

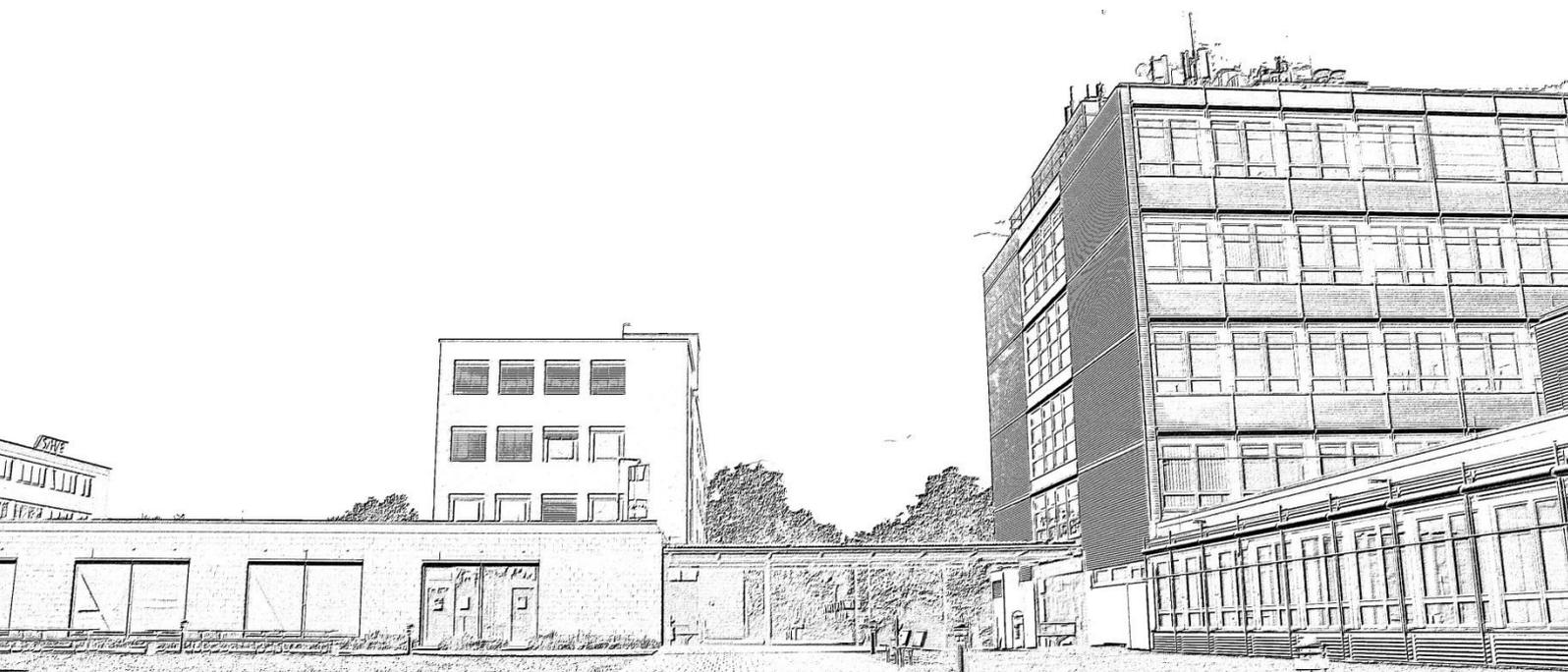
Arbeitspapier der Hochschule Ludwigshafen am Rhein

Nr. 5 – September 2017

What gets measured gets managed!?

Messbarkeitsprobleme des Leistungscontrollings am Beispiel des Hochschulwesens

Harry Müller



ISSN: 2365-7111

Kontakt:

Prof. Dr. Harry Müller

Professur für Allgemeine BWL, insbes. Rechnungswesen und Investitionswirtschaft

Fachbereich II

Hochschule Ludwigshafen am Rhein

Ernst-Boehe-Str. 4

67059 Ludwigshafen

Email: harry.mueller@hs-lu.de

www.hs-lu.de/arbeitspapiere

What gets measured gets managed!?

Messbarkeitsprobleme des Leistungscontrollings am Beispiel des Hochschulwesens

Abstract:

Das Controlling von Leistungsmaßen ist für das Management von betrieblichen Wertschöpfungsprozessen von großer Bedeutung, jedoch auch mit wesentlichen Messbarkeitsproblemen behaftet. Der vorliegende Beitrag untersucht dies am Beispiel des Hochschulwesens, wo sich die Schwierigkeiten und Grenzen des Performance Measurements anschaulich beschreiben lassen. Dazu werden zunächst allgemein die Rolle des Leistungscontrollings und seine Probleme in der Praxis herausgearbeitet. Am Beispiel der Hochschule wird für die Bereiche Lehre und Forschung gezeigt, dass das Abbilden guter Lehre und guter Forschung mit Hilfe von griffigen Kennzahlen häufig problematisch ist. Dabei wird deutlich, dass auch vermeintlich simple Indikatoren und erprobte Ansätze Fehlanreize für Hochschullehrer produzieren können, die zu relevanten Steuerungsproblemen führen können. Dieser Befund ist nicht in den speziellen Rahmenbedingungen des Hochschulbereichs begründet, sondern verweist auf ein grundsätzliches Problem des Leistungscontrollings. Hier fordert der Beitrag einen sensibleren Umgang mit den Methoden des Performance Measurements gerade auch in der privatwirtschaftlichen Praxis.

