

Das Demokratiebarometer: "basically theory driven"?

Jäckle, Sebastian; Wagschal, Uwe; Bauschke, Rafael

Postprint / Postprint

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Jäckle, S., Wagschal, U., & Bauschke, R. (2012). Das Demokratiebarometer: "basically theory driven"? *Zeitschrift für Vergleichende Politikwissenschaft : German Journal of Comparative Politics*, 6(1), 99-125. <https://doi.org/10.1007/s12286-012-0133-6>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:
<https://creativecommons.org/licenses/by/4.0>

Das Demokratiebarometer: „basically theory driven“?

Sebastian Jäckle, University of Freiburg
Uwe Wagschal, University of Freiburg
Rafael Bauschke, University of Freiburg

Zusammenfassung: Das Demokratiebarometer ist einer der neusten Ansätze, um die Qualitätsunterschiede zwischen etablierten Demokratien vergleichend messbar zu machen. Neben positiven Aspekten, wie der Sensibilität was de jure und de facto Indikatoren für Demokratie angeht oder der gründlichen Dokumentation von Methodik und verwendeten Daten, zeigen wir in unserer kritischen Diskussion dieses Ansatzes, dass auf den Ebenen der Konzeptualisierung, der Operationalisierung und Messung sowie der Aggregation weiterhin Probleme vorliegen. So geht erstens das Spannungsverhältnis der drei grundlegenden Demokratieprinzipien Freiheit, Gleichheit und Kontrolle im Laufe der Indexbildung verloren, zweitens wird die schrittweise Ausdifferenzierung des Konzeptbaums nicht bis auf die untersten Ebenen sauber durchgeführt, wodurch es zu Fehlanpassungen zwischen Konzept und Operationalisierung kommt, und drittens führt die verwendete Skalierung und Aggregationsregel zu unerwünschten Nebenwirkungen. Auch können die von den Autoren des Demokratiebarometers selbst durchgeführten Validierungsversuche bei genauerer Betrachtung nur bedingt überzeugen.

Schlüsselwörter: Demokratie Messung Indexbildung

„basically theory driven“? – a critical analysis of the Democracy Barometer

Abstract: The Democracy Barometer is a newly established approach towards measuring differences in the democratic quality of widely recognized democracies. Apart from positive aspects like its sensitivity for de jure and de facto indicators of democracy and its thorough documentation, the critical examination reveals problems in terms of conceptualization, operationalization and measurement as well as aggregation of the data. First, the theoretically meaningful tension between the three basic democratic principles of freedom, equality and control vanishes in the course of the index construction. Second, the step-by-step development of the concept tree from its stem towards its leaves remains unclear to some extent, making it prone to possible mismatches between concept and operationalization. Third, the used scaling method and aggregation rule result in unintended side effects. Finally, on close inspection, the attempts of the Democracy Barometer' authors at validating their approach is not entirely convincing.

Keywords: democracy measurement index construction

1 Einleitung: Ein neuer Fokus der Demokratiemessung?

In den vergangenen zwei Jahrzehnten entwickelte sich die empirische Demokratieforschung von einem politikwissenschaftlichen Randthema zu einem der wichtigsten Forschungsbereiche der vergleichenden Politikwissenschaft (Abromeit 2004; Berg-Schlosser 1999: 9; Fuchs 2004; Lauth 2004, 2008; Merkel et al. 2003; Merkel et al. 2006; Müller und Pickel 2007). Gerade in den letzten Jahren wurden eine ganze Reihe neuer Demokratiemaße entwickelt.¹ Innerhalb dieses Forschungsfelds lässt sich aktuell eine Verschiebung der Schwerpunktsetzung beobachten. In der aktuellen Debatte rücken zunehmend Fragen der Messmethodik sowie der Binnendifferenzierung demokratischer (und autoritärer) Systeme in den Vordergrund. Eine zunehmende Zahl von Beiträgen und Forschungsprojekten untersucht dabei, inwiefern Qualitätsunterschiede innerhalb von etablierten Demokratien existieren und wie diese einer Messung zugänglich gemacht werden können. Die Ursprünge dieser Debatte lassen sich primär in den stärker theoretisch orientierten Arbeiten zur „Quality of Democracy“ (Beetham et al. 2009; Diamond und Morlino 2005) verorten. Parallel hierzu lässt sich eine deutliche Zunahme an Beiträgen beobachten, die die Auseinandersetzung über die adäquate „Vermessung“ von Demokratien sowohl aus theoretisch-konzeptioneller als auch aus empirisch-messmethodischer Sicht deutlich vorangetrieben haben (Abromeit 2004; Fuchs 2004; Fuchs und Roller 2008; Lauth 2004, 2011; Müller und Pickel 2007; Stoiber 2007) oder zumindest auf den fahrenden Zug aufgesprungen sind (Campbell und Barth 2009).

Als einer der Auslöser für dieses gestiegene Interesse kann hierbei die Tatsache gelten, dass bestehende Konzepte zwar eine größtenteils robuste Differenzierung von Systemtypen ermöglichen², jedoch nicht die Qualitätsunterschiede abbilden können, die sich bei einer genaueren, qualitativen Betrachtung zwangsläufig feststellen lassen (Lauth 2004, 2011; Lauth et al. 2000; Merkel et al. 2003). Vorhandene Demokratiemaße geben für die generell als entwickelte Demokratien bezeichneten Länder eine extrem geringe Varianz an (vgl. Tabelle 1).³ Trotz dieses Befundes sind Zweifel an der Ähnlichkeit des demokratischen Niveaus in den etablierten Demokratien durchaus angebracht. Oft zitierte Beispiele bilden Italien unter Berlusconi oder die USA unter der Bush Administration (Merkel et al. 2006). Während sich einige wenige Autoren aufgrund der geschilderten Problematik für eine generelle Abkehr von der quantitativen Messung aussprechen (Abromeit 2004), versuchen andere die Schwierigkeiten

¹ Im Online-Anhang zu diesem Beitrag findet sich in Tabelle A1 ein Überblick von 12 Demokratiemaßen in Bezug auf ihre Autoren, das jeweils zugrunde liegende Demokratieverständnis, ihre ländermäßige sowie zeitliche Abdeckung sowie die verwendete Operationalisierung.

² Doch auch im Hinblick auf die Leistungsfähigkeit bei der Unterscheidung zwischen Autokratien und Demokratien sehen sich die Messansätze etwa aufgrund der Diskussion um hybride Systeme (Diamond 2002) oder defekte Demokratien (Merkel et al. 2003) einer zunehmenden Kritik ausgesetzt.

³ Selbst ein Index wie die SGI (*Sustainable Governance Indicators*) der Bertelsmann Stiftung, der dezidiert für die OECD-Welt entwickelt wurde, dokumentiert immer noch eine vergleichsweise geringe Streuung in der Demokratiekomponente. Nicht zuletzt aufgrund der ähnlichen Zielsetzung kann und soll der SGI im Rahmen dieses Beitrags als Referenzpunkt herangezogen werden.

mehr oder minder pragmatisch und konstruktiv anzugehen (Foweraker und Krznaric 2000; Fuchs 2004; Stoiber 2007, 2011).

Der jüngste in dieser Tradition stehende Beitrag ist das von einer Forschergruppe um Marc Bühlmann und Wolfgang Merkel entwickelte Demokratiebarometer. Laut den Autoren ist es das erklärte Ziel ihres Messansatzes „die konzeptionellen und methodologischen Schwächen bisheriger Demokratiemaße zu überwinden, um so die Qualitätsunterschiede von etablierten Demokratien messen und analysieren zu können“ (Bühlmann et al. 2011a). Betrachtet man die Maßzahlen in Tabelle 1, so gelingt es dem Demokratiebarometer auf den ersten Blick durchaus eine vergleichsweise gute Unterscheidbarkeit etablierter Demokratien zu erzielen. Gemessen am Variationskoeffizienten CV liegt es in etwa im Bereich von Vanhanens Index (welcher sich diese Varianz ja bekanntermaßen durch eine nur begrenzte Validität „erkaufte“). Wir werden allerdings zeigen, dass die Varianz des Demokratiebarometers ähnlich problematisch zu bewerten ist.

Tabelle 1: Vergleich etablierter Demokratieindizes für die entwickelten Demokratien

	FH PR (1-7)	FH CL (1-7)	PolityIV 2010 (0-10)	Vanhanen 2000 (0-100)	SGI 2011 (1-10)	Democracy Ranking 2011 (0-100)	Democracy Index 2011 (1-10)	UDCS 2010 (-1,5-2,0)	DB 2007 (0-100)
Maxima	1	1	10 (z.B. NOR, GER)	42,75 (ITA)	9,43 (NOR)	88,2 (NOR)	9,80 (NOR)	2,0 (z.B. DEN, FIN)	87,2 (DEN)
Minima	1	2 (ITA, JAP)	8 (CZE BEL)	18,96 (SUI)	6,26 (ITA)	68,4 (HUN)	7,04 (HUN)	1,11 (CZE)	41,5 (FRA)
Arith. Mittel	1	1,08	9,79	32,21	8,09	78,95	8,51	1,58	67,95
SD	0	0,27	0,59	6,97	0,92	5,38	0,77	0,34	14,22
Spannweite	0	1	2	23,79	3,17	19,80	2,76	0,91	45,7
Var. Koef. CV	0	25,2	6,93	22,0	11,36	6,93	10,1	34,8	20,9

Anmerkungen und Quellen:

Es wurden nur die Daten verwendet, für die im Vergleich hinreichend Datenpunkte vorlagen (n = 26). *Demokratiebarometer* (DB): <http://www.democracybarometer.org/>; *Freedom House* (FH; PR = Political Rights, CL = Civil Liberties): <http://www.freedomhouse.org/report/freedom-world/freedom-world-2011>; Polity IV: <http://www.systemicpeace.org/polity/polity06.htm>; *Vanhanen*: <http://www.prio.no/CSCW/Datasets/Governance/Vanhanens-index-of-democracy/>; *SGI Sustainable Governance Indicators* (Demokratiewert): <http://www.sgi-network.org/index.php?page=category&category=SA>; *Democracy Ranking*: <http://www.democracysranking.org/en/>; *Economist Democracy-Index*: http://www.eiu.com/public/topical_report.aspx?campaignid=DemocracyIndex2011; *UCDS Unified Democracy Scores* (ein über einen Bayesianischen latenten Variablen Ansatz aus 10 bestehenden Demokratiemaßen synthetisierter kombinierter Demokratiewert (vgl. Pemstein et al. 2010)): <http://www.unified-democracy-scores.org/>.

Die Zielsetzung des Demokratiebarometers ist zu begrüßen. Gleichzeitig muss einem solchen Vorhaben im Hinblick auf die zahlreichen theoretischen und praktischen Schwierigkeiten, die bis dato Messversuche unterminierten, mit gesunder Skepsis begegnet werden. Der vorliegende Beitrag möchte vor diesem Hintergrund analysieren, welche Strategie das Demokratiebarometer zur Überwindung des Problems der mangelnden Varianz vorschlägt, und seine theoretischen Grundlagen sowie die eigentliche Messung einer kritischen Begutachtung unterziehen. Statt die Leistungsfähigkeit des Demokratiebarometers jedoch ausschließlich im Vergleich zu den Ergebnissen herkömmlicher Messansätze zu evaluieren, schlagen wir eine Überprüfung des Konzepts anhand der von Gerardo Munck und Jay Verkuilen entwickelten Kriterien vor (2002). Anhand dieser Kriterien lässt sich ein

Raster aus Konzeptualisierung, Operationalisierung und Messung sowie Aggregation entwickeln, das es ermöglicht, die Qualität des Demokratiebarometers im Hinblick auf die Indexbildung entsprechend systematisch einzuschätzen.

Bevor wir uns dem Aufbau und der Zielsetzung des Demokratiebarometers en détail zuwenden, soll zunächst knapp skizziert werden, welche Aspekte und Kriterien für die Evaluation von Messansätzen laut Munck und Verkuilen (2002) (im folgenden M&V) bedeutend sind. Im Anschluss an diese theoretischen Vorüberlegungen, wird das Demokratiebarometer auf den drei Ebenen Konzeption, Messung und Aggregation diskutiert und bewertet. Bevor im letzten Abschnitt die gewonnenen Erkenntnisse synthetisiert sowie potentielle Verbesserungen im Hinblick auf die zugrunde liegende Messung diskutiert werden, widmet sich der vierte Abschnitt noch den von den Autoren des Demokratiebarometers selbst vorgeschlagenen Validierungsmöglichkeiten.

2 Theoretische Vorüberlegungen: das Konzept von Munck & Verkuilen

Konzeptionell baut unsere Kritik am Demokratiebarometer auf dem von M&V (2002) erarbeiteten Vorschlag zur Bewertung von Messansätzen auf. Hierfür sprechen insbesondere zwei Gründe. Erstens hat sich diese Herangehensweise in der Debatte um die Güte von Messvorschlägen mittlerweile als zentrales Analyseraster etabliert (Jäckle und Bauschke 2009, 2010; Müller und Pickel 2007). Zweitens beziehen sich die Autoren des Demokratiebarometers selbst – wenn auch, wie zu zeigen sein wird, selektiv und damit in der Konsequenz unzureichend – auf die von M&V als relevant erachteten „Konstruktionsempfehlungen“ (vgl. Bühlmann et al. 2011a, S. 2). M&V folgend lassen sich drei wesentliche Ebenen der Messung unterscheiden: Die *Konzeptualisierung*, die *Operationalisierung* mit der eigentlichen *Messung* im Anschluss, sowie die *Aggregation*.

Auf der Ebene der *Konzeptualisierung* muss zunächst das zu messende Konstrukt präzisiert werden. M&V (vgl. S. 10-14) identifizieren in diesem Zusammenhang mehrere Anforderungen. Zunächst gilt es ein Konzept zu entwickeln, das weder zu maximalistisch noch zu minimalistisch ausgestaltet ist. In einem zweiten Schritt gilt es die verschiedenen Bestandteile des Konzepts ihrem Abstraktionsgrad entsprechend logisch anzuordnen. Hier führen M&V das intuitiv eingängige Bild des Konzeptbaums (*concept tree*) ein. Das abstrakte Konzept der Demokratie – versinnbildlicht als Stamm eines Baumes – lässt sich demnach in einzelne Äste aufteilen, die wiederum in Zweige ausdifferenziert werden können. Die Äste repräsentieren hierbei die das Gesamtkonzept konstituierenden Dimensionen bzw. Komponenten, während die Zweige als deren Subkomponenten aufgefasst werden können. Zum Teil können mehrere Ebenen einer solchen Verzweigung notwendig sein, um ein Konzept adäquat in all seinen Feinheiten erfassen zu können. An den Enden dieses Konzeptbaums finden sich schließlich die Blätter (*concept leaves*). Im Rahmen der Operationalisierung werden diesen einzelne messbare Indikatoren zugeordnet. Die eigentliche Messung findet entsprechend auf der Ebene der Blätter statt. Die Aufgabe des Indexentwicklers ist es, zum einen die richtige horizontale und vertikale

Differenzierung sicherzustellen, zum anderen die beiden Gefahren der Redundanz (*redundancy*) sowie der Aufblähung (*conflation*) zu verhindern. Unter *redundancy* verstehen M&V, dass die auf einer Abstraktionsebene befindlichen Komponenten das genau über ihnen angesiedelten Attribut komplett erfassen, sich dabei jedoch nicht überlappen sollten. Redundante Komponenten könnten ein Attribut „überstrapazieren“. Bei der *conflation* fließt hingegen eine Komponente in ein Attribut mit ein, die diesem nicht zugehörig ist, wodurch das Attribut künstlich aufgebläht wird. Eventuell fehlt exakt diese Komponente dann an anderer Stelle, was für ein solches Attribut bedeuten würde, dass es unzureichend ausgefüllt wird. Allgemein kann man diese Gefahren in dem Maße umgehen, „in dem die Attribute und Komponenten erstens jeweils eine eigenständige Bedeutung haben und zweitens zueinander trennscharf sind“ (Fuchs und Roller 2008, S. 89).

Auf der Ebene der *Operationalisierung* und *Messung* lassen sich neben der grundsätzlichen Anforderung einer theoretisch begründeten und nachvollziehbaren Indikatorenwahl drei zentrale Anforderungen bzw. Gütekriterien an einen Messvorschlag formulieren (M&V 2002, S. 15-22). Erstens sollten die Blätter des Konzeptbaums möglichst durch mehrere Indikatoren abgebildet werden. Hierdurch können einerseits systematische Verzerrungen, die durch eine zu starke Fokussierung auf einzelnen Indikatoren verursacht würden, ausgeschlossen und andererseits der durch zufällige Messfehler entstehende bias reduziert werden. Für transnationale Vergleiche fordern M&V, dass die gewählten Indikatoren unabhängig vom jeweiligen kulturellen, sozialen oder politischen Kontext sind, d.h. in unterschiedlichen Ländern dasselbe messen. In Anlehnung an Przeworsky und Teune (1970) sprechen sie von „cross-system equivalence“, die bei der Auswahl der Indikatoren zu gewährleisten sei (M&V 2002, S. 16). Zweitens sollte die Auswahl des Messniveaus theoretisch begründet sein, der prinzipiellen externen Überprüfung offenstehen und sicherstellen, dass weder zu grob noch zu differenziert gemessen wird. M&V schlagen eine Maximierung der Homogenität innerhalb der Messklassen mit einem Minimum an notwendigen Unterscheidungen vor. Drittens gilt es die Replizierbarkeit der Messung sicherzustellen.

