

高回收率更好吗？

——对调查回收率的另一种认识

风笑天

提要: 论文结合国内外的研究实例,对调查回收率的衡量标准及其依据提出了自己的看法。作者指出,一方面,调查回收率所代表的实际样本的大小,是衡量一项调查结果所具有的代表性的最终指标,因而回收率越高越好。但另一方面,由于实际调查中存在着许多制约和影响回收率的因素,如果不切实际地单纯追求高回收率可能会影响到调查资料的质量,因此回收率并非越高越好。研究认为,样本的完整与资料的质量是我们在调查中应该同等关注的两个方面,既要尽可能地提高回收率,也要尽可能地防止和避免低质量的高回收率。

关键词: 社会调查 回收率 方法论

一、一个意外的提问

2005年8月,美国社会学协会(ASA)第100届年会在费城召开。在分组会议上,一篇由笔者与合作研究者(美国教授及其博士生)合写的论文进行了宣读。论文报告完后,在场的一位美国学者提出了一个这样的问题:“你们的研究报告指出,调查的回收率为96%,请问,回收率怎么会有这么高?”

这一提问让笔者感到十分意外。国内学者——包括笔者——以往所进行的一些调查,回收率达到90%以上可以说是相当常见的现象(实际的统计结果证实了笔者的印象,具体情况可参见本文第三部分的内容)。更为重要的是,在笔者(以及国内许多同行)看来,调查回收率的高低,是衡量一项调查质量高低的主要标准。而在实际从事社会调查的过程中,努力追求高回收率也早已成为许多国内同行的共识和行动。难道高回收率有什么不好吗?难道我们不应该追求高回收率吗?高回收率有什么不正常吗?正是这种意外的感觉以及由此所带来的一系列疑问,促使笔者对调查中的回收率问题进行了重新思考,并对调查

回收率或高或低的现象背后所具有的深刻含义进行了分析。本文的目的就是围绕这一看似简单、实际上却关系重大的方法论问题谈谈笔者新的认识,以就教于学界同行。

二、调查回收率及其意义

社会调查中的回收率(response rates, 也称作回答率或应答率)指的是“调查者实际调查的样本数与计划调查的样本数之比”(风笑天, 1987),也就是社会调查过程中研究者成功完成调查询问的个案数占计划完成的样本总个案数的百分比。它通常包括自填式问卷调查中的问卷回收率和结构式访问调查中的访问回答率。回收率的计算方法是:

$$\text{回收率} = \frac{\text{实际完成询问的个案数}}{\text{计划完成的样本总个案数}} \times 100\%$$

比如,一项调查从总体中抽取了400名调查对象作为样本,研究者采用自填问卷的方法收集资料,发出问卷400份,实际收回问卷380份。那么,根据上述公式计算,该项调查的回收率为:

$$(380/400) \times 100\% = 95\%$$

同样,如果研究者采用的是结构式访问的方法来收集资料,而在实际访问400名调查对象的过程中,有10位调查对象由于地址错误、外出打工、生病住院等客观原因而无法接触到,还有10位调查对象拒绝接受访问。结果,实际完成访问的对象为380人,回收率亦是95%。

实际调查中,由于收回的问卷里还可能会有一部分不合格的问卷,所以,真正严格意义上的调查回收率(也称为有效回收率)指的是通过对问卷的审核,剔除那些填答不全或明显乱填的废卷后所剩下的问卷数(即有效问卷数)占样本总个案数的百分比,即:

$$\text{(有效)回收率} = \frac{\text{实际完成询问的有效个案数}}{\text{计划完成的样本总个案数}} \times 100\%$$

比如,通过审核发现上述调查中所收回的380份问卷里,有20份不合格的废卷,将这些废卷剔除,该项调查的有效回收率就只有:

