

万卷方法

调查问卷的设计与评估

IMPROVING SURVEY QUESTIONS:
DESIGN AND EVALUATION

[美]弗洛德·J.福勒 (Floyd J. Fowler, Jr.) 著

蒋逸民 等译



重庆大学出版社

<http://www.cqup.com.cn>

万卷方法

调查问卷的设计与评估

IMPROVING SURVEY QUESTIONS:
DESIGN AND EVALUATION

[美]弗洛德·J.福勒 (Floyd J. Fowler, Jr.) 著

蒋逸民 田洪波 陆利军

杨海燕 邓 蕾 耿 文

安 钊 陈加荣 郭纪生

译

重庆大学出版社

Authorized translation from the English language edition, entitled *Improving Survey Questions: Design and Evaluation*, by Floyd J. Fowler, published by Sage Publications, Inc., Copyright © 1995 by Sage Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher. CHINESE SIMPLIFIED language edition published by CHONGQING UNIVERSITY PRESS, Copyright © 2008 by Chongqing University Press.

调查问卷的设计与评估,作者:弗洛德·J.福勒。原书英文版由美国 Sage 出版公司出版。原书版权属美国 Sage 出版公司。

本书简体中文版专有出版权由美国 Sage 出版公司授予重庆大学出版社,未经出版者书面许可,不得以任何形式复制。

版贸渝核字(2006)第39号。

图书在版编目(CIP)数据

调查问卷的设计与评估/(美)福勒著;蒋逸民等译。

—重庆:重庆大学出版社,2010.9

(万卷方法)

书名原文:Improving Survey Questions: Design and Evaluation

ISBN 978-7-5624-5153-2

I. 调… II. ①福…②蒋… III. ①问卷法—设计②问卷法—评估 IV. C91-03

中国版本图书馆 CIP 数据核字(2009)第 180282 号

调查问卷的设计与评估

弗洛德·J.福勒, Jr. 著

蒋逸民 等译

责任编辑:雷少波 邓飞燕 罗杉 版式设计:雷少波

责任校对:任卓惠 责任印制:赵晟

*

重庆大学出版社出版发行

出版人:张鸽盛

社址:重庆市沙坪坝正街174号重庆大学(A区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址:<http://www.cqup.com.cn>

邮箱:fxk@cqup.com.cn(营销中心)

全国新华书店经销

重庆升光电力印务有限公司印刷

*

开本:940×1360 1/32 印张:6.75 字数:185千

2010年10月第1版 2010年10月第1次印刷

印数:1—4 000

ISBN 978-7-5624-5153-2 定价:25.00元

本书如有印刷、装订等质量问题,本社负责调换
版权所有,请勿擅自翻印和用本书
制作各类出版物及配套用书,违者必究

作者简介

弗洛德·J. 福勒 (Floyd J. Fowler, Jr.) 美国密歇根大学 (University of Michigan) 社会心理学博士。自 1971 年以来为马萨诸塞-波士顿大学调查研究中心的资深研究员, 并任该中心主任长达 14 年之久。福勒博士曾在哈佛公共健康学院教授调查方法 7 年, 目前是达特茅斯医学院的研究人员。他参与了许多调查项目, 涵盖了一系列广泛的学科领域, 如犯罪、健康、住房、交通、政府服务和宗教等。他的主要研究兴趣是调查研究的误差来源。他最近出版的著作探讨了如何降低与访员有关的误差以及改进调查问题的设计和评估。其主要著作包括: 《标准化调查访问》(*Standardized Survey Interviewing*, 与 T. W. Mangione 合作)、《调查研究方法》(*Survey Research Methods*, 3rd ed.) 等*。

* 这两本书的中文版均由重庆大学出版社出版。

作者前言

我相信,调查问题的设计和评估是当前改进调查研究最富成果的方法论领域。粗劣的问题设计随处可见,而改进问题设计是提高调查数据质量所能采取的最容易、最有成本效益的步骤之一。

本书的一个重要来源是我与密歇根大学(University of Michigan)查尔斯·F.坎内尔(Charles F. Cannell)合作的一个研究项目,该项目探讨了对调查问题进行预测验的不同方法。在这项研究中,我们从政府和学界主要调查所使用的问题中选取了一个样本作为我们评估工作的焦点。虽然我们的主要目的不是评估目前调查问题的质量,但意外地发现,一些著名的调查机构使用了相当多的不符合最低标准的问题。在我看来,当前的调查研究工作显然可以做得更加出色。在过去十年里,有关问题设计的研究明显增多了,问题评估的技术也提高了。虽然很多工作仍在进行,但有不少关于如何设计和评估问题的合理原则,而这正是本书的主题。

