

教育研究方法

袁振国 主编

高等教育出版社

责任编辑	林丹璐
封面设计	杨立新
责任绘图	
版式设计	马静如
责任校对	夏 晔
责任印制	

内 容 提 要

本书是教育部“高等师范教育面向 21 世纪教学内容和课程体系改革计划”的研究成果，是面向 21 世纪课程教材。本书结构完整、内容新颖，既清晰明了地介绍了定量研究方法，又充分介绍了新起的定性研究方法。本书列举了两百多个成熟的研究事例，配备了大量的图、表，使得本书实用具体、操作性强。

本书可作为高等学校教育学科的教学用书，也可供广大的教育工作者和教育研究人员使用。

图书在版编目(CIP)数据

教育研究方法/袁振国主编．—北京：高等教育出版社，2000．6
面向 21 世纪课程教材
ISBN 7 - 04 - 008528 - 3

．教... ．袁... ．教育学 - 研究方法 - 教材
.G40 - 034

中国版本图书馆 CIP 数据核字 (2000) 第 17585 号

教育研究方法
袁振国 主编

出版发行	高等教育出版社		
社 址	北京市东城区沙滩后街55号	邮政编码	100009
电 话	010 - 64054588	传 真	010 - 64014048
网 址	http:// www hep .edu .cn		

经 销 新华书店北京发行所
印 刷
纸张供应 山东高唐纸业集团总公司

开 本	787 × 960 1/ 16	版 次	年 月第 版
印 张	14.5	印 次	年 月第 次印刷
字 数	260 000	定 价	13.10 元

凡购买高等教育出版社图书,如有缺页、倒页、脱页等质量问题,请在所购图书销售部门联系调换。

版权所有 侵权必究

目 录

第一章	教育研究的意义和过程	1
1	教育研究的意义	1
2	教育研究的过程	3
3	教育研究的分类	9
4	研究过程中的活动	13
第二章	研究问题的确定	18
1	研究问题的选择	18
2	研究问题的陈述	20
第三章	查阅文献	36
1	查阅文献活动	36
2	集中和总结信息	38
3	解释和利用信息	42
第四章	定量研究的设计	47
1	研究设计的目的	47
2	控制差异的概念	48
3	良好设计的特征	55
第五章	实验研究	59
1	实验设计的含义	59
2	“好设计”实验的标准	62
3	仅施后测控制组设计	66
4	前测—后测控制组设计	68
5	所罗门四组设计	70
6	因素设计	72
7	重复测量设计	74
8	时间延长设计	76
9	对实验结果的解释	78
10	随机性和代表性	81
第六章	准实验设计	84
1	效度问题	84
2	仅施后测、非对等控制组设计	85
3	前测—后测非对等控制组设计	87

	4 时间系列设计	90
	5 单个被试设计	95
	6 行动研究与准实验研究	104
第七章	调查研究	106
	1 调查研究的范围和描述	106
	2 调查设计	109
	3 调查研究方法论	112
	4 问卷调查	114
	5 访谈调查	127
	6 调查结果的分析和评论	133
第八章	定性研究	137
	1 定性研究的概念、意义和理论基础	137
	2 定性研究设计	142
	3 定性研究中的伦理问题	146
第九章	文献研究	149
	1 文献和文献研究法	149
	2 文献研究法的定性分析和定量分析	150
第十章	比较研究	158
	1 比较研究法的发展历史与作用	158
	2 比较研究法的概念及种类	161
	3 比较研究法的运用	163
	4 藏、汉儿童数学思维能力发展差异性的比较研究案例	166
第十一章	实地研究的方法	169
	1 现场研究	169
	2 参与观察	172
	3 访谈	176
	4 搜集资料的其他形式	179
第十二章	定性研究资料的分析	182
	1 通过分析发现主题	182
	2 定性资料的编码	184
	3 分析资料的过程	185
	4 定性研究报告的写作	187
第十三章	两个案例	190
	1 一个个案研究的实例	190
	2 一个课堂观察的实例	194
第十四章	行动研究	206
	1 行动研究的理论背景：观念的变化	206

2	行动研究的缘起与特征	210
3	行动研究的基本过程	213
4	行动研究对我国中小学教师的意义	219
后记		222

第一章

教育研究的意义和过程

1 教育研究的意义

教育研究是教育变革自身的要求

人类的活動可以粗略地分成三个层次：本能水平、经验水平和有意识的反思水平。本能活动是维持我们生存的基本活动，属于种族遗传性质；经验活动是人类将自己一代代活动的经验积累起来，传给下代，属于原始模仿性质；而有意识的反思则是将活动的主体和客体分离开来，对活动的特点、过程和规律进行理性的分析，属于研究性质。当一种活动进入研究水平，也就意味着我们从事这项活动的自主性达到了一个新的境界，具备了从自然王国向自由王国飞跃的可能性。教育研究正成为教育活动的有机组成部分，成为广大教育工作者的自觉要求，这充分反映了教育发展的时代特点。

教育活动可以说是与人类文明同步发展的，学校教育活动在我国也已有 3 千年的历史了。在漫长的教育活动史上，涌现出了许多著名的教育家，比如孔子、孟子、董仲舒、韩愈、朱熹、王夫之等等，他们提出了许多精辟的教育见解，对人们加深对教育意义和教育规律的认识有深刻的意义。但是，从整体上说，古代教育基本上是经验性教育，教育活动的內容和形式变化都比较缓慢，鲁迅在《从百草园到三味书屋》中所描绘的景象，很形象地反映了中国一以贯之的教育状况。20 世纪以后，随着社会变化的加剧，教育的变革也成为社会变革的一个重要方面。从而产生了一大批像蔡元培、陶行知、梁漱溟、晏阳初、陈鹤琴等教育改革家，产生了对教育性质、特点、效率的回顾和反思，形成了一大批教育改革理论。可以这样说，没有教育研究，就没有教育的理论，没有教育研究，就没有教育家。

现在，社会的发展更迅速，教育的差异性更明显，教育变革的要求更深刻，教育研究的迫切性也就更强烈。

教育研究是新世纪教育工作者的必备素质

1999年6月17日，中共中央国务院发布了《关于深化教育改革全面推进素质教育的决定》，《决定》将教师队伍建设列为一个部分专门予以阐述，指出，建设高质量的教师队伍，是全面推进素质教育的基本保证，系统提出了新世纪教师的必备素质，强调：教师“要遵循教育规律，积极参与教学科研，在工作中勇于探索创新”。把参与教育研究作为教师素质的一个必要内容提出来的确是以前从未有过的。之所以如此，是由时代要求所决定的。

教育研究是教师专业化的要求 1966年联合国教科文组织《关于教师地位的建议》明确指出：教师职业是一种“专业”。到了20世纪60年代后期，联合国还要特别强调教师是一种专业，说明对教师职业的专业性认同程度还不够。认同程度之所以不高既有主观的原因也有客观的原因。从主观上说，历来有一种看法，认为“学者即良师”，只要有知识、有学问就可以作教师，没有意识到一个合格的教师不仅要有知识和学问，还要有与教师职业相应的品格和技能，要有对教育规律和儿童成长规律的深刻认识，要有不断思考和改进教育工作的意识和能力；从客观上说，人们不会对律师、医生、会计的专业性发生任何怀疑，也不会有一名从未学过律师、医生、会计专业的大学生去直接谋求这些职业。而对于教师职业很多人就有胆量来尝试，因为它的专业化程度还没有达到与律师、医生、会计等职业同样的程度。但随着教师职业专业化程度的不断提高，必然要求教师不仅具有扎实的学科基础，而且要有教师职业的独特品格和能力，这就是《决定》所要求的“双师型”教师。在这其中，教育研究的意识和能力是非常重要的一个方面。另一方面，教师教育研究的意识和研究能力又是推进教师职业专业化的有力保证。

教育研究是教育创新的要求 传统的教育工作在相当大的程度上是知识传授、行为训练的工作。其实，教育工作者不仅要传授知识，更要启发智慧；不仅要训练行为，更要培养人格。我们现在实行的是班级授课制，但班级授课制是以这样的假设为前提的：相同年龄的儿童其发展水平和发展潜力相同，所以对相同年龄的儿童以同样的速度、同样的要求教授同样的内容，并以同样的标准进行统一的评价。然而，这一假设是不成立的，在智力上，7岁儿童的实际智力水平在4~11岁之间；在性格上、认知风格上，五颜六色，“人心不同，各如其面”；在家庭背景上，政治、经济、文化的差异十分巨大。这就注定了班级授课制一开始就种下了教育的一系列祸根：非针对性、去个性化。教育的理想是促进不同的儿童根据各自的特点得到更好的发展，如何根据不同儿童的不同特点进行教育，不仅需要有教育的智慧，更需要有对儿童细心的观察和精

心的研究。这样，才能不断突破别人和自己的经验，根据不断变化的教育对象和教育内容，因材施教。

教育研究是提高教育质量、形成独立教育教学风格的要求。随着社会要求的不断变化，教育的质量观也不断变化。如何不断提高我们的教育质量，没有现成的、千篇一律的经验可以模仿。事实上，绝对的教育质量是没有的，在现代社会，有特点才有质量，没有特点就没有质量。如何根据教育对象、教育要求和教师的自身特点，形成自己的教育教学风格，需要每一个地区、每一所学校、每一位教师自己去探索、去研究。

教育工作充满了复杂性和丰富性，教育工作没有“最好”，只有更好。任何一位教育家的成功经验都是有条件的，都是根据特定的对象、特定的要求、特定的内容和特定的自身条件创造出来的。一个人如果不加区别地模仿，即使模仿得再好，也只能形似神不似。一位教育工作者有没有自己对教育本质的思考，有没有自己对教育的个人见解，有没有自己对改进教育工作、提高教育质量的孜孜追求，并最终形成自己独特的教育、教学风格，是区分教书匠与教育家的分水岭。所有这些，都是以强烈的研究意识和自觉的研究活动为基础的。而掌握研究方法，则是少走弯路，尽快进入研究角色的捷径。了解和熟悉教育研究的方法是保证研究的科学性和高效率的必要条件。

教育研究是教育变革自身的要求，是新世纪教育工作者的必备素质。了解和熟悉教育研究的方法是保证研究的科学性和高效率的必要条件。

2 教育研究的过程

教育研究可以说是由一连串有因果关系的步骤构成的。学者们对科学研究方法的步骤可能有许多严格的分类，但典型的划分通常都是4~6个步骤。从科学方法的一致性，从研究过程提供的基本的、系统的要素看，可分为5个步骤：(1) 确定问题，(2) 查阅文献，(3) 收集资料，(4) 分析资料，(5) 推导结论。要使研究系统化，首先，待研究的问题的性质必须确定，即使只是一个轮廓。相关的知识确定后，重要的是，要建立一个进行研究的框架。与建立研究框架或基础紧密相关的，是明确必要的假设以及与所研究的问题有关的条件。

第2步是查阅别人研究类似问题的信息。毫无疑问，一个人可以从他人的研究中得到很大的收获。文献研究就是这样的信息来源。

系统研究的第3步是收集资料。收集资料不能带着实用的、随意的、无准备的态度进行，收集资料的过程需要适当的组织和控制，以便能够对手头的问题作出有效的决策。

第4步是针对问题作资料分析。

第5步是分析资料后，作出总结或得出具有普遍意义的结论。结论需在研究的工作框架里，在对资料的分析的基础上获得。

确定问题

查阅文献

收集资料

分析资料

得出结论

研究过程已被描述为是有系统的和有秩序的，但千万不要认为研究过程是死板的。研究步骤是可以跳跃或交叉的。有些研究，比如实验研究、验证假设的研究，问题已经被确定了；而在定性研究中，通常的假设可能在资料收集完以后也没有形成，直到对资料进行了分析后才得出。所以，特定的研究情况与上述的研究步骤可能是冲突的，但尽管如此，所有教育研究都是有系统的。

教育研究是有系统的，在一个大的框架中遵循着科学的方法步骤。但是，不同的研究类型在如何完成这些步骤时有较大的灵活性。

教育研究的效度

不管研究的形式是怎样的，也不管研究指向哪个目标，我们都期望研究是有效的，也就是具有效度。什么是研究的效度呢？通常，我们说某件事是有效的，是希望它基于事实或证据，是“能够被证明的”。更具体地说，效度同时包含两个概念，内在效度和外在效度。内在效度是指研究结果能被明确解释的程度，而外在效度是指研究结果能被推广到其他总体条件、时间和背景中的程度。请看下面解释效度概念的例子。

内在效度 假设一位体育教师要研究两种不同的练习方案对初二年级男生体育技能测试成绩产生影响的情况。这个教师计划用上午时间，并且可以随机在每个班级指定28个男生。他将这两种训练方案分别运用到两个班级中，时间同为16个星期。最后，由他本人负责对两个班进行同样的体育技能测试。这项研究具有较好的内在效度。这个研究的整个计划如图1.1所示。

为什么说这个研究具有好的内在效度呢？因为，如果两个班的体育技能测试成绩有差异，譬如说，2班的均数比1班高，我们可以充满自信地解释这个结果，并得出方案2比方案1更有效的结论。不论结果怎样，都可以较有把握地得出解释，因为总的来看，两个班级之间惟一不同的是练习方案。两个班级都是上午上课，同一个教师任教，在同一所学校，并且参加测试的每个班级都

由初二年级男生组成。肯定地讲，任何两个男孩在原有的体能方面存在不同，但是 56 个男孩是由每班随机指定的 28 个组成的。总的来看，两个班级中的其他因素与方案相比对体育技能的可能影响是非常相似的。虽然会有些个别差异，比如可能有些男孩比其他男孩睡得多一些，但总的平均来讲，两个班级大致是一样的。实质上，随机的指定使两个班级和其他因素变得相同了，因此我们说这两个组在随机涨落范围内是相当的。

学校 教师 A 方案 1	初二年级 1 班男生 随机指定	由教师 A 对两个班进行同样的体育技能测试
学校 教师 A 方案 2	初二年级 2 班男生 随机指定	

方案实施 16 周

图 1.1 有高内在效度的假设性研究的总体研究计划

请看另一个缺乏内在效度的代表性的例子。这一研究是关于不同类型的教材对初二年级自然科学课学习有什么不同影响。现抽出 3 位教师参与研究，他们分别在不同的学校教学；其中两名教师每人教 4 个班的初二年级自然科学课，一名教师教 3 个班。其中有一所学校的班级是按照能力分班的。碰巧，参与研究的这个学校的这名教师教的是能力高的班级。每位教师使用 1 种教材，试用教材的时间为 9 周。教师们使用不同的材料，并且没有教师使用超过一种以外的材料。9 周后，对学生进行自然科学的成绩测试，每位教师使用他（她）自己出的试卷。整个研究计划如图 1.2 所示。研究者计算出试用 3 种教材的自然科学的平均分数。对于 3 种教材的相对效果，我们能得出什么结论呢？简直可以说什么都不能得出。假设学校 III 的学生得分最高，是因为这些学生的能力高或教师 C 教得很出色吗？或教师 C 出的试卷比其他教师的容易？或者教材 3 比其他两种教材更好？我们无法从结果中得出有效的解释。不管是什么形式的结果，任何关于结果的雄辩的解释都不能得出“教材 3 效果最好”的结论。这样，这个研究就缺乏内在效度，因为结果不能被解释。

内在效度是指研究结果能被明确解释的程度。

学校 教师 A 教材 1	4 个班级 不同能力分布	自然科学成绩测试 教师 A 出的试卷
学校 教师 B 教材 2	3 个班级 不同能力分布	教师 B 出的试卷
学校 教师 C 教材 3	4 个班级 高能力分布	教师 C 出的试卷

教学 9 周

图 1.2 缺乏内在效度的假设性研究

外在效度 我们前面已经下过定义，外在效度是指研究结果能被推广到其他总体条件、时间和背景中的程度。在很大程度上说，内在效度是外在效度的先决条件。如果结果不能被解释，也就不大可能进行推广。请看这样一个例子：

某一地区有 5 所小学，进行一项关于学生家长们的课程质量、管理效果、纪律等看法的调查。事先形成了一个结构良好的电话交谈的条目细则，并且有足够条目覆盖住家长关心的学校特征。在每所小学有 25 个学生家长被随机选出并进行交谈，对于那些找不到或不愿意交谈的，有一个随机取代的规定程序。交谈结束后 125 个交谈者中只有 6 个父母因找不到而被重新替代。

