

Lässt sich Reformfähigkeit messen?

Eine kritische Würdigung der Sustainable Governance Indicators*

Kurzfassung

Während in der Good-Governance-Forschung bereits seit längerer Zeit quantitative Indizes und Benchmarkings für die Evaluation von Regierungsleistungen herangezogen werden, stand eine solche vergleichende Messung der Reformfähigkeit politischer Systeme bis dato aus. Mit den von der Bertelsmann Stiftung erstellten Sustainable Governance Indicators ist unlängst ein Messvorschlag erschienen, der einen Vergleich der OECD-Staaten ermöglichen soll. Grundsätzlich ist der Versuch, diese Forschungslücke zu schließen, sehr begrüßenswert. In diesem Beitrag soll nun evaluiert werden, inwiefern die Güte und Leistungsfähigkeit dieser neuen Messanlage den hohen Erwartungen standhält. Hierzu werden die Sustainable Governance Indicators anhand von Kriterien der Indexbildung, wie sie aus der Demokratiemessung bekannt sind, überprüft. Die Analyse offenbart Mängel im Rahmen der Konzeption, der Operationalisierung sowie der Aggregation der Daten. Als besonders problematisch auf allen drei Ebenen erweist sich die Inklusion der Komponente Demokratie in das verwendete Konzept der Reformfähigkeit. Zudem liefert eine statistische Ex-post-Evaluation keinen Hinweis darauf, dass wirklich Reformfähigkeit gemessen wird.

* Wir bedanken uns insbesondere bei Thomas Metz für intensive Diskussionen der Messproblematik sowie bei Paula Krieg für kritisches Korrekturlesen. Daneben gebührt unser Dank den beiden anonymen Gutachtern für deren wertvolle Hinweise und Tipps. Sämtliche verbliebenen Fehler gehen selbstredend auf unser Konto.

Inhalt

1. Einleitung	360
2. Konzeptualisierung: Was bedeuten Reformbedarf und Reformfähigkeit?	361
3. Operationalisierung: Wie misst man Reformbedarf und Reformfähigkeit?	368
a) Der Statusindex	368
(1) Qualität der Demokratie	368
(2) System- und Politikperformanz	370
b) Exkurs – die SGI im statistischen Vergleich mit anderen internationalen Indizes	371
c) Der Managementindex	373
(1) Strategiefähigkeit	373
(2) Demokratische Beteiligungskultur	375
4. Aggregation: Sinnvolle Transformation oder fabrizierte Varianz?	377
5. Statistische Evaluation des Managementindex	380
6. Schluss: Von der Schwierigkeit Reformfähigkeit zu messen	382

1. Einleitung

Seit einigen Jahren finden sich auch innerhalb der Politikwissenschaft vermehrt Benchmarkings unterschiedlichster Art. Insbesondere im Bereich der Good Governance spielen internationale Vergleiche der Regierungstätigkeit eine große Rolle (Kaufmann u. a. 2007), aber auch in der europäischen Governance-Debatte haben Benchmarkings nicht zuletzt seit dem Aufkommen der „Offenen Methode der Koordinierung“ einen prominenten Platz erobert (Benz 2008: 48). So existiert eine Vielzahl von vergleichenden Studien bzw. Messungen, die sich mit der Regierungsleistung in spezifischen Politikfeldern auseinandersetzen. Möchte man jedoch die grundsätzliche Reformfähigkeit oder den jeweiligen Reformbedarf und nicht die tatsächliche Reformtätigkeit untersuchen, sind die vorhandenen Ansätze weniger geeignet: Entweder sind sie thematisch zu spezifisch, um Rückschlüsse auf die grundlegende Reformfähigkeit politischer Systeme zu ermöglichen, wie etwa die PISA Studie (OECD 2006) oder die Gütermarkt- und Arbeitsschutz-Regulierungsindikatoren der OECD (Nicoletti u. a. 2000), oder sie sind zu allgemein gehalten, wie der Index of Economic Freedom der Heritage Foundation (Holmes u. a. 2008) oder diejenigen Ansätze, welche generell Wettbewerbsfähigkeit zu bemessen suchen (IMF

2008; IMD 2008). Diese Forschungslücke soll nun mit den von der Bertelsmann Stiftung neu entwickelten Sustainable Governance Indicators (SGI) geschlossen werden, mit deren Hilfe der Reformbedarf sowie die Reformfähigkeit der OECD-Staaten vergleichbar gemacht werden sollen.

In diesem Beitrag soll zum einen das Konzept der SGI näher erläutert und zum anderen dieses neue Instrument kritisch analysiert werden. Auch wenn bis dato keine Referenzmessungen vorliegen, lassen sich unter Rückgriff auf Messungen verwandter Bereiche, wie etwa der Demokratieforschung, Kriterien bzw. Analyseraster identifizieren, mit deren Hilfe die Güte einer solchen Messanlage bestimmt werden kann. Dieser Artikel folgt in seiner Argumentation primär dem von Munck und Verkuilen (2002) vorgeschlagenen Raster zur Beurteilung von Demokratieindizes. Basierend auf ihren Vorarbeiten werden die SGI zum einen entlang der drei Stufen der Indexbildung analysiert: der Konzeptionalisierung, der (eigentlichen) Messung, sowie der Aggregation (Munck/Verkuilen 2002: 8). Zum anderen werden im Rahmen der einzelnen Abschnitte die von Munck und Verkuilen erarbeiteten Kriterien als Referenz herangezogen, um vor diesem Hintergrund das Vorgehen der SGI kritisch zu diskutieren.¹

Während im Rahmen des zweiten Abschnitts die grundlegende theoretische Konzeption der SGI näher beleuchtet und mögliche Probleme der verwendeten Reformfähigkeitsdefinition diskutiert werden, erfolgt im dritten Abschnitt eine Analyse der Operationalisierung der SGI. In diesem Zusammenhang wird kritisch hinterfragt, inwiefern die gewählten Indikatoren tatsächlich geeignet sind, das entwickelte Konzept adäquat abzudecken, oder ob möglicherweise eine alternative Operationalisierung sinnvoller gewesen wäre. Der vierte Abschnitt widmet sich der Ebene der Aggregation. Im fünften Abschnitt wird das eigentlich innovative Element des Index, der Management-Teil, einer Ex-post-Evaluation unterzogen. Der sechste und letzte Abschnitt fasst die Evaluationsergebnisse nochmals zusammen und endet mit einer kritischen Würdigung des Messkonzepts.

2. Konzeptualisierung: Was bedeuten Reformbedarf und Reformfähigkeit?

Dem Gedankenmuster von Munck und Verkuilen folgend beginnt die Analyse der SGI auf der Ebene der Konzeptualisierung. Sie bildet das Fundament der

1 Hierzu auch Müller/Pickel 2007, deren Vergleich von Demokratiemessungen ebenfalls auf den von Munck und Verkuilen entwickelten Kriterien und logischen Ebenen der Messung beruht.

späteren Messung, da hier festgelegt bzw. definiert wird, was unter dem zu messenden Objekt zu verstehen ist. Grundlegend lassen sich einige Schwierigkeiten identifizieren: Dabei handelt es sich zunächst um die beiden antagonistischen Problematiken einer zu minimalen bzw. zu maximalen Definition. Im Falle der letzteren wird das grundlegende Konzept überfrachtet, sodass sich in der Empirie hierfür keine entsprechenden Fälle mehr identifizieren lassen. Im Gegensatz hierzu wird bei der minimalistischen eine zu sparsame und daher unvollständige Definition gewählt, die keine Differenzierung der zu untersuchenden Fälle mehr ermöglicht (Munck/Verkuilen 2002: 9). Auf der zweiten Stufe der Konzeptualisierung ist es notwendig, die einzelnen Subkomponenten im Hinblick auf Abstraktionsniveau und Zugehörigkeit passgenau an- bzw. den entsprechenden Komponenten zuzuordnen.

Vor dem Hintergrund der von Munck und Verkuilen angesprochenen Probleme bzw. der konzeptionellen Anforderungen an Messanlagen, kann nun die von den Autoren der SGI gewählte Definition näher betrachtet werden. Brusis definiert das Konzept der Reformfähigkeit wie folgt: „Reformfähigkeit bezeichnet die Fähigkeit der Politik, die zur Verbesserung des Status Quo notwendigen Veränderungen („Reformer“) zu identifizieren und umzusetzen“ (Brusis 2008: 93).

