

## Qualität des Hochschulstudiums: Messkonzepte und Datenquellen im Vergleich

Hillmert, Steffen; Kopecny, Silvia; Reimer, Maike

Postprint / Postprint

Sammelwerksbeitrag / collection article

### Empfohlene Zitierung / Suggested Citation:

Hillmert, S., Kopecny, S., & Reimer, M. (2022). Qualität des Hochschulstudiums: Messkonzepte und Datenquellen im Vergleich. In G. Brandt, & S. d. Vogel (Hrsg.), *Survey-Methoden in der Hochschulforschung* (S. 235-267). Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-36921-7\\_9](https://doi.org/10.1007/978-3-658-36921-7_9)

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**gesis**  
Leibniz-Institut  
für Sozialwissenschaften

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

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

Mitglied der  
  
Leibniz-Gemeinschaft

# Qualität des Hochschulstudiums: Messkonzepte und Datenquellen im Vergleich

Steffen Hillmert, Silvia Kopecny und Maike Reimer

## Zusammenfassung

Merkmale der akademischen Studienqualität werden für verschiedene Zwecke empirisch erhoben. Oft werden individuelle Bewertungen aus Befragungen aggregiert und als Merkmale zur Charakterisierung von Studiengängen oder anderen institutionellen Einheiten bis hin zu ganzen Hochschulen verwendet. Dieser Beitrag widmet sich der Frage, ob, wie und mit welchen Konsequenzen solche Aggregationen vorgenommen werden können und inwieweit aggregierte Informationen aus anderen Datenquellen die vorhandenen Befragungsdaten ergänzen können. Hierzu werden Bewertungen von zwei zentralen Aspekten der Studienqualität – Qualität der Betreuung und Studienorganisation – sowie die Studienzufriedenheit aus einer Absolvent\*innenstudie (BAP) und einem Hochschulranking (CHE) über dieselben Organisationseinheiten einander gegenübergestellt. Zusätzlich werden Informationen aus der amtlichen Hochschulstatistik einbezogen. Analysiert

---

S. Hillmert  
Institut für Soziologie, Universität Tübingen, Tübingen, Deutschland  
E-Mail: [steffen.hillmert@uni-tuebingen.de](mailto:steffen.hillmert@uni-tuebingen.de)

S. Kopecny  
Institut für Soziologie, Universität Tübingen, Tübingen, Deutschland  
E-Mail: [silvia.kopecny@uni-tuebingen.de](mailto:silvia.kopecny@uni-tuebingen.de)

M. Reimer  
Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung, München,  
Deutschland  
E-Mail: [reimer@ihf.bayern.de](mailto:reimer@ihf.bayern.de)

werden die Objektivität individueller Angaben, die Variation der Bewertungen innerhalb institutioneller Kontexte und zwischen ihnen, die direkte Übereinstimmung zwischen unterschiedlichen Datenquellen sowie die Konsistenz von weiteren Ergebnissen, welche auf diesen beruhen. Die Analysen zeigen u. a., dass Individualangaben trotz beträchtlicher Variation verlässliche Mittelwertschätzungen erlauben und die verfügbare Varianz von Kontextmerkmalen stark vom Aggregationsniveau abhängt. Das Ausmaß der Übereinstimmung zwischen Datenquellen wird dabei maßgeblich vom jeweiligen Datentyp bestimmt. Daraus folgende Implikationen für die gängige Praxis des Umgangs mit derartigen Daten werden diskutiert.

### Schlüsselwörter

Studienqualität · Operationalisierung · Datenquellen · Absolvent\*innenstudien · Hochschulrankings

## 1 Einleitung

Spätestens seit der Veröffentlichung von nationalen Hochschulrankings auch in Deutschland in den 1990er Jahren (vgl. etwa SPIEGEL, 1990) ist das Thema akademische Studienqualität Teil der öffentlichen Diskussion geworden. Im Zuge des Bologna-Prozesses entstand von Seiten der Politik ein erhöhtes Interesse an der lokalen Studienpraxis im Zusammenhang mit Fragen der Studierbarkeit und (Re-)Akkreditierung von Studiengängen (vgl. HIS, 2009). Einen empirischen Zugang zur Studienqualität stellen Befragungen von gegenwärtigen und ehemaligen Studierenden dar, in denen diese Aspekte ihres Studiums bewerten. Solche Befragungen sind zum einen, neben der Evaluation einzelner Lehrveranstaltungen, typischer Bestandteil von routinemäßig als Teil der Qualitätssicherung intern durchgeführten *institutionellen Evaluationen*. Weiterhin werden regelmäßig standortübergreifende *Erhebungen von Studienqualität zu wissenschaftlichen Zwecken* durchgeführt, etwa der Studienqualitätsmonitor (SQM; Woisch et al., 2013; DZHW, 2020a), der bundesweit seit 2007 jährlich Studienqualität und Studienbedingungen aus Sicht der Studierenden erfasst. Studienbewertungen werden auch im Rahmen *kommerzieller Hochschulrankings* erhoben, von denen, nach Initiativen mehrerer Wochenzeitschriften (DER SPIEGEL, Focus), aktuell das Ranking des Centrums für Hochschulentwicklung (CHE) in Zusammenarbeit mit der Wochenzeitung DIE ZEIT (CHE, 2020a) wohl

das bekannteste nationale Ranking ist. Ausgewählte Merkmale der Studienqualität sind schließlich auch Teil von *thematisch breiter angelegten Erhebungen* der Hochschulforschung, etwa der National Education Panel Study (Blossfeld et al., 2011), der Absolvent\*innenbefragungen des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (DZHW) (DZHW, 2020b) oder des Bayerischen Absolventenpanels (BAP) (Falk et al., 2007).

Viele Erhebungen dienen primär als evaluative qualitätssichernde Dienstleistung für die einzelnen Hochschulen, Studienbewerber\*innen oder Akteure der Bildungspolitik; entsprechende Merkmale finden aber auch Eingang in die wissenschaftliche Forschung, etwa als Prädiktoren für Studienzufriedenheit (Schiefele & Jacob-Ebbinghaus, 2006), Neigung zum Wechsel der Hochschule (Lörz et al., 2019) sowie Studienabbrüche oder deren Antizipation (Tinto, 1993; Rhee, 2008; Hillmert et al., 2017). Die individuellen Bewertungen der Studienqualität werden dabei oft auf verschiedenen institutionellen Ebenen (z. B. Studiengängen an Hochschulen oder ganzen Fakultäten) aggregiert, etwa als gewichtete oder ungewichtete Mittelwerte einzelner Items oder aus solchen gebildeter Indizes, und diese dienen dann der Charakterisierung der Studienqualität in den jeweiligen Einheiten. Informationen über die betreffenden Aggregate werden auch aus anderen Datenquellen (z. B. weiteren Befragungen oder Sammlungen statistischer Indikatoren) dem primär verwendeten Datensatz zugespielt. In vielen Fällen sind mit den Auswertungen weitreichende Implikationen verbunden: Im Fall von Rankings etwa werden persönliche Studienentscheidungen unterstützt, im Fall von Akkreditierungen mögliche Entscheidungen über die Veränderung oder Fortführung von Studienangeboten. Das Bildungsmonitoring informiert Hochschulentwicklungsmaßnahmen, und im Rahmen der Forschung werden Schlussfolgerungen hinsichtlich bildungswissenschaftlicher Zusammenhänge gezogen. Angesichts dieser Reichweite ist es bemerkenswert, dass über die methodischen Aspekte der betreffenden Daten vergleichsweise wenig bekannt ist. Dies betrifft etwa die Auswahl relevanter Dimensionen, ihre Operationalisierung und Erhebung sowie die Konsequenzen der Aggregation entsprechender Daten.

Im Folgenden formulieren wir vier zentrale methodische Fragen, denen im Zusammenhang mit der Erhebung von Studienqualität und der Verwendung derartiger Daten auf aggregierten Ebenen besondere Relevanz zukommt. Wir werden diese anschließend empirisch anhand dreier in verschiedenen Kontexten erhobenen und verwendeten Datenquellen näher untersuchen.

### **Leitfrage (1) – Objektivität von Bewertungen: Inwieweit sind Einschätzungen der Studienqualität vor allem individuelle Phänomene?**

Die erste Leitfrage richtet sich auf die grundlegende Problematik, inwieweit die individuellen Bewertungen von Studierenden objektiv vorhandene (Qualitäts-) Merkmale des Studiums übereinstimmend erfassen. Bewertungen und Einschätzungen können sowohl persönliche Erfahrungen darstellen als auch die kollektive Situation repräsentieren. Bereits bei der Erhebung wird allerdings häufig nicht eindeutig angegeben, worauf bzw. auf welche Aggregationseinheit genau die Antwortenden sich beziehen sollen. Darüber hinaus ist das Ausmaß der Konsistenz von individuellen Bewertungen eine empirische Frage.

### **Leitfrage (2) – Heterogenität der Hochschulkontexte: Wie einheitlich sind Hochschul- bzw. Fachkontexte?**

Daran anschließend beschäftigt sich die zweite Leitfrage damit, auf welchen Aggregationsebenen sich die von Befragten wahrgenommenen Kontextbedingungen des Studiums unterscheiden. Dabei stellt sich die Frage nach dem notwendigen bzw. sinnvollen Aggregationsniveau. Wie stark lassen sich die Bewertungskontexte ohne größeren Informationsverlust zusammenfassen, oder anders gefragt: Wie viel Aufwand muss in die Erhebung detaillierterer Informationen investiert werden, um sie als sinnvolle Kontextinformationen verwenden zu können?

### **Leitfrage (3) – Konsistenz unterschiedlicher Datenquellen: Welche analytischen Möglichkeiten bieten alternative bzw. zusätzliche Daten?**

Die dritte Leitfrage betrifft die direkten Übereinstimmungen zwischen Angaben über dieselben Untersuchungseinheiten in unterschiedlichen Datenquellen, etwa Absolvent\*innen- oder Studierendenbefragungen oder auch Daten der amtlichen Statistik. Unterschiedliche Datenquellen bieten Potenziale wechselseitiger Ergänzung. Auch für die Hochschulforschung lassen gerade Restriktionen in der Datenverfügbarkeit und Kostenargumente mögliche Alternativen interessant erscheinen, bei denen Informationen nicht im Rahmen des eigenen Surveys erhoben, sondern in Form externer Daten zugespielt werden. Hierfür wäre aber Voraussetzung, dass unterschiedliche Datenquellen vergleichbare Ergebnisse auf der entsprechenden Aggregationsebene zeigen.

### **Leitfrage (4) – Konsistenz in der Prädiktion: Welche Unterschiede in multivariaten Schätzungen resultieren aus der Verwendung von alternativen Datenquellen?**

Die vierte Leitfrage betrifft die praktischen Implikationen vorhandener oder fehlender Konsistenz. Insbesondere wenn verschiedene Datenquellen nicht

direkt übereinstimmen, bleibt zu fragen, ob dies tatsächlich zu unterschiedlichen inhaltlichen Schlussfolgerungen führt. Die Problematik lässt sich im Zusammenhang mit kriterialer Validität diskutieren, also der Frage, welche (individuellen) Merkmale sich jeweils durch welche Kontextinformationen vorhersagen lassen.

Im nachfolgenden, zweiten Abschnitt dieses Beitrags werden zunächst konzeptuelle Überlegungen zu Studienqualität und ihrer empirischen Erhebung in Befragungen skizziert sowie vorliegende Befunde dargestellt. Im dritten Abschnitt werden die verwendeten Daten und die Analysestrategie vorgestellt, in welcher Individualdaten aus einer Absolvent\*innenstudie auf verschiedenen institutionellen Ebenen aggregiert und systematisch mit Informationen aus einem Hochschulranking und der amtlichen Statistik angereichert werden. Im vierten Abschnitt werden die Ergebnisse der empirischen Analysen präsentiert, und im abschließenden fünften Abschnitt diskutieren wir deren praktische Implikationen, Begrenzungen sowie mögliche weiterführende Ansätze.