Keywords:

Leistungscontrolling, Performance Measurement, Hochschulmanagement, Rankings, Evaluation

Kontakt:

Prof. Dr. Harry Müller, Professur für Allgemeine BWL, insbes. Rechnungswesen und Investitionswirtschaft, Fachbereich II, Hochschule Ludwigshafen am Rhein, Ernst-Boehe-Str. 4, 67059 Ludwigshafen, Email: harry.mueller@hs-lu.de

www.hs-lu.de/arbeitspapiere

Inhaltsverzeichnis

1. Rolle des Leistungscontrollings	1
2. Messbarkeitsprobleme	3
3. Beispiel Hochschule	4
3.1. Lehrleistungen	5
3.2. Forschungsleistungen	6
4. Schlussfolgerungen	7
Literaturverzeichnis	9

1. Rolle des Leistungscontrollings

In dem Peter Drucker zugeschriebenen Diktum spiegelt sich ein zentraler Anspruch des Controllings und des wertorientierten Managements. „What gets measured gets managed“ verweist auf die Notwendigkeit der analytischen Durchdringung und Operationalisierung eines Sachverhalts als Voraussetzung für ein erfolgreiches, unternehmerisches Gestalten desselben. Betriebliche Funktionsbereiche müssen nicht nur verbalanalytisch und im Hinblick auf mögliche Gestaltungs- und Optimierungsoptionen untersucht werden. Die Ergebnisse einer solchen Analyse seien nach dieser in der Praxis weitverbreiteten Lesart der Unternehmensführung zudem in Form von numerischen Indikatoren, vorzugsweise skalierbar, abzubilden. In einer an klar definierten Zielen ausgerichteten Managementumwelt wird Messbarkeit zur Voraussetzung für Gestaltbarkeit im Sinne eines Management by Objectives (vgl. hierzu Odiorne 1965; Greenwood 1981). Ziele zu setzen und den Grad ihrer Erreichung zu bestimmen setzt voraus, dass diese in Form von Indikatoren oder besser noch Kennzahlen beschrieben werden können.

Auf diese Weise wird die Beziehung zwischen Measurement und Management zu einer gegenseitigen Bedingtheit. Was gemessen wird, wird gemanagt und damit abgearbeitet (so finden sich für das einleitende Zitat ebenfalls Belege in der Lesart „what gets measured gets done“). Kennzahlen und ihr Soll-Ist-Vergleich identifizieren Problembereiche, die es zu adressieren gilt. Umgekehrt wird Messbarkeit zur Voraussetzung für die Wahrnehmung eines Problems durch die Unternehmensführung. Was nicht anhand von Indikatoren abbildbar ist, ist als Gegenstand des Managements vielfach nicht existent. Sachverhalte, die sich nicht mit Hilfe von Kennzahlen operationalisieren lassen, finden nur schwer Eingang in das Reporting und werden dadurch von der Unternehmensleitung nicht oder nur sehr peripher wahrgenommen.

Zentrale Aufgabe des Controllings ist es in diesem Umfeld, zunächst das Management mit entscheidungsrelevanten Informationen zu versorgen. Auch wenn neuere Controllingkonzepte, wie bspw. die Betrachtung des Controllings als Rationalitätssicherung der Führung (vgl. Weber und Schäffer 1999), über diesen Anspruch hinausgehen, so schließen sie ihn doch als wesentlichen Bestandteil immer mit ein. Was nun mögliche Beobachtungs- und Messobjekte angeht, so liegt innerhalb des Controllings der Fokus traditionell stark auf der Kostenseite. Die Frage nach dem situationsadäquaten Umgang mit Gemeinkosten in betrieblichen Entscheidungsproblemen beschäftigt das Rechnungswesen bereits seit Schmalenbach (1928), während Interdependenzen auf der Leistungs- und Erlösseite traditionell eher wenig im Fokus stehen. Auch wenn es sich hier nicht um einen weißen Fleck handelt, ist das Leistungscontrolling verglichen mit der Kostenrechnung in Konzeption und Praxis eher rudimentär ausgebildet.

Dabei ist es in vielen Fällen Voraussetzung für ein wirksames Ergebniscontrolling, das nicht nur auf der Ebene von Umsatzerlösen stehenbleibt, sondern tiefergehende Erkenntnisse auch im Hinblick auf Vor- oder Hilfsprozesse im Betrieb generieren möchte. Während auf der Kostenseite eine Vielzahl von Konzepten etwa für eine vertiefende Abweichungsanalyse vorhanden ist, bleibt ein entsprechendes Pendant auf der Leistungsseite häufig noch unausgereift, wenngleich unter dem Stichwort Performance Measurement in den vergangenen Dekaden durchaus Schritte in diese Richtung gegangen wurden (vgl. Neely et al. 1995; Gleich 1997; Seiter 2006; Arnaboldi und Azzone 2010; Simons 2000).

Auf der Ebene von Reportings und Kennzahlen sind multidimensionale Perspektiven wie etwa in der etablierten Balanced Scorecard von Kaplan und Norton (1996) inzwischen üblich. Ein sinnvolles Verfolgen des ökonomischen Prinzips ist schließlich ohne Blick auf Leistungsmaße schlichtweg unmöglich, da ein reiner Fokus auf Kostengrößen etwaige Qualitätseinbußen auf der Leistungsseite unberücksichtigt ließe. Umso wichtiger ist es daher, die Datengrundlage des Leistungscontrollings kritisch zu untersuchen und qualitativ abzusichern. Wird nämlich im Performance Measurement das Falsche gemessen, können Kennzahlen und Balanced Scorecard auch zu falschen Managemententscheidungen führen.

Im Performance Measurement können konzeptionell vier Messebenen unterschieden werden (vgl. Brown und Svenson 1988; Weber und Großklaus 1995; Weber und Schäffer 2016):

1. Input
2. Prozess
3. Output
4. Outcome

Den weitesten Zusammenhang zur eigentlich im Fokus stehenden Leistung weisen die in den Erstellungsprozess einfließenden Inputgrößen auf. Beispiele sind Arbeitsstunden oder Materialaufwand. Die Prozessperspektive bildet die ordnungsgemäße Durchführung der Leistungserstellung ab, etwa ob alle notwendigen Arbeitsschritte richtig durchgeführt wurden. Der Output beschreibt das eigentliche Leistungsergebnis, beispielsweise die Stückzahl und/oder Qualitätseigenschaften eines produzierten Gegenstandes oder einer Dienstleistung. Der Outcome schließlich zielt auf hinter der Leistung stehende Ziele und Größen wie etwa eine langfristig profitable und stabile Kundenbeziehung.