Soll nun die Passgenauigkeit der gewählten Definition überprüft werden, bietet sich der Vergleich mit alternativen Konzeptionen an. Einen Anknüpfungspunkt liefert Lauth (2004: 22 ff.), der im Kontext der Demokratiemessung vorschlägt, über den Vergleich verschiedener Demokratiekonzepte und unter Zuhilfenahme des Verfahrens der Abstraktion nach Sartori (1970) den Kerngehalt der Demokratie zu bestimmen. Das sich herausbildende Konzept wäre auf der höchsten Abstraktionsstufe angesiedelt und folglich als eine „universelle Demokratiedefinition“ aufzufassen, wodurch es somit auch eine sinnvolle Grundlage für die Operationalisierung böte. Dieses Vorgehen lässt sich dabei nicht nur zur Identifikation des Kerngehalts eines Konzepts anwenden, sondern es kann auch im Sinne einer externen Validierung des Konstrukts verstanden werden. Lassen sich etwa im Zuge des Vergleichs verschiedener Definitionen in allen Konzepten dieselben Kernideen identifizieren, kann man dies als ein Kriterium für die Passgenauigkeit bzw. Validität des Konzepts auffassen (Lauth 2004: 25 f.; Wagschal 1999: 39 ff.). Ein solcher Vergleich wird im vorliegenden Fall zunächst dadurch erschwert, dass sich in der Literatur keine alternativen Konzepte von „Reformfähigkeit“ auffinden lassen. Daher erfolgt

eine Diskussion der verwendeten Definition unter Berücksichtigung alternativer Begriffe von „Reform“ allgemein.²

Schmidt definiert Reformen etwa als „gezielte, auf längere Frist angelegte Abweichungen der Politik (im Sinne von Policy) vom Status quo eines Bestandes oder einer Veränderungsrichtung“ (2005: 45). Eine weitere Definition findet sich bei Schultze, der Reformen als „bewußt und planvoll herbeigeführte Veränderung im politischen Institutionensystem und/oder dem Felde der Staats-tätigkeit“ (2002: 789) betrachtet. Im Vergleich mit diesen beiden alternativen Definitionsvorschlägen kann das Konzept der SGI als minimalistisch bezeich-net werden. Der direkte Vergleich verdeutlicht, dass zwei der ansonsten zu-meist als zentrale Elemente von Reformen geltenden Aspekte im Rahmen des SGI-Konzepts in den Hintergrund treten: Zum einen ist dies die langfristige Orientiertheit des politischen Handelns, der nach Schmidt eine zentrale Bedeu-tung zukommt, sowie andererseits die Ausrichtung des Handelns an einem be-stimmten Ziel. Als potenziell problematisch ist zudem die normative Aufladung der Definition zu betrachten, entsprechend derer nur eine Verbesserung und nicht jegliche Veränderung des Status quo als Reform zu verstehen ist. Brusis grenzt sich zwar von Ansätzen ab, die Reformfähigkeit danach bemessen, in-wiewfern ein bestimmter, zuvor auf normativer Basis erstellter Katalog an Reformen erfolgreich implementiert wird (Brusis 2009: 73), andererseits stellt die von ihm favorisierte indirekte Ex-post-Messung letztlich auch nichts anderes als ein normativ determiniertes Benchmarking dar. Trotz der genannten Ein-wände lässt sich jedoch von einer nachvollziehbaren und sinnvollen Definition der Reformfähigkeit im Falle der SGI sprechen, da die als wesentlich zu be-trachtenden Elemente des Reformbegriffs im Konzept der SGI enthalten sind.

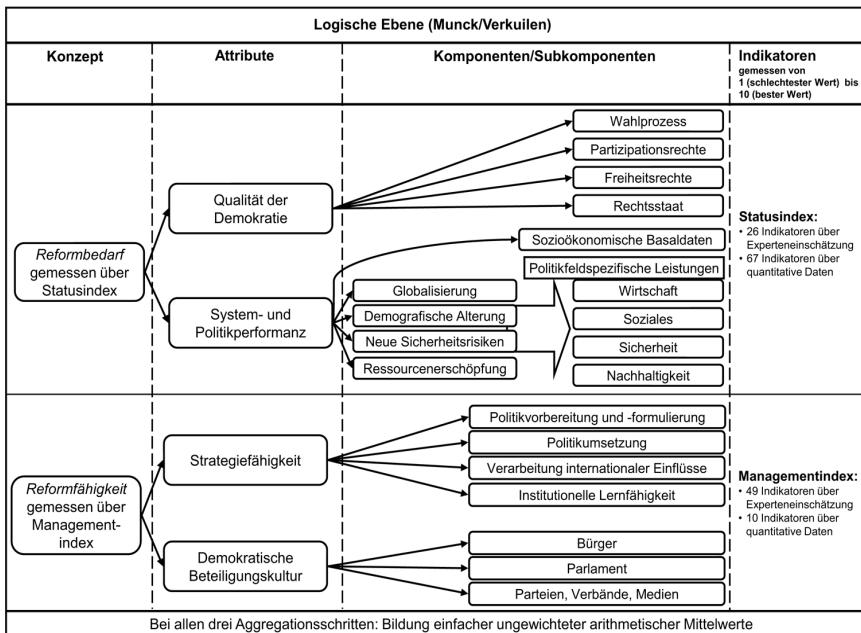
Wendet man sich nun der eigentlichen Konzeptualisierung im Sinne der Ausdifferenzierung des Modells zu, wird zunächst die hohe Komplexität des Messkonzepts deutlich. Munck und Verkuilen betonen im Hinblick auf die Konzeptionalisierung insbesondere die Notwendigkeit der geeigneten vertika-len Anordnung der Komponenten und Subkomponenten gemäß ihrem Abstrak-tionsniveau. Zusätzlich muss darauf geachtet werden, Subkomponenten nicht den falschen Komponenten zuzuordnen, bzw. keine Subkomponenten zu ver-wenden, die bereits in einer anderen Subkomponente enthalten sind (Munck/Verkuilen 2002: 12 ff.).

2 Dieses Vorgehen stellt zweifelsohne ausschließlich eine approximative Lösung dar, je-doch wird erwartet, dass ein sinnvolles Konzept der Reformfähigkeit ebenso zentrale Aspekte des Konzepts der Reform abdeckt.

Grundsätzlich unterscheidet das Konzept der Sustainable Governance Indicators zwei Dimensionen, die sich in Form des Status- und des Managementindex in den SGI wiederfinden. Während der Statusindex den Reformbedarf anhand von Politikergebnissen erfasst, beurteilt der Managementindex die Reformfähigkeit, deren Messung indirekt über eine Beurteilung der konkreten Regierungsführung erfolgt. Der Vergleich von Politikergebnissen ermöglicht somit eine Ex-post-Analyse des Erfolges bzw. weiteren Bedarfs von Reformanstrengungen, wohingegen der Vergleich von Regierungspraktiken Aufschluss über das jeweilige institutionelle Reformpotenzial geben kann (Brusis 2008: 93). Der Statusindex lässt sich nochmals in zwei weitere Subdimensionen untergliedern. Zum einen ist dies die Dimension der Demokratie, die nach der Leistungsfähigkeit der demokratischen Institutionen fragt, zum anderen die Dimension der Systemleistung, in der die ökonomische sowie politikfeldspezifische Performanz der Länder erfasst wird. Der Managementindex wird ebenfalls in zwei weitere Komponenten ausdifferenziert: Erstens die Dimension der strategischen Handlungsfähigkeit, die gewissermaßen die politikstrategische Steuerungskapazität messen möchte, und zweitens die demokratische Beteiligungskultur, im Sinne demokratisch verantwortlichen, transparenten Regierens (Brusis 2009: 79-85). Grundsätzlich stellt die gewählte Aufteilung in Status- und Managementindex eine sinnvolle Unterscheidung dar, die insbesondere im Vergleich mit anderen Messvorschlägen der Reformfähigkeit bzw. -tätigkeit als innovativ zu bezeichnen ist. Ein Großteil dieser Ansätze bezieht in ihre Messung lediglich die Ebene des Status quo ein, während die tatsächliche Aktivität – eben das Management der Reformen – nur selten erfasst wird.

Bei der in Abbildung 1 dargestellten Konstruktion der beiden Indizes entsprechend ihrer jeweils zwei Attribute stellt sich die Frage, inwiefern es sich bei diesen tatsächlich um (theoretisch begründete) konstituierende Merkmale des Reformbedarfs (Statusindex) bzw. der Reformfähigkeit (Managementindex) handelt. Sowohl die Zuordnung der System- und Politikperformanz als auch die der Strategiefähigkeit können in diesem Zusammenhang als unproblematisch erachtet werden. So ergibt sich aus der Zielsetzung des Statusindex fast zwangsläufig, dass Politikergebnisse zentraler Policyfelder bzw. makroökonomische Daten herangezogen werden (Brusis 2009: 76 ff.). Auch die Zuordnung der Strategiefähigkeit zum Managementindex ist plausibel, da Reformfähigkeit sicherlich in hohem Maße davon abhängt, inwiefern politische Gestalter zu langfristiger Planung imstande sind.

*Abbildung 1
Das SGI-Konzept des Reformbedarfs und der Reformfähigkeit*



Quelle: Eigene Darstellung nach (Brusis 2008: 104; 2009: 84, Munck/Verkuilen 2002: 13).

Problematisch ist jedoch die Integration der Komponenten *Qualität der Demokratie* und *demokratische Beteiligungskultur* zu beurteilen. Zum einen betrifft dies die *doppelte* Zuordnung der Demokratie zu Status- und Managementindex, zum anderen die grundsätzliche Sinnhaftigkeit der Aufnahme der Demokratie als konstituierendes Merkmal der Reformfähigkeit. Die *doppelte* Zuordnung der Demokratie über die beiden Attribute sowohl zum Management- als auch zum Statusindex wirkt problematisch, da hierdurch Demokratie einmal als Voraussetzung für Reformfähigkeit und zweitens als konstituierendes Element eines guten Reformstatus und damit als Indikator für einen geringen Reformbedarf gewertet wird. Aus Sicht der Konstruktionslogik der SGI lässt sich diese Kritik jedoch insofern entkräften, als es sich bei Status und Management

um zwei unterschiedliche Dimensionen handelt, wodurch sich auch die separate Erfassung der Demokratiekomponenten rechtfertigen lässt. Da es sich bei der Komponente *demokratische Beteiligungskultur* bei näherer Betrachtung der verwendeten Indikatoren außerdem eher um die Erfassung der Verantwortlichkeit bzw. der Transparenz des Regierens handelt, besteht zudem zwischen den beiden *Demokratie*-Komponenten inhaltlich eine ausreichende Differenz, die die vorgenommene Zuordnung zu den beiden Indizes zusätzlich rechtfertigt (Brusis 2009: 80 f.).