$$[(380 - 20) / 400] \times 100\% = (360 / 400) \times 100\% = 90\%$$

由于最终进入数据分析的问卷数目是有效问卷数,因而,一般情况

下研究者在研究报告中向读者报告的应该是有效问卷数和有效回收率。如无特别说明，本文中的调查回收率也指的是调查的有效回收率。

对于社会调查而言，回收率最常见、同时也是最基本的方法论意义在于：它是决定和影响调查样本代表性的重要因素。“回收率是受访者样本代表性的一项指标”（巴比，2000：331），“它是反映抽样调查结果对总体的代表性程度的重要指标之一”（风笑天，1987）。调查研究中，研究者虽然可以通过科学的抽样设计，达到从总体中抽取具有代表性的样本的目的，然而，这种抽样所得到的样本的代表性并不能最终反映调查结果的代表性。因为抽样所得样本对总体的代表性是一种处于调查开始之前的衡量指标。在进行调查的过程中，会有许多因素导致抽样样本中部分个案的缺失或失效，使得最终完成调查的样本只是抽样样本中的一个部分。而由于“未回答者通常与问卷的回答者有着相当的差别，他们通常是一些由于受教育程度低而看不懂问卷的人、一些年事已高而无法回答问卷的人、或者是一些流动性较大而无法找到的人”（Nachmias & Nachmias，2000：208），所以，当调查的回收率较低（即未回答者的数量较大）时，调查结果的总体代表性就会受到明显的影响和破坏。这方面最著名的例子是美国《文学文摘》杂志关于美国总统选举的调查。该杂志在1936年的总统选举调查中收回了200多万份调查表，其收回的问卷规模可谓巨大。但是，相对于其所发出的1000多万张调查表来说，200万的数字依然只是很小的一部分。正是这种20%左右的回收率（以及研究者在抽样框方面的失误），彻底葬送了《文学文摘》杂志的这次调查结果，也最终导致了这家杂志的关门。回收率过低而影响到调查结果质量的另一个著名例子是美国学者海蒂关于美国女性的性调查。尽管这项著名调查的有效样本规模也达到了4500份，但相对于其发出的10万份问卷来说则小得可怜（回收率仅为4.5%），其研究结果因此遭到学界猛烈批评。也正因为如此，“在调查研究中，回收率是研究者的一大担心”（Neuman，1994：239）。

那么，可接受的回收率，以及很好的回收率的标准是什么呢？对此学界目前还没有统一的认识。纳克米亚斯等人指出：“要确定一种可接受的回答率标准并不是一件容易的事情，因为科学家在最低回答率的标准上意见不一致”（Nachmias & Nachmias，2000：213）。美国社会学者巴比提出过一个简单的等级规则：“要进行分析和报告撰写，问卷回收率至少要有50%才是足够的，要至少达到60%的回收率才算是好的；

而达到 70% 就非常好。”但他同时也明确指出：“要记住，以上数据都只是概略指标，并没有统计上的基础”（巴比，2000：331）。

尽管没有统一的衡量标准，但笔者认为，一般情况下，如果无回答的比例超过样本总量的 $1/3$ ，即回收率低于样本总量的 $2/3$ 时，调查结果就有可能出现大的偏差。因此，当一项调查的回收率低于这一比例时，研究者有责任对未回收的调查对象所具有的基本特征及其所带来的误差进行分析和说明。因为这种误差在相当大的程度上将决定和影响调查研究结果的推广范围以及推广中的偏差范围。

由于回收率所代表的样本才是真正的样本，或者说，才是得出调查结果时实际用到的样本，因此，即使在抽样样本具有很好代表性的情况下，调查回收率的高低仍然可以影响到调查结果的代表性程度。正是这种由回收率所代表的调查样本的大小，成为衡量一项调查的结果所具有的代表性的最终指标。也正是在这种意义上，我们说，调查的回收率越高越好。

三、目前国内社会学界调查回收率的状况

正是基于上述认识，研究者在调查中一般都会尽可能地追求更高的回收率。国内社会学界调查回收率的状况也很好地说明了这一点。笔者对《社会学研究》最近两三年来（2004 年第 2 期—2006 年第 3 期）所刊登的、利用抽样调查数据（一项网络调查除外）撰写的全部研究论文中的回收率进行了统计，其结果表明，在最近两年半的时间内，《社会学研究》共发表了利用抽样调查所得数据撰写的研究论文 27 篇（涉及到 20 项调查）。其中，报告了回收率的 19 篇，没有说明回收率的 8 篇（其中有 4 篇通过其他方式了解到或计算出其回收率）。在最后得到的总共 23 篇（属于 16 项调查）有回收率统计的报告中，2 篇（属于同一调查）的回收率为 60% 左右，1 篇为 76%，3 篇为 82%—90% 之间，其他 17 篇（属于 11 项调查）的回收率都超过 90%。特别是其中两项全国范围的大规模调查（样本量均在 6000 份左右）的回收率更是达到了 99.2% 和 99.9%（见表 1）。