Die Frage, welche Ebene für das Controlling eines Geschäftsprozesses die richtige ist, ergibt sich aus dessen Struktur selbst, genauer aus dem Grad der Beobachtbarkeit des Ergebnisses. Ein präzises Abbilden des Outcomes wäre hier selbstverständlich ideal, dürfte jedoch in der Realität nur selten zu finden sein. In Unternehmen ist auf der Ebene des Umsatzes eine Messung des Outputs für bestimmte Produkte zumindest auf den ersten Blick unproblematisch. Eine Quantifizierung des Outcomes, also etwa der Kundenzufriedenheit und dem Potenzial für weitere profitable Geschäftsbeziehungen dürfte schwierig werden, obwohl dies durchaus lohnend erscheint. Geht man jedoch entlang der Prozesskette zurück, wird deutlich, dass Vorprozesse in der Produktion, z. B. die Qualität von Zulieferteilen, oder sekundären Geschäftsprozesse kaum auf der Ebene des Outputs beobachtbar sind. Hier muss hilfsweise auf die Prozess- oder sogar auf die Inputebene zurückgegriffen werden (vgl. Weber und Schäffer 2016). Nicht ohne Grund ist der Standardfall des Arbeitsentgelts der Zeitlohn und damit eine am Inputniveau orientierte Größe, da Prozessqualität, Output oder gar Outcome der von den Mitarbeiterinnen und Mitarbeitern¹ erbrachten Leistungen in der Regel nicht sinnvoll beobachtbar sind.

Die große Bedeutung von leistungsfähigen und auf einer situationsadäquaten Ebene aufsetzenden Indikatoren des Performance Measurements wird mit Blick auf die Aufgaben des Leistungscontrollings deutlich: Die Leistungsmessung ist notwendige Grundlage für ein effektives Planungs- und Kontrollsystem und damit für die Setzung von Anreizen für die erbringenden Bereiche, wirtschaftlich im Sinne

¹ Aus Gründen der besseren Lesbarkeit wird im Folgenden das generische Maskulinum verwendet. Sofern nicht anders gekennzeichnet, schließt dies die weibliche Form mit ein.

eines effizienten Ziel-/Mittelverhältnisses zu arbeiten. So werden Bereiche für ein zielorientiertes Management zugänglich, die sich einer unmittelbaren, monetären Quantifizierung mangels Marktumsätzen entziehen (vgl. Greenwood 1981; Gleich 1997; Rommelspacher et al. 2006).

Die Formulierung und Ermittlung von Leistungsgrößen hilft zudem, sogenannte Gemeinkostenbereiche analytisch besser zu durchdringen und bildet die Grundlage für eine Zurechnung der Kosten. Außerdem ist ein Leistungscontrolling näher an den operativen Steuerungsproblemen der jeweiligen Bereiche. Leistungsmaße verweisen unmittelbar auf betriebliche Abläufe und Prozesse, während etwa Prozesskosten lediglich deren kostenmäßige Folgen beschreiben.

2. Messbarkeitsprobleme

Ein wirksames Leistungscontrolling setzt jedoch voraus, dass geeignete Indikatoren existieren, die zwar nicht selbst mit der Leistung identisch sein müssen, diese jedoch adäquat abbilden. Infrage kommende Indikatoren müssen dabei drei zentralen Anforderungen genügen:

1. Beobachtbarkeit
2. Validität
3. Anreizkompatibilität

Beobachtbarkeit ist dadurch charakterisiert, dass der Leistungsindikator von einem unabhängigen Dritten (d. h. nicht allein durch den Leistungserbringer selbst) beobachtet und gemessen werden kann. Validität setzt voraus, dass die Messergebnisse reproduzierbar und damit für Außenstehende nachvollziehbar sind. Anreizkompatibilität als in der Praxis anspruchsvollste der drei Anforderungen fordert schließlich, dass die Indikatoren dazu geeignet sind, menschliches Verhalten in dem jeweiligen Leistungsbereich effektiv, d. h. in die angestrebte Richtung, zu steuern. Die zentrale Bedeutung dieser dritten Anforderung zeigt sich besonders anschaulich in den systemimmanenten Steuerungsproblemen der ehemaligen sozialistischen Zentralverwaltungswirtschaften Mittel- und Osteuropas (vgl. Lachmann 2004). Die aus mangelnder Anreizkompatibilität resultierenden Fehlsteuerungen werden unter Verweis darauf auch als „Tonnenideologie“ kritisiert.

Ebendieses Mess- und Anreizproblem besteht grundsätzlich immer dort, wo Transaktionen nicht durch Marktmechanismen gesteuert werden, d. h. sowohl in weiten Teilen der Volkswirtschaft als auch innerhalb von Unternehmen: Außerhalb des gewerblichen Bereichs, etwa bei vielen freien Berufen oder im Bereich von Non-Profit-Organisationen und öffentlichen Betrieben, sind die eingangs erläuterte Input- oder Prozessperspektive und darauf aufbauende Governance-Mechanismen tendenziell die Regel. In gewisser Hinsicht begründet die Nichtbeobachtbarkeit einer Leistung im Fall von sogenannten Vertrauensgütern die Verortung der Leistungserstellung im nicht rein marktorientierten Sektor oder zumindest eine strikte Regulierung desselben. Beispiele sind medizinische Leistungen, anwaltliche Tätigkeit, Kulturförderung oder der Bildungsbereich.

Außerhalb des Non-Profit- und des öffentlichen Sektors sind bei privatwirtschaftlichen Unternehmen die internen Vorleistungsprozesse und alle sekundären Geschäftsprozesse (d. h. Personal, Marketing, Finanzierung etc.) von dieser Problematik betroffen. Interne Verrechnungspreise lösen das Messbarkeitsproblem nicht, sondern setzen umgekehrt die Definition und Messung von validen Leistungseinheiten voraus, die dann in einem zweiten Schritt mit einem Preis versehen werden können. Neben den

genannten Bereichen stellt sich das Indikatorenproblem zudem bei Querschnittsfunktionen wie etwa einem Qualitätsmanagement oder der gezielten Entwicklung von Potenzialfaktoren und Beziehungen im Unternehmensumfeld, etwa zu Kunden, potenziellen Mitarbeitern, Öffentlichkeitsarbeit etc.