这项研究的结果可被推广的人群是那些孩子们在这 5 所小学就读的父母们。按照这种调查进行的方式和完成的交谈的数目，这个结果能很有把握地推广到这些人群中去。这项调查研究具有高的或是说好的外在效度。但如果把这项研究结果推广到其他学校系统的其他小学中就不会有多少意义了。如果要这样做，需要一个逻辑基础，如以这些学校小学生的父母与其他学区小学的学生的父母的相似性比较为条件，也就是说，要证明不同地区孩子父母的情况是相似的。

假如在上面所述的研究中，由于其他原因，只有 27 个学生的父母或在预计的 125 名家长中只有 22% 的交谈过。在这种情况下，这个研究就非常缺乏外在效度。不能说这 27 位交谈过的家长是预期 125 位家长以及这些学校中其他学生的家长或者其他对于这个问题的有关人员的代表性样本。

外在效度可能包含了对人群或条件的概括，再看看另一个更为极端一点的例子，缺乏外在效度的情况就可以更清楚了。

研究者要研究作用于视觉时间的长度对回忆无意义符号的影响（无意义符号可以是任意排列的字母）。研究对象（被试）是来自教育心理学专业的 10 个研究生志愿者，时间长度 5 分钟，每两个人接受一种检查。分别呈现给每个被试 20 个无意义符号，每次呈现后，被试即回忆符号，根据回忆的正确情况得出个人的表现分，整个计划见图 1.3。

呈现时间：1 组	志愿者 1 志愿者 2	总表部分
呈现时间：2 组	志愿者 3 志愿者 4	
呈现时间：3 组	志愿者 5 志愿者 6	
呈现时间：4 组	志愿者 7 志愿者 8	
呈现时间：5 组	志愿者 9 志愿者 10	

呈现 20 个符号

图 1.3 缺乏外在效度的假设性研究

假如表现分数随呈现时间延长而提高，但这一结果能推广到哪一人群和哪种条件呢？能推广到中小学生学习有意义材料的过程中去吗？能推广到在活动程序很严格的条件下工作的年轻人中去吗？结果甚至不能推广到研究生群体中，因为参与者是志愿的。总之，结果只适用于回忆无意义符号的 10 位志愿者，因而研究缺乏外在效度。

外在效度就是研究结果能被推广到其他总体条件、时间和背景中的程度。

当然这并不是说，为达到外在效度，结果必须能推广到许多不同的人群和条件中去。例如，如果研究只涉及天才学生，研究目的就只是推广到天才学生而不是所有学生。如果一个学校系统做一个需求评估的研究，其结果可以只推广到那个系统。特别是进行定性研究，研究者并不特别关心结果的广泛适用性，因此外在效度依赖于特定研究工作的条件及目的。

研究的效度只是一个度的问题，事实上也不可能得到纯粹的内在或外在效

度。以后几章将分析到，旨在提高内在效度的企图可能降低外在效度，反之亦然。研究者需要求得一种平衡，使某一结果得到合理的解释，又具有某种程度的推广价值。

研究的效度指结论能被明确解释的程度（内在效度）和结论的普遍性（外在效度）。

教育研究的信度

讨论效度时，有必要讨论一个相关概念——研究的可靠性，即信度。信度指研究的前后一致性以及研究能在多大程度上重复。有时我们区分出内在信度和外在信度。内在信度指在给定的相同条件下，资料收集、分析和解释能在多大程度上保持一致。如果使用多个人收集资料，内在信度的问题便是：收集人之间能达成一致吗？如果对教师行为进行研究，使用课堂观察方法收集资料，内在信度的问题便是：两个或更多的观察者在看待同一教师的表现时，能达成一致意见吗？这可称之为观察者之间的协同程度。如果缺乏内在信度，资料仅是收集者的一个函数，即随观察者眼光的不同，观察的结果不同，这就不能反映真实发生的情况。

外在信度涉及的是一个独立的研究者能否在相同或相似的背景下重复研究。如果能的话，那么结果是否前后一致？如果研究是可靠的，一个使用相同方法、相同条件的研究者应得到与先前研究相同的结果。为保证可重复性，研究中必须包括对研究过程和条件的充分界定，不同研究所需界定的方面可能不一样。

信度是效度的一个必要的保证，也就是说，一项研究不可能没有信度却有效度。如果一项研究是不可靠的，我们就不可能有信心去解释结果，并将它推广到其他的人群和条件中去。从本质上说，信度和效度共同构成了研究的可靠性。

研究的信度指研究的方法、条件和结果是否可重复，是否具有前后一贯性。

研究形式的多样性

研究的形式是多种多样的，有时我们听人讲：“我研究了这个问题”，意思

是说，他或她已经对同一问题的不同解决方案有所思考；一个小学生可能说：“教师要求我们研究巴西”，意思是说老师布置了作业，要求阅读有关巴西的材料；一个教育心理学专业的研究生可能正在实验室里做一个关于青年人完成复杂学习任务的实验；语文教师们在学校中正试用各种不同方法对那些阅读能力差的学生进行补偿阅读指导；一个教育社会学家也许正在一所高中研究社会相互作用的模式。这里都使用了“研究”一词，但显然目的、水平层次是大不相同的。

鉴别和描述不同形式和种类的教育研究将帮助我们形成对教育研究的范围及各种研究的不同差别的理解。不同形式和类型的研究包括多种分类系统，下面将对此问题进行细致讨论。

3 教育研究的分类

教育研究的分类可以有好几种方法。已经有人按事物的复杂程度来确立分类系统。本文介绍两种系统，一种基于研究的目标或目的，另一种基于研究的方法。后一种是两种层次分类的缩略。介绍教育研究方法的分类系统，为的是使读者对教育研究的性质的背景增加了解。

基础研究和应用研究

基础研究与应用研究是根据它们各自不同的目标或目的来区分的。应用研究的目的是解决当下的、实际的问题，这样的研究是指向特定的问题；基础研究的指向则具有普遍性，它可以为现有学科的知识体系增添新的东西。基础研究不一定要有当下的、实际的用处，即使有，那也是附带的，并非它最初追求的目的。另一方面，在解决实际问题的过程中，应用研究也有助于一般性领域的扩展，基础研究和应用研究都是重要的，不应以价值判断的等级来区别它们的高低。但是，有些人会对使用“基础”和“应用”研究这两个术语产生误解。其中之一是，许多人觉得，从方法论上讲，基础研究很复杂，而应用研究则较简单；另一种误解是，认为应用研究是由那些非专业化的实际工作者做的，而基础研究则是善于抽象思维的不切实际的思想家进行的；还有一种误解是，认为应用研究是粗糙的、无计划的，但又是有用的；而基础研究则是精细的、准确的，但却少有甚至没有实用价值。在前面我们已经说过，区分这两种研究，并不是从它们的复杂程度或价值，而是从它们的目标或目的来区分的。

基础研究和应用研究是根据它们的目的来区分的。基础研究的基本目的是扩展知识；应用研究的基本目的是解决当下的实际的问题。

行动研究可视为应用研究的一种，它是由教师或行政官员来操作的，以帮助地方学校决策和教师改进教学的方法。行动研究强调对基层的日常问题的解决。它很少关心研究结果是否对教育情境具有普遍适用性，它研究的人群常常也很小（如一所中学里上生物课的班级）。

假如一名初中自然科学的教师要考察在模拟实验中，是按计划布置练习册的作业还是应该布置附加作业，那他进行的行动研究只涉及上自然科学课的学生，以此来判定这两种方法的相对效率和效益。教师们只关心自己的情形，不必考虑能否将研究推广到其他学校去。

在方案设计和方法论方面，行动研究不如其他教育研究方法的要求严格，研究通常是整个群体参与，而在有些情况下，只是某一群体参与。然而，根据研究文献表明，行动研究能为基层的教育决策提供有用的、可行的方法。考虑到行动研究在我国正有越来越多的学校在运用，我们在本书的最后一章将作专门介绍。

行动研究通常是教师和教育管理人员为解决具体的问题，或为基层决策提供信息而使用的方法。

定性研究和定量研究

从极端典型的意义上说，定性研究是用文字来描述现象，而不是用数字和量度，定量研究是用数字和量度来描述对象，而不是用语言文字。当然，在实际教育研究过程中，绝对的、纯粹的定性研究或定量研究是不多的，它们通常是相互支持的。但任何一项研究在总体上必定有一个界定，要么是定性的，要么是定量的。这决定了研究设计的性质和基本思路。

尽管在定量研究和定性研究的概念和方法上存在不同，但这种区别就实施研究来说不是两分的，而是一个定性定量的连续统一体。定性研究基于描述性分析，本质上是一个归纳的过程，即从一般的特殊情境中归纳出一般性的结论。另一方面，定性研究与演绎法更相近，是从一般的原理推广到特殊的情境中去。

从认识论的角度看，定性研究与定量研究存在着根本的差别。它们基于不同的范例，范例就是由假设、概念和建议组成的一个模型。定性研究从最纯粹

的意义上讲从属于自然主义者的范例，即研究应在自然的情境中进行，研究所获得的意义也只适应于特定的情境和条件。定量研究根源于实证主义，与定性研究相比更接近于科学的方法。定量研究者对结果和产品预以极大的重视；而定性研究者比定量研究者更注重过程的影响。

定性研究并不强调在开始研究时对所研究的问题有一个理论基础。一个理论可以在研究的过程中形成，但随着研究的进行，理论又会被改变，被放弃或被进一步地精练。如果理论是基于数据而形成的，我们便有了一个“实在的理论”的概念，也就是理论是基于数据而不是基于一些已经形成的观点、想法或系统。如果没有理论出现，这个研究将是非理论的，但将获得描述性价值。在另一方面，由于定量研究自然具有的演绎性，从开始便更倾向于以理论为基础。事实上，当进行一项理论检验研究工作时，它很可能是定量的研究。当然在定量研究中，理论并不总是明晰确定的，但理论的基础存在于这样或那样的形式中。

总的看来，定量研究者比定性研究者更强调标准研究程序和预先设计。后者进入研究更灵活，定性研究比定量研究更经常地包含多种研究方法。定量研究比起定性研究来有更多的设计目录，即通常使用的设计方法和种类。其中最明显的区别是资料呈现的形式。定量研究主要是通过数据的展现说明统计结果，而定性研究主要是叙述性的说明。

定性研究和定量研究在目的上也是不同的。定性研究旨在理解社会现象，这里的“社会”一词是就其广义而言的，定量研究则旨在确定关系、影响、原因。当谈到教育研究时，这两种目的又是相关的，对教育的改进来说，确实非常相关。所以，这两种研究类型都是有价值的，事实上，对于理解影响教育的诸多因素，它们二者是相辅相成的。

尽管定性研究和定量研究具有不同的基础假设，并且具有不同的程序。但在方法论上，把它们视为连续体是有益的。因为，从实践的角度看，定量研究、定性研究的程序是经常混杂的。无论如何，各种方法论都可以被置于从定性研究到定量研究的连续体中。最典型的定量研究是实验研究，其次是准实验研究，介乎于定量研究与定性研究之间的是调查研究，然后是文献分析研究和比较研究等。80年代以后，实地研究方法日益兴起并有较快的发展。

定性和定量研究各有它们自己的特征，但在教育研究的运用中，它们的连续性多于他们的两分性。

实验研究 如果教育研究被称为“实验”的研究，那它至少有一个变量，即实验变量，它是由研究者人为地改变的，以便确定对变化的影响。这就是说，实验的变量是什么以及它的变化范围有多大，是研究者选择的。在一项实验中

实验变量可能不止一个，并且，为了使已经完成的研究被认为是一个“真实实验”，实验的参与者在实验处理中应是随机指定的。下面举 1 个例子加以说明。

假设一位研究健康教育的人，要研究 3 种不同锻炼时间（时间的长度）对稳定心率的作用。他随机抽取 60 名成年人，20 人为一组。60 名被试都按特定的要求每天锻炼，三组每天锻炼的时间分别为半小时、45 分钟、1 小时。每一时间长度的运动量是确定的，锻炼计划为期两个月。被试稳定心率的情况在实验前后分别测试，以便测量到变化。

这里实验的变量是时间的长度。这一变量——3 种水平的锻炼由研究者设定施加于被试，每 20 人一种水平。如果不同时间长度的锻炼对心率有影响，这种影响必须通过 3 个实验组（平均）心率的差异来反映。

准实验研究 准实验研究与实验研究一样包含一个或多个实验变量。但是，与参加者被随机地分配的实验处理不同，研究中被试是被“自然”地分配组别，比如班级。成员进哪个组是自我选择的。准实验研究中，也是运用包括实验处理在内的单一被试设计。因为随机分组的努力经常遇到困难，所以在教育中更经常地使用准实验研究。下面举 1 个例子加以说明。

假设一个研究者研究的是，一项教学计划在教学概念的测试中解决 6 级运算逻辑问题的效果。实验的方法是，每天进行半小时解决逻辑问题的教学，为期 10 周。研究者不可能随机分出 6 种水平的班级来，但有 8 个现有的班级接受了教学计划，另有 8 个班级作为对照组。10 周以后，有 16 个班的学生接受了普通数学概念的测试，以比较实验组与对照组的成绩差异。

调查研究 很多研究工作可以置于调查研究的名称之下。一般地说，调查研究涉及教育、心理、社会变量中的事件、分布和关系。不像实验研究中的变量是人为的。调查研究的变量是现有自然情境中的变量，有些调查只局限于现状，有些则试图确定变量之间发生的关系和影响。后一种情况我们将它称为回溯研究（ex post facto research）。下面举 1 个例子加以说明。

一位研究者要研究学校顾问的专业性的参与情况，就必须运用调查研究。方法之一是，设计一种合适的方法，多半是问卷的方法，调查部分被选定的顾问或对全部顾问进行调查。顾问们的回答对其专业性参与情况提供了一个基本的概貌。从这些信息中，研究者可以阐述相关的重要特征（顾问们认为的）以及参与的频率。这一研究强调的是参与的特征。

再看一个包括回溯研究的调查。一项研究涉及的是，在小学高年级学生数学、词汇等的认知能力和技能领域，对学校的态度与成绩之间的关系。研究者需要对学生进行分类，包括对测量态度和认知能力、技能成绩的手段进行适当的归类。没有变量是人为的，也就是说，研究者不能干扰学生或影响他们的态度或运算的得分。数据收集以后，研究者试图确定各种可能存在的影响和解释

这些影响是怎样作用的。比如，一个肯定要考虑的问题是：“是否某种态度模式总是与特定的成绩得分相联系？如果是，那么是这种态度影响成绩得分吗？”

文献研究 文献是记录已有知识的一切载体，是把人类知识用文字、符号、声像和视频等手段记录下来的所有资料。对这些资料进行分析研究以从中获得某一结论的方法就是文献分析法。教育研究人员常常倾向于对文献进行定性分析，应用逻辑推理探索事物之间的联系。

比较研究 比较研究是根据一定的标准，对两个或两个以上有联系的事物进行考察，寻找其异同，探求某些共同特点或发展规律的方法。比较研究是教育研究中的一种重要方法。

实地研究 (ethnographic research, 一译人种学研究或质的研究) 实地研究与人类文化学有密切联系，它在教育领域内的运用正日渐增多。实地研究是在广义的文化概念下对特定的文化作深入的、解析性的描述。就教育的情况而言，我们将实地研究定义为：提供对特定情况下的教育制度、过程和现象的确切描述的过程。由于实地研究是研究人员面对面的、直接参与活动过程的研究，故而得名。实地研究依赖于对研究现象的观察、描述、定性判断或解释。它发生在自然的过程中，重视过程，旨在获得整体画面。实地研究经常没有明确的理论指向和结论假设。

上面所描述的教育研究的分类的目的是为进行教育研究有效地组织信息。除了这种分类，也有其他不同层次和复杂性的分类系统。一些作者用诸如描述性研究和原因——比较研究的分类方法，而另外一些作者可能扩展方法论的范畴，如把个案研究也看做是一种基本方法。

本书没有把个案研究作为一般方法的单独类型，因为个案研究（对单个组或单一个人的仔细考察）可贯穿在其他的研究类型中。定性研究当对单个组进行深入研究时，也就是个案研究。个案研究通常与定性研究相联系，但也不限于定性研究。准实验研究中的单一被试研究就是个案研究。