表 1 2004—2006 年《社会学研究》刊发的抽样调查报告的回收率统计

序号	年份及期号	样本规模 (有效份数)	有效回收率 (%)	抽样方式 (未标出者为随机抽样)
1	2004 年第 2 期	383	93.4	配额抽样(非随机)
2	2004 年第 2 期	338	84.5	判断抽样(非随机)
3	2004 年第 4 期	6193	99.2 [*]	
4	2004 年第 4 期	280	无说明 ^{***}	偶遇抽样(非随机)
5	2004 年第 5 期	526/420/422	无说明 ^{***}	判断、偶遇抽样(非随机)
6	2004 年第 5 期	6193	99.2 [*]	
7	2004 年第 6 期	1004	59.9	
8	2004 年第 6 期	496	99.2	
9	2004 年第 6 期	812	90.0	
10	2004 年第 6 期	1039	94.5	偶遇抽样(非随机)
11	2004 年第 6 期	3824	76.3	
12	2005 年第 1 期	303	89.1	偶遇抽样(非随机)
13	2005 年第 2 期	6193	99.2 [*]	
14	2005 年第 2 期	976	91.2	判断抽样(非随机)
15	2005 年第 3 期	5894	99.9	
16	2005 年第 3 期	3293	82.3 ^{**}	
17	2005 年第 4 期	1004	59.9	
18	2005 年第 4 期	3183	84.2	
19	2005 年第 5 期	5894	99.9	
20	2005 年第 6 期	203	无说明 ^{***}	偶遇抽样(非随机)
21	2006 年第 1 期	3365	无说明 ^{***}	无抽样说明
22	2006 年第 2 期	913	91.3	
23	2006 年第 2 期	5894	99.9	
24	2006 年第 2 期	561	93.5	配额抽样(非随机)
25	2006 年第 3 期	5894	99.9	
26	2006 年第 3 期	973	97.3	
27	2006 年第 3 期	5894	99.9	

说明: * 原文未直接给出回收率,系笔者根据报告中同一调查的相关资料计算得到。

** 原文未直接给出计划样本数,系笔者根据原文中相关信息计算得到。

*** 原文未给出计划样本数,无法计算回收率。

上述统计说明,目前国内社会学界所进行的调查研究,绝大部分(70%左右)的回收率都在90%以上。如果那位提问的美国学者了解到这样一种十分普遍的高回收率状况,或许会更加大惑不解。然而,值得我们思考的问题是:他为什么会在这种较高的回收率感到不可思议?特别是为什么会以一种对高回收率表示怀疑和将高回收率看作缺点的方式来提问?

四、GSS: 调查回收率的一个参照

为了回答上述问题,我们先来看看美国社会调查方面的一些情况。在美国,有一项专门为全体社会科学的学者和学生所利用而设计的全国范围的大型抽样调查——综合社会调查(General Social Survey,简称GSS)。它由芝加哥大学的“全国民意研究中心”(NORC)具体负责实施。从1972年到1993年,GSS基本上每年进行一次。调查采用入户访问的方式进行。每次调查的样本都是单独抽取的,样本规模大约在1900—2500份之间。后来由于调查经费紧张,从1994年开始,改为双年调查一次,样本规模则相应增加了大约1倍,即在3800—4100份之间。1972年开始第一次GSS调查时,访问表格很短,仅仅由20个问题组成。从1973年开始,该调查得到美国国家科学基金的资助,问卷变成由上百个问题所组成。根据芝加哥大学NORC的统计,目前至少有7000份利用GSS的调查数据写成的研究报告发表。此外,GSS的数据资料还经常用在大学本科生和研究生的教学中,也经常被新闻媒体利用(Dorsten & Hotchkiss 2005: 350)。那么,这样一个著名的调查的回答率情况如何呢?从1972年到2002年,GSS在长达30多年的21次调查中,访问了将近44000位回答者,调查的回收率却始终比较稳定,基本上都处于70%—80%之间(最低的2000年和2002年是70%,最高的1993年是82%),全部21次调查的平均回收率为76%(见表2)。

将表2 GSS统计结果与表1国内调查回收率状况的统计结果相比较,我们就不难看出二者之间的差别,同时也就不难理解那位美国学者为什么会有如此的疑问。由此,给我们提出了这样几个值得思考的问题:首先,为什么在长达30年的时间范围内,GSS的回收率基本上没有什么大的波动?其次,为什么这种回收率的范围始终集中在76%左

