

### Rezension: David Colton & Robert W. Covert, 2007: Designing and constructing instruments for social research and evaluation

Fuchs, Marek

Veröffentlichungsversion / Published Version

Rezension / review

**Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:**

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Fuchs, M. (2008). Rezension: David Colton & Robert W. Covert, 2007: Designing and constructing instruments for social research and evaluation. [Rezension des Buches *Designing and constructing instruments for social research and evaluation*, von D. Colton, & R. W. Covert]. *Methoden, Daten, Analysen (mda)*, 2(1), 75-79. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-128473>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

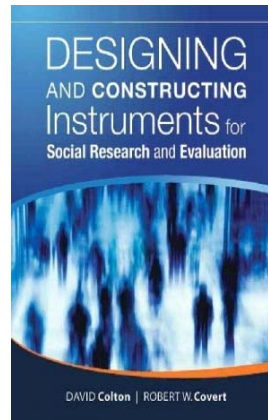
This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

und hinter die Ohren schreiben: Es gibt keinen Grund für irgendeinen „qualitativen“ induktivistisch-theoriefreien Dualismus auf der einen Seite und es gibt ebenso keinen Grund für die Annahme eines „quantitativen“ deduktivistischen Universalismus kausaler Gesetze auf der anderen. Es gibt einige Besonderheiten der Sozialwissenschaften, die es erzwingen, sich oftmals sehr genau die historisch und sozial sehr spezifischen (Mikro-)Verhältnisse anzusehen, aber das heißt noch lange nicht, dass damit auch die allgemeinen Regeln der Theoriebildung und -prüfung, wie sie für alle Wissenschaften gelten, etwa die Überprüfung der Geltungreichweite eines Zusammenhangs oder die Annahme von Kausalbeziehungen auch bei der Erklärung von Handlungen, außer Kraft gesetzt sind. Dem ist nichts hinzuzufügen, und man muss Udo Kelle sehr dankbar sein, dass er die Grenzen der beiden Extrempositionen und die Unhaltbarkeit der damit verbundenen Dualismen überzeugend begründet und unvoreingenommen und ausbalanciert vorträgt. Die Integration der qualitativen und der quantitativen Verfahren fällt dann wie eine reife Frucht vom Baume – und mancher wird sich fragen: warum gab es denn eigentlich darüber mal einen Streit?

HARTMUT ESSER, MANNHEIM

\* \* \* \* \*



DAVID COLTON and ROBERT W. COVERT, 2007: *Designing and Constructing Instruments for Social Research and Evaluation*. 1. Auflage. San Francisco: Jossey-Bass, 412. Seiten, ISBN: 978-0-7879-8784-8, 44,99 EUR.

Der Band tritt mit einem ambitionierten Titel an, der einen an der Umfrageforschung interessierten Leser zunächst zu der Annahme führt, es handele sich um ein Buch, in dem – zumindest auch – ausführlich über die Fragebogenkonstruktion informiert wird. Bei genauerer Inspektion zeigt sich jedoch, dass es sich eher um ein Einführungslehrbuch in die Methoden der empirischen Sozialforschung mit Schwerpunkt Instrumentenentwicklung (Beobachtungsbögen, diagnostische Instrumente, Fragebögen, Check-Listen, Tests, Beurteilungsbögen) handelt. Das Buch gliedert sich in 3 Hauptabschnitte, die mit „Grundlagen“, „Anwendung“ und „Organisation und Durchführung“ überschrieben sind, die intern in 15 Unterkapitel untergliedert sind:

Das einleitende Kapitel 1 liefert zunächst einige konzeptionelle Überlegungen zu den charakteristischen Differenzierungen verschiedener Arten von Instrumenten und ihrer Bestandteile. Hier wird auch der Prozess der Konstruktion eines Instruments entworfen, der im Verlauf des Bandes immer wieder aufgegriffen wird. In dieser Betonung des Prozesshaften liegt sicher eine der Stärken des Bandes, indem nicht nur die aus wissenschaftstheoretischer Sicht zu lösenden Aufgaben diskutiert, sondern auch die notwendigen Arbeitsschritte – und

dabei auftretende Fallstricke – thematisiert werden. Das zweite Kapitel stellt die Rolle der Instrumente und der Instrumentenentwicklung in den Kontext eines Forschungsprojekts. Dabei werden sehr kursorisch die Grundstrukturen quantitativer und qualitativer Forschungsmethodologie präsentiert, bevor im anschließenden Kapitel 3 die Skalenniveaus vermittelt werden.