上述数例从一般的方法论分类，简单、表面地描述了各种研究的类型，但它们已经能够说明各自所包含的意思。当然，任何一项具体的研究还需要对研究的问题和假设作适当的、更详细的解说。我们将在以后各章涉及到各种研究方法时再作详细例解。

4 研究过程中的活动

研究的系统过程和科学的方法导向一般的活动，包括进行一项学术研究。这些活动类似于前面描述的5个步骤，本节将作详细说明。这些活动并不局限

于某一类型的研究。

在总结一般活动，包括进行一项研究活动时，研究过程是严格的、有完整结构的。活动在有些时候可以有所重复，有时两项或两项以上的活动可能同时出现在过程中。例如，在定性研究中，形成假设可能贯穿于从收集资料开始的全过程。在许多研究中，初步的分析已经开始，但资料收集仍在进行。尽管如此，把活动排成序列还是有益的。

图 1.4 是活动序列模型的流程图，它描绘了不同研究活动的大致情况。上框表示的是一般活动，为了显示研究过程的灵活性和研究类型的多样性，活动之中是有重复的。比如，在进行实验研究之前，假设已经形成，要收集什么数据也早已一清二楚；但定性研究就可能在进入研究过程之后还在形成假设、增加要收集的资料。这种特征在图中用重叠的方框表示。

下面的(虚线)框不是表示活动，但也是研究的组成部分。箭头表示的是活动与现有的知识，与相关的理论，与被扩展、被修改，与新的理论和知识的关系。相关的理论是指部分，而没有必要是全部，是与研究的问题有关系的部分知识，被扩展、被修改，以及新知识，如果即将从研究课题中产生的话，就成了现存知识的一部分，但新产生的知识不被看做是理论。所有的一般活动都来自现有的知识，但在图中，我们只反映了知识的主要影响与要研究的问题的联系。

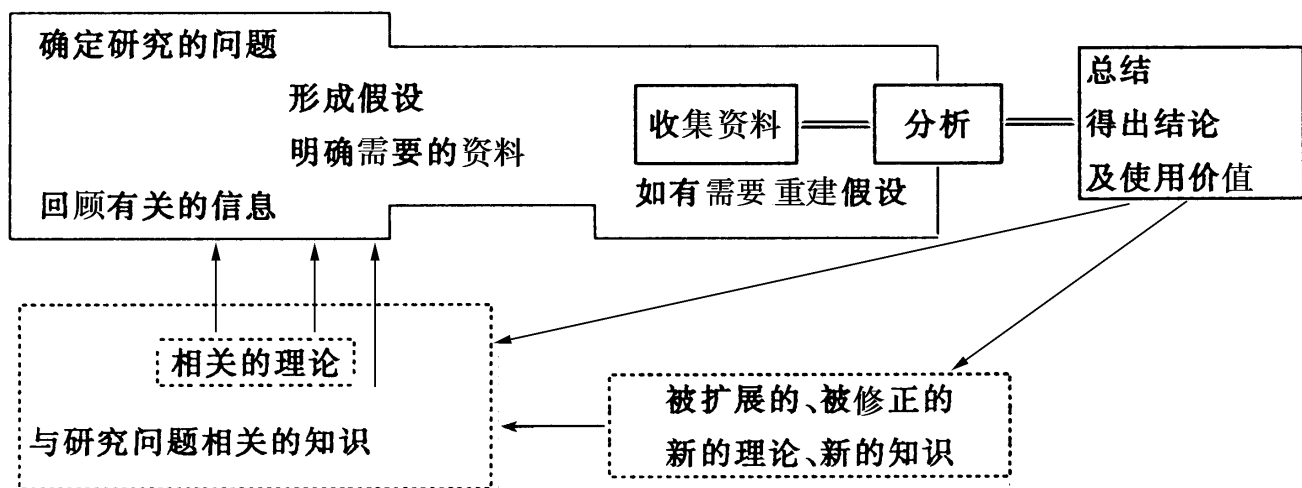


图 1.4 进行一项研究的一般活动的序列模型
以及这一活动与现有知识的关系

研究过程可以被看做是一个连续性的活动，其间，时有重叠和变化。

这样，活动的每一步都需得到详细的说明。但因为这里的讨论是导论性的，只准备讲述一个概貌。在以后的章节中将作详细介绍并有实例说明。

确定研究问题

这是研究过程的初始活动（至少应该是这样），有时是最困难的。问题必须能体现出特定的意义。此时假设就产生了。在确定需要的资料前，应首先确定变量，研究情境中这些变量的使用条件也应作充分的界定。通过阅读文献获得与所研究的问题或使用的方法有关的信息。也就是说了解别人已搞了什么研究或作出了什么有价值的发现。查阅文献是项重要的工作，本书的第3章将讨论这个问题。

收集资料

收集资料之前，一切测量的手段都必须明确；如果论述中要包含数据，研究者应组织好对数据的说明。如果正在使用实地研究的方法，研究者是通过访谈、观察等方法收集资料。如果正在进行一项实验，实验的处理或实施是在收集数据之前或之中。从本质上说，实验就是进行测量。在调查研究中，测量的手段是成就测验和问卷之类的方法。如果研究的手段得到了改进，必须在大规模收集资料之前试行。然后，必须对资料进行集中、编码，并为分析做准备。

分析

进行了分析，结果也就出来了。数据阐述也同时被组织和综合起来。经过对资料的总结、加工并进行抽象概括，就得到了论述和验证假设所需要的信息。如果有统计分析，也是在这一阶段完成。

总结结果和得出结论

分析资料得到结论后，研究者必须明确他准备提供什么信息。结果应该总结出来放在一起。分析必须有解释，结论必须明白提出。无疑要准备研究报告，这一工作常常需要反复。这项任务的重要性无论怎样强调都不过分。因为它是将研究结果归结为新的知识和理论，以及和已有理论相结合的途径。

从某种意义上说，最终的得出结论和启发性的工作是最重要的，因为从根本上说这是研究的目的。但是要得出结论需要解释、综合、洞察，这些工作是有一定难度的。说明得出结论之困难的一个明显的事实，就是许多论文的结尾与最初草稿相比较，读起来都像重新写过。

尽管我们不能保证得出的结论是明智的，但这里有一些建议，按照这些建议可以改善这一工作，提高研究的价值。这些建议是：

1. 弄清楚研究进行的领域，从实质上知道研究文献中有什么，不能只是一般性的了解。

2. 明确研究的外在效度（内在效度应先确定好）。讨论结果的普遍性。善于根据逻辑来扩展外在效度。为使结论普遍化，应考虑到所有可能的合理例证。

3. 集中注意结果的意义而不是结果本身。

4. 如果包含的理论已经存在或者是新的理论已经产生，要清楚地表明结果与理论之间的一致性。

5. 寻求研究结果与其他相似研究结果之间的一致性，并且解释这些相似结果在这个领域内扩展研究的外在效度。

6. 寻求研究结果与其他相似研究结果之间的不一致，弄清可能导致不一致的原因。

7. 使达到结果的推理环节明晰，不要假定读者能读出作者的想法。

8. 如果可能，注明本研究未解决的问题以便将来可以进一步研究，并提供下一步合理地逻辑延伸的知识。

总结结果，得出结论应该彻底地反映出学术成就，当然，我们希望避免没有根据的结论，但我们应尽量从研究结果中获取最大量的信息。让一个或更多的在这个领域中熟悉这个研究的同事对他（她）得出的结论进行再检查是一项非常有益的程序。

初搞研究的人可能试图从收集资料之类的工作开始当中插入研究（寻找研究问题的资料收集者），有时研究者在一些数据的基础上提出假设，并试图从假设中抽出问题。如此肢解整个研究的连续性会导致混乱和低效。虽然研究过程是灵活的，根据具体的研究需要，研究活动可能重叠或重复，但是，坚持图1.4所表示的研究程序可以提高研究的效率。

总结

教育研究的分类系统从范围和种类上描述了教育研究的一些思想。基础研究和应用研究的概念已出现了很长时间，它用于区分研究的目的；定性研究和定量研究是对研究进行分类的另一种方法，但它的功能并不是把这两种研究相割裂而是强调它们是相联系的。一般方法论意义上的研究种类都适用于教育研究。研究目的决定研究类型。教育研究可以放到科学方法的背景中去；但不管是教育研究还是科学研究方法，都不要看成是僵硬的、不变的、死板的程序设

定。

研究过程的一般活动已作了界定和说明，图 1.4 则表示了它们的内在联系。教育研究涉及许多活动，有些活动较为简单，有些则比较复杂。事实上从简单到复杂在整体上是连续的。研究可以在许多领域进行，如初步提到的课程、学习、教育管理等；研究可以在不同层次进行，从学校教师的行动研究到大规模的大学或其他机构的研究项目都是。对研究过程和研究结果应有足够的表述，即使是最有概括性的概念，置于现实中时，我们也会发现它们是由许多特定的东西构成的。这些特定的东西就是研究方法（活动、过程、逻辑、推理等）。这正是本书论述的主要内容。

第二章

研究问题的确定

1 研究问题的选择

明确研究问题，是研究过程中为了获得理想答案的重要一环。研究一开始就要求提出合适的问题或对所研究的现象有明确的认识，这可能是研究过程中困难的步骤之一，因为这时工作几乎还没有组织起来。充分地确定问题是研究过程展开的必要条件。因研究的类型不同，对问题的确认程度也有所不同。例如，实验研究通常是一种非常具体的研究，并伴随着假设，而定性的研究则更多是概括性的陈述，其假设是通过整个研究总结得出的。

对一个研究问题的确认不仅仅是就所涉及的领域提供一个具体的陈述或问题。确认的第一步是选择好研究的题目，然后是对题目的具体陈述。如果假设得到验证，这些假设及问题的陈述涉及到有关变量和条件的专门术语。教育与所有专业一样，也有专门用语（其中至少部分有时被称为行话）。研究问题范畴中术语的含义保持固定不变是必要的。本章中定义了大量的基本术语，读者理解这些术语十分必要，因为它们是教育研究基本用语的组成部分。

选择好合适的问题等于提出了好问题，也就是说，在教育领域中，问题是至关重要的。那么，如何着手选择一个研究问题呢？

选择研究问题有不同的途径。出自研究者的内部动机，不是所有的问题都是合适的。例如为完成毕业学位的论文，研究生选择研究问题往往注意以下的途径：他们密切结合某个导师或多个导师的研究方向，确定一个与之相关的问题，或确定一个研究方向分支方面的问题。这个假定的前提是出自彼此利益的考虑，学生和导师共同开展研究工作。第二种途径是通过与其他学生展开讨论，一些研究生深入到研究项目中，并对问题的选择有自己的见解。随着时间的推移，研究兴趣的增强，以及当某人开始考虑从事研究时，研究问题的选择随之而迫近。当然，研究问题不是在朝夕间便确定下来的。它是一个需要研究者对可能的问题进行全面的了解，并通过不断的思考，以及与他人探讨方能确定的过程。问题需要不断地修改和再确定。有时问题会扩展，有时被精练。一

个与之相关但方向不同的问题也有可能出现。总之，所有这些都针对问题本身。

如上所述，尽管研究问题的选择过程似乎有些“宽泛”，但过程中仍有某些特定因素。研究问题必须对研究者有所裨益，至少有些方面要涉及教育领域。应该确信研究的问题在教育领域中能占一席之地。特别是如果研究的题目涉及一个论点时，开创性也应该被考虑。但完全开创性的研究是很少的，更大的可能是已经完成课题的延伸。研究中重复研究的价值依具体的领域和研究条件而定。

考虑的另一因素是研究的问题应具有实践的或理论的意义。有些问题如小学生中穿布鞋或皮鞋的比例和鞋子的选择与成绩的关系等，在方法上是可以被研究的，但不论其结果如何，这些问题没有实际意义。研究的问题应通过有意义的方式增加现有的知识或有助于教育的发展。

教育中不是所有的问题都值得研究。有些问题就其性质而论是哲学上的，可以讨论但不可能研究。例如有这样一个问题：“高中历史是开一门课还是两门课？”如果要求的是两门课，那么学生将学更多的历史知识，但开两门课是否有必要，对这个问题的回答依不同价值判断而定。如果没有一定的条件为前提，这个问题就不值得研究。

即使是值得研究的问题，有时做起来也不可能实现。研究所需的资料有时过剩也有时很难得到。此外，像实验设备和资金等必需的资源也可能阻碍研究。诸如此类的很多条件使得某些具体问题的研究受阻。

可以说，大学的教育机构在有些具体的研究方面常常有所“突破”。一些机构可能对课程研究有兴趣，例如数学教学的研究。其他则可能对如教师行为，高等教育发展，或者体育训练等感兴趣。这些突破可作为研究问题的重要来源。但除此之外还有其他来源。研究者的职业经验和状况能使之联想到的问题，特别是在应用研究中。目前的教育期刊以及与教育相关的社会和政治刊物，也可启发出许多研究问题。研究文献中也包含一些引导人们继续从事某专题研究的暗示，例如，应用于教育的组织理论，需要有检验该理论的研究。总而言之，研究问题的选择，要依据与问题相关的各种信息和研究者的经验而定。

研究问题的选择涉及到阅读、讨论和构思等活动。它是一个与问题相关的因素被考虑时不断逼近问题的过程。

2 研究问题的陈述

一个研究问题的选择并不意味着这个问题有了恰当的陈述。通常的情况是，一个问题需要反复几次才能成为有效研究的恰当形式。在研究努力的初始阶段，问题可先粗略陈述一下，然后通过查阅文献系统地限制，这种做法比一开始陈述得很狭窄片面然后再来扩展的做法要好。

陈述问题可采用叙述或描述的形式，也可采用问题的形式。很多研究工作者，也许是大多数，偏向于采用问题的形式。其实两种形式都是可行的。陈述采用问题的形式对焦点问题的研究较好，尤其是当大问题中包括有小问题时效果更好。问题陈述最重要的特点是它必须为研究提供足够的焦点和方向。

基于这一点，指明某些满意和非满意的问题陈述也许是有益的。像“小学课程”这样太宽泛的陈述是不能充作问题的陈述的，它其实不含任何问题。“研究 A 城市小学 4 年级阅读课程对 4 年级学生阅读成绩的影响”可以算是一个满意的陈述。或者采用问题的陈述形式：“A 城市小学 4 年级的阅读课程对 4 年级学生的阅读成绩的影响是什么？”下面几个例子是原陈述和经过改进后更便于操作的再陈述，包括问题的陈述形式。

例 1：原陈述：小学生的创造能力。

再陈述：发散性思维的得分与所选择的 4、5、6 年级学生的特点间关系的研究。

问题的形式：4、5、6 年级学生发散性思维得分与一般智力测量分数的关系是什么？与阅读成绩测量分数的关系是什么？与动作灵巧测量的关系是什么？

例 2：原陈述：学习成绩与教学技术。

再陈述：关于 3 种教学技术对高中低年级学生的学科成绩影响的研究。

问题的形式：3 种不同教学技术是否对高中低年级学生的学科成绩产生不同影响？

例 3：原陈述：高中生的辍学。

再陈述：对正常的学习能力丧失的学生的学校环境进行实地研究，以确定有关潜在的辍学因素。

问题的形式：学校环境中的什么因素与正常的学习能力丧失的高中学生的潜在辍学有关？

例 4：原陈述：A 大学的历史。

再陈述：在 1955 ~ 1972 年期间，联邦政府对高等教育的资助对 A 大学扩展科学和教学课程影响的研究。

问题的形式：在 1955 年 ~ 1972 年期间，联邦政府的资助对 A 大学扩展学科和数学课程有何影响？

例 5：原陈述：高中指导顾问的职责。

再陈述：对 B 城市高中指导顾问的活动的调查。

问题的形式：下面四个问题说明了什么是更具体的问题：指导顾问的非指导活动占他工作时间的比例是多少？学生所感受到的指导顾问的主要优点是什么？学生所感受到的指导顾问的主要弱点是什么？作为向学生提供最有效的升学指导的指导顾问，他们自己认为应该从事什么活动？

问题陈述得好，可以为研究者提供从事该研究计划的方向。陈述应该指明总体的中心议题和问题的前后背景。陈述应辨认清楚研究计划中的关键因素，例如，在前面第一个再陈述中，“关系”一词暗示着某些程度、3 种年级水平的限制和研究中人数的限制；“发散性思维分数”无疑比“创造能力”更具体，“创造能力”是在原陈述中使用到的。发散性思维要求有一个操作性定义，如定义为在具体的发散性思维测验中的得分。