改进调查问题所面临的一个特殊挑战是:政治家、律师、经济学家、会计和新闻记者(只列举这些)等背景各异、缺乏特殊训练的人都在起草调查问题。每个人都以为他或她能写出好的调查问题。因此,本书是为所有想写调查问题的人以及所有想用调查结果的人而作的。人们只要受过良好的教育,就能阅读和领会本书的内容,而不必为统计问题和社会科学的术语所困扰。

本书最重要的部分大概就是强调问题评估的重要性。第5章论述了调查问题的预评估,这无疑是本书最重要的一章。目

打尽天下鸟”的诱惑。研究新手经常想用一次调查来收集许多方面的数据,问题覆盖面特别广,研究主题因此而变得模糊起来。

一般而言,调查问卷所收集的信息有六种类型:①当前行为,②过去行为,③对承担义务较大的问题的态度,④对承担义务较小的问题的态度,⑤未来的行为,⑥未来的态度。这六种类型的问题又可归纳为行为与态度两大类。当然,问卷也应该包括受访者人口统计学方面的信息,比如年龄、性别、民族、婚姻状况、教育水平、职业、收入、宗教信仰等,这些信息又称为“常见共变量”。行为和人口统计学信息是客观的问题,而感觉、态度和看法则是主观的问题。问题越敏感,问题的效度就越低。过去的行为由于涉及已经发生的事情,一般效度较高,而未来的态度由于涉及许多未知和不确定因素,效度通常较低。

问卷设计需要仔细、反复地推敲问题的措辞。问卷的问题应该是明晰和简洁的,与所设定的研究目标相一致。研究表明,问题的措辞会对回答问题产生很大的影响。如果要提高答案的准确度,在问题措辞方面的改进要远大于在抽样方面的改进。有试验表明,问题的模糊或不准确产生的误差可达20%~30%。

设计问卷还需要对问题本身进行评估。这一步骤是我国调查研究的一个薄弱环节。一方面,相当部分的研究者没有意识到问卷评估的必要性和重要性,另一方面,即使有人有这方面的意识,但囿于时间和资金,往往匆匆走过场,没有发挥问题评估本来应有的作用。近年来,西方研究者倾向于从认知和互动的视角来评估调查问题。他们认为,在调查之前,就应该弄清受访者是否能一致地理解问题?他们能否完成问题所交代的任务?访员能否逐字逐句地念出问题?这些都是起草高质量调查问卷的重要前提。

设计好的调查问卷是一项技术含量极高的专业活动。西方不少学者甚至为此耗尽了毕生的心血,有的学者的终身成就不过是设计了一份高质量的、广为使用的调查问卷。从专业的角度来看,什么是一份好的调查问卷?什么是一个好的调查问

题?如何设计一份好的调查问卷?如何设计一个好的问题?这些都是研究者正在探索的问题。美国著名调查研究方法专家福勒(Floyd J. Fowler, Jr.)试图系统地回答这些问题。他1995年出版了《调查问卷的设计与评估》一书。该书在西方出版后,备受瞩目,引用率极高。1999年我在香港大学读博士时,就参考了福勒的这本书。当时我要设计有关家庭环境的调查问卷。这本书给我的印象是,有关问题设计的原则简明扼要,所提供的案例丰富多样,说理透彻,针对性强,便于在实地调查中应用。

与其他有关调查方法的专著相比,这本十多万字的小册子并不算厚重。但是,它的内容却十分厚实,几乎涵盖了问卷设计的所有问题。比如它具体论述了如何起草调查问题并在经验的层面上加以评估,如何在撰写问题时选择合适的措辞,如何选择问题的格式,等等。该书专门用了两章的篇幅论述了如何撰写用来收集事实资料的问题以及如何撰写用来测量主观状态的问题。书中一一列举了“好的”和“坏的”调查问题的实例,并解释了“坏的”调查问题何以在一定程度上导致了有缺陷的答案。书中还讲解了如何对问卷问题进行评估,如何对问题理解和管理上的一致程度进行有效评估。可以毫不夸张地说,该书是社会科学研究者和调查工作者进行问卷设计的一本实用价值很大的案头参考书。此外,该书也可以帮助调查数据的使用者来正确地评估有关调查数据。