Auf dem Niveau eines Einführungslehrbuchs wird im anschließenden Kapitel 4 ein Überblick über Validitäts- und Reliabilitätskonzepte sowie über Methoden zur Überprüfung und Sicherstellung der beiden Gütekriterien gegeben. Einzelne Verweise auf weiterführende Literatur und Ressourcen reichen das Kapitel an, bevor es im darauffolgenden Kapitel 5 um die Explikation der Ziele der Studie geht, in deren Rahmen ein Instrument zu entwickeln ist. Diese Explikation soll den Forscher zu einer Vergegenwärtigung der zu messenden Konzepte und der Untersuchungseinheiten/Merkmalsträger anregen. Außerdem werden Entscheidungen über die Art der Administration nahegelegt. Um diese Explikation der Ziele anzuregen und herbeizuführen, werden Techniken für eine systematische Generierung von kreativen Ideen diskutiert, unter anderem: Brainstorming, Snowballing und Pyramiding, Delphie-Methode, was wiederum die sozialen Entwicklungsprozesse eines Instruments betont.

Im sechsten Kapitel werden Methoden des Pre-Testings diskutiert – was systematisch eigentlich nach den im Weiteren beschriebenen Arbeitsschritten liegt. Zudem bleiben die meisten Standardmethoden unberücksichtigt und werden z. T. nicht einmal genannt (Behavior-Coding, Interaction-Coding, Random-Probe usw.). Auch die einschlägige Standardliteratur wird nur spärlich erwähnt. Dafür unternehmen die Autoren einen interessanten Versuch, die Pre-Test-Methoden nach ihrer Stellung im Entwicklungsprozess eines Instruments zu ordnen: In der

Vorbereitungsphase, in der sich die Zielsetzung der Studie herauschält und präzisiert wird, in der eigentlichen Entwicklungsphase des Instruments, in der verschiedene Entwurfsfassungen schrittweise verfeinert und optimiert werden, und schließlich in der Testphase des vorläufig fertigen Instruments. Von daher erklärt sich dann auch, dass diese Methoden relativ früh in den Entwicklungsprozess eines Instruments eingeführt werden.

Jedoch krankt dieses Kapitel an der theoretischen Einbettung der Pre-Test-Methoden: Zwar wird zunächst auf im Prozess der Instrumententwicklung unentdeckte Validitätsprobleme verwiesen, dann aber folgt eine stichpunktartige Zusammenstellung von bekannten „Antwortfehlern“ (Context-Effekte, Antwortreihenfolgeneffekte, Effekte von sensitiven Items, Non-Attitude, Antwortinkonsistenzen und viele andere bekannte Effekte), deren Auftreten als Begründung für die Notwendigkeit eines Pre-Tests angeführt wird. Hier wird offensichtlich, dass die „kognitionspsychologische Wende“, die seit etwa 20 Jahren die methodische Forschung zu den verschiedenen Facetten des Measurement-Error wesentlich vorangebracht hat, in diesem Buch nur wenig Berücksichtigung gefunden hat.

Das anschließende Kapitel 7 bietet eine schöne, pragmatische Einführung in die Gestaltung von geschlossenen Antwortvorgaben. Einige basale Regeln, eine Systematisierung in Anlehnung an den Band von Fink und viele Beispiele von Standardskalen werden geboten. Allerdings fehlen wiederum die umfänglich existierenden Belege aus der experimentellen Forschung für die gegebenen Empfehlungen. Außerdem reicht manche „Guideline“ einfach zu kurz oder ist nur schwer nachvollziehbar: „By convention, response scales are ordered from low to high“ (S. 156). Weiter wird die Frage der numerischen Relative diskutiert, aber ohne Bezug zu den einschlägigen Texten, die z. B

