallstudie zum fachdidaktischen Coaching. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 231–261.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kasper, L., & Mikelskis, H. (2008). Lernen aus Dialogen und Geschichten im Physikunterricht – Ergebnisse einer Evaluationsstudie zum Thema Erdmagnetismus. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 7–25.
- Kelle, U. (2008). Qualitative vs. quantitative Forschung — die Debatte. In U. Kelle (Hrsg.), *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung: Theoretische Grundlagen und methodologische Konzepte* (2. Aufl., S. 25–55). Wiesbaden: VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-91174-8\\_2](https://doi.org/10.1007/978-3-531-91174-8_2)
- Krippendorff, K. (1980). *Validity in content analysis*
- Krüger, D., & Riemer, T. (2014). Die qualitative Inhaltsanalyse – eine Methode zur Auswertung von Interviews. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 133–145). Berlin, Heidelberg: Springer.
- Krüger, D., Parchmann, I., & Schecker, H. (Hrsg.). (2014). *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin, Heidelberg: Springer.
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz.
- Kühn, S.M. (2011). Weiterentwicklung der Aufgabenkultur im naturwissenschaftlichen Unterricht der gymnasialen Oberstufe und im Abitur. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 35–55.
- Kulgemeyer, C., & Tomczyszyn, E. (2015). Physik erklären – Messung der Erklärfähigkeit angehender Physiklehrkräfte in einer

- simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 111–126.
- Lamnek, S. (2010). *Qualitative Sozialforschung*. Weinheim: Beltz.
- Lamnek, S. (1989). *Qualitative Sozialforschung*. München: Psychologie Verlags Union.
- Leuders, T. (2015). Empirische Forschung in der Fachdidaktik. Eine Herausforderung für die Professionalisierung und die Nachwuchsqualifizierung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 33(2), 215–234.
- Lisch, R., & Kriz, J. (1978). *Grundlagen und Modelle der Inhaltsanalyse*. Reinbek: Rowohlt.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: replication in the education sciences. *Educational Researcher*, 43(6), 304–316.
- Marsch, S., Hartwig, C., & Krüger, D. (2009). Lehren und Lernen im Biologieunterricht: Ein Kategoriensystem zur Beurteilung konstruktivistisch orientierter Lernumgebungen. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 109–130.
- Mayring, P. (2002). *Einführung in die Qualitative Sozialforschung*. Weinheim: Beltz.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Beltz.
- Mayring, P. (2010). Qualitative Inhaltsanalyse. In G. Mey & K. Muck (Hrsg.), *Handbuch qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag. 601–613.
- Mayring, P. (2019). Qualitative Inhaltsanalyse – Abgrenzungen, Spielarten, Weiterentwicklungen. *Forum Qualitative Sozialforschung*, 20(3), 1–15.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. (2. Aufl.). Thousand Oaks, CA: SAGE.
- Mittelsten Scheid, N., & Hößle, C. (2008). Wie Schüler unter Verwendung syllogistischer Elemente argumentieren: Eine empirische Studie zu Niveaus von Argumentation im naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 14(2008), 145–165.
- Nehring, A., Stiller, J., Nowak, K. H., Upmeyer zu Belzen, A., & Tiemann, R. (2016). Naturwissenschaftliche Denk- und Arbeitsweisen im Chemieunterricht – eine modellbasierte Videostudie zu Lerngelegenheiten für den Kompetenzbereich der Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 77–96.
- Patton, M. (2001). *Qualitative research & evaluation methods*. Saint Paul: SAGE.
- Ramsenthaler, C. (2013). Was ist „Qualitative Inhaltsanalyse?“. In M. Schnell, C. Schulz, H. Kolbe, & C. Dunger (Hrsg.), *Der Patient am Lebensende: Eine Qualitative Inhaltsanalyse* (S. 23–42). Wiesbaden: Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-531-19660-2\\_2](https://doi.org/10.1007/978-3-531-19660-2_2)
- Reinisch, B., Helbig, K., & Krüger, D. (Hrsg.). (in Vorbereitung). *Biologiedidaktische Vorstellungsforschung: Zukunftsweisende Praxis*. Berlin: Springer.