Der Prozess der *Aggregation* führt die Messanlage wieder zurück zum Stamm des Konzeptbaums. Hierbei sieht man sich jedoch laut M&V mit einem generellen Trade-off konfrontiert (vgl. M&V 2002, S. 22-27). Mit jeder Verdichtung der Rohdaten bzw. der einzelnen Indikatoren geht ein Informationsverlust einher, den es theoretisch zu begründen gilt. So mag eine Aggregation in einen einzelnen Wert zwar im Sinne einer besseren Verständlichkeit und v.a. Vergleichbarkeit einleuchten, jedoch gleichzeitig die potentielle Mehrdimensionalität des zugrunde liegenden Konzepts ignorieren und die Differenzierungsmöglichkeiten der untersuchten Fälle reduzieren. Eine weitere Herausforderung stellt die Auswahl einer geeigneten Aggregationsregel dar, die wiederum theoriebasiert erfolgen sollte. In anderen Worten müssen sich die theoretisch postulierten Beziehungen zwischen den einzelnen Ebenen des Konzeptbaumes (Blätter – Zweige – Äste – Stamm) im

mathematischen Ausdruck der angewandten Aggregationsregel widerspiegeln.⁴ Eine weitere entscheidende Frage bei der Aggregation ist die Standardisierung der Variablen. Wenn einzelne Variablen nicht nur auf unterschiedlichen Skalen messen, sondern auch unterschiedliche Größenordnungen aufweisen, müssen sie zunächst vergleichbar gemacht werden, um eine sinnvolle Aggregation durchführen zu können. Schließlich ist die Frage der Gewichtung von Relevanz. Selbst wenn alle Variablen und Komponenten gleichgewichtig eingehen, ist dies eine inhaltliche Entscheidung zu Gunsten einer spezifischen Gewichtung, ohne dass dies jedoch explizit gemacht wird.

Die von M&V entwickelten Kriterien bieten uns nicht nur erste Anhaltspunkte für die Begutachtung der Messvorschläge sondern ebenso eine Strukturierungsmöglichkeit der nun folgenden Analyse. Vor der eigentlichen Evaluation soll jedoch das Demokratiebarometer knapp skizziert werden, um im Anschluss daran dessen zugrunde liegende Konzeption kritisch zu diskutieren und in den beiden darauf folgenden Abschnitten die Operationalisierung und Messung sowie die verwendete Aggregation zu evaluieren.

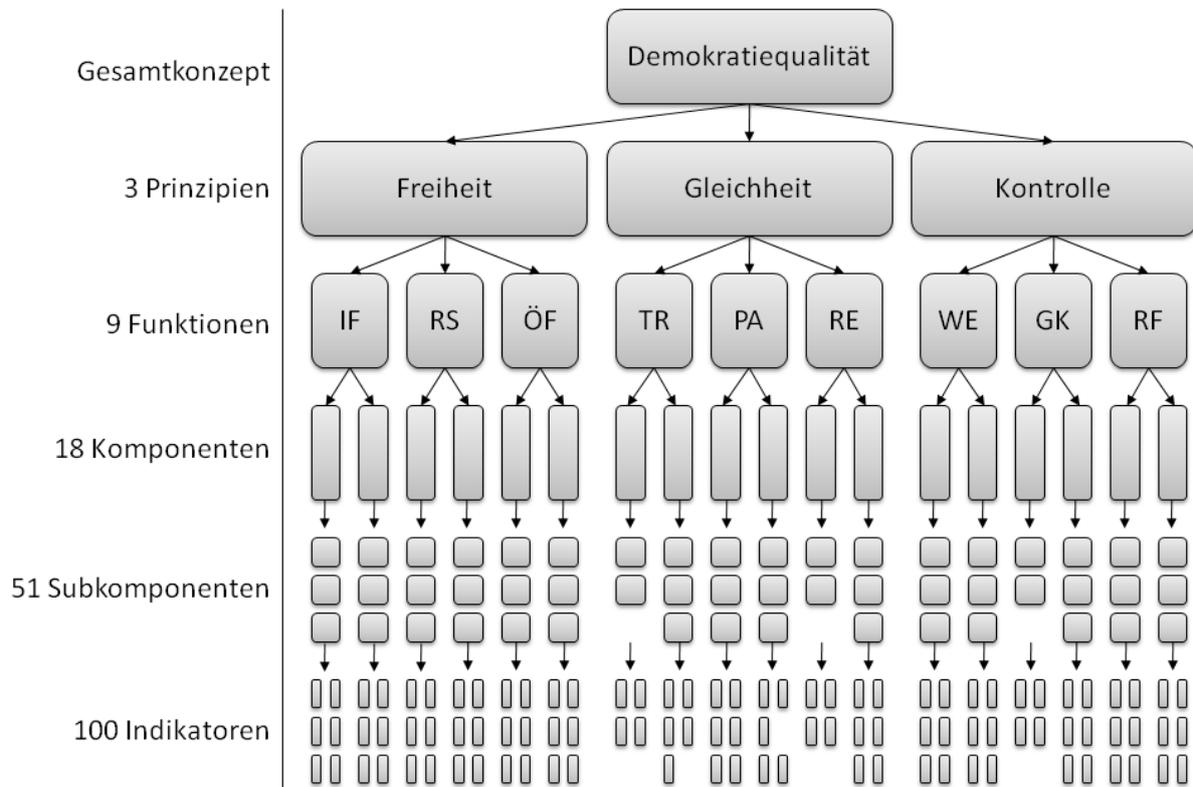
3 Das Demokratiebarometer – Zielsetzung und grundlegende Herangehensweise

Wie bereits weiter oben erläutert, ist das Hauptanliegen des Demokratiebarometers eine Binnendifferenzierung etablierter Demokratien zu ermöglichen. Ein weiteres und hiermit verknüpft Anliegen besteht laut den Autoren in der Identifizierung verschiedener (funktionaler) Realisierungsmöglichkeiten von Demokratie im Sinne einer „varieties of democracies“ (Bühlmann et al. 2011a, S. 9). Theoretisch basiert der Messvorschlag auf einer Demokratiekonzeption, die sich aus den Dimensionen – bzw. im Jargon des Demokratiebarometers – den Prinzipien Freiheit, Gleichheit und Kontrolle konstituiert.⁵ Diese drei Prinzipien lassen sich wiederum in jeweils 3 Funktionen aufgliedern (vgl. Abbildung 1).

⁴ Unseres Erachtens wird dieser Punkt zumeist nur wenig bei der Indexbildung reflektiert. So liegt der Entscheidung für eine additive Aggregation (z.B. Mittelwertbildung) oder multiplikative Aggregation (z.B. geometrisches Mittel) eine bedeutsame demokratietheoretische Entscheidung zugrunde. Bei einer Verwendung des arithmetischen Mittels ist die (oft unausgesprochene) Weichenstellung, dass sich die einzelnen Komponenten des Konzepts gegeneinander ausgleichen können, d.h. Defizite etwa bei der Freiheit können durch höhere Werte bei (sozialer) Gerechtigkeit kompensiert werden. Beim geometrischen Mittel bedingen sich dagegen die einzelnen Komponenten stärker, sie sind jeweils notwendige Bedingungen für eine Demokratie. Ist der Wert einer Teilkomponente (z.B. Freiheit) „null“ dann ist auch der Gesamtindex null. Je nach theoretischer Fundierung ist demzufolge entweder einer multiplikativen Verknüpfung über das geometrische Mittel oder einer additiven Verknüpfungslogik, wie sie dem arithmetischen Mittel zugrunde liegt, sachlogisch der Vorzug zu geben.

⁵ Das Demokratiebarometer bezieht sich damit implizit auf das von Hans-Joachim Lauth (2004) systematisierte Demokratieverständnis.

Abbildung 1: Der Konzeptbaum des Demokratiebarometers



Eigene Darstellung nach (Bühlmann et al. 2011c, S. 18-42). IF: Individuelle Freiheiten; RS: Rechtsstaatlichkeit; ÖF: Öffentlichkeit; TR: Transparenz; PA: Partizipation; RE: Repräsentation; WE: Wettbewerb; GK: Gewaltenkontrolle; RF: Regierungsfähigkeit;

Die Funktionen werden in einem weiteren Schritt in jeweils 2 Komponenten gegliedert, die über mehrere (zumeist drei) Subkomponenten ausdifferenziert werden, deren Operationalisierung zumeist über jeweils zwei Indikatoren erfolgt. Die Autoren folgen damit der von M&V eingeführten Metapher des Konzeptbaums.⁶ Die Messung erfolgt letzten Endes über insgesamt 100 Indikatoren, die nach einer Reskalierung in die Konstruktion des Demokratiebarometers einfließen. Bis dato liegen Daten für den Zeitraum 1995 bis 2007 für 30 etablierte Demokratien vor.

4 Kritische Diskussion des Demokratiebarometers

Nachdem im vorherigen Abschnitt die grundlegende Konstruktionslogik und die Zielsetzung des Demokratiebarometers erläutert wurden, soll das Konzept nun entlang der oben definierten Kategorien

⁶ Die Konzeptbäume für alle Funktionen bzw. Komponenten können in Bühlmann et al. nachgeschlagen werden (Bühlmann et al. 2011c)

Konzeptualisierung, Operationalisierung, Messung und Aggregation näher untersucht werden. Die Diskussion fokussiert dabei auf die aus unserer Sicht schwerwiegendsten Probleme.

4.1 Konzeptualisierung

Die Diskussion der konzeptionellen Grundlagen muss zwangsläufig an der verwendeten Demokratiedefinition ansetzen. Die Autoren schlagen hier eine „Demokratiekonzeption mittlerer Reichweite“ (Bühlmann et al. 2011a, S. 2) vor. Man könnte auch von einem dynamischen und durch Interaktionen geprägten Demokratiemodell sprechen, da die Autoren Demokratie auf der „Prämisse [aufbauen], dass demokratische Systeme eine Balance zwischen den interdependenten Werten ‘Freiheit’ und ‘Gleichheit’ herzustellen versuchen und [sie] sich dazu einer dritten demokratieinhärenten Dimension bedienen: Kontrolle“ (Bühlmann et al. 2011a, S. 4). Sie denken Demokratie somit nicht als starres Konzept.

4.1.1 Fundamentale Kritik an der grundlegenden Konzeption nach Victoria Kaina

Die drei Prinzipien stehen dabei laut Bühlmann et al. in einem Spannungsverhältnis. Einerseits werden sie als normativ gleichwertig erachtet und andererseits – und das ist das entscheidende Novum zu alternativen Messansätzen – stehen sie gleichzeitig in einem interdependenten Verhältnis zueinander. Dies bedeutet jedoch auch, dass die Maximierung einer Dimension zu Lasten einer oder der anderen beiden Dimensionen geht. Entsprechend kann man auch nicht den Fall als optimal für die Demokratiequalität bezeichnen, wenn alle drei Dimensionen vollständig bzw. maximal ausgeprägt vorliegen. Eine solche Konstellation ist in dem vorliegenden Demokratiekonzept an sich per definitionem ausgeschlossen. Die vom Demokratiebarometer vermutete, aber nicht explizit formulierte, grundlegende Vorstellung von Qualität ist insofern, wie es Victoria Kaina in ihrem Beitrag zum Demokratiebarometer pointiert herausgearbeitet hat, durchaus problematisch.⁷ Das Demokratiebarometer postuliert, dass sich „der Maßstab für die Demokratiequalität eines demokratischen politischen Systems an einem Optimum orientier[t], das sich für das Spannungsverhältnis jener drei Demokratieprinzipien sensibel zeigt“ (Kaina 2008, S. 521). Damit liegt aber eben gerade kein gemeinsamer Maßstab vor, der an alle drei Prinzipien immer in derselben Weise angelegt werden kann und die Bildung einer Rangfolge erlauben würde. Neben der praktischen Schwierigkeit ein allgemeingültiges Optimum zu bestimmen, verweist Kaina auf das Problem der mangelnden Kontextualität, da „sich das jeweilige Optimum in der Verwirklichung der drei Demokratieprinzipien in einem gegebenen demokratischen politischen System nicht kontextfrei bestimmen lässt“ (Kaina 2008, S. 522). In Ermangelung eines theoretischen Maßstabes läuft eine empirische Bestimmung des Optimums laut Kaina jedoch Gefahr arbiträr zu werden.⁸ Im Ergebnis

⁷ Die Autorin greift mit der Frage nach der Bedeutung von Qualität eine hochrelevante, jedoch weitestgehend ignorierte Frage im Rahmen der aktuellen Demokratieforschungsdebatte auf (Ausnahmen bilden Abromeit 2004; Diamond und Morlino 2005; Fuchs 2004; Fuchs und Roller 2008).

⁸ Ein drittes Problem sieht Kaina in der Verwendung eines „optimalen Erfüllungsgrades“ als Maßstab höchster Demokratiequalität. „Das Problem an diesem Argument ist, dass sich in der Sprachlogik der Worte „Optimum“ oder „optimal“ graduelle Abstufungen verbieten. Es geht hier also um Unterschiede der Art (und nicht des

bedeutet dies, dass das Demokratiebarometer dem eigenen Anspruch eine Rangfolge der Staaten entsprechend ihrer Demokratiequalität zu erstellen, nur teilweise gerecht werden kann. Stattdessen erscheint der Messvorschlag als ein Instrument zur Klassifikation verschiedener Typen der Demokratie oder wie Bühlmann et al. formulieren zur Identifikation der „Varieties of Democracies“.

Je nachdem für wie kritisch man die angesprochenen und zumindest von Kaina als durchaus fundamental erachteten theoretisch-konzeptionellen Defizite des Demokratiebarometers betrachtet, könnte man bereits an dieser Stelle jede weitere Evaluation abbrechen und den Daumen über dieses Messinstrument senken. Im Folgenden werden wir jedoch nicht den Weg einer solchen Fundamentalopposition einschlagen, sondern wollen vielmehr aufzeigen, dass abseits dieser konzeptionellen Ungereimtheiten auch noch weitere Kritikpunkte am Demokratiebarometer anzubringen sind, die zum Teil allerdings durchaus ihre Grundlegung in eben dieser problematischen Konzeption von demokratischer Qualität erfahren haben.

4.1.2 Zu maximalistisch? – die konzeptionelle Breite des Demokratiebarometers

Projiziert man das gewählte (funktionalistische) Demokratiemodell auf das von M&V erwähnte Kontinuum maximalistischer und minimalistischer Definitionen, lässt sich das Demokratiebarometer als eher maximalistisches Konzept auffassen, was in der Literatur bereits kritisiert wurde (Lauth 2010). Diese Entscheidung wird von den Autoren folgendermaßen begründet: Das Problem existierender Messansätze sei, „dass diese auf einer (zu) minimalistischen Demokratiekonzeption basieren.“ Und weiter: „Um das Phänomen ‘Demokratie’ jedoch in seiner ganzen Komplexität erfassen zu können, muss eine Demokratiequalitätsmessung auf einem breiten Demokratiekonzept fußen.“ (Bühlmann et al. 2011a, S. 2). Letzterer Aussage ist grundsätzlich zunächst zuzustimmen. Eine solche „Verbreiterung“ muss jedoch – wie auch M&V hervorheben – in einer Weise geschehen, die verhindert, dass Elemente aufgenommen werden, die für die Qualität einer Demokratie nicht konstitutiv sind (siehe hierzu auch: Fuchs 2004; Lauth 2004). Den Autoren des Demokratiebarometers kann jedoch bescheinigt werden, dass sie die von ihnen verwendete, vergleichsweise breite Demokratiedefinition über die drei Prinzipien Freiheit, Gleichheit und Kontrolle überzeugend begründen und damit zumindest auf der obersten Ebene des Konzeptbaumes kein überdehntes Demokratiekonzept aufstellen. Das bedeutet jedoch nicht, dass *conceptual stretching* nach Sartori für das gesamte Demokratiebarometer kein Problem darstellen würde. Vielmehr soll im Folgenden gezeigt werden, dass durch die weitere Ausdifferenzierung über die insgesamt fünf Ebenen, die unterhalb der Demokratiequalität anzusiedeln sind, das ursprüngliche

Grades), weil ein Optimum entweder erreicht ist oder nicht, etwas entweder optimal ist oder nicht.[...] Es können auf diesem Wege demzufolge nur dichotome Aussagen über die Demokratiegüte demokratischer politischer Systeme getroffen werden“ (Kaina 2008, S. 522). Rein semantisch betrachtet ist ihre Argumentation korrekt, allerdings erscheint uns ihre Kritik doch überzogen und letztlich am Punkt vorbei. Die Frage ist nicht, ob es eine optimalere Demokratiequalität als die optimale gibt, sondern wie man zu welchem Grad dem Optimum nahekommen kann.

Konzept Schritt für Schritt so ausgeweitet wird, dass letztlich ein deutlich breiteres Konzept auf der untersten Ebene gemessen wird als eigentlich intendiert.

4.1.3 Die Indikatorenwahl – Unzureichende Herleitung und schrittweise Ausweitung des Messkonzepts

Eine weitere Anforderung in der Phase der Konzeptualisierung ist die schrittweise Ausdifferenzierung des Demokratiemodells im Sinne des Konzeptbaums.⁹ Der hierfür notwendige Schritt der weiteren Ausspezifizierung der drei Prinzipien in jeweils drei Funktionen wird von den Autoren zwar relativ überzeugend geleistet, jedoch kritisiert bereits Lauth die nicht stimmige Zuordnung der Funktionen zu den Prinzipien, da sie „weder ausreichend begründet noch überzeugend sind“ (2010, S. 518). So kann die Funktion „Transparenz“ ohne weiteres auch unter dem Prinzip der „Kontrolle“ subsumiert werden und nicht nur unter dem der „Gleichheit“. Gleiches gilt, wenn man sich die Einsortierung des „Wettbewerbs“ unter dem Prinzip „Kontrolle“ ansieht, da dieser doch mindestens genauso plausibel unter „Gleichheit“ verortet werden könnte. Hier bereits zeigen sich mögliche Gefahren der *conflation*. Der Erfüllungsgrad der insgesamt neun Funktionen bildet dann den zentralen Maßstab der Demokratiequalität (vgl. Bühlmann et al. 2011a, S. 9). Noch deutlicher tritt der Mangel einer klaren, nachvollziehbaren und vor allem theoretisch begründeten Ausdifferenzierung allerdings unterhalb der Funktionenebene zutage. Die Verbindung der neun Funktionen mit den darunter liegenden und sie somit konstituierenden Komponenten bzw. Subkomponenten erfolgt wesentlich oberflächlicher als die ersten beiden Differenzierungsschritte vom „Demokratiestamm“ über die drei Prinzipien zu den neun Funktionen. Doch gerade auch auf den unteren Ebenen bedarf es einer differenzierteren Darstellung, um einerseits eine gewisse Nachvollziehbarkeit zu gewährleisten, und andererseits Anhaltspunkte für die Validität der Messung zu bieten. Schließlich ist es ein Unterschied, ob Indikatoren, d.h. die Blätter des Baumes beschrieben, oder zumindest in tabellarischer Form gelistet werden, wie dies beim Demokratiebarometer in der Regel geschieht, oder ob die jeweiligen Zusammenhänge zwischen Indikatoren, Subkomponenten und Komponenten bis zu dem eigentlich interessierenden Gesamtkonzept wirklich expliziert und somit die Verästelungen des Konzeptbaumes nachvollziehbar gemacht werden. Ohne hier der Diskussion der eigentlichen Messung vorgreifen zu wollen, lässt sich hinterfragen, ob der eigene Anspruch einer theoriegeleiteten, schrittweisen Entwicklung der Indikatoren ausgehend von einer grundlegenden Demokratiedefinition über mehrere Zwischenebenen beim Demokratiebarometer tatsächlich ausreichend eingelöst worden ist.

