

# 问卷编制的几个问题

石绍华 张梅玲

中国科学院心理研究所 (北京 100101)

**摘 要** 文章总结了问卷调查法研究中的经验和策略, 阐述了在问卷编制和预备研究过程中应遵循的原则、易犯的错误和常忽略的问题, 为研究人员具体操作提供了可遵循的方法。

**关键词** 标准化, 问卷, 量表, 预备研究

分类号 B841.7

问卷调查法已成为心理学研究的主要方法之一。对个性、气质、需要、态度、价值观等等的测量, 除了问卷法以外, 是什么精密的仪器也测不出来的。我们每个人可能都编写过问卷, 至少回答过各种问卷, 但是否我们每个人心理都很清楚, 如何编制一个标准化的问卷? 在编写问卷的过程中, 我们要注意哪些问题呢? 问卷的质量直接影响我们所获得的数据的质量, 最终影响研究的质量。从 Stanley Payne 有关调查问题设计的经典著作《问问题的艺术》(The Art of Asking Questions)<sup>[1]</sup>发表已经 38 年了。Payne 在这本书里向我们提示, 编制一个问卷不仅需要我们的直觉和经验, 热情和智慧, 也一定要注意实验的证据。关于问卷调查法的经验和实验研究得出了哪些结论呢<sup>[2]</sup>?

## 1 经验和研究策略

### 1.1 简单化的原则

这是一条永恒的基本原则。编写问卷要求语言简单、概念明确、任务容易操作、内容应是众所周知的。

#### 1.1.1 语言简单

要用大众化的、标准的语言, 句子在可能的情况下要短, 语言要简洁。曾有一段时间很强调用短句, Payne 甚至规定一句话不得超过 20 个词, 但现在发现在访谈时加一些承接语效果更好。语速要慢, 口齿要清楚, 让被试有更多的时间思考。以下容易发生混淆的情况是我们要尽力避免的:

①双重对象 (如: 你认为妇女和儿童是否应该……?)

②双重否定 有些被试回答这样的问题时容易迷惑, 在不得不使用双重否定的情况下, 应提醒被试特别注意。

③潜在否定 (如: 控制、限制、禁止、反对等词在某些积极的词前面)

④列举事项太多 在访谈中, 即使是“非常同意 比较同意 同意 不太同意 非常不同意”这样的 5 点或 7 点量表事项也太多, 应该做一个卡片, 让被试看着选 (show card)。如

果被试不识字，就要用图形表示。

### 1.1.2 概念明确

要使问题概念上清楚对心理学家来说可能是个很难的任务，因为心理学家已经被训练为习惯于抽象思维，他们体会不到没有经过这样训练的人在进行抽象思维时的困难。有些看起来很容易的问题，如概率、百分比，公众也不一定掌握得很好。有人<sup>[3]</sup>做了这样的调查：“你看电视新闻节目的时间在你晚上的闲暇时间里所占的比例是多少？”53 个人里只有 14 个知道这是什么意思。

### 1.1.3 任务容易操作

要将事实和态度区分开，因为回答有关事实的问题（婚否，有没有子女）比较容易，而回答关于意见和态度的问题比较难。有些事实的界定也有内在的困难，研究者提供一个参考系，而被试常常不从研究者提供的角度考虑问题。如对家庭的定义，有人认为家庭指核心家庭，有人包括了祖辈，有人包括更多的人。理论上说，所有的细节都应界定清楚，但这是做不到的，我们的定义往往是含糊的、有弹性的，我们应时刻记住这一点，从经验中不断完善。

需要回忆的问题是困难的，如果当时做决定时心不在焉，或很匆忙，或相隔时间太长，或问题较复杂，就更难回答。因此问题应为当前、特定、真实。如果问题是主要生活事件，对个人很重要，这样的问题就比较保险，但也需要一定的线索帮助回忆。有几个策略可以帮助回忆：

①限定回忆范围 Turner 和 Martin<sup>[4]</sup>指出，受害者关于事实的回忆如果经过较长的时间，就会没有价值。现在认为回忆的问题最远是 6 个月之内的，有些特定的事实应是一周，或昨天。