Das zweite und gravierendere Problem besteht in der grundsätzlichen Verknüpfung der Reformfähigkeit mit dem Konzept der Demokratie. Diese ist aus mehreren Gründen kritisch: Erstens liefern die Autoren keine schlüssige theoretische Begründung für die Aufnahme der Demokratie in ihr Konzept. Darüber hinaus wird durch diesen *a priori* gesetzten Zusammenhang eine Analyse der Beziehung zwischen Reformfähigkeit und Demokratie von vornherein verhindert.³ Dieser Zusammenhang – sollte ein solcher überhaupt vorliegen – und vor allem dessen kausale Richtung warten jedoch noch auf eine probate Untersuchung.

Drittens unterstellt das SGI-Konzept implizit, dass adäquate, umfassende Reformen nur in demokratischen Systemen möglich sind, da entsprechend der additiven Konstruktionslogik des Index (hierzu genauer in Abschnitt 4) ein maximaler Managementindex und damit Reformfähigkeitswert nur zu erzielen ist, sofern sowohl Strategiefähigkeit als auch demokratische Beteiligungskultur maximal ausgeprägt sind. Die Autoren des SGI folgen mit dieser Annahme der weit verbreiteten Auffassung, dass Demokratie gleichsam eine Voraussetzung von Reformen darstellt. In diesem Zusammenhang wird argumentiert, dass alleine eine gefestigte Demokratie in der Lage sei, Reformen und die damit einhergehenden Veränderungen und Machtverschiebungen strukturell zu verkraften bzw. auf gesellschaftliche Veränderungen mit Reformen zu reagieren. Im Gegensatz hierzu seien nicht konsolidierte Demokratien nur sehr begrenzt bzw. gar nicht zu Reformen fähig (Schultze 2002: 791). Zwar erscheint diese Argumentation plausibel, doch bezieht sie sich weitgehend auf politische Transformationsprozesse. Es ist in diesem Zusammenhang zumindest fraglich, inwieweit Demokratie als eine generelle Voraussetzung für jegliche Form der Reformfähigkeit betrachtet werden kann. Ferner muss kritisch gefragt werden, ob sich die Erkenntnisse der Transformationsforschung überhaupt auf die

3 Eine ähnlich gelagerte Problematik ist in der Verknüpfung von Demokratie und Marktwirtschaft zu sehen, wie sie im Rahmen der Demokratiemessung kritisch diskutiert wird (Munck/Verkuilen 2002: 9).

Reformfähigkeit und -fähigkeit etablierter Demokratien übertragen lassen. Im Falle der SGI stellt dies darüber hinaus eine klare Abkehr von der in der eigenen Definition vertretenen Auffassung von Reformfähigkeit als eine Fähigkeit eines politischen Systems bzw. der Politik, und eben nicht der Demokratie bzw. demokratischen Politik, dar (Brusis 2008: 93). Selbst wenn die Autoren der SGI im Rahmen ihrer Ausführungen eine theoretische Fundierung des Zusammenhangs zwischen Demokratie und Reformfähigkeit liefern würden, was sie nicht tun, sollte Demokratie trotzdem nicht als konstituierende Komponente von Reformfähigkeit verwendet werden, da der Umkehrschluss „undemokratische Staaten sind tendenziell weniger reformfähig“ nicht automatisch gelten muss. Genau dies wird jedoch implizit von den SGI unterstellt.

Diese theoretische Problematik lässt sich unabhängig davon, ob man Demokratie als behindernde oder ermögliche Komponente der Reformfähigkeit betrachtet, auch empirisch nachweisen. Betrachten wir etwa den Fall Singapur (auch wenn er nicht zum Sample der SGI gehört), so lässt sich eine hohe Reformfähigkeit konstatieren, die offensichtlich auch vor dem Hintergrund eines eher autoritären politischen Systems, welches in der politikwissenschaftlichen Debatte häufig als „guided democracy“ (Puri 2003) bezeichnet wird, realisierbar ist (Low 2001). Wird Demokratie bzw. demokratische Beteiligungskultur, wie im Falle der SGI (Brusis 2009: 81), hingegen als ermögliche Variable verstanden, kann die Schweiz als ein Gegenbeispiel angeführt werden (Castles u. a. 2005). Während die Qualität der Demokratie bzw. das Ausmaß politischer Teilhabe in der Schweiz im OECD-Vergleich als überdurchschnittlich betrachtet werden kann, wird in der Diskussion über die Schweizer Reformfähigkeit bzw. -tätigkeit immer wieder der durch die Diretdemokratie ausgelöste Reformstau angeführt, welcher mit dem durch die Autoren postulierten positiven Zusammenhang zwischen Demokratie und Reformfähigkeit nicht vereinbar ist.

Auch wenn es sich bei diesen beiden Ländern um Ausnahmen handeln sollte, lässt sich doch ein genereller Zusammenhang, wie er von den SGI angenommen wird, nur schwerlich nachweisen. Hält eine solche, für das Gesamtkonzept der Messung wichtige, grundlegende Beziehung der empirischen Überprüfung jedoch nicht stand, muss die Frage erlaubt sein, ob eine Einbeziehung der Demokratie in das Konzept als ein die Reformfähigkeit konstituierender Faktor wirklich angemessen ist. Aus den dargelegten Gründen würden wir im Sinne der theoretisch-konzeptionellen Passgenauigkeit eher dafür plädieren, die Komponente der Demokratie hier außen vor zu lassen, auch um sich selbst nicht der

Möglichkeit zu berauben, diesen (möglichen) Zusammenhang überhaupt erst adäquat untersuchen zu können.

3. Operationalisierung: Wie misst man Reformbedarf und Reformfähigkeit?

Nachdem im vorherigen Abschnitt die Ebene der Konzeptualisierung der SGI eingehend untersucht wurde, wenden wir uns nun der Operationalisierung zu. Munck und Verkuilen identifizieren auf dieser Ebene wiederum Kriterien, die eine Einschätzung der Messanlage im Hinblick auf ihre Qualität ermöglichen. Erstens sollte darauf geachtet werden, sofern möglich, mehrere Indikatoren zu verwenden, um die Gefahr einer unzureichenden Erfassung der Subkomponenten zu reduzieren. Werden mehrere Indikatoren verwendet, muss jedoch beachtet werden, dass ihre Äquivalenz insbesondere im Zuge der Aggregation weiterhin gewährleistet bleibt. Es ist ferner darauf zu achten, möglichst solche Indikatoren zu verwenden, die weniger anfällig sind für Messfehler, wie etwa Verzerrungen durch subjektive Bewertungen der Messenden (Munck/Verkuilen 2002: 15 f.). Dementsprechend sind quantitative Indikatoren eher jenen vorzuziehen, die auf subjektiven Einschätzungen von Experten beruhen. Eine Vielzahl an Fragestellungen lässt sich jedoch häufig gar nicht anders angehen, da Expertenbefragungen die einzige Möglichkeit darstellen, nicht quantifizierbare Parameter zumindest approximativ abzufragen (Laver/Hunt 1992: 34). Sofern deshalb auf das spezielle Wissen von Experten nicht verzichtet werden kann, sollte zumindest versucht werden, eine gewisse Reliabilität der Ergebnisse sicherzustellen.

a) Der Statusindex

(1) Qualität der Demokratie

Bevor die Komponente *Qualität der Demokratie* im Hinblick auf die Güte ihrer Operationalisierung näher betrachtet werden kann, muss zuerst ein grundlegendes Problem der Demokratiemessung angesprochen werden, das sich aus dem Sample der SGI ergibt und weitreichende Implikationen für die Messung hat. Vergleicht man die Ergebnisse bisheriger Demokratiemessungen, zeichnen diese durchweg ein einheitliches Bild für die OECD-Staaten (Stoiber 2008: 211). Primär liegt das daran, dass sie meist ein sehr großes Länder-Sample aufweisen, innerhalb dessen die untersuchten Staaten auf einem Kontinuum

von undemokatisch bis demokratisch zu verorten sind, und es demnach einer vergleichsweise groben Skala zur Messung bedarf. Ein Sample wie die OECD-Staaten verlangt dagegen, sofern man komparativ-differenzierende Aussagen treffen möchte, nach einem weitaus sensitiveren Messinstrumentarium. Demzufolge müssten zur Operationalisierung der Komponente *Qualität der Demokratie* Indikatoren gewählt werden, die von denjenigen der bekannten Demokratiemesskonzepte deutlich abweichen. Geschieht dies nicht, kann keine ausreichende Differenzierung demokratischer Qualität innerhalb der OECD-Staaten erfolgen, wodurch keine oder maximal eine sehr geringe Varianz zwischen den untersuchten Ländern zu erwarten wäre. Wenn sich die Werte der betrachteten Staaten allerdings in einer Komponente nicht voneinander unterscheiden, ergibt auch die Aufnahme dieser Komponente in den Gesamtindex keinen Sinn.

