

Forschungsevaluation auf Institutsebene – Der Einfluss der gewählten Methodik auf die Ergebnisse

Gerhard Reichmann, Christian Schlögl, Wolfgang G. Stock, Isabelle Dorsch

Der Artikel zeigt, wie sehr die Ergebnisse einer Forschungsevaluation auf Instituts-ebene durch die Wahl unterschiedlicher Vorgangsweisen beeinflusst werden können. Insbesondere wird dabei auf die folgenden neun potenziellen Einflussfaktoren für wissenschaftliche Leistung eingegangen: die herangezogene Datenquelle, die relevanten Institutsmitglieder, die Institutsgröße, die berücksichtigten Publikationsarten, die Wertigkeit der unterschiedlichen Publikationsarten, die vorhandenen Koautorinnen- und Koautorenschaften, die Publikationssprache, den Umfang der Publikationen in Seiten sowie die Berücksichtigung von Zeitschriftenrankings. Anhand einer Fallstudie zeigt sich, dass die relevanten Institutsmitglieder sowie die Berücksichtigung von Zeitschriftenrankings und Institutsgröße besonders ergebnisrelevant sind.

1 Einleitung

Evaluation ist eine Thematik, die sich im universitären Sektor mittlerweile weitgehend etabliert hat. Sie wird hier häufig als Instrument zur Leistungsmessung und zum Leistungsnachweis eingesetzt, und zwar in allen drei Kernbereichen universitärer Tätigkeit, also in Verwaltung (Jackson, 2015; Marina et al., 2015; Reichmann & Sommersguter-Reichmann, 2010), Lehre (Böttcher & Thiel, 2018; Kember & Ginns, 2012; Reichmann, 2007) und insbesondere Forschung (Jappe, 2020; de Rijcke et al., 2016; Gingras, 2016; Schlögl, 2014; Vinkler, 2010). Die im Zuge von Forschungsevaluationen erhobenen Leistungen, bei denen es sich in erster Linie um den Forschungsoutput von Personen oder Organisationseinheiten in Form von Publikationen handelt, dienen einerseits als Beleg für eine erfolgreiche Forschungstätigkeit, werden aber auch zunehmend als Basis für Leistungsvergleiche zwischen mehreren betrachteten Personen bzw. Einheiten herangezogen. Einzelne Forschende, aber auch Institute, Fachbereiche, Fakultäten oder ganze Universitäten, die bei solchen Vergleichen in Form von Rankings gut abschneiden, verfügen in der Folge nicht selten über Wettbewerbsvorteile (Fraja et al., 2019; Hicks, 2012). Bezieht sich ein solches Forschungsranking auf Institute, so kann ein gutes Abschneiden von großer Bedeutung für die zukünftige personelle und infrastrukturelle Ausstattung dieser Einheiten sein. Als primäre Indikatoren für Forschungsleistung gelten üblicherweise die hier untersuchten (teilweise gewichteten) Summen von wissenschaftlichen Publikationen, aber

auch die Anzahl von Zitationen sowie kombinierte Publikations-Zitations-Werte, wie der H-Index (Kamrani et al., 2020), werden im Rahmen von Forschungsevaluationen verbreitet herangezogen.

Auch im deutschsprachigen Raum hat es in den vergangenen Jahren eine Vielzahl von Aktivitäten im Bereich der Forschungsevaluation gegeben. Eine hohe Anzahl einschlägiger Publikationen in den unterschiedlichsten Publikationsmedien zeugt davon (Albers, 2015; Bommer & Ursprung, 1998; Fülbiel & Weller, 2011; Reichert, 2019; Sieweke et al., 2014; Wolf et al., 2005).

Eine auf Basis von Titel und Abstract erfolgte Analyse der Publikationen in drei – aus Sicht der Autorin und der Autoren – besonders wichtigen deutschsprachigen Zeitschriften für diesen Themenbereich, die sich auf die letzten zehn Jahre bezieht, legt offen, dass es hier insgesamt 43 Publikationen zum Thema „Forschungsevaluation“ gegeben hat (Tabelle 1). Diese sind relativ gleichmäßig auf die drei analysierten Zeitschriften verteilt, allerdings gibt es jahresweise Schwankungen. Die Betrachtung der Ausrichtung der Beiträge lässt erkennen, dass nur relativ wenige Forschungsevaluationen in Form von Leistungsvergleichen zwischen Organisationseinheiten durchgeführt wurden. Es überwiegen theoretische Beiträge, in denen auf Probleme im Zuge solcher Rankings bzw. auf Einzelaspekte im Zusammenhang mit diesen eingegangen wird.

Tabelle 1: Anzahl und Ausrichtung der Beiträge zum Thema „Forschungsevaluation“ in drei einschlägigen deutschsprachigen Zeitschriften in den vergangenen zehn Jahren

Zeitschrift	Ausrichtung der Beiträge	Anzahl der Beiträge										Gesamt – alle Jahre
		2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	
Beiträge zur Hochschulforschung	Theoretische Ausführungen	1	2	4	2			2	3		1	15
	Empirische Studien (Rankings)											0
	Gesamt	1	2	4	2	0	0	2	3	0	1	15
Information, Wissenschaft & Praxis	Theoretische Ausführungen		1		1			6	3	2	1	14
	Empirische Studien (Rankings)			1	1							2
	Gesamt	0	1	1	2	0	0	6	3	2	1	16
Zeitschrift für Evaluation	Theoretische Ausführungen			3	2			1		1		7
	Empirische Studien (Rankings)		1	1	1		1			1		5
	Gesamt	0	1	4	3	0	1	1	0	2	0	12
Gesamt – alle Zeitschriften		1	4	9	7	0	1	9	6	4	2	43

Eine nähere Betrachtung der in Tabelle 1 angeführten empirischen Studien lässt aufgrund der dort eingesetzten verschiedenen Vorgangsweisen vermuten, dass die Ergebnisse einer Forschungsevaluation sehr stark von der gewählten Vorgangsweise, die in der Regel am Anfang der Untersuchung festgelegt und danach meist nicht mehr variiert wird, abhängig sind. Um diese Vermutung zu überprüfen, soll nachfolgend eine beispielhafte vergleichende Forschungsevaluation für zwei fachverwandte Universitätsinstitute durchgeführt werden, bei der die Vorgangsweise mehrfach variiert wird. Insbesondere soll explorativ ermittelt werden, welche Aspekte der Vorgangsweise die Ergebnisse besonders stark beeinflussen. Diese Aspekte werden in der Folge als Einflussfaktoren (auf die Ergebnisse einer Forschungsevaluation) bezeichnet. Zudem wird noch darauf eingegangen, welche der vorgestellten Vorgangsweisen die Autorin und die Autoren bevorzugen würden. Damit ergeben sich folgende Forschungsfragen:

Hauptforschungsfrage: Welche der neun überprüften Einflussfaktoren haben die größten Auswirkungen auf die Ergebnisse einer Forschungsevaluation auf Institutsebene?

Nebenforschungsfrage: Welche der möglichen Vorgangsweisen erscheint der Autorin und den Autoren dieses Beitrags am sinnvollsten?

2 Forschungsevaluation in Form einer vergleichenden Publikationsanalyse auf Institutsebene

Eine Forschungsevaluation im Universitätsbereich kann sich auf einzelne Forschende (Rousseau et al., 2018), was in der Praxis sehr häufig vorkommt (Bornmann & Marx, 2013; Meyer et al., 2012), aber auch auf Forschungsgruppen, Institute bzw. Departments (Friedländer, 2014), Fakultäten bzw. Fachbereiche (Tost & Rindermann, 2016), gesamte Universitäten oder sogar Gruppen von Universitäten beziehen (Morgan, 2004; Reichert, 2019, S. 43–44). Im vorliegenden Fall erfolgt eine Evaluation auf Instituts-ebene, und zwar beschränkt auf die Gegenüberstellung der Forschungsleistung zweier fachverwandter Institute, da eine solche Beschränkung einen übersichtlichen Vergleich zwischen unterschiedlichen Vorgangsweisen ermöglicht. Die Institutsebene erscheint aufgrund der Sicherstellung einer gewissen Datenbreite geeignet, d.h., es sollte dadurch gewährleistet sein, dass sich die untersuchten Einflussfaktoren ausreichend in der Datenbasis widerspiegeln, also beispielsweise unterschiedliche Publikationsarten in verschiedenen Sprachen enthalten sind.

Forschungsleistungen können sich in Form von Publikationen (Röbken, 2011), deren Wirkung beispielsweise in Zitationen (Dilger & Müller, 2012; Dyckhoff et al., 2005; Moed, 2005), Auszeichnungen bzw. (Nobel-)Preisen, Mitgliedschaften in Editorial Boards (Rost & Frey, 2011) oder einer Zuerkennung von Drittmitteln manifestieren.

