

Ringel, Leopold; Werron, Tobias

Article — Published Version

Serielle Vergleiche: Zum Unterschied, den Wiederholung macht

KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie

Provided in Cooperation with:

Springer Nature

Suggested Citation: Ringel, Leopold; Werron, Tobias (2021) : Serielle Vergleiche: Zum Unterschied, den Wiederholung macht , KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie, ISSN 1861-891X, Springer Fachmedien Wiesbaden, Wiesbaden, Vol. 73, Iss. Suppl 1, pp. 301-331,
<https://doi.org/10.1007/s11577-021-00750-2>

This Version is available at:

<https://hdl.handle.net/10419/287402>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Serielle Vergleiche: Zum Unterschied, den Wiederholung macht

Anhand der Geschichte von Kunst- und Hochschulrankings

Leopold Ringel · Tobias Werron

Eingegangen: 4. Juni 2020 / Angenommen: 3. Juli 2020 / Online publiziert: 19. Mai 2021
© Der/die Autor(en) 2021

Zusammenfassung Wie entstehen und welche Wirkung entfalten Vergleichs- und Bewertungspraktiken, die Serien bilden, indem sie bestimmte Vergleichsereignisse regelmäßig wiederholen? Der Beitrag entwickelt diese Frage theoretisch und untersucht sie empirisch anhand der Geschichte von Kunst- und Hochschulrankings. Die konzeptionelle Analyse lenkt den Blick auf das „Wie“ des Vergleichens (performative Dimension) und betont, dass wiederholt und regelmäßig durchgeführte Vergleiche (serielle Vergleiche) neue Möglichkeiten erschließen, zeitliche Unterschiede festzustellen (temporales Vergleichen) sowie diese interpretativ miteinander zu verknüpfen (temporalisierendes Vergleichen). Die historische Analyse zeigt, wie sich diese Dimensionen der Zeitlichkeit von Vergleichen in der Entstehung der heutigen Kunst- und Hochschulrankings gegenseitig ermöglicht und stabilisiert haben. Dabei wird einerseits deutlich, dass ein Fokus auf Wiederholungs- und Regelmäßigkeitsmuster den Blick auf heute prominente Vergleichs- und Bewertungspraktiken grundlegend neu ausrichten kann. Zugleich zeigen sich bedeutende gesellschaftstheoretische Implikationen der Vergleichs- und Bewertungsforschung, etwa mit Blick auf die historische Ausdifferenzierung von Feldern.

Schlüsselwörter Bewerten · Felder · Institutionalisierung · Praktiken · Quantifizierung

Der Beitrag ist im Rahmen des von der Deutschen Forschungsgemeinschaft geförderten Projekts „Zur Institutionalisierung der Rankings: Diskurskarrieren tabellarischer Leistungsvergleiche zwischen 1850 und 1980“ entstanden.

L. Ringel (✉) · T. Werron (✉)
Fakultät für Soziologie, Universität Bielefeld
Universitätsstr. 25, Postfach 10 01 31, 33651 Bielefeld, Deutschland
E-Mail: leopold.ringel@uni-bielefeld.de; tobias.werron@uni-bielefeld.de

Serial Comparisons: How Repetition Matters

The Example of Art and University Rankings

Abstract How do regularly repeated practices of comparing emerge, and what are their consequences? The article discusses this question conceptually and empirically by tracing the history of rankings of artists and universities. The conceptual analysis draws attention to the performative dimension of comparison (the “how” of comparing) and suggests that regularly repeated acts of comparing (serial comparisons) can determine temporal differences (temporal comparisons) and spur the development of novel interpretations of these differences (temporalized comparisons). The historical analysis shows how these temporal dimensions of comparing have been mutually enabling and stabilizing in the history of art rankings and university rankings. We conclude by arguing that focusing on patterns of repetition and regularity not only allows for a new approach to the analysis of modern practices of comparing and valuation, but it also sheds light on general problems of social theory, such as the historical formation of fields.

Keywords Evaluation · Fields · Institutionalization · Practices · Quantification

1 Einleitung

Wie die interdisziplinäre Vergleichsforschung in den vergangenen Jahren zeigen konnte, sind Vergleiche mehr als kognitive Operationen. Sie sind nicht bloß „passive“ Beobachtungen, die eine unabhängig existierende soziale Wirklichkeit – mehr oder weniger „objektiv“ – abbilden, sondern Praktiken, die, einmal in die Welt gesetzt, mit häufig weitreichenden Folgen in die von ihnen beobachtete Welt intervenieren (Heintz 2010, 2016; Epple 2015; Heintz und Werron 2011; Steinmetz 2019; Werron 2010, 2012a, 2020). Auch Prozesse der Bewertung, wie die neuere Bewertungssoziologie sie untersucht (Krüger und Reinhart 2016; Meier et al. 2016; Nicolae et al. 2019), verbinden sich in vielen Fällen mit Vergleichspraktiken. Wenn in Schulen die Noten von Schülern¹ bewertet, in Dokumenten der Vereinten Nationen indigene Völker klassifiziert oder in Zeitungen das Bruttoinlandsprodukt von Nationalstaaten zum Thema gemacht wird, dann spielen immer auch Vergleiche eine wichtige Rolle. Diese, so die zentrale These der Vergleichsforschung, machen „etwas“: Schüler verwandeln sich in strategische Akteure, deren Ziel darin besteht, bessere Noten zu erhalten, Gruppen von Personen nehmen sich vor dem Hintergrund der Institutionalisierung von Indigenenrechten vermehrt als indigene Völker wahr und Politiker vergleichen die eigene nationalstaatliche Wirtschaftsleistung mit jener anderer Staaten und orientieren sich an „Erfolgsmodellen“. Kurz: Vergleiche sind im wahrsten Sinne des Wortes produktiv; sie sind jenseits aller methodischen Fragen ein soziales Phänomen und daher von soziologischem Interesse.

Wenig Aufmerksamkeit ist bisher hingegen darauf gerichtet worden, dass Vergleiche Phänomene sind, die auch eine zeitliche Dimension besitzen und gerade

¹ Das generische Maskulinum wird verwendet.

hierdurch wirksam werden – in einem doppelten Sinn: Man kann, erstens, unter zeitlichen Gesichtspunkten vergleichen. Wenn das „Frühwerk“ eines Schriftstellers mit seinem „Spätwerk“, der Zustand eines Autos vor und nach einem Unfall oder die „Entwicklung“ eines Nationalstaats über einen längeren Zeitraum verglichen wird, dann werden die Objekte des Vergleichs in ihrer Zeitlichkeit konstruiert. Vergleichspraktiken dieses Typs relationieren die verglichenen Objekte in zeitlichen Hinsichten. Diese Vergleiche möchten wir als temporal (Markierung zeitlicher Unterschiede) oder temporalisierend (Interpretation zeitlicher Verläufe) bezeichnen. Davon zu unterscheiden ist, zweitens, Zeitlichkeit, die sichtbar wird, wenn wir Praktiken des Vergleichens auf ihre eigene zeitliche Ordnung hin beobachten, etwa mit Blick darauf, ob seltener oder häufiger, regelmäßiger oder unregelmäßiger, in kürzerer oder längerer Taktung verglichen wird. Hier geht es um die Temporalität des Vergleichens als sozialer Praxis.

Der vorliegende Beitrag zeigt, dass Aufmerksamkeit für diese zweite Zeitlichkeitsdimension, die in der Literatur bisher noch kaum beachtet worden ist, den Blick auf heute prominente Vergleichs-, Quantifizierungs- und Bewertungspraktiken in grundlegenden Hinsichten neu ausrichten kann. Empirisch stützen wir uns auf ein Forschungsprojekt zur Geschichte der Rankings, dessen vorläufige Erkenntnisse wir vorsichtig generalisieren, indem wir einen allgemeineren Typus des Vergleichens identifizieren sowie am Beispiel von Rankings illustrieren: das *serielle Vergleichen*.

Serielle Vergleiche zeichnen sich – so unser Vorschlag – durch *Wiederholung* und *Regelmäßigkeit* aufeinander folgender, sinnhaft miteinander verknüpfter Vergleichsereignisse aus. Beispiele sind regelmäßig veröffentlichte Ranglisten, Ratings, Benchmarks oder die jährliche Benotung von Schülern. Serielle Vergleiche erlauben es, Ergebnisse von Vergleichen, die zu verschiedenen Zeitpunkten stattfinden, ihrerseits regelmäßig miteinander zu vergleichen. Das ermöglicht nicht zuletzt neuartige Vergleiche der „Entwicklung“ der verglichenen Einheiten über längere Zeiträume hinweg. Damit verknüpft unser Begriffsvorschlag beide Zeitlichkeitsdimensionen des Vergleichens und lässt vermuten, dass sich temporales und temporalisierendes Vergleichen nur verstehen lassen, wenn man auch die Serialität des Vergleichens untersucht – und umgekehrt.

Unser Beitrag entfaltet und prüft diese Vermutung in drei Schritten: Wir beginnen mit einem konzeptionellen Abschnitt, der unseren Begriff serieller Vergleiche genauer vorstellt. Der nächste Abschnitt erprobt diesen Vorschlag, indem er die (Vor-)Geschichte heutiger Rankings in Kunst und Wissenschaft in den Blick nimmt und zeigt, wie sich Praktiken des Vergleichens und quantitativen Bewertens in einer heute einflussreichen seriellen Vergleichspraxis (Rankings) verbunden haben, aber auch, wie sozial voraussetzungsreich und je nach Fall unterschiedlich die Durchsetzung dieser Vergleichspraxis verlaufen ist. Die empirischen Beobachtungen laufen auf zwei allgemeinere Thesen hinaus, die sie für die Vergleichs- und Bewertungsforschung auch über den Fall der Rankings hinaus interessant machen: (1) Untersuchungen serieller Vergleichspraktiken können dazu beitragen, *Beziehungen zwischen Praktiken des Vergleichens und quantitativen Bewertens* begrifflich und empirisch zu spezifizieren. (2) Beide Zeitlichkeitsdimensionen – temporales und temporalisierendes Vergleichen und Bewerten einerseits, Temporalität und Serialität von Vergleichs- und Bewertungspraktiken andererseits – kann man nur angemessen verstehen, wenn

man sie auch *auf ihre historischen Beziehungen hin untersucht*. Wie der dritte Abschnitt ausführt, eröffnen beide Thesen der Vergleichs- und Bewertungsforschung ein weites und noch kaum bearbeitetes Forschungsfeld.

2 Serielle Vergleichspraktiken: Zum Unterschied, den Wiederholung macht

Als Praktiken weisen Vergleiche stets zwei Dimensionen auf, eine informative und eine praktisch-performative (Werron und Ringel 2017): Die informative Dimension handelt vom „Was“ des Vergleichens, davon, welche Welt in Vergleichen konstruiert wird, welche Gegenstände sie beobachten, unterscheiden und bewerten; die praktisch-performative Dimension handelt dagegen vom „Wie“, also davon, wie sich Vergleiche in der sozialen Praxis vollziehen und sie in der sozialen Welt verwendet und institutionalisiert werden. Im empirischen Vollzug von Vergleichspraktiken hängen beide Dimensionen stets zusammen. Gleichwohl ist ihre analytische Unterscheidung hilfreich – nicht zuletzt, um die Zeitlichkeit des Vergleichens genauer in den Blick zu nehmen.

2.1 Zum „Was“ des Vergleichens: Die informative Dimension von Vergleichspraktiken

Die Vergleichsforschung hat sich theoretisch-begrifflich bislang primär auf die informative Dimension des Vergleichens konzentriert. Dabei kann sie sich auf eine grundlegende, von Bettina Heintz (2010) vorgeschlagene Definition stützen, wonach Vergleiche (1) die Vergleichbarkeit mindestens zweier Entitäten unter bestimmten Gesichtspunkten behaupten (Gleichartigkeits- oder Vergleichbarkeitsannahme), um (2) unter zusätzlichen Gesichtspunkten Ähnlichkeit oder Differenz zwischen diesen Entitäten festzustellen (Vergleichskriterien, auch *tertia comparationis*). Vergleiche behaupten also Gleichartigkeit, *um* Ähnlichkeiten oder Differenzen feststellen zu können. Mit diesem Vergleichsbegriff lassen sich selten reflektierte Voraussetzungen des Vergleichens explizieren und untersuchen. Deutlich wird beispielsweise, dass Behauptungen von Unvergleichbarkeit häufig auf Behauptungen von Unähnlichkeit hinauslaufen, die Vergleiche nicht ausschließen, sondern voraussetzen. So etwa im Fall von Äpfeln und Birnen, die einander, wie das Sprichwort betont, in wichtigen Hinsichten unähnlich sein mögen. Die Feststellung von Unähnlichkeit setzt jedoch gerade voraus, dass wir sie als Obst definieren und dadurch „verähnlichen“ (Gleichartigkeitsannahme) sowie hinsichtlich *tertia* wie Form, Farbe, Geschmack usw. unterscheiden (Vergleichskriterien). In solchen und vielen anderen Fällen erweist sich das Vergleichen als implizite, in der Alltags- wie Wissenschaftssprache selten reflektierte *Voraussetzung* des Bewertens. Mit anderen Worten: Eine Soziologie des Bewertens kommt ohne eine Soziologie des Vergleichens nicht aus.

An der in Bettina Heintz' Definition betonten informativen Seite des Vergleichens setzen auch die meisten Versuche der Bestimmung von *Typen* von Vergleichspraktiken an (Heintz 2016, S. 311). Wir greifen aus der Vielzahl von Vorschlägen drei Begriffspaare heraus, die für unsere Überlegungen zur Zeitlichkeit des Vergleichens

sowie für unsere empirische Analyse der Rankings wichtig sind: (1) komplexe im Unterschied zu einfachen Vergleichen; (2) steigernde Vergleiche im Unterschied zu bloßen Ähnlichkeits- oder Differenzfeststellungen; (3) quantitative im Unterschied zu qualitativen Vergleichen.

Einfache vs. komplexe Vergleiche. Man kann zunächst unterscheiden zwischen einfachen Vergleichen, die nur einen einzigen Vergleichsgesichtspunkt benutzen, um Ähnlichkeiten und Differenzen festzustellen, und komplexen Vergleichen, „die mehrere Partialvergleiche mit je eigener Hinsicht zueinander ins Verhältnis zu setzen versuchen“ (Grave 2015, S. 144). Beispiele für einfache Vergleiche sind jener zwischen Apfel und Birne unter dem Gesichtspunkt des Geschmacks oder von Universitäten unter dem Gesichtspunkt des Publikationsoutputs. Komplexe Vergleiche dagegen nutzen mehrere Vergleichshinsichten und relationieren sie zueinander, vergleichen etwa Apfel und Birne unter den Gesichtspunkten Geschmack *und* Form oder Universitäten unter den Gesichtspunkten Publikationsoutput *und* Zahl der Nobelpreise. Diese Unterscheidung ist wichtig, weil die Komplexität des Vergleichs auch das Verhältnis zwischen Vergleich, Zahlenverwendung und Bewertung beeinflusst. So mag noch relativ leicht festzustellen sein, welche Universität in einem gegebenen Zeitraum die meisten Nobelpreisträger aufweist. Wenn jedoch weitere Merkmale, wie Publikationsoutput oder Diversität der Studierendenschaft, berücksichtigt werden sollen, sind Verknüpfungs- und Metakriterien erforderlich, die diese Kriterien ihrerseits miteinander vergleichen, bewerten und verrechnen. Die entsprechenden Kombinationen verschiedener Vergleichs- und Bewertungskriterien und die historischen Voraussetzungen ihres Zustandekommens werden jedoch selten genauer reflektiert.