②让被试评价自己的平均状况

③用明显的标记性事件（如：“自我驻南联盟使馆被炸以来……”）帮助回忆是一个经实验证明的策略<sup>[2]</sup>。

④提供线索也是一个经大量实验证明的有用的策略

问一个假想的问题，被试回答是否也是假想的？也许不是，但研究表明，这样的问题是很难回答的。如果被试很认真地考虑，这个问题与考虑比率问题一样难。一些调查可能需要一些假想问题，有时假想的问题是很有价值的，因为被试真实的经验范围太广，研究者无法知道被试会把什么样的经验带到回答问题中，而假想的问题可将刺激标准化。假想的问题在使用之前要仔细审查，如果你认为某些假想问题是必要的，可以使用，同时请考虑两个策略：

①在几个假想问题之后至少加一个真实问题；

②至少在一个假想问题上加探针（probe），如：“你这样说是什么意思？”探明被试的参考系。这非常重要，因为对假想问题的研究很少，我们都需要这方面的数据。

### 1.1.4 内容众所周知

有很多信息是大众不知道的,在问这样的问题时,应做些解释,如:“我要问的问题所涉及到的事实是很多人都不知道的,你也可能不知道,问题是关于创新工程,我先解释一下创新工程……。”大多数人很难立刻运用这些新信息,如果新信息过于复杂,不要问这样的问题。

## 1.2 一些较复杂的方法

以上关于简单化的建议忽略了一些有趣的复杂的方法,而这些方法可用于提出重要的问题。

### 1.2.1 因素调查法 (Factorial Survey)

用小故事研究判断、决策或归因过程,这种方法的优点是它提供了具体详细的情境,被试容易判断,尽管故事是虚构的。可利用故事揭开复杂的现实世界的难题。这种方法存在的方法学问题是:1) 难度与其他调查如何比较? 2) 被试能接受一个故事的多少信息? 3) 一次调查中可以调查多少因素有没有限制? 4) 是否存在背景效应? 5) 故事中信息的顺序有没有影响? 由于这些问题都没有解决,这个方法不常用。

### 1.2.2 排序量表 (Ranking Scale)

这种方法有很长的历史,实验心理学家兼市场研究者 Daniel Starch<sup>[4]</sup>首次使用该方法,这种方法的缺点是任务难度大,今天的文献中已经见不到长的列表,但短的排序量表(4—6个项目)还是有的。Kohn<sup>[5]</sup>提出的部分排序法,他列举了男孩子的13个社会赞许品德,要求被试在13个选项中先选出3个最重要的,然后在这3个中间再选出一个最重要的。还要求被试选出3个最不重要的,然后在3个最不重要的选项中再选出一个最不重要的。

部分排序法使任务较容易操作,但普遍的观点是,它仍比等级量表(rating scale)难,后者不需要被试在多个项目之间进行比较,用的时间较短。排序量表的优点是被试作出更多的努力,赋予更多的注意。

排序量表的方法学问题是:它与等级量表是否能比较? Alwin 和 Krosnick<sup>[6]</sup>利用样本分半实验(split sample experiments)研究了这个问题。他们发现,在选择哪些选项重要,哪些不重要方面,两种方法是近似的,但用两种方法在确定潜在的维度与预测变量之间的关系上,结果是不一样的。排序量表要求被试作出某种比较和判断,而他们平时是不会进行这样的比较和判断的,因此排序量表不仅测量了被试的态度,也测量了被试认识选项之间逻辑对照的能力。

### 1.2.3 数值估计量表 (Magnitude Estimation Scale)

这种方法来自心理物理学,用数值估计来测量态度。如:“我要问你一些关于犯罪严重性的看法。第一个情境是,一个人在街上偷了一辆自行车,这样的罪行其严重性是10分,用这个情境判断其他情境。如果你认为某罪行比偷自行车严重20倍,就是200分,如果比偷自行车轻一倍,就是5分。以下情境是多少分:

- 父母打孩子,致使孩子住院。

- 在公共场所放炸弹，如果爆炸，将会炸死 20 人。

这是一种比率量表，更复杂，更难操作。它的缺点是任意规定意见的范围会失去很多信息，把这样得来的数据进行统计，准确性就不能保证。使用这种方法需训练被试学会比例判断，而被试是否能学会这种比例判断，方法学研究很少。有研究 (Lodge) [7] 表明它的优点是比一般的测量多解释了 12%—15% 的变异。

对这些复杂的方法还需要更多方法学的研究。

## 2 实验证据

有关方法学的实验得出的结果不很一致，也不很明确，因此长期以来，知识和个人经验在问卷设计中扮演着比实验结果更重要角色，但是，有些结果还是一致的。下面是一些实验结果。

