

Z Vgl Polit Wiss (2012) 6:207–231
DOI 10.1007/s12286-012-0127-4

AUFSÄTZE


ZEITSCHRIFT FÜR
VERGLEICHENDE POLITIKWISSENSCHAFT

Zur Messung von Staat-Kirche-Beziehungen: Eine vergleichende Analyse neuerer Indizes

Richard Traummüller

Zusammenfassung: Im Zuge der erhöhten Aufmerksamkeit, welche dem Faktor Religion in der Politik zuteil wird, ist das spannungsreiche Verhältnis von Staat und Kirche wieder vermehrt in den Blick geraten. Gleichwohl werden bisherige Forschungsbemühungen der quantitativ-vergleichenden Politikwissenschaft oftmals durch einen Mangel an geeigneten Messkonzepten gehemmt. Erst in allerjüngster Zeit sind verschiedene Vorschläge zur Messung von Staat-Kirche-Beziehungen hervorgebracht worden. Ziel des vorliegenden Beitrags ist es, diese neuen Messvorschläge vorzustellen und in einem kritischen Vergleich ihre jeweiligen methodischen Stärken und Schwächen herauszuarbeiten. Genauer gesprochen wird der Versuch unternommen, den von Munck und Verkuilen (*Comparative Political Studies* 35:5–24, 2002) im Rahmen der Demokratiemessung vorgeschlagenen und von Müller und Pickel (*Politische Vierteljahresschrift* 48:511–539, 2007) präzisierten Kriterienkatalog zur methodologischen Bewertung von Messkonzepten auf insgesamt fünf verschiedene Indizes zur *Messung von Staat-Kirche-Verflechtungen* anzuwenden. Konkret wird dabei anhand spezifischer Beurteilungskriterien untersucht, inwieweit die methodischen Herausforderungen der *Konzeptualisierung*, der *Messung* sowie der *Aggregation* bei der Konstruktion der betrachteten Indizes überzeugend bewältigt wurden. Neben der vergleichenden Analyse und Bewertung von neuen Messinstrumenten wird damit gleichzeitig die Brauchbarkeit dieses Evaluationsschemas, sowie seine Übertragbarkeit in alternative Anwendungskontexte erprobt.

Schlüsselwörter: Messung · Staat-Kirche-Beziehungen · Evaluation

Measuring church-state-relations: a comparative analysis of new indices

Abstract: The renewed interest in the religious factor in politics has also led to an increased concern for the entanglement of the church and the state. However, quantitative comparative research efforts in political science so far have been inhibited due to a lack of adequate measurement concepts. Only very recently several suggestions on how to measure church-state-relations have come up. The present article aims to introduce these new measurement instruments and to evaluate their methodological strengths and weaknesses in a comparative analysis. I draw on the criteria for the methodological evaluation of measurement instruments proposed by Munck and Verkuilen, (*Comparative Political Studies* 35:5–24, 2002) elaborated by Müller and Pickel (*Politische Vierteljahresschrift* 48:511–539, 2007) and apply them to a total of five different indices for

Online publiziert: 18.07.2012
© VS Verlag für Sozialwissenschaften 2012

R. Traummüller (✉)
Lehrstuhl für Politische Soziologie
Universität Bern, Bern, Schweiz
E-Mail: richard.traummueller@ipw.unibe.ch

the measurement of church-state-relations. In particular, I evaluate how these different measurement concepts perform in the three methodological steps of conceptualization, measurement and aggregation. Besides the comparative analysis of new measurement instruments this will also test the transferability and usefulness of this evaluation scheme in alternative contexts.

Keywords: Measurement · Church-State Relations · Evaluation

1 Einleitung

Dem Faktor Religion und seiner Wirkung auf die Politik wird seit einiger Zeit erhöhte Aufmerksamkeit zuteil.¹ Im Zuge dieser Entwicklung ist auch das spannungsreiche Verhältnis von Staat und Kirche erneut in den Blick geraten (vgl. Enyedi und Madeley 2003; Gill 2001; Liedhegener 2008; Minkenberg 2003; Minkenberg und Willems 2002, 2003). Dies liegt zum einen darin begründet, dass sich lange Zeit unangefochtene Auffassungen über die Rolle der Religion in modernen Gesellschaften – und insbesondere über ihr Verhältnis zur öffentlich-politischen Sphäre – eher als normativ grundierte Positionen, denn als plausible Realitätsbeschreibungen herausgestellt haben (Casanova 1994). Zum anderen sind angesichts des wachsenden religiösen Pluralismus und der verstärkten Sichtbarkeit religiöser Minderheiten religionspolitische Konflikte über Status und Rechte religiöser Gruppen, öffentliche religiöse Symbole, sowie über Religionsunterricht an staatlichen Schulen wieder zu zentralen Herausforderungen demokratischer Gemeinwesen avanciert (Bader 2007).

Darüber hinaus werden Staat-Kirche-Beziehungen gegenwärtig mit der Ausgestaltung wohlfahrtsstaatlicher Politik in Verbindung gebracht (Manow 2008; Minkenberg 2002; van Kersbergen und Manow 2009) und in ihren Wirkungen auf demokratische Zivilgesellschaften und bürgerschaftliches Engagement untersucht (Roßteutscher 2009; Traunmüller 2010, 2012; Traunmüller und Freitag 2011). Schließlich stellt die staatliche Regulierung von Religion im Rahmen der sogenannten *economics of religion* gar die entscheidende Schlüsselvariable zur Erklärung unterschiedlicher Religiositätsniveaus im internationalen Vergleich dar (Iannaccone 1991; Chaves und Cann 1992; Stark und Finke 2000).

Diese Beispiele mögen genügen, um zu verdeutlichen, dass es zweifellos zu den genuinen Aufgaben der vergleichenden Politikwissenschaft gehört, institutionelle Ausgestaltungen von Staat-Kirche-Arrangements in ihren vielfältigen empirischen Erscheinungsformen systematisch zu erfassen sowie nach ihren jeweiligen Ursachen

¹ Frühere Versionen dieses Aufsatzes wurden auf der Autorenkonferenz des AK *Demokratieforschung* in Verbindung mit der *Zeitschrift für Vergleichende Politikwissenschaft* an der *Universität Leipzig* (14.–15. Mai 2009) sowie im Rahmen der Frühjahrstagung des AK *Empirische Methoden der Politikwissenschaft* an der *Zeppelin University Friedrichshafen* (5. Juni 2009) vorgestellt. Ich möchte mich bei allen Anwesenden sowie bei den beiden anonymen Gutachtern für ihre detaillierte Kritik und hilfreichen Anmerkungen herzlich bedanken. Für die Hilfe bei der Erstellung des finalen Manuskripts danke ich Christina Eder, Birgit Jacob und Tobias Tober. Ein weiterer Dank gebührt Gert Pickel und Detlef Pollack, welche mit ihrem Enthusiasmus gegenüber der Idee, eine Evaluierung von Staat-Kirche-Indizes vorzunehmen entscheidend zu meiner Motivation beigetragen haben.

und Wirkungen zu fragen (Chaves et al. 1994; Minkenberg 2003).² Eine solche Untersuchungsperspektive ist vermutlich gut beraten, einen Wandel weg von der in rechtswissenschaftlichen oder politisch-theoretischen Konzepten üblichen Konzentration auf *differences in kind* – welche etwa idealtypisch eine strikte Trennung von Staat und Kirche von einem etablierten Staatskirchentum einerseits, sowie einem Kooperationsmodell andererseits, unterscheiden (z. B. Monsma und Soper 1997; Robbers 1995) – hin zu einer Anerkennung von *differences in degree* zu vollziehen. Denn nur so wird sie der tatsächlichen empirischen Komplexität und Vielschichtigkeit staatlich-kirchlicher Verflechtung im internationalen Vergleich gerecht. Staat-Kirche-Beziehungen sind nicht kategorial zu fassen, sondern stellen vielmehr ein graduelles Phänomen dar, welches sich aus einem vielschichtigen Bündel von politischen Regelungen und staatlichen Tätigkeiten zusammensetzt (Minkenberg 2003).³

Gleichwohl waren in diesem Sinne angelegte, quantitativ-vergleichende Forschungsbemühungen lange durch einen Mangel an adäquaten, diesen Ansprüchen genügenden, Messinstrumenten gehemmt. In der Tat sind erst in allerjüngster Zeit einige verschiedene Vorschläge zur differenzierten Messung von Staat-Kirche-Beziehungen hervorgebracht worden, die diesen Mangel beheben (z. B. Grim und Finke 2006; Fox 2006, 2008; Norris und Inglehart 2004). Entsprechend gering ist freilich der bisherige Erfahrungsschatz im Umgang mit diesen Messinstrumenten und v. a. das Wissen um die Qualität der durch sie generierten Daten. Angesichts der theoretischen und praktischen Relevanz, welche das Thema der Verflechtung von Kirche und Staat für sich beanspruchen kann, ist es daher von großer Wichtigkeit diese zur Verfügung stehenden Instrumente genauer zu untersuchen und kritisch zu beleuchten.

Das Ziel des vorliegenden Beitrags ist daher v. a. ein methodologisches. Ich werde diese neuen Messvorschläge zur Erfassung von Staat-Kirche-Beziehungen vorstellen, in einem kritischen Vergleich ihre jeweiligen methodischen Stärken und Schwächen herausarbeiten und sie damit hinsichtlich ihrer Messqualität bewerten. Genauer gesprochen wird der Versuch unternommen, den von Munck und Verkuilen (2002) ursprünglich im Rahmen der Demokratiemessung vorgeschlagenen und von Müller und Pickel (2007) präzisierten Kriterienkatalog zur methodologischen Bewertung von Messkonzepten auf insgesamt fünf verschiedene Indizes zur *Messung von Staat-Kirche-Verflechtungen* anzuwenden. Konkret wird für diese Indizes anhand spezifischer, aus messtheoretischen Überlegungen abgeleiteter Beurteilungskriterien untersucht, inwieweit die methodischen Herausforderungen der *Konzeptualisierung*, der *Messung* sowie der *Aggregation* überzeugend bewältigt wurden. Neben der vergleichenden Analyse und Bewertung von neuen Messinstrumenten in einem sich erst entwickelnden Forschungsfeld der vergleichenden Politikwissenschaft, wird gleichzeitig die Brauchbarkeit dieses Evaluationschemas

2 Ich halte hier an der gebräuchlichen Begrifflichkeit „Staat-Kirche-Beziehung“ fest, obwohl diese natürlich im eigentlichen Sinne des Wortes nur für den christlichen Kulturraum zutreffend ist. Gleichwohl stellt sich die Frage nach der staatlichen Regulierung von Religion grundsätzlich in allen Staaten der Erde.

3 Zu analogen Forderungen im Rahmen der Demokratieforschung siehe z. B. Elkins (2000), Pickel und Pickel (2006: 268) und Welzel (2000).

sowie seine Übertragbarkeit auf alternative Anwendungskontexte jenseits der Demokratie-messung erprobt.