3. Beispiel Hochschule

Dass die Frage der Messbarkeit von Outputindikatoren keineswegs eine rein akademische ist und sich die Problematik in ihrer Tiefe oft erst auf den zweiten Blick erschließt, lässt sich anschaulich am Beispiel des Hochschulwesens verdeutlichen. Die Leistungserbringung von Hochschulen lässt sich grundsätzlich in zwei bzw. drei große Bereiche gliedern:

1. Lehre
2. Forschung
3. Third Mission

In der Lehre vermitteln Hochschulen Wissen an ihre Studierenden, im Bereich Forschung generiert die Hochschule neues Wissen, wobei an Universitäten der Fokus stärker auf Grundlagenforschung liegt, während Fachhochschulen der anwendungsorientierten Forschung den Vorzug geben. Ob darüber hinaus noch eine Third Mission der Hochschulen existiert, die sich auf Interaktionen der Hochschule mit der Gesellschaft außerhalb von Forschung und Lehre bezieht, ist umstritten und soll an dieser Stelle nicht weiter thematisiert werden. Hier kann auf die einschlägigen Schriften im Umfeld des Centrums für Hochschulentwicklung der Bertelsmann-Stiftung (vgl. Roessler et al. 2015a, 2015b; Hachmeister et al. 2016a; Hachmeister et al. 2016b; Roessler et al. 2016; Schneidewind 2016) und auf die nicht unbegründete Kritik daran (vgl. Shore und McLauchlan 2012; Watson und Hall 2015; Bacevic 2017) verwiesen werden. Zur Untersuchung der Probleme des Leistungscontrollings werden sich die folgenden Ausführungen auf die unstrittigen Felder Forschung und Lehre konzentrieren.

Sinn und Unsinn eines Leistungscontrollings der an Hochschulen erbrachten Forschungs- und Lehrleistungen ist vielfach umstritten, in vielen konkreten Entscheidungs- und Führungssituationen jedoch notwendige Voraussetzung für ein transparentes Management. Zur Sicherung der Qualität in Forschung und Lehre, im Rahmen von Berufungsverfahren, zur Vergabe von Forschungs- oder Lehrpreisen, im Rahmen der W-Besoldung und der leistungsorientierten Mittelvergabe sind belastbare Aussagen zum erbrachten Output unverzichtbar. Daneben werden auch von Dritten, z. B. Studierenden bei der Hochschulauswahl oder von Unternehmen auf der Suche nach geeigneten Forschungspartnern, die Ergebnisse eines Leistungscontrollings nachgefragt.

Dies geschieht in Form von Rankings, die in der Terminologie des Controllings nichts anderes als ein aggregiertes Reporting auf Basis des Leistungscontrollings darstellen. Sie knüpfen dabei an verschiedenen Dimensionen der Leistungserbringung an, die als Rankingkriterium genutzt werden (vgl. Müller 2013). Lehrrankings können sich etwa auf die Ergebnisse von Lehrevaluationen oder Befragungen stützen, alternativ und insbesondere im angelsächsischen Raum beliebt ist auch das Anknüpfen an den späteren Verdienst der Absolventen als Leistungsindikator für den Lehrerfolg. Im Forschungsbereich werden entweder die erstellten Publikationen betrachtet oder es wird ihre Rezeption anhand der auf sie entfallenden Zitationen ermittelt. Beide Kennzahlen sind neben der Summe der eingeworbenen Drittmittel gängige Indikatoren für die Forschungsleistung, wobei die Drittmittel im Grunde keinen

Output, sondern einen Input darstellen und allenfalls indirekt als Leistungsmaß interpretiert werden können.

Jedes dieser Rankingkriterien kann nun auf verschiedenen Beobachtungsebenen betrachtet werden, die zugleich die Rankingobjekte bestimmen. Neben der Datenerhebung und Auswertung auf der Ebene einzelner Hochschullehrer/-innen können die Ergebnisse auf Fachbereichs-, Hochschul- oder Länder-ebene aggregiert werden, um eine Aussage zur Performance eines Kollektivs zu machen.

Im Folgenden soll für die Bereiche Lehre und Forschung exemplarisch untersucht werden, inwieweit in der Praxis gängige und teilweise seit Jahrzehnten etablierte Indikatoren den oben ausgeführten Anforderungen genügen. Messen die Indikatoren also wirklich das, was sie zu messen vorgeben?

3.1. Lehrleistungen

Das in der Hochschulpraxis wohl wichtigste Instrument des Lehrcontrollings sind die teilweise verpflichtend organisierten Lehrevaluationen. Als Methode hat sich hier die Befragung von Studierenden mittels eines standardisierten Fragebogens etabliert. Die Ergebnisse werden sowohl innerhalb der Fachbereiche erhoben als auch für ganze Hochschulen, beispielsweise vom Centrum für Hochschulentwicklung der Bertelsmann-Stiftung. Sie bilden eine wesentliche Grundlage des CHE-Hochschulrankings, das von der Wochenzeitung DIE ZEIT publiziert wird und dem im Hinblick auf potenzielle Studierende ein durchaus relevanter Einfluss nachgesagt wird. Die Aussagekraft dieser Rankings als Reporting- und Kennzahlensysteme hängt notwendigerweise von der Qualität der zugrundeliegenden Befragungsdaten ab.