Steigernde vs. Ähnlichkeit/Differenz feststellende Vergleiche. Der komplexe Universitätsvergleich ist zugleich ein Beispiel für einen Typus, den Steinmetz (2015, S. 88) als *steigernden Vergleich* charakterisiert. Er umfasst Vergleiche unter Gesichtspunkten wie kleiner/größer, mehr/weniger, besser/schlechter, kürzer/länger, höher/niedriger etc., die es gestatten, nicht nur auf Ähnlichkeit und Differenz zu beobachten, sondern zugleich hierarchisch zu differenzieren. Der Bezug zu Praktiken des Bewertens liegt hier auf der Hand. Wer steigernd vergleicht, bewertet das Vergleichen zugleich unter Gesichtspunkten wie Größe, Umfang oder Qualität, bewertet also vergleichend oder vergleicht bewertend. Die Plausibilität des steigernden Vergleichs dürfte zudem entscheidend davon profitieren, wenn Differenzen quantifiziert werden können – besonders, wie wir am Beispiel der Rankings sehen werden, wenn es um die Unterscheidung besser vs. schlechter geht.

Quantitative vs. qualitative Vergleiche. Die dritte Unterscheidung zielt auf den Unterschied zwischen numerischen Vergleichen – von einfachen Bezifferungen bis zu komplexen Statistiken – und solchen, die ohne Verwendung von Zahlen auskommen und in diesem Sinne qualitativ sind. Diese Unterscheidung ist schon deshalb wichtig, weil die Vergleichsforschung seit ihrem Aufkommen in den 1990er- und 2000er-Jahren mit der Forschung über Quantifizierungsprozesse eng verbunden ist (Espeland und Stevens 1998; Heintz 2010). Grundlegende Begriffe wie *commensuration* werden daher häufig in einer Bedeutung verwendet, die das Vergleichen mit *quantification* unauflöslich verknüpft (z. B. Espeland und Stevens 1998; Espeland und Sauder 2007). Mit der Amalgamierung von Vergleichen und Quantifizieren

gingen der Vergleichs- und Bewertungsforschung jedoch wesentliche Differenzierungsmöglichkeiten verloren. Die Unterscheidung zwischen Medien des Vergleichs – sprachlich, numerisch, bildlich – fängt diese Einsicht ein und betont, dass die soziale Wirkungsmacht des Vergleichens auch auf *seiner medialen Vielfalt* beruht, nicht allein auf seiner Affinität mit den Zahlen (Heintz 2010). Auf diese Weise lassen sich quantitative und qualitative Vergleiche unterscheiden, vor allem aber kann analysiert werden, wie und mit welchen Folgen quantitative und qualitative Vergleiche empirisch zusammentreffen und einander ermöglichen und verstärken (etwa im Sinne einer „fraktalen“ Logik; Heintz 2019, S. 63 ff.). Soweit eine Affinität zwischen Vergleichen und Zahlengebrauch besteht, sollte sie also nicht einfach unterstellt, sondern auf ihre empirischen Erscheinungsformen und historischen Möglichkeitsbedingungen hin untersucht werden. Auch dafür bietet die Geschichte der Rankings ein gutes Anschauungsbeispiel (s. u.).

Alle diese Typen von Vergleichen reflektieren grundlegende Einsichten der interdisziplinären Vergleichsforschung. Sie zeigen, dass der Charakter von Vergleichen sich in Abhängigkeit davon ändert, ob nur unter einem oder vielen Gesichtspunkten verglichen wird; das Verhältnis zwischen Vergleichen und Bewerten wird wesentlich davon geprägt, ob der Vergleich nur auf die Feststellung von Ähnlichkeiten und Differenzen abzielt oder auch steigernd ausgerichtet ist; und der Gebrauch von Zahlen sowie seine historischen Voraussetzungen sind wichtige Variablen, wenn es um die Voraussetzungen und Effekte von Vergleichen wie auch um das Verhältnis zwischen Vergleichen und Bewerten geht. Die in diesen typologischen Unterscheidungen eingefangenen Einsichten sind miteinander kombinierbar: Steigernde Vergleiche können insbesondere immer wieder aufs Neue und auf ähnliche Weise durchgeführt sowie mit komplexen und quantitativen Vergleichen kombiniert werden, um Entwicklungen auch in der Zeit zu diagnostizieren und zu bewerten. Entsprechend liegt die Vermutung nahe, dass die Temporalität solcher Vergleiche einen wichtigen Unterschied machen kann. Bevor wir diese Überlegung am Beispiel der Rankings vertiefen, diskutieren wir, was sich auf Grundlage der vorliegenden Vergleichsliteratur über die praktisch-performative Dimension sagen lässt.

2.2 Zum „Wie“ des Vergleichens: Die praktisch-performative Dimension von Vergleichspraktiken

Vergleiche variieren nicht nur in ihren Vergleichbarkeitsannahmen und Vergleichskriterien, sondern auch in der Art und Weise, *wie* sie vollzogen und als soziale Praktiken institutionalisiert werden. Mit Blick auf dieses „Wie“ scheint uns noch erheblicher Bedarf für begriffliche Präzisierungen zu bestehen. Unser theoretisches Anliegen lässt sich an Bettina Heintz' (2010) bereits erwähnter Unterscheidung von Vergleichsmedien – sprachlich, numerisch, bildlich – verdeutlichen. Wir hatten sie zustimmend zitiert, weil sie der begrifflichen Amalgamierung von Vergleichen und Quantifizierung vorbeugt und auf die Multimodalität von Vergleichen aufmerksam macht. Die Unterscheidung von Medien des Vergleichs liegt quer zu unserer Unterscheidung zwischen der informativen und praktisch-performativen Dimension von Vergleichspraktiken. Das Interesse an unterschiedlichen Medien lenkt die Aufmerksamkeit zwar einerseits immer auch auf das „Wie“: Zu fragen, wie sich

Vergleichspraktiken vollziehen, kann die Frage einschließen, ob sie sprachlich (z. B. narrativ), numerisch (z. B. statistisch) oder bildlich (z. B. grafisch) vollzogen werden. Andererseits steckt im medialen „Wie“ zugleich das eigentliche „Was“. Ohne Berücksichtigung der Erzählungen, Zahlen oder Bilder, in denen Vergleiche mitgeteilt werden, wüssten wir gar nicht, was verglichen und welche Ähnlichkeiten und Differenzen im Vergleich bestimmt werden. Kurz, Vergleichen ist *immer* sprachlich, quantitativ, bildlich oder sonst medial konstruiert.

Vor diesem Hintergrund wird deutlich, dass die performative Seite der Praxis des Vergleichens in der Bezeichnung von Vergleichsmedien nicht aufgeht, sondern zahlreiche weitere Fragen aufwirft. Um einige Beispiele zu nennen: Wie werden sprachliche, numerische oder bildliche Vergleiche kommuniziert und reflektiert? Geschieht dies öffentlich oder geheim? Richtet sich der Vergleich an ein räumlich situiertes Publikum, an die Mitglieder einer Organisation oder an eine potenziell unbegrenzte Zahl von Personen? Welche Rolle spielen die Räumlichkeit und Materialität der Vergleichssituationen? Wird die Vergleichsintention manifest oder bleibt sie implizit und der Interpretation des Empfängers überlassen? Und nicht zuletzt: In welcher zeitlichen Abfolge finden Vergleiche statt? Einmalig oder wiederholt? Regelmäßig und unregelmäßig? Selten oder häufig? All diese Fragen drängen auf eine Spezifikation dessen, was in der Vergleichsforschung häufig in einen nicht näher bestimmten „Kontext“ des Vergleichens verlagert wird. Praktiken des Vergleichens lassen sich jedoch von ihren sozialen Kontexten nicht sauber isolieren, ja manche Typen von Vergleichspraktiken sind überhaupt nur angemessen bestimmbar, wenn man ihre praktisch-performative Seite in die Begriffsbildung einbezieht.

Das *implizite* Vergleichen ist ein Typus, an dem diese Einsicht in der Literatur bereits produktiv gemacht worden ist. Dieser Typus verdankt sich der Unterscheidung zwischen Vergleichen, die in ihrem Vollzug signalisieren oder sonst kenntlich machen, dass es sich um einen Vergleich handelt (explizite Vergleiche), und solchen, die dies nicht tun (implizite Vergleiche). Davy et al. (2019) haben diese Unterscheidung weiter differenziert, indem sie „Grade des Kenntlichmachens und der Erkennbarkeit von Vergleichen“ (Davy et al. 2019, S. 14 ff.) unterscheiden, die von ausdrücklicher Thematisierung („ich vergleiche jetzt a mit b“) über indirekte sprachliche Markierungen („a ist größer als b“) bis zu „Schwundstufen“ der Kenntlichkeit reichen, in denen der Vergleichsvorgang nicht mehr im praktischen Vollzug selbst signalisiert wird, wohl aber von informierten Beobachtern aus dem Kontext erschlossen werden kann. Eine solche Analyse der praktischen Vollzüge des Vergleichens hat bemerkenswerte heuristische Konsequenzen. Wer nicht mit Graden der Explizitheit von Vergleichen rechnet, wird implizite Vergleiche gar nicht als solche erkennen und analysieren können. Hier wird exemplarisch deutlich, dass und wie die Aufmerksamkeit für die praktisch-performative Seite die Begriffsbildung schärft. Dasselbe möchten wir nun an der Zeitlichkeit des Vergleichens zeigen.

2.3 Zur Zeitlichkeit des Vergleichens: Serielle Vergleiche

Wir schlagen vor, zwischen drei typischen zeitlichen Vergleichspraktiken oder Zeitlichkeitsdimensionen von Vergleichspraktiken zu unterscheiden: (1) temporale, (2) temporalisierende und (3) temporalisierte Vergleiche. *Temporale Vergleiche*

stellen *zeitlich spezifizierte Ähnlichkeiten oder Unterschiede* fest. Das können vage Referenzen auf Vergangenheit, Gegenwart oder Zukunft sein („früher war $A=X$, jetzt ist $A=Y$ “; „gestern warst Du schlecht gelaunt, heute bist Du gut drauf“) oder zeitbegrifflich spezifizierte Referenzen („im April 2019 war $A=X$, im Januar 2026 wird $A=Y$ sein“). *Temporalisierende Vergleiche* dagegen stellen auch *interpretative Verknüpfungen zwischen Zeitbezügen* her. Sie stellen nicht nur zeitliche Ähnlichkeiten und Differenzen fest, sie interpretieren diese und betten sie in Narrative ein, die Kontinuität und Wandel über die markierten Zeiträume hinweg zu beobachten erlauben. Auch solche Interpretationen können mehr oder weniger spezifisch ausfallen und reichen von pauschalen historischen Diagnosen („die Menschheit hat sich seit der Steinzeit im Großen und Ganzen zum Guten entwickelt“) bis zu quantitativen, auf die Kommastelle genau bezifferten Entwicklungsvergleichen („die Universität A war im Ranking 2015 auf Platz 150 mit dem Leistungswert 3,75; sie hat sich aber durch Verbesserungen im Bereich der Forschung bis 2019 auf Platz 88 hochgearbeitet und dabei den Leistungswert 5,22 erreicht“). Temporales und temporalisierendes Vergleichen beziehen sich auf die informative Dimension des Vergleichens. Sie vergleichen unter zeitlichen Gesichtspunkten.

Das *temporalisierte Vergleichen* bezieht sich auf die performativ-praktische Dimension des Vergleichens: In welcher zeitlichen Abfolge finden Vergleiche statt? Einmalig oder wiederholt? Regelmäßig und unregelmäßig? Selten oder häufig? Akzeptiert man, dass sich Praktiken bereits an sich durch Repetitivität auszeichnen (vgl. Reckwitz 2015; Schäfer 2013), dann gilt dies für temporalisierte Vergleichspraktiken in einem doppelten Sinne: Es geht um Praktiken, die nicht nur auf Wiederholung beruhen, *sondern in sich repetitiv strukturiert sind*. Die temporale Ordnung von Vergleichspraktiken kann drei zusammenhängende Gesichtspunkte betreffen: Zunächst (1) die *Wiederholung* selbst, die gegeben ist, wenn einzelne Vergleichsereignisse auf ähnliche Weise erneut stattfinden, sodass die Ergebnisse einzelner Vergleichsakte aufeinander bezogen, ihrerseits für vergleichbar erklärt und (auch) unter zeitlichen Gesichtspunkten miteinander verglichen werden können (z. B. „Wirtschaftswachstum des Landes A im Jahr X [festgestellt zum Zeitpunkt XY]“ im Vergleich mit dem „Wirtschaftswachstum des Landes A im Jahr Y [festgestellt zum Zeitpunkt YX]“). (2) *Regelmäßigkeit*: Wiederholung kann mehr oder weniger regelmäßig stattfinden. So ermöglicht ein Ranking beispielsweise andere zeitliche Vergleiche, wenn es in regelmäßigen Abständen wiederholt wird, weil der regelmäßige Vergleich auch Leistungsentwicklungen in definierten Zeiträumen (wöchentlich, jährlich usw.) zu erfassen und zu bewerten erlaubt. (3) *Häufigkeit/Frequenz*: Schließlich kann für die Analyse der temporalen Ordnung von Vergleichspraktiken auch die Frequenz der Wiederholungen von Interesse sein. Fünf Rankings zwischen 2000 und 2015 erlauben weniger und andere zeitliche Vergleiche als 15 (jährliche) Rankings im selben Zeitraum.

Serielle Vergleiche verstehen wir nun als spezifischen Typus temporalisierter Vergleichspraktiken, der vorliegt, wenn wiederholt *und* regelmäßig stattfindende Vergleichsereignisse sich durch im Kern stabile Vergleichskategorien und -kriterien sowie wechselseitige Verweise miteinander verknüpfen und dadurch eine Serie zusammenhängender Vergleichsereignisse bilden. Wir schränken den Begriff damit auf temporalisierte Vergleiche ein, die wiederholt und regelmäßig stattfinden, ohne

ihn auf eine bestimmte Häufigkeit der Wiederholung festzulegen. Die Frequenz der Wiederholung serieller Vergleiche kann vielmehr variieren zwischen mehrjährigen Abständen und täglichen Wiederholungen bis hin zu ständigen Updates, in denen die Wiederholung selbst nicht mehr explizit zutage tritt, aber ständig praktiziert wird. Serielle Vergleiche gibt es zudem in vielfältiger medialer Gestalt, insbesondere in qualitativer (z. B. narrative nationale Entwicklungsvergleiche), quantitativer (z. B. statistische Häufigkeitsvergleiche) und visueller Form (z. B. Landkarten); auch ihr Beobachtungszweck kann variieren, etwa auf die Ermittlung von Häufigkeiten, Wahrscheinlichkeiten oder Leistungen abzielen.