## **2 Theoretische und praktische Aspekte der Erhebung von Studienqualität mittels Befragung**

### **2.1 Konzepte der Studienqualität**

In theoretischer Hinsicht existiert kein einheitliches Verständnis des Konzepts *Studienqualität*. Multrus (2013, S. 96) formuliert eine umfassende Definition von Lehr- und Studienqualität als „Effizienz der Vermittlung von fachlich und überfachlich relevanten Kenntnissen und Kompetenzen, die in einer günstigen Lernsituation unter Einbeziehung von Forschung und Praxis, Persönlichkeit, Wissenschaftlichkeit, Berufsbefähigung und bürgerliche Teilhabe ausbildet“. Der Begriff ist mehrdeutig, vielschichtig und seine Nutzung abhängig von den unterschiedlichen Erwartungen und Zielen der betroffenen hochschulinternen Statusgruppen (Studierende, Lehrende, Hochschulen) bzw. externer Stakeholder (Arbeitgeber, Bund/Länder; vgl. HRK, 2020). Vom Versuch einer allgemein verbindlichen Definition wird daher bisweilen grundsätzlich abgeraten (Kempgen, 2020).

Studienqualität wird typischerweise mehrdimensional verstanden (Multrus, 2013) und lässt sich in einen evaluativen Rahmen aus Struktur-, Prozess- und Ergebnisqualität einbetten (Donabedian, 1992; Pohlenz, 2008). Wir orientieren uns an der konzeptuellen Beschreibung von Grotheer und Kerst (2011, S. 3 f.), der zufolge Strukturqualität „alle personellen und materiellen Ressourcen sowie die organisatorischen und gesetzlichen Rahmenbedingungen der Hoch-

schullehre [umfasst]. (...) Prozessqualität beschreibt die Organisation und Verwaltung der Lehr- und Lernprozesse im engeren Sinne. Hierzu gehören z. B. die Koordination der Lehrplanung oder des Prüfungswesens, aber auch die inhaltliche und didaktische Gestaltung von Lehrveranstaltungen oder die Betreuung der Studierenden. Ergebnisqualität wiederum bezieht sich auf die Lehr- und Lernresultate, die vor dem Hintergrund zukünftiger Anforderungen aufgebaut und partiell erst in den späteren Erträgen für die Absolvent\*innen offenbar werden. Kompetenzen und Studienerfolg sowie (erfolgreiche) Berufseinstiege und der weitere berufliche Verbleib sind entsprechende Outputindikatoren“. Die Erfassung von Studienqualität über Befragungen von aktuellen oder ehemaligen Studierenden ist eine verbreitete Praxis, die sich u. a. dadurch rechtfertigt, dass vorhandene Indikatoren nicht für ausreichend fein definierte Einheiten verfügbar sind und dass es gerade Studierende oder Absolvent\*innen sind, die direkte und qualifizierte Rückmeldung geben können.

## **2.2 Methodische Herausforderungen bei der Erhebung und Aggregation von Individualbewertungen**

*Messmodell und Erhebungsmethode:* In der Literatur werden Studierendenurteile als subjektive Wahrnehmungen beschrieben, die von variierenden individuellen Voraussetzungen, Interessen und Zielen beeinflusst werden, aber auch Informationen über die Studienqualität als einem institutionellen Merkmal auf der Ebene der Studiengänge oder Fachbereiche enthalten (Blüthmann, 2012). Die spezifische Verwendung individueller Einschätzungen auf aggregierter Ebene ist davon abhängig, ob man die Angaben primär als individuelle Bewertung der persönlichen Situation oder „stellvertretend“ als Aussage über die kollektiv geteilten Studienbedingungen interpretieren kann (vgl. Chan, 1998). Nach Rottach et al., (2016, S. 537 f.) lassen sich idealtypisch zwei Arten der Erfassung unterscheiden: Nach dem Modell des *direct consensus* beurteilen Studierende Aspekte der Studienqualität bewusst aus ihrem persönlichen Blickwinkel (z. B. „Ich werde durch die Lehrenden in meinem Studiengang gut betreut.“). Eine Interpretation kann dann sowohl auf der individuellen Ebene als auch auf der Aggregatebene (der geteilten Umwelt) sinnvoll sein, dort aber im Sinne von Durchschnittswerten mit einer mehr oder weniger großen Variation innerhalb der Aggregate. Nach dem Modell eines Perspektivwechsels (*referent shift consensus model*) beurteilen Studierende stellvertretend für die betreffende Gruppe die gemeinsam geteilte Lernumwelt (z. B. „In unserem Studiengang sind die Prüfungstermine gut organisiert.“). Bei solchen sogenannten *Klimavariablen* ist nur eine Interpretation

auf der Aggregatebene sinnvoll, und man kann den Autor\*innen zufolge ein relativ hohes Maß an Konsistenz zwischen den individuellen Urteilen erwarten.

*Geeignete Aggregationsebene:* Sinnvolle Aggregationsebenen sind durch einheitliche, geteilte Lernumwelten und soziale Binnenstrukturen gekennzeichnet, wie sie etwa für die Schulforschung durch die Zugehörigkeit von Schüler\*innen zu einer Klasse entstehen (Blüthmann, 2012). Der Autorin zufolge können im Hochschulbereich einzelne Lehrveranstaltungen am ehesten geteilte Lernumwelten darstellen; eine soziale Binnenstruktur hingegen kann sich durch die kurze Laufzeit – in der Regel nur ein Semester – kaum einstellen. Grotheer und Kerst (2011) unterscheiden zwischen der Mikroebene (einzelne Lehrveranstaltung), der Mesoebene (Studiengänge und Fachbereiche) sowie der Makroebene (gesamte Hochschule). Die Autoren gehen davon aus, dass aus Individualurteilen aggregierte Indikatoren am besten auf der Mesoebene gebildet werden können. Rein technisch geht bei einer Aggregation von Merkmalen in der Regel Varianz verloren.

*Kompositionseffekte:* Unterschiede zwischen Aggregateinheiten können teilweise oder ganz auf die unterschiedliche Zusammensetzung der Urteilenden in den jeweiligen Einheiten zurückzuführen sein (vgl. Rottach et al., 2016). Da etwa männliche Studierende tendenziell negativer als Frauen bewerten, kann die Geschlechterverteilung einen Einfluss auf die mittlere Bewertung haben (Blüthmann, 2012). Bedeutsam ist ferner die Tatsache, dass die Urteile von Studierenden sich in Abhängigkeit vom Bewertungszeitpunkt verändern (Pohlenz, 2008). Aggregatunterschiede können daher davon beeinflusst werden, welchen jeweiligen Anteil Studierende niedrigerer und höherer Semester die Befragungen enthalten. Um Kompositionsunterschiede kontrollieren zu können, sind allerdings entsprechende (Individual-)Informationen erforderlich.

*Bewertungen von gegenwärtigen und ehemaligen Studierenden:* Dass schlechte Bewertungen und geringe Zufriedenheit mit einer erhöhten Wahrscheinlichkeit für Studienabbrüche oder Fach- bzw. Hochschulwechsel einhergehen (Rhee, 2008; Hillmert et al., 2017), kann das mittlere Niveau der Bewertungen, deren Variation sowie die Ergebnisse von Zusammenhangsanalysen beeinflussen. Bei Vergleichen zwischen den Bewertungen von Absolvent\*innen und Studierenden ist zu beachten, dass erstere per definitionem ihr Studium erfolgreich abgeschlossen haben.<sup>1</sup> Auch bewerten Absolvent\*innen rückblickend auf der Grundlage gewandelter Maßstäbe und Prioritäten. U. a. werden ihre Urteile

---

<sup>1</sup> Befragungen von Personen, die ihr Studium ohne Abschluss beendet haben (Abbruch oder Fach- bzw. Hochschulwechsel), sind auch aus praktischen Gründen eher selten.



von Erfahrungen beim Eintritt in den Arbeitsmarkt beeinflusst (Duarte et al., 2012; Pike, 1994)<sup>2</sup>, und es besteht die Möglichkeit einer Rationalisierung ex post (Beicht & Ulrich, 2010). Wenige Studien hierzu betrachten allerdings intra-individuelle Veränderungen der Bewertungen, sodass nicht klar ist, in welchem Ausmaß das Phänomen wann auftritt bzw. wie es Aggregatwerte beeinflusst.

## 2.3 Bisherige Befunde

Im Hinblick auf unsere Leitfragen (1) und (2) – *Objektivität von Bewertungen* und *Heterogenität der Hochschulkontexte* – ermöglichen statistische Mehrebenenanalysen eine Zerlegung der Varianz in den individuellen Urteilen in die Anteile, die jeweils auf Unterschiede zwischen und innerhalb von geteilten Kontexten (Aggregatebenen) entfallen. Damit können die Varianzanteile auch zur Bestimmung der Konsistenz der individuellen Angaben auf Ebene der Kontexte herangezogen werden. Rottach et al. (2016) etwa finden Reliabilitätswerte für nach (Master-)Studiengängen aggregierte Studierendenbewertungen zwischen 0,63 und 0,87 (ICC<sub>2</sub>) und bewerten diese als zufriedenstellend bis hoch. Bis zu 16 % der Varianz in Bewertungs- und Zufriedenheitsurteilen entfallen in dieser Studie auf Unterschiede zwischen den Studiengängen – unabhängig davon, ob für die Erhebung das Modell des *direct consensus* oder des *referent shift consensus* gewählt wurde. In ähnlicher Weise zeigen Grotheer und Kerst (2011) mit Fächergruppen an Hochschulen als übergeordneten Einheiten, dass der weitaus größte Teil der Varianz verschiedener Zufriedenheits- und Bewertungsdimensionen auf der Personenebene verortet werden kann. So entfallen an Universitäten nur etwa sieben bis zwölf Prozent der Gesamtvarianz dieser Merkmale auf die verwendete Aggregationsebene, an Hochschulen für angewandte Wissenschaften sogar nur vier bis sechs Prozent. Für Studienzufriedenheit ermittelt Blüthmann (2012) für eine Universität einen noch geringeren Anteil von 3,2 % der Varianz auf der Ebene der Studiengänge.

Im Hinblick auf unsere Leitfragen (3) und (4) – *Konsistenz unterschiedlicher Datenquellen und deren Konsequenzen* – existieren nur wenige Studien, die diese Fragen direkt beantworten. Grotheer und Kerst (2011) führen analoge multivariate Analysen mit zwei ähnlichen Studierenden surveys aus, die sich in

---

<sup>2</sup> Bei Absolvent\*innen stellen sich die Studienzufriedenheit und die Bewertung einzelner Aspekte relativ stabil dar, unabhängig davon, wie lange das Studium zurückliegt (Schmidt & Neßler, 2018).

ihrer Zusammensetzung sowie in der verwendeten Methode unterscheiden. Der Studienqualitätsmonitor umfasst eine bundesweit repräsentative Auswahl von relativ vielen Hochschulen und verwendet Zufriedenheitseinschätzungen, der Konstanzer Survey umfasst eine geringere Anzahl Hochschulen mit jeweils mehr Studierenden und lässt verschiedene Aspekte des Studiums bewerten. Die Ergebnisse zeigen, dass vor allem die deskriptiven Auswertungen – mittlere Tendenzen und Varianzen – unterschiedlich ausfallen, während die Relationen zwischen Hochschularten und Fächergruppen sowie korrelative Zusammenhänge in beiden Datenquellen ähnlich abgebildet werden. Die Autoren präsentieren auch Analysen, in denen individuelle Bewertungen von Lehre und Betreuung mit einigen Angaben aus der amtlichen Statistik auf Hochschul- bzw. Fächergruppenebene in Bezug gesetzt werden (Männeranteil, Studierendenanzahl und objektive Betreuungsrelationen in Fächergruppen); sie finden dort allerdings nur sehr geringe Zusammenhänge.