Die nachfolgenden Ausführungen und damit auch die empirische Studie sind auf Forschungsleistungen in Form von Publikationen beschränkt, da sie neben Zitationen in der Praxis sicherlich die größte Rolle spielen. Die Publikationen der beiden untersuchten Institute im Betrachtungszeitraum sollen möglichst umfassend berücksichtigt werden, weshalb in erster Linie auf Publikationslisten der institutszugehörigen Forschenden zurückgegriffen wird. Zitationen werden nicht berücksichtigt, da – wie später noch gezeigt werden wird – nur ein kleiner Teil der Publikationen laut Publikationslisten in Zitationsdatenbanken erfasst ist. Auch wird auf eine Erhebung altmetrischer Kennwerte (z. B. Tweets über Publikationen auf Twitter oder Zugriffszahlen auf Publikationen, etwa bei Mendeley (Priem, 2014)) verzichtet, da derzeit noch nicht exakt geklärt ist, was dabei eigentlich gemessen wird.

3 Einflussfaktoren auf die Ergebnisse einer Forschungsevaluation

3.1 Datenquelle

Die große Bedeutung der Wahl der Datenquelle für die Ergebnisse einer Forschungsevaluation wurde bereits in anderen Untersuchungen festgestellt (Halevi et al., 2017; Harzing & Alakangas, 2016; Martín-Martín et al., 2018). Meyer et al. (2012) zeigen etwa, dass im Rahmen einer Forschungsevaluation im Hinblick auf deutschsprachige Forschende aus dem Bereich der Betriebswirtschaftslehre die Datenquelle eine größere Rolle spielt als ein Wechsel zwischen Publikations- und Zitationsanalyse, wobei selbst im Falle einer Heranziehung von Google Scholar bei weitem nicht alle Publikationen erfasst werden. Möchte man den Publikationsoutput von Forschenden erfassen, so bietet sich ein Rückgriff auf allgemeinwissenschaftliche Informationsdienste, wie Web of Science, Scopus, Google Scholar, Microsoft Academic oder Dimensions, an (Dorsch et al., 2018). Am sinnvollsten erscheint der Autorin und den Autoren allerdings die Heranziehung von Publikationslisten, welche die Forschenden selbst erstellt haben. Diese ermöglichen am ehesten eine vollständige Erfassung aller für eine Forschungsevaluation benötigten Publikationen. In der Praxis ist es allerdings oftmals nicht möglich, an aktuelle und vollständige Publikationslisten zu gelangen. Positiv erscheint, dass es mittlerweile an vielen Universitäten, wie etwa an der Universität Graz, öffentlich zugängliche Forschungsdatenbanken gibt, in die alle dort beschäftigten Forschenden alle ihre Publikationen eintragen sollen. Diese Datenbanken können, bei entsprechender Qualität, durchaus als vollwertiger Ersatz für Publikationslisten angesehen werden.

Allgemeinwissenschaftliche Informationsdienste sind als Datenquellen nur bedingt geeignet, da – wie Untersuchungen gezeigt haben (Reichmann, 2020) – oftmals nur ein Bruchteil der relevanten Publikationen darin enthalten ist. Beschränkt man eine Forschungsevaluation aber auf die Publikationsart „Zeitschriftenbeiträge“, so sind die Differenzen wesentlich geringer, und im Falle einer nochmaligen Einschränkung auf

englischsprachige Beiträge (zur Berücksichtigung deutschsprachiger Zeitschriften in diesen Datenbanken: Clermont & Schmitz, 2008) sollten diese Differenzen nochmals deutlich sinken. In Google Scholar sind zwar wesentlich mehr Publikationen erfasst, diese sind aber infolge mangelnder Strukturiertheit der Datenbank nur schwer recherchierbar (Halevi et al., 2017). Um die gefundenen Publikationen zuverlässig den relevanten Forschenden zuordnen zu können, wäre ein Rückgriff auf Publikationslisten nötig. Sind diese vorhanden, kann man aber gleich auf Google Scholar als Datenquelle verzichten. Zudem ist die Datenqualität von Google Scholar schlechter als jene der übrigen erwähnten Datenquellen (zur Brauchbarkeit von Google Scholar: Dilger & Müller, 2012; Tost & Rindermann, 2016). Aufgrund der großen praktischen Bedeutung (Jappe, 2020, S. 11–12) und der Tatsache, dass das Web of Science infolge der meist geringsten Abdeckung (im Vergleich zu Scopus und Google Scholar) gewissermaßen den Gegenpol zu den Publikationslisten darstellt, wurde dieses – neben den Publikationslisten – im empirischen Teil der Arbeit als Datenquelle herangezogen.

3.2 Repräsentantinnen/Repräsentanten der Institutsleistung

Bei diesem Einflussfaktor geht es um die Klärung der Frage, wessen Forschungsleistung in die Forschungsevaluation einfließen soll. Die Beantwortung erscheint auf den ersten Blick simpel: Jene der evaluierten Einheiten, also im Falle einer Forschungsevaluation auf Institutsebene jene der betrachteten Institute (Rousseau et al., 2018). Könnte man nun auf eine Datenbank zurückgreifen, in der die gesamten Publikationen eines Instituts unter dessen Bezeichnung abrufbar sind, wäre die Recherche – zumindest aus erhebungstechnischer Sicht – einfach.

In vielen Fällen kann aber nicht in einem einzigen Schritt auf die gesamte Forschungsleistung eines Instituts zurückgegriffen werden, sondern es müssen die Publikationen auf Personenebene erfasst und anschließend aggregiert werden. Dabei stellt sich die Frage, welche Personen berücksichtigt werden sollen (Bornmann & Marx, 2013; Meyer et al., 2012). Möchte man die Forschungsleistung möglichst umfassend erheben, wird man auf alle Forschenden, d. h. alle wissenschaftlichen Mitarbeitenden, des Instituts zurückgreifen (Bornmann et al., 2012, S. 238). Allerdings weisen „Prae-Doc-Forschende“ bisweilen kaum Publikationen auf, da sie oft mehrere Jahre benötigen, um sich in die Forschung einzuarbeiten, und zudem vorwiegend an ihrer Dissertation arbeiten. Die Einbeziehung solcher Personen würde im Falle einer Berücksichtigung der Institutsgröße zu einer Verschlechterung der Forschungsleistung des Instituts führen. Das gleiche Problem ergibt sich auch hinsichtlich all jener anderen wissenschaftlichen Institutsmitarbeitenden, die kaum publizieren. Hinsichtlich der „Prae-Doc-Forschenden“ lässt sich die angesprochene Problematik lösen, indem nur „Post-Doc-Forschende“ als Repräsentantinnen und Repräsentanten eines Instituts berücksichtigt werden.

Wenn man Forschungsvergleiche auf Personenebene betrachtet, so werden diese häufig nur für Professorinnen und Professoren durchgeführt, da diese als wissenschaftliches Schlüsselpersonal angesehen werden. Folgt man dieser Strategie bei einer Forschungsevaluation auf Institutsebene, werden nur diese Personen als Repräsentantinnen und Repräsentanten der Institutsleistung herangezogen.

3.3 Institutsgröße

Vergleicht man die Forschungsleistung auf Institutsebene, so ist zu hinterfragen, ob es fair und sinnvoll ist, kleinere und größere Institute undifferenziert miteinander zu vergleichen, wie es auf Universitätsebene im Falle von Universitätsrankings häufig geschieht (Olcay & Bulu, 2017). Unterbleibt eine Berücksichtigung der Institutsgröße, dann hätte – bezogen auf die empirische Studie – das Düsseldorfer Institut, dem im Betrachtungszeitraum durchschnittlich mehr als elf wissenschaftliche Mitarbeitende angehörten, gegenüber dem Grazer Institut mit durchschnittlich nur etwas mehr als fünf wissenschaftlichen Mitarbeitenden einen erheblichen Größenvorteil. Bei internationalen Forschungsrankings, v. a. jenen auf Universitätsebene, wird diesbezüglich erstaunlicherweise oft undifferenziert vorgegangen (van Raan, 2005).

Möchte man die Institutsgröße berücksichtigen, so kann man – je nach einbezogenen Repräsentantinnen und Repräsentanten – die Forschungsleistung eines Instituts durch die Anzahl dieser Repräsentantinnen und Repräsentanten dividieren. Alternativ könnte man auch stets durch die Anzahl der Professorinnen und Professoren dividieren, wenn diese Anzahl als zentrales Merkmal für die Größe einer Einheit angesehen wird. Die von der Autorin und den Autoren als am sinnvollsten erachtete Variante einer Division der Forschungsleistung durch die Anzahl aller Forschenden (gemessen in Vollzeitäquivalenten (VZÄ)) nimmt in der empirischen Studie als Bestandteil der sogenannten Basisvariante 1 eine zentrale Stellung ein.

3.4 Publikationsarten

Als nächstes ist zu entscheiden, welche Arten von Publikationen für eine Forschungsevaluation herangezogen werden sollen (Kladroba et al., 2021; Rökken, 2011). Erfolgt diesbezüglich keinerlei Einschränkung, so zählt jeglicher Eintrag in Publikationslisten bzw. Forschungsdatenbanken als Publikation.

Im Falle einer Berücksichtigung dieses Einflussfaktors sind mögliche Publikationsarten zu definieren (Stock, 2000, S. 250–251). In der empirischen Studie wird diesbezüglich lediglich zwischen folgenden Arten unterschieden: Bücher (Monographien), Beiträge in wissenschaftlichen Zeitschriften, Beiträge in Tagungsbänden, Beiträge in sonstigen Sammelwerken sowie Sonstige (alle anderen Publikationen).