表 2 不同年份 GSS 的样本规模和回收率

年份	计划调查样本规模 * (份)	回收的有效样本规模 (份)	回收率 (%)
1972		1613	N/A
1973		1504	N/A
1974		1484	N/A
1975	1960	1490	76
1976	1999	1499	75
1977	1987	1530	77
1978	2070	1532	74
1980	1932	1468	76
1982	2385	1860	78
1983	2024	1599	79
1984	1865	1473	79
1985	1942	1534	79
1986	1934	1470	76
1987	2425	1819	75
1988	1923	1481	77
1989	1971	1537	78
1990	1854	1372	74
1991	1945	1517	78
1993	1959	1606	82
1994	3836	2992	78
1996	3821	2904	76
1998	3726	2832	76
2000	4024	2817	70
2002	3950	2765	70
合计		43698	

注: 此表参见 Dorsten, Linda Eberst & Lawrence Hotchkiss, 2005: 350。

说明: *最初几年无回收率统计; 表中计划调查样本规模系笔者根据表中有效样本规模和回收率数据计算得到。

右? 第三, 作为美国社会调查方面的著名机构, 芝加哥大学 NORC 的专家们显然明白高回收率对调查所具有的重要意义, 但他们为什么不努力、或不可能将调查的回收率提得更高些? 与这几个问题密切相关、同

时也更为本质的问题则是：调查中的高回收率何以可能或不可能？换句话说，有哪些因素在影响或决定着调查回收率的高低？

五、高回收率何以可能或不可能？

为了更好地提出这一问题，我们可以先看看国内调查与美国 GSS 调查之间的差异。在前面表 1 所列出的国内各项调查中，有两项以全国范围居民为总体的抽样调查与美国 GSS 调查的背景十分相似（其中一项是 2001 年对全国 12 个省及直辖市、总共 73 个区县的 6193 位城乡居民的抽样调查，另一项是 2003 年对全国 28 个省、直辖市、自治区、总共 92 个区县的 5894 位城镇居民的抽样调查）。同时，这两项调查恰恰又是目前国内大规模抽样调查中回收率最高的两项，其回收率分别为 99.2% 和 99.9%。

与美国 GSS 调查的回收率两相对照，反差似乎太强烈了。为什么同样是全国范围的、以普通居民为对象的抽样调查（样本规模还远大于 GSS 的样本），美国的回收率仅达到 75% 左右，而中国的回收率则接近 100%？或许我们对此能够找出多种不同的解释：比如，中美两国的文化传统和社会心理导致两国居民对调查的态度不同、对调查的接纳程度不同；或者，中国的调查能够通过官方的途径、得到基层政府的支持，等等。但是，一方面，所有这些可能的解释都还有待经验证据的证明。比如，对于“中国的调查能够通过官方的途径、得到基层政府的支持”这种解释来说，表 1 所统计的回收率超过 90% 的 10 项调查中，通过一定的官方途径、得到基层政府支持的调查可能只有 1 项。同时，国外有文献指出，即使是有着政府官方背景和资源基础所进行的大规模调查，同样也会遭遇到比较高的无回答比例。例如，2000—2001 年度的英国“综合家户调查”（General Household Survey，简称 GHS）的无回答比例是 33%，即该调查的回收率只有 67%；而 1999—2000 年度的英国“家庭消费调查”（Family Expenditure Survey，简称 FES）的无回答比例更是达到了 37%，即这一调查的回收率只有 63%（David & Sutton, 2004: 181）。另一方面，笔者认为，即便存在上述各种可能的解释，我们也不能忽视实际调查过程中所可能存在的其他方面的、同样重要的原因。

在影响调查回收率高低的各种因素中，人们比较熟悉的是调查方

式的影响。一般来说,采用入户方式进行的结构式访问调查的回收率往往最高,其次则是自填式的问卷调查;而采用邮寄方式进行的调查以及采用电话访问方式进行的调查,回收率(或应答率)往往是最底的。^①这也是为什么只要经费、时间等条件允许,研究者通常都会采用入户访问或自填问卷的方式进行调查的重要原因。然而,即便同样是入户访问调查,还是有一些其他方面的因素会影响到回收率的高低。这些因素主要包括调查对象的性质、调查抽样的方式、调查的工作量、调查人员的努力程度,以及调查资料的质量要求等等。