Zusammenfassend lässt sich anhand der, insgesamt im Umfang eher begrenzten, Literatur feststellen, dass Studierendenurteile theoretisch und empirisch nachweisbar eine Komponente enthalten, welche die Eigenschaften des jeweiligen Aggregats (Studiengang, fachliche Gruppierung oder Hochschule) widerspiegelt. Diese fällt jedoch im Vergleich zu der Komponente auf Personenebene, die sowohl individuelle Bedingungen als auch ggf. strukturelle Heterogenitäten innerhalb des Aggregats abbildet, gering aus. Die meisten Studien beziehen sich auf Studiengänge innerhalb von Hochschulen oder auf sehr grob aggregierte Fächergruppen über mehrere Hochschulen; bislang gibt es kaum systematische Vergleiche zwischen möglichen Aggregationseinheiten. Vergleiche zwischen mehreren Datenquellen liegen ebenfalls nur in geringem Umfang vor, und Ähnliches gilt für die Folgen der unterschiedlichen Operationalisierungen für die inhaltlichen Ergebnisse und Schlussfolgerungen.

### **3 Daten und Analysekonzept**

Um die Befundlage zu den vier Leitfragen zu verbessern, haben wir für diesen Beitrag eine ausgewählte Absolvent\*innenstudie mit Daten eines Hochschulrankings auf Basis von Studierendenbefragungen und Informationen aus der amtlichen Statistik angereichert. Der Datensatz ermöglicht eine systematische Auswertung der Folgen von unterschiedlich detaillierten Aggregationseinheiten für deskriptive Vergleiche und multivariate Analysemodelle auf Basis einer Vielzahl von Fächern und Hochschulen. Durch die Kombination von drei, auch im

Studiendesign, unterschiedlichen Datenquellen können Gemeinsamkeiten und Unterschiede aufgezeigt werden, und es kann das Potenzial für gegenseitige Validierung und Substituierung abgeschätzt werden.

### 3.1 Datenquellen und Analysesample

Die erste Datenquelle ist das *Bayerische Absolventenpanel* (BAP), eine repräsentative bayernweite Absolvent\*innenbefragung aller staatlichen Universitäten und Hochschulen für angewandte Wissenschaften (siehe Falk et al., 2007). Die BAP-Daten wurden als primäre Quelle gewählt, da sie eine gute Ausgangsbasis für die Bildung des Analysedatensatzes darstellen: Sie enthalten zahlreiche Hochschulen mit vergleichsweise hohen Fallzahlen pro Hochschule, und die detaillierten Angaben zum Studium erlauben die Zuordnung zu verschiedenen, hierarchisch geordneten Aggregationsebenen (Studiengänge, Fächer, Hochschulen). Zum abgeschlossenen Studium wurden im BAP drei Arten von wertenden Rückmeldungen erhoben: Bewertungen zu Aspekten der Studienqualität, die Gesamtzufriedenheit mit dem Studium sowie die Frage, ob die Befragten den jeweiligen Studiengang weiterempfehlen. Wir verwenden Daten aus der ersten Befragung der Abschlusskohorte 2013/2014 (Banschbach & Falk, 2017), die zwischen Dezember 2015 und April 2016 erhoben wurden. Es wurden alle 54.278 Absolvent\*innen der Grundgesamtheit eingeladen; 17.874 nahmen an der sowohl als Onlinebefragung als auch – optional – als Paper–Pencil durchgeführten Befragung teil (Rücklaufquote an Universitäten: 29,9 %; an Hochschulen für angewandte Wissenschaften: 32,9 %). Unter den Teilnehmenden sind Männer und Personen ohne deutsche Staatsbürgerschaft leicht unterrepräsentiert und geringfügig mehr Personen mit guten Abschlussnoten vertreten (vgl. Falk et al., 2017).

Die zweite Datenquelle sind Informationen aus dem *Hochschulranking des Centrums für Hochschulentwicklung* (CHE) zu den in den BAP-Daten enthaltenen Einheiten (zum aktuellen Ranking vgl. CHE (2020a), zur Methodik CHE (2020b)). Das CHE-Ranking verwendet ein im Dreijahresrhythmus rotierendes Stichprobendesign, in dem jedes Jahr jeweils nur ein Teil des Fächerspektrums in die aktuelle Erhebung einbezogen wird. Um das gesamte Fächerspektrum abzudecken, müssen daher mehrere Jahrgänge herangezogen werden. Da die CHE-Daten auf Stichproben von aktuell Studierenden ab dem dritten Semester bis zur Regelstudienzeit plus zwei Semester basieren, werden von uns die Jahrgänge 2012 bis 2014 herangezogen, um eine möglichst gute Entsprechung mit dem betreffenden Absolvent\*innenjahrgang des BAP in mög-

lichst vielen Studienfächern zu erhalten.<sup>3</sup> In der CHE-Studierendenbefragung werden subjektive Bewertungen der Studiensituation erhoben, die aber nicht als Individualdaten, sondern nur in aggregierter Form zugänglich sind. Aktuell werden allerdings im Rahmen des Methodenberichts – in grafischer Form – standortspezifische Konfidenzintervalle (offensichtlich auf dem 95 %-Niveau) sowie die zugehörigen Rücklaufzahlen ausgewiesen (CHE, 2020b). Diese machen bereits das oft hohe Ausmaß an Unsicherheit in den Mittelwertangaben deutlich.

Die dritte Datenquelle stellt die *amtliche Hochschulstatistik* für das Wintersemester 2013/2014 (Studierendenstatistik) bzw. das Jahr 2014 dar (Prüfungs- und Personalstatistik: ICEland, 2020a, 2020b; Statistisches Bundesamt, 2020). Diese liefert objektivierende Information über die Studienbedingungen, da sie nicht auf Befragungen beruht, sondern auf Sekundärdaten, die von Hochschulverwaltungen und Prüfungsämtern auf Grundlage vorhandener Unterlagen an die Statistischen Landesämter übermittelt werden. Es handelt sich um Vollerhebungen, und die Angaben sind gemäß dem Hochschulstatistikgesetz verpflichtend zu melden; Stichproben- und Itemausfälle sind also nicht zu erwarten. Der Zugang erfolgte über das vom DZHW bereitgestellte Informationssystem *ICEland*.

Um die unterschiedlichen Datenquellen zusammenzuführen, wird der Personendatensatz des BAP mit den aggregierten Indikatoren des CHE und der amtlichen Statistik angereichert. Dies erfolgt über die Information zu Studienfach und Hochschule der Befragten und bezieht sich daher nur auf das Studienangebot an bayerischen Hochschulen, das in allen Datenquellen vertreten ist. Die jeweils verwendeten Fächersystematiken wurden vorab soweit wie möglich harmonisiert und auf der jeweils niedrigsten Aggregationsebene (z. B. Studiengang in Form von Studienfach-Hochschul-Kombinationen) als Schlüsselindikator verwendet.<sup>4</sup> Das Fächerspektrum orientiert sich u. a. an den im CHE-Ranking enthaltenen Studienfächern und ist entsprechend eingeschränkt.<sup>5</sup> Wir beschränken

---

<sup>3</sup> Die entsprechenden Daten wurden uns freundlicherweise vom CHE bereitgestellt.

<sup>4</sup> Den verwendeten Fächerschlüssel stellen wir auf Anfrage gerne zur Verfügung.

<sup>5</sup> Zusätzlich ist zu berücksichtigen, dass im CHE-Ranking keine Ergebnisse für Studiengänge ausgewiesen werden, deren Rücklauf  $N < 15$  war. Kleine Studiengänge oder solche mit geringer Rücklaufquote im CHE sind daher ebenfalls nicht im Analysedatensatz repräsentiert. Bei zusätzlicher Einschränkung des Analysesamples auf Studiengänge mit  $N \geq 15$  auch im Rücklauf der BAP-Befragung bleiben die nachfolgend berichteten Ergebnisse stabil.

uns auf Bachelorabsolvent\*innen (64 % im BAP-Sample), da die anderen Abschlussarten (Master und Staatsexamen) entweder zu geringe Fallzahlen über die Hochschulen und Studienfächer aufweisen oder in den anderen Datenquellen nicht entsprechend vertreten sind. Nach diesen beiden Schritten verbleiben noch 36 % der Fälle im BAP. Weiterhin werden alle zu verwendenden Variablen um fehlende Werte bereinigt, wobei deren Anteil pro Variable zwischen einem Prozent und zehn Prozent liegt. Zuletzt steht auf Basis des BAP ein Analysesample von 4334 Fällen zur Verfügung, was 24 % aller erhobenen Fälle und 68 % der Bachelorabsolvent\*innen nach Ergänzung der CHE- und Destatis-Variablen entspricht. Eine Übersicht über die Fall- bzw. Kategorienzahlen auf den unterschiedlichen Aggregationsebenen gibt Tab. 1. Die Fälle verteilen sich unterschiedlich über die aggregierten Studienkontexte: Mit knapp 50 % der Absolvent\*innen

**Tab. 1** Datensatzstruktur: Hochschulen, Fächeraggregationen, Fallzahlen

	MW	Std.abw	Min	Max
<i>Individualebene/Absolvent*innen (N = 4334)</i>				
<i>Studiengänge (Hochschul-Fach-Kombinationen) (N =159)</i>				
Absolvent*innen	27,26	25,99	1	139
<i>Hochschulen (N=24)</i>				
Absolvent*innen an Hochschulen	180,58	147,98	30	537
Studienfächer an Hochschulen	6,63	4,77	1	16
Studienbereiche an Hochschulen	4,54	3,04	1	11
Fächergruppen an Hochschulen	2,08	0,88	1	4
<i>Studienfächer (N=46)</i>				
Absolvent*innen in Studienfächern	94,22	195,58	3	1290
Hochschulen pro Studienfach	3,46	3,71	1	21
<i>Studienbereiche (N=21)</i>				
Absolvent*innen in Studienbereichen	206,38	366,94	5	1716
Hochschulen pro Studienbereich	5,19	5,03	1	22
<i>Fächergruppen (N=4)</i>				
Absolvent*innen in Fächergruppen	1083,50	873,24	80	2118
Hochschulen pro Fächergruppe	12,50	9,15	3	23

MW: Arithmetischer Mittelwert; Std.abw: Standardabweichung; Min: Minimum; Max: Maximum.

Quelle: BAP 2013/2014 (1), angereichert mit Indikatoren des CHE-Hochschulrankings 2012–2014 und der Amtlichen Statistik (Destatis); eigene Berechnungen

ist die Fächergruppe der Rechts-, Wirtschafts- und Sozialwissenschaften am stärksten besetzt, gefolgt von den Ingenieurwissenschaften (30 %), Mathematik und Naturwissenschaften (18 %) und den Geisteswissenschaften mit einem sehr geringen Anteil von zwei Prozent. Die Rangfolge der vier größten Fächergruppen entspricht damit der Grundgesamtheit, wobei sich der Anteil von Mathematik und Naturwissenschaften mit jenem der Grundgesamtheit ziemlich genau deckt, Geisteswissenschaften stark unter- und die anderen beiden etwas überrepräsentiert sind. Von den 21 Studienbereichen stechen vor allem Wirtschaftswissenschaften (40 %) und Maschinenbau (12 %) durch hohe Fallzahlen heraus. Ähnliches gilt für die Studienfächer, bei denen die Betriebswirtschaftslehre (30 %), Maschinenbau (8 %) und Elektrotechnik (6 %) am stärksten besetzt sind.