In diesem Zusammenhang muss auf eine einschlägige Studie von Felton et al. (2008) verwiesen werden, die anhand eines umfangreichen statistischen Samples nachweist, dass die Ergebnisse von Lehrevaluationen („Overall Quality“) wesentlich von der „Easiness“ der Kurse (von den Studierenden wahrgenommenes Verhältnis von Noten und Lernaufwand) und der persönlichen Attraktivität der Lehrenden („Hotness“) abhängt. Dieser Zusammenhang wurde in weiteren Untersuchungen bestätigt (u. a. Timmerman 2008; Freng und Webber 2009; Spooen et al. 2013). Verallgemeinernd kann man daraus folgern, dass die Ergebnisse von Lehrevaluationen nicht ausschließlich veranstaltungsimmanente Eigenschaften widerspiegeln, sondern in wesentlichem Maße auch Ausdruck der Attraktivität des Lehrpersonals und seines Anspruchs in Prüfungssituationen sind.

Das Ergebnis mag auf den ersten Blick skurril wirken, Probleme ergeben sich jedoch vor allem im Hinblick auf die Anreizwirkungen für das Lehrpersonal. Die Erkenntnis, dass Studierende offenbar gutaussehende Lehrkräfte schätzen, ist dabei im Hinblick auf mögliche Anreizwirkungen weitestgehend unproblematisch, da sich die Attraktivität zu einem großen Teil auf von den Betroffenen weitestgehend unveränderliche Eigenschaften (wie Größe, Gesichtsform etc.) bezieht. Davon abgesehen handelt es sich um ein für viele Berufssparten typisches Phänomen (vgl. Frieze et al. 1991; Biddle und Hamermesh 1998; Judge und Cable 2004; Judge et al. 2009).

Deutlich problematischer ist die Tatsache, dass ein niedriges Niveau in Prüfungssituationen tendenziell gute Lehrevaluationen begünstigt. Sollten sich Lehrende anreizkonform verhalten und die Anforderungen in den Prüfungen absenken, drohen durchaus signifikante gesamtwirtschaftliche Schäden. Vom

einzelnen Prüfungskandidaten mag dies zwar als Vorteil empfunden und in der Evaluation entsprechend goutiert werden. In der Gesamtheit betrachtet handelt es sich jedoch um einen negativen externen Effekt. Zwar entstehen sowohl dem Geprüften in Form von weniger Arbeitsaufwand als auch dem Lehrenden durch die besseren Evaluationen jeweils einzelwirtschaftliche Vorteile, Wirtschaft und Gesellschaft als Ganzes können jedoch kein Interesse an einer Absenkung des Prüfungs- und Ausbildungsniveaus haben (vgl. Chan et al. 2007; Pressman 2007; Gaens 2013). In diesem Sinne besteht zumindest die begründete Vermutung, dass Lehrevaluationen auf Basis von Befragungsdaten (neben den an dieser Stelle nicht thematisierten und keineswegs bezweifelte positiven Effekten) auch gesamtgesellschaftlich negative Auswirkungen haben können. Die Frage, ob und wie man ggf. besser geeignete Indikatoren für diesen Bereich des Leistungscontrollings findet, drängt sich hier zumindest auf.

3.2. Forschungsleistungen

Neben dem Lehrbereich bildet die Forschung, d. h. die Generierung neuen Wissens, die zweite Hauptaufgabe hochschulischer Tätigkeit. Auch hier haben sich in den vergangenen Jahrzehnten gängige Methoden zur Messung des Outputs etabliert. Der Fokus liegt im Gegensatz zur Lehrevaluation stärker auf der Hochschul- als auf der Personenebene, die aber methodisch lediglich deren Aggregat ist. Die Ergebnisse der Outputmessung werden anhand von Kennzahlen zu Rankings verdichtet, die sowohl fachimmanent aber auch in der breiten Öffentlichkeit regelmäßig eine große Resonanz erzeugen. Beispiele dafür sind das internationale Ranking der Wissenschaftszeitschrift Nature, das F.A.Z.-Ökonomenranking, das CHE Ranking der Bertelsmann-Stiftung, das QS World University Ranking und insbesondere für unser Fach das Handelsblatt-BWL-Ranking (vgl. für eine systematische Übersicht Müller 2012; Dilger und Müller 2016).

Im Gegensatz zum Lehrbereich wird der Streit um die Validität und Sinnhaftigkeit von Forschungsleistungsrankings öffentlich und mit einiger Verve bis hin zu prominenten Boykottaufrufen geführt (siehe bspw. Albers 2009; Kieser 2010b, 2010a, 2012; Frey 2003; Erne 2007). Die Ergebnisse haben in der deutschsprachigen Fachöffentlichkeit auch deshalb für größere Resonanz gesorgt, da zumindest an den Universitäten der Zusammenhang zu Berufungen und zum beruflichen Fortkommen der Wissenschaftler noch deutlicher ist als bei den Lehrevaluationen, dem Thema also mehr personalpolitische Relevanz innewohnt.

Als Leistungsindikatoren kommen für den Forschungsbereich neben der Anzahl der Publikationen insbesondere Zitationskennzahlen in Betracht (vgl. grundlegend Hornbostel 2006). Zitationen messen die Anzahl der Verweise auf eine Publikation aus anderen Publikationen heraus und sollen die Wirkmächtigkeit des Inhalts abbilden. Sie versehen die Publikationen so mit einer Qualitätsgewichtung, die entweder publikationsindividuell oder auf Zeitschriftenebene über deren Impactfaktoren erfolgt (vgl. Müller 2012). Quantitative Untersuchungen zur Rezeption von wissenschaftlichen Publikationen haben sich unter dem Stichwort Bibliometrie zu einer eigenen Disziplin entwickelt.