应该强调的是，再次陈述问题时，很多术语必须界定。在第 1 个例子中，所选择的因素要求确认并定义，第 2 个例子中的 3 种教学技术要求根据特定情况予以定义。这些定义伴随问题的陈述，但问题通常不包括在陈述之中。因为这样使问题的陈述变得冗长。假定问题的陈述中有很多的定義，就不会造成对所调查的东西含糊不清。

对研究问题的陈述是否采用问题的形式在很大程度上取决于个人的爱好。如果问题的形式有助于研究，那就应该采用。实际上，陈述问题的形式相对来说不是重要的，重要的是陈述要精确和无可置疑，以免所研究的东西发生混乱。

结合例子来考虑往往有助于理解研究问题陈述的本质。下面一系列问题的陈述是在研究方法课上学生设计出来的。

1. 在上海所选定的学校中调查阅读方案的组成和实施。
2. 用脑过多对小孩的听力和自动化操作有何影响？
3. 在一所民办大学 2 年级，进行科学教学的实地研究。
4. 在小学，表扬的类型与学生对学校的态度有何关系？
5. 对 A 初中来自 9 所小学的学生学习成绩的调查。
6. 年龄、材料类型和可利用的信息量对掌握概念的成绩有何影响？

- 7. 关于 A 地区教师对能力测试方案的理解和态度的调查。
- 8. 1980 ~ 1990 期间，北京市给小学和初中增拨教育经费的历史研究。

研究问题的陈述应该简洁明了，并确定研究活动的关键因素（变量）。

常量、变量和操作性定义

问题的描述本身通常仅提供研究的大致方向，它不包括所有具体的信息。在我们如何表述有关研究问题和整体研究方面的具体信息时，有一些基本的术语相当重要。

常量是一个研究中所有个体都具有的相同的特征或条件。与此相反，变量是在一个研究中不同的个体体现出不同价值的特征和条件。如果一个研究是想比较两种不同教学方法对 5 年级学生的学科成绩的效果，那么年级水平则是一个常量，因为每个被研究的个体都是 5 年级的学生，5 年级这一特征对每个个体都是相同的，它是这个研究中不变的条件。两种不同教学方法实施后，将要对那些 5 年级的学生进行学科成绩测量。在测量中不可能每个学生的分数都相同，那么在这次学科成绩测量中，分数就是一个变量，因为不同的个体将会有不同的分数，至少不是所有个体都是同一分数。我们也可以说学科成绩是一个变量，但我们特别强调的是分数在学科成绩测量中是个变量。

常量是一个研究中所有个体都具有的特征或条件；变量是不同的个体具有不同价值或条件的特征。

在这个例子中还有一个变量——教学方法。与学科成绩测量的分数相比，教学方法无疑可以按多种价值标准来衡量，它是一个分类变量，包括两种教学方法。因而，教学方法是一个条件变量；也就是说，它决定这个研究的确定条件。这个研究中，所有 5 年级学生被分别分配到这种或那种教学方法中去，因此，将根据所参与的教学方法的不同对这些学生进行分类。为此，我们对不同的变量赋予不同的名称和分类。在现存的文献中有许多分类体系，其中有很多相互重叠和混淆的情况。我们采用的名称是描述性的，它们描述了变量在研究中所扮演的角色。以下描述的变量决不可能概括现存的所有分类体系和名称，但它们最适用于有关教育的研究。

自变量和因变量 一些重要的、并被广泛用于描述的变量，是自变量和因

变量。自变量和因变量源于数学。一般认为因变量的价值取决于自变量。因变量是我们可以测量的变量，如学科成绩。自变量仅仅是一个分类变量，它对研究的个体进行分类。在学科成绩的例子中，教学方法就是个自变量。研究中的每一个5年级学生都将接受两种教学方法中一种方法的教学，两种不同的教学方法构成这一变量的两种水平或两种类型（2是变量的最低水平数量）。学科成绩的分数——因变量，被假设为受教学方法影响。事实上，研究的目的在于弄清楚教学方法是否对学科成绩有影响。研究者试图从自变量出发解释因变量。

以下面的例子来进一步说明自变量和因变量的使用。在关于3种不同教学方法对学习初等代数的成绩影响的研究中，在同一所学校，由同一位教师用3种不同的教学方法分别教初一年级代数的内容，在研究中包括男生和女生。这里年级、学校和教师就是常量（这里的假设是除了教学方法，教师能确保教学效果相同）。研究中的自变量是教学方法和性别。教学方法有3种水平，可定为方法A、B、C；学生的性别当然有两种水平。经过一段教学后测得的代数成绩就是因变量。

再来看一看其他自变量和因变量的例子：

- 例 1：不同教育水平的师生课堂交流研究。
自变量：不同的教育水平，4 个类别：小学、初中、高中、大学。
因变量：测量出的师生课堂交流各项指标的分数。
- 例 2：学校地理分布影响初二年级学生对学校态度的研究。
自变量：学校地理分布，分 3 类：城市、城镇、乡村。
因变量：对学校态度的各项指标分数。
- 例 3：不同材料对概念理解的影响研究。
自变量：材料的种类：图片、语音。
因变量：理解概念所需的时间。
- 例 4：男、女教师职业态度研究。
自变量：教师性别：男、女。
因变量：职业态度各类指标分数。

自变量和因变量是教育研究中广泛采用的变量描述术语。自变量可影响因变量。在某种意义上，因变量取决于自变量。

变量的其他可能种类

在任何研究中，影响因素都以变量的形式表现出来。为了方便起见，我们给它们一个描述的名称，这些名称经常在文献中使用。

有机变量用来表示研究中个体的先天特征，它不是一个随机分配到个体中的变量。性别和智力都属于有机变量。

中间变量是可推断其存在的变量，但它不能够被控制或测量。如果它有影响的话，其影响只能从研究的自变量与因变量的关系之中推断出来。在代数教学的例子中，学生的学习风格就是一个中间变量。中间变量也可叫其他的名称，如无关变量、中性变量或复合变量，它们的存在会对其他变量的效果的解释产生不利影响。

控制变量是一个不同于起主要作用的自变量，它的效果可由研究者控制。在代数教学的例子中，如果男女性别对因变量影响的差异被确定，那么性别既是有机变量又是可控变量。

其他变量名称与上述名称部分重叠，减缓变量就是一个例子。假设某一研究是想比较3种关于4年级阅读成绩的阅读方案的效果，那么先前有较高阅读水平的学生与先前水平一般的学生存在相当大的差异，一种方案比另外两种方案更有效可能是因为那些学生先前具有较高的阅读水平。在这个研究中先前的阅读水平就是一个减缓变量，也就是说，方案的效果因学生先前有较高阅读水平而减小。如果减缓变量是不可控制的，那它实质上是一个中间变量；如果它是可控的，那它的效果能够确定，它就是一个控制变量。可见，对变量的描述并不是相互排斥的。有机变量也可成为控制变量。如果代数教学的例子扩展到包括两个或多个学校，且各学校的差异能够确定，那么学校这一变量就成了一个控制变量（但学校变量不可能成为有机变量）。

任何一个特定的研究都不会只局限于一个自变量、有机变量、控制变量、中间变量或因变量。假定在5年级学科成绩的例子中包括有5所学校，学生按以前的成绩进行高、中、低分组，男女性别也包括进去，且他们之间的差异被确定。一个自变量是教学方法，另一个自变量是材料类型，假如有两类材料被引进到研究中，进行更为广泛的成绩研究以观察同一组中的阅读成绩和拼写成绩。这些成绩分数也是因变量，这不是因为存在有两种或更多种的测量成绩，而是因为这些分数因成绩种类的不同而改变并且受自变量的影响，这时可能存在的中间变量就包括有学生的学习风格和教师的教学风格。这是一个比较复杂的例子，所有变量见图2.1。注意这里的控制变量是自变量。不能一概断定有机变量必然是控制变量，只有当它们的效果被确定，即能够被控制时，才成为

控制变量。学生的学习风格也许是一个有机变量，但它也可能是一个中间变量。

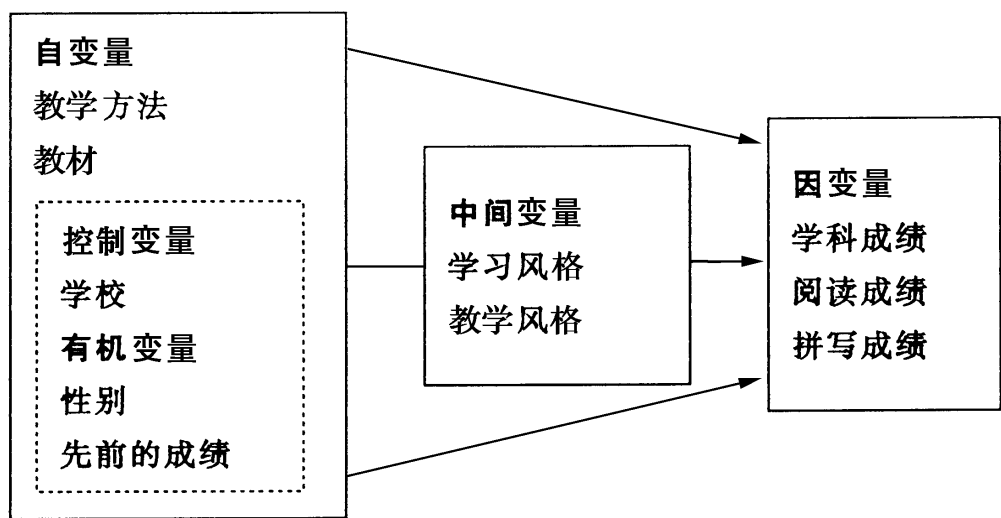


图 2.1 5 年级成绩研究中的变量类型

当把教育中可能的变量都考虑到时，就会发现所有的测量并不是相同的。在学科成绩的例子中，教学方法这个变量仅是个范畴变量（即关于教学方法的不同范畴的变量），不同的教学方法可用不同名称来命名。另一方面，从学科成绩测量得到的分数这个因变量来看，测量的不仅是范畴，这里还有对分数的排序和量化。

操作性定义 在所有研究中，所研究的变量和条件必须有可供操作的定义。教育者经常遇到一些难于表达清楚的变量。比如，如果学校的医生想了解一年级学生的体重，他就可以用一般的体重计来测量。相应地，我们如果想测出学生的学习能力或者阅读能力，并使之量化，就必须采用一些工具来实现这一目的。也许我们还要建立一系列为人们认可的有关学习能力方面的定义。为了成功的测量，我们的定义中还必须包括在研究现象或变量时将要用到的那些加工或操作过程，这样的定义称为操作性定义。本质上，变量的操作性定义是关于如何或用什么办法测量变量的描述。

操作性定义不仅是为满足教育研究所必须的，而且也是确定如何测量变量所必须的。我们知道，学龄儿童也包括所有其他人的创造能力是一个变量，但除非有一个操作性定义，否则创造能力无法测量。一个变量可能不只有一个操作定义，如创造能力，这取决于操作性定义的另一种价值。操作性定义能使研究者确认同样题目和问题研究间的相似点和差异。一些变量用不同的方法测量，可能导致明显矛盾的研究结果。操作性定义可用于重复的研究，并且能提高研究结果的可解释性。操作定义可以通过确定概括的范围，来帮助建立研究结果的外在效度（或缺陷）。

以下是操作性定义的几个例子：

- 1. 学习能力：在斯坦福—比奈（Stanford - Binet）智力量表 LM 表格中的分数。
- 2. 科研能力：衣阿华（IOWA）基本技能测试科学测量的分数。
- 3. 发散思维：砖的用途（Brick Uses）测试的分数。
- 4. 概念理解能力：要求准确理解 5 个概念的时间。

不仅对研究中的变量要下操作性定义，研究的有关条件也是如此。例如，如果要调查高中学生，那么什么叫高中学生？有必要先对此有个界定。高中学生的操作性定义也许可以这样下：即现在普通高中注册的还未毕业的学生。这里的高中学生的操作性定义是从确定的可观察或证实到的特征来下的。

操作性定义是一种规定，它使被确定的需要定义的变量和条件的操作或特征具体化。

假设与问题的陈述

在对变量、常量和操作性定义的概念阐述之后，我们再继续通过假设的运用详尽阐述研究的具体细节。

如果陈述所有的信息就会使其变得臃肿不便操作。要想内容更具体，方向更明确，就要运用假设这一概念。假设可直接从问题的陈述中产生，可从研究的文献中或从像实地研究的事例中产生，也可（至少部分可以）从资料的搜集和分析中产生。实地研究也涉及到预言的问题，尽管这些问题不是假设，但它们是对所从事的研究所要寻找的东西的陈述。预言的问题不是假设替代品，但在实地研究中它们是对假设的补充和提供研究的方向。

假设是一种推测或对问题答案及情况状态的一种猜测。一般地说，假设具有理论的某些特征，它通常被认为是关于某一现象的一大堆概括。这样，一个理论就可能包括几个假设。判定假设是否真实可靠，就叫检验假设，检验的结果不是支持假设就是否定假设。

伯格（Borg）和高尔（Gall）认为假设应具备以下 4 条标准：

- 1. 说明两个或两个以上变量间的期望关系。
- 2. 研究者应有该假设是否值得检验的明确的理由，这一理由有理论的或

事实的依据。

3. 假设应是可检验的。
4. 假设应尽可能简洁明了。

很多假设存在这样一个弱点，即太广太泛因而不能准确描述研究中的具体问题，请看下面这个假设：“聪明的学生对学校的态度好”。“聪明”、“好”和“态度”等词就是典型的笼统粗略而不针对具体研究问题的概念。聪明与好的态度之间隐含着某种模糊的关系，因而对研究没有什么指导作用。把陈述变换成可接受的假设，这段话给人的印象就可能发生变化：

IQ 测量分数排在前 25% 的 9 至 11 岁的学生在 X - Y - Z 态度测试中的平均分比那些 IQ 测量分数在同类中排在后 75% 的学生要高。

这个陈述详细具体，表明了一个期望的关系（IQ 得分排在前 25% 的学生比排在后 75% 的学生在态度上的得分要高）。假定这个测量能够做到，则假设是可以检验的。

另一种表述是：

在 9 ~ 11 岁的学生中，IQ 测量分数与态度测量得分之间存在正相关。

这个假设的陈述比初始的那种陈述要短。它包括一个期望的关系并且是可检验的。

这两个例子中，假设包含了有关变量的操作性定义——学习能力（聪明程度）和对学校的态度。这些变量用在具体测试和具体调查中的分数来定义。如果操作性定义使假设的表述过于笨重冗长，可分开来表述。但是，变量应清楚地确定以便使期望的关系能够清楚明确。也就是说，它肯定一组的平均分将比另一组高。假设是可检验的，这就是说，有一系列步骤用来分析那些将支持假设或排斥假设的资料上面给出的两种表述将有不同的方法来检验。假设预言期望的方向，可以相信，这不是毫无根据的猜测。如果有理由证明假设值得检验，那么总的说来，符合上述 4 条标准的假设就是一个很好的陈述。

假设是对问题的结果、两个或多个变量之间的关系或某些现象的性质的推测或提议。

假设的种类和形式

克林格 (kerlinger, 1989) 指出, 从广义上讲, 研究者使用的假设有两种: 实质性假设和统计假设。实质性假设也叫做研究假设, 比如 “在小学随着严厉惩罚的方法的增加, 学生的成绩将下降。” 在科学教育研究中, 则可以有这样的假设: “做实验的教学比起局限于讲课、讨论以及解决理论问题的教学方法来, 更能促进学生理解科学的过程”。

统计假设属于统计学的术语, 在技术性方面, 它是推断性统计, 是关于总体测量研究中一个或多个参数的描述。统计假设常采用定量的术语, 例如, “运用方法 A 和方法 B 教得的 3 年级学生的平均阅读成绩无明显差异”。

上述统计假设是用零形式 (null form) 陈述的假设的一个例子, 即假设平均分数没有差异。零假设有时被视为无差异或无关系的假设。从技术上说, 如果统计假设用推断统计检验, 接受检验的假设便是零假设。