该书文字精娴,文风嶙峋,语意隽永,表达朴涩。翻译该书不是一件轻松的事情。该书由蒋逸民与华东师范大学社会学系的部分研究生共同翻译,前后四易其稿。翻译的最初分工是:前言,蒋逸民;第1章,蒋逸民;第2章,杨海燕;第3章,邓蕾;第4章,耿文、田洪波;第5章,安钊;第6章,陈加荣、郭纪生;第7章,徐士林;附录A,徐士林;附录B—D,周燕;索引,蒋逸民。翻译初稿完成后,由田洪波进行了初次校阅。陆利军后来对其中的第2,5,6章等进行了再校阅,并对第2,6章做了较大的修改。陈建国协助陆利军校阅了第2章的部分内容。最后,蒋逸民对全书进行了修改、润色和校订,几乎对全书进行了重新翻译和校对。炼娟、马菱、孟维岩、蒋玥通读了最后的译

稿,提出了不少好的建议。由于译者的水平和学识有限,书中必有不少舛漏之处,但文责自负,恳请同仁和读者批评指正。

最后,衷心感谢重庆大学出版社的雷少波先生。他为本书的翻译提供了机会,始终关心本书的进展,正是在他的大力支持和帮助下,本书才得以面世。

蒋逸民

2010年1月于上海

译者前言

随着我国社会经济的快速发展,学术研究、政府机关、企事业单位和大众传媒等部门使用调查问卷的机会越来越多,然而,人们却不同程度地存在着对问卷的误解和误用。没有做过调查的人认为,起草一份问卷就像写文章一样容易。一些专业人士也认为,可以不费力气地在一个小时内搞定一份问卷。在他们眼中,问卷设计和调查似乎不需要任何专门的学习或培训,任何人都可以一试身手。于是,在设计问卷过程中,想到哪里就写到哪里,简单地罗列一些信手拈来的问题,既不系统,也不规范,遗漏的问题也不少,有些问题看似相关但实际上并不需要,也不能发挥什么作用。尽管设计问卷需要写作和常识等基本技能,但是作为一门科学的问卷设计和调查却远不止这些,需要更多、更深刻的东西。

问卷设计有着极其严格的操作程序。它不仅需要清晰地界定所研究的问题,仔细推敲问题的措辞,而且还需要对问题的信度和效度进行必要的评估。但是,在实际调查过程中,这些流程和细节往往被忽视或轻视。有些教科书对这些具体的操作步骤也是语焉不详。

设计问卷的第一步是界定研究问题。通过清晰、简要地界定问题来避免人们理解上的歧义。如果问题的范围和内涵界定不清,那么调查结果也会含糊不清。同样,也需要对所研究的事实、行为和态度进行清晰、简要的操作定义。没有经过界定的事实、行为和态度,就会产生含糊不清和难以解释的结果。清晰地界定研究问题可以廓清研究者的任务,从而避免“一石

前大多数调查机构对调查问题的评估还相当粗糙。我相信,如果第 5 章所概括的一些协议(protocols)成为调查研究的例行工作,就会极大地提高调查研究的质量。因此,如果我对该书还有什么最大希望的话,那就是,在调查中使用问题之前,该书对提高问题评估的水平发挥了作用。

杰克·福勒

目 录

| | |
|-----------------------|----|
| 1 作为测量的问题:概述 | 1 |
| 什么是好问题 | 1 |
| 影响测量的“问与答”特征 | 2 |
| 问题的评估 | 4 |
| 本书的构架 | 6 |
| 2 设计问题来收集事实资料 | 8 |
| 问题的目标 | 9 |
| 概念和术语的界定 | 13 |
| 知道答案与记得答案 | 20 |
| 答案的形式 | 27 |
| 减少社会想望对答案的影响 | 28 |
| 结 论 | 45 |
| 3 用来测量主观状态的问题 | 46 |
| 对人、地、事的描述和评估 | 47 |
| 测量对观点的回答 | 62 |
| 测量知识 | 68 |
| 多项目测量 | 70 |
| 有关主观状态答案的相对性 | 73 |
| 结 论 | 77 |
| 4 设计好的调查工具的一般原则 | 79 |
| 问什么问题 | 79 |
| 问题的措词 | 85 |