Kontrastiert man das Vorgehen des Demokratiebarometers mit den von M&V formulierten Regeln, lassen sich auf der Konzeptualisierungsebene folgende konkreten Defizite konstatieren: Zunächst lässt sich aufgrund der unzureichenden theoretischen Herleitung die Eignung mancher Indikatoren in Bezug auf deren Inhaltsvalidität hinterfragen, d.h. bezüglich der Passgenauigkeit der gewählten Indikatoren für

⁹ Bühlmann et al. unterschlagen allerdings diese explizit bei M&V erwähnte Anforderung in ihrer eigenen auf M&V aufbauenden Auflistung der „three critical tasks“ der Indexbildung (vgl. Bühlmann et al. 2011c, S. 3).

das zu messende Konzept. Es mag noch einleuchten, dass die öffentliche Unterstützung einer Regierung einen positiven Effekt auf die Regierungsfähigkeit hat,¹⁰ doch inwiefern sich diese Unterstützung über die persönliche Einstellung zum Schwarzfahren erheben lässt (vgl. Bühlmann et al. 2011b, S. 28), selbst wenn dies nur einen Teilindikator darstellt, ist zumindest fraglich und die Inhaltsvalidität muss an dieser Stelle angezweifelt werden. Durch die mangelnde Diskussion der Indikatoren in Bezug auf ihre vertikale Verknüpfung und ihre letztliche Eignung muss der Leser die vom Demokratiebarometer angenommenen Zusammenhänge zwischen Indikatoren und den übergeordneten Ebenen entweder unreflektiert hinnehmen oder versuchen sie sich selbst zu erschließen – die Ausführungen in den Publikationen zum Demokratiebarometer sind hierbei nur bedingt hilfreich.¹¹

Neben der mangelnden Stringenz in der Herleitung und der großen Anzahl an Indikatoren – was im Übrigen auch eine umfassende Diskussion und Evaluation im Hinblick auf die Problematik der *conflation* und *redundancy* erschwert – ist es der von den Autoren beschriebene Prozess der Indikatorenauswahl, der Zweifel an einer ausschließlich theoriegeleiteten Entwicklung aufkommen lässt. Zunächst wurden 300 potentielle Indikatoren kompiliert. In einem weiteren Schritt wurden über eine negative Auswahl dann 200 Indikatoren ausgeschlossen, so dass 100 für die weitere Analyse verblieben. Diese Reduzierung der Indikatorenzahl erfolgte anhand folgender Vorgaben: Erstens sollten Indikatoren, die auf Expertenbefragungen beruhen, aufgrund möglicher Reliabilitätsprobleme größtenteils ausgeschlossen werden. Zweitens wurde durch die Verwendung verschiedener Quellen und multipler Indikatoren versucht, systematische Verzerrungen möglichst auszuschließen sowie zufällige Fehler zu reduzieren. Drittens wurden jeweils Indikatoren ausgewählt, welche die de jure und die de facto Ebene abbilden können. Viertens wurden nur jene Indikatoren ausgewählt, die für einen längeren Zeitraum und für alle im Sample enthaltenen Fälle verfügbar waren. Sowohl der Auswahlprozess der ursprünglichen Indikatoren als auch ihre Reduktion erfolgte dabei laut den Autoren „basically theory driven“ (Bühlmann et al. 2011b, S. 4). Trotz dieses Hinweises drängt sich aufgrund der verfügbaren Informationen der Verdacht auf, dass es sich bei der Konzeption der Messung und der Identifikation von Indikatoren um teilweise voneinander losgelöste Prozesse handelt. Statt einer stringenten theoriegeleiteten Entwicklung über die verschiedenen Ebenen des Konzeptbaums, die in einer positiven Auswahl der letztlich gemessenen Indikatoren münden sollte, wird aus einer Vielzahl verfügbarer Indikatoren ein Indikatorenset zusammengestellt, dass dann mit den separat entwickelten Funktionen „gematcht“ wird. Der Konzeptbaum wächst damit einerseits vom Stamm der Demokratiedefinition

¹⁰ Sehr viel grundlegender kann man sich natürlich die Frage stellen, ob nicht auch nicht-demokratische Staaten, wie z.B. Singapur oder die VAE eine sehr gute Regierungsfähigkeit aufweisen können und inwiefern ein solcher Indikator, der sowohl in demokratischen als auch in autoritär geführten Systemen hoch ausgeprägt sein kann, als Maß für Demokratiequalität taugt.

¹¹ So ist beispielsweise auch der angenommene positive Zusammenhang zwischen der Dauer der Legislaturperiode und der Regierungsfähigkeit (vgl. Bühlmann et al. 2011b, S. 28) durchaus diskussionswürdig, nicht nur wenn man wiederum den Vergleich zu autokratischen Systemen anstrengt, in denen die „Legislaturperioden“ entweder unbegrenzt sind oder keinerlei Auswirkungen auf die Politik zeitigen.

ausgehend bis zu den neun Funktionen – dort bricht diese theoriegeleitete Vorgehensweise jedoch größtenteils ab – und andererseits wächst er ausgehend vom Universum potentieller Indikatoren (den Blättern) durch ein möglichst plausibles Zusammenfassen der Einzelindikatoren zu Subkomponenten und Komponenten gewissermaßen rückwärts in die andere Richtung. Die Gefahr dabei ist selbstverständlich, dass Äste und Zweige des Baums nicht zusammenpassen, dass es also zu einem gewissen „mismatch“ zwischen den Ebenen kommt.¹²

Grundsätzlich ist auch der Wunsch zu begrüßen, die einzelnen Komponenten und Subkomponenten jeweils durch mehrere Indikatoren abzubilden und so systematische wie zufällige Messfehler zu reduzieren. Gleichzeitig muss man sich aber aufgrund der sich hierdurch ergebenden sehr großen Anzahl von insgesamt 100 Indikatoren die Frage stellen, ob wirklich eine ausreichende Überprüfung auf *redundancy* und *conflation* stattgefunden hat. Eine intensivere Diskussion der einzelnen Komponenten, Subkomponenten und Indikatoren wäre genau aus diesem Grund wünschenswert, da an dieser Stelle einige Lücken des Demokratiebarometers sichtbar werden. Die Zusammensetzung der neun Funktionen aus Komponenten und Subkomponenten wird zwar noch beschrieben (Bühlmann et al 2011a, S. 9-16), über eine Nennung der einzelnen Subkomponenten reicht diese „Diskussion“ jedoch nur in den seltensten Fällen hinaus. Eine wirkliche Besprechung der untersten Ebene, also der verwendeten Indikatoren fehlt vollends.

Betrachtet man diese jedoch genauer, so tauchen durchaus potentielle Probleme auf: Allgemein scheinen die vielen Ebenen zu einer unbewussten Ausweitung des grundlegenden Demokratiekonzepts zu führen, so dass in die Messung vermehrt Aspekte eingehen, die mit dem eigentlichen Kern von Demokratie und ihrer Qualität wenn überhaupt nur noch peripher zusammenhängen. Auch wenn die Autoren mit dem Verweis auf ihr „grundsätzlich theoriegeleitetes Vorgehen“ behaupten das von Sartori problematisierte Überdehnen des Konzepts zu umgehen, scheint exakt dies der Fall zu sein. So erscheint die Verwendung der Mordrate und der Anzahl gewalttätiger Proteste zwar für die Gesamtkomponente „Recht auf körperliche Unversehrtheit“ durchaus notwendig, aber ob diese Komponente in ihrer Gänze wirklich einen notwendigen Teilaspekt der Demokratiequalität darstellt, kann wohl bezweifelt werden. An diesem Beispiel lassen sich zwei grundlegende Fragen verdeutlichen, die sich im Rahmen des Demokratiebarometers an unterschiedlichen Stellen immer wieder stellen: Ist es erstens möglich, dass dieselbe Ausprägung eines Indikators sowohl als Indiz für eine hohe Demokratiequalität als auch als Indiz für ein autoritäres Regime erachtet werden kann? Denn sowohl in autoritären Regimen, die Proteste jedweder Art von Grund auf nicht zulassen, als auch in Demokratien, in denen diese Proteste in der Regel friedlich ablaufen, würde der Indikator „Anzahl gewaltsamer Proteste“ dieselbe

¹² In eine ähnliche Richtung weist auch der mehrfach von den Autoren des Demokratiebarometers gegebene Hinweis (vgl. z.B. Bühlmann et al. 2011a, S. 20), dass sich die Messanlage nach Belieben durch Hinzufügen weiterer Indikatoren erweitern ließe. Ein in sich schlüssiges, kohärentes Bild gibt das Demokratiebarometer damit nicht ab.

Ausprägung annehmen. Ist er damit wirklich als Teilmaß für Demokratiequalität geeignet, kann er also um mit Przeworsky und Teune zu sprechen als Maß gesehen werden „that is reliable across systems and valid within systems“ (Przeworsky und Teune 1970, S. 114)? Zweitens stellt sich, selbst wenn man sich dazu entschließt Indikatoren, die eigentlich nicht zwischen Autokratien und Demokratien differenzieren, als Maß für demokratische Qualität zu verwenden, die Frage, ob man mit der Mordrate nicht etwas misst, was vielmehr ein Outcome guter Regierungsführung ist als ein konstituierendes Element demokratischer Qualität. Bühlmann et al. geben selbst an, dass sie dezidiert keine Outputs und Outcomes messen wollen: „Outputs [werden] als genuin politische Entscheidungen betrachtet, die durch Demokratie als Mittel hervorgebracht werden sollten, nicht aber Bestandteil von Demokratie sind.“

Ein weiteres Beispiel für ein Überdehnen des eigentlichen Konzepts findet sich bei der Komponente Öffentlichkeit. Eine funktionierende Öffentlichkeit ist bei Bühlmann et al. über die beiden Komponenten der Vereinigungs- und Meinungsfreiheit realisierbar. Dem ist soweit durchaus zuzustimmen. Wenn nun jedoch als Indikator für die real wahrgenommene Vereinigungsfreiheit die Anzahl gewerkschaftlich Organisierter sowie die Anzahl an Mitgliedern von Menschenrechts- und Tierschutzorganisationen herangezogen wird, dann handelt es sich hierbei um Indikatoren, bei denen der Konnex zur Demokratiequalität vergleichsweise schwach ist. Der gewerkschaftliche Organisationsgrad hängt stark mit der historischen Entwicklung des Wirtschafts- und Sozialsystems zusammen, die eben nur zum Teil durch die Politik und damit Demokratie geprägt war. Aus diesem Grund kann dieser Indikator in unseren Augen auch nicht als Maß für die Qualität einer Demokratie dienen. Die in einzelnen Ländern fast schon obligatorische Gewerkschaftsmitgliedschaft (in Island listet die auch vom Demokratiebarometer als Quelle verwendete ILO zum Teil mehr als 90%, ähnlich hohe Werte finden wir auch in den übrigen skandinavischen Ländern) deutet vielmehr auf einen gewissen sozialen Druck hin zumindest offiziell Gewerkschaftsmitglied zu sein, als darauf, dass diese Personen sich wirklich, gewissermaßen basisdemokratisch in der Gewerkschaftsumgebung ausleben. Die relativ betrachtet wenigen Gewerkschaftsmitglieder in anderen Staaten sind dagegen eventuell deutlich politisch-aktiver und tragen damit zur Qualität der Demokratie sogar eher mehr bei als die Masse der skandinavischen Gewerkschaftsmitglieder in ihren Ländern. Zudem finden sich auch ähnlich hohe Werte in postkommunistischen Staaten direkt nach dem Fall des Eisernen Vorhangs. In der tschechischen Republik waren 1990 noch 80,5 % der Arbeitnehmer gewerkschaftlich organisiert. Dieser Wert nahm im Laufe der demokratischen Konsolidierung kontinuierlich auf 20,5% im Jahr 2007 ab. Das Demokratiebarometer interpretiert dies als einen klaren Verlust demokratischer Qualität. Wie sinnvoll der Indikator Gewerkschaftsdichte damit ist, möge jeder selbst beurteilen. Gleichzeitig werden die für die demokratische Willensbildung wohl relevantesten Vereinigungen ausgeblendet. Wenn schon ein hoher Prozentsatz an in Tierschutzorganisationen engagierten Bürgern die Demokratiequalität erhöht, dann muss dieselbe Argumentation erst recht für die Mitgliederzahlen von Parteien gelten. Parteimitgliedschaft wird jedoch vom Demokratiebarometer nicht erfasst. Unserer Meinung nach böte

zudem ein Rückgriff auf Umfragedaten, die politische Partizipation als demokratisches Gut selbst in den Fokus nehmen (z.B. Eurobarometer EB 60.1), eventuell eine Alternative bzw. Bereicherung der reinen mitgliederzahlenbasierten Indikatoren, die augenblicklich Verwendung finden.

Betrachtet man die Konzeption des Demokratiebarometers abschließend, ergibt sich ein gemischtes Bild: Problematisch erscheint die mangelnde Detailliertheit in Bezug auf die Entwicklung des Konzeptbaums und die daraus resultierenden Unsicherheiten bei der Zuordnung von Indikatoren. Positiv ist hingegen zu bewerten, dass sich die Autoren im Hinblick auf die Auswahl von Indikatoren der de jure/de facto Problematik bewusst sind. Grundsätzlich wird der Versuch unternommen, Komponenten sowohl über institutionelle Variablen als auch Variablen, die die Verfassungswirklichkeit erfassen, abzubilden. Dies stellt eine der zentralen Weiterentwicklungen gegenüber bisherigen Messansätzen dar, die sich primär auf institutionelle Aspekte der Demokratie kaprizieren. Allerdings läuft man hier leicht Gefahr Aspekte in die Messung mit aufzunehmen, die kein Element der Demokratiequalität selbst sind, sondern ein Outcome guter demokratischer Qualität darstellen.

4.2 Messung: Wiegt die Flut an reliablen, quantitativen Indikatoren deren mangelnde Validität auf?

Bei der Messung beschränken sich die Autoren des Demokratiebarometers auf „objektive“ Daten aus statistischen Sekundärquellen oder repräsentativen Bevölkerungsumfragen“ (Bühlmann et al. 2011a, S. 17; Herv. im Orig.). Daten aus Expertenbefragungen oder Expertenbeurteilungen sollen aufgrund fraglicher Reliabilität und einer intransparenten Kategorienzueordnung hingegen gezielt keine Verwendung finden. Der diesem Vorgehen zugrundeliegende Wunsch nach nachvollziehbaren und dabei für die Messung wirklich geeigneten Indikatoren ist vollends verständlich, es bleibt allerdings fraglich, ob ein alleiniges Zurückgreifen auf „harte Daten“ wirklich ausreicht, um die Komplexitäten erfassen zu können, die das Konzept demokratischer Qualität auszeichnen. Denn die reliabelsten Indikatoren helfen nicht viel, wenn sie am eigentlich interessierenden Konzept „vorbeimessen“. Der vorherige Abschnitt hat bereits einige Indikatoren aufgezeigt, bei denen eine entsprechende Gefahr zumindest nicht auszuschließen ist. Insofern liegt beim Demokratiebarometer eventuell eine Überbetonung des Reliabilitätskriteriums bei gleichzeitiger Vernachlässigung des eigentlich noch wichtigeren Validitätskriteriums vor. Aus forschungspragmatischer Sicht verwundert dies nicht, sind doch die so beworbenen reliablen harte Daten-Indikatoren auch gerade diejenigen Maße, die sich gut und günstig über die Zeit wie über verschiedene Länder erheben lassen. Gerade bei der Messung latenter Variablen wie Demokratie und deren Qualität, die sich bereits auf konzeptioneller Ebene als vergleichsweise komplexe Phänomene präsentieren, und bei denen klassische Proxy-Maße nur eine unzureichende Passgenauigkeit aufweisen, kann jedoch auf das Wissen von Experten eventuell nicht

verzichtet werden.¹³ Entsprechend dieser Argumentation kombinieren beispielsweise die Sustainable Governance Indicators Expertenbefragungen und harte Daten.¹⁴

Zudem ist die pauschale Kritik an den Expertenbefragungen in Bezug auf deren Reliabilität ungerechtfertigt. Wird eine ausreichende Anzahl an Experten befragt – was freilich oftmals Probleme in Bezug auf die Kosten sowie die Identifizierung potentieller Experten aufwirft – lässt sich die interne Konsistenz ihrer Antworten problemlos mit statistischen Mitteln überprüfen und damit eine ähnlich reliable Datengrundlage schaffen, wie sie statistische Daten bieten.¹⁵

4.3 Aggregation: Problematische Reskalierung und unnötig komplexe Aggregationsregel

Nach der Messung der einzelnen Indikatoren ist es nötig, die Komplexität der gesammelten Informationen durch Aggregation zu reduzieren. Hierzu muss nach M&V zunächst entschieden werden, ob es sinnvoll ist, einen einzigen Gesamtindex zu bilden, oder ob hierdurch nicht zu viele Informationen über unterschiedliche Aspekte des Gesamtkonstrukts verloren gehen würden. Für das Untersuchungsobjekt Demokratiequalität ist es sicherlich sinnvoll, einen Gesamtindex – wie beim Demokratiebarometer – zu entwickeln. Gleichzeitig wäre es gerade im Hinblick auf die verwendete Konzeption von Demokratie als dynamisches Gleichgewicht der drei Dimensionen Freiheit, Gleichheit und Kontrolle, das eben unterschiedliche Schwerpunktsetzungen als durchaus gleichdemokratisch versteht, eventuell sinnvoll die Aggregation auf dieser Ebene abzurechnen, um die unterschiedlichen Demokratieprofile auf diese Weise besser vergleichend gegenüberstellen zu können. Durch Spinnengrafiken, die von den Autoren des Demokratiebarometers extensiv eingesetzt werden, um die neun Demokratiefunktionen für einzelne Länder abzubilden, wird rein visuell diesem Anspruch jedoch durchaus Rechnung getragen. Insofern ist auch das höchstmögliche Aggregationsniveau nicht als Problem aufzufassen.