Indem serielle Vergleiche die Merkmale der Wiederholung und Regelmäßigkeit kombinieren, erschließen sie die Möglichkeit, Ergebnisse von Vergleichen, die zu verschiedenen Zeitpunkten stattfinden, ihrerseits wiederholt und regelmäßig miteinander zu vergleichen und die Resultate selbst zum Thema zu machen. Das erlaubt nicht zuletzt die Beobachtung von Stabilität und Wandel mit Blick auf *einzelne* Vergleichseinheiten (z. B. Akteure) und *Relationen* zwischen verschiedenen Vergleichseinheiten (z. B. Leistungsdifferenzen zwischen Akteuren), wie auch eine Kombination akteurs- und feldbezogener zeitlicher Vergleiche (z. B. mit Blick auf Verschiebungen im globalen Feld der Universitäten). Aus diesen Gründen vermuten wir eine wechselseitig konstitutive Beziehung zwischen dem seriellen Vergleichen einerseits und dem temporalen oder temporalisierenden Vergleichen andererseits, der wir im nächsten Abschnitt an der Geschichte der Rankings exemplarisch nachgehen wollen.

3 Rankings als serielle Vergleichspraktiken: Ein historischer Abriss von Kunst- und Hochschulrankings

Rankings sind ein paradigmatisches Beispiel für die Wirkungsmacht serieller Vergleichspraktiken in der Moderne. Sie vereinen verschiedene Teilaspekte in sich: (1) Sie vergleichen und bewerten Leistungen (z. B. wissenschaftliche Exzellenz, Serviceorientierung von Hotels, menschliche Entwicklung, künstlerischer Ausdruck etc.); (2) diese Leistungsvergleiche liegen entweder in der Form von Zahlen vor oder werden in eine solche Form gebracht (zu Varianten der Kombination quantitativer und qualitativer Vergleiche in Ranglisten s. Heintz 2019); (3) sie werden visualisiert, oftmals, aber nicht nur, in Form von Tabellen; (4) und schließlich wird der tabellarische Leistungsvergleich wiederholt veröffentlicht, womit er (5) (Leistungs-)Konkurrenz zwischen den verglichenen Einheiten suggeriert (Brankovic et al. 2018; Ringel et al. 2020; Ringel und Werron 2019, 2020; Werron und Ringel 2017). Im Unterschied zu anderen konzeptionellen Beiträgen betont unser Rankingbegriff in den Punkten (3)–(5) die praktisch-performative Seite von Rankings. Die nachfolgende historische Analyse von Kunst- und Hochschulrankings macht selektiv von diesem Begriff Gebrauch, um die Genese von Serialität vergleichend zu untersuchen.

Hervorheben möchten wir im Anschluss an unsere allgemeinen Erörterungen zur Zeitlichkeit des Vergleichens die sachliche und zeitliche Unabgeschlossenheit der Vergleichsserien, auf die unser Rankingbegriff hinweist. Der öffentliche Leistungsvergleich kann im Prinzip beliebig häufig wiederholt werden und beliebig viele

Vergleichseinheiten zueinander in Bezug setzen, und der Vergleichsserie sind weder in zeitlicher noch sachlicher Hinsicht natürliche Grenzen gesetzt. Je länger sie andauert, desto umfangreicher wird zudem die Menge der Ereignisse („Daten“), die in das nächste Ranking eingehen können. Zugleich ist immer ein nächstes Vergleichsereignis möglich, das antizipiert werden kann oder muss. Damit schafft sich die Vergleichsserie ihre eigene Geschichte und Zukunft.

In ihrer institutionalisierten Form tragen Rankings zur Stabilisierung dynamischer Konkurrenzfelder bei, indem sie den von ihnen verglichenen Akteuren mit jeder Veröffentlichung aufs Neue suggerieren, dass sie um (künstlerische, wissenschaftliche, medizinische etc.) Reputation konkurrieren. Diese Eigenschaften verleihen Rankings ein spezifisches Unterhaltungs- und Streitpotenzial und haben sie zu einer der prominentesten aller Vergleichs- und Bewertungspraktiken werden lassen. Die Institutionalisierung von Rankings ist aber von Feld zu Feld unterschiedlich verlaufen, hat einmal mehr, einmal weniger Kritik hervorgerufen und ihnen einmal mehr, einmal weniger Einflussmöglichkeiten erschlossen. Wer heutige Rankings verstehen will, sollte daher von vorschnell verallgemeinernden Aussagen über ihre Wirkungsmacht absehen und stattdessen ihre historische Institutionalisierung sorgfältig rekonstruieren. Das gilt insbesondere mit Blick auf die uns hier interessierende Frage, wie und mit welchen Folgen die Serialität von Rankings „in die Gänge gekommen“ ist. Unsere empirische Analyse konzentriert sich auf Rankings und ihre historischen Vorgänger in Kunst und Wissenschaft. Exemplarisch untersuchen wir die Verknüpfung des seriellen Vergleichens mit dem quantitativen Bewerten, einschließlich der gegen sie gerichteten Kritik, und zeigen Gemeinsamkeiten und Unterschiede der Institutionalisierung von Kunst- und Hochschulrankings auf.

3.1 Kunst: Von Protorankings des 18. Jahrhunderts zu heutigen Kunstrankings

Temporales und temporalisierendes Vergleichen tritt im künstlerischen Feld vermehrt ab dem 19. Jahrhundert auf, und zwar vornehmlich in historisierender Absicht: Wer ist Schüler von wem? Wie hat sich das Werk von Künstler X im Laufe der Zeit entwickelt? Vergleiche dieses Typs sind seither zu einer normalen und vielfach gelebten Beobachtungspraxis geworden und nehmen unter anderem die Gestalt elaborierter Stammbäume der Kunst an (Schmidt-Burkhardt 2005). Diese Form der Analyse von Temporalität wandelte sich im Laufe der Zeit in eine häufig wissenschaftlich betriebene Praxis, insbesondere in der Kunstgeschichte. Aber nicht nur der klassifizierend-temporale Vergleich, auch die Bewertung der ästhetischen Qualität von Kunst findet im Allgemeinen explizit statt. Sie geht mit dem Aufstieg der Kunstkritik einher, die für sich in Anspruch nimmt, legitime Sprecherpositionen für die Beurteilung von Kunstwerken bereitzustellen. Insofern ist die Reflexion über Kunst von einer Vielzahl von wissenschaftlichen und professionellen, oftmals eng verzahnten Praktiken des Bewertens (wie ist ein Kunstwerk zu beurteilen?), Klassifizierens (welchem Stil, welchem Genre, welcher Schule ist ein Kunstwerk zuzuordnen?) und temporalen Vergleichens (wer tritt in wessen Fußstapfen?) durchzogen (Grave et al. 2020). Doch wie steht es um die Verwendung von Zahlen? Gibt es Versuche, quantitative Leistungsbewertungen in eine serielle Vergleichspraxis – und schlussendlich in Kunstrankings – zu überführen? Auf welche Reaktionen stoßen solche Versuche?

Wir diskutieren diese Frage anhand ausgewählter Beispiele, die einen relativ langen Zeitraum umfassen: von frühen Experimenten im 18. und 19. Jahrhundert über den seit den 1970er-Jahren veröffentlichten *Kunstkompass* bis hin zu Ranglisten jüngeren Datums.

Spoerhase (2014) zeigt, dass es bereits im 18. Jahrhundert einen regen Diskurs um und zahlreiche Experimente mit zahlenförmigen Praktiken des vergleichenden Bewertens gab, die vielfach auch einen rankingähnlichen Charakter hatten. Interessanterweise sind diese Versuche nicht Teil des institutionalisierten Gedächtnisses der Reflexion über Kunst geworden, sondern wurden schon bald wieder vergessen oder als Skurrilität abgetan. Gemeint sind quantitativ-vergleichende Bewertungen, die auf die 1708 von Roger de Piles veröffentlichte *Balance des Peintres* zurückgehen. De Piles verglich Kunstschaffende in verschiedenen Dimensionen und bewertete sie jeweils mit 0–20 Punkten, jedoch brachte er die Bewerteten nicht in eine Rangfolge, sondern ordnete sie alphabetisch; auch auf eine die Einzelbewertungen summierende oder verrechnende Gesamtbewertung verzichtete er. Offensichtlich traf er einen Nerv und in den darauffolgenden Jahrzehnten erschienen in verschiedenen europäischen Ländern und Sprachen eine Reihe ähnlicher Tabellen. Im Vergleich mit heutigen Rankings fällt vor allem auf, dass die Tabellen nahezu ausschließlich bereits verstorbene Kunstschaffende der vergleichenden Bewertung unterzogen. Die erste vollständige, nicht nur vergleichende, sondern auch nach ästhetischen Bewertungen von oben nach unten hierarchisierende, „steigernde“ (Steinmetz 2015) Rangliste, die *Balance des Peintres, rectifiée*, wurde im Jahr 1810 veröffentlicht, markierte jedoch gleichzeitig auch das Ende dieses frühen Diskurses (Abb. 1).

Wie sich anhand der Kommentare des Erstellers der Rangliste erkennen lässt, löste insbesondere die quantitative Bewertung großes Unbehagen aus – selbst bei jenen, die solche Bewertungen durchführten. Man solle „die Waagen den Händlern überlassen“, so Sobry (1810 zit. n. Spoerhase 2014, S. 125), damit demonstrierend, dass er quantitative Bewertungen mit ökonomischen Motiven, nicht aber der sachgemäßen Beurteilung von Kunstwerken assoziierte. Wohl auch deshalb konnten sich in diesem frühen Diskurs zwar einige, nicht aber alle Elemente moderner Rankings versammeln. Da die tabellarischen Leistungsvergleiche nur einmalig veröffentlicht wurden, fehlte ihnen insbesondere das serielle, auf regelmäßige Wiederholung angewiesene Element.

Quantifizierend-bewertende, steigernde Leistungsvergleiche verschwanden ab diesem Zeitpunkt offenbar weitestgehend aus dem künstlerischen Feld und tauchten erst in den 1970er-Jahren wieder auf. Das erste Kunstranking im engeren (oben definierten) Sinne ist der *Kunstkompass*, der noch immer in jährlichen Intervallen veröffentlicht wird. Der Publikation dieses Rankings geht eine Reihe von Problemdefinitionen voraus, von denen hinsichtlich des Aspekts der Serialität zwei besonders hervorzuheben sind (Wilbers 2019).² Erstens, die Intransparenz des Kunstmarktes, dessen opake Dynamiken die Produzenten des *Kunstkompasses* aufzudecken suchten (ein Motiv, das auch heutzutage immer wieder anzutreffen ist); zweitens, ein temporalisierender Vergleich, der von der Diagnose ausging, dass

² Die Analyse des *Kunstkompasses* hat stark von Wilbers (2019) profitiert, dem wir an dieser Stelle nochmal unseren Dank aussprechen möchten.

Abb. 1 *Balance des Peintres, rectifiée.* (Sobry 1810 in Spoerhase 2014, S. 121)

<i>Balance des Peintres, rectifiée.</i>					
N O M S.	Compo- sition.	Dessin.	Coloris	Expres- sion.	Résul- tat.
Raphaël	17	17	12	17	63
Poussin	18	18	9	18	63
Leueur	18	17	9	18	62
Dominiquin	14	17	10	16	57
Michel-Ange	16	18	6	15	55
Léonard-de-Vinci	15	16	10	14	55
Corrège	12	14	17	12	55
Titien	13	16	18	8	55
Rubens	14	14	17	9	54
Lorrain	17	10	17	7	51
Wandick	12	15	17	8	52
Lebrun	15	15	8	14	52
Carache	12	15	11	11	49
Albane	12	14	10	10	46
Pierre de Costone	12	12	12	6	42
Mignard	11	11	11	9	42
Burdon	12	11	10	9	42
Labire	11	12	12	7	42

die deutsche Kunstszene der amerikanischen hinterherhinke und behauptete, dass dieser Rückstand durch die Herstellung von Transparenz und eine verstärkte Dynamisierung des Marktes aufgeholt werden könne. Die quantitative Bewertung selbst übernahmen im Unterschied zu den frühen Experimenten nicht die Produzierenden des Bewertungsschemas, sondern als kompetent erachtete Akteure im Feld, die vom Herausgeber des Kunstkompasses, Willi Bongard, einen Fragebogen erhielten. Diese Einschätzungen dienten dann als Grundlage für die Entwicklung von Bewertungsmaßstäben für künstlerische Leistungen (Baumann 2001). Bewertet wurden nun auch nicht mehr die künstlerische Qualität des Werkes oder die Kunstschaffenden selbst, sondern deren „Ruhm“, d. h. ihre Reputation (vgl. auch Buckermann 2020). Der erste Kunstkompass, publiziert im Jahr 1970 in der Zeitschrift *Capital*, verteilte zwar Ränge, ordnete die Kunstschaffenden jedoch – insofern an die *Balance des Peintre* erinnernd – alphabetisch und nicht gemäß ihres Rangplatzes (Abb. 2).