对任何统计假设, 都有一个选择性假设, 它表示存在的可能结果。对阅读成绩的例子用选择假设表示就是: “运用方法 A 教得的 3 年级学生的平均阅读成绩与运用方法 B 教得的不相等。” 应说明的是, 在这个例子中, 无论两种方法的结果相等还是不相等, 零假设和选择假设包括了两种方法可能出现的所有结果。

如果我们最初假设 “接受方法 A 教学的 3 年级学生比接受方法 B 教学的学生的阅读成绩要高”, 为概括两种方法可能出现的所有结果, 我们可假设:

“接受方法 A 教学的 3 年级学生比接受方法 B 教学的学生阅读成绩低。”
和 “接受方法 A 和方法 B 教学的 3 年级学生阅读成绩相等。”

这里是两个选择性假设, 而后一个是零假设。我们可以把其中两个陈述合并为一个选择假设:

“接受方法 A 教学的 3 年级学生的阅读能力低于或等于接受方法 B 教学的学生的阅读能力。”

在此零假设被包含到一个选择性假设中, 这样所有可能的结果都被覆盖了。

假设还可区分为定向假设和非定向假设, 前者示意结果有一定趋向, 因此, 没有示意结果趋向的是后者, 正式的假设通常是定向的, 因为它们表示结

果的预期定向。

假定有位教师在试用新的技术教 3 年级学生拼写，他假设：采用新技术教 3 年级学生，他们的拼写成绩优于用传统方法教的学生，更具体地说，假设应是：

采用新方法教 3 年级学生，他们的 ABC 拼写测验的平均分超过用传统方法教的学生。

选择假设是

接受新方法教学的 3 年级学生，ABC 拼写测验的成绩低于或等于接受传统教学方法的学生。

原假设是一个定向假设，因为它预示了结果的方向：接受新方法教学的学生成绩更好。选择假设包含了零假设。

请思考以下有关假设和选择假设的附加例子：

1. 能力强的学生数学成绩等于能力一般的学生，或在能力一般和能力强的学生间，数学成绩没有区别。

1a. 能力强学生的数学成绩不等于能力一般的学生成绩。

2. 一年级女生的阅读水平与男生的阅读水平相同。

2a. 一年级女生的阅读水平不等于男生的阅读水平。

3. 采用归纳提问教学的成绩超过采用非归纳提问教学的成绩。

3a. 采用归纳提问教学的成绩低于或等于采用非归纳提问教学的成绩。

4. 初中阶段学生的学习态度分数与社会协调指标分数存在负相关。

4a. 初中阶段学生的学习态度分数与社会协调指标分数无关。

5. 参加方案 A 练习的 18 ~ 20 岁男生取得的 XY 运动行为测试分数，会高于参加方案 B 练习的 18 ~ 20 岁的男生分数。

5a. 参加方案 A 练习的 18 ~ 20 岁男生取得的 XY 运动行为测试分数，低于或等于参加方案 B 练习的 18 ~ 20 岁的男生分数。

其中 1、2 两个假设是零假设，且这些假设和它们的选择假设是非定向假设。3、4、5 这 3 个假设是定向假设，且每个假设的零假设都包含在选择假设之中。

假设的陈述是采用定向的形式还是采用非定向的形式？采用不同形式会产

生差别吗？这应由期望的结果来定。如果研究领域的文献表明我们能期望一个差异或一个有方向的结果，那么应采用定向假设；如果研究文献不能表明令人心服的证据有方向性，或者正在进行的是一个探讨性的研究，则应采用非定向假设。因为在推断统计学中强调零假设，教育研究和一般的行为科学研究，可能多采用非定向假设而少用定向假设。

统计假设用于资料的分析；实质性的假设或研究假设提示了结果的方向。假设可采用定向形式和非定向形式。零假设是没有差异或联系的假设。

与定性研究相比，定量研究更多地应用有关假设的这些陈述。在定量研究中，假设的确定一般在研究之初，资料收集之前，和问题修改变动不大的情况之下。另一方面，定性研究的假设，大多产生于研究进行的过程当中。在研究之初，或许没有尝试性的或一般性的假设。但当资料被收集和分析时，从事定性研究者希望增加、减少、修改和精练假设。定性研究的假设纯属推测，但不论它们是定向的还是零假设，关系不大。他们的表述适合研究中的现象背景。

预言的问题在前面提到过，它通常与实地研究有关，是对假设的补充，是对处在进行状态的研究的陈述。考虑一下前面所给的研究问题的陈述：“对正常的和学习能力丧失的学生的学校环境进行实地研究，以确定有关的潜在的辍学因素。”与这个陈述相关的预言的问题是：

1. 在教学中，正常的学生与学习能力丧失的学生之间的相互影响；
2. 教师的角色促进学生的学习；
3. 学生的社会系统；
4. 学生在学术科目中获得成功的机会。

要指出的是，预言的问题不是具体的期望的结果。实际上，它帮助研究者考察正在进行的研究确定因素。

与问题陈述有关的假设举例

研究问题的陈述可以有一个或多个（通常有多个）假设与之相联系。假设是在研究问题的陈述中形成的，并且应该直接来自于它。下面的每个例子提供了一个关于研究问题的陈述、假设和操作性定义或有关变量和条件的操作性定义的范例。

例 1

问题的陈述：

对某市高中学科的等级及等级关系的调查。

假设：

- 1. 化学和物理学科的平均等级高于生物和地理学科的平均等级。
- 2. 历史和其他社会学科的平均等级高于生物和物理学科的平均等级。
- 3. （高一年级的代数及更高年级的）高等数学的平均等级高于初等代数和应用数学的平均等级。
- 4. 英语课的得分等级与其他外国语课的得分等级之间存在正相关。
- 5. 代数的得分等级与化学的得分等级之间有正相关。
- 6. 随着高中规模的扩大，学科等级之间的相关程度也相应提高。

操作性定义：

学科：数学，英语，语文，历史，外语。

高中：某市所有获得高中学校资格的高一、高二、高三年级。

等级：可按照字母评分系统 A、B、C……来评定的类别，它可转换成数字型的分数。

等级关系：每一范畴内课程间等级的相关情况。

高中规模：高一至高三年级注册的总人数；范畴有：少于 200 个学生的，200 ~ 499，500 ~ 799，800 ~ 1 099，1 100 以上。

自变量：

学科

某些领域的具体课程

高中的规模

可能的中间变量：

学校类型

学生性别（有机的）

学校的地理位置

因变量：

等级

等级关系/ 按范畴来分的等级比例

可能的控制变量：

学校规模

例 2

问题的陈述：

在学校 C 中，研究两种阅读方案（A 和 B）对 3 年级学生阅读成绩的影响。

假设：

- 1. 对不同阅读成绩的学生，用方案 A 教学阅读成绩的平均增加分数与用

方案 B 教学的平均增加分数之间没有差异。

- 2. 按照原有的阅读成绩，对那些得分在后 30% 的学生用方案 A 来教，其平均增分大于用方案 B 来教的平均增分。
- 3. 按照原有的阅读成绩，对那些得分在前 30% 的学生用方案 A 来教，其平均增分大于用方案 B 来教的平均增分。
- 4. 根据原有的阅读成绩，对那些得分在中间 40% 的学生采用方案 A 与采用方案 B 来教，其平均增分没有差别。

操作性定义：

研究中包括的个体：C 学校所有 3 年级的学生。

方案 A：从 Y 出版商那里买来的一套材料及活动建议。

方案 B：从 X 出版商那里买来的一套材料及活动建议。

自变量：

阅读材料：A 和 B

因变量：

阅读成绩的分数增加，例如：进行两次标准阅读检验，一次在学习材料前，一次在学习材料后，前后两次的分数差即为增加的分数

可能的中间变量：

教师

教学风格（有机的）

学习风格（有机的）

学生的学习能力（有机的）

可能的控制变量：

原有的阅读成绩（有机的）

性别（有机的）

例 3

问题的陈述：

在 A 学校进行的一项关于实验教学在初中学生的理科课程教学中的作用
的实地研究。

预言的问题：

- 1. 实验中学生之间的相互影响。
- 2. 实验中学生与教师之间的相互影响。
- 3. 学生和教师的实验准备。
- 4. 实验与科学教育的其他方面的关系。

假设：

- 1. 实验的功用在于使学生参与合作活动。
- 2. 学术能力强的学生将垄断实验的操作。

操作性定义：

研究的参与者：A 学校中所有的学生。

理科课程教学：在物理科学、化学科学或生物科学中重视操作的所有课程。

实验工作：在实验室的活动，这些活动中学生直接（单个人或与其他人一起）进行实验，操作科学仪器，解剖生物样本等等。

变量：

实验的功用

在实验中发生的相互影响的类型

实验的程度

实验的类型，例如解剖动物，物理实验等与其他学科教学有关的实验的时间安排

例 4

问题的陈述：

对 C 城某地区教师对学校管理人员政策的看法的调查。

假设：

- 1. 对给在职人员予以补贴的政策持赞成态度的教师超过 75%。
- 2. 对教师调动政策持赞成态度的教师至少有 50%。
- 3. 小学教师与中学教师对病休政策持赞成态度的比例没有差异。
- 4. 对教师调动政策持赞成态度的小学教师，在郊区系统与城区系统之间存在差异。

操作性定义：

学校管理人员政策：记载在校委会备忘录或条例中的对学校工作程序（或条件）和活动的说明。

C 城某地区：C 城市及其他的郊区。

自变量：

教师类型：小学的、中学的（有机的）

教学或学校系统的位置——城市 and 不同郊区

可能的中间变量：

教师年龄

教师的婚姻状况（有机的）

教师的性别（有机的）

具体学校的规模

例 5

问题的陈述：

材料的种类、学生的年龄、性别和问题的复杂程度对概念理解成绩影响的研究。

因变量：

教师对政策的赞成比例

可能的控制变量：

教学的地点

假设：

- 1. 用图片和用语言材料对概念理解成绩的影响没有区别。
- 2. 20 ~ 25 岁和 26 ~ 30 岁两年龄段间成绩没有差异。
- 3. 男性与女性的成绩没有差别。
- 4. 对具有 2 个和 3 个维度的概念理解问题的把握同等困难。

操作性定义：

因为概念理解任务很具体，进行这项研究在方法上需要具体细节描述。

材料种类： 图片，表达概念的 3 × 3 图片。

语言，印有图片内容相对应的文字叙述的 3 × 3 卡片。

相关维度： 解释一概念所有卡片的共同特点。

成 绩： 理解概念要求的时间。

问题复杂程度： 概念中相关维度的数量，数量越多，问题越复杂。
(年龄和性别在假设中已被定义)

自变量：

材料种类

年龄（有机的）

性别（有机的）

问题的复杂程度

可能的中间变量：

个体的智力因素（有机的）

个体的空间想象力（有机的）

执行者的实施效率和准确度

因变量：

理解概念要求的时间

可能的控制变量：

年龄

性别

上述例子中，实地研究的例子包括预言的问题，例中给定的两个假设，即关于所研究现象的性质的假设没有基础，也没有最初假设。例中给定的变量也没有进行类型的划分，科学的实地研究采取描述的方法而不是因果分析的方法，它们不涉及自变量和因变量的差别，在实地研究中不存在任何控制变量，但有很多中间变量，如学生的小学学习背景。例 1 中的假设采用定向的形式；例 3 中的假设采用非定向的或零假设形式。在例 2 和例 4 中，两种假设类型都用到了，虽然为了前后一致采用了单一的假设形式，其实两种形式都是可采用的。例 5 有零假设，当然是非定向的。

如果某一学校想确定应采用某种方案，由 3 年级老师按例 2 实施，那么例 2 就是一个行动研究的例子。注意在这个例子中因变量是阅读成绩的增加分数，而不仅是阅读成绩的分数，为了测出增加分数，实施前的阅读成绩应知道；因而，实验前的阅读成绩就是一个可控变量。例 5 是一个尽可能在实验室背景下教学的实证研究的例子。

上述例子并没有包括所有可能存在的假设，它是依据研究的范围和条件而列举的。在确定变量的过程中，列出了可能的中间变量；当然，在具体研究中可能还有更多的中间变量。要注意的是性别在例 5 中是自变量，在例 2 中是可能的控制变量，在例 1 中是可能的中间变量。这是因为在例 2 和例 3 中，要确定男女性别之间是否存在成绩差异，而在例 1 中，假使性别有影响的话，它也不能从别的变量对结果的影响中分离开来。无论如何，性别总是一个有机变量，这在每个例子的括弧中都说明了。

总结

本章讨论的是问题的确定，即从问题的选择到研究问题陈述的产生这一过程。详细阐述了有关假设、变量的操作性定义和条件等方面的内容。指出研究者对问题的选择可采取不同途径，其中必须出自对某领域的研究兴趣和从事研究的内部动机。另外，外在因素也影响问题的选择。

确定研究问题的各个不同部分之间是相互联系的。问题陈述是设计假设的出发点，设计假设的过程不仅是把问题定义得更具体，而且还要有效地限定研究的问题。假设的陈述涉及到变量、操作性定义和条件。所有这些是在相关理论和现有知识的基础上得出的。

假设不是设计出来就结束了事了，而是要在研究过程中起作用。有时，一个研究方案的报告中似乎缺少假设的内容，那可能是研究者所从事的领域本来就非常缺乏背景信息，或者也许在很大程度上是理论发展所需要的。例如实地研究，原假设（最初假设）常常是粗略的且在数量上也有限，但随着研究的进行新的假设将产生，原有的假设可能被修订、保留或舍弃。如果可能，假设应尽量陈述得精确，并被当作研究的框架。

本章还描述了变量、常量和操作性定义的概念。但这些概念不仅仅用于确定问题，它们贯穿整个研究的过程。这些概念是相当重要的，因为它们是研究基本用语的组成部分。

1 查阅文献活动

查阅文献的意义

研究过程的前期工作之一，是查阅与研究课题相关的文献。在课题确定之后，了解相关的信息将有助于熟悉研究现状并使研究过程更趋有效。

要从各种不同的资料中获得大量有用的信息，阅读文献绝不是一件轻松的工作。它是一个需要认真、透彻了解并注意细节的系统阅读过程。在查阅文献时，研究者总是力图弄清楚他人对于类似的课题作了哪些研究，并尽可能搜集身边现有的材料，这一过程涉及到以下三个问题：

1. 信息从哪里发现的？
2. 在信息找到后，我们应做些什么？
3. 是什么构成这些信息。

第一个问题是有关调查报告或专门资料等来源的问题，对于大部分学生而言，这些资料常常是来自于图书馆。要获得信息一般是需要借助期刊文献索引等参考书目的，计算机查询也常常被用来挑选文献和鉴别那些最有用的材料。

第二个问题是有关信息如何组合和概括，如果一个报告的内容是有关正在研究的课题的，那么它必须以某种有效的方式被保存下来。

第三个问题在某种程度上比第一、第二个问题更为抽象。要回答第一、二个问题，研究者只要发现信息并通过适当方式使其保存下来即可，而要回答第三个问题，则须对报告中的信息作一判断，报告中的哪些信息与研究课题有关？报告的研究水平如何？因此，对第三个问题的回答某种程度上需要对报告有一些批判性的阅读，然后，才能将从相关报告中得来的信息组合在一起。

查阅文献有何意义呢？除了为研究提供内容参考外，它还有下列一个或所有方面的意义：

- 1. 更具体地限制和确定研究课题及假设；
- 2. 告诉研究者在本领域内已做了哪些工作；
- 3. 提供一些可能对当前研究有用的研究思路及方法；
- 4. 对研究方案提出一些适当的修改意见，以避免预想不到的困难；
- 5. 把握在研究中可能出现的差错；
- 6. 为解释研究结果提供背景材料。