Der Beitrag gliedert sich wie folgt: in einem ersten Schritt werden die zu evaluierenden Indizes zur empirischen Erfassung von Staat-Kirche-Verflechtungen kurz vorgestellt. Im zweiten Schritt erfolgt die methodologische Bewertung der Messkonzepte anhand des von Munck und Verkuilen (2002) vorgeschlagenen Evaluationschemas. Ein dritter Schritt nimmt eine Metaevaluation der Indizes vor, indem zum einen auf der Basis eines Metaindexes eine Rangfolge der Messkonzepte erstellt wird und zum anderen die Stärken und Schwächen der gegenwärtigen Messpraxis in der Gesamtschau betrachtet werden. Zuletzt werden zentrale Punkte zusammengefasst.

2 Vorstellung und empirische Reichweite der betrachteten Indizes

Bevor die eigentliche methodologische Bewertung der Indizes zur Messung von Staat-Kirche-Beziehungen vorgenommen wird, sollen die hier betrachteten fünf Messkonzepte zunächst kurz vorgestellt und in ihrer empirischen Reichweite beschrieben werden (s. Tab. 1). Ohne Anspruch auf Vollständigkeit wurden jene in der Literatur sichtbaren Indizes ausgewählt, welche a) explizit den Anspruch erheben, *Staat-Kirche-Beziehungen* (und nicht etwa nur irgendeinen Teilbereich) zu messen, b) einen erkennbar eigenständigen Messansatz vorlegen (und nicht bloß eine Anwendung bestehender Instrumente

Tab. 1: Empirische Reichweite der Indizes zur Messung von Staat-Kirche-Beziehungen

Indizes	Analyseeinheiten	Zeitpunkt
Religious Regulation Scale (Chaves und Cann 1992)	18 westliche Staaten	?
Index Kirche-Staat-Beziehung (Pollack 2003)	16 west- und osteuropäische Staaten	?
Religious Freedom Index (Norris und Inglehart 2004)	188 Staaten	2002
Government Regulation Index (GRI)	196 Staaten	2003
Government Favoritism Index (GFI)		
Social Regulation Index (SRI) (Grim und Finke 2006)		
Official Government Involvement in Religion	175 Staaten	1990–2002
Religious Restriction		
Religious Discrimination		
Religious Regulation		
Religious Legislation		
Overall Government Involvement in Religion (Fox 2006, 2008)		
Eigene Zusammenstellung		

darstellen) und sich c) aufgrund der Anzahl der erfassten Länder für eine quantitativ-statistische Analyse eignen.⁴

Den ersten zu evaluierenden Index stellt die *Religious Regulation Scale* von Chaves und Cann (1992) zur Messung des staatlichen Regulierungsgrads von Religion dar. Dieses Messkonzept ist das älteste der hier Betrachteten, welches allerdings neueren Messansätzen als Vorbild dient. Die *Religious Regulation Scale* setzt sich aus insgesamt sechs Items zusammen und wurde für insgesamt 18 westliche Staaten erhoben. Die geographische Reichweite und die damit einhergehende Möglichkeit auf Generalisierbarkeit abzielender Theorietests im Rahmen quantitativ-vergleichender Analysen ist daher eher beschränkt. Auf welchen Zeitraum sich diese Messung bezieht, ist unklar, da die Autoren darüber keine Aussage machen. Nachdem die Hauptquelle jedoch die *World Christian Encyclopedia* (Barrett 1982) darstellt, dürfte sie sich auf die Situation in den 70er Jahren beziehen.

Eine leicht modifizierte Version dieses Messkonzepts liegt mit dem von Pollack (2003; Pollack und Pickel 2000) konstruierten *Index Kirche-Staat-Beziehung* vor. Dieser umfasst jedoch nur fünf, z. T. neue Kriterien und wurde für eine andere Länderstichprobe von insgesamt 16 west- und osteuropäischen Staaten erhoben. Doch auch hier ist die Fallzahl und damit die Möglichkeit des quantitativ-statistischen Vergleichs sehr begrenzt und die zeitliche Reichweite der Messung nicht eindeutig festzustellen.

Erheblich erweitert wurde das ursprünglich von Chaves und Cann (1992) vorgeschlagene Messkonzept durch den von Norris und Inglehart (2004) entwickelten *Religious Freedom Index* zur Messung der Staat-Kirche-Beziehungen. Die ursprünglich sechs Kriterien wurden hier auf insgesamt 20 ausgeweitet und auch die Anzahl der Analyseeinheiten geht deutlich über das Original hinaus. Mit dem *Religious Freedom Index* lässt sich die Verflechtung von Staat und Kirche nun in insgesamt 188 Staaten für das Jahr 2002 betrachten und vergleichen. Der *Religious Freedom Index* eignet sich damit weitaus besser für statistische Analysen.

Einen eigenständigen und neu-konzipierten Messansatz stellen die *International Religion Indexes* von Grim und Finke (2006) dar. Dieses Messkonzept setzt sich aus insgesamt drei verschiedenen Indizes zusammen, welche jeweils fünf bis sechs Items umfassen: dem *Government Regulation Index* (GRI), dem *Government Favoritism Index* (GFI), sowie dem *Social Regulation Index* (SRI). Diese Indizes zu verschiedenen Aspekten des Staat-Kirche-Verhältnisses wurden für das Jahr 2003 erhoben und umfassen alle

4 So entspricht etwa die *Religious Freedom Scale* von Freedom House (2010) dem ersten Kriterium nicht, während dies für den *Religious Freedom Index* von Norris und Inglehart (2004) der Fall ist, da diese – trotz der irreführenden Bezeichnung ihres Index – ausdrücklich die Messung von Staat-Kirche-Beziehungen im Allgemeinen zum Ziel haben. Der Messansatz von Minkenberg (2003) wurde nicht aufgenommen, da es sich dabei im Prinzip um eine Anwendung der Skala von Chaves und Cann (1992) handelt und daher nicht als eigenständig gewertet werden kann. Bei Pollack (2003) ist die Eigenständigkeit gegeben, da neben der gänzlich anderen Länderauswahl zusätzliche Items in die Messung aufgenommen, andere dafür weggelassen wurden. Darüber hinaus hat jüngst Roßteutscher (2009) einen eigenständigen Vorschlag zur empirischen Erfassung von Staat-Kirche-Beziehungen gemacht, welcher jedoch aufgrund der geringen Fallzahl – acht europäische Staaten – das dritte Kriterium nicht erfüllt.

196 Staaten der Welt. Im Querschnitt haben die *International Religion Indexes* somit die größte empirische Reichweite aller betrachteten Messkonzepte.

Das letzte und jüngste Messkonzept von Staat-Kirche-Beziehungen liegt schließlich mit den Indizes des *Religion and the State* Projekts von Fox (2006, 2008) vor. Dieses großangelegte Projekt zur weltweiten Erfassung der Beziehung zwischen Staat und Religion umfasst insgesamt sechs verschiedene Skalen: *Official Government Involvement in Religion*, *Religious Restriction*, *Religious Discrimination* (bestehend aus 16 Komponenten), *Religious Regulation* (11 Komponenten), *Religious Legislation* (33 Komponenten) und schließlich ein diese fünf Maße integrierender Gesamtindex *Overall Government Involvement in Religion*. Diese Indizes wurden für insgesamt 175 Staaten erhoben und dies jährlich für den Zeitraum von 1990–2002. Es handelt sich damit um den einzigen Messansatz, welcher auch einen Vergleich über die Zeit zulässt.⁵

3 Vergleichende Bewertung der Indizes zur Messung von Staat-Kirche-Beziehungen

Im Folgenden soll der von Munck und Verkuilen (2002) im Rahmen der Demokratiemessung vorgeschlagene Kriterienkatalog zur methodologischen Bewertung von Messkonzepten auf die fünf soeben vorgestellten Indizes zur Messung von Staat-Kirche-Verflechtungen angewandt werden. Konkret sind Munck und Verkuilen (2002, S. 7–8) zufolge bei der Konstruktion von Indizes drei aufeinander aufbauende methodische Herausforderungen zu bewältigen: die Herausforderung der *Konzeptualisierung*, der *Messung* sowie der *Aggregation*.⁶ Jede dieser drei Herausforderungen beinhaltet wiederum bestimmte Aufgaben, die geleistet werden müssen und deren Erfüllung sich anhand spezifischer Evaluationskriterien bewerten lässt. Da die konkrete Operationalisierung dieser Beurteilungskriterien bei Munck und Verkuilen jedoch eher implizit bleibt, orientiere ich mich soweit wie möglich an den von Müller und Pickel (2007) vorgeschlagenen Kodierregeln⁷, um zu bestimmen, ob ein Kriterium erfüllt wurde oder nicht. Dies erlaubt zugleich einen quantifizierenden und damit systematischeren Vergleich der verschiedenen Messkonzepte (vgl. Müller und Pickel 2007, S. 518).

3.1 Die Phase der Konzeptualisierung

Der erste Schritt einer jeden Indexkonstruktion besteht in der Spezifikation des Konzeptes, welches überhaupt erfasst und abgebildet werden soll (Diekmann 2001; Miller 2007; Schnell et al. 2005). Dieser Schritt, der den Ausgangspunkt für alle darauf folgenden Schritte darstellt, beinhaltet zum einen die Aufgabe der *Identifizierung aller relevanten*

⁵ Gegenwärtig befindet sich das *Religion and the State* Projekt in der zweiten Erhebungsrunde (Fox 2009, 2011).

⁶ Für ein ähnliches Evaluationsschema, welches die Ebenen Konzeptspezifikation, Operationalisierung und Messung zur Bewertung von Messkonzepten berücksichtigt, s. Gaber (2000).

⁷ Siehe auch die daran anschließende Debatte in der PVS (Bühlmann et al. 2008; Kaina 2008; Müller und Pickel 2008).

Merkmale (Dimensionen) und zum anderen deren *logische Organisation* in einem Konzeptbaum, aus welchem sich dann in einem weiteren Schritt konkrete beobachtbare Indikatoren ableiten lassen (Munck und Verkuilen 2002; Müller und Pickel 2007, S. 525). In unserem Fall muss also geklärt werden, welche Merkmalsdimensionen Staat-Kirche-Beziehungen überhaupt umfassen und wie diese verschiedenen Dimensionen zueinander stehen bzw. richtig angeordnet werden können (s. Tab. 2). Dabei ist ganz offensichtlich, dass an dieser Stelle zunächst von den konkreten Vorgaben von Müller und Pickel (2007) abgewichen werden muss, da der konzeptionelle Anker eben nicht ein, wie auch immer geartetes, Demokratieverständnis sein kann, sondern die Definition von „Staat-Kirche-Beziehungen“. Zugleich ist es problematisch, auf objektive Weise festzulegen, was eine gute oder richtige Definition eines Phänomens ist. Diese hängt stets vom gegebenen Forschungsinteresse und den bestehenden Theorien ab und kann sich folglich auch nur vor diesem Hintergrund als nützlich oder eben weniger nützlich erweisen (vgl. Schnell et al. 2005, S. 128).