Erst ab der zweiten Ausgabe präsentierte sich der Kunstkompass als Ranking im engeren Sinne, indem er die Kunstschaffenden hierarchisch anordnete. Außerdem dynamisierte er die Vorstellung von Zeitlichkeit im Feld: Dem je aktuellen Rang stellte er den aus dem vergangenen Jahr gegenüber und verband damit das vorhergehende mit dem aktuellen Vergleichsereignis (eine Darstellungsform, die heutige Rankings – in ganz unterschiedlichen Feldern – häufig anwenden). Diese Praxis wurde seitdem beibehalten. Wie schon im ersten Jahr, so stellten auch die zweite und nachfolgende Ausgaben des Kunstkompasses eine explizite Verbindung zwischen dem „Ruhm“ von Kunstschaffenden und dem Marktwert ihrer Werke her; neu war jedoch, dass seit der zweiten Ausgabe auch eine Kategorie „Preisnote“

Name	Nation	Stil	Preise DM	Punkte	PPR Preis/Punkt-Relation	Rang
Albers	USA	Konstruktivismus	von 40 000 bis 50 000	2340	von 17 bis 21	25
Andre	USA	Minimal Art	bis 40 000	1290	bis 31	62
Antes	Deutschland	Neue Figuration	von 10 000 bis 20 000	2090	von 5 bis 10	31
Anuszkiewicz	USA	Op Art	bis 20 000	870	bis 23	93
Arakawa	Japan	Neuer Realismus	bis 7 500	1245	bis 6	69
Arman	Frankreich	Neuer Realismus	von 20 000 bis 40 000	3495	von 6 bis 11	7
Bell	USA	Minimal Art	von 25 000 bis 30 000	1535	von 16 bis 20	48
Berrocal	Spanien	Plastik	von 20 000 bis 30 000	1270	von 16 bis 24	67
Beuys	Deutschland	Aktionen, Prozeßkunst	von 60 000 bis 80 000	2000	von 30 bis 40	32
Bill	Schweiz	Konstruktivismus	bis 60 000	1660	bis 36	42
Blake	England	Pop Art	von 8 000 bis 35 000	1090	von 7 bis 22	82
Bontecou	USA	Plastik	von 40 000 bis 65 000	1080	von 37 bis 60	83
Brüning	Deutschland	Neuer Realismus	bis 12 000	1560	bis 8	46
Bury	Belgien	Kinetik	von 10 000 bis 30 000	1350	von 7 bis 22	59
Caro	England	Plastik	von 80 000 bis 100 000	1275	von 63 bis 78	66
César	Frankreich	Plastik	von 10 000 bis 20 000	1995	von 5 bis 10	33
Chamberlain	USA	Plastik	von 10 000 bis 20 000	1290	von 8 bis 15	62
Christo	Bulgarien	Neuer Realismus	von 10 000 bis 35 000	2100	von 5 bis 17	30
Chryssa	Griechenland	Lichtkunst	von 10 000 bis 40 000	1230	von 8 bis 33	72
Colombo	Italien	Kinetik	bis 10 000	940	bis 11	91
D'Arcangelo	USA	Pop Art	bis 30 000	980	bis 31	88
Demarco	Brasilien	Lichtkunst	von 2 000 bis 12 000	1050	von 2 bis 11	85
De Maria	USA	Land Art, Concept Art	bis 50 000	860	bis 58	95
Dine	USA	Pop Art	von 30 000 bis 45 000	3080	von 10 bis 15	10
Fahlström	Schweden	Neuer Realismus	bis 40 000	1420	bis 28	54
Flavin	USA	Minimal Art	von 12 000 bis 25 000	1910	von 6 bis 13	36

Abb. 2 Kunstkompass 1970. (Rohr-Bongard 2001, S. 37)

(„sehr billig“, „billig“, „preiswert“, „teuer“, „sehr teuer“) aufgenommen wurde, die lohnenswerte Investitionen auszeichnet (Abb. 3).

Die Verbindung von „Ruhm“ und Marktwert, so lässt sich vermuten, zielte darauf ab, die Dynamisierung des künstlerischen Feldes durch serielle Vergleiche weiter voranzutreiben. Explizit macht dies beispielsweise der Kunstkompass von 1976:

Sie möchten wissen, welche Künstler der Gegenwart sich mit ihren Werken internationale Anerkennung verschafft und entsprechend Aussicht auf Bestand haben? Dafür liefert die Rangliste der „100 Großen“ Anhaltspunkte. Je höher der Punktwert, der den einzelnen Künstlern aufgrund von Ausstellungenbeteiligungen, Museumskäufen, Literaturhinweisen und so weiter beigemessen wird, desto größer ist die Wahrscheinlichkeit, daß sie ihren Marktwert behalten oder gar noch steigern (zit. n. Rohr-Bongard 2001, S. 54).

Obwohl der Kunstkompass, insbesondere wegen seines Fokus auf den Kunstmarkt, zum Teil heftig kritisiert wurde, publizierte Willi Bongard ihn auch weiterhin in der Zeitschrift *Capital*. Seit den 1980er-Jahren scheint seine Legitimität nicht mehr fundamental hinterfragt zu werden; jedenfalls findet die kritische Diskussion öffentlich kaum noch statt und mitunter wird der Kunstkompass gar von feldinternen „Gatekeepern“ in Schutz genommen (Wilbers 2019). Die in den 1970er-Jahren einsetzende Temporalisierung des Kunstkompasses ging in den 1980er-Jahren weiter. Es wurden nicht mehr nur die Werte aus den vergangenen Jahren abgebildet,

Rangliste der Künstler (1–50)								Punkte
Rang 1971	Rang 1970	Name	Nation	Richtung	Preise für Grafik und Objekte DM	Preisnote	PPR	
1	1	Rauschenberg	USA	Neuer Realismus	1000–5000	preiswert	0,65	5895
2	5	Oldenburg	USA	Pop Art	600–8000	sehr teuer	1,28	4705
3	4	Johns	USA	Neuer Realismus	1600–10000	sehr teuer	2,13	4685
4	6	Tinguely	Schweiz	Kinetik	200–3000	preiswert	0,68	4420
5	8	Klein	Frankreich	Neuer Realismus	5500	sehr teuer	1,29	4260
6	10	Dine	USA	Pop Art	350–1500	billig	0,60	4170
7	7	Arman	Frankreich	Neuer Realismus	150–1200	sehr billig	0,29	4155
8	32	Beuys	Deutschland	Aktionen	bis 3000	preiswert	0,78	3935
9	9	Lichtenstein	USA	Pop Art	280–4500	sehr teuer	1,17	3930
10	11	Stella	USA	Neue Abstraktion	650–5000	sehr teuer	1,31	3820
11	13	Soto	Venezuela	Op Art	280–2000	billig	0,54	3700
12	14	Warhol	USA	Pop Art	500–2600	preiswert	0,71	3645
13	12	Paolozzi	England	Plastik	150–2000	billig	0,56	3585
14	16	Hockney	England	Pop Art	480–1800	billig	0,59	3405
15	15	Rosenquist	USA	Pop Art	200–2400	preiswert	0,78	3095
16	27	Morris	USA	Minimal	—	—	—	3080
17	18	Uecker	Deutschland	Zero	8–300	sehr billig	0,10	3075
18	30	Christo	Bulgarien	Neuer Realismus	160–800	sehr billig	0,27	2980
19	24	Noland	USA	Neue Abstraktion	1200	billig	0,41	2935
20	17	Louis	USA	Neue Abstraktion	—	—	—	2900
21	22	Rayese	Frankreich	Neuer Realismus	bis 3000	sehr teuer	1,07	2815
22	21	Kelly	USA	Neue Abstraktion	—	—	—	2725
23	19	Mack	Deutschland	Zero	140–990	billig	0,37	2710
24	20	Pienn	Deutschland	Zero	120–200	sehr billig	0,07	2690
25	29	Judd	USA	Minimal	70–900	billig	0,34	2685
26	33	Cesar	Frankreich	Plastik	200	sehr billig	0,08	2650
27	35	Riley	England	Op Art	350–1500	billig	0,57	2650
28	28	Segal	USA	Pop Art	400–4500	sehr teuer	1,71	2635
29	50	Kienholz	USA	Environment	—	—	—	2585
30	26	Jones	England	Pop Art	180–850	billig	0,33	2570
31	23	Kita	USA	Pop Art	450–1200	billig	0,47	2555
32	31	Antes	Deutschland	Neue Figuration	160–650	billig	0,27	2425
33	36	Flavin	USA	Minimal	3200	sehr teuer	1,32	2420
34	34	Le Parc	Argentinien	Kinetik	480–2000	teuer	0,86	2330
35	40	Wesselmann	USA	Pop Art	400–2000	teuer	0,87	2305
36	44	Pistoletto	Italien	Neuer Realismus	—	—	—	2285
37	62	Andre	USA	Minimal	—	—	—	2230
38	39	Tilson	England	Pop Art	200–600	sehr billig	0,27	2200
39	41	Indiana	USA	Pop Art	280–800	billig	0,36	2195
40	45	Smith, R.	England	Neue Abstraktion	400–800	billig	0,37	2155
41	37	Hamilton	England	Pop Art	150–4800	sehr teuer	2,24	2145
42	38	Schoeffar	Frankreich	Lichtkunst	200–350	sehr billig	0,17	2110
43	75	Spoerri	Schweiz	Neuer Realismus	—	—	—	2110
44	59	Bury	Belgien	Kinetik	550–1200	billig	0,58	2075
45	71	Naumann	USA	Minimal	300	billig	0,15	2040
46	61	Lewitt	USA	Minimal	130–400	billig	0,20	1995
47	47	Hundertwasser	Osterreich	Phant. Realismus	540–2800	sehr teuer	1,41	1990
48	43	Rivers	USA	Pop Art	185–2000	sehr teuer	1,01	1985
49	60	St. Phalle	Frankreich	Neuer Realismus	400–500	sehr billig	0,25	1985
50	83	Bontecou	USA	Plastik	550–750	billig	0,38	1965

Abb. 3 Kunstkompass 1971. (Rohr-Bongard 2001, S. 42)

sondern Auf- und Abwärtsbewegungen innerhalb des Rankings explizit hervorgehoben: Ausgezeichnet werden „Aufsteiger“, „Heimliche Gewinner“, „Punktegewinner“, „Nachrücker“ und „Die neuen Kompasskünstler“, um nur einige Beispiele zu nennen (Abb. 4).

Diese dynamisierenden Interventionen scheinen ihre Spuren hinterlassen zu haben. So hieß es z. B. in der *FAZ* (Firschnig 2002): „Alljährlich im Herbst fiebert die deutschsprachige Kunstgemeinde auf die Novemberausgabe der Zeitschrift ‚Capital‘. Im dortigen Kunst-Kompass kann man sehen, welche Künstler angesagt sind und welche auf der Abschlusliste stehen.“ Auch in ironischen Bezugnahmen wird dieser Effekt deutlich, beispielsweise in der folgenden Bemerkung des Künstlers Sigmar Polke: „Ich wollte nie die Nummer eins werden, denn dann kann ich ja nicht mehr steigen“ (Rohr-Bongard 2019). Ohne weiter ins Detail zu gehen, zeugen diese

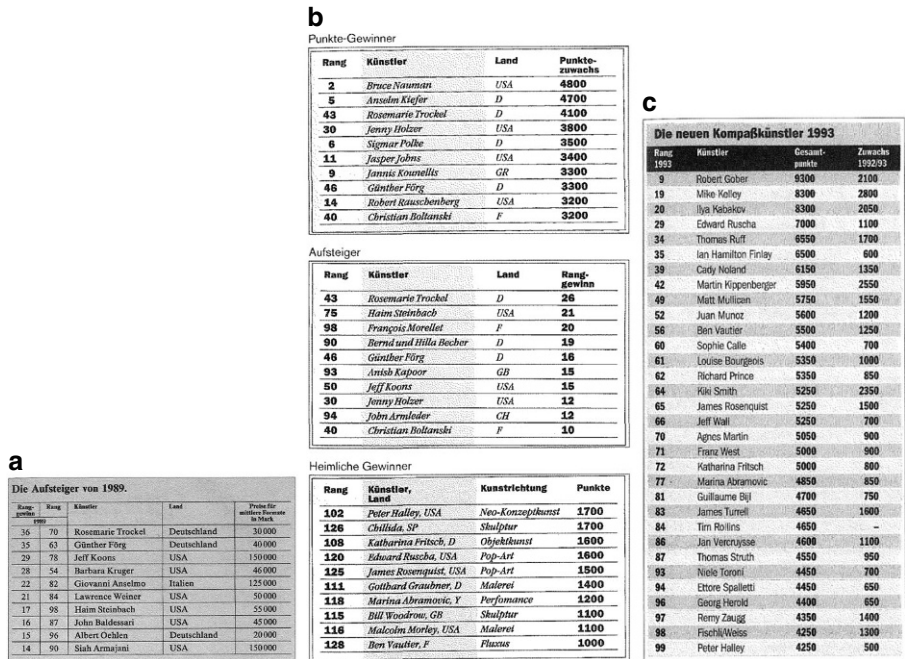


Abb. 4 a Kunstkompass 1989. (Rohr-Bongard 2001, S. 85), b Kunstkompass 1991. (Rohr-Bongard 2001, S. 93), c Kunstkompass 1993. (Rohr-Bongard 2001, S. 96)

und ähnliche Aussagen doch davon, dass sich im Feld Indizien für eine veränderte Wahrnehmung von Zeitlichkeit finden lassen.

Während der Kunstkompass, ähnlich Rankings in anderen Feldern, eine Zeitlichkeit etabliert, die auf jährliche, stabile Rhythmen abhebt, kommen ab den 2000er-Jahren im künstlerischen Feld weitere Experimente mit quantifizierend-bewertenden Vergleichen auf, die den seriellen Leistungsvergleich erneut beschleunigen. Hervorzuheben sind die Internetseiten *Artfacts* und *ArtRank*, die den vom Kunstkompass gestarteten Trend der Korrelation von Reputation und Marktwert weiterführen und im Fall von *ArtRank* noch expliziter verfolgen. Beide Rankings werden kontrovers diskutiert. *ArtFacts*, bestehend seit 2004, vermisst nicht nur je aktuelle Gegenwarten anhand von Ausstellungserfolgen, sondern dehnt den zeitlichen Horizont immer weiter in die Vergangenheit aus, indem beständig neue Informationen in die Datenbank aufgenommen werden (Claaßen 2018). Das Besondere an *ArtFacts* ist darüber hinaus, dass die Liste nicht jährlich (wie im Falle des Kunstkompasses), sondern wöchentlich auf den neuesten Stand gebracht wird. In einem Onlineportal melden Kunstschaffende sich an und tragen unablässig neue Informationen über ihre Ausstellungen ein. So werden die Gerankten selbst zu einer wichtigen Informationsquelle und nehmen beständig an der (Neu-)Kalkulation des Rankings teil. Und wie der Kunstkompass, so markiert auch *ArtFacts* im Ranking die zeitliche Dimension in der Gestalt von „Trends“, die durch Pfeile symbolisiert werden.

ArtRank zeichnet sich dadurch aus, dass es von allen Kunstrankings am deutlichsten den Marktwert hervorhebt. Wie der Kunstkompass will *ArtRank* nicht nur über den je aktuellen Zustand informieren, sondern auch zukünftige Entwicklungen prognostizieren, um Hinweise darauf zu geben, welche Kunstwerke lohnenswerte *Investitionsobjekte* sind. Dafür wird eine ganze Reihe von Daten gesammelt, von öffentlich verfügbaren, bei Auktionen erzielten Preisen, über die Sichtbarkeit in sozialen Medien (Facebook, Twitter, Instagram) bis hin zu Insiderinformationen über die „Produktionskapazitäten“ der Kunstschaffenden. Dabei kommen sogar Termini aus der Finanzwelt wie „buy now“, „early blue chip“, „sell now“ und „liquidate“ zur Anwendung, um gute oder schlechte Investitionen auszuzeichnen. Das Ranking teilt demzufolge nicht nur Beobachtungen und Vermessungen des Feldes mit, sondern gibt dem Publikum auch explizite Handlungsanweisungen. Der Fokus auf Preise und zukünftige Preisentwicklungen impliziert, so Velthuis (2014), eine sehr viel stärkere Dynamik in der Tabelle. Der *ArtRank* inhärenten Logik gemäß sollte sich die Rangordnung permanent in Bewegung befinden.