Betrachtet man die gewählten Indikatoren bzw. Teilfragen, wird jedoch deutlich, dass die Autoren die Problematik der mangelnden Varianz innerhalb ihres Samples bei der Operationalisierung nicht berücksichtigt haben. Insgesamt umfasst die Komponente *Qualität der Demokratie* vier Teilkreiche, die jeweils mit zwei bis drei von Experten zu beantwortenden Fragen erfasst werden. Dies sind die Bereiche: *Electoral Process, Access to Information, Civil Rights* und *Rule of Law* (Bertelsmann Stiftung 2009b: 5-11). Die Experten sind dazu aufgefordert, ihre Bewertungen nicht nur in Form einer Punktevergabe (von eins bis zehn) abzugeben, die in der eigentlichen Indexberechnung weiter Verwendung findet, sondern diese numerischen Einschätzungen auch jeweils zu begründen. Daneben fertigt jeder Experte auch einen Länderbericht an, der nochmals alle relevanten Teilaspekte abdecken soll. Die zur Erfassung der vier Teilkomponenten verwendeten Fragen geben wenig Anlass zu Kritik und scheinen das zugrunde liegende Konzept der *embedded democracy* ausreichend im Sinne der Autoren abzudecken (Brusis 2009: 76). Lediglich im Rahmen der Teilkomponente *Rule of Law* ließe sich diskutieren, inwiefern die Frage zur politischen Korruption tatsächlich im Rahmen dieses Konzepts notwendig ist – d. h. ob *Rule of Law* und Korruption als *interdependent* oder *independent* zu betrachten sind (Lauth 2004: 201; Kaufmann u. a. 2007: 4).

Betrachtet man jedoch die verwendeten Fragen genauer, wird das zentrale Problem der Operationalisierung der ersten Subkomponenten deutlich. So wurden zur Erfassung des Teilkreis *Electoral Process* Fragen konzipiert, die im Wesentlichen nicht über den Inhalt dessen hinausgehen, was in bisherigen Demokratiemessungen Verwendung findet. So lautet die Frage zur Fairness des

Wahlkampfes: „How fair are procedures for registering candidates and parties?“ (Bertelsmann Stiftung 2009b: 5). Vergleicht man diesen Indikator mit der entsprechenden Frage von Freedom House: „Is the registration of voters and candidates conducted in an accurate, timely, transparent, and nondiscriminatory manner?“ (Freedom House 2008), erweisen sich die beiden als weitestgehend deckungsgleich.

Auch für die übrigen Teilbereiche gilt das geschilderte Problem in ähnlicher Weise. Somit erscheint es zumindest zweifelhaft, inwiefern eine Operationalisierung, die sich größtenteils an bisherigen Demokratiemessungen orientiert, einen höheren Differenzierungsgrad unter den etablierten Demokratien erreichen kann. Die damit zu erwartende mangelnde Varianz der demokratischen Qualität spricht demnach ebenfalls eher dafür, die Komponente der Demokratie gar nicht erst in das Messkonzept einzubeziehen.

(2) System- und Politikperformanz

Die Operationalisierung der zweiten Komponente erfolgt im Gegensatz zur ersten bis auf wenige Ausnahmen nicht über Expertenbefragungen, sondern anhand ökonomischer Daten. Dieses Vorgehen hat im Sinne von Munck und Verkuilen den prinzipiellen Vorteil, die Wahrscheinlichkeit eines subjektiven Bias zu minimieren. Die gewählten Indikatoren sind letztlich nicht zu beanstanden, da es sich hierbei um vielfach erprobte und verwendete Größen handelt, die bereits im Rahmen ähnlicher Untersuchungen verwendet wurden.⁴ Doch genau aus diesem Grund erscheint eine erneute Rekombination dieser einzelnen Variablen im Rahmen der SGI nur bedingt sinnvoll. Vielmehr könnte beispielsweise im Sinne breiterer Analysemöglichkeiten, wie etwa der eines intertemporalen Vergleichs, auch auf bereits vorhandene Indikatorensets zurückgegriffen werden. Eine Option zur Auswahl potenziell hierfür geeigneter dritter Indikatoren bietet die Berechnung von Korrelationsmaßen zwischen den SGI und alternativen Messansätzen.

4 Im Gegensatz zu den im vorigen Abschnitt erwähnten Demokratiemaßen sind diese Indikatoren auch in der Lage, eine ausreichende Varianz zwischen den OECD-Staaten zu begründen.

**b) Exkurs – die SGI im statistischen Vergleich mit anderen internationa-
len Indizes**

Auch wenn die Macher der SGI selbst ihren Index als grundlegend verschieden von den bisherigen Indizes erachten, da diese die beiden für die SGI relevanten Dimensionen Reformstatus und Reformmanagement weder systematisch noch mit demselben Nachdruck messen (Wagschal/Jäckle 2009: 103), sollen einige dieser Maße im Folgenden trotzdem als Vergleichsmaßstab herangezogen werden. Es handelt sich dabei um:

- die Governance-Indikatoren (GI) der Weltbank,
- den Index of Economic Freedom (IEF) der Heritage Foundation,
- den Global Competitiveness Index (GCI) des Weltwirtschaftsforums (Genf),
- das World Competitiveness Yearbook (WCY) des International Institute for Management Development (Lausanne)
- sowie den Gütermarkt-Regulierungsindex der OECD.

Auch wenn keiner dieser Indizes den Fokus in gleicher Weise wie die SGI auf das Thema Reformen legt, so beschäftigen sie sich doch mit ähnlichen Sachverhalten, weisen demzufolge auch alle einen gewissen Reformaspekt auf, was sich trotz der Unterschiede in ihrer Konzeption sowie ihrer Methodik auch aus den berechneten Zusammenhangsmaßen herauslesen lässt.

So weist der Statusindex der SGI bei bivariater Betrachtung starke Korrelationen zu bereits vorhandenen Indizes auf. Besonders zu einzelnen GI der Weltbank (*Voice and Accountability*, *Government Effectiveness* und *Rule of Law*), sowie zu einigen der Dimensionen des IEF der Heritage Foundation (*Monetary Freedom* und *Property Rights*) bestehen besonders starke Zusammenhänge (vgl. Tabelle 1). Aber auch zu den beiden die Wettbewerbsfähigkeit eines Landes messenden Indizes, den GCI des Weltwirtschaftsforums sowie dem WCY des International Institute for Management Development, weist er beachtliche Korrelationen auf. Was von den Machern der SGI als Kreuzvalidierung ihrer Ergebnisse und damit letztlich als Gütemaß angesehen wird, kann aus einem anderen Blickwinkel betrachtet auch bedeuten, dass die Messung, die die Bertelsmann Stiftung mit dem Statusindex durchführt, letztlich unnötig ist, da genau dieselben Informationen auch über bereits vorhandene Maße erhoben werden könnten.

*Tabelle 1
Zusammenhang der SGI mit anderen Indikatoren*

Indikator	Bedeutung/was misst der Index?	Indikator		
		SGI Statusindex	SGI Statusindex (ohne Demokratie)	SGI-Managementindex
GI: Voice and Accountability (2006)	Politische und soziale Teilhabemöglichkeiten	0,875***	0,851***	0,475***
GI: Government Effectiveness (2006)	Effektivität von Regierungshandeln	0,921***	0,952***	0,577***
IEF: Gesamtscore (2006)	Ökonomische Freiheiten	0,685***	0,741***	0,637***
IEF: Monetary Freedom (2006)	Ausmaß an Preisstabilität	0,755***	0,756***	0,381***
IEF: Property Rights (2006)	Schutz von Privateigentum durch das Gesetz	0,887***	0,890***	0,563***
GCI (2006)	Wettbewerbsfähigkeit eines Landes	0,805***	0,799***	0,494***
WCY (2006)	Wettbewerbsfähigkeit eines Landes	0,772***	0,818***	0,540***
RKI (2006)	Kompositindex aus Governance-Indikatoren; misst den Reformbedarf	0,935***	0,956***	0,581***

Quelle: Dargestellt sind Spearman-Rangkorrelationen; SGI = Sustainable Governance Indicators; GI = Governance-Indikatoren der Weltbank; IEF = Index of Economic Freedom der Heritage Foundation; GCI = Global Competitiveness Index des Weltwirtschaftsforums; WCY = World Competitiveness Yearbook des IMD; RKI = Reformkompositindex; n = 30 (alle SGI-Länder); * = signifikant auf dem 0,1 Niveau; ** = signifikant auf dem 0,05 Niveau; *** = signifikant auf 0,01 Niveau (jeweils zweiseitiger Test).

Aus diesem Grund plädieren wir dafür, den Status des Reformbedarfs nicht extra über das relativ aufwendige und damit auch kostenträchtige Instrumentarium der SGI zu erheben, sondern stattdessen einen *Reformbedarf-Kompositindex* (RKI) aus bereits bestehenden Maßzahlen zu verwenden. Dieser kann sehr viel einfacher ohne die Notwendigkeit einer neuen Messung berechnet werden.