Gleichwohl lässt sich als ein mögliches und sinnvolles Beurteilungskriterium für die Identifizierung der Konzeptmerkmale die *Vermeidung von Maximal- und Minimaldefinitionen* anführen (Munck und Verkuilen 2002, S. 9). Mit anderen Worten zeichnet sich eine gute Konzeptualisierung dadurch aus, dass sie nicht zu viele, aber eben auch nicht zu wenige Attribute umfasst und somit insgesamt von analytischem Nutzen ist. Die Erfüllung dieses Kriteriums in einzelnen Messvorschlägen lässt sich dann konkret anhand der Kriterien *Sparsamkeit* und *Relevanz* der berücksichtigten Merkmale bzw. Dimensionen erfassen (Müller und Pickel 2007, S. 519).

Im thematischen Zusammenhang von Staat-Kirche-Beziehungen hat Minkenberg (2003, S. 120) exakt diesen Maßstab bei der Konzeptualisierung eingefordert: „[D]ie meisten Ansätze zur Klassifizierung des Staats-Kirche-Verhältnisses sind dadurch gekennzeichnet, dass sie entweder eine einzige, meist rechtliche Dimension von Kriterien bzw. eine kleine Zahl von zusammenhängenden juristischen Indikatoren, oder aber eine große Bandbreite von Indikatoren zu Grunde legen“. Eine für die vergleichende Politikwissenschaft brauchbare Konzeptualisierung müsse allerdings gleichermaßen politische, ökonomische und legale Kriterien berücksichtigen, dabei aber zugleich die Unterscheidungsmöglichkeit zwischen unabhängigen und abhängigen Variablen, also zwischen dem institutionellen Verhältnis selbst und seinen Folgen gewährleisten. Einen Weg um zu in diesem Sinne angemesseneren Konzeptualisierungen von Staat-Kirche-Beziehungen zu gelangen, welche Maximal- und Minimaldefinitionen gleichermaßen vermeiden, sieht Minkenberg in einer stärkeren theoretischen Orientierung an den Diskussionssträngen historisch-soziologischer und neo-institutionalistischer Theorieansätze sowie auch an der *economics of religion* Schule innerhalb der Religionssoziologie (2003, S. 120).

Tatsächlich konzeptualisieren, mit der Ausnahme von Fox (2006, 2008), alle der hier betrachteten Bemühungen Staat-Kirche-Beziehungen explizit vor dem theoretischen Hintergrund der *economics of religion* und müssen sich daher auch an diesen, ihren eigenen Ansprüchen messen lassen (z. B. Finke 1997; Iannaccone 1991; Stark und Finke 2000). In dieser Sichtweise repräsentieren Staat-Kirche-Beziehungen ein vielschichtiges Bündel von offiziellen Gesetzen, politischen Maßnahmen und administrativen Handlungen, denen gemein ist, dass sie darauf abzielen, religiöse Angelegenheiten zu regulieren. Dabei kann staatliche Regulierung von Religion zwei grundlegende Formen annehmen

Tab. 2: Die Phase der Konzeptualisierung

Aufgabe/ Bewertungskriterium	Chaves und Cann (1992)	Pollack (2003)	Norris und Inglehart (2004)	Grim und Fimke (2006)	Fox (2006)
<i>1.1 Identifizierung der Merkmale</i>					
Sparsamkeit	☺ (1 Dimension Unterstützung erfasst)	☺ (1 Dimension Unterstützung erfasst)	☺ (2 Dimensionen Restriktion und Unterstützung erfasst)	☺ (3 Dimensionen darunter soziale Regulierung von Religion und damit mehr als Staat-Kirche-Beziehungen gemessen)	☺ (2 Dimensionen Restriktion und Unterstützung erfasst)
Relevanz	☺ (Dimension staatliche Restriktionen von Religion fehlt)	☺ (Dimension staatliche Restriktionen von Religion fehlt)	☺	☺	☺
<i>1.2 Logische Anordnung der Merkmale</i>					
Redundanz	☺	☺	☺	☺	☺ (Überschneidung von <i>official hostility</i> , <i>general restrictions</i> und <i>religious regulation</i>)
Zuordnung/Vermischung	☺	☺	☺ (Dimensionen Restriktion und Unterstützung werden nicht ausdifferenziert)	☺	☺

☺ Kriterium erfüllt, ☹ Kriterium teilweise erfüllt, ☹ Kriterium nicht erfüllt

– *Restriktion* und *Unterstützung* (Finke 1997; Roßteutscher 2009). Einerseits kann ein Staat die religiösen Überzeugungen, Praktiken und Organisationen mit rechtlichen Restriktionen und administrativen Einschränkungen versehen. Andererseits kann ein Staat religiösen Gruppen und Organisationen aber auch konstitutionelle Privilegien und finanzielle Subventionierung gewähren.

In dem Maße, wie verschiedene Indexkonstruktionen diese der *economics of religion* entstammende Konzeption von Staat-Kirche-Beziehungen richtig umsetzen und die beiden Dimensionen der staatlichen Restriktion sowie der staatlichen Unterstützung von Religion beinhalten, kann ihnen die Erfüllung des Kriteriums der Relevanz und damit die Vermeidung einer Minimaldefinition attestiert werden. Insofern sie sich auf diese beiden Merkmale beschränken, erfüllen sie auch das Kriterium der Sparsamkeit und lösen damit das Problem der Maximaldefinition.

Legt man diesen Bewertungsmaßstab zugrunde, vermeiden sowohl die *Religious Regulation Scale* von Chaves und Cann (1992), der *Index Kirche-Staat-Beziehung* von Pollack (2003), als auch der *Religious Freedom Index* von Norris und Inglehart (2004) eine Maximaldefinition von Staat-Kirche-Beziehungen. Auch die von Fox (2006, 2008) konstruierten *Religion and the State* Indizes, welche sich zwar nicht explizit aus der Religionsökonomie herleiten,⁸ stützen sich auf ein Verständnis von Staat-Kirche-Beziehungen „which can be described as any government support for religion, or any government regulation or limitation placed on religion“ (Fox 2008, S.47) und erfüllen damit das Evaluationskriterium.

Lediglich der Konzeptualisierung von Grim und Finke (2006) könnte vorgeworfen werden, sie umfasse insgesamt zu viel. Neben den Dimensionen staatlicher Restriktion und staatlicher Unterstützung von Religion wird hier noch zusätzlich die Dimension der *sozialen Regulierung* von Religion durch andere religiöse Gruppen, Vereine oder die Kultur im Allgemeinen mit der Begründung aufgenommen, dass „[t]his form of regulation might be tolerated or even encouraged by the state but is not formally endorsed or implemented by government action“ (Grim und Finke 2006, S.8). Damit erweitern sie das Konzept jedoch um eine eigenständige Dimension, welche möglicherweise besser als Ursache oder auch Folge von Staat-Kirche-Beziehungen betrachtet werden sollte (Grim und Finke 2007, 2011).

Mit Blick auf die Vermeidung von Minimaldefinitionen erfüllen die Messkonzepte von Fox (2006, 2008), Grim und Finke (2006) sowie Norris und Inglehart (2004) das Kriterium der Relevanz, indem sowohl die Dimension der staatlichen Unterstützung als auch die der staatlichen Restriktion von Religion berücksichtigt wird. Die *Religious Regulation Scale* von Chaves und Cann (1992) und der daran angelehnte *Index Kirche-Staat-Beziehung* (Pollack 2003) erfassen allerdings lediglich den Aspekt staatlicher Förderung und vernachlässigen damit den wichtigen Aspekt der einschränkenden Regulierung durch den Staat.

Die sich unmittelbar an die Auswahl der relevanten Merkmale anschließende Aufgabe besteht in der *logischen Anordnung der Merkmale in einem Konzeptbaum* (Munck und Verkuilen 2002, S. 12). Es geht in diesem Schritt darum, die einzelnen Dimensionen und

⁸ Fox' Anliegen ist es vielmehr, mit den von ihm konzipierten Indizes eine ganze Reihe in der Literatur auffindbarer Definitionen von Staat-Kirche-Beziehungen operationalisieren zu können (2008, S. 47 und 57).

mögliche Subdimensionen trennscharf und nach ihrem Abstraktionsgrad zu ordnen und den Konzeptbaum entsprechend fein ausdifferenzieren. Inwieweit dies überzeugend geleistet wird, lässt sich an den konkreten Kriterien *Redundanz* und *Vermischung* festmachen (vgl. Müller und Pickel 2007, S. 525). Dies bedeutet, einzelne Merkmalsdimensionen von Staat-Kirche-Beziehungen sollten nicht mehrmals erfasst werden und etwaige Subdimensionen sollten dem richtigen Überkonzept zugeordnet sein.

Eine diesen Kriterien entsprechende Konzeptlogik weisen die Messkonzepte von Chaves und Cann (1992), Pollack (2003) und Grim und Finke (2006) auf. Während sich die beiden erstgenannten ohnehin nur auf eine einzige Dimension von Staat-Kirche-Beziehungen fokussieren und damit von vornherein wenig anfällig für Redundanz oder Vermischung sind, stellen die drei von Grim und Finke (2006) berücksichtigten Merkmalsdimensionen klar getrennte Aspekte dar und sind auch auf derselben allgemeinen Abstraktionsebene angesiedelt.

Als problematisch hinsichtlich der Redundanz der erfassten Dimensionen erweisen sich die von Fox (2006, 2008) vorgeschlagenen *Religion and the State* Indizes. Das Merkmal der staatlichen Restriktion von Religion wird hier danach ausdifferenziert, welche religiösen Gruppen (alle oder einzelne, Mehrheit oder Minderheit) jeweils davon betroffen sind. Eine solche Bildung von Subdimensionen ist theoretisch relevant und sinnvoll, etwa um die staatliche Diskriminierung religiöser Minderheiten erfassen zu können. Im vorliegenden Fall führt sie jedoch zur Überschneidung der Merkmale offizielle Feindseligkeit gegenüber Religion, generelle Restriktionen und Regulierung von Religion: „Unlike the *official hostility* variable, which measures whether states are hostile to all religions, the *general restrictions* variable measures whether the state is hostile to any religion. [...] The *religious regulation* variable measures the extent to which the government monitors, restricts, and regulates the majority religion or all religions in a state“ (Fox 2008, S. 50). Durch die mangelnde Trennschärfe werden hier prinzipiell gleiche Aspekte mehrfach erfasst, was zu einer gewissen Redundanz führt. Demgegenüber leidet der *Religious Freedom Index* von Norris und Inglehart (2004) aufgrund der fehlenden Ausdifferenzierung eines Konzeptbaumes an der Vermischung der Dimensionen staatlicher Restriktion und staatlicher Unterstützung von Religion, welche hier in eins fallen.