3.5 Wertigkeit unterschiedlicher Publikationsarten

Dieser Einflussfaktor kommt nur dann zum Tragen, wenn es eine Differenzierung nach Publikationsarten gibt. Uns erscheint eine Berücksichtigung der Wertigkeit unterschiedlicher Publikationsarten (Kladroba et al., 2021) durchaus als sinnvoll, da es beispielsweise meist wesentlich aufwendiger ist, eine umfangreiche Monographie zu verfassen als einen relativ kurzen Beitrag in einem Sammelwerk, wie einer Festschrift. Aus diesem Grund werden in der empirischen Studie auch zwei derartige Varianten (als Gewichtung 1 und 2 bezeichnet) untersucht. Varianten der Wertigkeit orientieren sich häufig am durchschnittlichen Umfang (deshalb erhalten Bücher dabei meist die höchste Punktzahl) sowie Renommee (deshalb erhalten Beiträge in Zeitschriften manchmal eine höhere Punktzahl als solche in Sammelwerken) der einzelnen Publikationsarten.

3.6 Koautorinnen-/Koautorenschaft

Die Autorin und die Autoren präferieren bei Mehrautorinnen- und Mehrautorenwerken die Berücksichtigung vorhandener Koautorinnen- und Koautorenschaften, also ein „fractional counting“. Ein solches spiegelt die erbrachte Leistung der einzelnen Autorinnen und Autoren wesentlich realistischer wider als die Variante „keine Berücksichtigung von Koautorinnen- und Koautorenschaften“ („full counting“). Dennoch wird im Rahmen der empirischen Studie auch die letztgenannte Variante analysiert, da diese in der Praxis durchaus verbreitet ist (Sivertsen et al., 2019, S. 679).

Für ein „fractional counting“ gibt es mehrere Möglichkeiten (Stock, 2000, S. 258), von denen die gleiche Berücksichtigung aller Autorinnen und Autoren nach der Formel $1/n$ (n = Anzahl der Autorinnen/Autoren) am verbreitetsten ist (Waltman, 2016, S. 378–380) und auch als am sinnvollsten erachtet wird. Die Variante mit der Formel $2/(n+1)$ hat im Sinne einer ausgewogenen Berücksichtigung von Koautorinnen- und Koautorenschaften im Verhältnis zu Alleinautorinnen- und Alleinautorenschaften gewisse Vorteile (Schneijderberg, 2018, S. 343), da berücksichtigt wird, dass Koautorinnen- und Koautorenschaften infolge der notwendigen Koordinierung einen Mehraufwand bedingen können, allerdings sind die mathematischen Probleme der Formel (die Summe der Autorinnen-/Autorenwerte ist größer als eins) ein entscheidender Nachteil.

3.7 Publikationssprache

Bei der Publikationssprache handelt es sich um einen Einflussfaktor, dessen Berücksichtigung von der Autorin und den Autoren – im Gegensatz zu den bisher behandelten Faktoren – durchaus in Frage gestellt wird. Diese Einschätzung gilt auch hinsichtlich einer Beschränkung auf englischsprachige Publikationen (Stockemer & Wigginton,

2019). Untersuchungen haben nämlich gezeigt, dass in manchen Fachbereichen deutschsprachige Organisationseinheiten in vorwiegend englischsprachigen Publikationsdatenbanken, wie Web of Science, nur relativ schwach vertreten sind (Schlögl, 2013), da dort nur relativ selten in englischsprachigen Zeitschriften publiziert wird. Eine Beschränkung auf deutschsprachige Publikationen, also auf Publikationen in jener Sprache, die am Standort der untersuchten Institute gesprochen wird, erscheint als noch weniger sinnvoll, da auf diese Weise oftmals die hochwertigsten Publikationen unberücksichtigt bleiben. Dennoch wird im empirischen Teil auch eine solche zu Demonstrationszwecken vorgenommen.

3.8 Umfang

Bei der Berücksichtigung des Umfangs einer Publikation (Moed, 2005) wird in einer ersten Variante auf die Anzahl der publizierten Seiten zurückgegriffen. Dabei kann es aufgrund von unterschiedlichen Formaten, Schriftgrößen und Layouts zu starken Verzerrungen kommen. Günstiger wäre eine Zählung des Umfangs in Wörtern oder Zeichen, allerdings sind derartige Daten nur bei elektronischen Publikationen verfügbar. In einer zweiten Variante finden nur Publikationen mit einem Mindestumfang von fünf Seiten Berücksichtigung, der Umfang einer Publikation dient in diesem Fall also als Filterkriterium. Es gibt zwar keine internationalen Standards für die optimale Länge einer Publikation, die sicherlich stark von Publikationsart und -medium abhängig sein wird, doch wissenschaftliche Beiträge, v. a. jene in Zeitschriften und Sammelwerken, sollten einen gewissen Mindestumfang aufweisen, um als „vollwertig“ angesehen werden zu können. Allerdings werden bei einer Festlegung auf mindestens fünf Druckseiten viele Beiträge in hochangesehenen Zeitschriften, wie „Nature“ oder „Science“, nicht mehr mitgezählt.

3.9 Zeitschriftenrankings

Dieser Einflussfaktor (Packalen & Bhattacharya, 2017) spielt v. a. bei einer Beschränkung der betrachteten Publikationsarten auf Zeitschriftenbeiträge eine Rolle. Eine in der Praxis verbreitete Gewichtungsmöglichkeit von Zeitschriftenbeiträgen ist jene nach dem „Wert“ der Zeitschrift, in welcher die Publikation erfolgte. Zur Bestimmung des „Werts“ einer Zeitschrift können facheinschlägige Zeitschriftenrankings (z. B. JOURQUAL 2 für den Bereich der Betriebswirtschaftslehre: Schrader & Hennig-Thurau, 2009) oder der Impact-Faktor dieser Zeitschrift laut Web of Science (Bauer, 2003; Garfield, 2006) herangezogen werden. Beiträge in Zeitschriften, die nicht im Web of Science enthalten sind und daher keinen Impact-Faktor aufweisen, werden nicht berücksichtigt.

Bei beiden in der empirischen Studie betrachteten Varianten wird so vorgegangen, dass als Basiswert für jede Zeitschriftenpublikation der Impact-Factor der Zeitschrift, in der publiziert wurde, verwendet wird; und zwar jener für das Jahr der Publikation. Bei der zweiten Variante kommt noch die Heranziehung eines Filterkriteriums für die Berücksichtigung eines Zeitschriftenbeitrags hinzu: Es werden nur Zeitschriftenbeiträge berücksichtigt, die in einer Zeitschrift mit einem Impact-Faktor von mindestens eins publiziert wurden.

4 Empirische Studie

4.1 Vorgehensweise

Um die zentrale Forschungsfrage, nämlich das Ausmaß des Einflusses der neun eben vorgestellten Faktoren auf die Ergebnisse einer Forschungsevaluation auf Instituts-ebene, beantworten zu können, wurde ein beispielhafter Leistungsvergleich im Hinblick auf die Forschungsleistung zweier Institute für einen Zeitraum von zehn Jahren (2009–2018) durchgeführt.

4.1.1 Untersuchungseinheiten

Der Institutsvergleich wurde für die beiden ehemaligen Institute für Informationswissenschaft an den Universitäten Düsseldorf und Graz durchgeführt. (Das Grazer Institut wurde Anfang Oktober 2020 – zusammen mit zwei weiteren Instituten – zum neu gegründeten Institut für Operations und Information Systems fusioniert; die Düsseldorfer Abteilung befindet sich gerade in Auflösung. Ebenso berücksichtigt wurde in Düsseldorf der formal an der Abteilung angesiedelte, jedoch unabhängig arbeitende Bereich „Web Science“.) Die Wahl fiel auf diese beiden Institute, da sie im gleichen Forschungsbereich angesiedelt sind, die Autorin und die Autoren dieses Beitrags Angehörige dieser Institute sind und daher die relevanten Forschungsaktivitäten und Rahmenbedingungen sehr gut kennen, und die Forschungsleistung beider Institute für den Betrachtungszeitraum vollständig dokumentiert und verfügbar ist.

Vergleicht man die personellen Ressourcen der beiden Institute, so zeigt sich, dass es im Betrachtungszeitraum in Düsseldorf immer nur einen Professor gegeben hat, wogegen in Graz stets vier Professoren tätig waren. Allerdings lag die Anzahl der für die Untersuchung relevanten sonstigen wissenschaftlichen Mitarbeitenden (vgl. dazu 4.1.2) in Düsseldorf im Durchschnitt bei 10 und in Graz bei nur 1. Daraus ergibt sich, dass für die vorliegende Evaluation die Forschungsleistung von durchschnittlich 11 VZÄ in Düsseldorf jener von durchschnittlich 5 VZÄ in Graz gegenübergestellt wird. Inhaltlich umfasste die Forschung an beiden Instituten v. a. informationswissenschaftliche Themen, wie Kommunikations- und Zitationsanalysen (Graz) oder Informationskompetenz und Social-Media-Analysen (Düsseldorf). Während es in Graz auch einige

Publikationen aus den Bereichen der Wirtschaftsinformatik sowie der Betriebswirtschaftslehre gegeben hat, lag ein Forschungsschwerpunkt in Düsseldorf auf informationellen Städten (Dorsch et al., 2017). Hinsichtlich der verwendeten Publikationskanäle lag der Schwerpunkt in beiden Fällen auf Zeitschriftenbeiträgen und Beiträgen in Sammelwerken, wobei es in Düsseldorf darüber hinaus auch mehrere Monographien gegeben hat. Zudem weist Düsseldorf eine größere Anzahl von Publikationen in englischer Sprache auf.