Entsprechend theoretischer Plausibilität und den hohen Korrelationen ist es hierfür ausreichend, zwei der Governance-Indikatoren – *Government Effectiveness* und *Rule of Law* –, reskaliert auf eine Skala von 0 bis 100, heranzuziehen:⁵

$$\text{RKI} = \frac{\text{Government Effectiveness} + \text{Rule of Law}}{2}$$

Der RKI korreliert mit Spearman Rho = 0,935 mit dem SGI-Statusindex. Damit ist die Ähnlichkeit bereits deutlich ausgeprägt. Klammt man des Weiteren den zuvor bereits als theoretisch unpassend identifizierten Demokratie-Teil, d. h. die Unterpunkte *Electoral Process*, *Access to Information* und *Civil Rights*, aus dem Statusindex aus, ist der Zusammenhang sogar noch stärker (Rho = 0,956). Hier lässt sich nun argumentieren, dass die Ähnlichkeit der beiden Maße so groß ist, dass eine eigene Erhebung, teilweise unter Mitwirkung von Expertenbefragungen, wie sie die SGI durchführen, gerade auch unter dem Aspekt der Kosten-Nutzen-Relation, überflüssig erscheint. Außerdem bestünde mit dem RKI auch die Möglichkeit, Reformbedarfswerte für die Vergangenheit zu erhalten, die als Basis für eine Ex-post-Evaluation des die Reformfähigkeit messenden Managementindex herangezogen werden können. Eine solche erfolgt im fünften Abschnitt dieses Beitrags.

c) Der Managementindex

(1) Strategiefähigkeit

Im Rahmen der ersten Komponente des Managementindex werden wiederum Fragebögen bzw. Expertenbefragungen zur Datengenerierung verwendet und mit bereits vorhandenen Datenquellen verknüpft (Brusis 2009: 80). Die Operationalisierung erfolgt entlang von vier Subkomponenten: *Policy Preparation*, *Policy Implementation*, *Incorporation of External Reform Impulses* sowie

5 Nach Kaufmann, Kraay und Mastruzzi lässt sich *Government Effectiveness* als die zentrale Komponente der Handlungsfähigkeit des politischen Systems und *Rule of Law* als notwendige Voraussetzung für die Implementation politischen Regelwerks verstehen. Damit stellen sie unserer Meinung nach den Kern dessen dar, was ein Reformbedarfssindex enthalten sollte. Anstelle der hier gewählten einfachen, nur aus zwei Komponenten der GI bestehenden Konzeption wurden verschiedene weitere Kompositindikatoren aus bereits bestehenden (Teil)Indizes getestet, die allerdings allesamt keine signifikant besseren Ergebnisse lieferten. Komplexere Indexbildungsverfahren via Faktorenanalyse wurden aus Gründen der Nachvollziehbarkeit bewusst nicht angewandt.

Institutional Learning. Insbesondere im Vergleich zur ersten Komponente *Qualität der Demokratie* fällt der hohe Detailgrad der gewählten Fragen auf. Dies ist nicht zuletzt der hohen Komplexität des Untersuchungsgegenstandes geschuldet. Insofern ist die genaue und detaillierte Fragestellung als sinnvolle Möglichkeit anzusehen, das Risiko von Fehlinterpretation der Fragen durch die Experten zu reduzieren. Gleichzeitig erscheint jedoch zweifelhaft, inwiefern die ausgewählten Fragen wirklich kompetent von Länderexperten beantwortet werden können, da es sich bei einem Großteil um administrative Interna handelt, die wohl kaum von *Außenstehenden* zu beantworten sind. So ist etwa die Frage nach der Häufigkeit der Treffen von Ministern und strategischen Planungseinheiten selbst von absoluten Länderexperten bestenfalls approximativ abzuschätzen (Bertelsmann Stiftung 2009b: 30). Auch die Frage nach dem Einfluss einer solchen Einheit für strategische Planung lässt sich schwer eindeutig beantworten bzw. öffnet Raum für partiell willkürliche Antworten, insbesondere wenn hierfür eine Einschätzung mittels Punkteskala, wie es die SGI-Fragen grundsätzlich vorsehen, erfolgen soll.

Allgemein zeigt sich hier ein bei Expertenbefragungen häufig auftretendes Dilemma: Die Befragung von Länderexperten ist einerseits sinnvoll, um komplexe und schwierige Fragen zu beantworten. Andererseits ist der Pool an qualifizierten Experten für einige Länder doch äußerst limitiert und auch wenn es theoretisch eine größere Menge an Experten gäbe, so setzen finanzielle Restriktionen der Untersuchung doch allzu häufig relativ enge Grenzen, was die Größe des Befragungssamples anbelangt. Expertenbefragungen, die auf relativ kleinen Samples aufbauen, weisen allerdings meist eine hohe Ausreißerproblematik sowie Reliabilitätsprobleme auf. Auf diese mithilfe statistischer Methoden zu kontrollieren, wäre jedoch erst ab einer gewissen Mindestanzahl an Befragungen möglich. Für die SGI geben pro Land jeweils drei Länderexperten ihre Einschätzungen ab. Das bedeutet zwar eine Steigerung im Vergleich zum von der Bertelsmann Stiftung schon seit mehreren Jahren erhobenen Transformations-Index (BTI), der nur auf den Bewertungen von je zwei Experten pro Land fußt, statistische Reliabilitätstests sind damit aber trotzdem noch nicht möglich. Die SGI versuchen deshalb, über semistandardisierte Fragebögen sowie ein mehrstufiges Reviewverfahren der Ländergutachten die Reliabilität zu verbessern. Gerade aber das Reviewverfahren gibt Anlass zur Kritik. So ist nicht nachvollziehbar, weshalb auf dessen erster Stufe der Regionalkoordinator, der die drei einzelnen Ländergutachten zu einem Synthesebericht pro Land bündelt, aus den Punktewerten, die von den Experten vergeben werden, den seiner Meinung nach passenden auswählt. Damit nehmen sich die SGI die,

wenn auch durch die geringe Anzahl an Länderexperten nur schwach ausgeprägte, Möglichkeit über Mittelwertbildungen der Reliabilitätsproblematik Herr zu werden. Auf der zweiten Stufe des Reviewprozesses kann zudem ein wissenschaftlich prominent besetzter Beirat, das SGI-Advisory-Board, sämtliche Indexwerte nochmals hinterfragen und gegebenenfalls korrigieren. In gewisser Weise eröffnet dieses als Kontrollinstanz gedachte Gremium damit jedoch auch die Möglichkeit zur Manipulation.

(2) Demokratische Beteiligungskultur

Die zweite Komponente des Managementindex wird ähnlich der vorangegangenen teilweise über Fragen, teilweise über quantitative Daten, etwa die Anzahl der Mitglieder der Legislative, erhoben. Auch zur Erhebung der *demokratischen Beteiligungskultur* werden vergleichsweise detaillierte Fragen gestellt. Somit gelten die bereits vorgebrachten Einwände auch hier entsprechend. Zudem bestehen Zweifel an der Eignung einzelner Indikatoren. Ziel der Autoren ist es, durch diese Komponente zu erfassen, inwiefern externe Akteure einen Einfluss auf bzw. Kontrolle über die Regierungsverantwortlichen ausüben (Brusis 2009: 79). Gleichzeitig sollen außerdem die Wechselbeziehungen zwischen Exekutive auf der einen und zivilgesellschaftlichen und politischen Akteuren (Parlament) auf der anderen Seite, sowie der Beitrag zur Verbesserung der Regierungstätigkeit dieser Akteure erfasst werden (Brusis 2009: 80 f.). Somit wird im Rahmen dieser Komponente nicht nur die Verantwortbarkeit der Exekutive, sondern darüber hinaus auch der Einfluss der genannten Akteure untersucht. An dieser Stelle muss allerdings gefragt werden, inwiefern das gewählte Konzept den Begriff der *Accountability* nicht unnötig überfrachtet, insbesondere wenn die von den SGI-Autoren selbst angeführte Konzeption von Schmitter als Vergleichsmaßstab angelegt wird, welcher *Accountability* primär im Sinne politischer Kontrolle und damit deutlich begrenzter begreift (Schmitter 2004: 47). Die von den SGI gewählte Konzeption ist demzufolge zumindest als vergleichsweise breit zu charakterisieren. Betrachtet man nun vor diesem Hintergrund die gewählte Operationalisierung, wird jedoch deutlich, dass diese breite Konzeption nur bedingt durch die Fragen der SGI abgedeckt wird.

Positiv lassen sich hier die Fragen des ersten Fragenblocks bewerten, welcher sich mit der Funktion der Bürger beschäftigt. Die hier gestellten Fragen zur öffentlichen Wahrnehmung der Regierungstätigkeit (*Policy Knowledge*), zum Vorhandensein von Konsultationsmechanismen sowie der politisch-kultu-

rellen Orientierung der Bürger (Bertelsmann Stiftung 2009b: 49), können im Sinne des SGI-Konzepts als durchaus angebracht betrachtet werden.

Problematischer gestaltet sich der zweite Fragenblock, der sich dem Einfluss des Parlaments auf die Exekutive widmet. Kritisch lässt sich in diesem Zusammenhang bemerken, dass zwar die Anzahl an Mitgliedern der Legislative abgefragt wird, diese jedoch ausschließlich für deskriptive Zwecke sowie scheinbar als Referenzgröße für die Größe des wissenschaftlichen Mitarbeiterstabes in den Index eingeht.⁶ Dabei legen aktuelle Forschungsergebnisse zur optimalen Anzahl an Abgeordneten nahe, diesen Parameter auch direkt in die Subkomponente der parlamentarischen Informations- und Kontrollressourcen einfließen zu lassen. Die Forschung hierzu geht bislang meist davon aus, dass sich über die Bevölkerungszahl sowie einige weitere Faktoren ein Optimum der Abgeordnetenanzahl bestimmen lässt (Auriol/Gary-Bobo 2007). Demzufolge würde es Sinn ergeben, Abweichungen von diesem Optimum negativ im Sinne mangelnder Kontrollmöglichkeiten des Parlaments in den Index einfließen zu lassen. Dadurch würden einerseits Staaten mit übergroßen Parlamenten, in denen sich die einzelnen Abgeordneten stark gegenseitig in ihrer Arbeit behindern, wie dies zum Beispiel in Italien oder Frankreich der Fall ist, genauso negativ bewertet wie Staaten, die über eine für ihre Verhältnisse zu kleine erste Kammer verfügen (USA) und in denen die Abgeordneten daher eher durch die Arbeitslast überfordert sind. Im Fall der Ausschüsse argumentieren die SGI auch sinnvollerweise entsprechend eben dieser Logik, indem sie annehmen, dass 13-25 Mitglieder das Optimum für den politischen Output darstellen (Bertelsmann Stiftung 2009b: 50).