3.2 Die Phase der Messung

Der zweite Schritt einer jeden Indexkonstruktion besteht darin, die in der Konzeptspezifikation festgelegten Merkmalsdimensionen zu operationalisieren, indem ihnen empirisch beobachtbare Sachverhalte (Indikatoren) zugeordnet werden (Diekmann 2001; Miller 2007; Schnell et al. 2005). In dieser Phase sind Munck und Verkuilen (2002, S. 15) zufolge insgesamt drei methodische Aufgaben zu bewältigen: die *Auswahl der Indikatoren*, die *Festlegung der Messniveaus* sowie die *Dokumentation des Kodiervorganges* (s. Tab. 3). Das Beurteilungskriterium für die ersten beiden Herausforderungen ist das der *Validität*, also das Ausmaß in dem tatsächlich das gemessen wird, was gemessen werden soll. Das Kriterium der dritten Aufgabe ist jenes der *Reliabilität* und *Replizierbarkeit*, also das Ausmaß in dem eine wiederholte Verkodung prinzipiell möglich ist und auch zu denselben Resultaten führt.

Tab. 3: Die Phase der Messung

Aufgabe/ Bewertungskriterium	Chaves und Cann (1992)	Pollack (2003)	Norris und Inglehart (2004)	Grim und Finke (2006)	Fox (2006)
<i>2.1 Auswahl der Indikatoren</i>					
Validität der Indikatoren	⊕ (Allerdings unklar was Ernennung von Geistlichen durch den Staat eigentlich bedeutet)	⊕ (Bezahlung von Kirchenpersonal nicht erfasst, Indikator für finanzielle Unterstützung zu grob)	⊕ (Einige Indikatoren messen etwas anderes, z. B. Einschätzung durch US Regierung, Staatliche Beilegung von Konflikt; Redundanz: Religionsfreiheit in Verfassung doppelt erfasst)	⊕ (Redundanz: Schutz und Gewährleistung von Religionsfreiheit doppelt erfasst, ungleiche Subventionierung und Finanzierung doppelt erfasst)	⊕ (Redundanz: optionaler und verpflichtender Religionsunterricht doppelt erfasst, Verminderung bei Restriktion von Konversion)
Breite der Quellenbasis	⊕ (Barrett 1982 validiert und ergänzt durch weitere Quellen)	⊕ („diverse Quellen“, aber nicht genannt welche)	⊕ (2002 <i>International Religious Freedom Report</i>)	⊕ (2003 <i>International Religious Freedom Report</i>)	⊕ (<i>International Religious Freedom Reports</i> , Barrett et al. 2001 und andere Quellen)
<i>2.2 Auswahl des Messniveaus</i>					
(Messniveau)	(Kategorial)	(Ordinal)	(Kategorial)	(Ordinal)	(Ordinal und kategorial)
Theoretische Begründung des Messniveaus	⊕	⊕	⊕	⊕	⊕ (Überlegungen zu ordinal vs. kategorial vorhanden)
<i>2.3 Dokumentation des Kodiervorgangs</i>					
Kodierregeln	⊕	⊕ (Unklar wann auf den einzelnen Items ein oder zwei Punkte vergeben werden)	⊕	⊕	⊕
Quellenangabe	⊕	⊕	⊕	⊕	⊕
Reliabilitätstests zwischen Kodierern	⊕	⊕	⊕	⊕	⊕
Disaggregierte Daten	⊕	⊕	⊕	⊕	⊕

⊕ Kriterium erfüllt, ⊕ Kriterium teilweise erfüllt, ⊕ Kriterium nicht erfüllt

Die erste Entscheidung in der Phase der Messung betrifft die *Auswahl valider Indikatoren*. Als praktische Richtlinie für diesen Schritt geben Munck und Verkuilen (2002, S. 15–17) erstens die Verwendung multipler Indikatoren an, welche zugleich auch die Vergleichbarkeit über kulturelle Kontexte hinweg sicherstellen müssen. Zweitens sollen Forscher Messfehlern vorbeugen, indem solche Indikatoren ausgewählt werden, die wenig anfällig für potentielle Verzerrungen sind und sich möglichst anhand mehrerer Quellen kreuzvalidieren lassen.

In Anlehnung an Müller und Pickel (2007, S. 526) soll hier davon ausgegangen werden, dass sich die kulturelle Äquivalenz der Indikatoren nur in konkreten Anwendungen testen lässt. Daher soll lediglich die Konsistenz der Indikatoren mit dem abzubildenden Konzept überprüft werden. Die Frage, ob die Indikatoren richtig messen, wird dann danach bewertet, ob sie erstens alle Merkmale des Konzepts erfassen, ob sie zweitens tatsächlich das messen, was sie messen sollen, ob sie drittens Redundanzen und Vermischungen vermeiden und schließlich, ob es sich überhaupt um einen Indikator im Sinne des beobachtbaren Sachverhalts handelt (Müller und Pickel 2007, S. 526).

Der einzige Index bei dem es keinerlei größere Bedenken bzgl. der Validität der Indikatoren gibt, ist die *Religious Regulation Scale* von Chaves und Cann (1992). Allerdings ist selbst hier nicht ganz klar, ob das Item „the state appoints or approves the appointment of church leaders“ tatsächlich als Indikator für die staatliche Unterstützung von Religion interpretiert werden kann, oder ob es sich nicht vielmehr um eine Form staatlicher Einmischung in religiöse Belange handelt. Bei den übrigen Messkonzepten lassen sich dagegen stärkere Mängel erkennen. Der *Index Kirche-Staat-Beziehung* von Pollack (2003) etwa verpasst es, den entscheidenden Aspekt der staatlichen Finanzierung Geistlicher und des Kirchenpersonals zu berücksichtigen. Der Indikator „Steuerliche Begünstigung der Kirchen, finanzielle Unterstützung“ erscheint zu grob und vermischt möglicherweise zwei getrennte Aspekte staatlicher Förderung von Religion. Die *Religious Freedom Scale* von Norris und Inglehart (2004) wiederum beinhaltet Indikatoren, welche etwas anderes als staatliche Regulierung von Religion messen. Dabei handelt es sich zum einen um einen Indikator über staatliches Versagen bei der Verhinderung von ethno-religiösen Konflikten und zum anderen um einen Indikator, ob die US-Regierung die Religionsfreiheit im gegebenen Land bedroht sieht. Weiterhin weist die Skala dahingehend Redundanzen auf, dass der Aspekt, ob die Verfassung Religionsfreiheit restringiert oder gewährleistet, doppelt erfasst wurde. Redundanzen und Vermischung lassen sich ebenfalls in den *International Religion Indexes* von Grim und Finke (2006) und den *Religion and the State* Indizes von Fox (2006, 2008) auffinden. Erstere berücksichtigen etwa den Schutz und die Gewährleistung der Religionsfreiheit durch den Staat sowie die ungleiche Subventionierung und Förderung von Religion doppelt. Fox (2008) erfasst das Vorhandensein optionalen und verpflichtenden Religionsunterrichts zweifach und ordnet einen Indikator zur Restriktion von Konversion einer Dimension zur staatlichen Unterstützung von Religion zu.

Inwieweit die Indikatoren in den verschiedenen Messansätzen anhand mehrerer Quellen validiert wurden, wird in Anlehnung an die Kodierregeln von Müller und Pickel (2007, S. 526) anhand der *Breite der Quellenbasis* beurteilt.⁹ Sowohl Chaves und Cann

9 Müller und Pickel (2007, S. 521) unterscheiden hierbei in Anlehnung an Lauth (2004) genauer, ob die Forscher mehrere Quellen herangezogen haben, ob darunter sowohl Quellen staatlicher

(1992) als auch Fox (2006, 2008) erfüllen dieses Kriterium, da sie neben Angaben aus der *World Christian Encyclopedia* (Barrett 1982; Barrett et al. 2001) noch weitere Quellen zur Kodierung der einzelnen Indikatoren heranziehen. Das Kriterium *nicht* erfüllen dagegen die Messkonzepte von Norris und Inglehart (2004) sowie von Grim und Finke (2006), da sich beide bei der Indikatorenbildung jeweils nur auf einen *International Religious Freedom Report* aus einem einzigen Jahr stützen. Die Bewertung des von Pollack (2003) vorgeschlagenen *Index Kirche-Staat-Beziehung* ist dahingehend problematisch, als er zwar durchaus diverse Quellen verwendet, diese jedoch nicht kenntlich gemacht werden. Das Kriterium wird aus diesem Grund als nur teilweise erfüllt bewertet.

Nachdem die Indikatoren ausgewählt wurden gilt es, sich für ein *adäquates Messniveau* zu entscheiden, auf welchem diese gemessen werden sollen. Munck und Verkuilen (2002, S. 17) zufolge sollten Forscher hierbei sicherstellen, dass die Messung insofern valide ist, als einerseits nicht feiner gemessen wird als überhaupt realistisch ist, andererseits aber die verschiedenen Messklassen tatsächlich möglichst gleiche Fälle beinhalten und damit auch nicht zu grob sind. In Anlehnung an Müller und Pickel (2007, S. 527) wird daher für jedes der fünf Messkonzepte überprüft, inwiefern eine *theoretische Begründung* bzw. überhaupt ein Problembewusstsein für das gewählte Skalenniveau vorhanden sind.

Dabei zeigt sich, dass dieser wichtige Aspekt der Indexkonstruktion in den hier betrachteten Bemühungen weithin unreflektiert bleibt. Der einzige Autor, der das Messniveau seiner Variablen explizit diskutiert, ist Fox (2008, S. 49), welcher etwa mit Blick auf sein *Official Government Involvement* Maß festhält, es sei „important to emphasize that this variable is ordinal and not categorial. Each category is a higher level of establishing a religion than the previous one“. Unabhängig davon, ob man diese Einschätzung im Einzelnen teilt, erfüllt er damit das Bewertungskriterium.

Zuletzt besteht ein wichtiger Bestandteil der Messphase in der ausreichenden *Dokumentation der Kodierregeln und des Kodiervorgangs* (Munck und Verkuilen 2002, S. 18). Nur so kann die Reliabilität und Replizierbarkeit des Messvorgangs gewährleistet und sichergestellt werden, dass eine Wiederholung des Kodierprozesses durch andere Forscher möglich ist und auch zu den gleichen Ergebnissen führt. Die Erfüllung dieses Beurteilungskriteriums wird daran festgemacht, ob in den einzelnen Indexkonstruktionen die Kodierregeln offengelegt, die verwendeten Quellen angegeben, Tests für die Interkoderreliabilität durchgeführt wurden und schließlich, ob die Daten in disaggregierter Form zur Verfügung stehen (Müller und Pickel 2007, S. 527).