Wenngleich *ArtRank* betont, keine ästhetische Bewertung vorzunehmen, wird dem Unternehmen doch unterstellt, neue – wirtschaftlich geprägte – Verständnisse von Kunst in das Feld einzuführen. Dementsprechend lautet der Titel eines Artikels: „Wie *ArtRank* sich zur meistgehassten Website in der Kunstszene entwickelte“ (artinfo24.com 2015). Tenor der Kritik in diesem wie auch in anderen Artikeln ist, dass das Ranking zur Kommerzialisierung des Kunstbetriebes beitrage und Bewertungsdimensionen stark mache, die nichts mit Kunst zu tun hätten. Ein zentraler Aspekt der Kritik ist die veränderte Zeitlichkeit des Auf- und Abbaus von Reputation. Kunstschaffende hätten plötzlich das „Etikett eines Verfallsdatums“ (artinfo24.com 2015). Dies wird dann besonders deutlich, wenn, wie in einem anderen Zeitungsartikel, die zeitliche Dynamik des Aufbaus von künstlerischer Reputation auf der einen und von Marktwerten auf der anderen Seite einander gegenübergestellt werden: „Art reputations take years to develop and moments to ruin; consume the young and you consume the future“ (Helmore 2014). Spätestens hier wendet sich die Kritik auch gegen die Serialität des Vergleichens selbst. Ob sie den Einfluss von Rankings im Kunstbetrieb abzubremesen vermag, bleibt jedoch offen.

3.2 Hochschulen: Zur allmählichen Entstehung von Hochschulrankings

Die Vorgeschichte der quantitativen Bewertung wissenschaftlicher Leistungen reicht im Unterschied zur Kunst nicht bis ins 18. Jahrhundert zurück. Dafür lassen sich im Verlauf des 20. Jahrhunderts in den USA, und zunächst nur dort, zahlreiche Einzelveröffentlichungen von Ranglisten beobachten, die Hochschulen miteinander vergleichen. Auch diese verstetigten sich jedoch die längste Zeit nicht in seriellen Vergleichspraktiken. Auffällig ist vor allem ihre große Menge: Zwischen 1898 und 1983 – also noch vor der explosionsartigen Vermehrung von Rankings in den vergangenen Jahrzehnten – wurden, so das vorläufige Ergebnis unserer Recherchen, mindestens 277 Klassifikationen und Ranglisten von Hochschulen veröffentlicht. Grob lassen sich vier Entwicklungsphasen unterscheiden: In der ersten Phase, von 1898–1949, erschienen 123 Rankings von 21 unterschiedlichen Autoren, jedoch nur in sieben verschiedenen Medien. Von 1950–1983 erschienen 154 Rankings von

106 unterschiedlichen Autoren, veröffentlicht in 57 verschiedenen Medien.³ Ab den 1980er-Jahren kommen, ausgehend von den USA, vermehrt regelmäßig erscheinende Hochschulrankings in verschiedenen westlichen (dann auch nichtwestlichen) Ländern auf. Und seit den frühen 2000er-Jahren werden schließlich auch globale Rankings publiziert, die in der Regel auf Universitäten fokussiert sind. Heute gibt es zehn globale und mehr als 150 nationale und/oder fächerspezifische, regelmäßig erscheinende Hochschulrankings (Stack 2018, S. 31).

Von besonderem Interesse ist, dass in den 1980er-Jahren ein Bruch stattfand, der zur Entstehung und Institutionalisierung serieller Ranglisten führte. Die Gründe für diesen Bruch sind jedoch nur zu verstehen, wenn man seine lange Vorgeschichte berücksichtigt. Historisch lassen sich im Wesentlichen zwei Formen des Vergleichs wissenschaftlicher Leistungen identifizieren: Bei der ersten Form handelt es sich um Klassifikationsschemata, die wiederholt und regelmäßig veröffentlicht wurden, aber darauf verzichteten oder daran scheiterten, hierarchische Leistungsvergleiche in der Gestalt von Ranglisten einzuführen; bei der zweiten Form handelt es sich um Ranglisten, die nicht wiederholt und/oder regelmäßig veröffentlicht wurden, also noch keine Rankings im oben definierten Sinne waren. Zugespitzt formuliert: Entweder Serie *oder* Rangliste, beides zusammen schien bis in die 1980er-Jahre nicht möglich.

Von *Außen* an Hochschulen herangetragene quantitative Vergleiche gab es bereits früh und auch in serieller Form. Das US-amerikanische „Bureau of Education“ veröffentlichte ab 1870 über mehrere Jahrzehnte regelmäßig statistische Klassifizierungen von Hochschulen, die z. B. Informationen zum Gründungsjahr, zu den aktuellen Präsidenten oder zur Ausstattung der Bibliothek enthielten; Aussagen über Qualität machten sie jedoch zunächst keine (Webster 1986). Im Jahr 1911 erteilte die US-Regierung dem „Bureau of Education“ den Auftrag, ein umfassendes Klassifikationsschema von Hochschulen zu erstellen, „to estimate the work and status of a large group of institutions“ (Babcock 1911, S. 3). Als Datengrundlage des Berichts, veröffentlicht unter dem Titel „A classification of universities and colleges with references to bachelor degrees“, dienten sowohl Statistiken, Kataloge, Register und Reports als auch der Eindruck, den der Ersteller des Schemas sich beim Besuch der Hochschulen machen konnte. Der Report wurde vor der finalen Fertigstellung an die Presse „geleakt“ und von Vertretern der Hochschulen, unter anderem in Leserbriefen, heftig kritisiert. Das Bureau gab schlussendlich nach und räumte ein, die Bezeichnung „classification“ sei „only tentative and confessedly imperfect“ (Claxton 1912, S. 8). Hochschulen hätten insbesondere an der Hierarchisierung und deren Betonung im Titel der Studie Anstoß genommen. Letztlich sah man von einer Veröffentlichung ab. Offensichtlich empfanden es die Hochschulen als besonders problematisch, von einer Bundesbehörde in über- und untergeordnete Leistungsklassen eingeteilt zu werden.

Überhaupt zeigte sich das wissenschaftliche Feld gegenüber solchen von außen herangetragenen Bewertungen skeptisch, wie das folgende Zitat veranschaulicht.

³ Die Erhebung dieser Zahlen, also auch ihre Einteilung in Phase eins und zwei, erfolgten im DFG-Projekt „Zur Institutionalisierung der Rankings: Diskurskarrieren tabellarischer Leistungsvergleiche zwischen 1850 und 1980“ durch Anna-Lena Grüner und Stefan Wilbers.

Alan Cartter, selbst Ersteller eines breit rezipierten Rankings in den 1960er-Jahren (s. u.), berichtet von einem Gespräch mit dem Vizepräsidenten der „National Science Foundation“, in dem dieser die Absicht äußerte, ein Ranking wissenschaftlicher Institute durch Privatanbieter durchführen zu lassen:

Logan Wilson and I hit the ceiling and alerted him ... to the fact that there had been a long-standing position of the Association of Graduate Deans that *nobody* [Hervorhebung im Original] ought to play around with evaluating the graduate programs who was not themselves [sic] responsible to the institutions themselves; either the AAU or some other group. The deans had always said that anyone who is going to accredit us or rate us has to be responsible to the institutions. We don't want any outside body doing this (Dolan 1976, S. 26).

Wissenschaftsinterne Experimente mit quantifizierten Leistungsvergleichen von Hochschulen wurden erstmals Ende des 19. Jahrhunderts durchgeführt, und zwar im Zusammenhang mit der damals breit diskutierten Diagnose, dass der Anteil von „eminent men“ an der Gesamtbevölkerung zurückgegangen sei (Ringel und Werron 2020; Webster 1986). Dabei stand die Frage im Mittelpunkt, warum bestimmte Personen zu wissenschaftlichen Spitzenleistungen in der Lage sind, andere hingegen nicht. Prominent beschäftigte sich Francis Galton mit diesem Thema. Er war der Auffassung, dass Leistungsfähigkeit vererbbar ist und nicht oder bestenfalls marginal mit der Qualität von Hochschulbildung zusammenhängt. So schreibt er, eine von ihm durchgeführte statistische Auswertung resümierend:

One-third of those who sent replies have been educated at Oxford or Cambridge, one-third at Scotch, Irish or London universities, and the remaining third at no university at all. I am totally unable to decide which of the three groups occupies the highest scientific position: they seem to me very much alike in this respect (Galton 1874, S. 236).

Aus einer solchen Sicht ist der zentrale Ort von „eminence“ das erbliche Potenzial der Gesamtpopulation, also die Abstammung („nature“), und nicht etwa die im Laufe eines Lebens durch Ausbildung an dafür vorgesehenen Einrichtungen erworbene und gesteigerte Leistungsfähigkeit („nurture“). Da Hochschulen nach dieser Auffassung keinen nennenswerten Einfluss auf die Entstehung von „eminent men“ haben, erscheint zwar temporalisierendes Vergleichen (in der Gestalt einer Niedergangserzählung), nicht aber serielles Vergleichen der Ausbildungsstätten sinnvoll.

Zunehmend entstand jedoch ein neues Verständnis von „eminence“, das gerade die Bedeutung von Hochschulbildung explizit hervorhob. Hatten in der Kunst Verweise auf den Markt eine Hebefunktion für die Etablierung serieller Ranglisten, so war es in der Wissenschaft die Idee, dass Leistungsfähigkeit von Ausbildung und daher auch den Ausbildungsstätten abhängt. Infolgedessen kam ein reges Interesse an der Vermessung der Qualität von Hochschulen auf. Hervorzuheben sind die Arbeiten des Psychologen James McKeen Cattell, der als prominenter Vertreter der „nurture“-These gelten kann. Entsprechend setzte er sich intensiv mit der Frage auseinander, welche Indikatoren am besten dazu geeignet sind, die Qualität von Hochschulen zu erfassen (Hammarfelt et al. 2017; Ringel und Werron 2020). War

a

Universities.	Humanities.	History and Economics.	Sciences.	Total.
Chicago.....	12	12	12	36
Yale	19	4	11	34
Johns Hopkins	11	3	19	33
Harvard	12	3	11	26
Pennsylvania	9	7	8	24
Columbia.....	7	5	10	22
Cornell.....	7	1	11	19
Clark.....	12	12
Michigan.....	6	1	7
New York.....	4	1	5
Wisconsin	2	1	2	5
Bryn Mawr.....	1	1	1	3
Leland Stanford, Jr.....	2	2
Nebraska.....	2	2
Brown.....	1	1
California.....	1	1
Columbian.....	1	1
Minnesota.....	1	1
Total number of Ph.D. degrees conferred.....	91	38	105	234

b

TABLE XI. THE SCIENTIFIC STRENGTH OF THE LEADING INSTITUTIONS

	Weighted Number.	Gain or Loss.
Harvard.....	146.0	+16.3
Chicago.....	94.6	+18.0
Columbia.....	79.3	-13.3
Hopkins.....	63.4	+ 4.2
Yale.....	61.7	+12.2
Cornell.....	57.6	+ 4.6
Wisconsin.....	49.0	+22.3
Geol. Survey.....	43.8	-12.2
Dept. Agric.....	40.9	- 4.9
Mass. Inst.....	37.7	+ 9.5
Michigan.....	37.1	- 3.5
California.....	32.4	- 5.0
Carnegie Inst.....	30.9	+19.4
Stanford.....	30.0	+ 4.8
Princeton.....	28.6	+ 7.5
Smithsonian Inst.....	26.0	- 7.3
Illinois.....	25.0	+16.7
Pennsylvania.....	24.4	- 4.5
Bur. of Standards.....	18.9	+ 0.1
Clark.....	16.0	+ 2.0

Abb. 5 a Total numbers of Ph. D. degrees conferred. (Cattell 1898, S. 198), b The scientific strength of the leading institutions. (Cattell 1910, S. 591)

es in frühen Publikationen die Menge an vergebenen Dokortiteln (Cattell 1898), so führte er später Umfragen unter Kollegen durch (Cattell 1910; Abb. 5).

In der Folge beschäftigte sich eine Vielzahl wissenschaftlicher Publikationen bis in die frühen 1980er-Jahre mit der rankingförmigen Vermessung von „eminence“, „quality“ und schließlich „excellence“, ohne jedoch einzelne Ranglisten zu veröffentlichen (Webster 1986). Insofern die verglichenen Eigenschaften der vermessenen Hochschulen als dynamisch erlebt wurden, meldeten die Publikationen mitunter zwar durchaus Bedarf für Wiederholungen an; in der Regel blieb es aber bei Einzelexperimenten. Was fehlte zur Durchsetzung serieller Ranglisten?

Ein Faktor scheint von besonderer Bedeutung zu sein. Während die Idee, Leistung zu vermessen, in der modernen Wissenschaft hohe Plausibilität genießt (schließlich ist vermessendes, quantifiziertes Vergleichen in vielen Disziplinen der zentrale Modus des Forschens), gerät die öffentliche Darstellung von Leistungshierarchien in Widerspruch zum wissenschaftlichen Ethos der selbstlosen Suche nach Wahrheit und der prinzipiellen Gleichheit aller Mitglieder der „scientific community“. Rankings müssen unter diesen Voraussetzungen mit Kritik rechnen und sind nur unter Auseinandersetzung mit oder Umgehung derselben durchzusetzen (Ringel und Weron 2020). Um diese Spannung zu veranschaulichen, diskutieren wir je ein Beispiel eines Rankings aus den Phasen 1898–1949 und 1950–1983, die explizit darauf abzielten, wiederholt produziert und veröffentlicht zu werden, sich jedoch letztlich nicht durchsetzen konnten.