| | |
|------------------------------|------------|
| 调查问卷的格式 | 93 |
| 培训受访者 | 100 |
| 结 论 | 103 |
| 5 问题的预调查评估 | 105 |
| 专题小组讨论 | 106 |
| 深度个人访谈 | 111 |
| 田野预调查 | 116 |
| 在自填式问卷中调整预调查策略 | 130 |
| 列表回答 | 133 |
| 结 论 | 136 |
| 6 评估调查问题的效度 | 138 |
| 研究相关的类型 | 139 |
| 用记录来验证 | 141 |
| 比较替代问题的形式 | 145 |
| 用一致性作为效度指标 | 147 |
| 结 论 | 148 |
| 7 问题设计和评估:回顾与结论 | 150 |
| 有关事实的问题 | 151 |
| 测量主观状态 | 151 |
| 测试问题 | 152 |
| 评估问题的效度 | 153 |
| 结 论 | 154 |
| 附 录 | 156 |
| 附录 A 常见测量的维度 | 156 |
| 附录 B 常见协变量测量 | 168 |
| 附录 C 关于开放式问题 | 180 |
| 附录 D 达成标准化访谈 | 184 |
| 参考文献 | 189 |
| 索 引 | 194 |

作为测量的问题：概述

什么是好问题

在社会科学界,重要的测量是以“问与答”过程为基础的,医学界更是如此。儿童回答标准化测试问题,其结果被用作对他们智力的测量;失业率是对“就业状况”和“从业意愿”答案的计算;对医疗结果的研究依赖于病人对其健康状况和生活质量等问题的回答。

仅仅通过向人们询问问题而获得所需和有用的信息,几乎无尽无休。在某些情况下,这是因为我们想要知道某些很难系统观察到的事实。例如,虽然有些犯罪报了警,但还有许多犯罪没有报警。因此,估算犯罪受害者比率的最佳办法是向受害者询问他们的受害经历。我们也经常想测量“人们在想什么”、“知道什么”或“感觉到什么”等只有个体本身才能察觉的一些现象。

显然,“问与答”是日常会话的一部分,也是我们社会生活结构的一部分。但是,本书特别的主题是:如何将日常过程转化为严格的测量。

把问题的答案用作测量,这一想法有以下几个含义:首先,我们对受访者出于私利的回答不感兴趣。相反,我们却对受访者给我们讲述的其他事情感兴趣。因此,一个好的“问与答”过程的重要标准是:它引出了答案,这些答案能提供有关我们所

设法描述东西的有意义信息。

其次,测量的目的通常是获得许多人或事的可比较信息。因此,当重复测量时,测量结果的前后一致性就显得格外重要。

这些观点引出了对“什么问题才算是一个好的问题”的思考和回答。所谓好的问题是指这样的问题:它所引出的答案是对我们想要描述的东西可靠的、有效的测量。

在经典心理测量的意义上,信度是指答案的一致性程度:当所描述的状态一致时,答案也是一致的(Nunnally,1978)。效度则是指,答案与我们所设法描述或测量的假想“真值(true value)”的一致程度(Cronbach and Meehl,1955)。

每年的政治竞选、报纸、营销组织、政府和大学学者组织的调查需要从成千上万人那里收集数据。这些调查是通过邮寄、电话和个人访谈等方式来完成的。实际上,它们几乎涵盖了每一个可以想象得到的主题。虽然调查数据有一些潜在的误差来源,但是这些调查结果的效度主要取决于所问问题的设计。不管搜集数据有什么模式或主题,问题总有一些共同的标准,如何达到这些标准也有一些共同遵循的原则。虽然本书主要关注人口调查问题的设计,但本书也阐明了当问题被用来提供有关群体或事件的系统信息时所运用的一般原则。

影响测量的“问与答”特征

当思考一个问题是否是一个好的测量时,就必然要考虑问题本身、问题形式、问题措辞以及问题设计所引出的答案类型。

好的问题的一个标准是:所有回答者应该用一致的方式来理解它,用与研究者所想要的表达相一致的方式来理解它。受访者可能不同于研究者,他们也可能在如何使用语言和理解语言上彼此各异,这一事实使之成为最难达到的一个标准。但是,它所达到的程度对最终测量的质量至关重要。

其次,好的问题必须能够用一致的方式来进行管理。用书写形式或让访员念问题的方式把问题呈现给受访者。如果用书写形式来呈现问题,好的问题的标准是,所有受访者都应该

能够读懂问题。如果由访员来念问题,好问题的标准是,访员能够并会根据问题的书写格式来念问题。

好的问题的第三个标准是,它能够与所有受访者就想要的、可接受的答案类型进行一致地沟通。如果受访者对“什么是适当的答案”看法不一,他们的答案就会五花八门,因为这些答案与我们所设法测量的东西风马牛不相及。

问题:你何时搬到堪萨斯城?