### 3.2 Merkmale und Messkonzepte der Studienqualität

Um Studienqualität mehrdimensional zu erfassen, werden Betreuung, Organisation und Zufriedenheit untersucht. Dadurch sind in den Befragungsdaten sowohl Aspekte der Prozess- als auch der Ergebnisqualität abgedeckt. Die *Bewertung der Betreuung* stellt eher auf individuelle Erfahrungen ab, die *Bewertung der Studienorganisation* hingegen auf formale Aspekte, welche eine stärkere interne Konsistenz lokaler Bewertungen erwarten lassen (Blüthmann, 2012; Rottach et al., 2016). Die Variablen unterscheiden sich in formaler Hinsicht, da es sich einerseits um einen Vergleich zwischen Indexwert (CHE-Indikator) und Einzelitem (BAP-Indikator) und andererseits um zwei aus mehreren Items gebildete Indexwerte handelt. Zusätzlich wurde als eine *Globalbewertung* die Angabe zur allgemeinen Zufriedenheit mit dem gesamten Studium ausgewählt. Die Spannweite der Merkmale wurde jeweils vereinheitlicht.<sup>6</sup> Eine Beschreibung der empirischen Merkmalsverteilungen sowie die Frageformulierungen und Antwortvorgaben findet sich in Tab.2.

Während die Bewertung der Organisation der Studiengänge in den unterschiedlichen Datenquellen im Mittel vergleichbar ausfällt, werden die Betreuung und die Gesamtsituation bzw. Zufriedenheit anhand der CHE-Indikatoren tendenziell besser bewertet. Auch wird in den BAP-Individualindikatoren die

---

<sup>6</sup> Die ursprünglich auf einer 6-er Skala erhobenen CHE-Indikatoren wurden mithilfe einer linearen Transformation entsprechend der Spannweite einer vergleichbaren 5-er Skala wie im BAP umgerechnet. Dies ist vor allem für die Varianzanalysen relevant, weniger für die Korrelations- und Prädiktionsanalysen.

**Tab. 2** Merkmale zur individuellen Bewertung der Studienqualität in den drei Datenquellen

	N	MW	Var	Std.abw	Min	Max	Variablentyp	Skalierung (wie in Analyse)
<b>BAP (Individualwerte)</b>								1 „sehr schlecht“ bis 5 „sehr gut“
Betreuung <sup>1</sup>	4334	3,67	0,94	0,97	1,00	5,00	Einzelitem	
Organisation <sup>1</sup>	4334	3,90	0,41	0,64	1,20	5,00	Index aus 5 Items	
<i>Zugang</i>	4334	4,08	1,03	1,01	1,00	5,00		
<i>Abstimmung Prüfungen</i>	4334	3,99	0,67	0,82	1,00	5,00		
<i>Transparenz</i>	4334	3,72	0,80	0,90	1,00	5,00		
<i>Koordination</i>	4334	3,94	0,82	0,91	1,00	5,00		
<i>Inhaltliche Abstimmung</i>	4334	3,78	0,70	0,84	1,00	5,00		
Zufriedenheit <sup>2</sup>	4334	3,93	0,76	0,87	1,00	5,00	Einzelfrage	1 „sehr unzufrieden“ bis 5 „sehr zufrieden“
<b>CHE-Ranking</b>								
Betreuung <sup>3</sup>	4334	4,03	0,10	0,31	2,92	4,60	Index aus 6 Items	1 „sehr schlecht“ bis 5“ sehr gut“
Organisation <sup>4</sup>	4334	3,83	0,10	0,31	2,68	4,44	Index aus 7 (8) Items	
Zufriedenheit <sup>5</sup>	4334	4,03	0,17	0,41	2,52	4,76	Einzelfrage	
<b>Destatis</b>								
Betreuungsrelation (in 100)	4334	0,23	0,03	0,17	0,02	1,62		
Studiendauer (in Jahren)	4334	3,70	0,16	0,40	2,90	5,20		

(Fortsetzung)

**Tab. 2** (Fortsetzung)

	N	MW	Var	Std.abw	Min	Max	Variablentyp	Skalierung (wie in Analyse)
<b>Weitere Variablen BAP</b>								
Weiterempfehlung <sup>6</sup>	4334	4,07	1,06	1,03	1,00	5,00	Einzelfrage	1 „nein, auf keinen Fall“ bis 5 „ja, auf jeden Fall“
Note der Hochschulzugangsberechtigung	4334	2,74	0,38	0,62	1,00	4,00		1 „ausreichend“ bis 4 „sehr gut“
Studienabschlussnote	4334	2,89	0,24	0,49	1,20	4,00		

<sup>1</sup> Wie beurteilen Sie die folgenden Aspekte Ihres Studiums? *Fachliche Beratung und Betreuung; Zugang zu Lehrveranstaltungen (verfügbare Plätze); Inhaltliche Abstimmung des Lehrangebots auf Prüfungsanforderungen; Transparenz der Studieninhalte; Anforderungen und Organisation; Zeitliche Abstimmung der Lehrveranstaltungen; Inhaltliche Abstimmung der Lehrveranstaltungen/Module*

<sup>2</sup> Wie zufrieden sind Sie aus heutiger Sicht mit dem Studium an Ihrer Hochschule insgesamt?

<sup>3</sup> Bitte beurteilen Sie die Qualität der Beratung und Betreuung durch das Lehrpersonal in Ihrem Studiengang! *Verhältnis zwischen Studierenden und Lehrenden; Engagement der Lehrenden; Erreichbarkeit der Lehrenden (Sprechstunden, per E-Mail etc.); Informelle Beratung und Betreuung; Besprechung von Klausuren, Hausarbeiten, Referaten etc.; Eignung der Skripte, Unterlagen und Beispielsammlungen*

<sup>4</sup> Bitte beurteilen Sie die folgenden Aspekte der Studierbarkeit Ihres Studienganges, *Studierbarkeit des Curriculums (Möglichkeit, in der Regelstudienzeit abzuschließen); Zugang zu Lehrveranstaltungen (keine Wartezeiten, keine zeitlichen Überschneidungen); Abstimmung des Lehrangebots auf Prüfungsanforderungen; Transparenz des Prüfungssystems; Feedback während des Studiums/ Leistungsrückmeldungen während des Semesters; Inhaltliche Abstimmung der Modulbestandteile aufeinander; Zeitliche Abstimmung der Lehrveranstaltungen (z. B. aus unterschiedlichen Modulen) aufeinander; (Zeitliche Belastung durch Veranstaltungsvor-/nachbereitung und Prüfungen nur 2012)*

<sup>5</sup> Wenn Sie einmal alles zusammen betrachten: Wie beurteilen Sie insgesamt die Studiensituation in Ihrem Studiengang an Ihrer derzeitigen Hochschule?

<sup>6</sup> Würden Sie jemandem empfehlen, Ihr Fach an Ihrer Hochschule zu studieren?

MW: Arithmetischer Mittelwert; Var: Varianz; Std.abw: Standardabweichung; Min: Minimum; Max: Maximum.

Quelle: BAP 2013/2014 (1), angereichert mit Indikatoren des CHE-Hochschulrankings 2012–2014 und der Amtlichen Statistik (Destatis); eigene Berechnungen



gesamte Spannweite der Skala ausgeschöpft, während in den CHE-Indikatoren die Randwerte nicht besetzt sind; die Varianz dieser Merkmale ist daher geringer. Vergleicht man die drei Indikatorengruppen, sticht insbesondere in den BAP-Daten die deutlich geringere Varianz in den Angaben zur Studienorganisation auf. Diese scheint den höchsten Grad an Objektivität beanspruchen zu können.

Aus den Daten der amtlichen Statistik wurden Indikatoren ausgewählt, die den beiden Bewertungen inhaltlich nahestehen und auch häufig zur Bewertung und Steuerung von Hochschulen herangezogen werden: *Betreuungsrelation* und *Studiendauer*. Mit der *Betreuungsrelation* kommt nun auch ein Indikator der Strukturqualität ins Spiel. Sie definiert sich als das zahlenmäßige Verhältnis von Studierenden zum wissenschaftlichen und künstlerischen Personal in Vollzeit-äquivalenten (Statistisches Bundesamt, 2018a, 2018b). Zur besseren Vergleichbarkeit mit den Effekten anderer Merkmale haben wir die Werte durch die Zahl 100 dividiert; der Mittelwert der *Betreuungsrelation* von 0,23 bedeutet also, dass im Durchschnitt 23 Studierende auf eine Lehrperson in Vollzeit entfallen. Dieser Indikator wurde ausnahmsweise nicht auf Ebene der Studiengänge, sondern für ganze Studienbereiche bzw. Lehr- und Forschungsbereiche pro Hochschule berechnet, da dies in diesem Fall die niedrigste ausgewiesene Ebene ist.<sup>7</sup> Die *Studiendauer* wird als mittlere Anzahl der Fachsemester auf der Aggregations-ebene der Studiengänge bestimmt und in Studienjahren ausgewiesen. Diese Dauer kann als Studierenerfolgsindikator ebenfalls der Ergebnisqualität zugeordnet werden.<sup>8</sup>

Als Kriteriumsvariable für die Abschätzung möglicher inhaltlicher Konsequenzen der Aggregation wird die BAP-Frage ausgewählt, inwieweit die Absolvent\*innen ihren eigenen Studiengang weiterempfehlen würden. In der Literatur, die sich mit Studierendenzufriedenheit aus einer Marketingperspektive beschäftigt und hierzu Modelle aus dem Bereich Service Marketing adaptiert, wird Zufriedenheit als eine Voraussetzung für die Loyalität von Kund\*innen bzw. Konsument\*innen gesehen, also dafür, ob diese den jeweiligen Service wieder in Anspruch nehmen oder weiterempfehlen würden. Auf individueller Ebene bestehen ausgeprägte Zusammenhänge zwischen den Bewertungen und der Zufriedenheit bzw. Loyalität (Duarte et al., 2012; Luque Martínez & Doña Toledo, 2013).

---

<sup>7</sup> Aus Gründen der Verständlichkeit sprechen wir im Folgenden auch bei diesem Indikator von der Ebene der Studiengänge.

<sup>8</sup> Alternativ wurden die Analysen mit dem verwandten Indikator des Anteils der Absolvent\*innen in Regelstudienzeit bzw. Regelstudienzeit plus zwei Semester berechnet. Die Ergebnisse bleiben dabei inhaltlich robust.

### **3.3 Analysestrategie**

Unsere empirische Analysestrategie orientiert sich an den vier eingangs formulierten Leitfragen.

#### **(1) Objektivität von Bewertungen**

Zunächst geht es grundlegend um das Verhältnis von individuellen Bewertungen und den durchschnittlichen Bewertungen in den jeweiligen aggregierten Einheiten („Within-Varianz“). Wir können dies hier auf der Ebene der Studiengänge an Hochschulen untersuchen, da den Absolvent\*innen im BAP darauf bezogene Fragen gestellt wurden. Mithilfe von Random-Intercept-Modellen lassen sich entsprechende „Within“- und „Between“-Varianzanteile bestimmen. Als Maß der Objektivität werden Varianzzerlegungen bzw. Intraklassenkorrelationen (ICCs), insbesondere die ICC\_2 herangezogen (vgl. Bliese, 2000; Rottach et al., 2016).

#### **(2) Heterogenität der Hochschulkontexte**

Im zweiten Schritt wird abgeschätzt, ob hinreichend Variation in den Bewertungen zwischen aggregierten Einheiten („Between-Varianz“) besteht, um eine Verwendung als Kontextmerkmale zu rechtfertigen. Dazu werden Schätzungen dieser Variation über die unterschiedlichen Aggregationsebenen und Studienqualitätsmerkmalen hinweg verglichen. Über die Studiengänge hinaus werden einerseits die gesamte Hochschule, andererseits – unabhängig von der Hochschule – Studienfach, Studienbereich und Fachgruppe sowie ein kreuzklassifiziertes Modell mit Hochschule und Studienfach verwendet. Letzteres unterscheidet sich insofern vom Modell mit dem Studiengang als aggregierter Ebene, als dass die Varianzanteile für beide Dimensionen getrennt ausgewiesen werden können, d. h. die durchschnittliche Varianz zwischen Hochschulen innerhalb von Studienfächern und vice versa (vgl. Rabe-Hesketh & Skrondal, 2012). Für einen weiteren Vergleich werden die Analysen zum einen mit den BAP-Daten anhand der dort erhobenen Studierendenbewertungen durchgeführt, zum anderen mit BAP-Daten, welche mit entsprechenden Angaben zur Studienqualität aus dem CHE-Hochschulranking angereichert wurden.