Alle genannten Punkte einer gründlichen Dokumentation erfüllen lediglich die *International Religion Indexes* von Grim und Finke (2006) und das *Religion and the State* Projekt von Fox (2006, 2008). Der Kodiervorgang ist in beiden Fällen sehr transparent und nachvollziehbar. Zu der von Norris und Inglehart (2004) konstruierten *Religious Freedom Scale* und der *Religious Regulation Scale* von Chaves und Cann (1992) wurden dagegen keine Reliabilitätstests durchgeführt. Beide Indizes sind nicht in disaggregierter Form vorhanden, wobei bei Chaves und Cann (1992) die Skalenpunkte der einzelnen Länder

und nicht-staatlicher Autoren berücksichtigt wurden und schließlich ob die Quellen aus unterschiedlichen nationalen bzw. kulturellen Hintergründen stammen. Ich erfasse diesen Aspekt größer, indem ich nur betrachte, ob mehrere Quellen verwendet wurden oder nicht – unabhängig von ihrem jeweiligen Hintergrund.

sogar mühsam aus einer Graphik erschlossen werden müssen. Pollack (2003) weist zwar die Werte der einzelnen Komponenten des *Index Kirche-Staat-Beziehung* für jedes Land aus, doch ist völlig unklar auf welcher Basis diese Werte vergeben und v. a. welche Quellen dazu verwendet wurden. Reliabilitätstests wurden ebenfalls nicht berichtet.

3.3 Die Phase der Aggregation

Im dritten und letzten Schritt der Indexkonstruktion gilt es, den vorangegangenen Prozess der Aufgliederung des Konzepts in einzelne, beobachtbare Bestandteile umzukehren und diese wieder zu einem Ganzen zusammenzufügen, d. h., die Indikatoren zu einem Gesamtmaß zu aggregieren (Diekmann 2001; Miller 2007; Schnell et al. 2005). Nach dem von Munck und Verkuilen (2002, S. 22–23) vorgeschlagenen Beurteilungskatalog stellen sich dabei drei methodische Aufgaben: erstens die *Auswahl des Aggregationsniveaus*, zweitens die *Festlegung der Aggregationsregel* und schließlich drittens die *Dokumentation des Aggregationsvorgangs* (s. Tab. 4). Abermals ist in den ersten beiden Punkten die Validität des Messkonzepts zu gewährleisten, während im dritten Punkt die Kriterien der Reliabilität und Replizierbarkeit sichergestellt werden müssen (vgl. Müller und Pickel 2007, S. 527).

Die *Angemessenheit des Aggregationsniveaus* lässt sich daran bemessen, inwieweit eine brauchbare Balance zwischen einerseits einem hohen Aggregationsniveau – und damit einem möglichst sparsamen, die Vergleichbarkeit garantierenden Maß – und andererseits einem niedrigen Aggregationsniveau – also einem differenzierten Maß mit möglichst wenig Informationsverlust – gefunden wird (Munck und Verkuilen 2002, S. 22). In Anlehnung an Müller und Pickel (2007, S. 523) soll das Beurteilungskriterium im Folgenden als erfüllt gelten, wenn die beiden widersprechenden Ziele der Vergleichbarkeit und Abbildungsgenauigkeit dahingehend erreicht werden, als mehrdimensionale Konzepte durch entsprechende Mehrskalendizes abgebildet werden.

Dies ist in der Tat für die Mehrzahl der hier evaluierten Messkonzepte von Staat-Kirche-Beziehungen der Fall. Unproblematisch sind dabei die *Religious Regulation Scale* (Chaves und Cann 1992) und der *Index Kirche-Staat-Beziehung* (Pollack 2003), welche ohnehin nur auf die Merkmalsdimension der staatlichen Förderung von Religion abzielen und somit richtigerweise aus einem entsprechenden Gesamtmaß in Form eines Einkalenindex bestehen. Das Messkonzept der *International Religion Indexes* (Grim und Finke 2006) wiederum basiert auf insgesamt drei Skalen, welche die drei theoretischen Dimensionen der staatlichen Förderung, der staatlichen Restriktion sowie der sozialen Regulierung abbilden. Die Dimensionalität wurde dabei auf vorbildliche Weise – und im Gegensatz zu allen anderen betrachteten Messkonzepten – eigens mittels konfirmatorischen Faktorenanalysen überprüft und empirisch bestätigt. Fox (2006, 2008) gelingt die Balance in seinem *Religion and the State* Projekt insofern, als sein Messkonzept zunächst auf fünf verschiedenen Indizes beruht, welche verschiedene Subdimensionen von staatlich-kirchlicher Verflechtung messen und damit eine hohe Abbildungsgenauigkeit gewährleisten. Zugleich stellt er aber noch einen hoch aggregierten Index für „Overall Government Involvement in Religion“ bereit, der diese fünf Skalen zu einem einzigen Gesamtmaß integriert. Auf diese Weise wird er beiden Anforderungen gerecht.

Tab. 4: Die Phase der Aggregation

Aufgabe/Bewertungskriterium	Chaves und Cann (1992)	Pollack (2003)	Norris und Ingelhart (2004)	Grim und Finke (2006)	Fox (2006)
<i>3.1 Auswahl des Aggregationsniveaus</i>					
Angemessenheit des Aggregationsniveaus	☺	☺	☺ (Zu hohes Aggregationsniveau, da Einskalenindex trotz zwei theoretischen Dimensionen)	☺	☺
<i>3.2 Auswahl der Aggregationsregel (Regel)</i>					
Theoretische Begründung	(Additiv, ungewichtet) ☺	(Additiv, ungewichtet) ☺	(Reskaliert, additiv, ungewichtet) ☺	(Reskaliert, additiv, ungewichtet) ☺ (Überlegungen zur Aggregation vorhanden)	(Additiv, ungewichtet, teilweise reskaliert) ☺ (Überlegungen zur Aggregation vorhanden)
<i>3.3 Dokumentation des Aggregationsvorgangs</i>					
Aggregationsregel und Anwendbarkeit	☺ (Regel bekannt aber ohne Daten nicht anwendbar)	☺	☺ (Regel bekannt aber ohne Daten nicht anwendbar)	☺	☺

☺ Kriterium erfüllt, ☹ Kriterium teilweise erfüllt, ☹ Kriterium nicht erfüllt

Lediglich die von Norris und Inglehart (2004) konstruierte *Religious Freedom Scale* ist auf einem zu hohen Aggregationsniveau angesiedelt. Sie ist als Einskalenindex angelegt, der die beiden theoretisch intendierten Dimensionen staatlicher Unterstützung sowie staatlicher Restriktion von Religion in ein einziges Maß zwängt – womit die Validität der Messung geschwächt wird.

Die zweite Aufgabe in der Aggregationsphase besteht in der *Auswahl der Aggregationsregel*, also der Rechenoperation, durch welche sich der Index aus den einzelnen Indikatoren berechnen lässt. Nach den Bewertungsmaßstäben von Munck und Verkuilen (2002, S. 24) gilt es hier zu gewährleisten, dass sich das theoretische Zusammenspiel der einzelnen Bestandteile adäquat im gewählten Aggregationsverfahren widerspiegelt – etwa nur dann addiert wird, wenn es theoretisch plausibel scheint, dass die Komponenten von gleicher Bedeutung für das zu messende Konstrukt sind. Dieses Kriterium soll im folgenden als erfüllt gelten, wenn die Autoren der hier betrachteten Indizes eine explizite theoretische Begründung für die angewandte Aggregationsregel liefern und damit (unabhängig von der tatsächlichen Angemessenheit) methodisches Problembewusstsein demonstrieren (vgl. Müller und Pickel 2007, S. 527).

Ein Blick auf die hier zu analysierenden Messkonzepte macht jedoch deutlich, dass ein solches Problembewusstsein nicht sehr stark entwickelt ist. Alle Messvorschläge zur empirischen Erfassung von Staat-Kirche-Beziehungen konstruieren einen Index durch das ungewichtete Aufaddieren der einzelnen Indikatoren – wobei letztere in einigen Fällen zuvor noch zum Zwecke der Gleichgewichtung reskaliert werden (Norris und Inglehart 2004; Grim und Finke 2006 und teilweise Fox 2006, 2008). Eine explizite Begründung für dieses methodische Vorgehen wird aber nur in zwei Fällen geliefert. Lediglich bei Grim und Finke (2006) und bei Fox (2006, 2008, 2011) findet sich eine Diskussion zu den Vor- und Nachteilen der Gewichtung einzelner Komponenten.

Zuletzt gilt wie schon bei der Messung, dass der *Vorgang der Aggregation ausreichend dokumentiert* sein muss, um die Nachvollziehbarkeit und damit die Replizierbarkeit zu gewährleisten. Dies erfordert ganz konkret, dass einerseits die Aggregationsregel offenlegt wird und andererseits Daten zu den einzelnen Komponenten veröffentlicht und zugänglich gemacht werden, um die Aggregationsregel überhaupt anwendbar zu machen (Munck und Verkuilen 2002, S. 25; Müller und Pickel 2007, S. 528).

Dieses allerletzte Evaluationskriterium wird immerhin von drei der fünf evaluierten Indexkonstruktionen erfüllt. Vorbildlich sind in dieser Hinsicht die Messkonzepte von Grim und Finke (2006) und Fox (2006, 2008), welche außer einer genauen Dokumentation der verwendeten Rechenregeln v. a. Daten zu den einzelnen Bestandteilen im Netz öffentlich und frei zugänglich bereitstellen. Bei Pollack (2003) lassen sich die einzelnen Komponenten und ihre Ausprägung immerhin aus einer Tabelle erschließen. Dagegen sind bei der *Religious Regulation Scale* von Chaves und Cann (1992) sowie der *Religious Freedom Scale* von Norris und Inglehart (2004) zwar jeweils die Aggregationsregeln ausbuchstabiert worden, doch lassen sich diese mangels disaggregierter Daten nicht anwenden und replizieren.

4 Metaevaluierung der Staat-Kirche-Indizes

In diesem Abschnitt soll der bisherige Vergleich der Indizes zur Messung von Staat-Kirche-Beziehungen zusammengefasst und eine Metaevaluierung vorgenommen werden. Dabei sind zwei prinzipielle Vorgehensweisen denkbar (vgl. Müller und Pickel 2007, S. 528). Zum einen lassen sich in einem *parallelen Vergleich* die Evaluationsergebnisse der einzelnen Indizes in Bezug zueinander setzen und auf diese Weise eine Rangfolge der Messansätze zu Staat-Kirche-Verflechtungen ermitteln. Eine solche Rangfolge könnte dann bei der Auswahl eines Indexes im Rahmen thematischer Anwendungen als Orientierung dienen. Weiterhin besteht eine zweite Möglichkeit im *Quervergleich* der Leistungen der Indizes in den einzelnen methodischen Schritten der Indexkonstruktion und damit in einer Einschätzung der allgemeinen Stärken und Schwächen gegenwärtiger Messansätze insgesamt. Daraus ließen sich insbesondere Lehren für die zukünftige Verbesserung und Weiterentwicklung von Messungen im Bereich der Staat-Kirche-Beziehungen ableiten. Im Folgenden werden beide Varianten verfolgt.