4.1.2 Datenerhebung

In einem ersten Schritt wurde der Publikationsoutput der beiden untersuchten Institute für den zehnjährigen Betrachtungszeitraum erhoben. Die Zeitspanne von zehn Jahren wurde gewählt, um eine ausreichende (im Hinblick auf die Anzahl und Vielfalt der Daten) und ausgewogene (im Hinblick auf jahresweise Schwankungen) Datenbasis zu gewinnen. Die Datenerhebung erfolgte anhand von Publikationslisten der in die Untersuchung einbezogenen Forschenden (für Graz wurde die als vollwertiger Ersatz erachtete Forschungsdatenbank der Universität verwendet, in die sämtliche Forschungsleistungen von den Forschenden selbst eingetragen werden). Alle relevanten Daten wurden in eine eigens für dieses Projekt erstellte Access-Datenbank eingegeben. Erfasst wurden sämtliche Publikationen aller während des Untersuchungszeitraums durchgehend oder auch nur zeitweise an diesen Instituten beschäftigten Forschenden, die sie während der Dauer dieser Beschäftigung verfasst haben.

In die Datenerhebung wurden zwei Gruppen von Forschenden als Repräsentantinnen und Repräsentanten der Institutsleistung einbezogen: 1.) Professorinnen/Professoren – darunter werden hier alle Forschenden verstanden, die einen Professorinnen-/Professorentitel tragen – und 2.) sonstige wissenschaftliche Mitarbeitende – darunter fallen hier alle übrigen Forschenden, die zumindest einen Bachelorabschluss haben und zu deren Aufgaben die Forschung zählt. Nicht einbezogen wurden demnach u. a. studentische Mitarbeitende und Lehrbeauftragte. Ebenso unberücksichtigt blieben Gastprofessorinnen und -professoren, da diese an den beiden betrachteten Instituten in erster Linie Lehrveranstaltungen abgehalten und kaum publiziert haben.

Bezüglich jeder relevanten Publikation wurden zunächst folgende Merkmale erfasst: 1.) Name des betrachteten Instituts (Düsseldorf oder Graz), 2.) Name der betrachteten Person (Autorin/Autor), 3.) Anzahl der Koautorinnen/Koautoren, 4.) Publikationsart, 5.) Sprache, 6.) Umfang in Seiten und – im Falle von Zeitschriftenbeiträgen – 7.) Name der Zeitschrift, in der publiziert wurde.

Danach wurde für die Beiträge in Zeitschriften erhoben, ob sie auch im Web of Science enthalten sind. Ergänzend wurde der Impact-Faktor der jeweiligen Zeitschriften im

Publikationsjahr ermittelt. Andere Datenquellen (z.B. Scopus und Google Scholar) wurden nicht in die empirische Erhebung einbezogen (vgl. 3.1).

4.1.3 Datenauswertung

Es wurden zwei Basisvarianten zur Ermittlung der Forschungsleistung sowie eine Reihe von darauf wirkenden Einflussfaktoren definiert. Basisvariante 1 ist jene, welche der Autorin und den Autoren dieses Beitrages am sinnvollsten erscheint. Es handelt sich dabei um die Ermittlung der Forschungsleistung eines Instituts unter Heranziehung der Publikationslisten (Datenquelle) aller Forschenden des Instituts (Repräsentantinnen und Repräsentanten der Institutsleistung) unter Berücksichtigung der Anzahl dieser Forschenden (Institutsgröße) auf Basis aller Publikationen (Publikationsart), wobei Bücher/Zeitschriftenbeiträge/Beiträge in Tagungsbänden und sonstigen Sammelwerken/Sonstige mit den durchaus willkürlich gewählten Faktoren 10/5/3/1 gewichtet werden (Wertigkeit unterschiedlicher Publikationsarten – Gewichtung 1). Das Ergebnis wird im Falle von mehreren Autorinnen/Autoren durch deren Anzahl dividiert (Koautorinnen-/Koautorenschaft). Publikationssprache, Umfang und Zeitschriftenrankings bleiben unberücksichtigt. In der Folge wird das auf Basis dieser Vorgangsweise ermittelte Ergebnis als Gesamtforschungsleistung pro VZÄ bezeichnet und – wie die Ergebnisse aller Varianten zur Berechnung der Forschungsleistung eines Instituts – in Form von Punkten dargestellt.

Basisvariante 2, in der Folge als Forschungsleistung auf Basis der Zeitschriftenbeiträge pro VZÄ bezeichnet, ist eine auf die Publikationsart „Zeitschriftenbeiträge“ eingeschränkte Version der Basisvariante 1: Diese Einschränkung hat zur Folge, dass die Berücksichtigung der Wertigkeit unterschiedlicher Publikationsarten entfällt. Alle übrigen Einflussfaktoren bleiben unverändert. Basisvariante 2 trägt dem Umstand Rechnung, dass in Teilen der Sozialwissenschaften Zeitschriftenbeiträge die mit Abstand größte Bedeutung haben.

Anschließend wurde auf Grundlage der zwei Basisvarianten die Bedeutung jedes einzelnen der neun betrachteten Einflussfaktoren ermittelt, indem jede Basisvariante jeweils im Hinblick auf diesen einen Faktor variiert wurde. Die so ermittelten Ergebnisse wurden mit der jeweiligen Basisvariante verglichen. Die dabei festgestellten Unterschiede wurden einerseits in Punkten und andererseits in Prozenten angegeben. In einem weiteren Schritt wurde für jeden Einflussfaktor der Durchschnittswert dieser prozentuellen Abweichungen ermittelt. Diese Durchschnittswerte wurden als Gesamtmaß für den Einfluss eines Faktors auf die Ergebnisse einer Forschungsevaluation betrachtet und in der Folge für eine Reihung der Einflussfaktoren nach deren Bedeutung herangezogen.

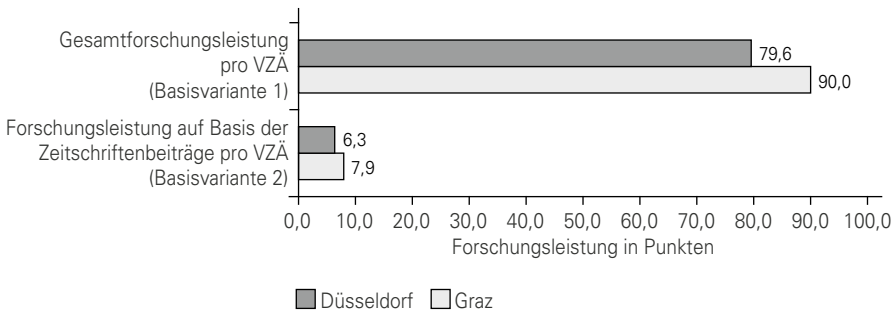
4.2 Ergebnisse

4.2.1 Institutsvergleich auf Grundlage der Basisvarianten

Aus Abbildung 1 geht hervor, dass bei Heranziehung der aus Sicht der Autorin und der Autoren sinnvollsten Vorgangsweise für eine Forschungsevaluation auf Instituts-ebene (Basisvariante 1) die Forschungsleistung von Graz jene von Düsseldorf übertrifft. Dies ändert sich auch nicht im Falle einer Beschränkung der berücksichtigten Publikationsarten auf Zeitschriftenbeiträge (Basisvariante 2).

Eine Analyse des Zustandekommens der Punktwerte im Hinblick auf Basisvariante 1 legt offen, dass Düsseldorf im Betrachtungszeitraum zwar insgesamt beinahe doppelt so viele Publikationen wie Graz aufweist, dafür aber mehr als die doppelten personellen Ressourcen (allerdings mehrheitlich Prae-Docs) zur Verfügung hatte. Hinsichtlich Basisvariante 2 ist dieses Verhältnis – im Zähler stehen diesmal allerdings nur die Zeitschriftenbeiträge, was zu deutlich geringeren Absolutwerten führt – ziemlich ähnlich, woraus sich eine im Vergleich zu Basisvariante 1 ähnliche Relation der Punktwerte zwischen Düsseldorf und Graz ergibt.