Problematisch dagegen sind zwei weitere Aspekte der Aggregation. Zunächst müssen sämtliche Indikatoren so standardisiert werden, dass sie an sich vergleichbar sind, und damit auch über eine geeignete Aggregationsregel kombinierbar werden. Sowohl die Standardisierung als auch die verwendete Aggregationsregel sind beim Demokratiebarometer jedoch suboptimal.

¹³ Bei der Messung von Parteiideologie, einem der Demokratieindikatoren in Bezug auf die Komplexität durchaus vergleichbaren Konstrukt, haben sich Expertenbefragungen als sinnvolles Instrument herausgestellt, gerade weil sie es erlauben einige Schlüsselparameter, die sich über härtere Datenquellen nur mangelhaft bestimmen lassen, gezielt abzufragen (vgl. Castles und Mair 1984; Laver und Hunt 1992, S. 34).

¹⁴ Auch wenn an der Art und Weise Kritik geübt werden kann, wie die SGI die Expertenbefragungen im Prozess der Indexbildung verarbeiten (Jäckle und Bauschke 2009, 2010), so ist der grundlegende Gedanke komplexe Sachverhalte über Expertenwissen anzunähern und diese Informationen mit den statistischen Daten zu einem Gesamtindex zu integrieren nicht von der Hand zu weisen.

¹⁵ Zudem, wenn bereits Expertenbefragungen aufgrund einer mangelnden Reliabilität vermieden werden sollen, stellt sich die Frage, weshalb auf Daten des IMD zurückgegriffen wird, die aus einer Umfrage unter den Absolventen dieser Schweizer Business School gewonnen werden (vgl. http://www.imd.org/research/centers/wcc/research_methodology.cfm).

4.3.1 Künstlich erzeugte Varianz durch Standardisierung

Die Standardisierung erfolgt entsprechend empirischer Maximal- und Minimalwerte, was als „best-practice-Verfahren“ bezeichnet wird. Hierzu wird ein *blueprint-sample* aus den 30 etablierten Demokratien zwischen 1995 und 2005 betrachtet. Der höchste Wert, den jeder Indikator über alle 330 Länderjahre annimmt, wird auf den Wert 100 gesetzt, der niedrigste auf den Wert 0. Die ursprünglichen Indikatorwerte dazwischen werden entsprechend linearer Transformation auf der hierdurch entstehenden Skala zwischen 0 und 100 verteilt. Problematisch an diesem Verfahren sind aus unserer Sicht zwei Aspekte: Erstens werden entgegen der ursprünglichen Konzeption, die davon ausgeht, dass Demokratie aus unterschiedlichen Kombinationen von Gleichheit, Freiheit und Kontrolle bestehen kann, doch wieder jeweils ein möglichst hoher Wert bei jedem Indikator als besonders gut für die Demokratiequalität betrachtet. Hierdurch wird implizit nichts anderes gemacht, als die Problematik der theoretischen Bestimmung der Maxima und Minima¹⁶ auf ein theoretisch unreflektiertes „der Wert sollte möglichst groß sein“ auszulagern. Am Wahlbeteiligungsbeispiel hieße das, dass eine Demokratie, in der Wahlpflicht herrscht, den Standard für gute Demokratie setzt. Damit ist es die partizipatorische Demokratietheorie, die über andere Demokratieverständnisse dominiert.

Zweitens spreizt das gewählte Skalierungsverfahren die Daten künstlich auf. Wenn beispielsweise einzelne Indikatoren nur eine minimale Varianz in den Rohdaten aufweisen, werden diese Unterschiede so aufgeblasen, dass sie dieselbe Spannweite aufweisen, wie Indikatoren, in denen sich die Länder wirklich deutlich unterscheiden. Diese aufgespreizte Varianz fließt dann in gleicher Weise in die Indexberechnung ein wie die reale Varianz anderer Indikatoren. Dies ist einerseits ein konzeptionelles Problem, da eigentlich irrelevante Faktoren plötzlich wichtig erscheinen und Länder, die nur marginal schlechter abschneiden, überproportional abgestraft werden und andererseits ein praktisches Problem, da durch die Transformation die Transitivität des Index verletzt wird. Dies bedeutet, dass aus einer vollständig transitiven Ordnung der ursprünglichen Rohindikatorwerte keinerlei Rangfolge bei den transformierten Werten mehr herausgelesen werden kann. Tabelle 2 verdeutlicht diese Situation:

Tabelle 2: Beispiel zur Verdeutlichung der Skalierungsproblematik

	In Blueprint Sample		Rohdaten			Transformierte Werte		
	Min	Max	Land A	Land B	Land C	Land A	Land B	Land C
Indikator X	10	30	10	20	30	0	50	100
Indikator Y	20	80	80	50	20	100	50	0
Mittelwert	-	-	45	35	25	50	50	50

¹⁶ Dies wird von den Autoren des Demokratiebarometer sehr anschaulich anhand des Beispiels Wahlrecht/Wahlbeteiligung dargestellt (Bühlmann et al. 2011a, S. 18-19).

Lässt sich aus den ursprünglichen Indikatorwerten noch eine eindeutige Rangfolge mittels Durchschnittsbildung ablesen¹⁷ ($A > B > C$), ist dies nach erfolgter Transformation der Werte nicht mehr möglich ($A = B = C$). Damit macht die Transformation der Werte die Situation tendenziell unklarer, was im Gegensatz zu der Hauptaufgabe eines Index allgemein steht, die Komplexität der Realität klarer geordnet darzustellen. Im ungünstigsten Fall könnte dies sogar dazu führen, dass sich auch die Rangfolge der Länder im Demokratiebarometer aufgrund der Skalierung verändert. Dadurch, dass für die Transformation stets auf dasselbe blueprint-sample zurückgegriffen wird, umgeht das Demokratiebarometer ein Problem, das bei den Sustainable Governance Indicators deutlich zu Tage tritt. Die SGI verwenden dieselbe „best-practice“-Skalierung, allerdings werden dort für jede Welle die Minima und Maxima neu bestimmt, wodurch ein Vergleich der Indexwerte über die Zeit, d.h. über mehrere Erhebungswellen hinweg nicht möglich ist. Beim Demokratiebarometer ist dieser, solange die Transformation auf Basis derselben Vergleichsjahre durchgeführt wird, möglich, allerdings mit der Konsequenz, dass sich Werte über 100 und unter 0 ergeben können – problematisch ist dies jedoch nicht, so lange deutlich – wie bei den Autoren des Demokratiebarometers – darauf hingewiesen wird, dass der Wert 100 kein absolutes Optimum darstellt.

Für das 2007 publizierte Demokratiebarometer haben wir zu Vergleichszwecken alle Rohdaten der 100 Indikatoren z-transformiert und dann wie beim Demokratiebarometer auch als arithmetisches Mittel bis hin zur Ebene der Komponenten aggregiert.¹⁸ Der Zusammenhang zwischen den entsprechend der beiden Skalierungsverfahren berechneten Komponentenwerten ist zwar durchaus hoch (zwischen $r = 0,89$ und $r = 0,99$) gleichzeitig zeigen sich aber auch einige Abweichungen. Von den insgesamt 540 Komponentenwerten bleiben die Rangplätze nur bei 170 unverändert. Im Durchschnitt verändern die Länder ihre Positionen auf der Ebene der Komponenten um 1,44 Rangplätze. Bei den gravierendsten Fällen gibt es sogar Abweichungen zwischen den beiden Skalierungsverfahren von bis zu 10 Rangplätzen (Spanien bei der Komponente *substantive representation*).¹⁹ Insgesamt muss aber festgestellt werden, dass die von uns beschriebene verzerrende Wirkung des vom Demokratiebarometer verwendeten Aufspreizungsverfahrens empirisch bei den vorliegenden Daten nicht allzu stark ins Gewicht fällt, obgleich die grundlegende Problematik weiterhin offensichtlich ist.

Ähnlich wie bei den SGI lassen sich die skalierten Werte des Demokratiebarometers aber sinnvollerweise lediglich ordinal interpretieren (vgl. Brusis 2009b, S. 544-545), die verwendeten Aggregationsregeln (arithmetisches Mittel und Arkustangens-Funktion) benötigen jedoch metrisches

¹⁷ Die Aggregation von den Indikatorwerten zu den Subkomponenten erfolgt auch beim Demokratiebarometer mittels arithmetischem Mittel.

¹⁸ Die weitere Aggregation erfolgt beim Demokratiebarometer mittels der Arkustangens-Funktion (vgl. Kapitel 3.3.2). Diese ist nur mit positiven Ausgangswerten möglich. Da die z-Standardisierung auch negative Werte hervorbringt brechen wir die zu Vergleichszwecken durchgeführte Aggregation nach dem zweiten Aggregationsschritt auf der Ebene der Komponenten ab.

¹⁹ Eine Übersicht über die Veränderungen der Rangplätze findet sich im Onlineanhang (Tabelle A2).

Skalenniveau und auch die letztlich vom Demokratiebarometer berichteten Indexwerte werden nicht als Ranking präsentiert, sondern als Intervallskala implizierende Komma-Werte.

4.3.2 Arkustangensfunktion als Aggregationsregel

Die Aggregation erfolgt auf den untersten beiden Ebenen von den Indikatoren über die Subkomponenten zu den Komponenten mittels des arithmetischen Mittels. Für die folgenden drei Aggregationsschritte bis hin zum Gesamtindex wird die Arkustangens-Funktion verwendet. Zunächst wird also davon ausgegangen, dass die einzelnen Indikatoren (Subkomponenten) für die entsprechende Subkomponente (Komponente) alle dasselbe Gewicht haben und dass es sich um jeweils hinreichende Bedingungen handelt. Im späteren Verlauf werden dann die Komponenten, Funktionen und Prinzipien jeweils als notwendige Bedingungen für die jeweils höhere Ebene betrachtet, weshalb ein multiplikatives Aggregationsverfahren kombiniert mit der Arkustangens-Funktion gewählt wurde.²⁰ Warum gerade ab der dritten Ebene die Aggregation auf der Logik notwendiger Bedingungen beruht, darunter jedoch die einzelnen Indikatoren/Subkomponenten nicht notwendig sind, wird theoretisch nicht begründet.

Betrachtet man dieses Phänomen jedoch aus empirischer Warte, wird diese Entscheidung verständlicher: Würde man bereits auf der untersten Ebene die einzelnen Indikatoren als notwendige Bedingungen auffassen und entsprechend multiplikativ aggregieren, so würden dadurch, dass jedes der betrachteten Länder mindestens bei einem Indikator den schlechtesten Wert Null bekommt, alle Länder sowohl während des Zeitraums des blueprint-samples als auch für das Jahr 2007 den Wert Null auch für den Gesamtdemokratiewert erhalten. Wenn man den ersten Aggregationsschritt mittels arithmetischem Mittel durchführt und erst ab dem zweiten multiplikativ aggregiert, dann weisen immer noch 24 von 30 Ländern einen Demokratiewert von Null auf. Wenn man hingegen – wie das Demokratiebarometer – erst ab dem dritten Aggregationsschritt die Logik notwendiger Bedingungen ansetzt, fällt über den gesamten Beobachtungszeitraum kein einziges Land auf den Wert Null. Dies würde nur geschehen, wenn ein Land in allen in der Regel sechs Indikatoren, die eine Komponente auszeichnen, einen genauso schlechten Wert hätte wie das schlechteste Land während des blueprint-sample Vergleichszeitraums.

Die Arkustangensfunktion soll eine begrenzte Substituierung ungleicher Werte ermöglichen oder anders ausgedrückt, es soll eine möglichst gleichmäßige Verteilung der jeweiligen zu aggregierenden Subkomponenten (beispielsweise der drei Prinzipien Freiheit, Gleichheit und Kontrolle) belohnt werden. Letztlich vergrößert dieses Aggregationsverfahren mittels Progression die Messunterschiede und es wird wiederum eine künstliche Varianz geschaffen. Die Annahme dahinter ist, dass die Erhöhung eines an sich schon sehr hohen Wertes lediglich einen marginalen zusätzlichen Nutzen für die

²⁰ Um den Wertebereich in etwa auf 0 bis 100 festzusetzen werden an der eigentlichen Arkustangens-Funktion noch gewisse arithmetische Umformungen vorgenommen: $(\arctan(\text{Komponente1} * \text{Komponente2}) * 1,2/4000) * 80$). Den Grund für diese Umformung muss sich der Leser jedoch selbst herleiten, da sich in den Publikationen des Demokratiebarometers hierzu keine Informationen finden.

Demokratiequalität hat. Eine Annahme, die so aus dem grundlegenden Konzept von Demokratie als dynamisches Gleichgewicht zwischen Freiheit, Gleichheit und Kontrolle nur bedingt abgeleitet werden kann. Tabelle 3 verdeutlicht den Effekt der vom Demokratiebarometer verwendeten Arkustangens-Aggregationsregel im Vergleich zu den weitaus üblicheren Varianten des arithmetischen und geometrischen Mittels.

Tabelle 3: Unterschiedliche Aggregationsregeln im Vergleich

	Wert 1	Wert 2	arithmetisches Mittel	geometrisches Mittel	Arkustangens-Funktion
Land A	1	99	50,00	9,95	2,38
Land B	25	75	50,00	43,30	40,99
Land C	49	51	50,00	49,99	51,46
Land D	50	50	50,00	50,00	51,48

Es wird deutlich, dass die Arkustangens-Formel noch stärker als die reine Multiplikation, welche dem geometrischen Mittel zugrunde liegt, ungleiche Verteilungen der Basiswerte bestraft und gleichzeitig ähnliche Verteilungen so stark belohnt, dass der aggregierte Wert letztlich sogar größer sein kann als die Werte, aus denen er gebildet wurde.

Ferner ist auch die Gewichtung ein Einfallstor für Verzerrungen. Die Autoren des Demokratiebarometers möchten allen 100 Indikatoren in der Messung ein gleiches Gewicht zuweisen. Über die unterschiedlichen Besetzungszahlen der Subindikatoren und Indikatoren werden diese jedoch in letzter Konsequenz (vgl. Abbildung 1) nicht gleich gewichtet. Zudem ist die angestrebte Gleichgewichtung inhaltlich teilweise nicht sonderlich plausibel: So wird in der gegenwärtigen Konstruktion den Einstellungen zum Schwarzfahren das gleiche Gewicht für die Demokratiequalität zugewiesen wie der Möglichkeit direktdemokratisch abzustimmen.

Insgesamt ist die Aggregationsmethode durchaus ausführlich beschrieben und damit nachvollziehbar, allerdings weist sie, wie gezeigt wurde, gewisse Differenzen zur grundlegenden Konzeption auf. Zudem, und dies ist ein Aspekt, der nicht unterschätzt werden sollte, dürfte die Aggregation mittels Arkustangensfunktion von einem Großteil der Leser mathematisch nur bedingt nachzuvollziehen sein. Eine simplere Aggregationsregel, beispielsweise über das geometrische Mittel, würde zwar keine so starke Abstrafung von Ländern mit unterschiedlich ausgeprägten Demokratiekomponenten ermöglichen, sie wäre jedoch weitaus besser nachzuvollziehen.²¹

5 Die Qualität der Qualitätsmessung – zur Validierung des Demokratiebarometers

²¹ Da das geometrische Mittel positive Werte voraussetzt müsste die nach der aktuell verwendeten Skalierungsmethode potentiell negative Werte aufweisende Skala in den positiven Wertebereich verschoben werden, oder eine Skalierung angewandt werden, die erst gar keine negativen Werte hervorbringt.

Die Autoren des Demokratiebarometers sind sich durchaus der Notwendigkeit bewusst, ihre eigene Messung anhand gewisser Gütekriterien zu überprüfen. Sie beschränken sich hierbei jedoch auf die drei Validitätskriterien der Konstrukt-, Kriteriums- sowie der endogenen Validität. Die Inhaltsvalidität, die abprüft, ob das theoretisch entwickelte Konzept auch passgenau über messbare Indikatoren abgedeckt wird und die klassischerweise bei einer Indexkonstruktion unter Validität zu verstehen ist, wird hingegen vernachlässigt. Dabei bestehen in Anbetracht der von uns vorgenommenen Analyse (Kapitel 3.1) des gleichzeitig vom Stamm und den Blättern aufeinanderzuwachsenden Konzeptbaumes gerade hier in unseren Augen Zweifel bezüglich der Passgenauigkeit der gewählten Indikatoren.