belegen, dass diese die vom Befragten vorgenommene Verankerung der Skala zu verändern vermögen. Entscheidend ist aber, dass die Konstruktion der Antwortkategorien bei geschlossenen Fragen auf das Kriterium reduziert wird, welche Kategorien „most effective for capturing the data you need“ sind. Damit bleibt der zentrale Ansatz der sozialwissenschaftlichen Methodenforschung, der die Qualität einer Frage am Umfang der auftretenden Antwortfehler misst, unberücksichtigt. In Kapitel 8 folgt erneut eine Zusammenstellung typischer Probleme bei der Formulierung von Fragebogenfragen und Antwortkategorien. Dabei wird auf das Problem der sozialen Erwünschtheit und der sensitiven Fragen sowie auf Probleme der Bedeutungsäquivalenz bei interkulturellen Befragungen eingegangen. Hinzu kommen einige Regeln für die Formulierung von Antwortvorgaben bei Rating-Skalen. Daran schließt sich im Kapitel 9 eine Darstellung von Mehrfachnennungen und besonderen Formen der Einfachauswahl an, sowie eine Diskussion der Exklusivität der Antwortvorgaben.

Zwar treten an verschiedenen Stellen im Buch Redundanzen auf, die mit Blick auf selektive Leser, die nicht das ganze Buch zur Kenntnis nehmen, durchaus sinnvoll sein können. In den Kapiteln 7, 8 und 9 nehmen diese aber überhand. Eine systematischere Aufbereitung und Gliederung dieser drei für den Fragebogenkonstrukteur zentralen Kapitel hätten sicher zu einer gelungeneren, weniger redundanten Präsentation der an sich kenntnisreichen Darstellung führen können.

Im Kapitel 10 werden offenen Fragen und ihre Auswertung diskutiert, was zu einer relativ ausführlichen Passage über inhaltsanalytische Verfahren führt. Die Formulierung der offenen Frage wird kurz dargestellt; die aktuelle Diskussion über die „Visual Design Language“ der offenen Frage (u. a. Dillman, Smyth, Christian) wird hinge-

gen nicht aufgegriffen, auch klassische Studien zur unterschiedlichen Mächtigkeit von geschlossenen und offenen Items bleiben unberücksichtigt. Gelungener ist demgegenüber das anschließende Kapitel 11, in dem eine Beschreibung der Funktionsweise von verschiedenen Skalentypen geleistet wird: Neben dem semantischen Differenzial und der allfälligen Likert-Skala werden zwei Varianten von Thurstone und die Goal-Attainment-Scale (GAS) diskutiert. Die Darstellung gibt einen Einblick in die Grundzüge der Funktionsweise und der Konstruktion der jeweiligen Skala, der für den Leser dieses Einführungsbuches durchaus aufschlussreich ist.

Der abschließende dritte Teil des Buches beginnt mit einem Kapitel über die Zusammenführung verschiedener Items und Skalen zu einem Gesamtinstrument und zu seiner formalen Gestaltung. Hier werden Titel, Einleitungsstatement und Formulierungen zum Datenschutz diskutiert, aber auch – kurz – das Design von Filterinstruktionen in selbst-administrierten Fragebögen. Ausführlich gehen die Autoren auf die sozialstatistischen Fragen ein, wobei vor allem auch Ethnizität und ‚Race‘ behandelt werden. Außerdem wird auf die Reihenfolge von Fragetypen und Themen im Fragebogen eingegangen sowie – erneut – auf die Länge des Fragebogens. Die Frage des visuellen Designs wird nur auf wenigen Seiten abgehandelt und vor allem mit Bezug zur Schriftart und Schriftgröße.

Schließlich werden in Kapitel 13 die Administration der Instrumente getrennt für Beobachter-/Rater-/Interviewer-administrierte und selbst-administrierte Instrumente diskutiert. Dabei werden im Bereich der nicht selbst-administrierten Instrumente vor allem Beobachtungsbögen und das mit der Beobachtung einhergehende soziale Setting besprochen – interviewer-administrierte Befragungen kommen nur am Rande vor, eine ausführliche Diskussion des Modus der Admi-

nistration entsprechend ebenfalls nicht. Dafür wird über die Auswahl der Orte für eine Beobachtung, die Auswahl der Beobachteten und die Schulung der Beobachter gesprochen.