Eine weitere diskussionswürdige Frage ist die nach der Zeit, welche die Debatte über den Haushalt in Anspruch nimmt. Die SGI werten eine möglichst lange Dauer von der ersten Präsentation bis zum Beginn des Haushaltsjahres positiv, da das Parlament hierdurch seine Kontrollkompetenzen gegenüber der Exekutive länger und damit exzessiver wahrnehmen könne. Grundsätzlich lässt sich einerseits fragen, ob ein langer Gesetzgebungsprozess allgemein wirklich als ein Zeichen hoher Reformfähigkeit gelten solle, und andererseits, inwiefern gerade die Dauer der Haushaltsberatungen, die bekanntlich in vielen Ländern besonderen institutionellen Arrangements unterliegt, als Proxymaß für parlamentarische Kontrolle geeignet ist. Hier wäre es eventuell sinnvoller,

6 Letzteres lässt sich jedoch weder aus dem Codebook (Bertelsmann Stiftung 2009b), den Artikeln von Brusis (2008, 2009) noch der SGI-Website (Bertelsmann Stiftung 2009a) eindeutig herauslesen.

die durchschnittliche Dauer zu verwenden, die ein Gesetz bis zur Verabschiebung benötigt.⁷

4. Aggregation: Sinnvolle Transformation oder fabrizierte Varianz?

Aufbauend auf den Ausführungen zur Operationalisierung der einzelnen Komponenten der SGI soll nun mit der Aggregation ein weiterer, eher technischer Teilaспект angesprochen werden, welcher die dritte Ebene des Munck und Verkuilen'schen *concept trees* repräsentiert. Aus den unterschiedlichen Indexkomponenten wird in den meisten politikwissenschaftlichen und ökonomischen Indizes ein einzelner aggregierter Wert berechnet. Auch wenn es bei den SGI zwei sind, bleiben die potenziellen Probleme dieselben. Es sind dies die der Gewichtung sowie der Aggregationsregel, die das Ergebnis – also beispielsweise die Reihenfolge in einem Länderranking – beeinflussen können. Zu beachten ist, dass bei der Aggregation von einer niedrigen zu einer höheren Ebene kein systematischer Informationsverlust und auch keine Verzerrungen stattfinden sollten. Schließlich sollten auch hier die zugrunde liegenden Berechnungsregeln offengelegt werden, um damit die Aggregation replizierbar zu machen (Munck/Verkuilen 2002: 8). Im Hinblick auf die Transparenz der für die SGI verwendeten Aggregationsregel besteht keinerlei Grund zur Kritik, da sie einer einfachen additiven Grundlogik folgt und vor allem über die Internetpräsenz der SGI problemlos nachzuvollziehen ist. Die Aggregation der einzelnen Subkomponenten erfolgt ungewichtet und über die Bildung arithmetischer Mittelwerte, wobei jeweils eine unterschiedliche Anzahl an Subkomponenten, die entweder quantitativ oder qualitativ über Expertenbefragungen erhoben werden, zur nächsthöheren Ebene aggregiert werden. Auch auf den weiteren Ebenen bis hin zum letztlichen Status- und Managementindex bleiben die SGI dieser Aggregationsregel treu (Brusis 2008: 107). Die Aggregation zum Status- und zum Managementindex folgt somit zumindest vordergründig dem Prinzip einer möglichst einfach gehaltenen Indexbildung, durch die eine gute Nachvollziehbarkeit erreicht werden soll. Die praktische Vorgehensweise birgt jedoch wiederum Grund zu Kritik. Zunächst werden die quantitativen Sekundärdaten auf eine gemeinsame Skala gebracht, um vergleichbare Werte für die weitere Aggregation zu erhalten. Dies geschieht, indem sie auf eine Punkteskala von eins bis zehn transformiert werden, wobei der Staat mit dem besten empirischen Wert zehn Punkte erhält, der mit dem schlechtesten einen Punkt. Den

7 Vergleiche hierzu für Deutschland die Untersuchungen von König/Bräuninger (2005: 61-70) sowie Manow/Burkhart 2009.

übrigen 28 OECD-Ländern werden proportional Punktwerte zwischen diesen beiden Extremen zugewiesen. Damit schöpfen alle quantitativen Indikatoren per definitionem die Punkteskala vollständig aus.

Die qualitativen Daten, d. h. die Einschätzungen der Experten, sind hingegen bereits auf einer Zehnerskala angeordnet und könnten somit ohne weitere Transformation gemeinsam mit den standardisierten quantitativen Werten zu den Gesamtindizes aggregiert werden. Hier zeigt sich nun aber das bereits theoretisch prognostizierte Problem einer zu geringen Varianz unter den Expertenscores einiger Teilkomponenten. So bewegen sich insbesondere die Werte im Bereich der Demokratiequalität häufig maximal zwischen fünf oder sechs (Türkei) und zehn Punkten (für eine größere Gruppe an Ländern, darunter vor allem die skandinavischen) und schöpfen damit die maximal mögliche Skala nur äußerst unzureichend aus. Der Problematik der mangelnden Varianz begegnen die SGI dabei auf äußerst pragmatische Art und Weise. Analog zum Verfahren bei den quantitativen Sekundärdaten werden auch hier die maximale Expertenbewertung gleich zehn und die minimale gleich einem Punkt gesetzt. Alle übrigen Bewertungen werden proportional zwischen diesen Extremen verteilt. Durch diese *Aufspreizung* wird eine in den Originaldaten nicht vorhandene Varianz künstlich erzeugt. So erhält beispielsweise die Türkei, auch wenn sie von den Experten bei den Fragen zur Demokratie durchweg im mittleren Bereich der ursprünglichen Skala eingeordnet wird, letztlich meist jeweils nur einen Punkt, da sie im Vergleich zu den anderen Staaten etwas schlechtere Einzelwerte aufweist. Damit werden auch sämtliche Demokratisierungsanstrennungen in einem solchen Land, das von einem deutlich niedrigeren Demokratieniveau aus startet, in keiner Weise honoriert, da sich selbst eine deutliche Verbesserung in den Expertenbewertungen für den Letztplatzierten nicht in einer Steigerung des transformierten Wertes widerspiegeln würde – dieser bliebe für den Letzten stets beim Minimalwert von eins. Auch wenn auf den ersten Blick dieses Vorgehen die Varianzproblematik beseitigt, führt es doch zu ungewollten Nebeneffekten, die gegen eine solche artifizielle Varianzerzeugung sprechen. Einerseits lassen sich so Fälle konstruieren, in denen die Transitivität des Gesamtindikators verletzt wird, d. h. im ungünstigsten Fall kann es passieren, dass aus einer vollständig transitiven Ordnung der ursprünglichen Expertenbewertungen keinerlei Rangfolge der transformierten Werte mehr herausgelesen werden kann. Tabelle 2 verdeutlicht diese Situation:

*Tabelle 2
Beispiel zur Verdeutlichung der Transformationsproblematik*

	Ursprüngliche Expertenwerte			Transformierte Werte		
	Land A	Land B	Land C	Land A	Land B	Land C
Frage X	1	2	3	1	5,5	10
Frage Y	8	5	2	10	5,5	1
Mittelwert	4,5	3,5	2,5	5,5	5,5	5,5

Quelle: Eigene Darstellung. Die einzelnen Werte der linken Tabellenhälfte stellen hypothetische Werte für die ursprünglichen Punktzuweisungen der Experten für zwei Fragen dar. In der rechten Hälfte sind diese Werte nach der SGI-Methode transformiert.

Lässt sich aus der ursprünglichen Punktevergabe der Experten noch eine eindeutige Rangfolge des mittels Durchschnittsbildung aggregierten Indexwertes ablesen ($A > B > C$), 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, nämlich die Komplexität der Realität klarer geordnet darzustellen und somit schneller und einfacher direkte Vergleiche zwischen den untersuchten Objekten zu ermöglichen. Schlimmstenfalls kann es sogar sein, dass sich die Reihenfolge der Länder in den SGI, welche über die standariserten Werte berechnet wird, von derjenigen unterscheidet, die auf den eigentlichen Bewertungen und Punktzuweisungen der Experten beruht.