5.1 Konstruktvalidierung über Clusteranalyse?

Als Möglichkeit einer Konstruktvalidierung schlagen die Autoren eine Clusteranalyse auf Basis der neun Funktionen vor. Sie nehmen an, dass sich unterschiedliche Cluster bilden, die jeweils abweichende Demokratieprofile repräsentieren. Dies erscheint vor der Annahme, dass die zugrundeliegende Theorie korrekt ist, als sinnvolle Option, jedoch kann aus methodischer Perspektive eingewendet werden, dass alternative Clusterungsalgorithmen zu vollkommen anderen Gruppierungen führen und dieses Verfahren sehr sensitiv sein kann (Wagschal 1999). Der Versuch die Clusterlösung für das Jahr 2005 auf Basis des im Internet verfügbaren Demokratiebarometer-Datensatzes (standardisierte Version) und der von Bühlmann et al. angegebenen Clusterungsvorgaben (quadrierte Euklidische Distanz, Ward-Verknüpfung) zu replizieren führte allerdings nicht zu der von Bühlmann und seinen Kollegen präsentierten Clusterlösung (2011a, S. 27).²²

Wichtiger noch als die Frage was für Cluster sich bilden, ist jedoch für die Macher des Demokratiebarometers die Tatsache, dass sich überhaupt unterschiedliche Gruppen identifizieren lassen. Ob sich diese Gruppen aber entsprechend der neun Funktionen so stark unterscheiden, dass es auch inhaltlich sinnvoll ist von distinkten Demokratieprofilen zu sprechen, kann eine Clusteranalyse allein nicht klären. Hierzu ist es sinnvoll die Mittelwerte der Gruppen zu vergleichen. Während Bühlmann et al. dies mittels einer deskriptiven Tabelle darstellen, soll hier nun durch einen Mittelwertvergleich untersucht werden, ob sich die Cluster entsprechend der Mittelwerte der neun Funktionen über die in einer Gruppe enthaltenen Länder wirklich signifikant unterscheiden.²³

Tabelle 4 zeigt, dass nur 21 der 54 Gruppenvergleiche auf der Ebene der Funktionen signifikant unterschiedliche Mittelwerte aufweisen. Auch auf den höher aggregierten Ebenen lassen sich die Gruppen zum Teil nicht signifikant differenzieren. Die von uns durchgeführte Clusteranalyse kann

²² Im Online-Anhang findet sich zu Vergleichszwecken neben der von uns gefundenen Clusterlösung (Abbildung A1) auch die von Bühlmann et al. berichtete (Tabelle A3).

²³ Da sich die Ergebnisse der Clusteranalyse von Bühlmann et al. nicht replizieren lassen, wird für den Mittelwertvergleich auf die sich in unserer Clusteranalyse ergebende Gruppeneinteilung zurückgegriffen. Es werden ebenfalls vier Cluster betrachtet: Gruppe 1: Island, Norwegen, Belgien, Dänemark, Niederlande, Slowenien, Finnland, Schweden, Neuseeland, Australien, Kanada, Österreich, Deutschland, Spanien; Gruppe 2: Tschechische Republik, Ungarn, Italien, Costa Rica, Polen, Malta, Südafrika; Gruppe 3: Luxemburg, Schweiz, Zypern; Gruppe 4: Frankreich, Großbritannien, Irland, Portugal, Japan, USA.

insofern nur sehr bedingt als eine Konstruktvalidierung des Demokratiebarometers gelten. Zudem zeigt sich dass es notwendig ist die Clusteranalyse durch einen systematischen Mittelwertvergleich der Gruppen zu ergänzen um sie als Konstruktvalidierung einsetzen zu können.

Tabelle 4: Mittelwertvergleich zwischen den gefundenen Clustern

	Signifikante Unterschiede zwischen den Gruppen vorhanden? (T-Test für unabhängige Stichproben, 2-seitig, 95% Signifikanzniveau)					
	Gruppe 1 vs. Gruppe 2	Gruppe 1 vs. Gruppe 3	Gruppe 1 vs. Gruppe 4	Gruppe 2 vs. Gruppe 3	Gruppe 2 vs. Gruppe 4	Gruppe 3 vs. Gruppe 4
Individuelle Freiheiten	Nein	Nein	Nein	Nein	Nein	Nein
Rechtsstaatlichkeit	Ja	Nein	Nein	Ja	Ja	Nein
Öffentlichkeit	Nein	Nein	Nein	Ja	Nein	Nein
Wettbewerb	Ja	Nein	Ja	Nein	Nein	Ja
Gewaltkontrolle	Nein	Nein	Nein	Ja	Nein	Nein
Regierungs- und Implementationsfähigkeit	Ja	Nein	Ja	Ja	Nein	Nein
Transparenz	Ja	Ja	Nein	Nein	Ja	Ja
Partizipation	Ja	Ja	Ja	Nein	Nein	Nein
Repräsentation	Ja	Ja	Ja	Nein	Nein	Nein
Freiheit	Ja	Nein	Nein	Ja	Nein	Nein
Kontrolle	Ja	Nein	Ja	Nein	Nein	Nein
Gleichheit	Ja	Ja	Ja	Nein	Ja	Nein
Demokratiequalität	Ja	Ja	Ja	Nein	Nein	Nein

5.2 Externe Validität

Bei der Überprüfung der externen Validität, die unverständlicherweise als endogene Validität bezeichnet wird, greifen die Autoren des Demokratiebarometers auf die Worldwide Governance Indicators (WGI) der Weltbank zurück und zeigen, dass generell starke Korrelationen zwischen einem aus den sechs Dimensionen aggregierten Governance-Gesamtindex und dem Demokratiebarometer bestehen, wobei dessen einzelne Funktionen unterschiedlich stark mit den Governance Indicators zusammenhängen. Insbesondere die Kontrollfunktionen würden weniger stark korrelieren, was laut Bühlmann et al. dafür spricht, dass die Governance Indicators diese Dimension ähnlich wie andere Demokratiemaße und im Gegensatz zum Demokratiebarometer allzu stiefmütterlich behandeln würden (Bühlmann et al. 2011c, S. 30-31). Fraglich ist allerdings, weshalb für den Test der externen Validität auf ein Maß zurückgegriffen wird, das gar nicht beabsichtigt Demokratie bzw. demokratische Qualität zu messen.

Ein weiteres weniger theoretisch sondern praktisch gelagertes Problem der externen Validierung über die WGI ist zudem in der potentiellen Quellenüberschneidung der beiden Messungen zu sehen. Der direkte Vergleich der Quellen des Demokratiebarometers und der WGI deutet auf die Existenz deutlicher Überschneidungen hin. Von den 30 Quellen, die in die WGI einfließen, finden sich 10 auch

in den Quellen des Demokratiebarometers.²⁴ Zwar lässt sich aufgrund der komplexen Indikatorenzuordnung und Konstruktion des Demokratiebarometers nicht genau rekonstruieren, welche Indikatoren aus anderen Messungen übernommen wurden, allerdings kann basierend auf dem Codebuch zumindest für die 9 Funktionen die Quellenüberschneidung aufgeschlüsselt werden (vgl. Tabelle 5). In der Gesamtbetrachtung liegt der Anteil der sich überschneidenden Quellen bei knapp einem Sechstel, und in manchen Funktionen – z.B. Rule of Law – bei über einem Drittel. Der gefundene Zusammenhang zwischen den beiden Maßen kann vor diesem Hintergrund weniger verwundern und ist daher als externe Validierung nur noch bedingt geeignet.

Tabelle 5: Identische Quellen von Demokratiebarometer und Governance Indicators

Komponente	Identische Quelle / Gesamtanzahl der Quellen
Individuelle Freiheiten	6/18
Rechtsstaatlichkeit	11/30
Öffentlichkeit	2/45
Wettbewerb	0/31
Gewaltkontrolle	0/19
Regierungs- und Implementationsfähigkeit	7/22
Transparenz	6/26
Partizipation	12/76
Repräsentation	3/35
Gesamtanteil	47/302

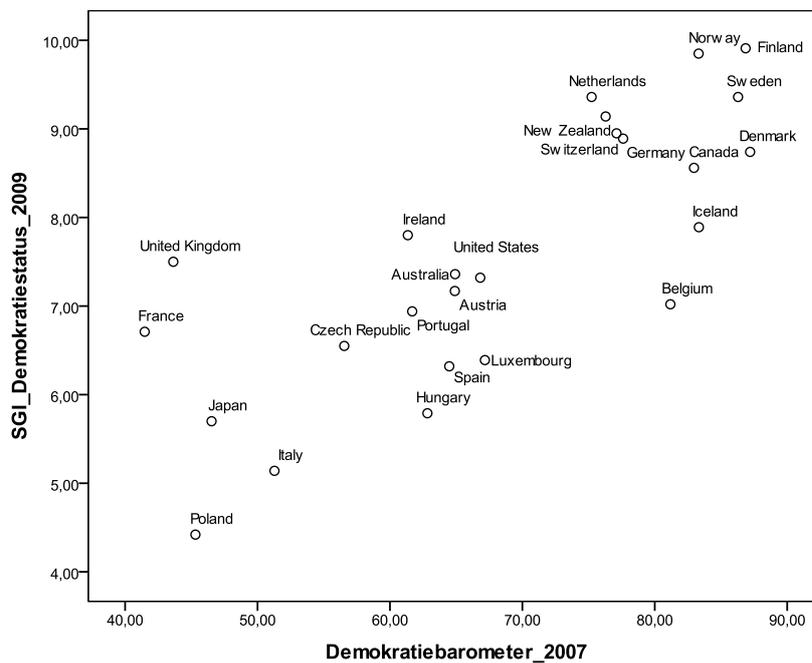
Eigene Darstellung basierend auf Bühlmann et al. (2011). Democracy Barometer. Codebook for Blueprint Dataset Version 2. Aarau: Zentrum für Demokratie.) und WGI (<http://info.worldbank.org/governance/wgi/sources.htm>)

Ein Rückgriff auf etablierte Maße (FH, Polity IV usw.) stellt zwar aus offensichtlichen Gründen auch keine Alternative dar, aber ein Vergleich mit den Sustainable Governance Indicators, die zumindest vorgeben ein feineres Messinstrumentarium für Demokratie zu verwenden, erscheint angebracht. Abbildung 2 trägt den SGI Status der Demokratie für 2009 gegen den Demokratiebarometerwert für 2007 ab. Das Ergebnis dürfte die Macher des Demokratiebarometers durchaus erfreuen. Ein Zusammenhang zwischen den beiden Maßen lässt sich nicht leugnen. Auch wenn aufgrund unterschiedlicher Länderabdeckung nur 25 Staaten in den Vergleich einbezogen werden können so ergibt sich doch ein Spearman Rangkorrelationskoeffizient von 0,744 (signifikant auf dem 99%-Niveau).

Allerdings zeigt das Streudiagramm auch an, welche Fälle beim Demokratiebarometer im Vergleich zum Demokratiestatus des SGI unterbewertet sind. Vor allem überrascht das schlechte Abschneiden von Frankreich und dem Vereinigten Königreich, die man als lange etablierte Demokratien nicht auf den beiden letzten Plätzen vermutet hätte (vgl. auch Tabelle 1). Dagegen ist Belgien beim Demokratiebarometer im Vergleich zum SGI überbewertet. Interessant wäre bei einem solchen Abschneiden von zentralen Fällen die Überprüfung der externen Validität der Indikatoren, die für diesen Befund verantwortlich sind.

²⁴ Darüber hinaus werden die WGI selbst als Quelle im Demokratiebarometer ausgewiesen.

Abbildung 2: Demokratiebarometer vs. SGI-Demokratiestatus



5.3 Adäquate Präsentation der Daten im Internet fördert Nachvollziehbarkeit

Als grundsätzlich gelungen, insbesondere im Vergleich zu etablierten Indizes wie Freedom House oder Polity IV, ist die Präsentation der Daten zu bezeichnen. Auf der Homepage des Demokratiebarometers wird nicht nur der Gesamtindex dargestellt, sondern gleichzeitig auch die Methodik beschrieben, die Ergebnisse für alle Staaten in einzelnen Länderstudien verdeutlicht sowie die Datenbasis des Demokratiebarometers sowohl in standardisierter als auch in unstandardisierter Rohdatenform zum Download zur Verfügung gestellt. Zudem bietet sie die Möglichkeit eigene auf spezifische Fragestellungen angepasste Vergleichsgrafiken direkt zu erstellen. Die umfassende Dokumentation und Datenoffenlegung erhöht zweifellos die Nachvollziehbarkeit. Einzig die Zuordnung der Daten zu den ursprünglichen Quellen gestaltet sich schwieriger als nötig. Die kurzen Beschreibungen, die zusammen mit der Angabe der Sekundärquelle der Daten in den Tabellen des Methodologie-Papiers (Bühlmann et al. 2011c, S. 20-42) aufgeführt sind, erlauben zumeist keine einfache Zuordnung. Eine zusätzliche Spalte mit der Bezeichnung des Indikators im ursprünglichen Datensatz würde die Nachvollziehbarkeit

erhöhen. Zudem könnte der geneigte Leser so auch einfacher für sich entscheiden, ob der verwendete Indikator wirklich passend ist.

5.4 Objektivität unklar und nicht thematisiert

Das dritte Kriterium der Objektivität bezeichnet, die Unabhängigkeit des Indikators von normativen Verzerrungen, insbesondere etwa vom Einfluss derjenigen, welche die Datenerhebung durchführen bzw. das Konzept des Indikators definieren. Bei Umfragen kann beispielsweise durch Intercoderreliabilitätsmaße überprüft werden, wie übereinstimmend Länder von den Kodierenden bewertet werden. Objektivität ist jedoch auch von Bedeutung, wenn statistische Daten externer Quellen herangezogen werden. Prima vista gilt oftmals die Überlegenheit solcher Sekundär- und Aggregatdaten gegenüber Individualdaten und eigenen Primärdaten. Doch auch im Falle von Sekundärdaten ist die Qualität der Daten nicht immer gesichert (Elklit 1994). Bei 100 Variablen im Datensatz kann eine Qualitätsprüfung nur stichprobenartig erfolgen, aber anhand der direktdemokratischen Beteiligungsvariablen DIRDEM soll das Problem verdeutlicht werden.

Die Autoren verwenden einen zusammengesetzten Indikator, der aus Daten für das obligatorische und fakultative Referendum besteht. Am Ende liegt die Schweiz wie erwartet an der Spitze, auf Platz 2 findet sich jedoch Spanien, in dem seit 1990 nur ein einziges nationales Referendum stattfand. Damit befindet sich Spanien in der Gruppe der Länder, die am wenigsten direktdemokratisch abstimmen („Rules in Use“, Quelle C2D Aarau). Auch bei den konstitutionellen Möglichkeiten ist Spanien bestenfalls im Mittelfeld der europäischen Länder. Diese „Rules in Form“ werden vom IRI-Index (Kaufmann und Waters 2004) gemessen und Spanien wird hier der Gruppe der Vorsichtigen zugerechnet. So liegt Italien, welches bei beiden Indikatoren vor Spanien liegt (die aus den gleichen Datenquellen wie bei Demokratiebarometer kommen) im Demokratiebarometer hinter diesem Land, obgleich die Datenbasis identisch ist. Solche Fehler können nur durch Qualitätskontrolle ausgemerzt werden, was durch die lobenswerte Transparenz immerhin für Externe möglich ist. Problematisch ist jedoch das Negieren der Objektivitätsproblematik bei Bühlmann et al. (2011a; 2011b).

6 Fazit: Ein vielversprechender Ansatz – mit deutlichen Defiziten

Betrachtet man das Problem der mangelnden Varianz innerhalb etablierter Demokratien als eine der zentralen Herausforderungen der Demokratiemessung, kommt das Demokratiebarometer dem eigenen Anspruch nach zu urteilen einer Komplettlösung gleich. Zum einen sollen methodische Defizite bisheriger Ansätze umgangen und zum anderen eine wesentlich präzisere Analyse etablierter Demokratien erreicht werden. Sowohl die Vorgehensweise bei der Entwicklung und Konstruktion des Messvorschlages als auch die eigentlichen Ergebnisse des Demokratiebarometers lassen hierbei zunächst vermuten, dass dieses neue Instrument seine hehren Ziele auch erfüllt. Im Hinblick auf das „Design“ muss man dem Demokratiebarometer insbesondere im Vergleich zu vielen „alten“ Messansätzen eine stärkere Sensitivität für die Problematik und Relevanz der theoretischen

Verankerung und entsprechender Konzeptualisierung zu Gute halten. Dies wird vor allem in der relativ breiten Erläuterung des Demokratiekonzepts deutlich.

Ein weiterer wichtiger Aspekt, den das Demokratiebarometer explizit adressiert, ist die bis dato eher wenig beachtete Diskrepanz zwischen de jure und de facto Indikatoren in der Demokratiemessung (vgl. Lauth 2004). Ebenso positiv schneidet die Messanlage im Hinblick auf die Verfügbarkeit und Präsentation der Daten über die Projekthomepage ab. Schließlich lässt der Blick auf die bisherige Probemessung den Schluss zu, dass das Demokratiebarometer seine eigentliche Kernaufgabe, die Binnendifferenzierung etablierter Demokratien, scheinbar erfüllt.

Auf den zweiten Blick treten jedoch zahlreiche problematische Aspekte zu Tage, die uns trotz der genannten Vorzüge dazu veranlassen die Leistungsfähigkeit des Demokratiebarometers kritisch zu bewerten. Neben des bereits von Kaina (2008) und Lauth (2010) adressierten Problems, dass das grundsätzliche, vom Demokratiebarometer verwendete Konzept von (demokratischer) Qualität nur mangelhaft expliziert wird und dabei dessen inhärentes Spannungsverhältnis zwischen den drei Demokratieprinzipien verloren geht, lassen sich sowohl auf der Ebene der Konzeption, der Messung als auch der Aggregation Defizite feststellen.

Auf der konzeptionellen Ebene lässt sich festhalten, dass die Wahl einer maximalistischen und damit zwangsläufig anspruchsvolleren Demokratiedefinition zwar als zielführend anzusehen ist, sie jedoch in Verbindung mit der gewählten Operationalisierung zu einem potentiellen Problem wird. Erstens verliert die logische Verbindung der verschiedenen Konzeptebenen umso weiter wir uns im Konzeptbaum nach oben bewegen an Argumentationstiefe und Überzeugungskraft. Gerade das Vorgehen in der Verbindung der Subkomponenten und der eigentlichen Indikatoren hätte hierbei mehr Raum einnehmen sollen und müssen. Zweitens, und dies erscheint auch als das schwerwiegendere Problem, wird die Binnendifferenzierung über einen potentiell überladenen Indikatorenapparat erreicht. Gerade in Verbindung einer mangelnden Erläuterung der Indikatorenauswahl drängt sich der Verdacht auf, dass das Ziel der Varianzerzeugung hierbei zumindest teilweise höher als das der theoretischen Eignung der Indikatoren gewichtet wurde. Denn auch wenn der Auswahlprozess wie von den Autoren selbst hervorgehoben „basically theory driven“ erfolgt, lassen sie den Leser an diesem Prozess nicht ausreichend teilhaben, um die theoretische Geeignetheit der Indikatoren individuell zu bewerten und nachzuvollziehen. Die Messanlage läuft hierbei Gefahr Indikatoren aufzunehmen, die fraglich im Hinblick auf ihre „cross-systemische“ Äquivalenz sind und teilweise nur mit viel gutem Willen als Proxy für demokratische Qualität angesehen werden können.