Die Frage, wie sich individuelle Forschungsleistungen mit Hilfe bibliometrischer Methoden abbilden lassen und welche weiteren Einflussfaktoren sich in den ermittelten Ergebnissen niederschlagen, wurde für die deutschsprachige BWL in einer von Müller und Dilger 2016 vorgelegten Studie eruiert. Untersuchungsgegenstand waren die Publikationen aller Hochschullehrer im Verband der Hochschullehrer für Betriebswirtschaft. Neben der typischen Ungleichverteilung der Publikationstätigkeit und der Zitationen gemäß der Pareto-Regel (vgl. Pareto 1964) fiel hier insbesondere der Zusammenhang

zwischen dem inhaltlichen Forschungsfeld eines Hochschullehrers und dem Abschneiden in Rankings in den Blick. Im Rahmen von multivariaten Analysen konnte gezeigt werden, dass die inhaltlichen Schwerpunkte eines Autors einen signifikanten Einfluss auf dessen Platzierungen in Rankings haben. Als Gründe dafür wurden unterschiedliche Zitations- und Publikationskulturen in den Teilfächern der BWL identifiziert, die nicht mit Unterschieden in Anspruch oder Qualität verwechselt werden dürfen. Es ist also weder ein Zufall noch ein Qualitätsindikator, dass die Spitzenpositionen in Forschungsleistungsrankings in der BWL überdurchschnittlich häufig von Vertretern bestimmter Teilfächer dominiert werden.

Zweifelsohne spiegeln Rankings Quantität und Qualität des wissenschaftlichen Schaffens eines Hochschullehrers wider. Die Rangfolge wird aber durch die bearbeiteten Themenfelder systematisch verzerrt, d. h. es wird Ungleiches miteinander verglichen. Die Frage, ob jemand auf Platz fünf Größeres geleistet hat als jemand auf Platz 105 ist damit schlichtweg nicht beantwortbar. Als Mittel des Outputcontrollings erzeugen Rankings eine Scheinobjektivität, auf deren Basis dann aber möglicherweise reale (Fehl-)Entscheidungen getroffen werden (vgl. Albers 2009; Müller 2013; Müller und Dilger 2016; Kieser 2012).

4. Schlussfolgerungen

Die beiden referierten Beispiele aus dem Hochschulbereich sollen abschließend im Hinblick auf das Thema Leistungscontrolling im Allgemeinen diskutiert werden. Dazu werden auf ihrer Basis sieben Themen formuliert.

1. Die Beispiele aus Forschung und Lehre verweisen auf ein allgemeines Problem des Leistungscontrollings. Es handelt sich keineswegs um ein Spezifikum des Hochschulbereichs, sondern kann immer dann auftreten, wenn das Ergebnis eines Leistungsprozesses außerhalb einer Markttransaktion gemessen werden muss. Stehen keine marktbasieren Umsatzdaten zur Verfügung, werden Indikatoren benötigt, die potenziell anfällig für Störeffekte sind.
2. Auch vermeintlich simple Indikatoren oder jahrzehntelang gelebte Praktiken des Outputcontrollings sind kritisch zu hinterfragen. Sowohl bei den Evaluationen als auch bei den Rankings handelt es sich um erprobte und im Falle der Lehrevaluationen jahrein jahraus tausendfach durchgeführte Standardverfahren.
3. Problematische Leistungsindikatoren können Fehlanreize für die Leistungserbringer setzen und produzieren in der Folge negative Externalitäten. Wenn Leistungsindikatoren nur einen Ausschnitt des Leistungsbereichs und diesen ggf. auch noch verzerrt abbilden, reduziert sich in der Folge die erbrachte Leistung oft auf diesen Ausschnitt.
4. Durch die Fehlanreize entsteht der eigentliche Schaden. Das Problem sind nicht die ggf. irrelevanten Messwerte an sich, sondern deren anreizbasierten Folgen. Die betriebs- und volkswirtschaftlichen Schäden sind in diesen Fehllenkungen zu suchen.
5. Verbesserte Messtechniken können das Problem allein nicht lösen. Eine größere Anzahl von Indikatoren oder immer differenziertere, methodisch aufwendigere Messverfahren (bspw. eine statistische Normalisierung der identifizierten Störgrößen) verringern das Problem nicht automatisch. Je differenzierter die Methodik, desto schwieriger sind Validitäts- und Anreizprobleme erkennbar.

6. Daraus folgt keine Forderung zu einem generellen Verzicht auf Leistungscontrolling. Dies wäre nur dann gerechtfertigt, wenn die Schäden durch Fehlanreize den Nutzen des Controllings überwiegen, was zumindest in der Regel nicht der Fall sein dürfte.
7. Ein reflektiertes Output-Controlling muss hingegen zwischen den Controllingzielen und etwaigen Fehlanreizen abwägen. Dies erfordert eine genauere Analyse der im Einzelfall vorgeschlagenen Indikatoren hinsichtlich Beobachtbarkeit, Validität und Anreizwirkungen. Ein Leistungscontrolling mit imperfekten Indikatoren bleibt dort sinnvoll, wo auch bei der Berücksichtigung von Fehlanreizen ein positiver Ergebnisbeitrag des Ansatzes zu erwarten ist.

Mit Blick auf das Postulat „What gets measured gets managed!“ empfehlen diese Schlussfolgerungen eine gewisse Bescheidenheit. Die Fallbeispiele aus dem Hochschulwesen haben gezeigt, dass Performance Measurement, neben seinen hier nicht weiter thematisierten positiven Wirkungen, in bestimmten Fällen auch mit betriebs- und volkswirtschaftlich negativen Wirkungen einhergehen kann, die dann gegen erstere abgewogen werden müssen. Der Grund dafür liegt nicht in den Spezifika der Hochschule als Organisationsform, sondern ist genereller Natur. Der Output wird nicht in einem Marktprozess mit Preisindikation versehen, sondern muss anhand von Indikatoren quantifiziert werden. Diese Indikatoren können naturgemäß amorphe Leistungsverständnisse wie „gute Lehre“ oder „relevante Forschung“ aber nicht adäquat abbilden und führen so unweigerlich zu Problemen.