Abbildung 1: Institutsvergleich – Basisvarianten



Erläuterungen:

Die Punktwerte repräsentieren die Forschungsleistung eines Instituts: Dabei wurde so vorgegangen, dass bei Basisvariante 1 jede Publikation zunächst mit einem Basiswert von 1 bewertet wurde. Dieser Basiswert wurde dann in Abhängigkeit von Publikationsart und Koautorinnen-/Koautorenschaft modifiziert. Beispielsweise wurde im Falle eines Zeitschriftenaufsatzes, der von zwei Mitarbeitenden des betrachteten Instituts sowie von zwei externen Autorinnen/Autoren verfasst wurde, der Basiswert mit 5 (= gewählte Gewichtung für Zeitschriftenbeiträge) multipliziert und anschließend durch 2 (= 50 % der Koautorinnen/Koautoren gehören nicht zum Institut) dividiert, was einen Punktwert von 2,5 ergibt. Bei Basisvariante 2, die nur Zeitschriftenbeiträge umfasst, entfiel demzufolge die Modifikation in Abhängigkeit von der Publikationsart, sodass der Basiswert von 1 lediglich in Abhängigkeit von vorhandenen Koautorinnen-/Koautorenschaften modifiziert wurde. Dies hat zur Folge, dass die Punktwerte bei Basisvariante 2 im Durchschnitt um ca. 80 % geringer sind als bei Basisvariante 1; vgl. dazu auch 4.1.3).

4.2.2 Bedeutung der Einflussfaktoren – Einzelbetrachtung

Tabelle 2 lässt erkennen, dass die hier untersuchten neun Faktoren die Ergebnisse einer Forschungsevaluation sehr stark beeinflussen können. Während bei beiden Basisvarianten Graz vor Düsseldorf liegt, ist dies bei den Variationen in 12 von 26 Fällen

nicht mehr der Fall. So kommt ein Wechsel der Datenquelle von Publikationslisten auf das Web of Science Düsseldorf deutlich zugute: Bei beiden Varianten ist Düsseldorf nun vor Graz gereiht. Dies liegt vor allem daran, dass die Publikationen von Düsseldorf im Web of Science besser repräsentiert sind als jene von Graz, was wiederum dem größeren Anteil an englischsprachigen Zeitschriften- und Proceedingsbeiträgen geschuldet ist. Allerdings verdeutlicht ein Vergleich der hier relevanten Forschungspunkte mit jenen laut Basis, dass im Hinblick auf beide Institute nur ein eher geringer Anteil aller Publikationen im Web of Science enthalten ist, wodurch ein Vergleich der Gesamtforschungsleistung auf Basis dieser Datenquelle nicht wirklich sinnvoll erscheint. Dennoch dürfte in der Praxis auch in solchen Fällen nicht selten ausschließlich auf das Web of Science zurückgegriffen werden, da keine detaillierteren Daten vorhanden oder verfügbar sind.

Tabelle 2: Bedeutung der einzelnen Einflussfaktoren

Vorgangsweise	Forschungsleistung in Punkten				Differenz in %	
	Basisvariante 1		Basisvariante 2		Basisvariante 1	Basisvariante 2
	Düsseldorf	Graz	Düsseldorf	Graz		
Basis	79,6	<u>90,0</u>	6,3	7,9	13,1 %	25,4 %
Variation des Einflussfaktors 1 (Datenquelle)						
Web of Science (statt Publikationslisten)	<u>13,5</u>	12,7	<u>2,7</u>	2,4	-6,1 %	-11,1 %
Differenz zur Basis					-19,3 %	-36,5 %
Variation des Einflussfaktors 2 (Repräsentantinnen/Repräsentanten der Institutsleistung)						
Professorinnen/Professoren (statt alle Forschenden)	<u>193,2</u>	86,7	<u>20,0</u>	7,0	-55,1 %	-64,9 %
Differenz zur Basis					-68,3 %	-90,2 %
Variation des Einflussfaktors 3 (Institutsgröße)						
Keine Berücksichtigung (statt Division durch VZÄ)	<u>900,5</u>	461,2	<u>71,0</u>	40,3	-48,8 %	-43,2 %
Differenz zur Basis					-61,9 %	-68,6 %
Variation des Einflussfaktors 4 (Publikationsart)						
Alle mit Ausnahme der Sonstigen (statt alle)	74,7	<u>75,8</u>	X	X	1,5 %	X
Differenz zur Basis					-11,6 %	X
Bücher und Zeitschriftenbeiträge (statt alle)	38,0	<u>41,0</u>	X	X	7,8 %	X
Differenz zur Basis					-5,3 %	X
Zeitschriftenbeiträge (statt alle)	31,4	<u>39,3</u>	X	X	25,4 %	X
Differenz zur Basis					12,2 %	X
Beiträge in Proceedings (statt alle)	<u>27,1</u>	20,0	X	X	-25,9 %	X
Differenz zur Basis					-39,0 %	X

Fortsetzung nächste Seite

Fortsetzung von **Tabelle 2**

Vorgangsweise	Forschungsleistung in Punkten				Differenz in %	
	Basisvariante 1		Basisvariante 2		Basisvariante 1	Basisvariante 2
	Düsseldorf	Graz	Düsseldorf	Graz		
Variation des Einflussfaktors 5 (Wertigkeit unterschiedlicher Publikationsarten)						
Keine Berücksichtigung (statt Gewichtung 1)	24,0	<u>33,8</u>	X	X	40,8 %	X
Differenz zur Basis					27,7 %	X
Gewichtung 2 (statt Gewichtung 1)	66,7	<u>76,4</u>	X	X	14,4 %	X
Differenz zur Basis					1,3 %	X
Variation des Einflussfaktors 6 (Koautorinnen-/Koautorenschaft)						
Keine Berücksichtigung (statt 1/n)	103,8	<u>128,0</u>	8,3	<u>12,1</u>	23,3 %	45,7 %
Differenz zur Basis					10,2 %	20,3 %
2/(n+1) (statt 1/n)	102,8	<u>105,0</u>	7,9	<u>9,5</u>	2,1 %	19,8 %
Differenz zur Basis					-11,0 %	-5,6 %
Variation des Einflussfaktors 7 (Publikationssprache)						
Deutschsprachige Publikationen (statt keine Berücksichtigung)	28,2	<u>50,6</u>	2,9	<u>5,2</u>	79,2 %	78,9 %
Differenz zur Basis					66,0 %	53,5 %
Englischsprachige Publikationen (statt keine Berücksichtigung)	<u>51,3</u>	39,4	<u>3,4</u>	2,7	-23,2 %	-20,1 %
Differenz zur Basis					-36,3 %	-45,4 %
Variation des Einflussfaktors 8 (Umfang)						
Seitenzahlen (statt keine Berücksichtigung)	<u>599,4</u>	439,7	75,5	<u>79,0</u>	-26,6 %	4,7 %
Differenz zur Basis					-39,7 %	-20,7 %
Mindestumfang von 5 Seiten (statt keine Berücksichtigung)	65,4	<u>72,4</u>	5,5	<u>7,7</u>	10,6 %	40,7 %
Differenz zur Basis					-2,5 %	15,3 %
Variation des Einflussfaktors 9 (Zeitschriftenrankings)						
Impact-Faktor (statt keine Berücksichtigung)	X	X	<u>2,2</u>	1,2	X	-46,8 %
Differenz zur Basis					X	-72,1 %
Impact-Faktor, sofern mind. 1 (statt keine Berücksichtigung)	X	X	<u>2,1</u>	1,1	X	-47,1 %
Differenz zur Basis					X	-72,5 %

Erläuterungen:

Zu den Punktwerten: vgl. Erläuterungen zu Abbildung 1. In den vier Spalten zur „Forschungsleistung in Punkten“ sind jeweils die Werte jenes der beiden verglichenen Institute unterstrichen, das besser abgeschnitten hat. Weiters sind in der Tabelle jene fünf Varianten hell (dunkel) unterlegt aufgeführt, bei denen Düsseldorf (Graz) am weitesten vor Graz (Düsseldorf) liegt. Zudem geben die in den beiden rechten Spalten der Tabelle angeführten Werte – mit Ausnahme jener in den Zeilen „Differenz zur Basis“ – im Hinblick auf beide Basisvarianten die prozentuellen Ergebnisdifferenzen zwischen Düsseldorf und Graz an; und zwar in der Form, dass die Werte jeweils ausdrücken, um wie viel Prozent Graz besser (= positive Werte) bzw. schlechter (= negative Werte) abscheidet als Düsseldorf. Die in den Zeilen „Differenz zur Basis“ angeführten Werte geben an, um wie viele Prozentpunkte sich die Ergebnisdifferenz bei Heranziehung der jeweiligen Variation gegenüber der Ergebnisdifferenz laut Basis (vgl. Zeile 1) verändert hat. Dabei zeigt ein positiver (negativer) Wert an, dass die jeweilige Variation Graz (Düsseldorf) im Vergleich zur Basisvariante begünstigt.

Die vorgenommene Variation des Einflussfaktors 2 in Form einer ausschließlichen Berücksichtigung der Forschungsleistung der Professorinnen und Professoren führt ebenfalls zu einer drastischen Ergebnisveränderung. Ursache dafür ist, dass es in Düsseldorf während des Betrachtungszeitraums stets nur eine Professorinnen-/Professorenstelle gegeben hat. In Graz waren dagegen immer vier Professoren beschäftigt, deren Einzelproduktivität unter jener des Düsseldorfer Professors lag. Dabei ist allerdings auch die große Anzahl der wissenschaftlichen Mitarbeitenden in Düsseldorf zu beachten, die bei vielen Publikationen des Professors als Koautorinnen und Koautoren aufscheinen und auch aktiv mitgearbeitet haben. Diese Anzahl betrug im Durchschnitt mehr als zehn, gegenüber durchschnittlich einer/einem einzigen wissenschaftlichen Mitarbeitenden in Graz.