4.1 Rangfolge der Indizes zur Messung von Staat-Kirche-Beziehungen

Um eine Rangfolge der Leistungsfähigkeit und Messqualität der einzelnen Indizes zu erstellen, bietet es sich an, in Anlehnung an Müller und Pickel (2007) einen *Metaindex* zu konstruieren, welcher die Erfüllung der einzelnen Kriterien in einer Maßzahl quantifiziert und somit einen systematischen Vergleich zwischen den Messkonzepten ermöglicht. Um auf möglichst unkomplizierte und transparente Weise zu einem solchen Gesamtmaß zu kommen, wird zunächst für jedes erfüllte Bewertungskriterium ein Punkt vergeben. Ist ein Kriterium dagegen nur teilweise erfüllt, wird entsprechend nur ein halber Punkt vergeben. Auf Basis der ermittelten Gesamtpunktzahl lassen sich die Indizes dann miteinander vergleichen (s. Tab. 5).

Selbstverständlich müsste nun aber eigentlich auch der vom Autor konstruierte Metaindex jenen strengen methodologischen Anforderungen genügen, die er bei anderen so dringend einfordert – indem er etwa die Gewichtung der Evaluationskriterien theoretisch begründet oder auch Reliabilitätstests für die Verkodung der Kriterien berichtet. Insbesondere die Frage der Gewichtung der einzelnen Kriterien – deren konkrete Operationalisierung darüber hinaus u. U. streitbar ist – scheint hier von zentraler Bedeutung, da diese die Gesamtbeurteilung der Indizes möglicherweise in erheblichem Maße beeinflussen oder gar verzerren kann.

Während Müller und Pickel (2007, S. 537) auf eine Gewichtung der Kriterien gänzlich verzichten, da ihnen eine solche „nur möglich erscheint, wenn hohe Begründungskosten in Kauf genommen werden,“ schlage ich eine andere Lösung vor. Im Rahmen einer *Sensitivitätsanalyse* sollen verschiedene plausible Gewichtungsoptionen für den Metaindex miteinander kontrastiert und auf diese Weise geprüft werden, inwiefern sich verschiedene Prioritätensetzungen bei der Bewertung der Messansätze auf das Gesamturteil über die Indizes auswirken. Da darüber hinaus die Bewertung der einzelnen Evaluationskriterien offengelegt wurde, steht es weiterhin jedem Interessenten zusätzlich frei, eine den eigenen Bedürfnissen und Präferenzen entsprechende Gewichtung vorzunehmen und auf

Tab. 5: Metaevaluation der Indizes zur Messung von Staat-Kirche-Beziehungen

	Chaves und Cann (1992)	Pollack (2003)	Norris und Inglehart (2004)	Grim und Finke (2006)	Fox (2006)
<i>1. Konzeptualisierung</i>					
<i>1.1 Identifizierung der Merkmale</i>					
Sparsamkeit	1	1	1	0	1
Relevanz	0	0	1	1	1
<i>1.2 Logische Anordnung der Merkmale</i>					
Redundanz	1	1	1	1	0
Zuordnung/Vermischung	1	1	0,5	1	1
<i>2. Messung</i>					
<i>2.1 Auswahl der Indikatoren</i>					
Validität der Indikatoren	1	0	0	0,5	0,5
Breite der Quellenbasis	1	0,5	0	0	1
<i>2.2 Auswahl des Messniveaus</i>					
Theoretische Begründung	0	0	0	0	1
<i>2.3 Dokumentation der Kodierregeln und des Kodiervorgangs</i>					
Kodierregeln	1	0	1	1	1
Quellenangabe	1	0	1	1	1
Reliabilitätstests zwischen Kodierern	0	0	0	1	1
Disaggregierte Daten	0	1	0	1	1
<i>3. Aggregation</i>					
<i>3.1 Auswahl des Aggregationsniveaus</i>					
Angemessenheit des Aggregationsniveaus	1	1	0	1	1
<i>3.2 Auswahl der Aggregationsregel</i>					
Theoretische Begründung der Aggregationsregel	0	0	0	1	1
<i>3.3 Dokumentation des Aggregationsvorgangs</i>					
Aggregationsregel und Anwendbarkeit	0	1	0	1	1
<i>Gesamt</i>	8	6,5	5,5	10,5	12,5
<i>Anteil erfüllter Bewertungskriterien (%)</i>	57,1	46,2	39,3	75,0	89,3

diese Weise zu einem eigenen Urteil über die relativen Vorzüge der einzelnen Indizes zu gelangen.¹⁰

Die erste und einfachste Gewichtungsoption besteht darin, die insgesamt 14 kodierten Evaluationskriterien einfach ohne jegliche Gewichtung aufzuaddieren und ihnen allen damit die gleiche Bedeutung für die Bewertung der Indizes einzuräumen. Nach einem

¹⁰ Außerdem lädt der Autor ausdrücklich dazu ein, den Kodiervorgang für die hier präsentierten Messkonzepte zu wiederholen, um auf diese Weise zu einem Maß für die Interkoderreliabilität zu gelangen.

auf solche Weise berechneten Metaindex lassen sich die *Religion and the State* Indizes von Fox (2006, 2008) als insgesamt bestes Messkonzept zur Erfassung von Staat-Kirche-Beziehungen ausmachen. Von insgesamt 14 möglichen Punkten erzielt dieser Messansatz immerhin 12,5 Punkte. Er erfüllt damit etwa 89% aller Bewertungskriterien und kann als durchaus vorbildliches Beispiel für eine gelungene Indexkonstruktion bezeichnet werden. Dicht darauf folgt das von Grim und Finke (2006) entwickelte Messkonzept der *International Religion Indexes*, welches mit 10,5 Punkten immer noch 75% der Kriterien einer methodisch korrekten Indexkonstruktion erfüllt. Die von Chaves und Cann (1992) ursprünglich vorgeschlagene *Religious Regulation Scale* wiederum erreicht mit einer Punktzahl von 8 nur etwas mehr als die Hälfte (57%) der Bewertungskriterien. Den beiden übrigen Messkonzepten, welche explizit an die *Religious Regulation Scale* anknüpfen und diese weiterentwickeln, ist es – unter rein methodologischen Gesichtspunkten – nicht gelungen, das Original zu übertreffen. So erzielt etwa der von Pollack (2003) konstruierte *Index Kirche-Staat-Beziehung* nur 6,5 von 14 möglichen Punkten und genügt damit insgesamt weniger als der Hälfte der methodischen Anforderungen (46%). Das Schlusslicht bildet schließlich die an Chaves und Cann (1992) angelehnte *Religious Freedom Scale* von Norris und Inglehart (2004), welche lediglich 39% der Evaluationskriterien erfüllt (5,5 von 14 Punkten) und insgesamt am schlechtesten abschneidet.

Inwiefern verändert sich aber diese Gesamtbeurteilung und damit die Rangfolge der verschiedenen Messinstrumente zur Erfassung von Staat-Kirche-Beziehungen, wenn alternative Gewichtungen vorgenommen werden? Im Folgenden werden neben der Gleichgewichtung aller Kriterien insgesamt fünf weitere Gewichtungsoptionen für den Metaindex miteinander verglichen (Abb. 1).¹¹ Ein Problem der Gleichgewichtung besteht darin, dass für die verschiedenen Aufgaben der Indexkonstruktion z. T. unterschiedlich viele Bewertungskriterien vorliegen und auf diese Weise Aufgaben, welche anhand mehrerer Kriterien operationalisiert werden – etwa die Dokumentation der Kodierregeln (Aufgabe 2.3) – auch stärker in die Gesamtbewertung einfließen, als Aufgaben, die lediglich anhand eines einzelnen Kriteriums beurteilt werden – etwa die Auswahl des Messniveaus (Aufgabe 2.2). Dies ist natürlich nicht unproblematisch.

Eine mögliche Gewichtungsoption bestünde daher darin, anstelle aller Evaluationskriterien alle sieben *Aufgaben* gleichzugewichten, indem ihnen jeweils einheitlich ein Gewicht von eins zugewiesen wird. Wie in Abb. 1 ersichtlich ist, ergeben sich bei dieser Gewichtungsoption (O) im Vergleich zum ungewichteten Metaindex (*) jedoch nur leichte Unterschiede in der Gesamteinschätzung der relativen Messgüte der verschiedenen Indizes. Lediglich der *Index Kirche-Staat-Beziehung* von Pollack (2003) würde deutlich besser bewertet und mit der *Religious Regulation Scale* von Chaves und Cann (1992) gleichziehen. Die relative Bewertung der anderen Messkonzepte würde sich dagegen nicht verändern.

Eine andere Bestrebung könnte sein, alle *Phasen* der Indexkonstruktion gleichermaßen in die Gesamtbewertung einfließen zu lassen und diese daher jeweils auf den Wert eins zu setzen, um die unterschiedliche Anzahl an Aufgaben und Kriterien auszugleichen (Δ in

11 Um einen solchen Vergleich zu gewährleisten wurden die verschieden gewichteten Metaindizes normalisiert, also in die gleiche Skala transformiert, ohne dabei jedoch die relativen Skalenabstände zu beeinträchtigen.

Abb. 1). Wenngleich die generelle Rangfolge gleich bleibt, so würde diese Gewichtungsoption abermals den Index von Pollack (2003) der Skala von Chaves und Cann (1992) gleichsetzen. Außerdem würden die *International Religion Indexes* von Grim und Finke (2006) etwas näher an die von Fox (2006) eingenommene Spitzenposition heranrücken. Eine Variation dieses spezifischen Metaindexes, welche der Phase der Messung – und damit der Schwachstelle der meisten der hier betrachteten Indizes – stärkeres Gewicht einräumt und diese daher doppelt gewichtet (∇ in Abb. 1), kommt in der Gesamtbewertung wiederum zu demselben Schluss wie ein völlig ungewichteter Index.

Weiterhin sind feingliedrigere Gewichtungspräferenzen denkbar. Würde man etwa lediglich die Evaluationskriterien der Phase der Messung doppelt gewichten, um diesem Aspekt der Indexkonstruktion besondere Wichtigkeit zu verleihen (in Abb. 1), würden sich die Ansätze von Chaves und Cann (1992) sowie von Grim und Finke (2006) in der Gesamtbewertung leicht annähern. Der Index von Pollack (2003) würde dagegen in der Beurteilung etwas nach unten abrutschen. Ein Metaindex, welcher demgegenüber der Validität der verwendeten Indikatoren große Priorität einräumt und dieses Kriterium mit dem Faktor drei eingehen lässt, jedoch die Dokumentation des Kodiervorgangs als unwichtiger erachtet und allen vier Kriterien lediglich das gemeinsame Gewicht eins zuweist, würde die *Religious Regulation Scale* (Chaves und Cann 1992) und die *International Religion Indexes* (Grim und Finke 2006) gleichbewerten (in Abb. 1). Das obere und untere Ende der Rangfolge bliebe abermals unverändert.