Für diese allgemeine Erkenntnis bieten sich zahlreiche mögliche Transferbereiche auch und gerade im Controlling von privaten Wirtschaftsunternehmen an. Die Frage, wie naturgemäß unscharfe Leistungsbegriffe adäquat operationalisiert werden können, ist eine zentrale Herausforderung des Leistungscontrollings. Das Konzept des Performance Measurements in die Unternehmenspraxis zu transferieren bedeutet, sich die Frage nach Beobachtbarkeit, Validität und Anreizkompatibilität von Leistungsindikatoren zu stellen. Dies geschieht auch in der Privatwirtschaft vielfach tendenziell zu unüberlegt, so dass sich auf Basis der hier umrissenen Problematik ein vielfältiges Programm für eine anwendungsorientierte Forschung ergibt. Es gilt, das Konzept des Performance Measurements mit dem Blick auf konkrete Anwendungssituationen nicht nur weiter zu detaillieren, sondern kritisch auf mögliche Probleme hin zu prüfen. Die Fragestellung bietet zahlreiche Anknüpfungspunkte sowohl im Hinblick auf eine konzeptionelle Weiterentwicklung des Leistungscontrollings als auch für konkrete Projekte in Unternehmen. Als mögliche Themenfelder können exemplarisch und ohne den Anspruch auf Vollständigkeit etwa die Humanvermögensrechnung, Sozialbilanzen, das Controlling von Corporate Social Responsibility-Aktivitäten, Markenmanagement, das Controlling des Forschungs- und Entwicklungsbereichs oder die Messung von nicht-stochastischen Risiken, bspw. in den Bereichen Recht und Reputation, genannt werden. Darüber hinaus stellt sich die Frage der Leistungsmessung in all den Bereichen der Volkswirtschaft, in denen Leistungsempfänger und –zahler nicht identisch sind, etwa im Gesundheitsbereich, der Daseinsfürsorge, dem Schulwesen usw.

In allen diesen Feldern gilt es, die Validität von Leistungsindikatoren kritisch zu hinterfragen und Fehlanreize aufzudecken, wo sie existieren. Auf dieser Basis können dann reflektierte Gestaltungsvorschläge für eine verbesserte Controllingpraxis entwickelt und praktisch implementiert werden.

Literaturverzeichnis

- Albers, Sönke (2009): Misleading Rankings of Research in Business. In: *German Economic Review* 10 (3), S. 352–363.
- Arnaboldi, Michela; Azzone, Giovanni (2010): Constructing Performance Measurement in the Public Sector. In: *Critical perspectives on accounting* 21; Jg. 2010 (4), S. 266–282.
- Bacevic, Jana (2017): Beyond the Third Mission. Toward an Actor-Based Account of Universities' Relationship with Society. In: Hakan Ergül und Simten Coşar (Hg.): *Universities in the Neoliberal Era. Academic Cultures and Critical Perspectives*. London: Palgrave Macmillan, S. 21–39.
- Biddle, Jeff E.; Hamermesh, Daniel S. (1998): Beauty, Productivity, and Discrimination. Lawyers' Looks and Lucre. In: *Journal of Labor Economics* 16 (1), S. 172–201.
- Brown, Mark G.; Svenson, Raynold A. (1988): Measuring R&D Productivity. In: *Research-Technology Management* 31 (4), S. 11–15.
- Chan, William; Hao, Li; Suen, Wing (2007): A Signaling Theory of Grade Inflation. In: *International Economic Review* 48 (3), S. 1065–1090.
- Dilger, Alexander; Müller, Harry (2016): Outputanalyse betriebswirtschaftlicher Fachbereiche. Ein zitationsbasiertes Ranking deutschsprachiger Hochschulen. In: Heinz Ahn, Marcel Clermont und Rainer Souren (Hg.): *Nachhaltiges Entscheiden. Beiträge zum multiperspektivischen Performancemanagement von Wertschöpfungsprozessen*. Wiesbaden: Springer Gabler, S. 405–425.
- Erne, Roland (2007): On the Use and Abuse of Bibliometric Performance Indicators. A critique of Hix's 'Global Ranking of Political Science Departments. In: *European Political Science* 6 (3), S. 306–314.
- Felton, James; Koper, Peter T.; Mitchell, John; Stinson, Michael (2008): Attractiveness, Easiness and Other Issues. Student Evaluations of Professors on Ratemyprofessors. com. In: *Assessment & Evaluation in Higher Education* 33 (1), S. 45–61.
- Freng, Scott; Webber, David (2009): Turning Up the Heat on Online Teaching Evaluations. Does "Hotness" Matter? In: *Teaching of Psychology* 36 (3), S. 189–193.
- Frey, Bruno S. (2003): Publishing as Prostitution? Choosing Between One's Own Ideas and Academic Success. In: *Public Choice* 116 (1), S. 205–223.
- Frieze, Irene Hanson; Olson, Josephine E.; Russell, June (1991): Attractiveness and Income for Men and Women in Management. In: *Journal of Applied Social Psychology* 21 (13), S. 1039–1057.
- Gaens, Thomas (2013): Von einem, der auszog, einen Leistungsindikator zu erheben. Durchfallquoten und die Problematik ihrer Bildung. In: *Das Hochschulwesen* 61 (6), S. 200–206.
- Gleich, Ronald (1997): Performance Measurement. In: *Die Betriebswirtschaft* 57, S. 114–118.
- Greenwood, Ronald C. (1981): Management by Objectives. As Developed by Peter Drucker, Assisted by Harold Smiddy. In: *Academy of Management Review* 6 (2), S. 225–230.