Ein zweiter ungewollter Nebeneffekt der Skalenaufspiegelung ist, dass keinerlei Vergleichbarkeit der SGI-Werte über die Zeit mehr gegeben ist. Das Beispiel aus Tabelle 3 verdeutlicht dieses Problem: Angenommen, in einer weiteren Ausgabe der SGI blieben bei einem qualitativen Indikator die Bewertungen für alle untersuchten Länder unverändert, lediglich das in der vorherigen SGI-Ausgabe am niedrigsten eingestufte Land A verbesserte sich etwas, bliebe aber immer noch hinter den anderen Ländern zurück. Durch die Transformation aller Bewertungen auf die Punkteskala von eins bis zehn erhielt Land A in der zweiten SGI-Ausgabe erneut den niedrigsten Wert zugewiesen; gleichzeitig würden aber durch die Transformation alle anderen Länder (außer dem Spitzenreiter, der wiederum zehn Punkte erhielte) im Vergleich zur vorigen SGI-Ausgabe herabgestuft, ohne dass sich ihr Zustand tatsächlich verändert hätte. Gemäß der Formel, die hinter der Transformation steckt, betreffen diese Veränderungen der transformierten Werte die einzelnen Länder zusätzlich in einem unterschiedlichen Maß. So verliert Land B deutlich stärker als

beispielsweise Land D. Insofern sind die aggregierten Ergebnisse in den Status- und Managementindizes aufeinander folgender SGI-Ausgaben nicht miteinander vergleichbar, weil sie auf relativen Extremwerten beruhen, die im Zeitverlauf Verschiebungen unterliegen können.

Tabelle 3

Beispiel zur Problematik intertemporaler Vergleiche der SGI-Werte

	Ursprüngliche Expertenwerte					Transformierte Werte				
	Land					Land				
	A	B	C	D	E	A	B	C	D	E
Erste SGI-Ausgabe	3	6	7	8	9	1	5,5	7	8,5	10
Zweite SGI-Ausgabe	5	6	7	8	9	1	3,25	5,5	7,75	10

Quelle: Eigene Darstellung. Die einzelnen Werte der linken Tabellenhälfte stellen hypothetische Werte für die ursprünglichen Punktuweisungen der Experten für eine Frage dar. In der rechten Hälfte sind diese Werte nach der SGI-Methode transformiert.

Das Verfahren der künstlichen Varianzerzeugung ist aus den dargelegten Gründen deutlich abzulehnen. Statt das Problem der mangelnden Varianz einzelner Komponenten der Indizes über rein technische Manipulationen auf der Ebene der Aggregation anzugehen, sollten stattdessen bereits bei Konzeption bzw. Operationalisierung entsprechende Änderungen erfolgen. Eine Option bestünde etwa darin, den aus theoretischer Sicht problematischen Demokratie- teil entweder ganz zu streichen oder passgenauer zu operationalisieren.

5. Statistische Evaluation des Managementindex

Nachdem in den vorigen Abschnitten das Konzept, die Operationalisierung sowie die Aggregation der SGI eingehend untersucht wurden, soll nun noch der eigentlich innovative Teil der SGI, nämlich der Managementindex, einer weiterführenden empirischen Evaluation unterzogen werden. Hierfür sei zunächst auf Tabelle 1 verwiesen, in der sich für den Managementindex ein grundlegend anderes Bild bietet als für den Statusindex. Es lassen sich zwar ebenfalls signifikante Korrelationen zu bestehenden Indizes (und Teilstörungen) feststellen, diese erreichen aber bei Weitem nicht die Stärke wie beim Statusindex. Aus diesem Befund lässt sich ableiten, dass der Managementindex zwar in gewisser Hinsicht mit dem Status der Reformfähigkeit (egal ob über den SGI-Statusindex

oder den RKI gemessen) zusammenhängt, gleichzeitig jedoch auch ein deutlicher Unterschied zwischen dem Management- und dem Statusindex bzw. den anderen bisherigen Maßen besteht. Demzufolge erscheint es aus statistischer Perspektive als legitim, von einer wirklichen neuen Perspektive zu sprechen, die der Managementindex dem Thema Reformen gegenüber einnimmt.

Um zu überprüfen, inwiefern es sich tatsächlich um die Reformfähigkeit handelt, die dieser misst, wird der Reformbedarf der letzten Jahre, gemessen über den RKI, herangezogen. Die Differenzen zwischen aufeinanderfolgenden Jahren lassen sich dann als Ex-post-Evaluation der Reformfähigkeit verstehen. Die Logik hierbei ist einfach: Sofern der Reformbedarf zwischen zwei Zeitpunkten gesunken ist, müssen in der Zwischenzeit Reformen stattgefunden haben und damit muss eine gewisse Reformfähigkeit vorhanden gewesen sein. Da nur schwer festgestellt werden kann, wie lange es dauert, bis eine Reform Wirkung zeigt und sich somit auf den Indexwert des RKI auswirkt, wurde eine Reihe an Differenzen (beginnend mit dem Jahr 1996) zur jüngsten sowie zweitjüngsten Ausgabe der Governance-Indikatoren gebildet, aus denen sich der RKI berechnet.

*Tabelle 4
Zusammenhang zwischen RKI-Jahresdifferenzen und dem SGI-Managementindex*

		Basisjahr für die Differenzbildung der RKI					
		1996	2000	2002	2004	2005	2006
RKI Vergleichsjahr	2006	0,274	0,267	0,287	0,145	0,353*	-
	2007	0,256	0,251	0,214	0,146	0,291	-0,136

Quelle: Dargestellt sind Spearman Rangkorrelationskoeffizienten; n = 30 (alle SGI-Länder); * = signifikant auf dem 0,1 Niveau; ** = signifikant auf dem 0,05 Niveau; *** = signifikant auf 0,01 Niveau (jeweils zweiseitiger Test).

Wenn man diese Differenzen mit dem Managementindex korreliert, sollten wenigstens einige einen Zusammenhang ergeben. Allerdings zeigt Tabelle 4, dass ein solcher nicht besteht. Damit bietet die Ex-post-Evaluation ein klares Bild. Eine Veränderung des Reformbedarfs der letzten Jahre lässt sich nicht durch die vom Managementindex gemessene Reformfähigkeit erklären. Dies mag entweder daran liegen, dass der jeweilige betrachtete Zeitraum nicht analog ist, d. h. die Governance-Indikatoren und damit der RKI nicht weit

genug reichen, um die Auswirkungen der vom Managementindex gemessenen Reformfähigkeit schon abzudecken, oder aber der Managementindex misst doch nicht das, was hier unter Reformfähigkeit verstanden wird. Die letzte Interpretation deutet darauf hin, dass wohl doch bei Konzeption und Operationalisierung gewisse Unzulänglichkeiten vorliegen, die einer genaueren Überprüfung bedürfen.

6. Schluss: Von der Schwierigkeit Reformfähigkeit zu messen

Ziel des vorliegenden Beitrags war es, den prinzipiell zu begrüßenden Messvorschlag des Reformbedarfs sowie der Reformfähigkeit in Form der Sustainable Governance Indicators kritisch zu evaluieren. Das Ergebnis dieser Evaluation bestätigt für die SGI dieselbe Problematik, die sich auch bereits bei einem Großteil der Messansätze im Rahmen der empirischen Demokratieforschung herausgestellt hat: Fehler im Rahmen der Konzeptionalisierung ziehen zumeist auch Fehler bei der Operationalisierung und bei der Aggregation nach sich (Munck/Verkuilen 2002: 28).

Im Falle der SGI lassen sich auf der Ebene der Konzeptionalisierung zwei Probleme identifizieren, die zu weiteren Komplikationen auf nachgelagerten Ebenen bei der Indexbildung führen. Das erste besteht in der zugrunde liegenden Definition, welche äußerst minimalistisch angelegt ist, jedoch in der weiteren Konzeptualisierung deutlich mit einer Reihe von Faktoren, die aus der ursprünglichen Definition keineswegs hervorgehen, überfrachtet wird. Damit eng verknüpft ist die zweite Problematik, entsprechend derer die minimalistische Definition nicht nur überfrachtet wird, sondern auch Elemente enthält, die selbst bei wohlwollender Betrachtungsweise nicht als relevant für die Reformfähigkeit anzunehmen sind. Hierbei ist primär die Aufnahme der Demokratie als ein konstituierendes Merkmal des Reformfähigkeitskonzepts zu kritisieren. Durch diesen a priori gesetzten Zusammenhang wird eine Analyse der eigentlich erst noch zu untersuchenden Beziehung zwischen Reformfähigkeit und Demokratie von vornherein verhindert. Darüber hinaus unterstellt das SGI-Konzept somit implizit, dass adäquate, umfassende Reformen nur in demokratischen Systemen möglich sind.

Auf der Ebene der Operationalisierung erweist sich die Aufnahme der Demokratie als Merkmal der Reformfähigkeit ebenfalls als diskussionswürdig. Schließlich hätte für diese Komponente aufgrund der untersuchten Länder eine Operationalisierung gewählt werden müssen, die dem Problem der mangelnden Varianz innerhalb der OECD-Staaten Rechnung trägt. Die Missachtung dieses Umstandes führt zu einer Operationalisierung, die zu einer Differenzierung der

zu untersuchenden Länder im Hinblick auf ihre Demokratiequalität schlichtweg nicht in der Lage ist. Doch auch in Bezug auf die anderen Teilkomponenten wäre eine alternative Operationalisierung von Vorteil gewesen. Dies wird insbesondere im Fall der System- und Politikperformanz deutlich. Hier gibt weniger die Art und Weise, wie diese Komponente operationalisiert wird, Anlass zu Kritik, sondern vielmehr die Tatsache, dass die SGI diese überhaupt selbst erheben. Es konnte gezeigt werden, dass die Operationalisierung dieser Komponente auch weitaus effizienter und ohne sonderliche inhaltliche Einbußen unter Zuhilfenahme bestehender Indizes anhand eines Reformbedarf-Kompositindexes (RKI) erfolgen kann.