Raymond Hughes beschäftigte sich intensiv mit der Vermessung der Ausbildungsleistungen von Hochschulen. Bereits 1925 begründete er die Notwendigkeit von Hochschulrankings damit, dass „(d)uring the past fifty years there has been an

Abb. 6 Economics. (1925, S. 14)

ECONOMICS									
Ratings submitted by: John H. Ashworth, Lloyd V. Ballard, Gilbert H. Barnes, Clarence E. Bonnett, John E. Brindley, E. J. Brown, J. W. Crook, Ira B. Cross, Edmund E. Day, Herbert Feis, Frank A. Fetter, Eugene Gredier, Lewis H. Haney, Wilbur O. Hedrick, Floyd N. House, Walter E. Lagerquist, W. E. Leonard, L. C. Marshall, W. C. Mitchell, C. T. Murchison, Tipton A. Snavely, E. T. Towne, J. H. Underwood, M. S. Wildman.									
Combined Ratings: (24)									
	1	2	3	4-5		1	2	3	4-5
Harvard.....	20	4	0	0	Illinois.....	0	5	4	4
Columbia.....	11	9	2	1	Cornell.....	0	2	7	5
Chicago.....	9	7	3	2	Princeton.....	2	1	4	4
Wisconsin.....	8	7	4	2	California.....	0	3	4	5
Yale.....	3	3	9	3	Minnesota.....	0	2	4	6
Johns Hopkins.....	2	4	8	3	Northwestern.....	0	2	3	6
Michigan.....	0	6	4	5	Stanford.....	0	1	4	6
Pennsylvania.....	0	3	6	8	Ohio State.....	0	1	2	8
					Toronto.....	0	2	2	3

amazing development of the colleges and universities of this country. They have grown in numbers, in size of faculties, in enrollment of students, in scope of work“ (Hughes 1925, S. 3). In der zunehmend unübersichtlichen Hochschullandschaft sei es immer schwieriger, geeignetes Personal auszuwählen; man wisse nicht, welchen „Wert“ der Abschluss einer Hochschule besitze. Daher brauche es ein Bewertungsschema, das in Entscheidungsprozessen verwendet werden könne. Wie Cattell, so sah auch Hughes Bewertungen durch Experten – d.h. Professoren – als bestmögliche Form der Feststellung der Qualität von Hochschulen. Die Darstellung der Ergebnisse erfolgte getrennt nach Disziplinen (hier das Beispiel Wirtschaftswissenschaften) (Abb. 6). Weil das akademische Feld immer in Bewegung sei, bestehe die Notwendigkeit, eine solche Bewertung nicht nur einmalig durchzuführen, sondern zu wiederholen, und zwar etwa alle drei Jahre; nur dann seien sie „of any value“ (Hughes 1925). Dementsprechend rät Hughes zwei unterschiedlichen professionellen akademischen Vereinigungen bis 1928, ein „re-rating“ (Hughes 1925, S. 7) in Auftrag zu geben. Jedoch folgten sie seinem Aufruf nicht. Daher wiederholte er die Studie erst neun Jahre später – diesmal aber *ohne die einzelnen Bewertungen in ein Ranking zu überführen*. Hervorgehoben wurden im Report von 1934 nur „most distinguished“ Institute (s. hier auch wieder am Beispiel Wirtschaftswissenschaften) (Abb. 7).

Im Jahr 1946, also 12 Jahre später, veröffentlichte Hughes schließlich noch einen Report, der die Zahl der von Hochschulen zwischen 1937 und 1942 verliehenen Dokortitel verglich. Da er keine institutionelle Unterstützung bekommen habe, sei er gezwungen gewesen, sich auf publiziertes Material zu beschränken, so Hughes – Umfragen unter Experten stünden nicht zur Verfügung. Der Report basiere daher nicht, wie seine Vorgänger, auf „the discriminating judgement of eminent men“ (Hughes 1946, S. iii). Die Ergebnisse der Studie solle man dementsprechend mit Vorsicht genießen und nicht als Wiederholung seiner früheren Arbeiten begreifen. Erneut betonte Hughes, man müsse in regelmäßigen Abständen (diesmal spricht

Abb. 7 Economics. (1934, S. 204)

Brown University	* University of Chicago
* Columbia University	University of Illinois
* Cornell University	University of Iowa
* Harvard University—	* University of Michigan
Radcliffe College	* University of Minnesota
Johns Hopkins University	* University of Missouri
New York University	* University of Pennsylvania
Northwestern University	University of Texas
Ohio State University	University of Virginia
* Princeton University	* University of Wisconsin
Stanford University	* Yale University
* University of California	

er von fünf Jahren) eine Evaluation amerikanischer Hochschulen publizieren, und wieder suchte er – vergeblich – nach einer Organisation, die ihn dabei unterstützt.

Als zweites instruktives Beispiel dient der in den 1960er-Jahren veröffentlichte Cartter-Report (Cartter 1966). Dieser kann insofern als Zäsur gelten, als es sich hierbei um eine breit rezipierte,⁴ auf umfangreichen, vom Autor und seinem Team selbst erhobenen Surveydaten basierende Rangliste von Graduiertenprogrammen handelte. Der Report zitiert explizit die genannten Studien von Hughes und auch eine Studie von Keniston (1959), stellt sich also selbst in eine zeitliche Reihe. Finanziell unterstützt und herausgegeben wurde er vom „American Council on Education“. Im Vorwort, verfasst von Logan Wilson, dem damaligen Präsidenten des „Council“, findet sich folgende Problemdefinition:

Excellence, by definition, is a state only the few rather than the many can attain. Striving for academic excellence, however, is a worthy ideal for colleges and universities, and it can be reasonably argued that every educational institution should meet minimum qualitative standards, and particularly if it offers graduate work. A present problem is the need for a better general understanding of what quality signifies (Cartter 1966, S. vii).

Die Orientierung an „excellence“ and „quality“ wird gekoppelt an einen temporalisierenden Vergleich, mit dem Cartter ein Argument von Hughes wieder aufgreift: Im Verlauf der vergangenen Jahrzehnte habe eine drastische Expansion des Hochschulsektors stattgefunden; daher sei es essenziell herauszufinden, welche Hochschulen gut und welche schlecht ausbilden. Die Nation, so heißt es weiter, müsse „the adequacy of the future supply of college teachers“ (Cartter 1966, S. 1) sicherstellen, insbesondere, da unter amerikanischen Graduiertenprogrammen in qualitativer Hinsicht eine hohe „diversity“ (Cartter 1966, S. 3) zu vermuten sei, die auch quantitativ abgebildet werden müsse. Denn: „evaluation by rumor and word of mouth is far from satisfactory, particularly in advanced training for scholarship and the professions“ (Cartter 1966). Hughes’ und Kenistons Studien seien diesbezüglich zwar wichtige Meilensteine, jedoch müssten sie „up to date“ (Cartter 1966) gebracht werden. Ähnlich wie Hughes forderte Cartter eine Wiederholung der Studie nach fünf Jahren, um zu gewährleisten, dass Qualitätsmessungen auch tatsächlich dem aktuellen Stand entsprechen: „Because the results of a study such as this one tend to be ‘writ in stone,’ it has been planned from the beginning to do a follow-up within the next five years. ... To avoid ‘freezing’ the reputations of various universities, the American Council on Education intends to repeat the study not later than 1970“ (Cartter 1966, S. 9). Auch Kritik an der Studie vonseiten jener, die an ihr teilnahmen, wird angeführt und ausführlich diskutiert; Kritiker bezögen sich insbesondere auf die vom Ranking unterstellte Hierarchie – die „pecking order“ (Cartter 1966, S. 8) zwischen Hochschulen.

Wie gefordert, erschien tatsächlich innerhalb von fünf Jahren eine vom „Council“ in Auftrag gegebene und finanzierte Folgestudie, durchgeführt von Roose und Andersen (1970). Die Autoren beziehen sich bereits in der Einleitung auf den Cartter-

⁴ Insgesamt seien, so Wilson im Vorwort zur Replikationsstudie von Roose und Andersen (1970), mehr als 26.000 Exemplare des Reports im Umlauf gewesen (Cartter 1966, S. ix).

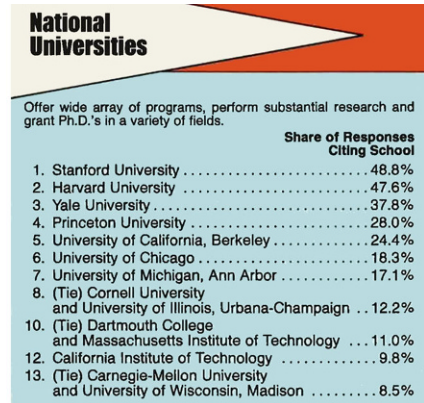
Report und behaupten einen sachlich-zeitlichen Zusammenhang zwischen den beiden Studien, die sie damit als Vergleichsereignisse miteinander in Beziehung setzen. Man verfolge das Ziel „to fulfill the commitment made when the earlier survey report was published“ (Roose und Andersen 1970, S. xi); und: „In updating the earlier study, we have been mindful of the various purposes and publics served by ratings“ (Roose und Andersen 1970). Auch hier wird also wieder auf die dem akademischen Feld inhärente Dynamik Bezug genommen, die es notwendig mache, Studien zu wiederholen. Jedoch halten die Autoren weder sich noch das „Council“ für zuständig: „If there is an ongoing purpose to be served by such ratings, the Council urges discipline areas or perhaps the Council of Graduate Schools in the United States to repeat the survey as necessary“ (Roose und Andersen 1970).

Wie schon bei Hughes fällt auf, dass die Replikationsstudie davon Abstand nimmt, die Ergebnisse erneut hierarchisch als Rangliste darzustellen. Dass dies kein Zufall ist, wird im folgenden Zitat deutlich, das einen direkten Verweis auf eine im Cartter-Report zitierte kritische Rückmeldung (die bereits genannte „pecking-order“) enthält: „In this new survey we have tried to de-emphasize the pecking-order relationships inherent in most scoring systems, for it is not our purpose to bolster or deflate egos“ (Roose und Andersen 1970). Neben den negativen Folgen für das akademische Feld bereitet den Autoren auch die Art und Weise Sorgen, in der außerwissenschaftliche Publika Ranglisten konsumieren. Insbesondere Studienanfänger tendierten dazu, Hierarchien für bare Münze zu nehmen; daher wolle man diesen nur mitteilen, ob ein gewisser Mindeststandard erreicht wurde oder nicht. Eine zweite Einschränkung betrifft die Benennung der Leistung, die der Report zu messen vorgibt. Während der Cartter Report noch explizit davon sprach, Aussagen über Qualität („quality“) zu machen, gibt sich die Replikationsstudie bescheidener und nimmt von diesem Begriff Abstand.

Parallel lassen sich ab den 1950er-Jahren erste zaghafte Versuche der massenmedialen Herstellung und Thematisierung von Hochschulrankings beobachten, so z. B. eine in der *Chicago Sunday Tribune* erschienene Rangliste mit dem Titel „Greatest Schools in Nation“ (Manly 1957). Rankings dieses Typs nahmen im Allgemeinen nicht oder bestenfalls peripher auf die Vielzahl von wissenschaftsintern produzierten Ranglisten Bezug und umgekehrt. Die Verknüpfung beider Diskurse glückte erstmals im Jahr 1983 dem Hersteller eines der wohl erfolgreichsten und wirkungsmächtigsten Rankings aller Zeiten, dem *US News and World Report*, einer Zeitschrift, die ihre Ranglisten entgegen aller Kritik und Skepsis durchgesetzt hat und zu einem etablierten Akteur im US-amerikanischen Hochschulfeld geworden ist (Sauder 2008). Es dauerte insgesamt also mehr als 80 Jahre, bis es zur Veröffentlichung von auf Hochschulen bezogenen Rankings i. e. S. – als *seriellen* öffentlichen Leistungsvergleichen – gekommen ist (Abb. 8).

Ab diesem Zeitpunkt änderte sich zweierlei: Erstens, Hochschulrankings werden ganz explizit nicht mehr ausschließlich als wissenschaftliche Kommunikation verstanden, sondern auch als Instrument zur Verknüpfung breiter Massenpublika mit der internen Beobachtung des Hochschul- und Wissenschaftsfeldes. Das war zwar ebenso von ihren Vorgängern intendiert, faktisch richteten diese sich jedoch vor allem an ein wissenschaftliches Expertenpublikum, was daran zu erkennen ist, dass sie als Reports oder in Fachzeitschriften erschienen und wissenschaftliche Kritik an

Abb. 8 National Universities.
(U.S. News Education 2014
[1983])



den negativen Folgen serieller Rankings respektierten. Die neueren, zumeist massen-medial aufbereiteten Varianten von Hochschulrankings orientieren sich zwar auch an wissenschaftlichen Standards (Lim 2018), legen es jedoch vor allem auf die Maximierung von Aufmerksamkeit an, z. B. durch aufwändige Visualisierungen, PR-Maßnahmen, Veranstaltungen und dergleichen (Brankovic et al. 2018; Stack 2018). Zweitens erzeugt ihr regelmäßiges Erscheinen, wie Espeland und Sauder (2007) für den amerikanischen Fall eindrucksvoll zeigen konnten, mitunter einen großen Konkurrenz- und Handlungsdruck, der sich darin äußert, dass die verglichenen Akteure die jährliche Veröffentlichung der Ergebnisse mit Spannung erwarten und sogar ihre organisationale Planung an die Logik der Datenerhebung von Rankingunternehmen anpassen. Wie schon im Falle des Kunstkompasses explizieren und reflektieren Hochschulrankings ihre eigene Temporalität: Die Ranglisten enthalten immer auch Informationen über die Ergebnisse der Vorjahre und die Internetseiten der Hersteller sind voll von Narrativen über auf- und absteigende Hochschulen, aber auch über Auf- und Abstiegstrends von Ländern oder Regionen. Die Temporalität der Rankings ist zudem prominent in ihre Visualisierung eingebettet und lässt Benutzer zumeist auch interaktiv darauf zugreifen. Im Unterschied zum künstlerischen Feld gibt es jedoch bislang keine Beschleunigung der Temporalisierung: Sowohl nationale als auch globale Hochschulrankings erscheinen in der Regel in jährlichen Abständen.

Für die erfolgreiche Durchsetzung serieller Hochschulrankings ab den 1980er-Jahren dürfte es mehrere Gründe geben, die hier nicht im Detail erörtert werden können und deren genauere Untersuchung noch aussteht. Ein wichtiger Faktor mag gewesen sein, dass es, wie der historische Abriss zeigt, im akademischen Feld schon seit dem frühen 20. Jahrhundert Rufe nach der regelmäßigen Vermessung der Qualität von Hochschulen gibt. Insofern reagieren heutige Rankings auf Bedürfnisse, die seit jeher im Feld artikuliert werden. Ein weiterer Grund scheint der neuartige Umgang mit Kritik zu sein. Wie schon der Cartter-Report, so sehen sich auch von Medienunternehmen hergestellte Rankings mit Kritik konfrontiert; im Unterschied zur Replikationsstudie von Roose und Anderson reagieren sie jedoch nicht mit der Einstellung ihrer Ranglisten, sondern geloben, ihr „Produkt“ zu reformieren und es im nächsten Jahr besser zu machen (Lim 2018). Insofern diese Kritik oftmals als Me-

thodenkritik geäußert wird, schließt sie selbst an die lange Tradition der Betrachtung von Rankings als einer wissenschaftlichen Methode an. Dies wird mitunter sogar explizit geäußert, beispielsweise vom einflussreichen britischen Hochschulforscher Simon Marginson in einem Aufsatz, der den Titel „University Rankings and Social Science“ (Marginson 2014) trägt. Überhaupt fällt auf, dass (Hochschul-)Forscher und Rankingproduzenten auf Diskussions- und anderen Veranstaltungen regelmäßig interagieren und (methodologische) Fragen erörtern; mitunter publizieren sie auch, wie die prominente Hochschulforscherin Ellen Hazelkorn, in Sammelwerken, die Beiträge von Wissenschaftlern *und* Rankingproduzenten enthalten (s. z. B. Marope et al. 2013). Durch die Aufnahme von Kritik in die jährliche Produktion der Hochschulrankings entsteht eine eigentümliche Dynamik der *thematisch-temporalisierten* Institutionalisierung (Ringel und Werron 2020), die nicht primär auf Konsens angewiesen ist, sondern auf die Fähigkeit, Kontroversen aufzunehmen und in der ständigen Reform der Vergleichskriterien zu reflektieren.