#### **(3) Konsistenz unterschiedlicher Datenquellen**

Im dritten Schritt geht es um eine direkte Prüfung der Übereinstimmung in der Erfassung von Studienqualität anhand der Indikatoren des BAP und der zugespielten Indikatoren des CHE-Hochschulrankings. Dazu werden Korrelationskoeffizienten der Indikatoren auf den verschiedenen Aggregationsebenen verglichen. Da die BAP-Daten als Individualdaten verfügbar sind, können hier

die Bewertungen sowohl als Individualvariablen als auch in Form von Aggregatmerkmalen verwendet werden; im Fall des CHE ist dies nur in Form der vorgegebenen Aggregationen möglich. Auf der Studiengangsebene und allen höher aggregierten Ebenen liegt in beiden Datenquellen dieselbe Aggregationsebene zugrunde. Zusätzlich werden Zusammenhänge mit den Indikatoren aus der amtlichen Statistik analysiert. Die Aggregationsebenen werden dabei mit der individuellen Fallzahl pro Aggregationseinheit im BAP gewichtet.

#### **(4) Konsistenz in der Prädiktion**

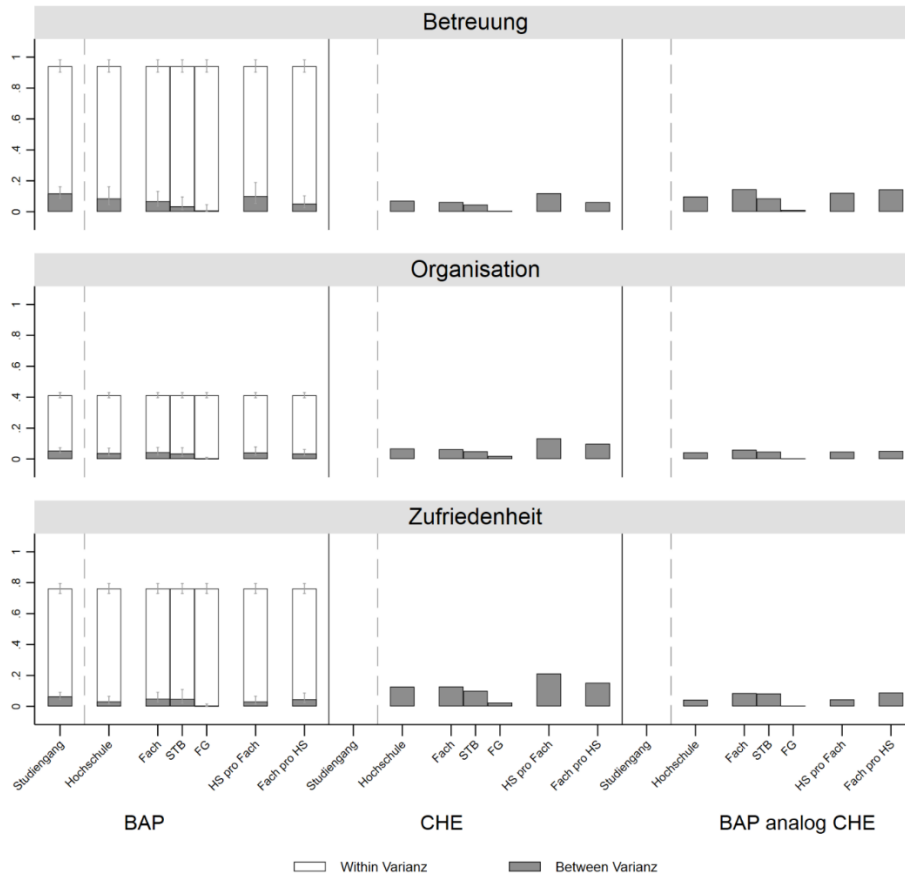
Im letzten Schritt wird untersucht, ob Bewertungen unterschiedlichen Aggregationsgrades bzw. aus unterschiedlicher Quelle zu ähnlichen inhaltlichen Ergebnissen führen, wenn sie Teil von prädiktiven Analysen sind. Hierzu wird erneut auf Random-Intercept-Modelle zurückgegriffen, wobei die Kreuzklassifizierung aus Fach und Hochschule als höhere Ebene verwendet wird. Im Gegensatz zu den bisherigen Analysen stellen nicht die Studienqualitätsbewertungen die Kriteriumsvariable dar, sondern die im BAP erhobene individuelle Bereitschaft, das Studium an der jeweiligen Hochschule weiterzuempfehlen. Diese wird einmal mit den auf den verschiedenen Ebenen aggregierten Bewertungen aus dem BAP, einmal mit jenen aus der CHE-Studie und einmal mit den Merkmalen aus der amtlichen Statistik vorhergesagt, und die entsprechenden Modellkoeffizienten werden verglichen. Eine Variation der Aggregationsebenen erfolgt hier also anders als bisher auf Seite der unabhängigen Variablen.

## **4 Empirische Ergebnisse**

Auch die Darstellung der vorliegenden Ergebnisse erfolgt entlang der vier Leitfragen.

### **(1) Objektivität von Bewertungen**

Wir beginnen mit den Resultaten statistischer Mehrebenenanalysen, in welchen Individuen als disaggregierte und Studiengänge an Hochschulen als aggregierte Analyseeinheiten fungieren (Abb. 1, jeweils erste Säule in der Darstellung der drei Merkmale; die eingezeichneten Konfidenzintervalle beziehen sich jeweils auf das 95 %-Niveau). Bei den im Rahmen des BAP erhobenen Studienbewertungen können wir hier zwischen dem jeweiligen Ausmaß der Varianz *innerhalb* und



**Abb. 1** Varianzkomponenten in den Angaben zur Studienqualität nach Aggregationsebene, Merkmal und Datenquelle. Angegeben sind absolute Varianzen und – für die BAP-Individualdaten – die jeweiligen 95 %-Konfidenzintervalle. STB: Studienbereich; FG: Fachgruppe; HS: Hochschule. Quelle: BAP 2013/2014 (1), angereichert mit Indikatoren des CHE-Hochschulrankings 2012–2014 und der Amtlichen Statistik (Destatis); eigene Berechnungen

*zwischen* den aggregierten (Kontext-)Einheiten unterscheiden.<sup>9</sup> Eine hohe Variation innerhalb der Kontexte bedeutet einen geringen Grad an Konsistenz in den Bewertungen gemeinsamer aggregierter Einheiten. In den vorliegenden Analysen ist die interne Variation hoch, und der Anteil an der Gesamtvarianz, der auf Gruppenunterschiede zurückzuführen ist (ICC\_1; vgl. auch Bliese, 2000), liegt bei zehn Prozent oder weniger. Wie erwartet ist die interne Variation (und auch die Gesamtvariation) beim Merkmal Studienorganisation am geringsten. Im Hinblick auf Aggregatvergleiche (Gruppenmittelwerte) liegen die Maße der Reliabilität (ICC\_2, berechnet gemäß Bliese, 2000, S. 357) über Werten von 0,7 bis 0,85 und damit im Bereich dessen, was als verlässlich gelten kann (LeBreton & Senter, 2008). Bei der explizit subjektiven Dimension der Zufriedenheit sind die Werte besonders niedrig.

## **(2) Heterogenität der Hochschulkontexte**

Die in Abb. 1 zusammengefassten Analysen erlauben auch Vergleiche der in den Daten jeweils beobachteten Varianz *zwischen* aggregierten Einheiten, wenn das Aggregationsniveau verändert wird (weitere Säulen, jeweils schattierte Flächen). Die Aggregationsebenen repräsentieren – in der Abfolge der Säulen – jeweils Hochschulen, Fächer bzw. Zusammenfassungen davon (Studienbereiche, Fächergruppen), Hochschulen im Rahmen bestimmter Studienfächer und Studienfächer im Rahmen bestimmter Hochschulen. Auch diese Werte wurden hier als Between-Varianzkomponenten in Mehrebenenmodellen geschätzt. (Im BAP enthalten jeweils höher aggregierte Einheiten nun als „interne Varianz“ sowohl interpersonale Unterschiede als auch Heterogenitäten zwischen jeweils kleineren aggregierten Einheiten.) Tab. 3 zeigt exemplarisch die zugrunde liegenden Modelle für den linken oberen Teil von Abb. 1, also Bewertungen der Betreuung im BAP. Die Berechnung der anderen Werte erfolgte analog. Wie erwartet nimmt die Varianz umso mehr ab, je stärker die Facheinheiten aggregiert werden

---

<sup>9</sup> Im Fall der verfügbaren CHE-Daten ist die „interne“ Varianz jeweils definitorisch 0, da nur Durchschnittswerte und keine Individualdaten vorliegen. Auch Informationen über Schätzfehler standen uns nicht zur Verfügung. Um zumindest die ungefähre Größenordnung der Varianzkomponenten abzuschätzen, haben wir die vom CHE in grafischer Form veröffentlichten Konfidenzintervalle „ausgemessen“ und auf Basis dieser sowie der ebenfalls veröffentlichten Fallzahlen entsprechende Varianzen geschätzt. Auch hier gelten für die veröffentlichten Werte Mindestgrenzen für die Fallzahlen, was zu einer verringerten Anzahl der ausgewiesenen Hochschulen führt. Fachspezifisch und auf das gesamte Bundesgebiet bezogen sind die aggregierten Einheiten etwas homogener als in den BAP-Daten, ansonsten sind die Ergebnisse aber vergleichbar.

**Tab. 3** Beispieltabelle: Betreuung – Bestimmung von Varianzkomponenten und ICCs über Multilevel-Modelle, verschiedene Aggregationsebenen (BAP-Indikator)

	M1 <i>Studiengang</i>	M2 <i>Hochschule</i>	M3a <i>Fach</i>	M3b <i>Studienbereich</i>	M3c <i>Fachgruppe</i>	M4 <i>Hochschule &amp; Fach klassifiziert</i>
Konstante	3,748*** (0,033)	3,774*** (0,063)	3,702*** (0,046)	3,751*** (0,048)	3,714*** (0,051)	3,853*** (0,079)
<i>Random-Effects-Teil</i>						
Var (Level 2)	0,118*** (0,019)	0,085*** (0,028)	0,068*** (0,023)	0,035*** (0,018)	0,008*** (0,007)	0,100*** (0,032)
Var (Residual)	0,832*** (0,018)	0,879*** (0,019)	0,907*** (0,020)	0,923*** (0,020)	0,932** (0,020)	0,848*** (0,018)
Var (Fach)						0,052*** (0,018)
Log Likelihood	-5858,595	-5903,101	-5972,803	-5993,214	-6000,567	-5855,904
ICC_1 Studiengang	0,124					
ICC_1 Hochschule		0,089				0,100
ICC_1 Fach			0,070			0,052
ICC_1 Studienbereich				0,036		
ICC_1 Fachgruppe					0,009	
<i>N</i>	4334	4334	4334	4334	4334	4334
<i>N Studiengang</i>	159					
<i>N Hochschule</i>		24				24
<i>N Fach</i>			46			46
<i>N Studienbereich</i>				21		
<i>N Fachgruppe</i>					4	