Im Weiteren wird dann die Durchführung einer selbst-administrierten Befragung thematisiert, wobei Stichprobenverfahren und Nonresponse problematisiert aber keinesfalls erschöpfend abgehandelt werden. In diesen Passagen wird deutlich, dass das Buch eben nicht nur den Konstruktionsprozess von Instrumenten beschreiben, sondern eher einen umfassenden Blick auf den Forschungsprozess in der empirischen Sozialforschung liefern will. Diesem überambitionierten Ziel kann der Band auf knapp 400 Seiten jedoch kaum gerecht werden. Diese Ausführungen über Stichproben und Feldarbeit eignen sich daher eher als Ausblick und Problemaufriss, allerdings fehlt für eine sinnvolle Weiterführung der Verweis auf die einschlägige Literatur, mit deren Hilfe die Leser die Probleme von Coverage, Sampling und Nonresponse gründlicher aufarbeiten könnte. Ebenfalls im Schlussteil des Bandes findet sich ein Kapitel (Kapitel 14) zum Einsatz von Computern bei der Erstellung, der Administration und Auswertung von Instrumenten. Diese sehr knappe Darstellung – Web Surveys zum Beispiel werden auf 1-2 Seiten abgehandelt – vermag immerhin eine gewisse Sensibilität für die Folgen der Computerisierung zu wecken; einige Vor- und Nachteile der Computer-Administration werden diskutiert. Allerdings fehlt der Verweis auf die Standardterminologie (CASI, CSAQ usw.) und auf die entsprechende Literatur. Das abschließende Kapitel 15 diskutiert Fragen des Datenzugangs und Copyright-Fragen bei von abhängig beschäftigten Mitarbeitern erstellen Instrumenten und gesammelten Daten. Schließlich werden einige Seiten auf die Berichterstattung und die Dokumentation der Studie verwendet. Auch hier gilt, dass dieses Kapitel mehr einem Pro-

blemaufriss gleicht, als dass es Antworten liefert.

Eine Gesamtwürdigung kommt zu einem uneinheitlichen Bild: Sehr erfreulich sind die vielen Beispielinstrumente im Text. Am Ende jedes Kapitels finden sich meist mehrere Instrumente aus allen Bereichen der empirischen Forschung (Beobachtung, psychologische Diagnose, Befragung), an denen jeweils verschiedene Aspekte oder Schritte der Instrumentenentwicklung verdeutlicht werden. In diesem Anwendungsbezug, der dem Leser die Möglichkeit gibt, die diskutierten Aspekte der Instrumententwicklung nachzuvollziehen, sind sicher die Stärken des Bandes zu verbuchen. Doch auch wenn diese Beispielfragebögen recht gut demonstrieren, was im Text erläutert wird, enthalten einige von ihnen Unzulänglichkeiten oder zumindest zweifelhafte Elemente, die bei einem am kognitionspsychologischen Konzept orientierten Vorgehen aufgefallen wären (z. B. die „not sure“ Option im Fragebogen auf Seite 124, unterschiedliche Spaltenbreite bei Matrixfragen auf den Seiten 294/295, alternierende Graustufen in aufeinanderfolgenden Spalten einer Matrixfrage auf Seite 339 – um nur einige dieser Stellen zu nennen). Mehrere kleine Flüchtigkeitsfehler (S. 174, 212, 209) z. B. bei der Exklusivität der Antwortkategorien, die ein erfahrener Fragebogenkonstrukteur als solche übersehen kann, bzw. die ihn nicht daran hindern, das, was im Beispiel verdeutlicht werden soll wahrzunehmen, sind nicht nur ärgerlich, sondern bedenklich, weil unerfahrene Leser diese Ungenauigkeiten möglicher Weise als intendiert missverstehen können.