4 Zusammenfassung und Ausblick: Serielle Vergleiche in der Vergleichs- und Bewertungsforschung

Unsere beiden Fallstudien zur Geschichte von Rankings sollten exemplarisch zeigen, wie sich serielle Vergleiche historisch entwickelt und auf unterschiedliche Art und Weise etabliert haben. Zum Abschluss möchten wir die zentralen Gemeinsamkeiten und Unterschiede beider Fälle unter Bezug auf die eingeführten Vergleichstypen noch etwas genauer herausarbeiten und in zwei Thesen zusammenfassen. Auf dieser Grundlage begründen wir, weshalb sich die Vergleichs- und Bewertungsforschung auch über den Fall der Rankings hinaus für serielle Vergleiche und die Temporalität von Vergleichspraktiken interessieren sollte.

4.1 Thesen

These 1 Serielles und temporales bzw. temporalisierendes Vergleichen

Die Fallstudien lassen darauf schließen, dass beide Zeitlichkeitsdimensionen des Vergleichens – das temporale oder temporalisierende Vergleichen einerseits, die Temporalität und Serialität von Vergleichspraktiken andererseits – nur verstanden werden können, wenn man ihre systematischen und historischen Beziehungen untersucht. In beiden Feldern gingen dem seriellen Vergleichen *temporalisierende Narrative* voraus, die auf die Beobachtung von dynamischen Leistungs- und Wertentwicklungen abzielen: wechselnde Reputationszuschreibungen und Preise im Fall der Kunst, Qualität der Ausbildung und der wissenschaftlichen Exzellenz als flüchtige Größen im Fall der Hochschulen. Diese temporalisierende Vergleichsorientierung verlangte offenbar nach Vergleichs- und Bewertungspraktiken, die mit der Dynamik der zu beobachtenden Entwicklungen Schritt halten konnten. Die Idee, alle wichtigen Künstler und alle (amerikanischen) Hochschulen wiederholt tabellarisch miteinander zu vergleichen, ist folglich als Reaktion auf sich verfestigende temporalisierende Narrative zu verstehen. Auch die Entstehungsgeschichte dieser Narrative wurde am

Beispiel der Hochschulrankings angedeutet. Die Vorstellung, dass die Ausbildung für das Hervortreten von „eminent men“ wichtig sei, musste sich um 1900 erst gegen Vererbungslehren durchsetzen („nature“ vs. „nurture“), was eine entscheidende Voraussetzung für das Interesse an der Leistungsfähigkeit von Hochschulen als Stätten der Ausbildung und Wissensproduktion war.

Der Übergang vom temporalisierenden zum temporalisierten und schließlich seriellen Vergleichen lässt sich an beiden Fällen auch an den Vergleichskriterien nachvollziehen: Die quantifizierenden Kriterien des Kunstkompasses reflektierten Anfang der 1970er-Jahre das Interesse an temporalisierenden Preis- und Reputationsbeobachtungen. Und das Interesse an temporalisierenden Leistungs- und Reputationsbeobachtungen von Hochschulen mündete im Laufe des 20. Jahrhunderts mehrfach in den Vorschlag, Ranglisten unterschiedlicher Fachbereiche wiederholt und regelmäßig zu produzieren. Am Fall der amerikanischen Hochschulrankings fällt zugleich auf, dass die frühen Anläufe auf Widerstand stießen und die „Serialisierung“ von Ranglisten sich erst in den 1980er-Jahren durchsetzen konnte. Wie wir an Beispielen aus beiden Feldern gezeigt haben, spielte sich dann jedoch eine robuste Steigerungsbeziehung zwischen dem temporalisierenden und dem seriellen Vergleichen ein, die nicht zuletzt auch mit Rückwirkungen des seriellen auf das temporalisierende Vergleichen verbunden war. Die Rankings vergleichen und reflektieren Preis-, Reputations- und Leistungsentwicklungen von Kunstschaffenden und Hochschulen, und die Rankingorganisationen integrieren die Beobachtung solcher Entwicklungen in die visuelle Präsentation der Ranglisten wie auch in die öffentliche Kommentierung der Ergebnisse. Kurz, temporalisierender Vergleich und serielle Veröffentlichung stützen sich gegenseitig ab, plausibilisieren und stabilisieren einander.

Darüber hinaus drängt sich der Eindruck auf, dass der Durchbruch der seriellen Veröffentlichung von Rankings einer Orientierung an den Erwartungen eines breiteren (Laien-)Publikums („der Öffentlichkeit“) zu verdanken war, wodurch es gelang, Vorbehalte von Experten zu relativieren oder „standhaft“ zu bleiben, sodass es seit den 1980er-Jahren zur Institutionalisierung regelmäßig veröffentlichter Ranglisten kommen konnte. Die Kritik hörte darum nicht auf. Aber statt sich „einschüchtern“ zu lassen, pflegen die neueren Rankings einen offensiven Umgang mit ihr und nutzen die selbstkritische Auseinandersetzung und die ständige Reform der Vergleichskriterien zur eigenen Legitimierung und Stabilisierung.

These 2 Vergleichen und Bewerten

An Rankings, so unsere zweite These, lässt sich exemplarisch zeigen, dass Untersuchungen der Serialität von Praktiken des Vergleichens dazu beitragen können, die *Beziehungen zwischen Praktiken des Vergleichens und Bewertens begrifflich und empirisch zu spezifizieren*. Die Geschichte der Rankings bietet zunächst idealtypische Beispiele für die Entstehung und Verbreitung *steigernder* Vergleiche (Steinmetz 2015), die im Unterschied zu bloßen Ähnlichkeits- und Differenzbeobachtungen auf die Feststellung hierarchisch bewerteter Differenzen abzielen, hier vor allem entlang der Unterscheidung besser oder schlechter. Vergleich und Bewertung sind folglich in der Geschichte der Rankings von Anfang eng miteinander verbunden. Die historische Rekonstruktion dieser (und weiterer; vgl. Ringel und Werron 2020) Rankings zeigt

zudem, dass sich die Bewertung von Kunstschaaffenden, Hochschulen und anderen Akteuren häufig auf *Leistungsvergleiche* abstützt und somit von der Zuschreibung besserer und schlechterer Leistungen abhängig macht. Das mag heute selbstverständlich erscheinen, ist aber tatsächlich, wie die beiden untersuchten Fälle zeigen, historisch kontingent und kann ein Produkt jahrhundertelanger Entwicklungen sein.

Im Kunstfeld ereignete sich die Einführung des Kunstkompasses in den 1970er-Jahren vor dem Hintergrund von Experimenten mit ästhetischen Rankings und rankingähnlichen Vergleichsformen vom frühen 18. bis zum frühen 19. Jahrhundert, die ästhetische Werturteile quantitativ auszudrücken und tabellarisch darzustellen versuchten. Diese Vergleiche waren noch nicht seriell (es fehlten Wiederholungen). Sie gerieten zudem mit dem Aufkommen der professionellen Kunstkritik und des romantischen Geniekonzepts im 19. Jahrhundert in Verruf und waren zur Entstehungszeit des Kunstkompasses weitgehend vergessen. Gleichwohl spiegeln die Vergleichskriterien des Kunstkompasses und anderer Rankings auch die im Kunstfeld institutionalisierte Kritik an quantitativen ästhetischen Werturteilen implizit wider, insofern sie nicht versuchen, die ästhetische Qualität von Kunstwerken *unmittelbar* zu vergleichen. Vielmehr geht es (nur noch) um Preise und Reputationszuschreibungen, die quantitativ erhoben, in Mittelwerten verrechnet und tabellarisch dargestellt werden. Diese Umstellung trägt nicht nur den im Feld institutionalisierten Vorbehalten gegenüber quantitativen ästhetischen Werturteilen Rechnung, sie spiegelt auch Interessen an temporaler bzw. temporalisierender Beobachtung von Preis- und Reputationsentwicklungen und verbindet dies mit der Ansprache allgemeiner Laienpublika. Im Moment ihrer „Serialisierung“ wurden die Vergleiche populärer und zugleich komplexer: An die Stelle der einfachen, ästhetisch-quantitativen und einmaligen Ranglisten des 18. bis 19. Jahrhunderts, produziert von Experten für Experten, treten seit den 1970er-Jahren Ranglisten, die Reputations- und Preisinformationen in Gesamtbewertungen verrechnen und für ein breites (Laien-)Publikum wiederholt produziert und regelmäßig veröffentlicht werden.

Auch im Fall von Hochschulen lässt sich im 20. Jahrhundert die allmähliche Durchsetzung komplexer, quantitativ vergleichender serieller Ranglisten beobachten, die auf Leistungs- und Reputationsinformationen basieren. Frühe Experimente waren zumeist als Beitrag von individuellen Wissenschaftlern zu primär wissenschaftlichen Debatten ausgeflaggt. Außerdem finden sich Hinweise, dass zwar die Idee, Leistung zu vermessen, bereits in der ersten Hälfte des 20. Jahrhunderts Plausibilität genoss, die öffentliche Darstellung von Leistungshierarchien aber lange Zeit auf Widerstand stieß – in zeitgenössischen Diskursen reflektiert in der Ablehnung von durch solche Tabellen künstlich erzeugten „pecking orders“. Im Mittelpunkt der Kritik stand also weniger die quantitative Bewertung als solche, sondern die öffentliche Inszenierung von Leistungs- und Reputationsunterschieden. Während die neueren Kunstrankings die Kritik am quantifizierten ästhetischen Vergleich unterlaufen, indem sie ästhetische Kriterien durch Preis- und Reputationskriterien ersetzen, umgehen die neueren Hochschulrankings die Kritik an öffentlichen „pecking orders“, indem sie (a) methodologische Diskussionen mit (kritischen) Experten in die Produktionslogik inkorporieren, (b) mit ihren Veröffentlichungen ein allgemeineres Publikum adressieren und (c) der Tendenz nach immer mehr Indikatoren aufnehmen, um – so die Selbstbeschreibung – ein möglichst umfassendes Bild des Leistungs-

vermögens von Hochschulen zu zeichnen. Die Endprodukte sind sich jedoch letztlich auffallend ähnlich: Kunst- wie Hochschulrankings haben sich seit den 1970er-/1980er-Jahren zu seriellen, an breitere Publika gerichteten Vergleichspraktiken entwickelt und erfolgreich in beiden Feldern festgesetzt.

4.2 Forschungsperspektiven

Der Begriff des seriellen Vergleichens macht auf Wiederholungs- und Regelmäßigkeitsmuster von Vergleichs- und Bewertungspraktiken aufmerksam, die bisher kaum gesehen und untersucht worden sind. Wir haben solche Muster hier als Merkmal moderner Rankings identifiziert und historisch anhand der Fälle Kunst und Wissenschaft untersucht. Rankings sind jedoch nur eines von vielen Beispielen für serielle Vergleiche. Zum Abschluss sei daher angedeutet, welche grundlegenden Forschungsperspektiven wir mit dem Begriff serieller Vergleiche verbinden.

Wiederholungsmuster identifizieren und analysieren. Zunächst gilt es, Wiederholungs- und Regelmäßigkeitsmuster an möglichst zahlreichen und vielfältigen Vergleichs- und Bewertungspraktiken zu untersuchen. Dabei könnte es beispielsweise um „Verwandte“ von Rankings, wie sogenannte Ratings oder Blacklists, gehen, die mit ihnen einzelne Merkmale, wie das steigernd-bewertende Vergleichen, teilen, während ihnen andere Merkmale, wie die präzise quantitative Bewertung von Leistungen oder die tabellarische Visualisierung der Leistungshierarchie, fehlen. Weitere interessante Beispiele wären beschreibende Bevölkerungsstatistiken, denen das bewertende Element fehlt, oder jährlich vergebene Schulnoten, denen das tabellarische und das öffentliche Element fehlen, die aber wiederholt und regelmäßig produziert werden und daher auch als serielle Vergleiche begriffen werden können. Alle diese Vergleichspraktiken ließen sich darauf hin untersuchen, wie sie serielles und temporalisierendes Vergleichen aufeinander beziehen sowie Vergleichen mit Bewerten kombinieren. Über serielle Vergleiche im engen Sinne hinaus erscheint es uns zudem vielversprechend, die Temporalität von Vergleichs- und Bewertungspraktiken insgesamt genauer in den Blick zu nehmen, von der Rolle, die Wiederholungsmuster der vergleichenden Bewertung in sozialen Konflikten spielen, bis zum wiederholenden Gebrauch von Metaphern in Kunst und Literatur. Der Unterschied, den Wiederholung macht, könnte so an den unterschiedlichsten Vergleichspraktiken sichtbar gemacht, historisch untersucht und theoretisch reflektiert werden.

Historisch-soziologische Untersuchungen. Auf Basis eines grundlegenden Verständnisses der Serialität und Temporalität von Vergleichspraktiken ließen sich auch die Voraussetzungen und Effekte der Institutionalisierung solcher Praktiken genauer in den Blick nehmen. Wir haben dies hier exemplarisch an der Frage diskutiert, wie Kunst- und Hochschulrankings aufgegriffen und institutionalisiert wurden. Es ist deutlich geworden, dass bestimmte Typen von Vergleichs- und Bewertungspraktiken in Feldern auf unterschiedliche Rezeptionsbedingungen getroffen sind. Zugleich zeigten sich aber auch überraschende Ähnlichkeiten. An neueren Rankings fiel in beiden Fällen nicht zuletzt der trickreiche Umgang mit (Experten-)Kritik auf, der dazu führte, dass Rankings gewissermaßen an der kritischen Expertise vorbei oder durch die Auseinandersetzung mit dieser Kritik institutionalisiert wurden.

Dieser Eindruck schärft sich in der vergleichenden Analyse der Karrieren von Rankings in weiteren Feldern, etwa im Wettkampfsport. Während sich Rankings in Kunst und Wissenschaft durchsetzen konnten, aber hoch umstritten geblieben sind, stießen sie im Sport des späten 19. bis frühen 20. Jahrhundert rasch auf breite Akzeptanz, sind seitdem mehr oder weniger unangefochtener Bestandteil des Feldes und machen einen erheblichen Anteil des Unterhaltungswerts des modernen Wettkampfsports aus (Ringel und Werron 2020; Werron 2012b). Um diese unterschiedlichen Feldkarrieren serieller Vergleiche genauer zu verstehen, gilt es, Felder als soziale Kontexte zu untersuchen, die seriellen Vergleichen spezifische Legitimitäts- und Plausibilitätschancen bieten, umgekehrt aber auch durch serielle Vergleiche geformt und unter Veränderungsdruck gesetzt werden. Feldstrukturen ermöglichen serielle Vergleiche; serielle Vergleiche prägen Feldstrukturen.