Auch auf der Ebene der Aggregation pflanzt sich das Problem der Aufnahme der Demokratie als konstituierendes Element der Reformfähigkeit fort. Das Problem einer zu geringen Varianz unter den Expertenscores wird von den SGI durch eine Manipulation im Vorfeld der Aggregation angegangen. Durch eine Aufspreizung der Skalen wird eine in den Originaldaten nicht vorhandene Varianz künstlich erzeugt, was zu nicht intendierten Nebeneffekten wie der Verletzung der Transitivität des Gesamtindikators sowie der Reduktion der Vergleichbarkeit der SGI-Werte über die Zeit führt.

Zwar beziehen sich die genannten Probleme überwiegend auf den Statusindex der SGI, doch ist auch der Managementindex durchaus kritisch zu sehen. So konnte statistisch nicht nachgewiesen werden, dass über den RKI gemessene Veränderungen des Reformbedarfs sich, wie zu erwarten wäre, im Managementindex niederschlagen. Dies wirft Zweifel daran auf, inwiefern der Managementindex wirklich die Reformfähigkeit bemisst. Fasst man die genannten Kritikpunkte zusammen, ergibt sich ein eher negatives Gesamurteil für die SGI: Der vorgestellte Index weist eine mangelnde theoretische Fundierung und ein überdehntes Konzept, eine nicht konsequent zu Ende gedachte Operationalisierung sowie eine hochproblematische Aggregation auf.

Die Lösung dieser Probleme wäre jedoch vergleichsweise einfach. Betrachtet man die Komponente der Demokratie im Zusammenhang mit den SGI konzeptionell als die Wurzel des Problems, läge es praktisch, und im Übrigen theoretisch fundiert, nahe, Reformfähigkeit ohne Demokratie als konstituierendes Merkmal zu definieren. Die Erfassung des Statusindex könnte dann über eine Rekombination bereits vorhandener Indizes, wie hier am Beispiel des RKI geschehen, erfolgen. Dies brächte zudem den Vorteil einer intertemporalen Vergleichbarkeit der Reformbedarfswerte mit sich, da sich solche Kompositindikatoren aufgrund der guten Datenlage problemlos auch für vergangene Jahre berechnen lassen.

Trotz der vorgebrachten Kritik und Verbesserungsvorschläge sollte nicht gänzlich in den Hintergrund treten, dass es sich bei den SGI prinzipiell um einen sinnvollen Messansatz handelt, der gerade mit seiner Unterscheidung zwischen Reformbedarf (Statusindex) und Reformfähigkeit (Managementindex) eine deutlich innovative Komponente im Vergleich zu anderen verfügbaren Indikatoren bietet. Das Beispiel der SGI verdeutlicht jedoch, welche weitreichenden Konsequenzen eine mangelhafte bzw. überfrachtete Konzeptualisierung haben kann. Eine adäquate Lösung dieser Problematik sollte jedoch nicht, wie bei diesem Index geschehen, in einer notdürftigen Korrektur im Zuge der Operationalisierung bzw. der Aggregation gesucht werden, sondern in der erneuten Auseinandersetzung mit den theoretisch-konzeptionellen Grundlagen.

Literatur

- Auriol, Emmanuel/Gary-Bobo, Robert J., 2007:* On the Optimal Number of Representatives, CEPR Discussion Paper 6417, 1-48.
- Benz, Arthur, 2008:* Entwicklung von Governance im Mehrebenensystem der EU, in: Ingeborg Tömmel (Hrsg.), Die Europäische Union. Governance und Policy-Making, Wiesbaden, 36-57.
- Bertelsmann Stiftung, 2009a:* SGI Sustainable Governance Indicators 2009, <http://www.sgi-network.org/> (Stand: 03.08.09).
- Bertelsmann Stiftung, 2009b:* Sustainable Governance Indicators: Codebook. Gütersloh, http://www.bertelsmann-stiftung.de/cps/rde/xbr/SID-0A000F0A-25396AA5/bst/xcms_bst_dms_27451_27452_2.pdf (Stand: 07.03.09).
- Brusis, Martin, 2008:* Reformfähigkeit messen? Konzeptionelle Überlegungen zu einem Reformfähigkeitsindex für OECD-Staaten, in: PVS 49, 92-113.
- Brusis, Martin, 2009:* Designing Sustainable Governance Indicators. Criteria and Methodology, in: Bertelsmann Stiftung (Hrsg.), Sustainable Governance Indicators 2009, Gütersloh.
- Castles, Francis G./Obinger, Herbert/Leibfried, Stephan, 2005:* Bremst der Föderalismus den Leviathan? Bundesstaat und Sozialstaat im internationalen Vergleich, 1880–2005, in: PVS 46, 215-237.
- Freedom House, 2008:* Freedom in the World 2008. Checklist, Questions and Guidelines, http://freedomhouse.org/template.cfm?page=351&ana_page=341&year=2008 (Stand: 07.03.09).

- Holmes, Kim R./Feulner, Edwin J./O'Grady, Mary A., 2008:* 2008 Index of Economic Freedom, Washington, D. C./New York.
- IMD, 2008:* IMD World Competitiveness Yearbook 2007, Lausanne, International Institute for Management Development.
- IMF, 2008:* IMF Country Reports by Date, http://www.imf.org/external/pubs/cat/scr1_sp.cfm?s_year=1997&e_year=2008&brtype=default (Stand: 07.03.09).
- Kaufmann, Daniel/Kraay, Aart/Mastruzzi, Massimo, 2007:* Governance Matters VI: Governance Indicators for 1996-2006, Washington, <http://ssrn.com/paper=999979> (Stand: 07.03.09).
- König, Thomas/Bräuninger, Thomas, 2005:* Gesetzgebung im Föderalismus, Speyer.
- Lauth, Hans-Joachim, 2004:* Demokratie und Demokratiemessung, Wiesbaden.
- Laver, Michael/Hunt, W. Ben, 1992:* Policy and Party Competition, New York.
- Low, Linda, 2001:* The Singapore developmental state in the new economy and polity, in: Pacific Review 14 (3), 411-441.
- Manow, Philip/Burkhart, Simone, 2009:* Die Dauer der Gesetzgebungstätigkeit und die Herrschaft über den parlamentarischen Zeitplan – eine empirische Untersuchung des Legislativprozesses in Deutschland, in: Steffen Ganghof/Christoph Hönnige/Christian Stecker (Hrsg.), Parlamente, Agendasetzung und Vetospieler – Festschrift für Herbert Döring, Wiesbaden.
- Müller, Thomas/Pickel, Susanne, 2007:* Evaluierung der Evaluierung von Demokratie – Ergebnisse und Grenzen. Operationalisierung, Messung, Auswertung sowie Schwächen des Evaluierungskonzeptes für Indizes der Demokratiemessung von L. Munck & Jay Verkuilen, in: PVS 48, 511-539.
- Munck, Gerardo L./Verkuilen, Jay, 2002:* Conceptualizing and Measuring Democracy. Evaluating Alternative Indices, in: Comparative Political Studies 35, 5-34.
- Nicoletti, Giuseppe/Scarpetta, Stefano/Boylaud, Olivier, 2000:* Summary Indicators of Product Market Regulation with an Extension to Employment Protection Legislation, <http://ideas.repec.org/p/oec/ecoaaa/226-en.html> (Stand: 07.03.09).
- OECD, 2006:* The Program for International Student Assessment, Paris.
- Puri, Anjali, 2003:* Dancing on Tables in Singapore's Guided Democracy, in: Contemporary Review 282, 351-353.
- Sartori, Giovanni, 1970:* Concept Misformation in Comparative Politics, in: American Political Science Review 64 (4), 1033-1053.
- Schmidt, Manfred G., 2005:* Politische Reformen und Demokratie. Befunde der vergleichenden Demokratie- und Staatstätigkeitsforschung, in: Hans Vorländer (Hrsg.), Politische Reform in der Demokratie, Baden-Baden.

- Schmitter, Philippe C.*, 2004: The Ambiguous Virtues of Accountability, in: Journal of Democracy 15 (4), 47-60.
- Schultze, Rainer-Olaf*, 2002: Reform, in: Dieter Nohlen (Hrsg.), Lexikon der Politikwissenschaft. Theorien, Methoden, Begriffe, München.
- Stoiber, Michael*, 2008: Ein neues, kontextualisiertes Maß für Demokratie. Konzeptualisierung und Operationalisierung, in: ZPol 18, 209-231.
- Wagschal, Uwe*, 1999: Statistik für Politikwissenschaftler, München/Wien.
- Wagschal, Uwe/Jäckle, Sebastian*, 2009: The Sustainable Governance Indicators in Cross-National Comparison, in: Bertelsmann Stiftung (Hrsg.), Sustainable Governance Indicators 2009, Gütersloh.

Korrespondenzanschrift:

Sebastian Jäckle
Seminar für wissenschaftliche Politik
Werthmannstr. 12
79098 Freiburg
E-Mail: sebastian.jaeckle@politik.uni-freiburg.de

Rafael Bauschke
Institut für Politische Wissenschaft
Bergheimer Str. 58
69115 Heidelberg
E-Mail: r.bauschke@uni-heidelberg.de
Web: http://www.uni-heidelberg.de/politikwissenschaften/personal/wagschal/personal/m_bauschke.html