- Reitschert, K., & Hößle, C. (2007). Wie Schüler ethisch bewerten: Eine qualitative Untersuchung zur Strukturierung und Ausdifferenzierung von Bewertungskompetenz in bioethischen Sachverhalten bei Schülern der Sek. I. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 125–143.
- Riemeier, T. (2005). Schülervorstellungen von Zellen, Teilung und Wachstum. *Zeitschrift für Didaktik der Naturwissenschaften*, 11(1), 52–72.
- Riemeier, T., Jankowski, M., Kersten, B., Pach, S., Rabe, I., Sundermeier, S., et al. (2010). Wo das Blut fließt. Schülervorstellungen zu Blut, Herz und Kreislauf beim Menschen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 77–93.
- Riemeier, T., Jankowski, M., Kersten, B., Pach, S., Rabe, I., Sundermeier, S., & Gropengießer, H. (2010). Wo das Blut fließt. Schülervorstellungen zu Blut, Herz und Kreislauf beim Menschen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 77–93.
- Riess, W. (2003). Die Kluft zwischen Umweltwissen und Umwelthandeln als pädagogische Herausforderung – Entwicklung und Erprobung eines Prozessmodells zum Umwelthandeln in alltäglichen Anforderungssituationen. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 147–159.
- Sandmann, A., & Schmiemann, P. (Hrsg.). (2016). *Biologiedidaktische Forschung: Schwerpunkte und Forschungsgegenstände*. Berlin: Logos.
- Schmelzing, S., Wüsten, S., Sandmann, A., & Neuhaus, B. (2010). Fachdidaktisches Wissen und Reflektieren im Querschnitt der Biologielehrerbildung. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 189–207.
- Schönfelder, S., & Bögeholz, S. (2009). Bewertungskompetenz in der reflexiven Leitbildarbeit eines Umweltbildungszentrums – Ein Beitrag zur Professionalisierung des pädagogischen Personals. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 247–263.
- Schreier, M. (2012). *Qualitative content analysis in practice*. London: SAGE.
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *Forum qualitative Sozialforschung*, 15(1), 1–27.
- Schwedler, S. (2017). Was überfordert Chemiestudierende zu Studienbeginn? *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 165–179.
- Seale, C. (2002). Quality issues in qualitative inquiry. *Qualitative Social Work*, 1(1), 97–110.
- Stamann, C., Janssen, M., & Schreier, M. (2016). Qualitative Inhaltsanalyse – Versuch einer Begriffsbestimmung und Systematisierung. *Forum qualitative Sozialforschung*, 17, 1–16.
- Staraschek, E. (2006). Der Einfluss von Textkohäsion und gegenständlichen externen piktoralen Repräsentationen auf die Verständlichkeit von Texten zum Physiklernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 127–157.
- Steinke, I. (2013). Gütekriterien qualitativer Forschung. In U. Flick, E. von Kardorff & I. Steinke (Hrsg.), *Qualitative Forschung: Ein Handbuch*. Reinbek bei Hamburg: Rowohlt.
- Uhden, O. (2016). Verständnisprobleme von Schülerinnen und Schülern beim Verbinden von Physik und Mathematik. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 13–24.
- Weitzel, H., & Gropengießer, H. (2009). Vorstellungsentwicklung zur stammesgeschichtlichen Anpassung: Wie man Lernhindernisse verstehen und förderliche Lernangebote machen kann. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 287–305.
- Welzel, M., Haller, K., Bandiera, M., Hammelev, D., Koumaras, P., Niedderer, H., Paulsen, Albert; Robinault, Karine; & von Aufschneider, S. (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden – Ergebnisse einer europäischen Umfrage. *Zeitschrift für Didaktik der Naturwissenschaften*, 4(1), 29–44.
- Wernecke, U., Schwanewedel, J., Schütte, K., & Harms, U. (2016). Wie wird Energie im Biologieschulbuch dargestellt? – Entwicklung eines Kategoriensystems und exemplarische Anwendung auf eine Schulbuchreihe. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 215–229.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.