查阅文献有着多方面的意义，对于一项设计很好的研究而言是一项基础的工作，它一般在研究过程的前期进行，它对于研究的各个部分都将提供有益的信息。

查阅文献过程

在研究课题已初步明确以后，下面的活动通常就如图 3.1 的流程图所示的，按照箭头所指方向依次进行。

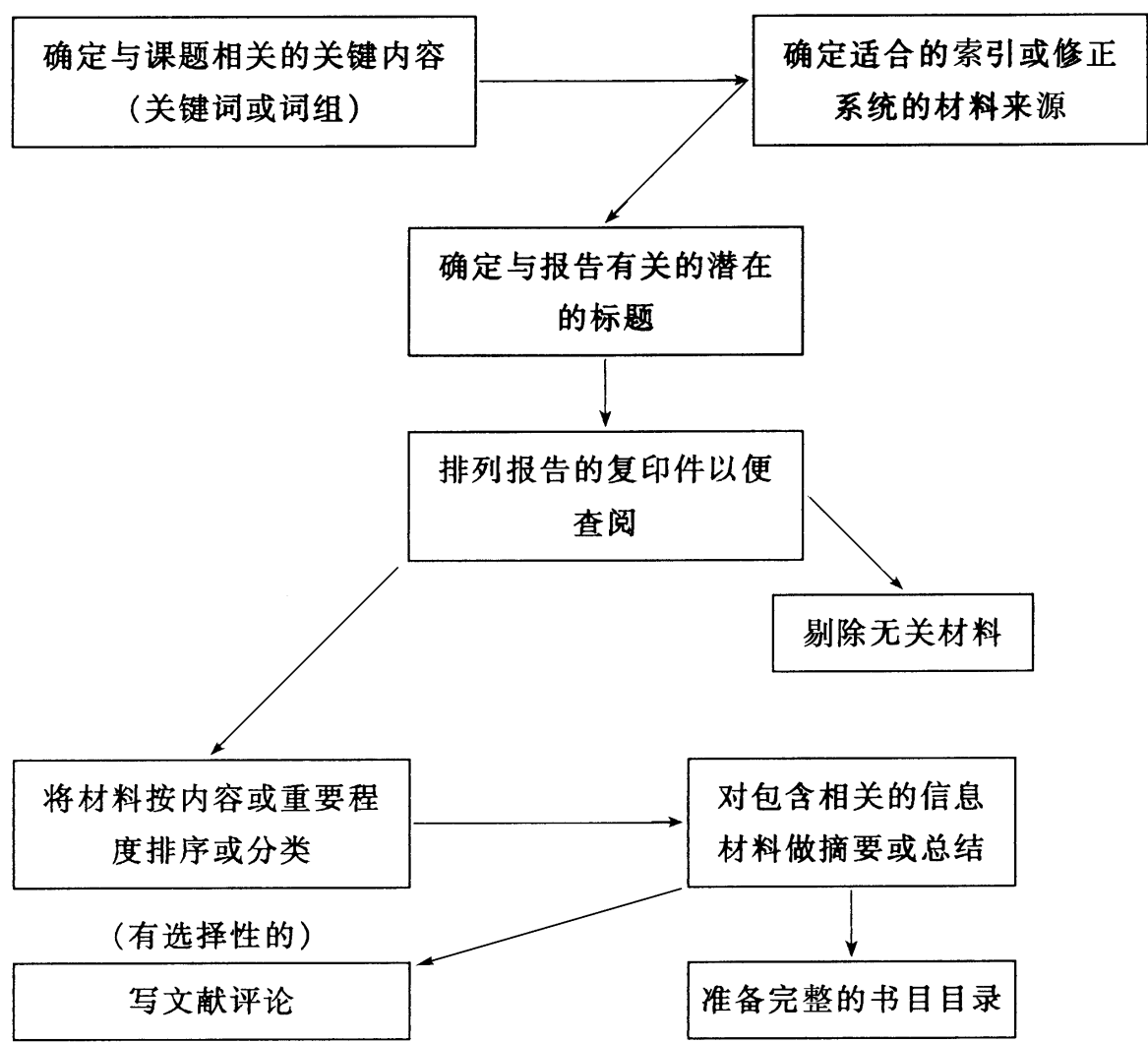


图 3.1 查阅文献活动流程图

与一项工作中的多数活动或步骤一样，在查阅文献的活动中也有高效与低效之分，研究者查阅文献时应循着流程图有系统地进行，而不是走到图书馆东抄一段，西摘一段。尽管在收集材料和报告时也可能有某些无效劳动，但只要循此进行，最终还是能提高工作效率的。研究中另一个重要的程序化观点是，不管做什么事，最初的工作做得越完备、越准确越好。比如，当相关的报告都已排了序，摘要也已准备好，就应该做好目录登记。这样免得最后写参考文献时再回过头来编目录。如果文献中有很多的报告可以提供有用的信息，做好了目录以后就不会使已有的（条件、方法、作者等）内容或结果产生混乱。查阅文献根据这样的建议进行，不仅可以减少失误，而且可以节省时间。

2 集中和总结信息

在通过文献搜索查出和阅览有关材料后，研究者就必须考虑“我该如何处理这些信息？”最初的决定就是应判断这篇报告（文章等）的内容是否与正在研究的课题有关，如果无关，可以将它删去，如果有关，就应对它所包含的信息进行总结，并将该信息以某种形式保存以便研究者一旦需要，就可以取出。在图 3.1 中，报告的各部分内容往往依其重要性程度被主观分开。而事实上，有时要做好这项工作往往是很困难的，因为重要性的差别可能并不很清楚。因此，这项活动可能被省略或者是在摘要或概括工作完成后进行。

有许多记录信息的形式或程度能帮助我们z从已有的研究中集中信息，大多数研究者都有适合于他们自己的系统做法。例如，简单的记笔记的做法就可以运用，如果不考虑专门形式的话，它是一种最好的用来不断记录信息并使重要部分不至于遗漏的系统方法，以下将介绍几种形式或做法。

书目登记

对每一篇阅读过的相关报告应做好完全并准确的书目登记，这种登记一般是摘要或总结的一部分；为便于快速查找，将书目登记写在 3×3 的卡片上或许是最理想的，如果没有其他的目的，这样的卡片对于文献阅读完之后编辑一份完整的书目摘要是非常有益的。它们可按作者姓氏字母顺序依次排列。

书目登记可以如图 3.2 那样制作，因为在图书馆中不可能携带打字机，因此大多数登记都是手写的。

一份书目登记包含下列信息：作者（们）的姓名，报告（文章、书籍等）的标题，关于出版的情况；如果文章是在一本期刊上的话，它所在的页码。关

于出版的情况包括地点、出版者、出版日期，或者一篇文章所在期刊的卷数及日期。

王某某和张某某
班级能力水平对学生成就和班级行为的影响
教育研究
1984 年第 3 期，29 ~ 34

图 3.2 书目摘要卡片举例

不论什么时间使用报告中的信息，一开始我们就应做好完整并准确的书目登记，这将有利于节省时间和减少混乱，如果书目登记不完整，又不再查阅原文，势必会有所损失，这是一件令人非常遗憾的事情。

摘要或总结

阅读文献的研究必须以适宜的形式从相关的报告中抽取信息并对这些信息加以总结。为有效地利用研究结果，必须从报告中获得相当多的信息。例如，一般而言，了解结果产生的条件是必要的，已研究过的问题也应清楚，从而能使它与正在研究的问题相联系。因此，阅读一份报告时，我们应能判断出哪些信息应该保留，并记下来。

研究者通常使用一种摘要形式，而不是作一些关于研究报告的笔记，一份摘要就是一份对包含各种信息的研究报告的总结。在阅读各种报告时，信息的内容应基本一致，由不同作者所建议的摘要的组成部分稍有些不同，但一般而言，下列内容都是应被包括的：

书目登记：摘要之前往往是一份准确和完整的书目登记。

问题：这是对正在阅读的报告中的研究课题的说明，它也可能包括对假设的说明。

对象：被研究的个体通常被称为研究的“对象”，例如，“在基础教育部登记的 50 名大学 2 年级学生，25 名男生和 25 名女生。”

方法：这部分描述研究是如何进行的，它包括测量和分析的方式等，这部分也被称之为“方法论”。

结果和结论：这部分主要指出研究的结果和结论，我们应区分结果和结论——结果指发生了什么，比如一定的统计数字；结论则是指研究者如何运用结果。在一篇报告中，如有很多结果和结论，最好将它们标上数字。

摘要是根据具体领域对从研究报告中获得的相关信息的总结。

摘要的长度往往也应考虑, 包含过多的信息反而是无效的。许多杂志在每一篇报告之前都有一份极其简短的摘要, 但由于在长度上又过分限制, 这也常常不能满足研究者的要求, 研究者所写摘要应介于这两者之间。

在概括一篇报告时, 研究者一方面要试图尽可能浓缩主要的相关信息, 另一方面又要注意包括所需要的所有细节。尽管摘要应是简要的, 但它包含所有必需信息的重要性也不可忽视。当我们发现一份摘要中的信息是不完全的并且不得不重新查找报告时, 那将是一件令人不快并且是毫无效益可言的事情。有几个因素影响摘要的长度, 比如报告的长度, 研究的复杂性及结果的广阔性。一些作者建议应将摘要内容限制在一页或 5×8 的卡片上, 这些建议有一定的道理, 因为这样做将有利于利用参考信息, 但要将摘要局限在如此有限的范围内可能会有一些困难, 摘要内容应尽可能地简短, 但不同报告的摘要, 由于所包含的必要信息的不同, 也常常引起摘要内容的长短不一。

以下是两个有关摘要的例子, 第 1 例是一篇 9 页长的文章的摘要, 第 2 例则是 16 页长的文章的摘要。当然, 报告的长短并不是影响摘要内容多少的惟一因素, 得出结论的数量、研究的复杂性、与手头研究课题的相关程度都影响着摘要的长度。

例 1

杨格 (R. E. Yager) 和波斯特 (R. J. Bonnstetter) “学生对学科教师、课堂及教学内容的看法”, 学校科学和数学, 84, 1984, 5 ~ 6 月, 第 5 期, 第 406 ~ 414 页。

问题: 了解各年龄阶段学生及青少年对他们科学教育中各因素的看法。

对象: (总计) 将近 700 名芦瓦区的学生和 9、13、17 岁的青少年及 24 ~ 35 岁的成年人, 为了比较, 运用了全国教育进步评价研究 2 500 个样本的数据。

方法: 用有 13 个项目的问题对他们进行调查, 这 13 个项目是 1977 年全国教育评价研究会所运用的。芦瓦区教育机构的科学顾问也参与数据搜集, 对各个项目的反应以百分比的形式总结出来, 并与 1977 年研究的结果进行对比。

结果和结论: 1982 年在芦瓦区调查的结果与 1977 年的研究有高度的一致性, 几乎一半初级水平的教师承认不知道如何回答学生的提问, 而且随着学生年龄的增加, 认为教师能有趣介绍科学的比例下降 (4 个年龄组比例依次为 68%, 56%, 45%, 34%)。研究还显示教师在科学上越有学问, 他们在生动有趣地向学生介绍科学知识方面就越成功。随着学生年级提高, 认为科学能使人感到乐趣、兴奋及充实的比例也逐步降低, 在青少年间最低比例达到大约 30%。认为教学内容有用的观念, 无论是现在还是将来, 随着学生年龄的增长, 都有某种程度的下降, 但最大的下降幅度在 17 岁和年轻的成年人当中。

例 2

雷得曼 (D. J. Veldman) 和杉福得 (J. P. Sanford) “ 班级能力水平对学生成就和班级行为的影响 ”, 美国教育研究杂志, 21, 1984 年秋, 第 9 期, 第 629 ~ 644 页。

问题: 这项研究主要提出三个问题:

1. 在较高和较低能力的班级间, 学生的课堂行为和成就水平是否有规则地变化?
2. 在一个班级内, 较高能力和较低能力学生的课堂行为是否有规则地变化?
3. 学生的能力水平与班级能力水平的相互影响是否会有规则地引起学生的课堂行为和成就的变化?

对象: 从得克萨斯州的 58 个数学和 78 个英语班级中抽取的初中学生 (7、8 年级); 数学班级近 500 人, 英语班级近 650 人。

方法: 回溯研究; 对观察者估计的高水平的 25 名学生进行数学和英语成绩测量, 对低水平的 25 名学生的行为进行测量; 与以前的 (平均) 模式 (数学和英语分开) 进行比较, 以确定学生能力的相互影响对年级、班级平均能力、学生能力及班级能力的促进作用; 标准变量是班级平均数或学生, 成就或行为测量。结果:

A. 在学生成就方面:

1. 高能力学生在数学和英语 (所期望的) 方面都比低能力学生表现好。
2. 在数学和英语方面都存在相互影响 (班级能力与学生能力之间); 在数学方面, 高能力班级的成员对低能力学生的影响比对高能力学生的影响要大; 在英语方面, 结论相同, 只是其程度上差距更大一些。

B. 在课堂行为方面, 观察者打分定级

1. 25 名被观察的学生中, 14 名在数学和英语两方面显示了与班级和班级内学生能力水平的显著相关; 级别的确定与这样一些行为有关: 工作习惯、动机、忍耐性、自信、依赖性、学术领导和班级参与; 高能力的班级和学生在学生行为问题、侵犯性、粗话及学术领导方面级别较低; 与外向性、和教师的交往、生理成熟、不快、喜欢合作和与邻座同学说话方面没有显著相关。

2. 在相互作用方面, 数学有 8 个变量, 英语有 4 个变量, 其中只有 3 个变量是共有的, 一般而言, 班级能力水平对低能力学生行为的影响要比对高能力学生的影响大, 高、低能力学生在一起的话, 往往对低能力学生带来消极影响。

C. 在课堂过程中的变量

1. 高能力班级较少有固定的接触, 学生有较少的失误。

2. 高能力学生得到更多的回答机会，有更多正确的和较少错误的答案，与教师敌对的交往及错误行为较少。

3. 在数学和英语课上，4/5 的有效交往的方式是一致的。一般而言，低能力学生在低能力班级比在高能力班级更容易捣乱；在英语而不是在数学课上，低能力学生在高能力班级较少做出错误行为。

每个摘要例子都有一个完整的书目登记，第 2 篇摘要的文章有 3 个需要提出的研究问题，这就增加了问题说明的长度。因为研究的问题是比较广泛的，自然也就产生了相当多的结果，这些结果比第一篇文章的结果更加明确地被分列出来。在第 2 篇摘要中，最后的结果是笼统地进行了综合。如果结果是对应于 3 个问题分列出来，那摘要就太长了。

摘要只是反映结果的意思，而不是结果本身，它们可以写在卡片上或有规格大小的纸上，重要的是它们应是有用的，并能包含必要的信息。尽管摘要的版式有一定要求，但研究者在做摘要时仍有充分的灵活性以满足自身的需要。

3 解释和利用信息

当我们阅读报告时，对其结果或对在研究报告中发现的其他信息的解释也就开始了。阅读报告时，先浏览一下，获得一个总体印象，而不要过多地关心细节，也许是一个比较好的做法。有了一个总体印象，我们就能判断该报告与正在研究课题之间的相关程度。如果两者是相关的，那我们再重点阅读并做好摘要。

批判性阅读

众所周知，教育研究文献的各种报告在其质量及综合性方面有很大的差别。因此，研究者在阅读报告时，就应带有某种程度的批判性。但一个人如何批判性地阅读？阅读研究报告需要读者智慧上的努力，作为读者，至少也应具备研究方法方面的知识并熟悉本研究领域（不一定熟悉该领域的某一研究）。事实上，学习研究方法的目的之一就在于更好地理解研究文献。

无论如何，读者所寻找的报告应具备几项特定内容，如果这些内容比较薄弱或根本没有时，那该报告就值得怀疑。史密斯（Smith）和格拉斯（Glass）（1987）称这些内容为标准和工具。他们从合乎逻辑的效度规则入手，关注效度的不同形态。这种效度决定最终的研究与报告如何有机地融为一体。报告的各部分是否协调，其意思是否已充分表达？研究成果得出的结论是否符合情

理？两者是否一致？研究方法与研究过程是否一致？

我们还要继续阐述研究的内在与外在效度。内在效度主要强调多样和适当的研究方法以使其结果令人信服，研究程序进行得是否恰当？研究过程是否带有偏见？研究程序是否已充分说明？是否足以使读者明了其研究过程？分析得是否得当？

外在效度主要关心研究结果的概括程度，有关的内容在报告中是否谈到？如果有所论述，那研究结果是否有充分的推广价值？结果的概括是否有充分的理由？如果没有，那在研究结果的代表性方面是否有争论之处？

以上讨论了研究者在阅读研究报告过程中必须牢记的几个问题。许多更具体的问题可能被提出来，它们是阅读报告的副产品。这是批判性阅读中的智力训练。如果一篇文章在讨论中是跳跃的，或没有遵守逻辑规则，那它对研究者是没有什么帮助的。内在效度在解释最终结果方面必须有充分的证据和理由。