Diese Beispiele zeigen, dass sich auch klassische Fragen der soziologischen Gesellschaftstheorie neu stellen, achtet man auf die Rolle serieller Vergleiche. Das betrifft erstens die Frage, was eigentlich ein modernes „Feld“ ausmacht und durch welche Eigenschaften sich Wissenschaft, Wirtschaft, Politik, Kunst, Sport, Erziehung u. a. voneinander unterscheiden. Sie lässt sich nun spezifizieren durch die Frage, *welche Rolle serielle Vergleichspraktiken in der historischen Ausdifferenzierung von Feldern gespielt haben*, etwa, indem sie zur Institutionalisierung moralisch-ethischer, meritokratischer, probabilistischer und anderer zukunftsbezogener Vergleichs- und Bewertungskriterien beitragen. Und es betrifft zweitens die Frage, *wie sich serielle Vergleichspraktiken mit Diskursformationen historisch verbunden haben*, die alle Felder gleichermaßen unter Veränderungsdruck setzen. Dazu zählen insbesondere Leistungs-, Konkurrenz- und Öffentlichkeitsdiskurse, die, wie wir hier an Rankings zeigen konnten, Spielräume für die Institutionalisierung serieller Vergleiche in allen Feldern geschaffen und diese zugleich auf je spezifische Weise limitiert haben. Das Interesse an seriellen Vergleichspraktiken bietet insofern auch einen Anlass, grundlegende gesellschaftstheoretische Fragen neu zu stellen und anders zu beantworten.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

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

Literatur

- artinfo24.com. 2015. Wie ArtRank sich zur meistgehassten Website in der Kunstszene entwickelte. <https://www.artinfo24.com/kunstmarkt/news-1415.html> (Zugegriffen: 06. Apr. 2020).
- Babcock, Kendric Charles. 1911. *A classification of universities and colleges with references to bachelor degrees*. Washington: Government Printing Office.
- Baumann, Margret. 2001. Die Entstehung des Kunstkompass. In *Kunst = Kapital. Der Capital Kunstkompass von 1970 bis heute*, Hrsg. Linde Rohr-Bongard, 9–16. Köln: Salon Verlag.
- Brankovic, Jelena, Leopold Ringel und Tobias Werron. 2018. How rankings produce competition. The case of global university rankings. *Zeitschrift für Soziologie* 47:270–288.
- Buckermann, Paul. 2020. *Die Vermessung der Kunstwelt. Quantifizierende Beobachtungen und plurale Ordnungen der Kunst*. Weilerswist: Velbrück.
- Cartter, Allan M. 1966. *An Assessment of Quality in Graduate Education*. Washington, DC: American Council on Education.
- Cattell, James McKeen. 1898. Doctorates Conferred by American Universities for Scientific Research. *Science* 8:197–201.
- Cattell, James McKeen. 1910. *A statistical study of American men of science*. New York, NY: Science Press.
- Claaßen, Marek. 2018. Art History is Exhibition history—the story behind ArtFacts.Net. <https://blog.artfacts.net/art-history-is-exhibition-history-the-story-behind-artfacts-net/> (Zugegriffen: 06. Apr. 2020).
- Claxton, Philander Priestly. 1912. *An Explanatory Statement in Regard to 'A Classification of Universities and Colleges with Reference to Bachelor's Degrees'*. Washington, D.C.: U.S. Bureau of Education.
- Davy, Ulrike, Johannes Grave, Marcus Hartner, Ralf Schneider und Willibald Steinmetz. 2019. *Grundbegriffe für eine Theorie des Vergleichens. Ein Zwischenbericht*. Working Paper 3. Praktiken des Vergleichens. Bielefeld: Working Paper des SFB 1288.
- Dolan, W. Patrick. 1976. *The Ranking Game*. Lincoln: University of Nebraska Printing and Duplicating Service.
- Epple, Angelika. 2015. Doing Comparisons. Ein praxeologischer Zugang zur Geschichte der Globalisierung/en. In *Die Welt beobachten. Praktiken des Vergleichens*, Hrsg. Angelika Epple und Walter Erhart, 161–199. Frankfurt a. M./New York: Campus.
- Espeland, Wendy N., und Michael Sauder. 2007. Rankings and Reactivity. How Public Measures Recreate Social Worlds. *American Journal of Sociology* 113:1–40.
- Espeland, Wendy N., und Mitchell L. Stevens. 1998. Commensuration as a social process. *Annual Review of Sociology* 24:313–343.
- Firschnig, Ulrich R. 2002. Schweizer analysieren ihren Kunstmarkt sorgfältiger als Deutsche. <https://www.faz.net/aktuell/feuilleton/kunst-markt-schweizer-analysieren-ihren-kunstmarkt-sorgfaeltiger-als-deutsche-169464.html> (Zugegriffen: 06. Apr. 2020).
- Galton, Francis. 1874. *English men of science. Their nature and nurture*. London: Macmillan.
- Grave, Johannes. 2015. Vergleichen als Praxis. Vorüberlegungen zu einer praxistheoretisch orientierten Untersuchung von Vergleichen. In *Die Welt beobachten. Praktiken des Vergleichens*, Hrsg. Angelika Epple und Walter Erhart, 135–159. Frankfurt a. M./New York: Campus.
- Grave, Johannes, Joris C. Heyder und Britta Hochkirchen. 2020. *BlickWechsel. Praktiken des Vergleichens von Bildern, Kunstwerken und Artefakten*, Bielefeld: Bielefeld University Press.
- Hammarfelt, Björn, Sarah De Rijcke und Paul Wouters. 2017. From Eminent Men to Excellent Universities. University rankings as Calculative Devices. *Minerva* 55:391–411.
- Heintz, Bettina. 2010. Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs. *Zeitschrift für Soziologie* 39:162–181.
- Heintz, Bettina. 2016. „Wir leben im Zeitalter der Vergleichung.“ Perspektiven einer Soziologie des Vergleichs. *Zeitschrift für Soziologie* 45:305–323.
- Heintz, Bettina. 2019. Vom Komparativ zum Superlativ. Eine kleine Soziologie der Rangliste. In *(Be)Werten. Beiträge zur sozialen Konstruktion von Wertigkeit*, Hrsg. Stefan Nicolae, Martin Endreß, Oliver Berli und Daniel Bischur, 45–79. Wiesbaden: VS Verlag.
- Heintz, Bettina, und Tobias Werron. 2011. Wie ist Globalisierung möglich? Zur Entstehung globaler Vergleichshorizonte am Beispiel von Wissenschaft und Sport. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 63:359–394.

- Helmore, Edward. 2014. Buy! Sell! Liquidate! How ArtRank is shaking up the art market. <https://www.theguardian.com/artanddesign/2014/jun/23/artrank-buy-sell-liquidate-art-market-website-artists-commodities> (Zugegriffen 06. Apr. 2020).
- Hughes, Raymond M. 1925. *A Study of the Graduate Schools of America*. Oxford, Ohio: Miami University Press.
- Hughes, Raymond M. 1934. Report of the committee on graduate instruction. *Educational Record* 15:192–234.
- Hughes, Raymond M. 1946. *A study of American graduate schools conferring the doctorate, 1937–38 to 1941–42*. Ames: Iowa State College Press.
- Keniston, Hayward. 1959. *Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania*. Philadelphia: University of Pennsylvania Press.
- Krüger, Anne K., und Martin Reinhart. 2016. Wert, Werte, (Be)Wertungen. Eine erste begriffs- und prozesstheoretische Sondierung der aktuellen Soziologie der Bewertung. *Berliner Journal für Soziologie* 26(3/4):485–500.
- Lim, Miguel Antonio. 2018. The Building of Weak Expertise. The Work of Global University Rankers. *Higher Education* 75:415–430.
- Manly, Chesly. 1957. 'Greatest Schools in Nation'—a New Survey by Tribune. *Chicago Sunday Tribune* 21. April:1–4.
- Marginson, Simon. 2014. University Rankings and Social Science. *European Journal of Education* 49:45–59.
- Marope, Mmantsetsa, Peter Wells und Ellen Hazelkorn. 2013. *Rankings and accountability in higher education. Uses and misuses*. Paris: UNESCO Publishing.
- Meier, Frank, Thorsten Peetz und Désirée Waibel. 2016. Bewertungskonstellationen. Theoretische Überlegungen zur Soziologie der Bewertung. *Berliner Journal für Soziologie* 26(3):307–328.
- Nicolae, Stefan, Martin Endreß, Oliver Berli und Daniel Bischur. 2019. *(Be)Werten. Beiträge zur sozialen Konstruktion von Wertigkeit*. Wiesbaden: Springer VS.
- Reckwitz, Andreas. 2015. Die Reproduktion und die Subversion sozialer Praktiken. Zugleich ein Kommentar zu Pierre Bourdieu und Judith Butler. In *Doing Culture. Neue Positionen um Verhältnis von Kultur und sozialer Praxis*, Hrsg. Karl H. Hörning und Julia Reuter, 40–54. Bielefeld: transcript Verlag.
- Ringel, Leopold, und Tobias Werron. 2019. Soziologie der Rankings. Neue Perspektiven. In *Rankings – Soziologische Fallstudien*, Hrsg. Leopold Ringel und Tobias Werron, S. V–XXVI. Wiesbaden: Springer VS.
- Ringel, Leopold, und Tobias Werron. 2020. Where Do Rankings Come From? A historical-sociological perspective on the history of modern rankings. In *Practices of Comparing: Towards a New Understanding of a Fundamental Human Practice*, Hrsg. Angelika Epple, Walter Erhart und Johannes Grave, 137–170. Bielefeld: Bielefeld University Press.
- Ringel, Leopold, Jelena Brankovic und Tobias Werron. 2020. The organizational engine of rankings. Connecting “new” and “old” institutionalism. *Politics and Governance* 8(2):36–47.
- Rohr-Bongard, Linde. 2001. *Kunst = Kapital: Die 100 Größten. 30 Ausgaben Capital-Kunstkompass*. Köln: Salon Verlag.
- Rohr-Bongard, Linde. 2019. Top 20: Olymp der Kunst. <https://www.capital.de/leben/top-20-olymp-der-kunst> (Zugegriffen: 06. Apr. 2020).
- Roose, Kenneth D., und Charles J. Andersen. 1970. *A Rating of Graduate Programs*. Washington, DC: American Council on Education.
- Sauder, Michael. 2008. Interlopers and Field Change. The Entry of U.S. News into the Field of Legal Education. *Administrative Science Quarterly* 53(2):209–234.
- Schäfer, Hilmar. 2013. *Die Instabilität der Praxis. Reproduktion und Transformation des Sozialen in der Praxistheorie*. Weilerswist: Velbrück.
- Schmidt-Burkhardt, Astrit. 2005. *Stammbäume der Kunst. Zur Genealogie der Avantgarde*. Berlin: Akademie Verlag.
- Spoerhase, Carlos. 2014. Das Maß der Potsdamer Garde. Die ästhetische Vorgeschichte des Rankings in der Europäischen Literatur- und Kunstkritik des 18. Jahrhunderts. *Jahrbuch der Deutschen Schillergesellschaft* 58:90–126.
- Stack, Michelle. 2018. *Global university rankings and the mediatization of higher education*. Houndmills, Basingstoke: Palgrave MacMillan.
- Steinmetz, Willibald. 2015. „Vergleich“ – eine begriffsgeschichtliche Skizze. In *Die Welt beobachten. Praktiken des Vergleichens*, Hrsg. Angelika Epple und Walter Erhart, 85–134. Frankfurt/New York: Campus.

- Steinmetz, Willibald. 2019. *The Force of Comparison. A New Perspective on Modern European History and the Contemporary World. New German Historical Perspectives. Vol 11.* New York/Oxford: Berghahn Books.
- U. S. News Education. 2014 [1983]. National Universities. <https://twitter.com/USNewsEducation/status/501791231647952896> (Zugriffen: 07. Apr. 2020).
- Velthuis, Olav. 2014. ArtRank and the Flippers. *Apocalypse now? Texte zur Kunst* 24(96):34–49.
- Webster, David S. 1986. *Academic Quality Rankings of American Colleges and Universities.* Springfield: University of Michigan.
- Werron, Tobias. 2010. *Der Weltsport und sein Publikum. Autonomie und Entstehung des modernen Sports.* Weilerswist: Velbrück.
- Werron, Tobias. 2012a. Worum konkurrieren Nationalstaaten? Zu Begriff und Geschichte der Konkurrenz um ‚weiche‘ globale Güter. *Zeitschrift für Soziologie* 41(5):338–355.
- Werron, Tobias. 2012b. Kontingenz in Serie. Zur „Spannung“ des modernen Sports. In *Die Möglichkeit des Sports. Kontingenz im Brennpunkt sportwissenschaftlicher Analysen*, Hrsg. Swen Körner und Peter Frei, 25–48. Bielefeld: transcript.
- Werron, Tobias. 2020. Global Publics as Catalysts of Global Competition. A Sociological View. In *Global Publics. Their Power and Their Limits, 1870–1990*, Hrsg. Valeska Huber und Jürgen Osterhammel, 343–366. Oxford: Oxford University Press.
- Werron, Tobias und Leopold Ringel. 2017. Rankings in a comparative perspective. Conceptual remarks. In *Geschlossene Gesellschaften. Verhandlungen des 38. Kongresses der Deutschen Gesellschaft für Soziologie*, Hrsg. Stephan Lessenich. Essen: DGS.
- Wilbers, Stefan. 2019. Grenzarbeit im Kunstbetrieb. Zur Institutionalisierung des Rankings Kunstkompass. In *Rankings – Soziologische Fallstudien*, Hrsg. Leopold Ringel und Tobias Werron, 57–87. Wiesbaden: Springer VS.

Leopold Ringel 1984, Dr., akademischer Rat a. Z., Universität Bielefeld. Forschungsgebiete: Organisationen, Felder, Quantifizierung, Rankings, Transparenz, Vergleichen. Veröffentlichungen: *Transparenz als Ideal und Organisationsproblem. Eine Studie am Beispiel der Piratenpartei Deutschland.* Wiesbaden, 2017. *Toward permeable boundaries of organizations?* (Hrsg.), Bingley, 2018 (mit P. Hiller und C. Zietsma).

Tobias Werron 1970, Dr., Professor für soziologische Theorie, Universität Bielefeld. Forschungsgebiete: Soziologie der Konkurrenz, Nationalismus, Globalisierungs- und Weltgesellschaftsforschung, Praktiken des Theoretisierens. Veröffentlichungen: *What in the World? Understanding Global Social Change.* Bristol 2020 (als Hrsg., mit M. Albert); *Der globale Nationalismus.* Berlin, 2018.