Im Hinblick auf die eigentliche Messung erscheint die weitgehende Ablehnung auf Expertenbefragungen beruhender Indikatoren teilweise willkürlich. Der Fokus auf harte Daten im Sinne einer gesteigerten Reliabilität überlagert hierbei die Sicherstellung der Validität der Messung. Die Ebene der Aggregation bietet aus unserer Sicht ebenfalls eine große Angriffsfläche. Erstens erscheint

die vorgenommene Reskalierung als reine Aufspreizung mangelnder Unterschiede, die darüber hinaus im ungünstigsten Fall sogar zu einer Verschiebung der Rangfolge der Länder im Demokratiebarometer führen kann. Vor diesem Hintergrund entpuppt sich die erreichte höhere Differenzierung als fabrizierte Varianz. Alternativen sollten zumindest durchgespielt und diskutiert werden. Zweitens steht es im Widerspruch zu der ursprünglichen Konzeption des Demokratiebarometers, nach der hohe Demokratiequalität durch unterschiedliche Kombinationen aus Gleichheit, Freiheit und Kontrolle verwirklicht werden könne, wenn ein möglichst hoher Wert bei jedem Indikator als besonders gut für die Demokratiequalität betrachtet wird. Drittens wird auch die Frage der Gewichtung nicht thematisiert. Schließlich bedeutet die Verwendung des Arkustangens eine unnötige Verkomplizierung ohne besonders großen Mehrwert, die dabei die Nachvollziehbarkeit des Aggregationsprozesses erschwert. Die Alternative wäre das geometrische Mittel.

Die von den Autoren vorgenommenen Überprüfungen der Güte des eigenen Messvorschlages sind begrüßenswert, jedoch bietet das gewählte Vorgehen erneut Raum für Kritik. Neben der prinzipiellen Vernachlässigung der Inhaltsvalidität, erweist sich die Konstruktvalidität nur als bedingt gewährleistet. Plausibilitätschecks, statistische Sensitivitätsanalysen und die nicht diskutierte Objektivitätsproblematik geben Raum zur Verbesserung. Die Heranziehung der World Governance Indicators im Sinne einer externen Validierung betrachten wir nicht nur aufgrund der fragwürdigen theoretischen Eignung sondern auch aufgrund der teils deutlichen Quellenüberschneidungen als wenig zielführend und schlagen daher die Verwendung alternativer Maße wie der SGI vor, die, auch wirklich wie das Demokratiebarometer, den Anspruch formulieren demokratische Qualität zu bemessen.

In der Gesamtschau ziehen wir somit ein überwiegend skeptisches Resümee, denn auch wenn das Demokratiebarometer Optimierungen gegenüber vorhergegangenen Messvorschlägen beinhaltet, zeigt es doch auch noch eine Reihe an Schwächen sowohl in theoretisch-konzeptioneller als auch in praktischer Hinsicht. Um ein letztes Mal die Metapher zu bemühen, die uns durch den bisherigen Artikel begleitet hat: Der Konzeptbaum des Demokratiebarometers mag zwar über stärkere Wurzeln verfügen als bisherige Demokratiemaße, diese können aber nicht verhindern, dass er aufgrund eines fragilen Stammes und einer vor Blättern strotzenden Baumkrone Gefahr läuft umzuknicken.

Literatur

- Abromeit, Heidrun. 2004. Die Messbarkeit von Demokratie: Zur Relevanz des Kontextes. *Politische Vierteljahresschrift* 45 (1): 73 - 93.
- Beetham, David, Edzia Carvalho, Todd Landman und Stuart Weir. 2009. *Assessing the Quality of Democracy - A Practical Guide*. Stockholm: International IDEA.
- Berg-Schlosser, Dirk. 1999. *Empirische Demokratieforschung - Exemplarische Analysen*, Frankfurt a.M.: Campus Verlag.
- Brusis, Martin. 2009a. Designing Sustainable Governance Indicators - Assessment Criteria and Methodology. In *Sustainable Governance Indicators 2009*, Hrsg. Bertelsmann-Stiftung, 71-100. Gütersloh: Verlag Bertelsmann Stiftung.

- Brusis, Martin. 2009b. Konzepte, Messansatz und Validierung der Sustainable Governance Indicators. Eine Replik auf Sebastian Jäckle und Rafael Bauschke. *Zeitschrift für Politikwissenschaft* 19 (4): 537 - 552.
- Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, Heiko Giebler und Bernhard Weßels. 2011a: Demokratiebarometer – ein neues Instrument zur Messung von Demokratiequalität. http://www.democracybarometer.org/publications_de.html (gesehen 12.09.2012).
- Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, Heiko Giebler und Bernhard Weßels. 2011b: Democracy Barometer. Codebook for Blueprint Dataset Version 1. Aarau. http://www.democracybarometer.org/Images/Codebook_only_blueprints_JAN2011.pdf (gesehen 12.09.2012).
- Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, Heiko Giebler und Bernhard Weßels. 2011c: Democracy Barometer. Methodology. Aarau. http://www.democracybarometer.org/Images/Methodical_Explanatory_Note_JAN_2011.pdf (gesehen 12.09.2012).
- Campbell, David F.J. und Thorsten D. Barth. 2009. Wie können Demokratie und Demokratiequalität gemessen werden? Modelle, Demokratie-Indices und Länderbeispiele im globalen Vergleich. *SWS-Rundschau* 49 (2): 208 - 233.
- Castles, Francis G. und Peter Mair. 1984. Left-Right Political Scales. Some 'Expert' Judgements. *European Journal of Political Research* 12 (1): 73-88.
- Diamond, Larry J. 2002. Thinking about Hybrid Regimes, in: *Journal of Democracy*, 13 (2), 21-35.
- Diamond, Larry J. und Leonardo Morlino. 2005. *Assessing the Quality of Democracy*. Baltimore: Johns Hopkins University Press.
- Elklit, Jorgen. 1994. Is the degree of electoral democracy measurable? Experiences from Bulgaria, Kenya, Latvia, Mongolia and Nepal. In *Defining and Measuring Democracy*, Hrsg. David Beetham, 89-111. London/Thousand Oaks/New Delhi: Sage.
- Foweraker, Joe und Roman Krznaric. 2000. Measuring Liberal Democratic Performance: A Conceptual and Empirical Critique. *Political Studies* 48 (4): 759 - 787.
- Fuchs, Dieter. 2004: Konzept und Messung von Demokratie. Eine Replik auf Heidrun Abromeit. *Politische Vierteljahresschrift* 45 (1): 94 - 106.
- Fuchs, Dieter und Edeltraud Roller. 2008. Die Konzeptualisierung der Qualität von Demokratie. Eine kritische Diskussion aktueller Ansätze. In *Bedrohungen der Demokratie*, Hrsg. André Brodocz, Markus Llanque und Gary S. Schaal, 77-96. Wiesbaden: VS-Verlag für Sozialwissenschaften.
- Jäckle, Sebastian und Rafael Bauschke. 2009. Lässt sich Reformfähigkeit messen? Eine kritische Würdigung der Sustainable Governance Indicators. *Zeitschrift für Politikwissenschaft* 19 (3): 359 - 386.
- Jäckle, Sebastian und Rafael Bauschke. 2010. Die Problematik bleibt bestehen. Antwort auf die Replik von Martin Brusis zu den Sustainable Governance Indicators. *Zeitschrift für Politikwissenschaft* 20 (1): 79 - 88.
- Kaina, Viktoria. 2008. Die Messbarkeit von Demokratiequalität als ungelöstes Theorieproblem. *Politische Vierteljahresschrift* 49 (3): 518 - 524.
- Kaufmann, Bruno und M. Dane Waters, (Hrsg.). 2004. *Direct Democracy in Europe. A Comprehensive Reference Guide to the Initiative and Referendum Process in Europe*. Durham, North Carolina: Carolina Academic Press.
- Lauth, Hans-Joachim. 2004. *Demokratie und Demokratiemessung. Eine konzeptionelle Grundlegung für den interkulturellen Vergleich*. Wiesbaden: VS-Verlag für Sozialwissenschaften.
- Lauth, Hans-Joachim. 2008. Demokratieentwicklung und demokratische Qualität. In *Die EU-Staaten im Vergleich: Strukturen, Prozesse, Politikinhalt*, Hrsg. Oscar W. Gabriel und Sabine Kropp, 33-61. Wiesbaden: VS-Verlag für Sozialwissenschaften.
- Lauth, Hans-Joachim. 2010. Möglichkeiten und Grenzen der Demokratiemessung. *Zeitschrift für Staats- und Europawissenschaften* 8 (4): 498-529.
- Lauth, Hans-Joachim. 2011. Qualitative Ansätze der Demokratiemessung *Zeitschrift für Staats- und Europawissenschaften* 9 (1): 49-77.

- Lauth, Hans-Joachim, Gert Pickel und Christian Welzel, Hrsg. 2000. *Demokratiemessung*. Konzepte und Befunde im internationalen Vergleich. Opladen: Westdeutscher Verlag.
- Laver, Michael und W. Ben Hunt. 1992. *Policy and Party Competition*. New York: Routledge.
- Merkel, Wolfgang, Hans-Jürgen Puhle und Aurel Croissant, (Hrsg.). 2003. *Defekte Demokratien. Bd.1: Theorien und Probleme*. Wiesbaden: VS-Verlag für Sozialwissenschaften.
- Merkel, Wolfgang, Hans-Jürgen Puhle, Aurel Croissant und Peter Thiery, (Hrsg.). 2006. *Defekte Demokratie. Bd. 2: Regionalanalysen*. Wiesbaden: VS-Verlag für Sozialwissenschaften.
- Müller, Thomas und Susanne Pickel. 2007. Wie lässt sich Demokratie am besten messen? Zur Konzeptqualität von Demokratie-Indizes. *Politische Vierteljahresschrift* 48 (3): 511 - 539.
- Munck, Gerardo L. und Jay Verkuilen. 2002. Conceptualizing and Measuring Democracy. Evaluating Alternative Indices. *Comparative Political Studies* 35 (1): 5 - 34.
- Przeworsky, Adam und Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley.
- Stoiber, Michael. 2007. Eine neues, kontextualisierten Maß für Demokratie. Konzeptualisierung und Operationalisierung.
https://www.dvpw.de/fileadmin/user_upload/sek_vp/Vortrag%20Stoiber%2011_11_07.pdf
 (gesehen 12.09.2012).
- Stoiber, Michael. 2011. *Die Qualität von Demokratien im Vergleich: Zur Bedeutung des Kontextes in der empirisch vergleichenden Demokratietheorie*. Baden-Baden: Nomos.

Online Anhang: Jäckle, Sebastian/Wagschal, Uwe/Bauschke, Rafael: Das Demokratiebarometer: „basically theory driven“?

Tabelle A 1: Demokratiemaße im Vergleich

Name	Länderzahl	Zeitraum	Autoren	Demokratiedefinition	Operationalisierung	Link
Vanhanen-Index	187	1810-2000	Tatu Vanhanen	Demokratie="[A] political system in which ideologically and socially different groups are legally entitled to compete for political power and in which institutional power holders are elected by the people and are responsible to the people."	Indikator Competition=100-Stimmenanteil der größten Partei; Indikator Participation=Wahlbeteiligte als Bevölkerungsanteil; Demokratie=(Competition x Participation) / 100	http://www.prio.no/CS/CW/Datasets/Governance/Vanhanens-index-of-democracy/
Polity IV Polity Score	164	1800-2010	Monty G. Marshall, Keith Jagers, Ted R. Gurr (Initiator); unterstützt von Political Instability Task Force, Societal-Systems Research und Center for Systemic Peace	"Democracy is conceived as three essential, interdependent elements. One is the presence of institutions and procedures through which citizens can express effective preferences about alternative policies and leaders. Second is the existence of institutionalized constraints on the exercise of power by the executive. Third is the guarantee of civil liberties to all citizens in their daily lives and in acts of political participation. Other aspects of plural democracy, such as the rule of law, systems of checks and balances, freedom of the press, and so on are means to, or specific manifestations of, these general principles."	DEMOC (0 bis +10): Institutionalisierte Demokratie gemessen durch Indikatoren für Wettbewerb bei politischer Partizipation, Offenheit und Wettbewerb bei der Bestellung der Exekutive und Beschränkungen der Spitze der Exekutive; AUTOC (0 bis +10): Autokratiegrad gemessen durch Indikatoren für Wettbewerb bei politischer Partizipation, Regulierung politischer Partizipation, Offenheit und Wettbewerb bei der Bestellung der Exekutive und Beschränkungen der Spitze der Exekutive; POLITY (-10 bis +10): Kombiniertes Wert durch Subtraktion AUTOC von DEMOC	http://www.systemicpeace.org/polity/polity4.htm http://www.systemicpeace.org/inscr/p4manualv2010.pdf
Freedom House Index (Freedom in the World)	195	1972-2011	Freedom House, unter Führung von David Kramer	Freedom="Freedom House measures freedom according to two broad categories: political rights and civil liberties. Political rights enable people to participate freely in the political process through the right to vote, compete for public office and elect representatives who have a decisive impact on public policies and are accountable to the electorate. Civil liberties allow for the freedoms of expression and belief, associational and organizational rights, rule of law, and personal autonomy without interference from the state. Freedom House's definition of freedom is derived in large measure from the Universal Declaration of Human Rights that was adopted by the United Nations General Assembly in 1948. The rights enumerated in the Universal Declaration include freedom of religion, expression, and assembly; freedom from torture; and the right to take part in the government of his or her country. "	Dimensionen Civil Rights und Political Freedoms, beiden gemessen auf einer Skala von 1 (am meisten frei) bis 7 (am wenigsten frei); Die Einstufung basiert auf einer Gesamtpunktzahl bis 100 für jedes Land auf der Grundlage von 10 Fragen zu "political rights" und 15 Fragen zu "civil liberties"; Die Vergabe von Punkten beruht auf "a multilayered process of analysis and evaluation by a team of in-house and consultant regional experts and scholars"; Freiheit wird letztlich verstanden als der Durchschnitt der beiden genannten Dimensionen (freedom rating): Free (1.0 - 2.5), Partly Free (3.0 - 5.0), Not Free (5.5 - 7.0)	http://www.freedomhouse.org/report-types/freedom-world http://freedomhouse.org/report/freedom-world-2012/methodology
Neuer Index der Demokratie (NID)	57 bzw. 27	1996-2002 (bis 2004 für die EU-Länder)	Hans-Joachim Lauth	Das Demokratiekonzept des NID stützt sich auf vorangehende Überlegungen bei Freedom House, Vanhanen, Dahl und Polity. Es beansprucht eine Ergänzung durch die Berücksichtigung von "horizontaler Accountability und Rechtsstaatlichkeit" vorzunehmen.	Variable NID3D: beruht auf Polity: DEMOC (0 bis +10), der Freedom House Political Rights Skala (auf 0 bis 10 skaliert) und dem Weltbank governance indicator Rule of Law"; NID3D = dritte Wurzel des Produktes der drei Komponenten; Variable Staatlichkeit=Indikator "Political Stability" der Weltbank governance indicators; NID=Wurzel aus dem produkt von NID3D und Staatlichkeit	http://www.springerlink.com/content/v450j642357w4681/fulltext.pdf http://www.politikwissenschaft.uni-wuerzburg.de/lehrber-eiche/vergleichende/forschung/neuer_index_der_demokratie_nid/

Sustainable Governance Indicators	31 (OECD)	2009, 2011	Martin Brusis, Jörg Siegmund; Bertelsmann Stiftung	"The SGI's concept of democracy [...] includes not only the rights of political participation and electoral competition, but also the rule of law [...] There are a series of questions designed to address whether citizens face discrimination in the electoral process, how citizens can access public information, the degree to which the media are independent and diversified, how well states protect civil rights, and whether the government and administration act predictably and in accordance with the law."	Dimensionen: Wahlprozess (vier Indikatoren), Informationszugang (drei Indikatoren), Bürgerrechte (drei Indikatoren), Rechtsstaatlichkeit (vier Indikatoren); diese von Experten beurteilen Dimensionen/Indikatoren werden ergänzt durch objektive Variablen für Rechtsstaatlichkeit und Korruption; Alle Indikatoren (teils nach linearer Transformation) messen auf einer Skala von 1 bis 10 und werden gleichgewichtet addiert.	http://www.sgi-network.org/pdf/SGI1_1_Criteria_Methodology_EN.pdf http://www.sgi-network.org/pdf/SGI1_1_Codebook_EN.pdf
Demokratiebarometer	30	1990-2007	Daniel Bochsler, Wolfgang Merkel	"We define democracy as a political system that continuously redefines and alters itself, depending on ongoing political as well as societal deliberation"	Demokratiequalität wird gemessen über die drei Dimensionen Freiheit, Kontrolle und Gleichheit; Jede der Dimensionen beinhaltet drei unterfunktionen, diese wiederum Unterkategorien, welchen schließlich wiederum Indikatoren zugeordnet sind; insgesamt setzt sich das Demokratiequalitätsmaß aus 100 Indikatoren zusammen; Bei der Aggregation werden die empirischen Minima und Maxima von Ausprägungen für die Skalierung der Werte herangezogen; dazu werden bei der Normierung sogenannte blueprint Demokratien betrachtet die Freedom House Werte 1.5 oder niedriger und Polity IV Werte von mindestens 9 zwischen 1995 und 2005 aufweisen; bei der Aggregation wird innerhalb der jeweiligen Ebene/Gruppe von Elementen das gleiche Gewicht angenommen.	http://www.democracymeter.org/concept_de.html http://www.democracymeter.org/Papers/Methodology%201990-2007.pdf
Polyarchie and Contestation Scales	196	1985, 2000	Michael Coppedge, Wolfgang Reinecke	Stützt sich auf den Polyarchiebegriff von Robert Dahl; "Polyarchy is defined as the set of institutional arrangements that permits public opposition and establishes the fight to participate in politics. In these two respects-public contestation and inclusiveness-polyarchy is similar to the concept of democracy. However, polyarchy is not, and was not intended to be, exactly equivalent to democracy" (Coppedge/Reinecke 1990: 51).	Indikatoren: Faire Wahlen, Organisationsfreiheit, Meinungsfreiheit und Medienpluralismus; Kodierung durch Autoren und Studenten gemäß Vorgabe von Codes.	http://www.nd.edu/~m_coppedg/crd/data/crd.htm http://cps.sagepub.com/content/35/1/5.full.pdf+html
Democracy and Dictatorship (DD)	202	1946-2008	José Antonio Cheibub, Jennifer Gandhi, James Raymond Vreeland	"[T]he minimalist conception of democracy we adopt here is procedural in the sense that it sees democracy simply as a method for choosing rulers."	DD: basiert auf 'objektiven Kriterien Austragen von Wahlen, die Existenz von mehr als einer politischen Partei und Veränderung in der Regierungsbesetzung; Alle Merkmale müssen vorliegen, damit es als Demokratie klassifiziert wird; Die Klassifizierung von Ländern ist entsprechend dichotom: Demokratie oder Diktatur.	https://netfiles.uiuc.edu/cheibub/www/DD_page.html https://netfiles.uiuc.edu/cheibub/www/DD_page_files/Cheibub%20Gandhi%20Vreeland%20DD%20Revisited.pdf
Index of Democracy	165	2006, 2008, 2010, 2011	Laza Kekic, The Economist Intelligence Unit	Das Demokratieverständnis stützt sich auf bestehende Messkonzept, der Index des Economist ergänzt die bisherigen Messung der Demokratie jedoch durch stärker substantielle Elemente neben formalen Merkmalen von Demokratie: "Freedom is an essential component of democracy, but not sufficient. In existing measures, the elements of political participation and functioning of government are taken into account only in a marginal way."	Fünf Kategorien: 1. electoral process and pluralism; 2. civil liberties; 3. the functioning of government; 4. political participation; 5. political culture; Insgesamt werden 60 Indikatoren verwendet, dichotom oder mit drei Ausprägungen. Die Indikatoren der einzelnen Kategorien werden gleichgewichtet addiert. Grundlage der Indikatoren sind neben Expertenurteilen auch Daten aus Befragungen (z.B. WVS)	www.economist.com/media/pdf/DEMOCRACY_INDEX_2007_v3.pdf