4.2 Die Messung von Staat-Kirche-Beziehungen im Quervergleich

Die vorangegangene Analyse der einzelnen Indizes hat bereits deutlich gemacht, dass die Indexkonstruktion im Bereich der empirischen Erfassung von Staat-Kirche-Beziehungen in mancher Hinsicht zwar durchaus gute Ergebnisse erzielt, gleichzeitig aber auch z. T. methodische Defizite aufweist. In diesem abschließenden Analyseschritt werden nun in einem *Quervergleich* der Stand der gegenwärtigen Messung von Staat-Kirche-Beziehungen und ihre Stärken und Schwächen in der Gesamtschau betrachtet. Ganz allgemein gesprochen zeigt sich dabei, dass keine der insgesamt drei Phasen der Indexkonstruktion

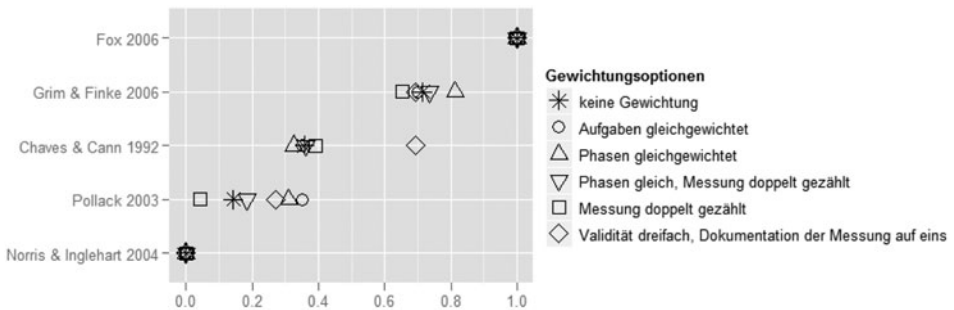


Abb. 1: Sensitivitätsanalyse unterschiedlicher Gewichtungen. (Gesamtbeurteilung der Indizes mit sechs verschiedenen Gewichtungen der Evaluationskriterien; die verschiedenen Gewichtungsoptionen werden im Text beschrieben; zum Zweck der Vergleichbarkeit wurden sie jeweils normalisiert)

– weder Konzeptualisierung, Messung noch Aggregation – gänzlich frei von Mängeln ist. Keine einzige der innerhalb dieser drei Phasen jeweils zu bewältigenden, methodischen Aufgaben ist von allen Messkonzepten zufriedenstellend erfüllt worden. Gleichzeitig ist aber auch festzuhalten, dass die einzelnen Schritte der Indexkonstruktion durchaus unterschiedlich stark mit Mängeln behaftet sind und dass es v. a. kein Bewertungskriterium gibt, welches nicht doch zumindest in einem Fall erfolgreich gemeistert wurde. Es besteht also durchaus die Möglichkeit, von guten Beispielen zu lernen.

Am insgesamt unproblematischsten gestaltet sich ganz offensichtlich die *Phase der Konzeptualisierung*. Von den von den fünf betrachteten Indizes insgesamt zu erzielenden ($5 \times 4 =$) 20 Punkten konnten immerhin 15,5 Punkte bzw. rund 78% eingefahren werden. Die explizit oder implizit erfolgende Orientierung an dem zweidimensionalen Konzept von Staat-Kirche-Verflechtung wie es in der *economics of religion* Schule vertreten wird, hat sich dabei als nützlicher Ausgangspunkt erwiesen, um auf sehr sparsame Weise alle relevanten Dimensionen zu erfassen und diese auch ohne große Probleme logisch anzuordnen. Voraussetzung dafür ist freilich, dass man sich auch tatsächlich daran orientiert und nicht etwa wie Chaves und Cann (1992) oder Pollack (2003) – obgleich explizit auf die Überprüfung von Thesen der Religionsökonomie abzielend – nur die Dimension der staatlichen Unterstützung von Religion berücksichtigt, die Dimension staatlicher Restriktion dagegen vernachlässigt, oder aber im Gegenteil dem Konzept gar noch weitere Dimensionen hinzufügt (z. B. Grim und Finke 2006). Letzteres birgt durchaus die Gefahr, Redundanzen bei der Anordnung des Konzeptbaums zu erzeugen, wie im Falle von Fox (2006, 2008) geschehen.

Allerdings ist es nicht ohne weiteres gelungen, die insgesamt gute Konzeptualisierungsleistung in der *Phase der Messung* fortzuführen und dort überzeugend umzusetzen. Tatsächlich handelt es sich bei diesem entscheidenden Schritt um den größten Schwachpunkt der gegenwärtigen Messpraxis im Bereich der Staat-Kirche-Indizes überhaupt. Von insgesamt ($5 \times 7 =$) 35 möglichen Punkten haben es die hier betrachteten Messkonzepte auf lediglich 18,5 Punkte gebracht. Damit wurden gerade einmal die Hälfte (53%) der methodischen Kriterien für eine valide und reliable Messung erfüllt.

Schon die *Auswahl valider Indikatoren* stellt sich als defizitär dar. Die häufigsten Probleme bestehen dabei einerseits in der fehlerhaften Zuordnung von einzelnen Indikatoren zu theoretischen Dimensionen – es wird also oftmals schlicht nicht das gemessen, was zu messen vorgegeben wird (z. B. Norris und Inglehart 2004). Andererseits ist v. a. die doppelte Erfassung der gleichen Indikatoren weit verbreitet, was zu einer impliziten Gewichtung führt und ebenfalls die Validität der Messung beeinträchtigt (Norris und Inglehart 2004; Grim und Finke 2006; Fox 2006, 2008). Gleichzeitig fällt eine weit verbreitete Unbedarftigkeit bzgl. der *Auswahl des Messniveaus* auf. Eine theoretische Begründung für das gewählte Messniveau liefert lediglich Fox (2008). Bei den restlichen Messkonzepten wird dagegen unreflektiert auf kategorialem oder ordinalem Skalenniveau gemessen. Zuletzt ist mit Blick auf die *Kodierregeln und den Kodiervorgang* zu konstatieren, dass es im Bereich der Messung von Staat-Kirche-Verflechtungen offenbar nicht selbstverständlich ist, dass die Kodierung der Indikatoren auf ihre Interkoderreliabilität hin überprüft wird. Diese Maßnahme zur Gewährleistung der Messgüte wurde nur in zwei Fällen erwähnt (Grim und Finke 2006; Fox 2008).

Die *Phase der Aggregation* wurde von den Autoren der hier evaluierten Messkonzepte methodisch insgesamt wieder etwas besser bewältigt. Dennoch sind offensichtliche Mängel nicht zu übersehen. In der Tat wurden nur 9 von ($5 \times 3 =$) 15 Kriterien und damit 60 % erfolgreich erfüllt, was auf einen Spielraum für Verbesserungsanstrengungen hindeutet. Am deutlichsten fällt auch in dieser Phase wieder das Fehlen einer *theoretischen Begründung für die Auswahl der Aggregationsregel* ins Auge. Nur in zwei der berücksichtigten Fälle lassen sich Überlegungen darüber finden, wie die einzelnen Indikatoren am besten zu einem Gesamtmaß zusammengefügt werden können und eine Begründung für das gewählte Verfahren gegeben werden kann (Grim und Finke 2006; Fox 2006). In den übrigen Fällen wird fraglos auf die Konstruktion eines ungewichteten Summenindex zurückgegriffen.

5 Zusammenfassung

Der vorliegende Beitrag hat sich zum Ziel gesetzt, neuere Indizes zur Messung von Staat-Kirche-Beziehungen vorzustellen und diese in einem kritischen Vergleich auf ihre methodologischen Stärken und Schwächen hin zu untersuchen. Zu diesem Zweck wurde auf ein von Munck und Verkuilen (2002) im Rahmen der Demokratiemessung vorgeschlagenes und von Müller und Pickel (2007) präzisiertes Evaluationsschema zurückgegriffen und auf einen alternativen Anwendungskontext übertragen. Anhand eines methodologischen Kriterienkatalogs lassen sich damit Messkonzepte zur empirischen Erfassung von Staat-Kirche-Verflechtungen hinsichtlich ihrer methodischen Leistungsfähigkeit bei Konzeptualisierung, Messung und Aggregation beurteilen.

Die Relevanz eines solchen methodenkritischen Unterfangens liegt zunächst v. a. darin begründet, dass letztlich nur ein methodisch sorgfältiges Vorgehen auch valide Indikatoren und verlässliche Daten für die vergleichende Erforschung von Staat-Kirche-Beziehungen liefern kann. Gerade für die Theorieentwicklung in diesem erst am Anfang stehenden Bereich der vergleichenden Politikwissenschaft ist dies von großer Bedeutung – nur gute Messungen liefern auch überzeugende Ergebnisse und Befunde. Angesichts der Dringlichkeit, mit denen sich religionspolitische Fragen in gegenwärtigen demokratischen Gesellschaften unter den Bedingungen eines wachsenden religiösen Pluralismus stellen, gewinnt ein solches Bemühen ganz praktische Relevanz (Bader 2007).

Der kritische Vergleich der Indizes offenbart in der Tat z. T. deutliche Unterschiede in der methodischen Sorgfalt im Rahmen der Indexkonstruktion und damit einhergehend Differenzen hinsichtlich ihrer messtheoretischen Güte. Dabei haben sich insbesondere die Indizes des von Fox (2006, 2008) initiierten *Religion and the State* Projekts als methodologisch überzeugendes Instrumentarium zur empirischen Erfassung von Staat-Kirche-Beziehungen herausgestellt, welches als durchaus vorbildliches Beispiel für eine methodisch saubere Indexkonstruktion gelten kann und darüber hinaus durch seinen hohen Differenzierungsgrad und seine breite empirische Reichweite überzeugt.

Wenngleich die hier geleistete Bewertung der verschiedenen Messkonzepte natürlich Orientierungswissen für die Anwendung im Rahmen thematischer Fragestellungen bereitstellen möchte, so soll dennoch von expliziten Anwendungsempfehlungen für oder gegen den einen oder anderen Index abgesehen werden. Wie eine Sensitivitätsanalyse verdeut-

licht hat, ergeben sich je nach Gewichtung der einzelnen Evaluationskriterien z. T. andere Schlussfolgerungen über die Gesamtbewertung einzelner Messkonzepte. Insbesondere die Urteile über die relativen Vorzüge der *Religious Regulation Scale* (Chaves und Cann 1992), dem *Index Kirche-Staat-Beziehung* (Pollack 2003) oder der *International Religion Indizes* (Grim und Finke 2006) hängen durchaus maßgeblich von solchen Gewichtungspferenzen ab. Vielmehr soll dem Anwender durch die Offenlegung der methodischen Stärken und Schwächen der einzelnen Messansätze ein reflektierter Umgang mit diesen Instrumenten und eine den eigenen Bedürfnissen entsprechende Wahl erleichtert werden.