首先,在其他条件相同时,调查对象是城市居民还是农村居民,是青年人还是中老年人,回收率会大不一样。一般而言,农村居民相对比较朴实,受教育程度较低,待人更为热情厚道,对调查员的接纳程度相对较高,对调查往往会比较重视。因而,以农村居民为对象的调查,其回收率往往会比同类的城市居民调查要高。相反,城市居民文化程度相对较高,且见多识广、比较精明,而待人相对比较冷漠,戒备心较强,对调查员的接纳程度及对调查的关注、重视程度也较低。因而,调查要获得较高的回收率自然会更为困难。而同样是在城市居民中做调查,也往往会呈现出城市越大调查越困难、调查的回收率越低的趋势。国外的调查也有类似情况发生。比如日本家庭社会学学会于1998年进行了日本第一次“全国家庭调查”(National Family Research,简称NFR),其样本规模达到10500份,而回收率只有66.5%。在谈到回收率较低的原因时,研究者特别指出:“在大城市和低年龄层中的回收率低,这与其他社会调查出现的问题相似”(石原邦雄,2000)。

其次,在其他条件相同时,调查抽样的方式是随机抽样还是非随机抽样、是一次性抽样还是多阶段抽样、是简单随机抽样还是分层随机抽样或整群随机抽样,对回收率的影响也各不相同。随机抽样程序的严格实施与调查的回收率似乎具有天然的反比关系。这是因为,“随机抽样的科学原则及其严格的程序要求,具有很强的理想化色彩。而抽样的实际过程又是处于各种现实条件的限制之中的,科学原则与现实条件之间充满了矛盾”(风笑天,2000:177)。真正严格的随机抽样在具体

^① 当然,也有特殊情况。例如,2003年有研究者以电话访问的方式所进行的一项有关“SASS”社会心态的调查,借助当时暂时的封闭性的社区生活状态下的特殊情境及被访者急于与外界沟通的特殊心态之便利,获得了较高的应答率。

实施中肯定会遇到一些无法获取调查的个案。越是严格地按随机抽样的方案去寻找和调查每一个对象,调查的难度就会越大,现实中面临的不可能情况就会越多,较高的回收率也往往越不容易达到。比如,本文开头所提到的笔者的那项调查,其获得较高回收率的原因,除了调查对象大部分为农村居民,采用的又是入户访问的方式以外,还有一个重要的原因,就是其采用的是以村(居委会)为单位的整群抽样(即对所抽中的村或者居委会中的所有居民户进行调查),这种方法在获取调查对象上比其他抽样方法显然相对容易一些。同时,在实际调查中,又由于实施上的客观困难(如有的居民不在家)而放弃了样本的完整性,导致实际调查的样本并不是严格意义上的整群抽样样本(或者说,其实际的整群样本要大于调查的整群样本)。这样看来,96%的高回收率实质上是在牺牲了样本抽取严格性的条件下得到的。

另外,表1的统计结果也从另一个侧面反映出随机抽样方式的这种影响。在表1统计的总共20项调查研究中,明确为非随机抽样方式的高达9项,加上一项抽样方式不明的,共有一半的调查没能实现随机抽样。这种结果的出现显然不是研究者不知道或者不愿意采取随机抽样的方法所导致。其中的原因,除了一些客观上的困难外,保证调查的可行性和获得较高的回答率(即避免过低的回收率)恐怕也是一个重要的考虑因素。总之,抽样方式的随机性、抽样调查实施中的严格性与调查的高回收率之间存在着一定的矛盾。实际从事调查研究的学者往往会有这样的切身体会:我们在抽样方法的随机性或者在具体实施的严格性方面稍微做出一点点“退让”,往往就会为调查的“顺利”实施和完成带来极大的现实性和可行性。

第三,回收率还与调查的总工作量、调查资料的质量要求以及调查人员的努力程度等因素密切相关。调查的总工作量由调查样本的规模、样本分布的范围、调查每一个个案所需时间三者的乘积构成。不难理解,无论是样本规模的扩大,或是样本分布范围的扩大,还是调查每一个个案所需时间的增加,都会增大调查的总工作量。而在其他条件不变的情况下,一项调查的总工作量越大,它要获得高回收率的难度也往往越大。另外,一个不容忽视的客观事实是:任何调查最终都是由调查员具体完成的。影响调查回收率的因素中,调查员对调查(包括实地抽样和具体访问)的实施其实是最为关键的因素。在面临各种调查和抽样困难时,调查员的努力程度以及由此所形成的处理困难的方法,既