The Democracy Ranking of the Quality of Democracy	110	2006-2007, 2009-2010	Sándor Hasenöhr, Christa Pözlbauer, Peter Arbenz	Qualität der Demokratie = (Freiheit + andere Merkmale des politischen Systems) + (Leistung der nicht-politischen Dimensionen). Untersucht werden nur diejenigen Länder, die 1) von Freedom House als mindestens "partly free" eingestuft werden, 2) eine Bevölkerung von mindestens einer Million aufweisen und c) eine bestimmte Zahl fehlender Werte für die verwendeten Indikatoren nicht überschritten wird.	Sechs Dimensionen der Qualität der Demokratie: Politik, Wirtschaft, Geschlechtergleichheit, Gesundheit, Umwelt, Wissen; jede Dimension umfasst verschiedene Indikatoren, die nach bestimmten, unterschiedlichen Gewichten aggregiert werden;	http://www.democracyranking.org/en/ranking.htm http://www.democracyranking.org/downloads/Key-findings_Democracy-Ranking_2011_en-A4.pdf
Unified Democracy Scores (UDS)	197	1946-2008	Daniel Pemstein, Stephen Meserve, James Melton	Der Index basiert auf Dahls Polyarchie-/Demokratiekonzept mit den zwei Grundkomponenten Wettbewerb und Partizipation. Es wird jedoch keine theoretische Weiterentwicklung angestrebt, sondern: "In a sense, the focus of this paper is not on measuring democracy, it is on modeling how other researchers rate or judge democracy across polities."	Die UDS-Werte beruhen auf den Daten von zehn Demokratiemessungen: Arat (1991), Bowman, Lehoucq, and Mahoney (2005) (BLM), Bollen (2001), Freedom House (2007), Hadenius (1992), Przeworski et al. (2000) (PACL), Polity scores by Marshall, Jaggers, and Gurr (2006), Polyarchy scale by Coppedge and Reinicke (1991), Gasiorowski's (1996) Political Regime Change measure (PRC), and Vanhanen (2003). Es wird unterstellt, dass allen zehn Messungen/Indikatoren ein latentes Konzept "Demokratie" zugrunde liegt. Diese latente Variable wird geschätzt über die Mittelung der Indikatoren. Dabei werden deren jeweilige Werte als mit stochastischen Fehlern behaftetes Urteil behandelt und die Ungenauigkeit dieser Urteile wird in Form von Varianzen proportional als Gewicht bei der Mittelung herangezogen. Die Modellierung der Demokratiewerte geschieht mittels Bayesianischer Schätzung unter Verwendung des Monte Carlo Verfahrens. Daraus resultiert schließlich eine Verteilung, an der die einzelnen Messkonzepte auf ihre Reliabilität gemessen werden können.	http://pan.oxfordjournals.org/content/early/2010/08/26/pan.mpq020.full.pdf+html http://www.unified-democracy-scores.org/uds.html
Varieties of Democracy (V-Dem)	bisher 12 ¹	1900-2011	Coppedge, Michael and John Gerring, with David Altman, Michael Bernhard, Steven Fish, Allen Hicken, Matthew Kroenig, Staffan I. Lindberg, Kelly McMann, Pamela Paxton, Holli A. Semetko, Svend-Erik Skaaning, Jeffrey Staton, and Jan Teorell	Die Autoren zielen explizit nicht auf eine bestimmte Definition von Demokratie: "Our intention here is to capture various possible conceptions of democracy without making judgments about how they might be combined or how they might contribute to a summary index." Sie gehen davon aus, dass der semantische Möglichkeitsraum des Begriffs Demokratie durch die sechs Elemente Wahlen, Liberalismus, Partizipation, Majoritätspinzip, Deliberation und Egalitarismus weitestgehend abgedeckt wird.	Der Index umfasst sechs Prinzipien der Demokratie: "electoral, liberal, participatory, majoritarian, deliberative, and egalitarian"; 33 Komponenten; 188 Indikatoren. Die Werte beruhen auf den Urteilen von sogenannten "country experts" und "institutional experts". Über deren zusätzliche Angabe des Vertrauens in ihres Urteile werden diese aggregiert und mit einem Wert für die Sicherheit (Grad der Übereinstimmung der Experten) des resultierenden Aggregatwerts versehen.	https://v-dem.net/about http://people.bu.edu/jgerring/documents/MeasuringDemocracy.pdf http://kellogg.nd.edu/projects/vdem/index.shtml

¹ (Albanien, Burma, Ägypten, Ghana, Japan, Mexiko, Russland, Südafrika, Surinam, Schweden, Schweiz und Jemen), zukünftig sollen alle Länder erhoben werden.

Tabelle A2: Veränderungen durch z-Standardisierung im Vergleich zu der vom Demokratiebarometer angewandten Skalierungsmethode

Land	Right to physical integrity [IL_PHIN]					Right to free conduct of life [IL_SELFU]				
	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge
Australia	0,0265362	74,7008469	16	19	3	-0,05929916	64,6774353	18	21	3
Austria	-0,1840744	69,3017336	22	22	0	0,02298686	66,1349492	15	17	2
Belgium	0,06130671	77,2968829	15	18	3	-0,35085831	50,7925055	24	28	4
Canada	-0,20958668	69,0441907	26	26	0	0,31270895	83,0303939	9	6	3
Costa Rica	-0,1268025	73,1529139	21	21	0	-1,19134954	51,2428472	30	27	3
Cyprus	0,18085062	85,7959232	12	8	4	0,3747139	85,4570893	7	5	2
Czech Republic	0,16581414	85,644132	13	9	4	0,11392026	79,1217026	14	11	3
Denmark	0,02099897	79,8247761	17	14	3	-0,01788418	65,4917591	16	19	3
Finland	0,65834769	99,4072794	2	2	0	0,90009755	98,2164269	1	1	0
France	-1,09246032	59,1170009	29	29	0	-0,72528113	40,3976819	27	29	2
Germany	0,31468538	83,1277341	8	11	3	0,5307553	80,9768852	3	8	5
Hungary	-0,1040096	77,4022377	20	16	4	0,51961877	89,5277256	4	3	1
Iceland	0,31932216	83,1745417	7	10	3	-0,03531823	57,6005862	17	23	6
Ireland	-0,20216973	69,1190637	24	24	0	-0,18525467	54,6524545	22	25	3
Italy	-0,1939299	69,2022437	23	23	0	-0,85629949	39,0429656	28	30	2
Japan	0,48416043	93,2344344	4	3	1	0,27992945	82,3858655	11	7	4
Luxembourg	0,4903459	91,3061636	3	6	3	0,21968093	72,5918519	13	15	2
Malta	0,08507462	77,5368167	14	15	1	0,49413016	87,8051169	5	4	1
Netherlands	0,31348312	83,1155974	9	12	3	-0,12534984	55,8303357	21	24	3
New Zealand	0,21943133	83,0532575	10	13	3	0,31687952	79,4114708	8	10	2
Norway	0,70118896	99,8397561	1	1	0	0,22680435	72,7319165	12	14	2
Poland	-0,2033037	69,1076165	25	25	0	-0,3906531	67,9929866	25	16	9
Portugal	-0,02106565	77,3524453	18	17	1	-0,0883502	65,3276559	19	20	1
Slovenia	0,40924579	91,3745692	6	5	1	-0,12493225	73,2177342	20	13	7
South Africa	-1,65518742	48,7920937	30	30	0	-0,50390345	65,7661966	26	18	8
Spain	-0,50477232	63,678404	27	27	0	-0,2633834	61,7256195	23	22	1
Sweden	0,46379331	93,0288311	5	4	1	0,81498136	96,542826	2	2	0
Switzerland	0,19562257	85,945044	11	7	4	0,3005364	74,0212376	10	12	2
United Kingdom	-0,08388634	73,5156987	19	20	1	-0,96509082	51,9837833	29	26	3
United States	-0,52895935	62,5471392	28	28	0	0,45546402	79,6569055	6	9	3

Land	Equality before the law [RL_EQL]					Quality of the legal system [RL_QUAL]				
	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Range	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Range
Australia	0,40454268	75,1687319	10	11	1	0,56063982	77,9924068	6	7	1
Austria	0,58898326	80,6328729	5	6	1	0,88596725	89,2188805	4	3	1
Belgium	0,13990312	71,6166583	14	12	2	-0,25865802	59,1008378	18	17	1
Canada	0,60006469	82,2708944	4	4	0	0,43826556	75,0505209	9	9	0
Costa Rica	-0,37163882	57,2813497	22	21	1	-1,55569195	20,6001731	30	30	0
Cyprus	-0,11366962	62,9389696	20	17	3	-0,41723772	50,0126455	22	26	4
Czech Republic	-0,8312716	43,2125474	27	27	0	-0,84929952	50,4997949	27	24	3
Denmark	0,950721	92,1261443	2	2	0	1,3042281	98,3534941	1	1	0
Finland	0,8874545	90,5726406	3	3	0	0,95521162	86,1275544	2	4	2
France	-0,49502957	48,5506995	25	26	1	-0,32067604	53,8324378	20	21	1
Germany	0,26788391	71,0183354	12	13	1	0,38376767	69,6465944	11	13	2
Hungary	-0,65321951	50,6288323	26	25	1	-0,30751531	62,1343737	19	16	3
Iceland	-0,04481902	58,820123	19	20	1	0,43827779	71,8364184	8	12	4
Ireland	0,19282378	68,4600315	13	16	3	0,11154367	67,9524564	15	15	0
Italy	-1,38332754	27,7291821	30	30	0	-0,67780489	45,7106415	25	28	3
Japan	0,30952508	75,7881985	11	9	2	0,14452935	68,7464384	14	14	0
Luxembourg	0,49207097	79,4408718	8	7	1	0,53471409	81,112522	7	6	1
Malta	-0,32849039	55,2470708	21	23	2	-0,34254952	50,8352844	21	23	2
Netherlands	0,52400679	77,8371161	7	8	1	0,37039257	73,9497207	12	11	1
New Zealand	0,40903921	75,4386908	9	10	1	0,40553945	74,7599124	10	10	0
Norway	-0,0082824	59,7172767	18	19	1	0,90813047	89,3689623	3	2	1
Poland	-1,32270334	28,9336768	29	29	0	-0,9021498	48,5289705	28	27	1
Portugal	0,0664445	69,8319685	15	14	1	-0,48429337	56,408298	24	19	5
Slovenia	-0,85023633	41,8179293	28	28	0	-0,70492349	52,7979568	26	22	4
South Africa	-0,46633681	56,3888732	23	22	1	-0,92486032	42,8847313	29	29	0
Spain	-0,49487898	53,4893096	24	24	0	-0,46471178	50,0700885	23	25	2
Sweden	0,96770362	93,7497763	1	1	0	0,65936225	84,3281592	5	5	0
Switzerland	0,54931738	81,8802192	6	5	1	0,02100535	55,1840014	16	20	4
United Kingdom	0,00862386	62,4495759	16	18	2	-0,20552855	57,0730395	17	18	1
United States	0,00479557	68,8511889	17	15	2	0,29432529	75,4690239	13	8	5

Land	Freedom to associate [PS_FRAS]					Freedom of opinion [PS_FROP]				
	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge
Australia	0,02186778	29,382635	12	18	6	-1,40430614	5,317844	30	30	0
Austria	-0,73520788	14,2338576	29	29	0	-0,60557052	23,0085689	28	28	0
Belgium	0,15571689	39,8829502	11	9	2	0,4645884	52,3877984	7	7	0
Canada	0,70121005	47,5842649	4	5	1	-0,1277638	36,4434205	17	17	0
Costa Rica	-0,57462989	21,3164561	27	28	1	-0,55876659	24,3707834	27	27	0
Cyprus	-0,04556174	34,1103072	16	13	3	1,0016004	61,6127694	1	2	1
Czech Republic	-0,46262111	24,1424893	23	23	0	0,04288595	40,7086474	13	13	0
Denmark	0,23670256	38,6161755	10	10	0	-0,25266739	34,2958775	20	19	1
Finland	0,499304	45,5068141	6	6	0	0,7466254	56,8294365	4	5	1
France	-0,96298873	10,3394697	30	30	0	-0,47982614	26,396361	26	26	0
Germany	-0,2556373	28,1100288	20	21	1	-0,18649228	35,6913502	18	18	0
Hungary	-0,02240178	36,5243638	15	12	3	0,59168401	55,7858262	6	6	0
Iceland	0,3284944	42,281163	8	7	1	0,44452679	50,3247205	8	8	0
Ireland	-0,47183492	22,1729629	24	26	2	-0,3851323	28,1800654	24	25	1
Italy	-0,05923725	32,8284596	17	14	3	-0,27048883	33,8634541	21	21	0
Japan	-0,22294324	28,4387378	19	19	0	-0,40068318	31,0731649	25	22	3
Luxembourg	0,27088003	41,1061017	9	8	1	0,26748596	46,0329373	11	10	1
Malta	-0,31797223	28,131921	21	20	1	0,44249936	45,4232528	9	11	2
Netherlands	0,34957878	38,5132394	7	11	4	-0,07272561	38,2510284	15	16	1
New Zealand	0,91191288	50,3397379	2	3	1	-0,35013254	28,9198222	22	24	2
Norway	-0,01107806	30,5641915	14	17	3	0,73589396	56,9762406	5	4	1
Poland	-0,50965746	22,8390498	26	25	1	-0,12382627	39,7179799	16	15	1
Portugal	-0,49536462	23,8885571	25	24	1	-0,23537616	34,153092	19	20	1
Slovenia	-0,10451165	32,1040171	18	16	2	0,29797758	46,051876	10	9	1
South Africa	-0,00661007	32,40942	13	15	2	0,0449043	41,0458102	12	12	0
Spain	-0,58239816	21,5061394	28	27	1	-0,37196839	29,1103521	23	23	0
Sweden	0,69798958	49,7767588	5	4	1	0,00533602	40,3678696	14	14	0
Switzerland	1,15309337	57,2145799	1	1	0	0,87002586	61,8663561	2	1	1
United Kingdom	-0,32122063	25,193542	22	22	0	-0,89761863	18,1879775	29	29	0
United States	0,8351264	51,5374734	3	2	1	0,76731078	60,229027	3	3	0

Land	Vulnerability (Competitiveness of elections) [CO_COMP]					Contestability (Openness of elections) [CO_OPEN]				
	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge
Australia	-0,57168496	40,3694509	25	25	0	-0,3024837	57,0482362	24	23	1
Austria	0,81539767	85,3025028	3	3	0	-0,38236432	54,4815809	26	25	1
Belgium	0,63864714	69,4116532	5	10	5	0,78369885	78,3408091	2	2	0
Canada	-0,25571926	47,9346155	21	22	1	-0,24694309	56,5884037	22	24	2
Costa Rica	-0,09490862	56,2172474	19	18	1	-0,37792842	54,0754123	25	26	1
Cyprus	0,29168609	62,1588157	12	14	2	0,34462133	69,7777995	8	6	2
Czech Republic	0,37524247	71,6964815	10	8	2	-0,53802462	52,7856305	27	27	0
Denmark	1,05068599	90,7921254	1	1	0	0,38458281	69,4248419	6	8	2
Finland	0,50572062	66,6744171	7	12	5	0,56952387	73,2079833	5	5	0
France	-0,68936852	40,137023	26	26	0	0,08127155	67,3590305	13	10	3
Germany	0,50447202	74,5523851	8	5	3	-0,25582138	57,8789707	23	21	2