可能的答案:

在 1948 年。

当我 10 岁时。

在我大学毕业以后。

所有这些回答都是可能和合情合理的,因为这一问题没有提供有关“给出什么样的答案”的线索。“你何时搬到堪萨斯城?”不是一个好的问题。“你是在哪一年搬到堪萨斯城的?”是一个更好的问题,因为它详细说明了想要的答案类型。

好的问题的第四个标准是,受访者能够回答它们的程度。当然,在某些情况下,例如,当研究者试图测量知识或能力时,人们能否回答这些问题实际上是正要探寻的信息。当目标不是测量受访者的知识,如果询问人们不知道答案的问题,它就成为测量误差的一个来源。

最后一个标准是,我们应该询问受访者愿意提供正确和有效答案的问题。如果有些受访者有意修饰了答案,就降低了测量的效度。

因此,一个好的测量过程的“问与答”有五个基本特征:

1. 对问题的理解要一致。
2. 对问题的管理要一致,或者与受访者的沟通要一致。
3. 对“什么是一个充分的答案”应该有一致的沟通。
4. 除非问题的目标是测量知识,所有受访者都应该获得准确回答问题所需要的信息。
5. 受访者必须愿意提供问题所需要的答案。

在调查研究历史上,曾把问题设计看作是一门艺术,而不是一门科学(Payne,1951)。科学和艺术的主要区别是:在艺术那里,好(goodness)是“情人眼里出西施”;而在科学那里,“好”有统一的标准。就问题测量而言,“好”的最高标准是:一个问题引出测量某物答案的信度和效度。在过去几十年里,上述有关提供有效数据问题特征的概括有了进一步的发展。此外,我们对如何把这些概括运用到个别问题的设计上也有了更多的了解。本书阐述了我们目前所知道的有关如何设计符合上述这些标准问题的知识。

问题的评估

虽然有一些关于怎样设计好的问题的概括,但调查研究科学的一个重要环节是对调查问题的实证评估。像所有科学测量一样,调查研究测量的质量也会参差不齐。好的科学要设法将误差降至最低,并采取措施来测量剩余误差(remaining error),好让我们知道我们的数据究竟有多好而且我们可以继续改进我们的方法。

问题评估有两种类型:第一种类型旨在评估我们所要询问的问题究竟在多大程度上符合上述五条标准,这五条标准可以被视为过程标准;第二种类型旨在评估所获得的答案的效度。

为了评估这些问题符合过程标准的程度,可以采取以下三种可能的步骤:

1. 专题小组(focus group)讨论。
2. 认知访谈,在访谈中探查和评估人们对问题的理解及其回答问题的方式。
3. 在真实情境下的田野预调查。

从这些活动所提供的问题信息类型来看,每一种活动各有其优劣。然而,在过去十年里,越来越重视在研究项目使用问

题之前,就对问题进行评估。而且,对如何使用这些技术来获得有关问题的系统信息,也有了很大的了解。

还有一套评估策略用来发现问题答案到底在多大程度上产生了有效的测量。它们包括:

1. 分析所获得的数据,来评估答案之间可预测关系的强度以及答案与受访者其他特征之间可预测关系的强度。
2. 对用措辞各异的问题询问类似样本所得到的数据进行比较。
3. 将答案与纪录进行比较。
4. 测量同一个受访者在两个时间节点上答案的一致性。

如果某个特定问题的答案构成一种有效的测量,那么通常就存在着某些可预测的方式,通过这些方式,这些答案应该联系着其他问题的答案或联系着受访者的已知特征。分析结果所提供的证据表明,这些答案是对研究者所设法测量东西的测量。

以上所概括的其他评估策略需要专用的数据收集程序,通常超出了调查的范围。但是,这些策略却提供了不同的、非常重要的方法来评估调查测量的误差。如果用同样问题的两种不同形式来询问同一个样本或可比样本,如果没有测量误差,那么这两个答案的分布应该是相同的。如果两个可比问题的答案没有得出相同的结果,那么这两个问题中的一个或两个结果就含有误差。

把调查所获得的答案与某些纪录或其他标准进行核对并不常见。当这些研究完成时,通常对住院者等有住院记录的人进行抽样。然后就这些事件对这些人进行访谈,并把调查答案与记录数据进行比较。当然,纪录也会有误差,因此,调查答案与纪录之间的差异并不是调查误差明确无误的指标。尽管有这样的局限以及这些研究往往因花费不菲而屈指可数,但是许多方法论上的重大进步却源自记录核查研究(record-check studies)。

最后,对受访者样本进行回访,两次询问相同的问题并比