### 2.1 具体特定的问题比一般性问题好

同一个问题对不同的人其意义是不同的，一般性问题尤其如此，问题越一般，越抽象，对它的解释越多样，行为和态度也越不易区分。特定的、具体的问题的意义对每个人都会比较明确。

一般性问题不能完全排除，因为有时要比较一般观点和具体观点，有时要了解一般状况。

### 2.2 何时用开放式问题，何时用封闭式问题？

我们最终还是要构成封闭问卷，否则谈不上标准化。对封闭问题最多的批评是，它强迫被试选择，而不是被试用自己的语言表达。但也正因为封闭问题提供备选答案，使问题成为具体而特定的，而且使所有的人都使用统一的参照系，有助于确定特定的问题。

当你对某一方面的问题不是专家时，要用开放问题。例如，在美国 1981 年有一项大型研究，调查公众认为美国当时的最大问题是什么，用开放式问卷时，很多人认为最大问题是能源短缺，而用封闭式问卷时，由于调查者没有考虑到能源问题，99% 的被试在问卷提供的 5 个备选答案中选择。可见这时应该用开放式问卷<sup>[2]</sup>。

在有关敏感问题、社会不赞许行为方面，尽量用开放式问卷。有研究<sup>[2]</sup>表明，在关于饮酒和性生活的封闭式问卷调查中，被试报告的频率都低于实际频率，但使用开放式问卷，被试报告的频率高于封闭式问卷。因为如果你的备选项为：从不 一年一次或更少 几个月一次 一月一次 几周一次 一周一次 每周几次 每日一次 被试倾向于选择那些位于中间的选项。

在调查突出的问题或捕捉时尚时，开放式问卷也好于封闭式问卷。

### 2.3 提供“不知道”或“没有意见”选项

研究表明，如果提供“不知道”或“没意见”选项，有 1/8—1/3 的人会说不知道。最好加一句话：“对这个问题并不是每个人都持有看法，如果你有看法，请回答下面的问题。”

### 2.4 取消中间立场的选项

对这个问题的看法一直不一致，有人认为提供中间立场选项是危险的，而有人提倡要

提供中间立场选项。有一项样本分半实验研究了这个问题,当不提供中间选项时,只有21.7%的被试自己加上这个选项,但提供这个选项时,40.2%选择了它。有20%选择中间立场的被试意见不强烈,但他们还是有倾向性,如果不提供中间选项,他们就能表达出意见。

### 2.5 迫选法优于同意-不同意陈述

通用的态度问卷是用“同意-不同意”陈述,这也是个在方法学上受到最多攻击的方法,批评之一是所谓“默认反应定势”(acquiescence response set)。被试有时(特别在疲劳时)作出全选“同意”。默认反应定势与受教育程度有关,受教育程度越低,越倾向于默认反应定势,态度截然相反的两道题,他们都会选择同意。

### 2.6 措辞的影响

有关措辞的实验研究的结果是有趣的。在有些情况下,措辞上微小的改变就会影响很多人,而这种影响事先是不能预测的;在另一些情况下,在问题上加一些消极的措辞并不影响被试的反应。解决措辞问题可用以下方法:

①做样本分半比较,对同一问题采用两种提问方式,可以看出措辞的影响;

②在封闭式问题后面跟上开放式问题;

③对同一个题目多问几个问题,寻找不同的角度,用多元测量可以避免偶然性,使研究者不将单一问题的回答作为结论。

## 3 问卷编制

最终,每个问题都设计者写出来的,设计者必须修改、调整、尝试、看它看起来如何,听起来如何,人们对它如何反应,再修改、再尝试、减一些问题,加一些问题。编制问卷有两个步骤:问卷开发和预备研究。

### 3.1 开发问卷

开发一个问卷可以从向专家和普通公众请教开始。一个探索性研究要求研究者走出自己的圈子,向其他领域的专家请教,如果只是与背景相似的人讨论,他们的局限性和我们自己是一样的,而不同的角度,不同的经验,可以带来新的信息。

可对从目标人群中选择几个人进行深入访谈,这样可以开阔研究者的眼界。召开稍微正式一点的讨论会效果更好。不幸的是,我们大多数人都忽略了探索研究的这个基本步骤,太快地直接进入写问卷或从其他问卷量表中借用问题。