Hungary	0,13272541	65,707554	15	13	2	-0,66391595	50,7387817	28	28	0
Iceland	-0,03872829	55,3417749	18	19	1	0,24702065	67,0345873	9	11	2
Ireland	-0,74583621	38,3456965	27	27	0	0,17999201	66,530778	11	12	1
Italy	0,24833261	68,7424768	13	11	2	0,64653778	75,9768418	4	3	1
Japan	-0,90998461	36,0196066	29	28	1	-0,19949805	59,5100762	20	19	1
Luxembourg	-0,21968435	52,2759782	20	21	1	0,35970313	69,7468812	7	7	0
Malta	-0,47885163	47,0262405	24	23	1	-0,94942222	40,3096173	29	29	0
Netherlands	1,02374976	90,2621809	2	2	0	0,72668496	75,6956918	3	4	1
New Zealand	0,05926813	61,3746978	16	15	1	0,12049385	65,6509979	12	13	1
Norway	0,656731	82,8365511	4	4	0	0,0493663	63,436615	15	15	0
Poland	0,21095534	60,5773373	14	17	3	-0,11493344	61,9168317	17	16	1
Portugal	-0,39844833	52,3001062	22	20	2	-0,10252049	60,7563164	16	18	2
Slovenia	0,42059618	69,7503015	9	9	0	0,0671141	64,5152673	14	14	0
South Africa	-2,22798996	16,0526537	30	30	0	-0,15530689	57,4711551	19	22	3
Spain	0,04473481	60,5834395	17	16	1	-0,14471268	61,0085755	18	17	1
Sweden	0,34015044	71,8367123	11	7	4	0,20947683	67,5017011	10	9	1
Switzerland	0,59395095	72,5658907	6	6	0	0,92441699	81,1399211	1	1	0
United Kingdom	-0,41285806	45,6547261	23	24	1	-0,2414444	59,104397	21	20	1
United States	-0,86898382	34,5938943	28	29	1	-1,01918537	39,2351878	30	30	0

Land	Executive-legislature relationship [MC_EXLERE]					Additional checks of powers [MC_ACHE]				
	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge
Australia	0,0350701	69,2142022	16	17	1	1,18125055	85,6042865	3	2	1
Austria	-0,10813468	72,2095549	18	16	2	-0,1413879	39,5296734	16	17	1
Belgium	0,19043091	75,3200835	14	13	1	0,06653025	47,1229117	12	10	2
Canada	0,818076	88,5248303	3	3	0	1,05994235	74,8672434	4	5	1
Costa Rica	-0,1330456	62,8100835	19	20	1	-0,44128767	34,9867597	24	23	1
Cyprus	-0,3631142	57,1731602	23	24	1	-0,42949771	35,2825586	23	21	2
Czech Republic	0,55589627	81,209397	7	8	1	0,16380586	52,6635968	10	8	2
Denmark	0,98599814	91,9718998	2	1	1	0,11082511	46,1831627	11	12	1
Finland	0,22927569	75,9800258	13	10	3	-0,23974812	34,485464	20	24	4
France	-0,54830352	57,381679	26	23	3	-0,31779433	35,0197007	22	22	0
Germany	-0,82548955	50,9942236	27	27	0	1,28613515	88,2309078	1	1	0
Hungary	1,05423561	91,6927329	1	2	1	-0,12397087	42,9838156	15	14	1
Iceland	0,0542912	75,1400656	15	14	1	-0,07116801	41,5778919	14	15	1
Ireland	0,50480593	83,0406093	8	7	1	-0,20482441	40,9495093	17	16	1
Italy	0,66736049	86,3258676	6	5	1	0,25107473	54,5786878	7	7	0
Japan	-1,19804393	39,999303	29	29	0	0,21589739	48,5167946	8	9	1
Luxembourg	-0,19404823	65,5416094	21	19	2	-0,74737427	21,769926	27	27	0
Malta	-0,4618641	53,3824905	25	26	1	-0,50555766	33,3706313	25	25	0
Netherlands	0,70496482	87,0783326	4	4	0	-0,85954759	14,6474763	28	28	0
New Zealand	0,69950912	85,0739826	5	6	1	-0,9935578	12,7737556	30	30	0
Norway	-0,34562986	56,7442626	22	25	3	-0,02861403	45,1090913	13	13	0
Poland	0,23682651	72,6298544	12	15	3	-0,23475829	37,1195296	19	19	0
Portugal	0,24958688	75,4734018	11	11	0	-0,29073792	38,7829207	21	18	3
Slovenia	0,29212146	75,3835709	10	12	2	-0,63948001	24,4875278	26	26	0
South Africa	-1,02859044	50,5698705	28	28	0	0,87289382	76,6575528	5	4	1

Spain	0,43544149	78,8582896	9	9	0	0,16532844	46,9056006	9	11	2
Sweden	-0,02922085	65,5662729	17	18	1	-0,21000941	35,2904054	18	20	2
Switzerland	-1,9297105	20,5552395	30	30	0	0,76811866	64,1482675	6	6	0
United Kingdom	-0,40425574	59,8770293	24	22	2	-0,91009347	13,4032282	29	29	0
United States	-0,14443937	60,4694572	20	21	1	1,24760715	84,5410176	2	3	1

Land	Government resources [GC_GORE]					Conditions for efficient implementation [GC_CEIM]				
	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge
Australia	-0,28546781	64,2857072	22	19	3	0,71267556	89,663708	6	4	2
Austria	0,29202091	73,838793	7	7	0	0,96499428	88,3073582	3	5	2
Belgium	-0,5124731	54,3519248	25	24	1	-0,56066175	69,2768015	22	18	4
Canada	0,2914984	69,1588178	8	16	8	0,54970246	87,8625626	9	6	3
Costa Rica	-0,65133618	52,4628703	28	25	3	-0,7504766	52,5731485	25	29	4
Cyprus	0,63159221	79,363249	3	3	0	0,06890692	70,019106	15	17	2
Czech Republic	-0,38569116	54,6865365	23	23	0	-0,67103217	65,9631181	23	22	1
Denmark	0,47694082	76,8527232	4	4	0	0,97927223	93,9004558	2	1	1
Finland	0,19761495	72,0511735	12	11	1	0,51695169	87,4935116	10	7	3
France	-0,45824836	50,8683529	24	27	3	-0,33339277	59,292056	19	26	7
Germany	0,18806054	71,9983938	14	12	2	0,0904153	84,0466692	14	12	2
Hungary	-0,61210363	50,0201204	27	28	1	-0,48544073	65,9353932	20	23	3
Iceland	-0,65768787	48,2905319	29	29	0	0,40362234	86,2272467	13	9	4
Ireland	0,19053047	69,8044393	13	15	2	0,41215282	82,169281	12	13	1
Italy	0,14044377	64,8547173	16	18	2	-0,91733733	56,9729013	27	27	0
Japan	-0,24601628	58,0081862	21	22	1	0,01458632	75,6649639	16	15	1
Luxembourg	0,69724814	80,3088048	2	2	0	0,74519132	90,0203562	4	2	2
Malta	0,77619359	81,7410989	1	1	0	-0,00636091	69,1758488	17	19	2
Netherlands	0,28585061	72,998881	9	8	1	0,59208522	73,759157	7	16	9
New Zealand	0,04993449	71,6074217	18	13	5	0,73088079	89,8778601	5	3	2
Norway	0,22102801	72,7181199	10	9	1	0,46741231	84,8587544	11	11	0
Poland	-0,79372955	46,0124604	30	30	0	-0,99862696	59,4482884	30	25	5
Portugal	-0,59816967	51,0664761	26	26	0	-0,76255342	64,131295	26	24	2
Slovenia	0,21999779	72,6415383	11	10	1	-0,26491612	66,3108172	18	21	3
South Africa	0,16610352	66,7832713	15	17	2	-0,95302672	46,1615905	29	30	1
Spain	0,35923207	75,8977325	5	5	0	-0,9304361	53,9407236	28	28	0
Sweden	0,31114725	74,6989554	6	6	0	0,56815258	85,971725	8	10	2
Switzerland	0,0926756	70,0370027	17	14	3	1,0722117	87,4133472	1	8	7
United Kingdom	-0,1972828	58,444689	20	21	1	-0,5190369	75,9866372	21	14	7
United States	-0,18990671	63,4663518	19	20	1	-0,73591537	66,9910586	24	20	4

Land	No secrecy [TR_NOSEC]					Provisions for transparent political process [TR_PTPP]				
	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge
Australia	0,52558888	64,729245	6	8	2	0,16879323	64,6585329	13	11	2
Austria	-0,25370599	38,320692	23	24	1	-0,18474068	57,6609275	18	16	2
Belgium	0,46266531	69,0296039	7	5	2	0,38214119	65,4637762	10	10	0

Canada	0,59789824	70,2524971	4	4	0	0,18528374	62,2263228	11	12	1
Costa Rica	-0,70889898	31,1197773	27	28	1	-0,95643849	35,5084204	29	29	0
Cyprus	-1,13281537	17,1429459	29	29	0	-0,94902577	34,3902599	28	30	2
Czech Republic	-0,74975373	32,9545144	28	26	2	-0,73107829	40,6627512	25	27	2
Denmark	0,67228988	67,3574631	2	6	4	1,16654686	84,1559653	1	1	0
Finland	0,31235438	56,6618824	9	14	5	0,6079581	70,8196214	6	7	1
France	0,58531874	72,0688154	5	3	2	-0,34088121	52,3242977	21	20	1
Germany	0,05093238	52,2209828	16	16	0	-0,06865898	58,6565812	16	15	1
Hungary	-0,64893552	35,4812112	25	25	0	-0,39330903	51,4687384	22	22	0
Iceland	0,01532113	44,693808	18	19	1	1,04973334	80,6650165	3	3	0
Ireland	0,44444516	64,6641287	8	9	1	0,17737949	61,112106	12	13	1
Italy	0,01609726	59,2418061	17	11	6	-1,03323866	40,6232157	30	28	2
Japan	0,15779413	56,5957582	14	15	1	-0,57930413	43,6700423	24	24	0
Luxembourg	-0,21805867	39,0968043	22	23	1	0,00375275	51,9390116	14	21	7
Malta	-0,68950121	32,7084519	26	27	1	-0,48371544	41,623029	23	25	2
Netherlands	0,15352312	49,8234899	15	17	2	0,50687931	69,4202214	8	8	0
New Zealand	0,94857024	78,8849393	1	1	0	0,84254159	76,9078244	5	5	0
Norway	-0,06710595	44,3300713	20	20	0	0,59364714	71,5779325	7	6	1
Poland	-0,44652742	47,129004	24	18	6	-0,77014975	45,7442704	26	23	3
Portugal	0,22860749	64,0850277	11	10	1	0,39068705	67,1293268	9	9	0
Slovenia	0,17877509	59,04634	13	12	1	-0,1720786	57,2111575	17	17	0
South Africa	-1,3785029	10,9661846	30	30	0	-0,28451357	56,5165616	20	18	2
Spain	0,29683279	64,9778933	10	7	3	-0,8043336	41,3628405	27	26	1
Sweden	-0,06236871	42,6682877	19	21	2	0,89187297	78,1920351	4	4	0
Switzerland	-0,15145864	40,5293101	21	22	1	1,07583908	83,1264426	2	2	0
United Kingdom	0,65011932	72,4755065	3	2	1	-0,24341312	54,2993311	19	19	0
United States	0,21049957	58,0727706	12	13	1	-0,04817651	59,117641	15	14	1

Land	Equality of participation [PAR_EQPA]					Effective participation [PAR_EFPA]				
	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge
Australia	0,79766551	75,9882618	4	3	1	1,28225763	68,7525479	1	1	0
Austria	0,07852858	60,7293867	14	18	4	-0,15565571	44,7428461	17	16	1
Belgium	0,19143193	63,7143621	12	12	0	0,25170593	53,6882309	12	10	2
Canada	0,643977	72,4117542	5	5	0	0,44663928	49,8509122	7	12	5
Costa Rica	0,05391175	61,1677969	17	15	2	-0,70673933	22,3957193	28	30	2
Cyprus	-1,30625418	35,6027318	29	29	0	0,10612987	49,7757734	13	13	0
Czech Republic	-0,01681295	59,3050247	21	21	0	-0,33484859	41,2186057	22	19	3
Denmark	0,34654973	66,6643037	6	6	0	0,32184428	55,5000606	9	8	1
Finland	0,07507661	61,2543988	15	14	1	-0,22288114	43,869912	18	17	1
France	0,02964688	59,9095902	19	20	1	0,27343601	52,940174	11	11	0
Germany	0,22089531	64,0700333	11	10	1	0,44786984	58,491323	6	4	2
Hungary	-0,96800117	41,099751	28	28	0	-0,52235556	36,8478897	25	22	3
Iceland	1,05373005	80,3333075	1	1	0	0,28591389	54,0962087	10	9	1
Ireland	0,3225678	65,7532393	7	8	1	-0,74897031	23,2297188	30	29	1
Italy	0,08005753	61,8727848	13	13	0	-0,29806242	40,6927075	21	21	0
Japan	0,00325692	60,4955528	20	19	1	-0,39641525	32,048075	24	27	3
Luxembourg	-1,31841695	33,3215267	30	30	0	0,47680229	57,3684021	5	7	2

Malta	-0,49859231	50,1507814	25	24	1	-0,24558104	34,2313566	19	24	5
Netherlands New Zealand	0,05145585	60,9561763	18	17	1	0,05737125	49,724238	14	14	0
Norway	0,80002206	75,6915639	3	4	1	0,95262276	62,5847481	2	2	0
Poland	0,32255915	66,181759	8	7	1	0,41905287	57,5958445	8	5	3
Portugal	-0,63042299	47,8840098	26	26	0	-0,7122766	32,8190044	29	26	3
Portugal	-0,82186622	45,2056446	27	27	0	-0,64299967	34,2227324	27	25	2
Slovenia	0,05531022	60,9768649	16	16	0	-0,36209005	41,0329783	23	20	3
South Africa	0,26282175	64,0545995	9	11	2	-0,03071141	36,7914464	15	23	8
Spain	0,2570228	65,2097628	10	9	1	0,50708673	57,5350428	3	6	3
Sweden	0,91066021	77,2998258	2	2	0	0,50137033	59,9786545	4	3	1
Switzerland	-0,45135546	50,0354719	24	25	1	-0,25965123	42,5484332	20	18	2
United Kingdom	-0,23088839	54,9477122	22	22	0	-0,15404937	45,6971981	16	15	1
United States	-0,31453703	53,9123502	23	23	0	-0,53681527	28,2481348	26	28	2

Land	Substantive representation [REP_SR]					Descriptive representation [REP_DR]				
	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge	Z-Standardisierung	DB Standardisierung	Rang Z	Rang DB	Differenz Ränge
Australia	-0,05163104	57,859676	16	11	5	-0,16499029	61,4829805	17	17	0
Austria	0,60534986	68,5116632	6	7	1	0,63436252	80,7121904	8	8	0
Belgium	-0,82028873	42,542592	28	27	1	0,53937211	78,212398	10	11	1
Canada	-0,48799139	45,5632336	25	25	0	-0,09089955	62,9689878	16	16	0
Costa Rica	-0,01598833	58,8143118	14	9	5	0,41917238	75,8596236	12	12	0
Cyprus	-0,01979258	50,7739483	15	20	5	-0,69930148	46,5372907	26	27	1
Czech Republic	0,07459845	54,8740928	12	15	3	-0,72397078	48,2742148	27	26	1
Denmark	0,79485364	70,5722598	2	3	1	0,79037609	84,8625986	7	7	0
Finland	0,25143409	57,6782527	9	12	3	1,02539674	90,7947973	2	2	0
France	-0,98851229	38,8574549	30	29	1	-0,50861267	53,4687976	22	21	1
Germany	-0,69469983	39,3218304	27	28	1	0,49522239	79,696347	11	10	1
Hungary	-0,51792516	50,6477516	26	21	5	-0,51517107	52,7956967	23	23	0
Iceland	1,38333472	85,9218274	1	1	0	0,82884405	86,4310226	5	5	0
Ireland	0,51798916	68,18809	7	8	1	-0,36203003	55,4193995	19	20	1
Italy	-0,15782593	52,6731381	19	18	1	-0,86900393	43,2495503	28	29	1
Japan	-0,35961227	48,8455644	23	23	0	-1,93857193	20,530649	30	30	0
Luxembourg	0,77866938	69,0900978	3	5	2	-0,89868432	44,011796	29	28	1
Malta	0,62295112	73,5416919	4	2	2	-0,19969871	60,3744292	18	18	0
Netherlands New Zealand	0,17202404	55,6450315	11	14	3	0,8115292	85,4144182	6	6	0
Norway	-0,13003706	53,1672964	18	17	1	0,0561532	67,2929985	13	13	0
Norway	-0,33510556	49,9946535	21	22	1	1,02879024	91,1942715	1	1	0
Poland	-0,07184563	51,5109248	17	19	2	-0,43564539	53,4241041	21	22	1
Portugal	-0,36395495	48,0385209	24	24	0	0,00348708	65,3726304	14	14	0
Slovenia	0,60814065	69,7417538	5	4	1	-0,06298102	64,3139615	15	15	0
South Africa	4,0848E-05	53,6948778	13	16	3	0,61697147	80,011208	9	9	0
Spain	-0,16642519	58,2094664	20	10	10	0,9368195	89,3765913	4	3	1
Sweden	0,21222557	56,8602161	10	13	3	0,97242107	88,8620451	3	4	1
Switzerland	0,47271171	68,7822271	8	6	2	-0,67822032	49,8680543	25	24	1
United Kingdom	-0,96602729	38,4062895	29	30	1	-0,368158	55,9079148	20	19	1
United States	-0,34665999	43,6682355	22	26	4	-0,64297857	48,5066591	24	25	1

**Tabelle A3: Clusterlösung in
Bühlmann et al. (2011a, S.27)**

Abweichungen zu der von uns gefundenen
Clusterung sind gelb markiert

Cluster	Enthaltene Laender
1	Iceland Norway Belgium Denmark Netherlands Finland Sweden New Zealand
2	Costa Rica Italy Malta Poland Czech Republik South Africa
3	Luxembourg Switzerland Cyprus
4	Australia Germany France United Kingdom Ireland Portugal Japan Canada Austria Slovenia Spain Hungary USA

Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, Heiko Giebler und
Bernhard Weßels. 2011a: Demokratiebarometer – ein neues
Instrument zur Messung von Demokratiequalität.
http://www.democracybarometer.org/publications_de.html (gesehen
12.09.2012).

Abbildung A1: Von uns gefundene Clusterlösung