直接影响到调查资料的质量，又影响到调查的回收率。严格按照抽样方案或调查计划的程序要求、不折不扣地执行调查和抽样方案，往往可以获得较高质量的资料，但同时需要花费调查员更多的时间、精力、体力和耐心。因而在调查的效率和质量面前，作为普通人的调查员会有着不同的处理方式。据笔者有限的了解，正是实地调查中调查员在面临各种困难时的种种“临时处理”，帮助一些调查（包括笔者以前从事过的某些调查）达到了相对较高的回收率。调查员的这种“临时处理”可能包括下列一些情况：遇到样本户拒访或无人在家时，临时改为就近方便抽取邻居家庭；为了节省时间（免得再次登门）或方便调查，将入户后的随机抽取临时改为谁在家或者谁愿意接受调查就调查谁；因调查工作量太大，调查员为加快进度，有意识跳过问卷中的若干问题不问并代为填答；问卷中有些问题不易为被访者理解，或者调查起来比较困难，调查员主动代为填答以保证问卷的完整性，等等。应该认识到，越是困难的情况下，调查员所看重和关心的越是“完成一份问卷”，而不是“严格执行调查或抽样的要求”。前者的结果是较高的回收率，后者的结果才是调查资料的高质量。

概括地说，在由调查的总工作量、调查员的努力程度以及对调查资料的质量要求这三者与调查回收率所构成的相互关系中，调查的总工作量和对调查资料质量的要求往往与调查回收率的高低成反比，而调查人员的努力程度则与调查回收率成正比。在其他条件相同的情况下，一项调查的总工作量越大、对调查资料质量的要求越高、调查人员的努力程度越低，调查的回收率就会越低。特别是对调查资料的质量要求与对高回收率的要求常常在客观上形成冲突并相互制约。而任何一项调查的研究者都无一例外地要在这样一种不可兼得同时又相互制约的关系面前做出自己的选择。

六、高回收率更好吗？

分析了调查中的高回收率何以可能或不可能的各种影响因素，我们就不难回答高回收率是不是更好的问题。表面上看，回收率越高，调查样本的代表性越大。因此，当然是回收率越高越好。但是，这只是一种理论上的回答。如果加上现实的因素，或许过高的回收率并不一定

是件好事。这是因为,在现实社会中,要获得高质量的高回收率并不是一件容易的事情。现实中的社会调查总是会遇到这样那样的困难和挑战。在这些困难和挑战面前,为了保证调查的高回收率,研究者最经常采取的做法就是在调查对象抽取、调查访问实施等操作环节中进行若干改变或替换,以一种易于进行的方式来实施。而正是这些改变和替换,虽然保证了较高的回收率,却在不知不觉中降低了调查资料的质量。

回到前面提出的两个问题。笔者认为,在长达近 30 年的时间内,美国 GSS 调查的回收率没有大的波动,说明这种调查在客观上具有某种内在的规律性:只要调查的总体、对象、程序、方式、方法等条件不发生改变,调查的回收率就是一种相对稳定的事物。换句话说,如果在同一个社会中采用同样的方法进行类似的调查研究,正常的回收率的范围就是相对稳定和可以预知的。一旦调查的回收率超出了这种正常的范围(无论是显著地低于还是显著地高于),那么就一定是有某种特殊的条件发生了改变。至于为什么 GSS 调查回收率的范围一直稳定在 76% 左右,笔者的理解是:这是在美国那样的社会中,开展这种调查所受到的各种外部因素制约的结果,同时也是研究人员特别是调查员严格遵循抽样设计、严格按调查设计的要求,实事求是地进行实地调查的一种客观反映。一方面,美国社会中所特有的政治制度、文化传统、社会心理、生活方式、风俗习惯,特别是各种社会事件,构成了影响和制约问卷调查这种研究方式顺利开展的外部因素,它通常会导致回收率的下降。另一方面,在各种外部阻碍因素面前,研究者和调查员对于调查的实施又具有相当大的自主性。是冒着较低回收率的风险严格执行研究设计的要求,还是为保证较高的回收率而采取各种变通的方式开展调查,是每一位研究者和调查员都必须面临和选择的问题。笔者认为,虽然相对较低但却真实可靠的回收率,比相对较高但却不够真实、不太可靠的回收率更有价值。无论是相对于过低的回收率(如 50% 以下),还是相对于过高的回收率(如 90% 以上),76% 的回收率或许正是 GSS 调查质量优良的一种客观标志。有的美国学者在评价 GSS 调查时指出:该调查“能达到如此高的回收率,就是全国民意研究中心(NORC)的调查小组才能与奉献精神的良好证明”(Dorsten & Hotchkiss, 2005: 350)。为什么 76% 左右的回收率就是“如此高的”回收率?为什么要特别强调调查小组的“才能与奉献精神”?细细品味这段话的意思,我们不难明白:如果不是调查人员的才能与奉献精神,GSS 调查的回收率可能还