- Hachmeister, Cort-Denis; Henke, Justus; Roessler, Isabel; Schmid, Sarah (2016a): Die Vermessung der Third Mission. Wege zu einer erweiterten Darstellung von Lehre und Forschung. In: *Die Hochschule* 25 (1), S. 7–13.
- Hachmeister, Cort-Denis; Möllenkamp, Moritz; Roessler, Isabel; Scholz, Christina (2016b): Katalog von Facetten von und Indikatoren für Forschung und Third Mission an Hochschulen für angewandte Wissenschaften. Gütersloh: Centrum für Hochschulentwicklung gGmbH (CHE Arbeitspapier).
- Hornbostel, Stefan (2006): Leistungsmessung in der Forschung. In: Hochschulrektorenkonferenz (HRK) (Hg.): Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung. Bonn (Beiträge zur Hochschulpolitik, 1), S. 219–228.
- Judge, Timothy A.; Cable, Daniel M. (2004): The Effect of Physical Height on Workplace Success and Income. Preliminary Test of a Theoretical Model. In: *The Journal of applied psychology* 89 (3), S. 428–441.
- Judge, Timothy A.; Hurst, Charlice; Simon, Lauren S. (2009): Does it Pay to Be Smart, Attractive, or Confident (or All Three)? Relationships among General Mental Ability, Physical Attractiveness, Core Self-Evaluations, and Income. In: *Journal of Applied Psychology* 94 (3), S. 742–755.
- Kaplan, Robert S.; Norton, David P. (1996): *The Balanced Scorecard. Translating Strategy into Action.* Boston, MA: Harvard Business Press.
- Kieser, Alfred (2010a): Unternehmen Wissenschaft? In: *Leviathan* 38 (3), S. 347–367.
- Kieser, Alfred (2010b): Die Tonnenideologie der Forschung. In: *Frankfurter Allgemeine Zeitung* 130, 09.06.2010, N5.
- Kieser, Alfred (2012): JOURQUAL-der Gebrauch, nicht der Missbrauch, ist das Problem. Oder: warum Wirtschaftsinformatik die beste deutschsprachige betriebswirtschaftliche Zeitschrift ist. In: *Die Betriebswirtschaft* 72 (1), S. 93–110.
- Lachmann, Werner (2004): *Volkswirtschaftslehre 2. Anwendungen.* Zweite, vollständig neu bearbeitete und erweiterte Auflage. Berlin, Heidelberg: Springer.
- Müller, Harry (2012): Zitationen als Grundlage von Forschungsleistungsrankings. Konzeptionelle Überlegungen am Beispiel der Betriebswirtschaftslehre. In: *Beiträge zur Hochschulforschung* 34, S. 68–92.
- Müller, Harry (2013): Zur Ethik von Rankings im Hochschulwesen. Eine Betrachtung aus ökonomischer Perspektive. In: *Hochschulmanagement* 8 (2/3), S. 41–46.
- Müller, Harry; Dilger, Alexander (2016): Wie der Forschungsschwerpunkt den Zitationserfolg beeinflusst. Eine empirische Untersuchung für die deutschsprachige BWL. In: *Betriebswirtschaftliche Forschung und Praxis* 68 (1), S. 36–52.
- Neely, Andy; Gregory, Mike; Platts, Ken (1995): Performance Measurement System Design. A Literature Review and Research Agenda. In: *International journal of operations & production management* 15 (4), S. 80–116.

- Odiorne, George S. (1965): *Management by Objectives. A System of Managerial Leadership*. New York: Pitman.
- Pareto, Vilfredo (1964): *Cours d'Économie Politique*. Genf: Librairie Droz.
- Pressman, Steven (2007): The Economics of Grade Inflation. In: *Challenge* 50 (5), S. 93–102.
- Roessler, Isabel; Duong, Sindy; Hachmeister, Cort-Denis (2015a): Teaching, Research and More?! Achievements of Universities of Applied Sciences with regard to Society. Gütersloh: Centrum für Hochschulentwicklung (CHE Arbeitspapier).
- Roessler, Isabel; Duong, Sindy; Hachmeister, Cort-Denis (2015b): Welche Missionen haben Hochschulen? Third Mission als Leistung der Fachhochschulen für die und mit der Gesellschaft. Gütersloh: Centrum für Hochschulentwicklung.
- Roessler, Isabel; Hachmeister, Cort-Denis; Scholz, Christina (2016): Positionierung durch Profilierung- Stärkung der Third Mission an HAW. Gütersloh: Centrum für Hochschulentwicklung.
- Rommelspacher, Jonas; Burmester, Lars; Goeken, Matthias (2006): Performance-Measurement- und Analyse-Konzepte im Hochschulcontrolling. In: Hans-Dietrich Haasis, Herbert Kopfer und Jörn Schönberger (Hg.): *Operations Research Proceedings 2005. Selected Papers of the Annual International Conference of the German Operations Research Society (GOR), Bremen, September 7-9, 2005*, Bd. 2005. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg (Operations Research Proceedings, 2005), S. 539–544.
- Schmalenbach, Eugen (1928): *Buchführung und Kalkulation im Fabrikgeschäft*. Unveränd. Nachdr. aus der Dt. Metallindustriezeitung, 15. Jg. 1899. Leipzig: Gloeckner.
- Schneidewind, Uwe (2016): Die "Third Mission" zur "First Mission" machen? Wuppertal: Wuppertal Institut für Klima, Umwelt, Energie.
- Seiter, Mischa (2006): Performance Measurement. In: *Wissenschaftsmanagement* 12; Jg. 2006 (5), S. 36–38.
- Shore, Cris; McLauchlan, Laura (2012): 'Third Mission Activities, Commercialisation and Academic Entrepreneurs. In: *Social anthropology* 20 (3), S. 267–286.
- Simons, Robert (2000): *Performance Measurement & Control Systems for Implementing Strategy. Text & Cases*. Upper Saddle River, NJ: Pearson Education.
- Spooren, Pieter; Brockx, Bert; Mortelmans, Dimitri (2013): On the Validity of Student Evaluation of Teaching. In: *Review of Educational Research* 83 (4), S. 598–642.
- Timmerman, Thomas (2008): On the Validity of RateMyProfessors.com. In: *Journal of Education for Business* 84 (1), S. 55–61.
- Watson, Derek; Hall, Lynne (2015): Addressing the Elephant in the Room. Are Universities Committed to the Third Stream Agenda. In: *International Journal of Academic Research in Management* 4 (2), S. 48–76.
- Weber, Jürgen; Großklaus, Armin (1995): *Kennzahlen für die Logistik*. Stuttgart: Schäffer-Poeschel.

Weber, Jürgen; Schäffer, Utz (1999): Sicherstellung der Rationalität von Führung als Aufgabe des Controlling. In: *Die Betriebswirtschaft* 59 (6), S. 731–747.

Weber, Jürgen; Schäffer, Utz (2016): Einführung in das Controlling. 15. Aufl. Stuttgart: Schäffer-Poeschel.