Zusammenfassend lässt sich für die gegenwärtige Messpraxis im Bereich der vergleichenden politikwissenschaftlichen Forschung zu Staat-Kirche-Verflechtungen sagen, dass hier gute Konzeptualisierungsleistungen mit eindeutigen Umsetzungsschwächen in den darauf aufbauenden Schritten der Messung und Aggregation einhergehen. An diesen kritischen Stellen herrscht ein oftmals unbedarfter Gebrauch von Standardlösungen vor, welcher kaum methodenkritisch reflektiert wird. Dieser Befund eines entwicklungsfähigen Methodenbewusstseins deckt sich weitgehend mit jenen im Bereich der vergleichenden Demokratieforschung gemachten (Munck und Verkuilen 2002; Müller und Pickel 2007), so dass dies möglicherweise auf ein generelles Problem gegenwärtiger Messpraxis in der vergleichenden Politikwissenschaft hindeutet, in dessen Behandlung zukünftig stärker investiert werden sollte.

Das hier herangezogene Evaluationsschema hat sich in diesem Zusammenhang als nützliches Werkzeug zur Aufdeckung von Stärken und Schwächen erwiesen und kann für ein solches Unterfangen als hilfreiche Orientierung dienen. Da es sich aber stets selbst an den eigenen Ansprüchen messen und kritisch hinterfragen lassen muss, stellt es darüber hinaus eine geeignete Grundlage für die generelle methodologische Verständigung innerhalb der vergleichenden Politikwissenschaft dar.

Literatur

- Bader, Veit. 2007. *Secularism or democracy? Associational governance of religious diversity*. Amsterdam: Amsterdam University Press.
- Barrett, David B. 1982. *World christian encyclopedia*. Oxford: Oxford University Press.
- Barrett, David B., George T. Kurian, und Todd M. Johnson. 2001. *World christian encyclopedia*. 2. Aufl. Oxford: Oxford University Press.
- Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, und Bernhard Weßels. 2008. Wie lässt sich Demokratie am besten messen? Zum Forumsbeitrag von Thomas Müller und Susanne Pickel. *Politische Vierteljahresschrift* 49:114–122.
- Casanova, José. 1994. *Public religions in the modern world*. Chicago: University of Chicago Press.
- Chaves, Mark, und David E. Cann. 1992. Regulation, pluralism, and religious market structure: Explaining religion's vitality. *Rationality and Society* 4:272–290.
- Chaves, Mark, Peter J. Schraeder, und Mario Sprindys. 1994. State regulation of religion and muslim religious vitality in the industrialized West. *Journal of Politics* 56:1087–1097.
- Diekmann, Andreas. 2001. *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek: Rowohlt.
- Elkins, Zachary. 2000. Gradations of democracy? Empirical tests of alternative conceptualizations. *American Journal of Political Science* 44:287–294.

- Enyedi, Zsolt, und John T. S. Madeley, Hrsg. 2003. *Church and state in contemporary Europe. The chimera of neutrality*. London: Frank Cass.
- Finke, Roger. 1997. The consequences of religious competition: Supply-side explanations for religious changes. In *Rational choice theory and religion. Theory and assessment*, Hrsg. Lawrence A. Young, 46–56. London: Routledge.
- Fox, Jonathan. 2006. World separation of religion and state into the 21st century. *Comparative Political Studies* 39:537–569.
- Fox, Jonathan. 2008. *A world survey of religion and the state*. Cambridge: Cambridge University Press.
- Fox, Jonathan. 2009. Quantifying religion and state: Round two of the religion and state project. *Politics and Religion* 2:444–452.
- Fox, Jonathan. 2011. Building Composite Measures of Religion and State. *Interdisciplinary Journal of Research on Religion* 7: Article 8.
- Freedom House. 2010. *Freedom in the world 2010*. Freedom House.
- Gaber, Rusanna. 2000. Demokratie in quantitativen Indizes: Ein mehr- oder eindimensionales Phänomen? In *Demokratiemessung. Konzepte und Befunde im internationalen Vergleich*, Hrsg. Hans-Joachim Lauth, Gert Pickel und Christian Welzel, 112–131. Wiesbaden: Westdeutscher.
- Gill, Anthony. 2001. Religion and comparative politics. *Annual Review of Political Science* 4:117–138.
- Grim, Brian J., und Roger Finke. 2006. International religion indexes: Government regulation, government favoritism, and social regulation of religion. *Interdisciplinary Journal of Research on Religion* 2:2–40.
- Grim, Brian J., und Roger Finke. 2007. Religious Persecution in Cross-National Context: Clashing Civilizations or Regulated Religious Economies. *American Sociological Review* 72: 633–658.
- Grim, Brian J., und Roger Finke. 2011. *The Price of Freedom Denied. Religious Persecution and Conflict in the 21st Century*. New York: Cambridge University Press.
- Iannaccone, Laurence R. 1991. The consequences of religious market structure: Adam Smith and the economics of religion. *Rationality and Society* 3:156–177.
- Kaina, Viktoria. 2008. Die Messbarkeit von Demokratiequalität als ungelöstes Theorieproblem. Zum PVS-Forums-Beitrag von Marc Bühlmann, Wolfgang Merkel, Lisa Müller und Bernhard Weßels. *Politische Vierteljahresschrift* 49:518–524.
- Lauth, Hans-Joachim. 2004. *Demokratie und Demokratiemessung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Liedhegener, Antonius. 2008. Religion in der Vergleichenden Politikwissenschaft: Begriffe – Konzepte – Forschungsfelder. In *Der Begriff der Religion*, Hrsg. Mathias Hildebrandt und Manfred Brocker, 179–196. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Manow, Philip. 2008. *Religion und Sozialstaat. Die konfessionellen Grundlagen europäischer Wohlfahrtsstaatsregime*. Frankfurt a. M.: Campus.
- Miller, Bernhard. 2007. Maßvoll messen: Zur konzeptorientierten Entwicklung von Messinstrumenten. In *Forschungsdesign in der Politikwissenschaft. Probleme – Strategien – Anwendungen*, Hrsg. Thomas Gschwend und Frank Schimmelfennig, 123–148. Frankfurt a. M.: Campus.
- Minkenberg, Michael. 2002. Religion and public policy. Institutional, cultural, and political impact on the shaping of abortion policies in Western democracies. *Comparative Political Studies* 35:221–247.
- Minkenberg, Michael. 2003. Staat und Kirche in westlichen Demokratien. In *Politik und Religion (Sonderheft 32 der Politischen Vierteljahresschrift)*, Hrsg. Michael Minkenberg und Ulrich Willems, 115–138. Wiesbaden: Westdeutscher.
- Minkenberg, Michael, und Ulrich Willems. 2002. Neuere Entwicklungen im Verhältnis von Politik und Religion im Spiegel politikwissenschaftlicher Debatten. *Aus Politik und Zeitgeschichte* B42–43:6–14.

- Minkenber, Michael, und Ulrich Willems. 2003. Politik und Religion im Übergang – Tendenzen und Forschungsfragen am Beginn des 21. Jahrhunderts. In *Politik und Religion (Sonderheft 32 der Politischen Vierteljahresschrift)*, Hrsg. Michael Minkenber und Ulrich Willems, 13–41. Wiesbaden: Westdeutscher.
- Monsma, Stephen V., und Christopher J. Soper. 1997. *The challenges of pluralism. Church and state in five democracies*. Lanham: Rowman & Littlefield.
- Müller, Thomas, und Susanne Pickel. 2007. Wie lässt sich Demokratie am besten messen? Zur Konzeptqualität von Demokratie-Indizes. *Politische Vierteljahresschrift* 48:511–539.
- Müller, Thomas, und Susanne Pickel. 2008. Antwort auf die Replik von Marc Bühlmann, Wolfgang Merkel, Lisa Müller und Bernhard Weßels zum Forumsbeitrag von Thomas Müller und Susanne Pickel. *Politische Vierteljahresschrift* 49:123–126.
- Munck, Gerardo L., und Jay Verkuilen. 2002. Conceptualizing and measuring democracy. Evaluating alternative indices. *Comparative Political Studies* 35:5–34.
- Norris, Pippa, und Ronald Inglehart. 2004. *Sacred and secular. Religion and politics worldwide*. Cambridge: Cambridge University Press.
- Pickel, Susanne, und Gert Pickel. 2006. *Politische Kultur- und Demokratieforschung. Grundbegriffe, Theorien, Methoden. Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Pollack, Detlef. 2003. *Säkularisierung – ein moderner Mythos? Studien zum religiösen Wandel in Deutschland*. Tübingen: Siebeck.
- Pollack, Detlef, und Gert Pickel. 2000. *The vitality of religion-church integration and politics in Eastern and Western Europe in comparison*. Discussion Paper Frankfurter Institut für Transformationsstudien 13/00. Frankfurt/Oder.
- Robbers, Gerhard. Hrsg. 1995. *Staat und Kirche in der Europäischen Union*. Baden-Baden: Nomos.
- Roßteutscher, Sigrid. 2009. *Religion, Zivilgesellschaft, Demokratie. Eine international vergleichende Studie zur Natur religiöser Märkte und der demokratischen Rolle religiöser Zivilgesellschaften*. Baden-Baden: Nomos.
- Schnell, Rainer, Paul B. Hill, und Elke Esser. 2005. *Methoden der empirischen Sozialforschung*. München: Oldenbourg.
- Stark, Rodney, und Roger Finke. 2000. *Acts of faith: Explaining the human side of religion*. Berkeley: University of California Press.
- Traunmüller, Richard. 2010. Segen oder Fluch? Zum Einfluss von Staat-Kirche-Beziehungen auf die Vitalität religiöser Zivilgesellschaften im europäischen Vergleich. In *Religion zwischen Zivilgesellschaft und politischem System*, Hrsg. Antonius Liedhegener und Ines Werkner, 138–161. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Traunmüller, Richard, und Markus Freitag. 2011. State support of religion: Making or breaking faith-based social capital? *Comparative Politics* 43:253–269.
- Traunmüller, Richard. 2012. Religion und Sozialkapital. Ein doppelter Kulturvergleich. Wiesbaden: Springer VS.
- van Kersbergen, Kees, und Philip Manow. Hrsg. 2009. *Religion, class coalitions, and welfare states*. New York: Cambridge University Press.
- Welzel, Christian. 2000. Humanentwicklung und Demokratie: Welcher Index erfasst die „humane“ Dimension der Demokratie am besten? In *Demokratiemessung. Konzepte und Befunde im internationalen Vergleich*, Hrsg. Hans-Joachim Lauth, Gert Pickel und Christian Welzel, 132–162. Wiesbaden: Westdeutscher.