Aufgrund der wesentlich umfangreicheren personellen Ressourcen profitiert Düsseldorf natürlich auch von einer Variation des Einflussfaktors „Institutsgröße“ in Form einer unterbleibenden Berücksichtigung. Nach dieser Variante wird die Forschungsleistung von Düsseldorf – im Hinblick auf Basisvariante 1 und 2 – als beinahe doppelt so hoch wie jene von Graz ausgewiesen.

Von den vier vorgenommenen Variationen im Hinblick auf die Publikationsart (Einflussfaktor 4) führt nur eine zu einer Ergebnisdrehung gegenüber der Basisvariante 1: Lediglich bei einem Forschungsvergleich auf alleiniger Basis der Beiträge in Proceedings liegt Düsseldorf vor Graz. Allerdings kommt es auch bei Heranziehung der beiden erstgenannten Varianten zu einer Verbesserung der Stellung von Düsseldorf. Dies deshalb, da in Düsseldorf verhältnismäßig wenige sonstige Beiträge und viel mehr Bücher als in Graz publiziert wurden. Eine Stärke von Graz sind dagegen die Zeitschriftenbeiträge; eine Beschränkung auf diese Publikationsart lässt den Vorsprung von Graz gegenüber der Basisvariante um mehr als zwölf Prozentpunkte steigen.

Berücksichtigt man den Einflussfaktor „Wertigkeit unterschiedlicher Publikationsarten“ überhaupt nicht, so gibt es sogar eine Steigerung zugunsten von Graz um mehr als 27 Prozentpunkte. Ursache dafür ist vor allem die verhältnismäßig hohe Anzahl an sonstigen Publikationen in Graz, die bei dieser Variante beispielsweise den Büchern gleichgestellt wären, was nicht ausgewogen erscheint. Kaum Änderungen gegenüber der Basisvariante gibt es im Falle einer abweichenden Bewertung der einzelnen Publikationsarten, bei der Bücher/Zeitschriftenbeiträge/Beiträge in Tagungsbänden/Beiträge in sonstigen Sammelwerken/Sonstige mit den Faktoren 5/4/3/2/1 gewichtet werden (Gewichtung 2).

Im Hinblick auf die Koautorinnen-/Koautorenschaft profitiert Graz stets von einem „full counting“, da es dort verhältnismäßig mehr Koautorinnen-/Koautorenschaften mit institutsfremden Autorinnen und Autoren gegeben hat als in Düsseldorf, wogegen Düsseldorf die Variante „ $2/(n+1)$ “ entgegenkommt. Dies deshalb, da dort im Falle von Mehrfachautorinnen-/Mehrfachautorenschaften die Anzahl der Autorinnen und Autoren im Durchschnitt höher war, was wiederum zu einer weit über eins liegenden Punkte-summe für eine einzelne Publikation führt.

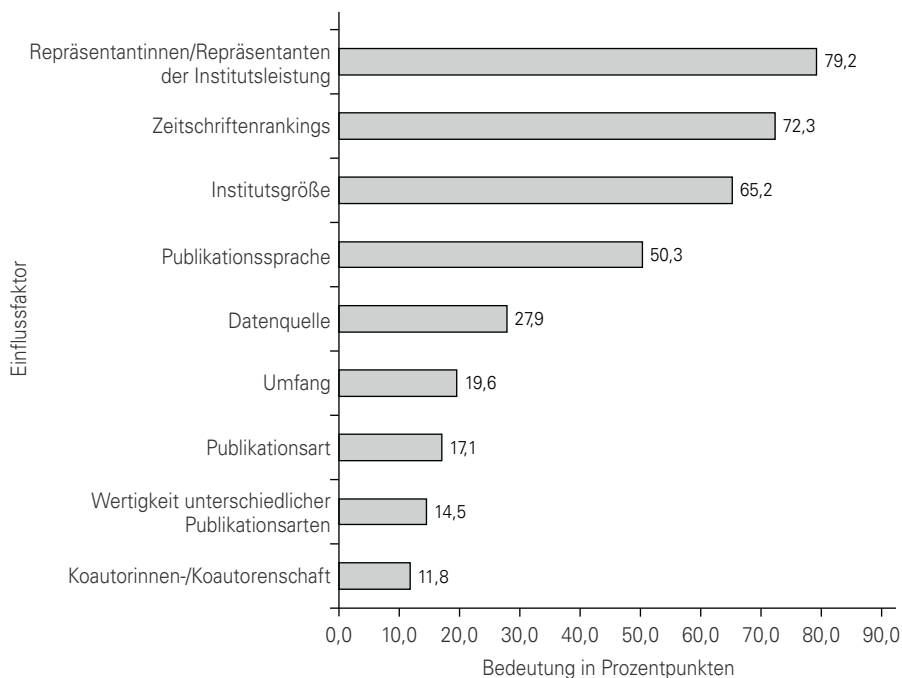
Die vorgenommenen Variationen des Einflussfaktors „Publikationssprache“ haben sehr unterschiedliche Folgen. Hinsichtlich Basisvariante 1 (2) führt eine Beschränkung auf deutschsprachige Publikationen zu einer Erhöhung des Vorsprungs von Graz um 66 (mehr als 53) Prozentpunkte, eine solche auf englischsprachige Publikationen zu einer Verbesserung der Position von Düsseldorf um mehr als 36 (45) Prozentpunkte und damit zu einer Ergebnisdrehung. Insgesamt war im Betrachtungszeitraum nur ca. ein Drittel der Düsseldorfer, aber mehr als die Hälfte der Grazer Publikationen deutschsprachig.

Kommt es zu einer Berücksichtigung des Umfangs der Publikationen (Einflussfaktor 8), so führt dies in drei von vier betrachteten Fällen zu einer Verbesserung für Düsseldorf. Während sich ein Mindestumfang von 5 Seiten als Filter für die Berücksichtigung einer Publikation für Düsseldorf im Hinblick auf Basisvariante 1 nur leicht positiv und im Hinblick auf Basisvariante 2 sogar deutlich negativ auswirkt, führt ein Rückgriff auf die Seitenzahlen in beiden Fällen zu einer starken Verbesserung für Düsseldorf, im Hinblick auf Basisvariante 1 sogar zu einer Ergebnisdrehung. Ursache dafür ist, dass Bücher hier aufgrund ihrer hohen Seitenanzahl von großer Bedeutung sind und in Düsseldorf im Betrachtungszeitraum zehn derartige Publikationen verfasst wurden; in Graz dagegen nur zwei.

Nachdem die Düsseldorfer Publikationen im Web of Science, welches für die Variationen des Einflussfaktors „Zeitschriftenrankings“ als Datenquelle herangezogen wurde, wesentlich besser vertreten sind, ist das gute Abschneiden im Hinblick auf beide Variationen wenig überraschend. In beiden Fällen kommt es zu einer Verbesserung der Positionierung von Düsseldorf um mehr als 70 Prozentpunkte inklusive Ergebnisdrehung.

4.2.3 Bedeutung der Einflussfaktoren – Gesamtbetrachtung

Abbildung 2: Bedeutung der Einflussfaktoren



Erläuterungen:

Die in der Abbildung enthaltenen Werte sind die Durchschnittswerte aller in Tabelle 2 zu einem Einflussfaktor in der/den Zeile/n „Differenz zur Basis“ enthaltenen Werte.

Gemäß Abbildung 2 haben im Hinblick auf das Anwendungsbeispiel vor allem vier Faktoren einen besonders großen Einfluss auf das Ergebnis des Forschungsvergleichs auf Institutsebene: die Repräsentantinnen/Repräsentanten der Institutsleistung sowie die Berücksichtigung von Zeitschriftenrankings, Institutsgröße und Publikationssprache. Diese Reihung ist in erster Linie eine Folge der vorliegenden Datenkonstellation. Zwischen den beiden miteinander verglichenen Instituten lagen erhebliche Unterschiede im Hinblick auf Anzahl und Zusammensetzung der institutzugehörigen Forschenden vor. Während es in Graz stets wesentlich mehr Professoren gab, war die Gesamtzahl der Forschenden in Düsseldorf deutlich höher. Dementsprechend ist es für den Vergleich von großer Bedeutung, ob man einerseits nur die Professorinnen/Professoren oder alle Forschenden als Repräsentantinnen/Repräsentanten der Institutsleistung heranzieht, und ob man andererseits den Forschungsoutput zur Anzahl der Forschenden in Relation setzt oder nicht. Weiters gab es ein sehr unterschiedliches Publikationsverhalten in Bezug auf Publikationssprache und Medienwahl. Während in Graz zwar in Summe mehr Zeitschriftenbeiträge publiziert wurden als in Düsseldorf, waren die Düsseldorfer Zeitschriftenbeiträge viel häufiger in englischer Sprache ver-

fasst und dementsprechend in Zeitschriften publiziert, die einen (höheren) Impact-Faktor aufweisen.