会更低。

正是由于十分了解严格执行随机抽样和实地调查所必然面临的许多客观障碍和所必然造成的回收率下降的现象，所以，当听到高回收率时，有经验的研究者就会很自然地对调查实际实施的严格性提出疑问。这就是那位美国教授为什么会以一种对高回收率表示怀疑和将高回收率看作缺点而提问的主要原因。在他眼里，过高的回收率实际上意味着调查过程的不严格性，其结果则是调查资料的低质量。

七、寻求高回收率与高质量的平衡

如何正确看待调查的回收率？笔者认为，这要以一项调查的全部结果质量为标准来衡量。高回收率只是衡量调查结果质量的一个指标，尽管是一个比较重要的指标；另一个更为重要、同时与回收率密切相关的指标则是调查实施过程的严格性以及调查所得资料的质量。追求高回收率只是提高调查质量的一种努力，并不是保证调查质量的一种充分条件；而且在某些情况下，对高回收率的片面追求，有可能形成调查过程的不严格，造成低品质的资料，从而极大地伤害到调查结果的质量。因为在实际调查过程中，无论采用什么样的抽样方法和调查方法，总是会有一部分被调查者由于外出、生病、搬迁、语言不通等原因而无法接受调查，更有一些被调查者会以各种方式和理由拒绝接受调查。任何一个实际从事过调查研究（特别是当过调查员）的研究者，都会对实际调查过程中所存在的各种现实困难深有感受。从一定意义上说，调查达不到100%的回收率是绝对的；要保证调查过程的严格性、保证调查资料的高品质，达不到非常高的回收率也是很正常的。

总之，样本的完整与资料的质量是我们在调查中应该同等关注的两个问题。作为研究者，我们要在“高回收率”和“高质量”两个方面同时下功夫。一方面要设法尽可能地提高回收率，另一方面也要尽可能地防止和避免低质量的高回收率。当高回收率与高质量不可兼得的时候，研究者的任务就是要在“高质量的低回收率”与“低质量的高回收率”之间，寻找到最佳的平衡点，达到调查研究的最佳效果。在这方面，我们或许应该记住巴比教授的忠告：“明显不存在回收偏差比有偏差的高回收率重要得多”（Babbie, 1986: 221）。特别是当对高回收率的追求

有可能造成低质量资料的情况下,这种提醒就显得更加重要。

对于如何才能做到高回收率与高质量平衡的问题,目前并没有可以依据的答案。但笔者认为,要做到这一点,最为关键的一环是要提高调查员的工作质量。实际上,一项调查中大量的制约因素都是客观的,或者说都是难以改变的(比如调查的目标、调查的内容、抽样的方法、样本的性质、样本的规模等等)。真正可以发生变化的影响因素则主要是建立在调查员的工作态度和工作能力基础上的调查员工作质量。这也是调查研究中,研究者十分注意调查员培训、调查员报酬、调查督导和检查等工作的根本原因。这些方式都是从不同的方面在为提高调查员的工作质量做出努力。但是,研究者往往会忽视另一个重要的因素:调查员是普通人。作为普通人,他不可能超出某种生理的和心理的极限达到研究者所期待的工作质量。笔者多次遇到这样的情况:一些受过系统的社会调查训练、具有多次调查实践经验(即有很好的工作能力)、做事认真负责(即有很好的工作态度)的调查员,用自责和内疚的语气对笔者说:由于调查任务实在太大,为了完成调查,后来只好马虎应付了。这样的结果必然是回收率保证了,资料的质量却下来了。针对这一现象,笔者从自己的调查实践中总结出一点具体操作方法上的心得,提出来供广大同行参考。

笔者的基本指导思想是将调查员当作普通人。从这一前提出发,将调查员的日平均工作量以及在一项调查中的总工作量都限定在一定的范围内,从生理和心理上保证调查员的工作质量。笔者提出的调查员日平均工作量的标准是:一般规模的问卷(半小时左右的访问工作量),调查员每天完成的数量最好定在4—6份的范围内。而调查员的总工作量标准是:对于一般中小规模的调查来说,最好限制在30份问卷的范围内,对于较大规模的调查也应该限制在50份问卷的范围内。无论是超出了日平均工作量,还是超出了调查的总工作量,都有可能带来调查质量的下降。