国内外大规模调查越来越多,积累了很多问题,在设计问卷时应参照,可以节省大量人力和时间。但借来的问题要修改。

### 3.2 预备研究

预备研究的重要性毋庸置疑,但做起来有很多问题,主要是预备研究不正式,靠直觉,缺乏统一的原则,缺少系统化的实践。在应该如何进行预备研究、研究者应从预备研究中了解什么、如何在预备研究的基础上重新修改问卷等问题上研究不多,还没有形成规范。

### 3.3 预备研究设计

预备研究设计分为被试参与设计(participating)和被试不参与设计(undeclared)两种。在

被试参与设计中，被试知道这是预备研究。可以让被试解释他的反应，可以问一些这样的问题：

你对整个问卷感觉怎样？

当你说“……”时，你心里是怎样想的？

这个问题如果这样问“……”你会怎样回答？

如果怎样问“……”你回答起来会有什么困难吗？

在被试不参与设计（undeclared）中，被试不知道这是预备研究，就像一个正式的访谈，主试可以从被试的反应知道被试的参照系和他们的意思，但是不如被试参与设计那样深入。最好的方法是先做参与设计，后做不参与设计。

在预备研究中，主试的角色是不同的，有些研究者在第一次预备研究时喜欢自己做主试，问一些即兴的问题，也可以让那些很有经验有技巧的主试这样做，准确记录下他们问的每个问题。几个关键的研究者在第一次预备研究时可以偏离问卷结构，问些即兴问题，一般主试，特别是新手、学生，则按照写下的问卷问题提问。还有一种变式，所有的主试在访谈最后都提几个非结构问题（离开问卷的问题）。

### 3.4 研究的目的

检验被试对项目的反应的分布情况是预备研究最通常的目标，如果回答是/否的比例是 95/5，我们就应该寻找更值得研究的问题。

检验被试对问题的理解是否与问卷设计者心里想的意义一样。Belson<sup>[3]</sup>曾做了一个研究，他构成了一个小问卷，50 个题目，访谈结束后，一个训练有素的主试再去找那些人访谈，发现 50 个问题中没有一个是所有的被试和问卷设计者在问题的每一个部分理解都是一样的，最好的问题有 50% 的被试在问题的所有部分与设计者理解一致，最差的问题没有一个被试与设计者理解完全一致，平均为 29%。当然，他们挑选的问题都是容易造成困惑的，一般的问卷不会这么差，但他们提出的问题是十分重要的。

检验任务难度是否适当。

检验被试的兴趣和注意，发现哪些问题被试最感兴趣，哪些问题被试觉得枯燥。

为保证问卷的流畅性，在问卷编好后要多看几遍，把自己当成被试，如果是访谈，要自己多读几遍，在问卷的不同部分之间要插入一些提示，如“现在，我要问你几个关于教育子女的问题”。有时需要几个冗余词，使语速减慢，被试听得更清楚。

对问题顺序研究不多，一般认为应以有趣的问题开始。在电话访谈中，初始问题应直接与主题有关，并做自我介绍。

一般电话访谈时，被试的回答都比面对面时简短，因此被试在电话里多说几句话是很有价值的。开始不要用开放式问题，因为开放式问题不太好回答，会使被试感到气馁。人口学问题可以在开始问，也可以在最后问，也可以分开来问，有关收入等敏感问题最好放在最后问。

跳题模式（例如，若回答“是”，请跳到第 37 题，若回答“否”，请跳到第 38 题）有时

很复杂，应反复核对。

时间一般应限制在 50 分钟左右，每个部分都要控制，电话访谈要稍短。

被试对整个问卷的兴趣和注意也是预备研究时要考察的，超过 2 小时的问卷测的是被试的耐心，而被试一般不是很有耐心的。枯燥的问卷也会引起疲劳和注意问题，解决的方法之一是，问卷不同的部分用不同的格式，引起被试注意。我们可能做到的最生动的问卷是让被试做一个个不同的活动，先回答一些是否题，再从非常同意……非常不同意中挑选一个，然后在一个尺子上评价自己的感觉，然后用迫选法从几个选项中挑一个。任务的变化可以引起被试的兴趣，但变化带来的注意和重新启动一个新的格式的成本要衡量。