批判性阅读是一种智力锻炼，它要求读者判断报告效用的适用性。

写综述

当研究者觉得一篇报告可以利用，或至少大部分内容是可以利用的话，那有关的信息就应综合起来并将它们写进该文献的综述之中。研究者必须决定如何利用这些信息。查阅的文献摘要可以根据前述讨论进行分组。如果有明显的缺陷，我们还必须重新回到文献中去，并扩大搜索范围。正是在这一点上，研究者可以利用摘要并将阅读中的难题拼合起来。

综述的长度可以依据正在准备的研究报告的类型而定。查阅的文献的数目比文献综述的数目大得多。在一篇为专业性杂志所写的文献中，其综述是非常有限的，可能只有一页纸。当然，有关研究问题的来龙去脉及其他某些内容还是必须交待的。在这样的情形下，综述可以包含 6~8 本参考材料，有时会少些（在一些文章中，由于刊物版面的限制，只包含了 1~2 本参考书），这并不意味着研究者只读过 2~3 本报告，许多内容可能都已读过，但只有最相关的材料才被提及。当一份综述是为较大的研究而做的话，比如一篇博士论文，许多的报告，可能有 50 本或更多，都将被涉及。在此情况下，文献综述成为一项主要任务，所写的综述自身一般也含小标题，而且从摘要中所抽取的信息也是按照小标题组织起来的，这样的综述可能占全文的五分之一到七分之一。

如果不考虑综述长度的话，研究者应尽可能在综述中包含最新的信息，这并不是指过去的信息是不相关的，而是综述应超前于时代。如果没有一份参考材料是近 10 年来的话，这样的综述无疑有严重的缺陷和不足。

列参考文献

当一份材料的信息被引证时，它就必须被详细地说明。可以参考的说明样式有很多，其中之一即是在本页底部作脚注。例如，对某一研究有如下的叙述性评论：

杨格和波斯特 1982 年对学生的研究与 1977 年由全国教育进步评价研究会所得出的结果一致。

本页的底部就可以利用脚注注明：

杨格和波斯特“学生对学科教师、班级和教学内容的看法”。学校学科和教学，84，第 5 期（1984）P . 409。

为了叙述的方便，如果一篇文章有 3 个或更多的作者，那第一作者的姓可以列出，其后只加等人，而不是将所有姓名列出。介乎于这两者之间的做法是，第一次引用参考材料时将所有姓名列出，此后再次引用到该材料时，只要用“等等”即可，这是一种出于简短的目的而采取的做法。在脚注中，则通常应包括所有作者的姓名。

对于那些不需要在文章底部作脚注的参考材料可以运用一种较短的样式来说明，如在每一章或报告的最后可以列出参考材料的名单或目录，一种样式为作者——日期样式，该样式的例子如下：

杨格和波斯特（1984）得出了与……一致的结论

或者，如果在叙述中作者的姓名无需强调，则可以：

有研究者（杨格和波斯特，1984）得出了与……一致的结论

这种样式应将所有出现的参考材料都按字母顺序排列出来，如果杨格和波斯特的文章在目录中不止一次被提到，并且所用的参考材料就是他们 1984 年第 2 版上的文章，那参考材料注释方法就是“杨格和波斯特（1984b）”。如果研究者希望在参考材料中注明页码，则可在年份之后，当然在页码与年份间用逗号隔开。

参考材料的说明相对较短的方法是用作者——号码样式。在这种情况下，目录登记可以数字化，用参考材料的页码来代替出版物的年份。如果参考材料包括刊号的话，那页码可紧接其后。例如，如果杨格和波斯特的文章登记号为30，参考材料便是：

杨格和波斯特 (30 409) 得出了与.....一致的结果

或者

有研究者 (杨格和波斯特 30 409) 得出了与.....一致的结果

作者——日期样式一般而言要优于作者——页码样式，因为后一种样式如果参考材料的内容被增减，那参考材料所有的页码可能由于增减而相应地改变，在一篇好的评论已写完的情况下，如果要增加一篇参考材料，那必要的改变将是令人心烦并可能带来前后混乱的错误。

参考材料可以通过几种方式列出：传统的脚注或更简短的方法，如作者——日期样式或作者——页码样式。

准备文献目录

通常，查阅文献的最后一步即是将文献目录集中起来，一些杂志将参考材料名单与文献目录作了区分，参考材料名单仅限于在报告中引用的材料，而文献目录则可能还包括为了解背景或深入阅读提供信息的材料。尽管文献目录可以包括在评论中并没有直接引用的文献，但包含过多没有引用过的文献的目录并不是明智的。

文献目录中的项目是按照原始作者（列在第一位的）姓的字母顺序排列的。如果项目内容已被列入文献目录卡片，则卡片排列顺序是按字母顺序或其中的项目顺序，这种做法简化了誊写文献目录的任务，特别是并非所有原始资料都被运用时更是如此。摘要有时也可以利用，尽管它们或许会有少许的麻烦。

在英文中，文献目录的项目往往自第一作者的姓开始，然后是其他作者，再后是出版物的标题。文章或章节的标题常用引文标志注明。书名，期刊名及其他等等，则在其名称下面划上横线。期刊的出版日期及卷数也应注明。书籍

出版者的名称、地点及日期也须提到。最后，如果需要的话，文章在期刊杂志上的页码同样要注明。

中文文献的引用顺序是：作者姓名，书籍或刊物名称，出版者，出版时间，或期、卷号数，页码。

总结

本章介绍了查阅文献的过程。查阅文献具有几方面的意义：它明确了研究课题的地位；它提供了已有研究的信息；它常常还能提供如何进行研究的信息，包括对手段及研究设计的建议。

查阅文献的范围根据几方面的因素，但通常它需要我们将时间和精力花在细节方面。有时，一位研究者可能会这样说：“这份文献中没有任何与我们的研究课题有关的信息”，这句话是指他还没有发现与他正在考虑的课题极其相像的研究。文献查阅可能包括那些甚至并非与研究课程直接有关的研究，查阅者的任务就在于从查阅过的若干文献中发现它们的相关关系并综合有关信息。

图 3.1 在总体上列出了查阅研究文献的各项活动，尽管一位研究者有时因为增加的信息而回过头来查找文献或重写评论的某一部分。但一般情况下，查阅的各项活动都是按程序图 3.1 由顶部到底部按顺序进行的。文献阅览常常是在研究活动的早期进行；但如果研究需在很长的时间段内进行，则也要不断查阅，以便能紧跟时代的有关信息。

第四章

定量研究的设计

1 研究设计的目的

当研究者已经确定了研究的问题并且完成了一些文献资料的准备以后，接下来就是进行“研究的设计”——即引导研究进行的计划或策略了。作为一种计划，研究设计涉及诸如选择研究的参与者，数据收集的准备等研究过程的各种活动。本章着重于定量研究的研究设计的一般原则，而定性研究设计的一般原则，将在后面的章节讨论。定量研究和定性研究有着不同的研究过程。定量研究有它的实证研究的基础，与定性研究相比，定量研究的研究设计更倾向于结构化和规范性。同样，顾名思义，定量研究的结果通常是由大量的数据来表示的，研究设计是为了使研究者通过对这些数据的比较和分析作出有效的解释。

在多数情况下，研究设计的需要是不明确的。怎样才能更深入地研究？应该有一个研究计划，该计划能够产生具体的研究活动并获得成功的结果。研究设计有两个基本目的：(1) 提供研究问题的答案；(2) 控制差异。

第一个目的是一般的、直接的——即为具体的研究提供问题的答案。但只有进行研究的愿望或研究活动的努力并不一定能产生答案。研究应该是有效的，这包括它能解释结果，并且通过这些结果来回答或揭示研究的问题。好的研究设计有助于理解并解释研究的结论、保证研究者获得有用的结果。

进行研究的目的是为了解释差异——不是所有的个体都是一样的或具有同样的得分或量度。例如，在小学学生的学业成就、动机、态度、年龄和家庭背景等方面，差异是被看重的。同时，任何一个变量的差异也是被看重的，它可能受任何一种因素的影响。例如，在学习成就上，差异可能是由于态度和动机引起的，或者是同时由于这两种可能的因素引起的。为了控制差异，在定量研究中采用了设计程序。控制差异意味着能够解释或说明由研究变量而产生的差异。差异具有数量的意义，在某种程度上，差异的区分有助于考察研究中的变量。

解释和控制差异是定量研究的一个重要组成部分，研究设计将说明这个问题。

2 控制差异的概念

在定量研究中，我们试图将差异量化，并且尽可能按照各种理由或原因将差异区分开来。但是，什么是差异？我们来举一个例子。一个高中的化学教师正在研究3种不同的教学方法对于学生化学成绩的影响，这个问题可能是这样表述的：

一项关于某种教学方法影响高中生化学成绩的研究。

这个问题暗示了将要做一项实验，因为教学方法这个最感兴趣的自变量，将由化学教师来试用。教学方法有3个不同的种类（或水平）：M1、M2和M3。因变量是一学期以后学生化学成绩的表现。90名学生，全是同一学校由同一教师任教的高中低年级学生，参加这一研究。当测试了这90名学生的化学成绩后，得到了90个不同的分数，同时得到了这些分数的分布图，这一分布存在着差异。

为什么这90个分数并不都一样？这可能有多种原因。教学方法也许有不同的效果，既然方法是自变量，研究者当然希望能确定这3种教学方法是否有不同的效果。一些学生无疑是比其他一些同学更有学习能力，这与教学方法无关。可能教育的时间长短产生了效果。这里的变量也许是在第2章已经讨论过的控制变量或中间变量的例子。毫无疑问，在一定程度上学生对化学测试的反应是有内在差异的。任何一种因素对于导致因变量分数的差异都可能是起作用的。

让我们来考虑这90个化学测验分数的差异。正如我们将在统计分析一章中可看到的，差异可以被量化，它是一个正实数，比如103或109大于零或等于零，零差异表明没有差异，那就是在一个分布区域中所有分数都是相同的。数字越大，差异也就越大；然而，当分析差异时，我们更感兴趣的是对有关的差异进行比较而不是这些数字本身的价值。假定90个化学测验的分数的差异是360，尽管差异并不一定发生在一个圆圈内，但可以对差异进行量化。

差异可以被量化为一个正实数。零差异表明在某一区域所有的分数都是相同的。

控制差异的方法

这一节我们将继续讨论化学测验的例子，并解释通过加强控制差异所产生的结果。教学方法是独立于兴趣的变化，同时学生的能力将作为一个可控制的变量来考虑。有 4 种控制差异的基本方法：

- 1. 随机化。
- 2. 设计中建立一些条件或因素作为自变量。
- 3. 保持条件或因素的不变。
- 4. 统计调节。

前 3 种方法直接包含在研究设计的结构中。第 4 种方法运用于试图进行控制的统计矫正中，它在研究的每一步骤都有运用，尽管研究设计中已经作了准备。

随机性 假设在上面的化学教学研究中同样的教师对这 90 名学生进行教学，以 30 个学生为一个班分 3 个班，每一个班用不同的方法进行教学。假设这 90 名学生代表了一个总体中的典型样本，而不是不同能力水平的不同类型的组合，那么对于化学测试水平就可能有一些影响。30 个学生被任意地安排到某一种方法的教学班中，见图 4.1，这样，能力水平被随机地分布在 3 种方法的班级中。人们期待在每一个班级中的效果是一样的。因为，能力水平的影响已经广泛地渗透在这 3 个班级中，并且这一影响对于所有 3 个班级都是相同的。尽管这一做法只是分散了能力水平的影响，但由能力水平引起的差异不是由研究者决定的。

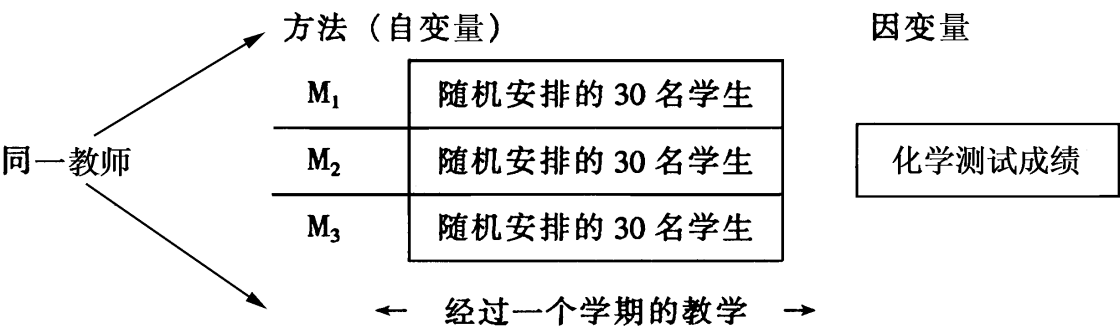


图 4.1 研究设计中用随机方法控制能力水平

记住化学测验分数是因变量，90 个测验分数的差异是 360。假定对应每种

方法 M1、M2 和 M3 的 30 个测验分数的平均分数分别是 89，75，和 96，同时假定由于平均分数的不同表明的差异是 360 中的 50。那么差异的量化可以以图 4.2 来表示。整个 360 差异被分成两部分，由方法引起的差异和其余的差异(随机或固有差异)。量化的话，前一部分是 50，而后一部分是 310。随机差异中包含由学生的能力带来的差异，但因为随机化，这个差异被平均分布在 3 个学生组之中了。

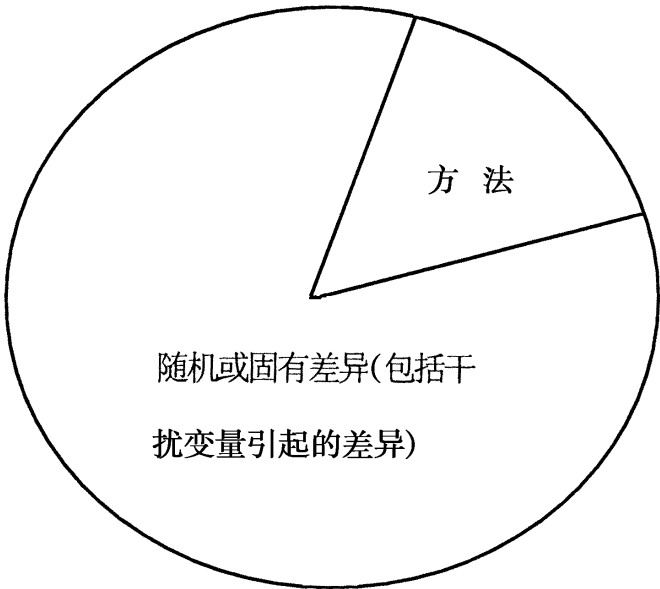


图 4.2 90 个化学测试分数差异的定量表示

在这个例子中，我们可以看到，随机性也可以控制其他变量——与学生有关的最初的、有机的变量。例如，动机的水平将是随机地分布在 3 个班级的学生中，影响化学测试成绩的教学和技能也是随机地分布的。

随机性将变量的影响平均地分配到研究组群中去。通常，有机变量就是这样被控制的。

在这里讨论一个有关的术语。随机或固有差异，也被称为误差，特别在作统计分析时经常使用。误差不是指发生了错误或类似的情况。它的最直接的意思是由随机分配或随机选择所产生的固有的或自然的误差。

设定自变量因子 在如图 4.1 的设计中，按照教学方法将学生分成 3 个组。如果能力水平测量是有效的（可能是最近的一次智商测试中的得分）学生也就能按能力水平进行分组。如果把在能力水平测量中比较好的 45 名学生称之为能力较高者，其余 45 名学生称为能力较低者，那么“能力水平”对于这两组学生来说则是一个自变量，将这两个能力组中的 45 名学生按高、低能力组随机安排到 3 种方法的教学班中（这些组中的能力水平的任何差异都是随机