~p<0,10, \*p<0,05, \*\*p<0,01, \*\*\*p<0,001; geschätzte Standardfehler jeweils in Klammern  
Quelle: BAP 2013/2014 (1), angereichert mit Indikatoren des CHE-Hochschulrankings 2012–2014 und der Amtlichen Statistik (Destatis); eigene Berechnungen

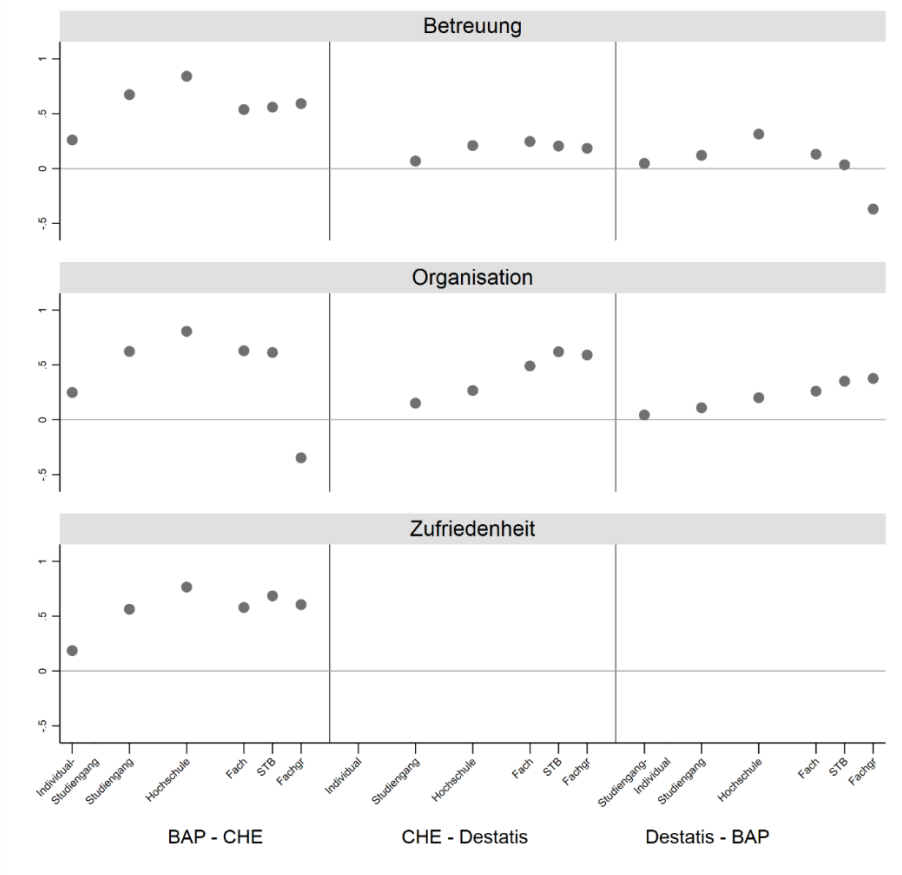
(Studienfächer, Studienbereiche, Fächergruppen). Das bedeutet, dass wenn Studienqualität als entsprechendes Kontextmerkmal verwendet wird, dort jeweils immer weniger Variation zur Verfügung steht – was die Eignung als Kontextmerkmal umso mehr in Frage stellt, je stärker aggregiert wird. Die Muster sind für alle drei betrachteten Merkmale ähnlich, sie unterscheiden sich jedoch im Niveau: Am geringsten ist das Ausmaß der „externen“ Varianz beim Merkmal Studienzufriedenheit, am höchsten hingegen bei der Betreuung. Im Vergleich der Operationalisierungen fällt auch auf, dass die externe Varianz bei Verwendung der Angaben aus dem CHE-Ranking (mittlere Säulengruppe in Abb. 1) deutlich höher ausfällt als bei Verwendung der im BAP selbst erhobenen Bewertungen (Säulengruppe links). Die größten Unterschiede zwischen den beiden Operationalisierungen (Datenquellen) ergeben sich bei der Studienzufriedenheit. Solche Abweichungen können mehrere Ursachen haben, einerseits Unterschiede in den konkreten Stichproben bzw. im Antwortverhalten, andererseits aber auch das unterschiedliche Aggregationsniveau der verfügbaren Daten. Um dies näher zu untersuchen, haben wir die BAP-Individualangaben entsprechend den vorliegenden CHE-Daten auf Studiengangsebene aggregiert und diese als Basis für weitergehende Aggregationen verwendet (Ergebnisse in Säulengruppe rechts). Die Unterschiede zwischen BAP und CHE sind jetzt deutlich geringer ausgeprägt, was auf die Relevanz unterschiedlicher Datenstrukturen hinweist.

### **(3) Konsistenz unterschiedlicher Datenquellen**

Im Anschluss an die bereits deutlich gewordenen Unterschiede prüfen wir nun direkt die Korrelationen, die zwischen den Angaben aus den unterschiedlichen Datenquellen entlang der verschiedenen Merkmale und Aggregationsebenen bestehen (siehe Abb. 2).

Nur im BAP stehen Individualdaten zur Verfügung. In den anderen Datenquellen gibt es keine direkte Entsprechung, daher wird auf die nächsthöhere Datenebene zurückgegriffen. Dies wird beim jeweils ersten Korrelationswert der Kombinationen mit BAP-Beteiligung („Individual – Studiengang“ bzw. „Studiengang – Individual“) deutlich. Die Aggregation der Individualangaben und die weiteren Aggregationen folgen dem Prinzip der (verfügbaren) CHE-Daten, es sind also jeweils Durchschnittswerte ohne Fehlerangabe.

Generell korrelieren die Ergebnisse auf Basis von Angaben in den beiden Befragungen BAP und CHE in allen drei Aspekten der Studienqualität und auf den unterschiedlichen Aggregationsebenen deutlich positiv – die Einheiten werden im Mittel also in beiden Datensätzen tendenziell ähnlich bewertet.



**Abb. 2** Korrelationen der Indikatoren zur Studienqualität nach Aggregationsebene, Merkmal und Datenquelle  
 Beim Betreuungsindikator von Destatis ist wieder zu beachten, dass die niedrigste Fachaggregationsebene den Studienbereichen entspricht (sowohl im „Studiengang“ als auch in der Aggregationsebene „Fach“ steckt in diesem Fall der Studienbereich). STB: Studienbereich; Fachgr: Fachgruppe. Quelle: BAP 2013/2014 (1), angereichert mit Indikatoren des CHE-Hochschulrankings 2012–2014 und der Amtlichen Statistik (Destatis); eigene Berechnungen



Ein typisches Problem zeigt sich allerdings beim Aspekt „Organisation“ auf der Ebene der vier Fachgruppen, wo plötzlich ein negativer Zusammenhang erscheint. Das Problem ist hier offensichtlich der sehr hohe Aggregationsgrad. Ungeachtet der Gewichtung mittels der zahlreichen Individuen handelt es sich ja letztlich um nur noch vier unterschiedliche Datenpunkte, sodass die Ergebnisse sehr anfällig für jede Art von Ausreißern, Hebeleffekten u. Ä. sind. Generell mahnen solche Befunde zu großer Vorsicht bezüglich zu starker Aggregation.

Die Zusammenhänge zwischen den Angaben aus den beiden Befragungen (BAP und CHE) und den objektiven Informationen aus der amtlichen Statistik (Destatis) sind ebenfalls überwiegend positiv; dies ist aber inhaltlich anders zu bewerten, denn *ungünstigere* Betreuungsrelationen und Fachstudiendauern gehen hier also tendenziell mit *besseren* Bewertungen einher. Dieser kontraintuitive Befund wird vor allem von der Gruppe der Universitäten getrieben; auch stellen einige Fächer numerische Ausreißer dar. Außerdem stehen sich Betreuungsrelationen und Bewertungen der Betreuung zwar inhaltlich nahe, sie können aber formal unterschiedlichen Dimensionen von Studienqualität zugeordnet werden, einerseits der Strukturqualität (Betreuungsrelationen), andererseits der Prozessqualität (Bewertungen). Insgesamt zeigen sich also konsistente Übereinstimmungen nur zwischen den Befragungsdaten.

#### **(4) Konsistenz in der Prädiktion**

Im letzten Schritt der empirischen Analyse führen wir multivariate Mehrebenenmodelle mit Hochschule und Fach als kreuzklassifizierten Faktoren auf aggregierter Ebene durch, wobei die Merkmale der Studienqualität als Prädiktorvariablen dienen und die Weiterempfehlung als Kriteriumsvariable. Zusätzlich werden die Note der Hochschulzugangsberechtigung und die Studienabschlussnote als potenziell konfundierende Faktoren auf Individual-ebene kontrolliert (vgl. auch Grotheer & Kerst, 2011). Beispiele für die Modelle finden sich in Tab. 4; Abb. 3 gibt einen Überblick über die Ergebnisse der Modellschätzungen. Dargestellt sind die Punktschätzungen der betreffenden Koeffizienten und die zugehörigen 95 %-Konfidenzintervalle.

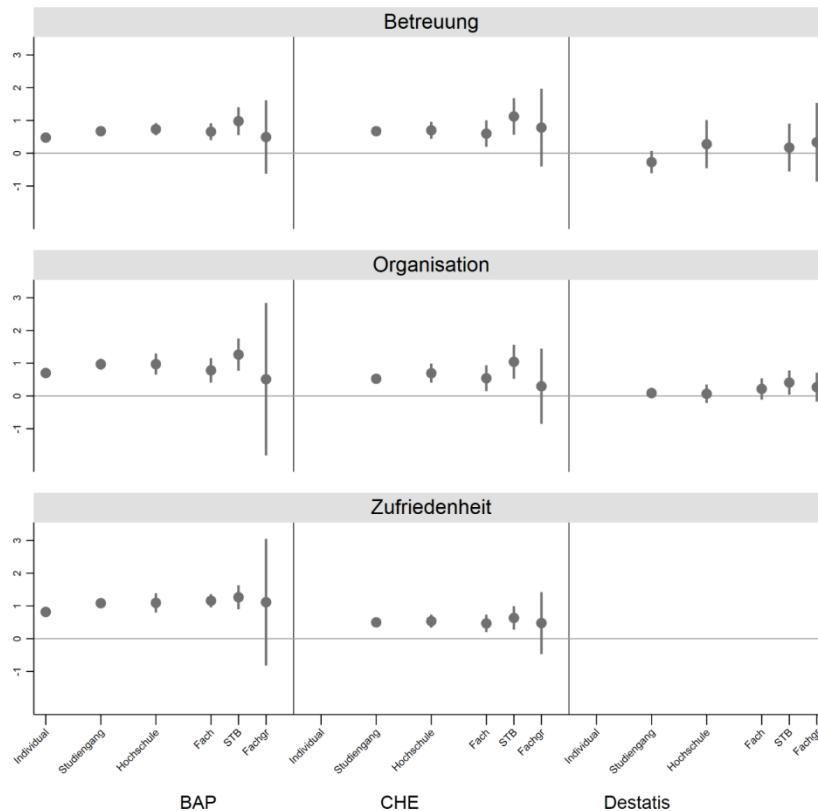
Die erste Operationalisierung („Individual“), die nur mit den BAP-Daten durchgeführt werden kann, enthält als einzige die gesamte Varianz der Individualangaben und führt daher zu einer besonders präzisen Schätzung. Auf den aggregierten Ebenen finden sich eher stärkere, aber auch deutlich unpräziser geschätzte Effekte. Am deutlichsten sind sie auf der Ebene der Studienbereiche (allerdings auch mit größerer Schätzunsicherheit), während die wenigen Fächergruppen offensichtlich zu hoch aggregiert sind, um noch relevante Unterschiede abzubilden.