由于受到每人日平均工作量和总工作量两个因素的限制,所以,研究者在研究设计中,要以这两个条件为基础,对调查的时间和调查员规模进行相应的计划和调整。具体做法是:研究者先根据调查的样本规模和上述调查员日平均工作量,计算出一项调查所需要的总的工作日数量;然后,结合上述调查员的总工作量标准来决定合适的调查员数量以及集中调查的时间长短。以笔者前面谈到那项调查为例,其样本规

模比较大,达到1500份,按人均日工作量5份计算,共需要300个工作日。对于这300个工作日如何分配调查员数量和集中调查的时间,又会有多种不同的选择。比如可以用10个调查员做30天、或者用20个调查员做15天、或者用30个调查员做10天,还可以用60个调查员做5天等等。在其他条件相同的前提下,比较可取的方案是第三种,而第一种、第二种和第四种都不可取。原因是,第一种方案不仅集中调查的时间太长(这样会带来严重的生理疲劳和心理疲劳),更重要的是每个调查员的总工作量太大(每人为150份,大大超出上述标准,即超出了普通人的生理和心理极限);第二种虽然在时间上有所缩短,但每个调查员的总工作量仍然大大超出上述标准(每人为75份);而第四种虽然调查的时间最短,每个调查员的总工作量也最小(每人为25份),但因所需的调查员数量过多而带来一系列新的问题,如加大组织与管理的难度,调查员人数过多还增加了调查员能力差别控制的复杂性,等等,以致从另一个方面大大增加调查的偏差。

参考文献:

- 艾尔·巴比, 2000,《社会研究方法》, 丘泽奇译, 北京: 华夏出版社。
- 风笑天, 1987《浅谈当前抽样调查中的若干失误》,《天津社会科学》第3期。
- , 2000,《论社会调查方法面临的挑战》,中国社会科学院社会学研究所编,《中国社会学年鉴(1995—1998)》,北京: 社会科学文献出版社。
- 石原邦雄, 2000,《日本家庭研究动向及日中家庭研究课题》,《社会学研究》第6期。
- Babbie, Earl 1986, *The Practice of Social Research*. 4th edition. Wadsworth, Inc.
- David, Matthew & Carole D. Sutton 2004, *Social Research: The Basics*. Sage Publications Ltd.
- Dorsten, Linda Eberst & Lawrence Hotchkiss 2005 *Research Methods and Society*. Pearson Education, Inc.
- Nachmias, Chava Frankfort & David Nachmias 2000, *Research Methods in the Social Sciences*. Sixth edition, Worth Publishers.
- Neuman, W. Lawrence 1994, *Social Research Methods*. Allyn and Bacon.

作者单位: 南京大学社会学系
责任编辑: 张宛丽

deeply shaped by the state's policies and institutional arrangements. Meanwhile, the state's institutional arrangements of consumption have much to do with the situation of incentives to labor. Furthermore, the relationships between consumption institutions and incentives to labor vary with the state's legitimacy resources. Thus, there is a logical path underlying the change of the institutional arrangements regarding consumer lives and labor motivation, and this path is closely related to the change of the state's legitimacy resources.

The Medieval Cities and the Origin of Modern Citizenship: The heritage of Weber's sociological study on city *Wang Xiaozhang* 99

Abstract: Citizenship is a core concept of modern politics, and the modern citizenship of some special western cities originated in the late middle ages. Through a reconstructive reading and interpretation of Weber's study on cities this article points out that by comparative studies between the eastern cities and the western cities, between medieval cities and ancient cities, and via the hidden dimension of religion in his sociological study on city, Weber showed to us the structural, economical and cultural conditions under which the western medieval cities were capable of giving birth to the early shape of the form of modern citizenship. Weber's analytic framework with respect to the medieval cities and the origin of the modern citizenship is of enlightening significance to our understanding and reading of the recent theories related to citizenship, including the disputes between Communitarianism and liberalism.

Is High Response Rate Better? Another understanding on response rates for social survey *Feng Xiaotian* 121

Abstract: Based on the query of American scholar, the paper discusses the problem of response rates for social survey and its methodological significance. Using the results of American GSS for reference, the author points out that on the one hand, as the size of a sample represented by the response rate is the final indicator of representativeness of a survey result, so the higher response rates the better. On the other hand, there are many factors restricting and affecting the response rates for social survey, and if seeking an impractical high response rate it would be likely to affect the quality of the survey. In this case the higher does not equal to better. We should pay attention to both the completeness of the sample and the quality of the information to increase response rates as well as avoid high response rates with lower quality.

Is Falsification Possible in Social Science? *Zhang Yang* 136

Abstract: This article discusses the methodology of Popper's methodological falsificationism