Obwohl die übrigen fünf Einflussfaktoren im Anwendungsbeispiel von geringerer Bedeutung sind, sollten die theoretischen Ausführungen sowie jene zur Bedeutung der einzelnen Einflussfaktoren im Anwendungsbeispiel doch veranschaulicht haben, dass diese Bedeutung in erster Linie von der jeweiligen Datenkonstellation abhängig ist. Je größer die Unterschiede zwischen den verglichenen Untersuchungseinheiten in Bezug auf einen einzelnen Faktor sind, desto stärker ist sein Einfluss auf das Ergebnis eines Forschungsvergleichs.

5 Resümee

Zusammenfassend lässt sich sagen, dass die Ergebnisse einer Forschungsevaluation auf Institutsebene tendenziell sehr stark von der gewählten Vorgehensweise abhängig sein können. Welche der neun hier untersuchten Einflussfaktoren von besonderer Ergebnisrelevanz sind (Hauptforschungsfrage), lässt sich nicht generell beantworten, da dies in erster Linie von der jeweiligen Datenkonstellation abhängig ist. Grundsätzlich gilt, dass in jedem Anwendungsfall jene Faktoren von großer Bedeutung sind, bezüglich der es zwischen den untersuchten Einheiten große Differenzen gibt. Wird beispielsweise die Forschungsleistung von mehreren Instituten miteinander verglichen, die eine unterschiedliche Größe aufweisen und zudem in unterschiedlichen Sprachräumen beheimatet sind, so werden wahrscheinlich die Einflussfaktoren „Institutsgröße“ und „Publikationssprache“ von erheblicher Bedeutung sein. Werden dagegen mehrere ungefähr gleich große Institute miteinander verglichen, die alle im englischsprachigen Raum angesiedelt sind, allerdings unterschiedlichen Fachbereichen mit divergenter Publikationskultur zugehörig sind, so werden die Einflussfaktoren „Publikationsart“, „Wertigkeit unterschiedlicher Publikationsarten“ und „Koautorinnen-/Koautorenschaft“ von besonderer Bedeutung sein.

Im hier analysierten Beispiel, das aus erhebungstechnischen und ressourcenökonomischen Gründen auf zwei Institute beschränkt war, gab es erhebliche Unterschiede hinsichtlich Anzahl und Zusammensetzung des wissenschaftlichen Personals sowie hinsichtlich Publikationskultur. Dementsprechend erwiesen sich vor allem die Einflussfaktoren „Institutsgröße“, „Repräsentantinnen/Repräsentanten der Institutsleistung“, „Publikationssprache“ und „Zeitschriftenrankings“ als besonders bedeutsam. Wählt man hier die Varianten „keine Berücksichtigung der Institutsgröße“, „Beschränkung auf Professorinnen/Professoren als Repräsentantinnen/Repräsentanten der Institutsleistung“, „Beschränkung auf englischsprachige Publikationen“ sowie – im Falle einer Einschränkung auf Zeitschriftenbeiträge – „Berücksichtigung von Zeitschriftenrankings“, so weist Düsseldorf eine wesentlich höhere Forschungsleistung als Graz auf.

Die hier als Basisvariante 1 bezeichnete Vorgangsweise erscheint der Autorin und den Autoren in ihrer Disziplin am sinnvollsten (Nebenforschungsfrage), also eine Forschungsevaluation auf Basis von Publikationslisten unter Einbeziehung aller Forschenden des jeweiligen Instituts, die sich auf alle Arten von Publikationen – diese werden jedoch unterschiedlich gewichtet – bezieht, unabhängig von Publikationssprache, Umfang und – im Falle von Zeitschriftenbeiträgen – Zeitschriftenrankings. Berücksichtigung sollen dagegen die Institutsgröße sowie vorhandene Koautorinnen-/Koautorenschaften finden.

Gäbe es eine Vorgangsweise, die von einer entsprechend großen Anzahl von Fachkundigen im Rahmen einer Disziplin als am sinnvollsten erachtet würde, könnte man versuchen, diese als Standard für Forschungsevaluationen auf Institutsebene zu etablieren, um gleiche und transparente Rahmenbedingungen zu schaffen und eine gute Vergleichbarkeit verschiedener Evaluationen zu gewährleisten. Ein solcher Standard würde gewissermaßen eine Heuristik zur Durchführung von Forschungsevaluationen in dieser Disziplin darstellen (Bornmann & Marewski, 2019). Es gibt zwar durchaus Ansätze in eine derartige Richtung (Bornmann et al., 2012), von einem etablierten Standard kann aber bei weitem noch nicht gesprochen werden. Dagegen gibt es im Hinblick auf eine wissenschaftstheoretisch korrekte Vorgangsweise bei empirischen Studien durchaus etablierte Standards, wie etwa hinsichtlich Validität und Reliabilität (Farrington, 2003). Eventuell würde sich die in diesem Beitrag von der Autorin und den Autoren präferierte Vorgangsweise für Institutervergleiche auf Basis der Forschungsleistung in Form von Publikationen – zumindest für den Bereich der Informationswissenschaft – durchaus als Standard eignen.

Literatur

Albers, S. (2015). What drives publication productivity in German business faculties? *Schmalenbach Business Review*, 67(1), 6–33.

Bauer, B. (2003). Habilitationskriterium Impact-Factor. Wie evaluieren medizinische Fakultäten wissenschaftliche Leistungen von Habilitanden? *medizin-bibliothek-information*, 3(2), 40–43.

Bommer, R. & Ursprung, H. (1998). Spieglein, Spieglein an der Wand: Eine publikationsanalytische Erfassung von Forschungsleistungen volkswirtschaftlicher Fachbereiche in Deutschland, Österreich und der Schweiz. *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 118(1), 1–28.

Bornmann, L., Bowman, B., Bauer, J., Marx, W., Schier, H. & Palzenberger, M. (2012). Standards für die Anwendung der Bibliometrie bei der Evaluation von Forschungsinstituten im Bereich der Naturwissenschaften. *Zeitschrift für Evaluation*, 11(2), 233–260.

Bornmann, L. & Marewski, J. (2019). Heuristics as conceptual lens for understanding and studying the usage of bibliometrics in research evaluation. *Scientometrics*, 120(2), 419–459.

Bornmann, L. & Marx, W. (2013). Vorschläge für Standards zur Anwendung der Szientometrie bei der Evaluation von einzelnen Wissenschaftler(inne)n im Bereich der Naturwissenschaften. *Zeitschrift für Evaluation*, 12(1), 103–127.

Böttcher, F. & Thiel, F. (2018). Evaluating research-oriented teaching: a new instrument to assess university students' research competences. *Higher Education*, 75(1), 91–110.

Clermont, M. & Schmitz, C. (2008). Erfassung betriebswirtschaftlich relevanter Zeitschriften in den ISI-Datenbanken sowie der Scopus-Datenbank. *Zeitschrift für Betriebswirtschaft*, 78(10), 987–1010.

Dilger, A. & Müller, H. (2012). Ein Forschungsleistungsranking auf der Grundlage von Google Scholar. *Zeitschrift für Betriebswirtschaft*, 82(10), 1089–1105.

Dorsch, I., Askeridis, J. & Stock, W. (2018). Truebounded, overbounded, or underbounded? Scientists' personal publication lists versus lists generated through bibliographic information services. *Publications*, 6(1), 1–9.

Dorsch, I., Schlögl, C., Stock, W. & Rauch, W. (2017). Forschungsthemen der Düsseldorfer und Grazer Informationswissenschaft (2010 bis 2016). *Information. Wissenschaft & Praxis*, 68(5–6), 320–328.

Dyckhoff, H., Thieme, A. & Schmitz, C. (2005). Die Wahrnehmung deutschsprachiger Hochschullehrer für Betriebswirtschaft in der internationalen Forschung. Eine Pilotstudie zu Zitationsverhalten und möglichen Einflussfaktoren. *Die Betriebswirtschaft*, 65(4), 350–372.

Farrington, D. (2003). Methodological quality standards for evaluation research. *The ANNALS of the American Academy of Political and Social Science*, 587(1), 49–68.

Fraja, G., Facchini, G. & Gathergood, J. (2019). Academic salaries and public evaluation of university research: Evidence from the UK Research Excellence Framework. *Economic Policy*, 34(99), 523–583.

Friedländer, M. (2014). Informationswissenschaft an deutschsprachigen Universitäten – eine komparative informetrische Analyse. *Information. Wissenschaft & Praxis*, 65(2), 109–119.

Fülbier, R. & Weller, M. (2011). A glance at German financial accounting research between 1950 and 2005: A publication and citation analysis. *Schmalenbach Business Review*, 63(1), 2–33.

Garfield, E. (2006). The History and Meaning of the Journal Impact Factor. *Journal of the American Medical Association*, 295(1), 90–93.

Gingras, Y. (2016). *Bibliometrics and Research Evaluation: Uses and Abuses*. Cambridge: The MIT Press.