**Tab. 4** Beispieltabelle: Betreuung Indikator des BAP als Determinante der Weiterempfehlung – Multilevel-Modelle mit verschiedenen Aggregationsebenen der Studienqualität

	M1a <i>Individual</i>	M1b <i>Studiengang</i>	M1c <i>Hochschule</i>	M1d <i>Fach</i>	M1e <i>Studienbereich</i>	M1f <i>Fachgruppe</i>
<b>Betreuung</b>	<b>0,479*** (0,014)</b>	<b>0,675*** (0,055)</b>	<b>0,734*** (0,093)</b>	<b>0,658*** (0,131)</b>	<b>0,980*** (0,217)</b>	<b>0,495 (0,573)</b>
Studienabschlussnote	0,209*** (0,034)	0,299*** (0,037)	0,340*** (0,037)	0,359*** (0,037)	0,360*** (0,037)	0,363*** (0,037)
Note der Hochschulzugangsberechtigung	-0,044 (0,029)	-0,074* (0,031)	-0,075* (0,032)	-0,093** (0,032)	-0,093** (0,032)	-0,094** (0,033)
Konstante	1,729*** (0,112)	0,812*** (0,227)	0,512 (0,364)	0,814 (0,500)	-0,411 (0,821)	1,408 (2,133)
<i>Random-Effects-Teil</i>						
Var (Hochschule)	0,013*** (0,006)	0,010*** (0,005)	0,008*** (0,004)	0,047*** (0,017)	0,050*** (0,018)	0,052*** (0,019)
Var (Fach)	0,061*** (0,018)	0,050*** (0,016)	0,084*** (0,024)	0,051*** (0,016)	0,056*** (0,018)	0,091*** (0,027)
Var (Residual)	0,762*** (0,017)	0,931*** (0,020)	0,948* (0,021)	0,947* (0,021)	0,947* (0,021)	0,947* (0,021)
Log Likelihood	-5609,538	-6034,756	-6078,244	-6084,219	-6085,996	-6094,189
ICC_1 Hochschule	0,016	0,010	0,007	0,045	0,048	0,048
ICC_1 Fach	0,073	0,051	0,081	0,049	0,053	0,084
<i>N</i>	4334					
<i>N Hochschule</i>	24					
<i>N Fach</i>	46					

~p < 0,10, \*p < 0,05, \*\*p < 0,01, \*\*\*p < 0,001; geschätzte Standardfehler jeweils in Klammern

Quelle: BAP 2013/2014 (1), angereichert um Indikatoren des CHE-Hochschulrankings 2012–2014 und der Amtlichen Statistik (Destatis); eigene Berechnungen



**Abb. 3** Effekte von Merkmalen der Studienqualität auf die Weiterempfehlung – unterschiedliche Aggregationsebenen und Datenquellen der Studienqualität

Die Angaben sind jeweils nicht immer auf allen Ebenen verfügbar, und im Fall der Destatis-Betreuerbeziehungen entspricht „Studiengang“ der Kombination aus Hochschule und Studienbereich statt des Fachs. Alle Koeffizienten geschätzt unter Kontrolle der Noten von Hochschulzugangsberechtigung und Studienabschluss sowie Hochschule und Fach; abgebildet mit den zugehörigen 95 %-Konfidenzintervallen. STB: Studienbereich; Fachgr: Fachgruppe. Quelle: BAP 2013/2014 (1), angereichert mit Indikatoren des CHE-Hochschulrankings 2012–2014 und der Amtlichen Statistik (Destatis); eigene Berechnungen

Die Modellierungen mit Aggregatdaten, einmal anhand der BAP-Indikatoren und einmal anhand der CHE-Indikatoren, fallen ähnlich aus und zeigen überwiegend positive und statistisch signifikante Effekte. Unabhängig davon, aus welcher Datenquelle die Bewertung stammt, wird das Studium an der Hochschule umso eher weiterempfohlen, je besser das jeweilige Studienqualitätsmerkmal im Mittel bewertet wird. Die Verwendung der objektiven Merkmale aus der amtlichen Statistik hingegen führt zu deutlich veränderten Modellkoeffizienten. Während sich im Fall der Betreuung (Betreuungsrelationen) überhaupt keine statistisch signifikanten Effekte zeigen, gibt es im Fall der Organisation (Studien-dauer) nur eine leicht positive Tendenz.

## 5 Resümee

### 5.1 Zusammenfassung zentraler Ergebnisse

In diesem Beitrag haben wir uns entlang von vier Leitfragen mit der Verwendung von individuellen und aggregierten Indikatoren der akademischen Studienqualität beschäftigt. In der ersten Leitfrage ging es um die *Objektivität von Bewertungen der Studienqualität* bzw. die Frage, inwiefern sich Individualangaben zur Konstruktion von (kollektiven) Kontextmerkmalen eignen. Dies setzt voraus, dass im Hinblick auf den Gegenstand der Frage die Vorstellung „objektiv vorhandener“, geteilter Kontextbedingungen angemessen ist – was in Bezug auf die hier betrachteten Aspekte vor allem für die Qualität der Studienorganisation angenommen werden kann. Tatsächlich ist der Anteil der geteilten Variation bei diesem Merkmal etwas größer als bei der Bewertung der Betreuung und der Zufriedenheit. Insgesamt sind die geteilten Anteile aber recht gering, wie auch in bisherigen Studien liegt der weitaus größere Teil der Varianz jeweils *innerhalb* der bewerteten institutionellen Einheiten. Gleichwohl ermöglichen die Angaben eine ausreichend stabile Bestimmung von Gruppenmittelwerten auf den unterschiedlichen Aggregationsebenen.

Unsere zweite Leitfrage betraf die *Variation der Studienqualitätsbewertungen zwischen unterschiedlich aggregierten Hochschulkontexten*. Auf der Ebene von Studiengängen gibt es durchaus nennenswerte Unterschiede, bei stärkeren Aggregationen wird die Varianz aber tendenziell eingeebnet, was dann eine sinnvolle Verwendung als Kontextmerkmal zunehmend in Frage stellt.

In der dritten Leitfrage haben wir nach der *Konsistenz der Informationen aus unterschiedlichen Datenquellen* gefragt. Hierbei wurde nicht von einem grundsätzlich „wahren“ Maßstab ausgegangen, sondern (nur) nach Übereinstimmungen

und damit Potenzialen einer wechselseitigen Substitution gesucht. Es zeigt sich, dass die Aggregatinformationen auf der Basis der Studierenden- und Absolvent\*innenangaben zwar nicht genau übereinstimmen, aber insgesamt in dieselbe Richtung weisen. Von Studierenden besser bewertete Einheiten werden tendenziell auch von ehemaligen Studierenden besser bewertet. Es handelt sich hierbei allerdings nicht um dieselben Personen, das Ergebnis stellt also nicht unbedingt einen Beleg für die Konsistenz individueller Einschätzungen über die Zeit dar. Die verfügbaren objektiveren Indikatoren aus der amtlichen Statistik hingegen weisen keine oder teilweise sogar inhaltlich negative Zusammenhänge mit korrespondierenden Bewertungen auf, sodass hier offenbar nur wenige Alternativen bei der Datenauswahl bestehen.

In der vierten Leitfrage schließlich ging es darum, inwieweit Aggregatinformationen zur Studienqualität, die sich auf dieselben Einheiten beziehen, aber aus unterschiedlichen Datenquellen stammen, zu *substanziell unterschiedlichen inhaltlichen Ergebnissen* führen. Bei der Verwendung als Prädiktoren in Mehrebenenmodellen hängt auch hier die Güte der Schätzungen davon ab, ob es gelingt, Aggregate ausreichend detailliert abzubilden; spätestens bei den Fächergruppen ist offensichtlich eine Grenze erreicht. In den Befunden auf Basis der beiden Befragungen zeigen sich erneut Übereinstimmungen auch im Hinblick auf substanzielle Zusammenhänge, während die auf Basis der amtlichen Statistik gebildeten Indikatoren keine entsprechende Prädiktion ermöglichen. Dies legt nahe, dass sich diese Indikatoren relativ weit weg von der Studienrealität befinden. Beispielsweise muss sich eine insgesamt günstige quantitative Betreuungsrelation nicht notwendig in einer individuell positiv wahrgenommenen Betreuung niederschlagen, und es hat sich auch gezeigt, dass viele wesentliche Gründe für lange Studiendauern nicht aufseiten der Studienorganisation zu finden sind (Penthin et al., 2017).

## **5.2 Folgerungen für Hochschulforschung und Hochschulpraxis**

Unsere Analysen relativieren zunächst die allgemeine Aussagekraft von Fächer- und Institutionenvergleichen anhand von aggregierten Bewertungen. Die Mehrebenenmodelle mit BAP-Daten zeigen, dass der Anteil der Varianz zwischen Studienkontexten, also jener Anteil, der von Hochschulen jenseits von reinen Selektionseffekten überhaupt nennenswert beeinflusst werden kann, offenbar eher gering ist – selbst wenn die Betrachtungseinheiten sehr fein gewählt werden. Zum anderen wird die Relevanz der Aggregationsebene bei solchen Vergleichen deut-

lich; Hochschulevaluationen und kommerzielle Hochschulrankings zur Unterstützung der Studienentscheidung sollten die fachliche Dimension stets mit einbeziehen (wie auch vielfach bereits umgesetzt).

Grundsätzlich bestehen große Unterschiede im Informationspotenzial von Individualdaten und nur aggregiert verfügbaren Informationen, insbesondere in der Form einfacher Durchschnittswerte ohne (verfügbare) Fehlerangabe. Diese verschleiern die zum Teil beträchtliche interne Varianz und suggerieren ein höheres Ausmaß an Aussagekraft, als angesichts der relativ geringen Anteile geteilter Varianz gerechtfertigt ist. Verwendet man sie im Rahmen von Kontextanalysen, dann wird auch die Fehlervarianz von entsprechenden Effekten falsch geschätzt, denn es wird i. d. R. nicht berücksichtigt, dass die verwendeten Werte selbst mit Unsicherheit behaftet sind. Mit Individualdaten hingegen, bei denen man die Informationen selbst aggregiert, kann man zumindest prinzipiell die Auswertungen so gestalten, dass die Schätzunsicherheit beim Übergang von niedrigeren auf höhere Aggregationsebenen erhalten bleibt, auch wenn man bewusst nicht die Einzelwerte, sondern beispielsweise aggregierte Mittelwerte als Prädiktoren nimmt. Die (zusätzliche) Bereitstellung von Angaben zur Schätzunsicherheit der ausgewiesenen Mittelwerte, wie sie teilweise für das CHE-Ranking erfolgt, ist daher positiv zu werten, allerdings sollten solche Angaben auch unmittelbar in numerischer und nicht nur in grafischer Form zur Verfügung stehen. Weiterhin bieten unsere Analysen nur wenig Anhaltspunkte dafür, dass die gewählten Indikatoren der amtlichen Statistik aussagekräftig für erlebte Studienbedingungen oder -qualität stehen und damit als Substitute für individuelle Einschätzungen dienen können. Dies hängt sicher auch damit zusammen, dass sie für Zwecke der Hochschulverwaltung und der Hochschulpolitik erhoben werden. Trotz inhaltlicher Nähe sind hier Unterschiede eher zu erwarten als bei Abbildung derselben Qualitätsdimensionen in Befragungsdaten.

Schließlich lassen sich aus dieser Studie methodische Implikationen für die Anreicherung von Datenquellen mit externen Indikatoren ableiten. Zunächst sollten für eine Zusammenführung unterschiedlicher Datenquellen die Aggregationseinheiten möglichst übereinstimmen. Im Fall der Hochschule kann es hierbei zu Komplikationen kommen, wenn die Hochschulangaben oder feingliedrige Fachangaben aufgrund von Datenschutzvorgaben nicht zugänglich sind. Allerdings kann auch eine unterschiedliche Fachgliederung zu Unschärfen oder zumindest zusätzlichem Harmonisierungsaufwand führen – wie in der vorliegenden Studie etwa im Fall der vom BAP und der amtlichen Statistik verwendeten Studienfachsystematik des Statistischen Bundesamts im Vergleich zur eigenen Fächerklassifikation des CHE. Um die Vergleichbarkeit zu erhöhen, ist darüber hinaus eine möglichst gute Übereinstimmung hinsichtlich der zeit-

lichen Dimension sowie der inhaltlichen Gestaltung der erhobenen Konstrukte notwendig. Dem Vorteil, die Informationen nicht im eigenen Survey erheben zu müssen, steht folglich der Nachteil einer möglicherweise geringen Übereinstimmung gegenüber, die auch mit einer nennenswerten Reduktion der Fallzahlen auf Individual- und Aggregationsebenen verbunden sein kann.