Doch abgesehen von diesen Kleinigkeiten fällt die Würdigung des Bandes noch aus einem anderen Grund nicht rundum positiv aus: Während die internationale Methodenforschung die Fragebogenentwicklung in den letzten beiden Jahrzehnten definitiv von einer ‚Kunst‘ zu einer ‚Wissenschaft‘

weiterentwickelt hat, die auf experimentell gewonnenen Aussagen und theoretischen Modellen beruht, wird die Entwicklung eines Instruments von Colton und Covert eher als Gleichklang von ‚Kunst‘ und ‚Wissenschaft‘ diskutiert – mit einem beträchtlichen Schwerpunkt auf ‚Kunst‘. Auf dem Weg von der ‚Kunst‘ der Instrumententwicklung zur ‚Wissenschaft‘ ist dieser Text daher nur einen ersten Schritt gegangen. Die in fast allen Kapiteln bemühte Analogie zur Malerei und zur Komposition eines Gemäldes bekräftigt den Eindruck, dass die Instrumentenentwicklung eben noch wesentlich vom Genius des Konstrukteurs mitbedingt wird. Entsprechend finden sich explizit die Formulierungen „Purposeful Creativity“ (S. 97) und „The Art of Instrument Construction“ (S. 43), die nicht nur der manchmal überstrapazierten Gemälde-Analogie geschuldet sind, sondern tatsächlich Programm sind.

Die dem Buch zugrunde liegende Literatur enthält zwar einzelne Titel aus den Jahren 2006 und 2007; ein systematisches Review des State of the „Art“ scheint dem Buch aber nicht vorausgegangen bzw. bereits vor einigen Jahren durchgeführt worden zu sein. Z. T. werden einzelne wichtige Autoren genannt (Dillman, Fowler, Krosnick, Schwarz ...) andere werden übersehen (Tourangeau, Conrad, Couper ...). Daher entsteht der Eindruck, dass dieses Buch vor allem auf dem Erfahrungsschatz der Autoren basiert, der durch ausgewählte Publikationen abgesichert wurde. Positiv zu würdigen ist die Tatsache, dass manche abgelegene Quelle benannt und verarbeitet wird, die im Kanon der Instrumentenentwicklung und insbesondere der Fragebogenforschung bisher nicht berücksichtigt wird, aber durchaus interessante Teilbefunde – allerdings für den Experten – liefert (etwa zu den offenen Fragen).

Nach allem eignet sich das Buch nicht unbedingt als vertiefende Einführung in die Fragebogenkonstruktion. Es scheint eher für

eine Leserschaft geschrieben zu sein, die noch gar keine Erfahrungen mit der empirischen Sozialforschung hat. Das Fehlen der Bezüge zu der immens angewachsenen Literatur zum Total Survey Error, zum Measurement Error im Besonderen lässt einem mit dem Gefühl zurück, dass die Konstruktion eines Instruments zwar nicht mehr freihändig „aus dem Bauch heraus“ vorzunehmen sei, aber eben auch nicht auf der Basis der umfangreichen Fachliteratur zu verschiedenen Facetten des auf das Messinstrument zurückgehenden Messfehlers sowie der vorhandenen theoretischen Konzepte zu Integration und Erklärung dieser Effekte.

Positiv zu würdigen ist hingegen die Betonung des Prozesscharakters der Instrumentenentwicklung, die das iterative Vorgehen bei der Erstellung eines Fragebogens hervorhebt und betont. In diesem Zusammenhang soll explizit die allgegenwärtige Aufforderung zu reflektieren, zu überdenken und Transparenz herzustellen, gelobt werden. Damit kann dem Buch als Verdienst die Betonung des sozialen Entstehungsprozesses eines Instruments zugeschrieben werden, die in einer stärker wissenschaftstheoretisch oder kognitionspsychologisch zugeschnittenen Einführung sicher weniger sichtbar geworden wäre.

Die Stärken des Buchs liegen aber eindeutig darin, dass es den Anfänger in einer verständlichen Sprache einige Grundprobleme der Instrumentenentwicklung nahe bringt. Dadurch werden diese für die auftretenden Untiefen bei der Planung und Durchführung einer quantitativen empirischen Studie sensibilisiert. Die gegebenen Antworten bleiben aber hinter dem zurück, was man Personen, die erstmals eine Befragung durchführen, mit auf den Weg geben möchte.

MAREK FUCHS, KASSEL

\*\*\*\*\*