- Halevi, G., Moed, H. & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation – Review of the Literature. *Journal of Informetrics*, 11(3), 823–834.
- Harzing, A.-W. & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251–261.
- Jackson, B. (2015). University Rankings: How Well Do They Measure Library Service Quality? *Portal: Libraries & the Academy*, 15(2), 315–330.
- Jappe, A. (2020). Professional standards in bibliometric research evaluation? A meta-evaluation of European assessment practice 2005–2019. *PLoS ONE*, 15(4), e0231735.
- Kamrani, P., Dorsch, I. & Stock, W.G. (2020). Publikationen, Zitationen und H-Index im Meinungsbild deutscher Universitätsprofessoren. *Beiträge zur Hochschulforschung*, 42(3), 78–98.
- Kember, D. & Ginns, P. (2012). *Evaluating Teaching and Learning: A practical handbook for colleges, universities and the scholarship of teaching*. New York: Routledge.
- Kladroba, A., Buchmann, T., Friz, K., Lange, M. & Wolf, P. (2021). *Indikatoren für die Messung von Forschung, Entwicklung und Innovation. Steckbriefe mit Hinweisen zur Anwendung*. Wiesbaden: Gabler.
- Marina, S., Davis-Hamilton, Z. & Charmanski, K. (2015). Evaluating Research Administration: Methods and Utility. *Journal of Research Administration*, 46(2), 95–114.
- Martin-Martín, A., Orduna-Malea, E., Thelwall, M. & López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1160–1177.
- Meyer, M., Waldkirch, R. & Zaggl, M. (2012). Relative performance measurement of researchers: the impact of data source selection. *Schmalenbach Business Review*, 64(4), 308–330.
- Moed, H. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.
- Morgan, K. (2004). The research assessment exercise in English universities, 2001. *Higher Education*, 48(4), 461–482.
- Olcay, G. & Bulu, M. (2017). Is measuring the knowledge creation of universities possible?: A review of university rankings. *Technological Forecasting and Social Change*, 123(C), 153–160.
- Packalen, M. & Bhattacharya, J. (2017). Neophilia ranking of scientific journals. *Scientometrics*, 110(1), 43–64.

Priem, J. (2014). Altmetrics. In B. Cronin & C. Sugimoto (Hrsg.), *Beyond bibliometrics: harnessing multidimensional indicators of performance* (S. 263–287). Cambridge, MA: MIT Press.

van Raan, A. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics* 62(1), 133–143.

Reichert, S. (2019). Chancen und Grenzen der Hochschulevaluation. *fteval Journal for Research and Technology Policy Evaluation*, 49(Dec.), 42–47.

Reichmann, G. (2007). Evaluierung von Studienplänen. Ermittlung von Möglichkeiten zur Verbesserung von Studienbedingungen. *Zeitschrift für Hochschulrecht, Hochschulmanagement und Hochschulpolitik*, 6(2), 37–45.

Reichmann, G. (2020). Was zählt eine Publikation? Möglichkeiten zur Durchführung einer Forschungsevaluierung im Bereich der Betriebswirtschaftslehre. In I. Dorsch, K. Fietkiewicz, A. Ilhan, C. Meschede & T. Siebenlist (Hrsg.), *Facetten von Wolf Stock und ihre Bedeutung für die Informationswissenschaft. Festschrift zu Ehren von Wolfgang G. Stock* (Schriften zur Informationswissenschaft, Band 73, S. 15–34). Glückstadt: Verlag Werner Hülsbusch.

Reichmann, G. & Sommersguter-Reichmann, M. (2010). Efficiency measures and productivity indexes in the context of university library benchmarking. *Applied Economics*, 42(1–3), 311–324.

de Rijcke, S., Wouters, P., Rushforth, A., Franssen, T. & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use – a literature review. *Research Evaluation*, 25(2), 161–169.

Röbken, H. (2011). Forschungsproduktivität von Wissenschaftlern und Wissenschaftlerinnen – Eine empirische Analyse von Publikationsaktivitäten vor und nach der Berufung. *Beiträge zur Hochschulforschung*, 33(3), 62–81.

Rost, K. & Frey, B. (2011). Quantitative and qualitative rankings of scholars. *Schmalenbach Business Review*, 63(1), 63–91.

Rousseau, R., Egghe, L. & Guns, R. (2018). *Becoming Metric-wise: A Bibliometric Guide for Researchers*. Cambridge, MA: Chandos.

Schlögl, C. (2013). Internationale Sichtbarkeit der europäischen und insbesondere der deutschsprachigen Informationswissenschaft. *Information. Wissenschaft & Praxis*, 64(1), 1–8.

Schlögl, C. (2014). Usage versus citation behaviours in four subject areas. *Scientometrics*, 101(2), 1077–1095.

Schneijderberg, C. (2018). *Promovieren in den Sozialwissenschaften: Eine sozialisations-theoretische Erschließung des Forschungsfeldes Promotion*. Wiesbaden: Springer.

Schrader, U. & Hennig-Thurau, T. (2009). VHB-JOURQUAL2: Method, Results, and Implications of the German Academic Association for Business Research's Journal Ranking. *Business Research*, 2(2), 180–204.

Sieweke, J., Muck, J., Süß, S. & Haucap, J. (2014). Forschungsevaluation an Universitäten – Ergebnisse einer explorativen Studie rechts- und wirtschaftswissenschaftlicher Fakultäten. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 66(4), 274–305.

Sivertsen, G., Rousseau, R. & Zhang, L. (2019). Measuring scientific contributions with modified fractional counting. *Journal of Informetrics*, 13(2), 679–694.

Stock, W. (2000). Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung. In K. Fuchs-Kittowski, H. Laitko, H. Parthey & W. Umstätter (Hrsg.), *Wissenschaft und Digitale Bibliothek. Wissenschaftsforschung Jahrbuch 1998* (S. 239–282). Berlin: Gesellschaft für Wissenschaftsforschung.

Stockemer, D. & Wigginton, M. (2019). Publishing in English or another language: An inclusive study of scholar's language publication preferences in the natural, social and interdisciplinary sciences. *Scientometrics*, 118(2), 645–652.

Tost, C. & Rindermann, H. (2016). Evaluation psychologischer und pädagogischer Forschungsleistungen mithilfe der bibliometrischen Datenbanken Scopus und Google Scholar. *Zeitschrift für Evaluation*, 15(2), 241–267.

Vinkler, P. (2010). *The Evaluation of Research by Scientometric Indicators*. Oxford: Chandos Publishing.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391.

Wolf, J., Rohn, A. & Macharzina, K. (2005). Institution und Forschungsproduktivität. Befunde und Interpretationen aus der deutschsprachigen Betriebswirtschaftslehre. *Die Betriebswirtschaft*, 65(1), 62–77.

Manuskript eingegangen: 05.03.2021

Manuskript angenommen: 08.10.2021

Angaben zur Autorin und den Autoren:

ao. Prof. DDr. Gerhard Reichmann
Universität Graz, Institut für Operations und Information Systems
Universitätsstraße 15/F3
8010 Graz
Österreich
E-Mail: gerhard.reichmann@uni-graz.at

ao. Prof. Dr. Christian Schlögl
Universität Graz, Institut für Operations und Information Systems
Universitätsstraße 15/F3
8010 Graz
Österreich
E-Mail: christian.schloegl@uni-graz.at

Prof. Dr. Wolfgang G. Stock
Heinrich-Heine-Universität Düsseldorf, Abteilung für Informationswissenschaft
Universitätsstraße 1
40225 Düsseldorf
Universität Graz, Institut für Operations und Information Systems
Universitätsstraße 15
8010 Graz
Österreich
E-Mail: wolfgang.stock@hhu.de

Dr. Isabelle Dorsch
Heinrich-Heine-Universität Düsseldorf, Abteilung für Informationswissenschaft
Universitätsstraße 1
40225 Düsseldorf
E-Mail: isabelle.dorsch@hhu.de

Gerhard Reichmann ist seit 1997 an der Sozial- und Wirtschaftswissenschaftlichen Fakultät der Karl-Franzens-Universität Graz beschäftigt. Seine aktuellen Forschungsschwerpunkte sind Evaluation (Evaluation von universitärer Forschung, Lehre und Verwaltung, Evaluation im Gesundheitsbereich), Informationskompetenz und Informationsrecht (Datenschutzrecht, Urheberrecht und Plagiatsproblematik).

Christian Schlögl ist seit 1990 an der Sozial- und Wirtschaftswissenschaftlichen Fakultät der Karl-Franzens-Universität Graz beschäftigt, seit 2005 ist er stellvertretender wissenschaftlicher Leiter des an den Universitäten Wien, Innsbruck und Graz eingerichteten interuniversitären Universitätslehrgangs „Master of Science (MSc) in Library and Information Studies“. Seine aktuellen Forschungsgebiete sind Bibliometrie und Szientometrie, Informationskompetenz sowie Informations- und Wissensmanagement.

Wolfgang G. Stock ist Informationswissenschaftler. Er war bis zu seiner Pensionierung Mitte 2019 Leiter der Abteilung für Informationswissenschaft in Düsseldorf, danach wechselte er als Gastprofessor nach Graz. Eines seiner Forschungsgebiete sind Infometrie und Szientometrie.

Isabelle Dorsch ist wissenschaftliche Mitarbeiterin der Abteilung für Informationswissenschaft in Düsseldorf. Sie forscht in den Bereichen Social Media und Szientometrie.