### **5.3 Limitationen und Ausblick**

Unsere Analysen haben sich auf eine kleine Anzahl ausgewählter Variablen beschränkt, für die jeweils zu den Angaben aus unserer primären Datenquelle sinnvolle Entsprechungen in den anderen Datenquellen gefunden werden konnten. Mit den Bewertungen von Betreuung und Studienorganisation sind allerdings zwei Aspekte vertreten, die häufig als zentrale Komponenten der Prozessqualität verwendet werden (Grotheer & Kerst, 2011), und die Ergänzungen durch Betreuungsrelationen als Merkmal der Strukturqualität sowie Zufriedenheit, Studiendauern und Weiterempfehlung als Aspekte der Ergebnisqualität ermöglichen eine Anknüpfung an weitere sozialwissenschaftliche Analysen. Gerade die Tatsache, dass häufig eben keine hinreichenden Informationen zu bestimmten Aspekten der Studienqualität übereinstimmend in mehreren Datenquellen vorliegen, unterstreicht die Wichtigkeit, mehr über die Potenziale alternativer Quellen zu forschen und auch über neue Indikatoren nachzudenken.

Eine weitere mögliche Einschränkung stellt die regionale Beschränkung der Datenbasis dar. Wir können letztlich nicht feststellen, ob sich die bayerischen Hochschulen oder Studierenden bzw. Absolvent\*innen von jenen in anderen Bundesländern in einem für die Generalisierbarkeit dieser Befunde relevanten Maße unterscheiden. Allerdings ist die Grundgesamtheit, die in der vorliegenden Datenbasis widerspiegelt wird, prinzipiell sehr vielfältig. Unter den Hochschulen finden sich Volluniversitäten, eine technische Universität und mehrere technische Hochschulen sowie Einrichtungen mit wirtschafts- oder sozialwissenschaftlichen Profilen, dazu sowohl Hochschulen mit jahrhundertelanger Geschichte als auch Neugründungen aus den 1970er Jahren. Auch derartige Unterschiede ließen sich noch näher untersuchen. In den Daten gibt es beispielsweise Hinweise darauf, dass die Bewertungen von Betreuung und Organisation in Universitäten und großen Hochschulen heterogener ausfallen; ebenso in den Geistes-, Rechts-, Sozial- und Wirtschaftswissenschaften im Vergleich zu den mathematisch-naturwissenschaftlichen Fächern und den Ingenieurwissenschaften. Wir konnten dies im Rahmen dieses Beitrags allerdings nicht vertiefen.

Eine dritte Limitation ergibt sich daraus, dass aus den CHE-Befragungen keine Individualdaten verfügbar waren und somit nicht alle Analysen analog mit den beiden Befragungsdatensätzen durchgeführt werden konnten. Für zukünftige Analysen ist daher auch an andere Informationsquellen zu denken, etwa den SQM. Insgesamt bieten sich bereits mit den vorliegenden, erst recht aber mit erweiterten Daten zahlreiche Möglichkeiten für vertiefte Analysen zu verschiedenen Aspekten der Erhebung und Verwendung von Merkmalen der Studienqualität.

## Literatur

- Banschbach, V., & Falk, S. (2017). Warum in die Ferne schweifen? Der Übergang vom Bachelor- ins Masterstudium an bayerischen Hochschulen. *IHF kompakt August 2017*. IHF.
- Beicht, U., & Ulrich, J. G. (2010). When graduates assess the use of their vocational training: What do their ratings stand for? A critique of evaluative research in VET. *Swiss Journal of Sociology*, 36(1), 161–185.
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Hrsg.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (S. 349–381). Jossey-Bass.
- Blossfeld, H.-P., von Maurice, J., & Schneider, T. (2011). The National Educational Panel Study: Need, main features, and research potential. *Zeitschrift für Erziehungswissenschaft*, 14(2), 5–17.
- Blüthmann, I. (2012). Individuelle und studienbezogene Einflussfaktoren auf die Zufriedenheit von Bachelorstudierenden. *Zeitschrift für Erziehungswissenschaft*, 15(2), 273–303.
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, 83(2), 234–246.
- CHE (Centrum für Hochschulentwicklung). (2020a). CHE Hochschulranking. <https://www.che.de/ranking-deutschland/>. Zugegriffen: 28. Okt. 2021.
- CHE (Centrum für Hochschulentwicklung). (2020b). CHE Ranking-Methodik. <https://methodik.che-ranking.de/>. Zugegriffen: 28. Okt. 2021.
- Donabedian, A. (1992). The role of outcomes in quality assessment and assurance. *QRB – Quality Review Bulletin*, 18(11), 356–360.
- Duarte, P. O., Raposo, M. B., & Alves, H. B. (2012). Using a satisfaction index to compare students' satisfaction during and after higher education service consumption. *Tertiary Education and Management*, 18(1), 17–40.
- DZHW (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung). (2020a). Studienqualitätsmonitor. <https://www.dzhw.eu/forschung/governance/sqm>. Zugegriffen: 28. Okt. 2021.



- DZHW (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung). (2020b). DZHW-Absolventenpanel. [https://www.dzhw.eu/forschung/projekt?pr\\_id\\_467](https://www.dzhw.eu/forschung/projekt?pr_id_467). Zugegriffen: 28. Okt. 2021.
- Falk, S., Müller, C., Reimer, M., & Wieschke, J. (2017). Feldbericht der ersten Erhebung des Absolventenjahrgangs 2013/2014. IHF. <https://www.bap.ihf.bayern.de/absolventenjahrgaenge/2013-2014>. Zugegriffen: 28. Okt. 2021.
- Falk, S., Reimer, M., & Hartwig, L. (2007). Absolventenforschung für Hochschulen und Bildungspolitik: Konzeption und Ziele des „Bayerischen Absolventenpanels“. *Beiträge zur Hochschulforschung*, 29(1), 6–33.
- Grotheer, M., & Kerst, C. (2011). *Studienqualität in system- und hochschulbezogener Perspektive. Auswertungen mit Daten des Studienqualitätsmonitors und des Konstanzer Studierendensurveys*. HIS.
- Hillmert, S., Groß, M., Schmidt-Hertha, B., & Weber, H. (2017). Informational environments and college student dropout. In J. Buder & F. W. Hesse (Hrsg.), *Informational environments: Effects of use, effective designs* (S. 27–52). Springer.
- HIS (Hochschul-Informationen-System GmbH). (2009). *Perspektive Studienqualität: Themen und Forschungsergebnisse der HIS-Fachtagung Studienqualität*. HIS.
- HRK (Hochschulrektorenkonferenz) (2020). Stichwort „Studienqualität“. <https://www.hrk-nexus.de/glossar-der-studienreform/begriff/studienqualitaet/>. Zugegriffen: 28. Okt. 2021.
- ICELand. (2020a). Das hochschulstatistische Informationssystem der Bildungsministerien der Länder. <https://iceland.dzhw.eu/>. Zugegriffen: 28. Okt. 2021.
- ICELand. (2020b). Datenbestandsübersicht Hochschulstatistik. <https://iceland.dzhw.eu/iceproject/www/app/land/stat/datastock.html#T4>. Zugegriffen: 28. Okt. 2021.
- Kempgen, S. (2020). *Studienqualität: Begriffe – Modelle – Analysen*. University of Bamberg Press.
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815–852.
- Lörz, M., Quast, H., Roloff, J., & Trennt, F. (2019). Determinanten des Übergangs ins Masterstudium. Theoretische Modellierung und empirische Überprüfung. In M. Lörz & H. Quast (Hrsg.), *Bildungs- und Berufsverläufe mit Bachelor und Master: Determinanten, Herausforderungen und Konsequenzen* (S. 53–93). Springer VS.
- Luque Martínez, T., & Doña Toledo, L. (2013). What do graduates think? An analysis of intention to repeat the same studies and university. *Journal of Marketing for Higher Education*, 23(1), 62–89.
- Multus, F. (2013). *Referenzrahmen zur Lehr- und Studienqualität: Aufarbeitung eines facettenreichen Themensfeldes. Hefte zur Bildungs- und Hochschulforschung 67*. Arbeitsgruppe Hochschulforschung, Universität Konstanz.
- Penthin, M., Fritzsche, E. S., & Kröner, S. (2017). Gründe für die Überschreitung der Regelstudienzeit aus Studierendensicht. *Beiträge zur Hochschulforschung*, 39(2), 8–31.
- Pike, G. R. (1994). The relationship between alumni satisfaction and work experiences. *Research in Higher Education*, 35(1), 105–123.
- Pohlentz, P. (2008). *Datenqualität als Schlüsselfrage der Qualitätssicherung an Hochschulen*. Universitätsverlag Potsdam.
- Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata*. Stata Press.

- Rhee, B.-S. [Yi, P.-s.]. (2008). Institutional climate and student departure: A multi-nomial multilevel modeling approach. *The Review of Higher Education*, 31(2), 161–183.
- Rottach, A., Blüthmann, I., & Watermann, R. (2016). Studienqualität und Verbleibsquote im Masterstudium. Eine Analyse auf Studiengangsebene zum Zusammenhang von Qualitätsmerkmalen von Masterstudiengängen mit der Verbleibsquote. *Empirische Pädagogik*, 30(3/4), 531–556.
- Schiefele, U., & Jacob-Ebbinghaus, L. (2006). Lernermerkmale und Lehrqualität als Bedingungen der Studienzufriedenheit. *Zeitschrift für Pädagogische Psychologie*, 20(3), 199–212.
- Schmidt, U., & Neßler, C. (2018). Absolventenstudien: Aussagekraft und Handlungsrelevanz retrospektiver Bewertungen des Studiums. In S. Falk, M. Reimer & U. Schmidt (Hrsg.), *Absolventenstudien und Qualitätsmanagement: Best Practices an deutschen und österreichischen Hochschulen* (S. 173–188). Waxmann.
- SPIEGEL (1990). *Studieren heute – Welche Uni ist die beste? SPIEGEL-Rangliste der westdeutschen Hochschulen*. SPIEGEL Spezial 1/1990. SPIEGEL-Verlag.
- Statistisches Bundesamt. (2018a). Bildung und Kultur: Personal an Hochschulen. Fachserie 11, Reihe 4.4. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/personal-hochschulen-2110440177004.pdf?blob=publicationFile>. Zugegriffen: 28. Okt. 2021.
- Statistisches Bundesamt. (2018b). Hochschulen auf einen Blick. Ausgabe 2018b. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/broschuere-hochschulen-blick-0110010187004.pdf?blob=publicationFile>. Zugegriffen: 28. Okt. 2021.
- Statistisches Bundesamt. (2020). Bildung, Forschung und Kultur: Hochschulen. [https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/\\_inhalt.html?jsessionid = EBAD7EBE8F24EB45ACD55E6C6D8F045F.internet8721](https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/_inhalt.html?jsessionid = EBAD7EBE8F24EB45ACD55E6C6D8F045F.internet8721). Zugegriffen: 28. Okt. 2021.
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago Press.
- Woisch, A., Ortenburger, A., & Multrus, F. (2013). *Studienqualitätsmonitor 2012: Studienqualität und Studienbedingungen an deutschen Hochschulen*. HIS.