很多问题涉及个人问题。美国很多大学有“人类被试审查委员会”(Human Subjects Review)，研究者要向人类被试审查委员会提交申请，经他们批准的问卷才能使用。美国有一项关于抑郁的大规模调查<sup>[2]</sup>，在预备研究阶段，一些有经验的主试认为预备研究使被试沮丧，因为这项调查只涉及生活中的坏消息和心理疾病的症状，这些主试坚持问卷必须修改。这些智慧的意见和经验胜利了，设计进行了修改，增加了积极生活事件，也增加了一些关于被试如何应付困难的问题，这是这个问卷的第五次修改，研究者和被试都从这里获益。对这个问题并没有太多的数据支持，这主要靠研究者和主试的广阔的专业经验、判断力、对人的关爱、自信、才能和科学良知。

预备研究是定量研究的定性阶段，小型的预备研究被试需 25—75 人，有了 25—75 人的数据就可以对数据进行编码。这样数量的被试已经可以包括边缘分布。在某些情况下，边缘分布可以支持某些预感，而否定另一些预感，也使研究者知道自己的经验可能不具代表性，不具有普遍性。有时被试的反应比列联表的边缘频率更有建议性，这更说明预备研究的定性的性质。

### 3.5 研究程序

在预备研究中，大多数问题应该是简化的，简化的问题不可能包括所有的状况，做过调查的人都知道人类的多样性，研究者用一个“其他”就把一些性质完全不同的东西归为一类，漏掉了很多生动的生活特征和非常情况，开放式问题的确在这方面好得多，但开放式问卷的效度很有问题，它不可避免的一个问题是心照不宣假设(tacit assumption)。

Paul Lazarsfeld<sup>[2]</sup>在 50 年前调查奥地利儿童时发现了这个问题。他问孩子们生活中最希望的是什么，他们写的都是钱、农场、幸福等，没有一个人写智慧。在封闭式问卷中有智慧这个选项，大多数孩子都选择了智慧。回答开放式问卷的孩子假设问的是通过努力达到的东西，而不是天生的东西。

如果还要做第二次预备研究，那么第一次预备研究的长度应是正式问卷的 2 倍，如果没有第二次，预备研究不应超过 1 小时 15 分钟。预备研究的被试应与目标人群相似。如果主试是有经验的访谈者，预备研究的被试人数  $N=10-25$ ，如果主试是新手， $N=75$ 。近年的第一次预备研究一般  $N=30$ ，由 6 个有经验的访谈者进行，每人访谈 5 个被试。

访谈时最好两人一起，一人问问题，一人观察被试和主试的反应，哪个问题使被试感

到不舒服？被试容易误解哪个问题？主试是否不得不重复哪个问题？哪个部分拖延了时间？哪个部分被试似乎有更多的话要说？将观察结果记录在问卷上。对反应进行编码，计算边缘频次。曾一度要求访谈者严格按照问卷朗读，不回答被试的问题，近期的研究表明向被试解释清楚问题能大大提高效率<sup>[8]</sup>。

第二次预备研究是为第一次预备研究润色。它是彩排，不是修改错误，也不探索新东西，而是精简、调整、连接、重新排列、一切与正式调查一样。

在实测之前可以请局外人批评。同事，尤其是做过类似工作的同事可能有好的意见。在你的问卷未形成时，他们很难提意见，这时是好机会。

对问卷的信度、效度和项目分析是所有测量的共同课题，而不是问卷调查的特殊课题，这里不再赘述。

经过这些步骤，相信你已经获得了一个标准化的问卷。

## 参考文献

- [1]Payne S L. The Art of Asking Questions. Princeton, NJ: Princeton University Press,1951.
- [2]Converse J M, Presser S. Survey Questions: Handcrafting The Standardized Questionnaire, Beverly Hills, CA: Sage ,1986.
- [3]Belson W R.The Design and Understanding of Survey Question. Aldershot, England: Gower, 1981.
- [4]Starch D. Principles of Advertising, New York: McGraw-Hill, 1923.
- [5]Kohn M L. A Study of Values. Homewood, IL: Dorsey, 1969.
- [6]Alwin D F, Krosnick J. The measurement of values in surveys: a comparison of ratings and rankings, Public Opinion Quarterly, 1985, 49: 535-552.
- [7]Lodge M. Magnitude Scaling: Quantitative Measurement of Opinions. Beverly Hills, CA: Sage, 1981.
- [8]Krosnick JA. Survey Research. Annual Review of Psychology, 1999, 50:537-67.