分布的)。这一设计体现在图 4.3 中。

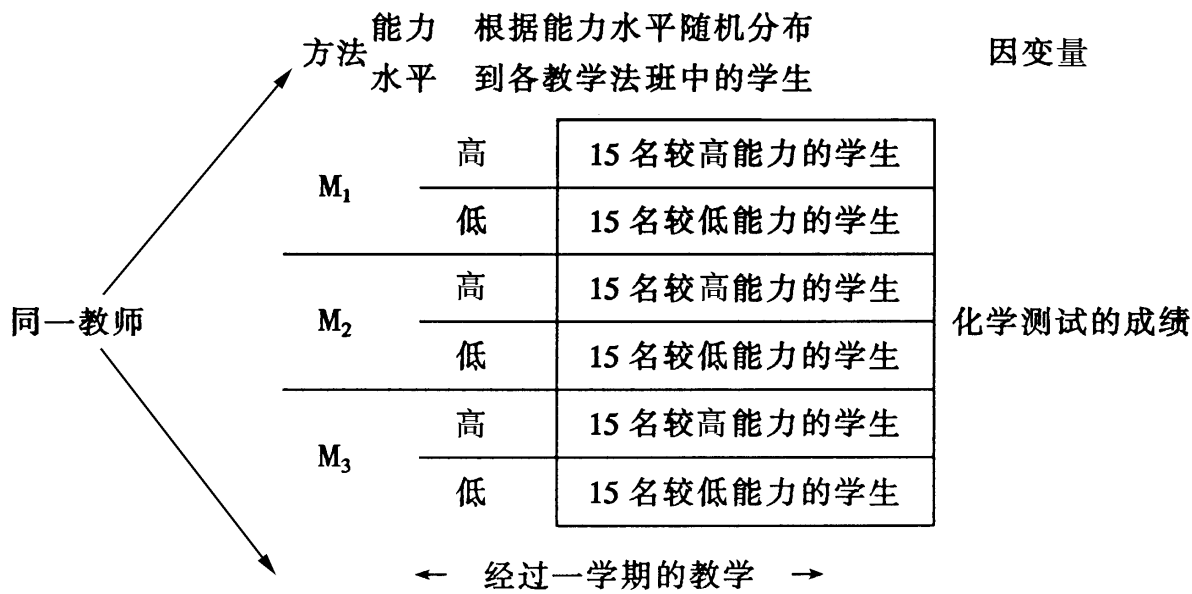


图 4.3 把能力水平作为自变量的研究设计

那么，这时测试成绩的差异不仅是由方法之间的不同来决定，而且还由能力较高和能力较低之间的差异来决定。可能由于能力水平影响而导致的差异能够得到说明，假定能力水平产生的差异是 75（这将由化学测验分数中两个能力水平组的不同来决定），教学方法仍占整个差异中的 50，那么随机差异现在就是 235（ $360 - 50 - 75 = 235$ ）。图 4.4 反映了对整个差异的划分。

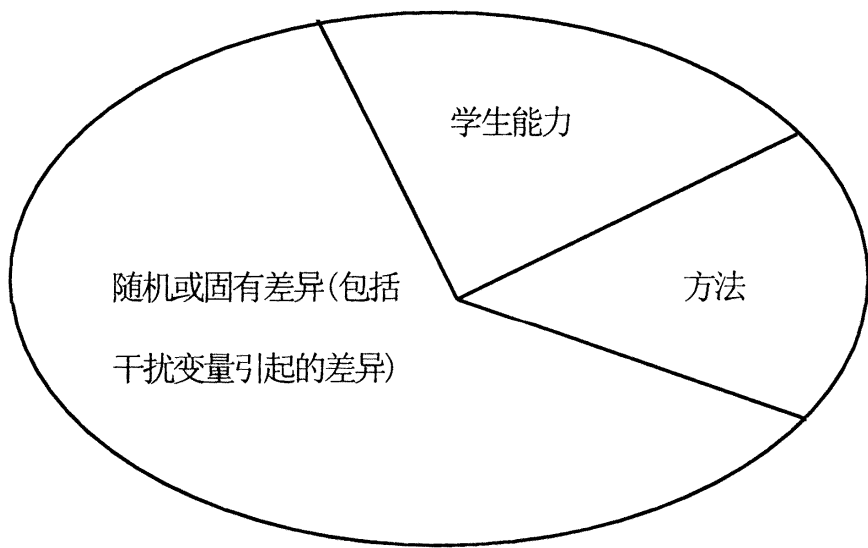


图 4.4 把能力水平作为独立变量的 90 个化学测试成绩
分数差异的定量分析

在这一例子中，将学生分成两个而不是更多的能力水平是随机的（也可以分成高、中、低三个能力水平组）。应该注意的是，研究者随机地分成两组以保持每组人数的相等。为此，我们没有运用智商测验的设计者提出的关于能力水平高低的定义（假设这些定义是可行的），因为已经设定好的定义无疑会形成每一能力水平学生的不平等。事实上对于化学测试来说，无论是谁都希望更

少的学生属于低能力水平组。

建立变量时应允许研究者确定变量的影响，也就是说可以确定可归因于哪个变量的差异，或者确定哪一变量能解释差异。然而，为什么并没有一直运用这一做法呢？首先，个体得以被区分开的测量并非总是可行的，在上面的研究中学生的区分也是这样。例如，动机水平也许有影响，但由于这一变量而引起的学生学业成绩的变化是很难测量的。再者，动机水平可以作为一个中间变量，但如果引用随机分布的话，则它是随机分布的。同时，引进更多的自变量可能会使研究设计变得不必要地复杂起来。因此，一般考虑的因素通常是能引起因变量变化的一些因素。

在研究设计时建立一些作为自变量的因素能使研究者确定这些因素的影响。然而，太多的自变量也会使研究设计变得不必要地复杂起来。

保持因子不变 保持某一因素恒常不变，实际上就是减少变量的变化使其趋于稳定。在上面化学测试的例子中，如果教师能够把能力水平的影响降低到恒常状态，如果仅仅具有一种确定的能力水平的学生的话，换句话说如果智商测验成绩都在 100 到 108 之间，如果能力水平趋向于在因变量意义上影响学生的成绩，那么它的作用将肯定比它已经起作用的范围更小。由于能力水平而引起的化学测试成绩的差异大多数将会缩小。研究设计将是与图 4.1 基本相同，除非那些处于确定的智商范围的学生被随机地分布到各教学方法的班级。如果教师已经开始了对 90 名学生进行教学，那么这一数字可能是每一组少于 30 人。存在于学生中的化学分数的差异将会发生变化。来自于学生中的差异将倾向于减少，因为现在学生在能力方面更一致了。

保持因素恒常也会带来一些缺陷。其一是已经被提示的：研究过程中个体数量的减少可能引起计算上的问题，或减少了作为因变量的可行的数据的数量。同时，这种结果仅仅是概括了限制组。化学教学的例子已经有一些恒量运用在研究设计中，例如教师和教学的时间。

当一个因素保持不变，一个潜在的变量就被保持了恒常。这就削弱或至少降低了对因变量的影响因素。

统计控制 当我们使用统计控制时，它是指适用于数据分析时的计算机处理程序，但其中的控制变量是在研究进行前就计划好的。对于化学教学例子中的能力水平变量来说，如果我们能很方便地获得在最近可控的智商测验中的

90 名学生的能力数据，能力对所有 90 名学生就是变量。很可能在智商测验中的得分和在化学测验中的得分存在某种关系，因此可以说，在一方面获得高分很可能在另一方面也获得高分，同时，低分也可能是同时存在于某一个体。

如果我们能以某种方式来调整因能力水平不同而获得的化学测验成绩，我们就能控制学生的能力对化学测验成绩的影响。这可以由相对成熟的统计方法来完成。依据化学测验和智商测验之间的关系的强度，如果学生有较高的能力水平，那么化学测验成绩应该降低（通过统计处理）；同时具有较低能力水平的学生则相应提高他们的化学成绩。这样做调整了化学测验成绩，能力水平的独立性就可以进行分析了。能力水平的独立性指的是能力水平的影响已经被去除了。

统计控制过程可以通过图 4.5 来表示。另外，我们可以对 90 个化学测验成绩的全部差异作具体分析。与将学生能力水平分为两个组的自变量的方法相比，统计控制将可能说明更多的差异。以智商测验这一被认为具有差别等级的测量作为衡量学生的标准，要比把学生分成能力较高和能力较低两类能更好地反映差异。假定统计控制表明在化学测验分数中，能力水平所产生的差异占差异 360 中的 130，教学方法仍占 50，这样随机或固有差异现在减少到 180 ($360 - 50 - 130 = 180$)。图 4.5 表明了这一结果。

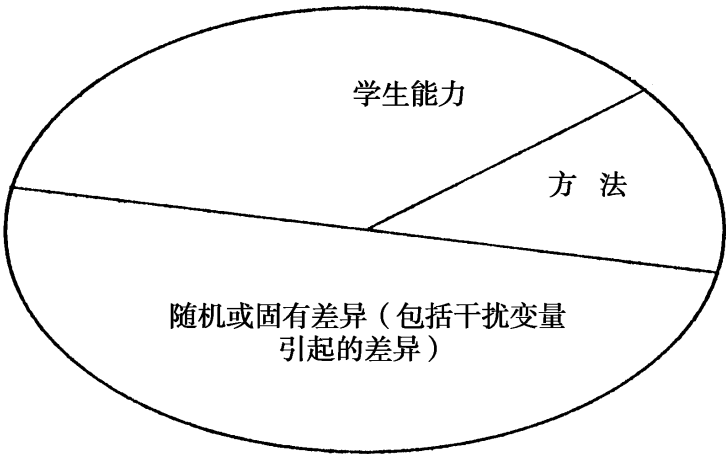


图 4.5 把能力水平进行统计控制的 90 个化学测试成绩分数的
差异的定量分析

实质上，统计控制也就是调整因变量的分数以便去除控制变量的影响。

如果说统计控制是一种简单地通过计算机进行分析的活动，为什么它没有频繁地被运用呢？一个原因是获得适当的控制变量的分数通常是困难的。分数对于所有的研究参与者来说必须是易得到的。控制变量的分数应该在研究进行

前获得，这样才可能避免分数受自变量影响和可能性。在上述化学教学的例子中，自变量也就是教学方法。再一个原因就是控制变量只有在它们与因变量相联系并且考虑到变量分数差异时才是有效的。另外，一定的统计假设也应该是适当的，如果统计假设站不住脚，那么应避免使用统计分析的方法。

控制方法的组合运用 加强控制的 4 种方法可以单独运用，也可以彼此组合运用。一个程序可以用作控制一个变量，另外一个程序可以控制第 2 个变量。化学测试例子的扩展可在某种程度上帮助解释这一点。

如果不是只有 90 名学生，而是一个更大规模的高中有至少 180 名学生参加测试。还是 3 种教学方法，而且仍以此作为自变量。由于有了 6 个教学班（每班 30 人），所以使用了两名教师 T1 和 T2。据了解，学生来自不同的学科，他们的小学阶段和已经接受的中学阶段的数学背景也不一样。例如，一些学生已经学过《代数》，而其他一些人则没有学过。当学生们都是初二年级时对他们进行智商测验获得了智商分数。在这一学区有其他两所中学，在这两所中学之间，学生的写作水平可能由于社会经济背景而有所不同。除了自变量之外的 4 个变量及其控制方法如下：

变量	控制方法
1. 学生的学科背景	1. 随机化。参加该研究的 180 名学生随机分布，每种方法有 60 人。
2. 教师	2. 建立自变量，每一个教师用 3 种方法教学。
3. 学校	3. 保持恒常，仅仅采用一个学校的学生。
4. 能力水平	4. 统计控制，运用智商测试成绩。

这个研究设计可以由图 4.6 表示。

随着控制程度的相对提高，如果在使用 3 种教学方法的学生的化学成绩之间出现足够大的差别，那么这个结果将表明差别来自于不同的教学方法。这个研究设计具有高度的内在效度；结果也能十分肯定地得到解释。然而，结论的推广性在某种程度上受到了限制。因为仅仅在一所学校，仅仅由两名教师进行了试验。这个结果仅仅适合于对这个学校的学生和教师的推广，他们是一个更大范围的总体的代表。它的意义也许仅仅是概括了这个学校的学生。但如果涉及到更广泛的一般概念研究的外在效度必须在逻辑基础上予以阐明。当要决定有关自变量的影响时，要将自变量所产生的差异与随机差异或误差进行比较。如果随机差异由于包括了由无关变量所引起的差异，并有较大的部分，那将使结果难以解释或可能做出错误的解释，因此内在有效性将受到危害。尽管随机性将一个变量的影响均匀地散布开，但它不能消除随机或固有的差异。有些变量，如学生的学习方式，在可能与它有关的大多数情况下实际上都不可能用其他的方式控制。所以，我们想防止有太多的差异无法解释的局面，就必须将控

制差异的方法组合起来运用。

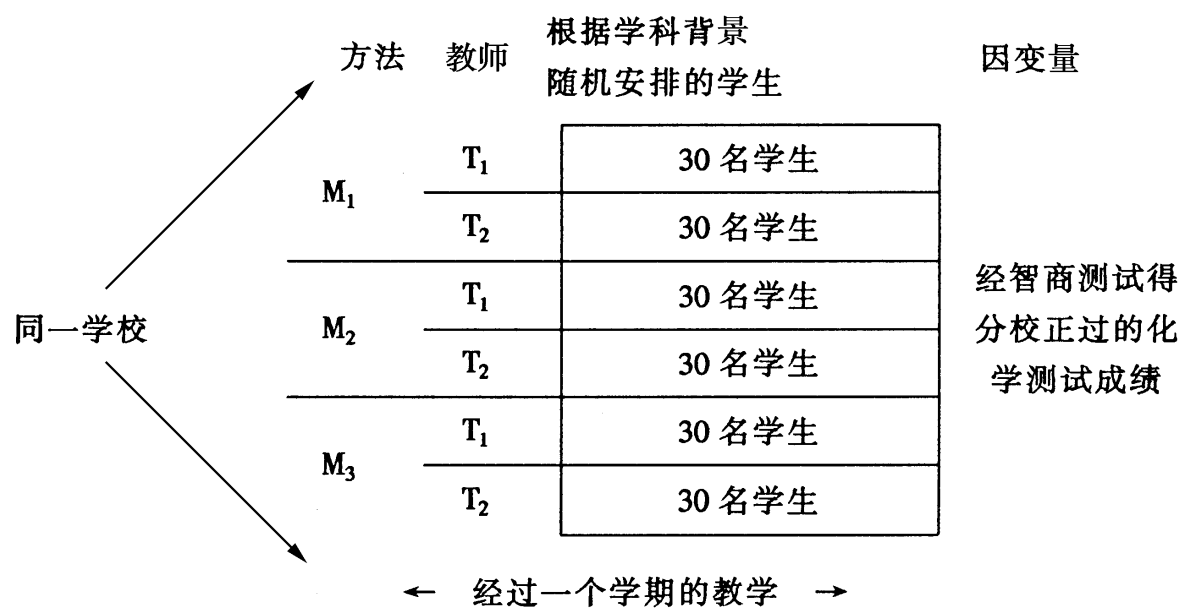


图 4.6 由不同控制方法控制的学校、学科背景、教师
和能力水平的研究设计示意图

定量研究的研究设计是为了控制或解释差异。它通过增加某些研究条件，例如增加附加变量为自变量，使这些变量得以控制来实现。

控制差异的目的是为了增加对结果的解释，以便研究者能够告诉我们变量具有什么效果。研究设计被推进的时间应当是考虑控制的时间。仔细的设计研究有助于提高研究的效度；如果不是这样，也许可能会无法解释或得不出结论。

3 良好设计的特征

良好的研究设计是怎样构成的？有一些一般性的回答，如设计应该是适合于假设的，或者在可利用的资源条件下是行得通的。然而，更具体地说，以下 4 个特征将保证研究设计是好的。一个良好的研究设计不仅必须是适当的和易操作的，而且应该是能从研究中产生出能够充分解释的结论。

排除偏见

良好的研究设计的第一个特征是研究能提供没有偏见的资料。这意味着获

得的资料和运算过的统计在任何系统中都不会有很大的变化，而只是基于随机波动的期望值。因此，出现的任何差异都将归因于研究中的自变量。如果在资料中存在某种偏见，发现的差异就不能归因于随机变量，也就不可能决定这些差异是由于偏见的作用还是由于自变量的影响。

偏见影响数据的方式很多，包括实验程序中的个体的有偏见的分布。例如，在上述化学教学的例子中，如果较高能力的学生都被安排在一种教学方法的班级中，这将会引起偏见。

资料中的偏见也许会由于个体的随机分布或随机抽样而被缩小，但偏见可能会由于独立的随机处理而产生。在化学教学的例子中，随机安排学生将会缩小由于高能力学生被安排在一种教学方法班级而引起的偏见（胡乱分布的可能性是存在的，但这种可能性是很小的）。当然，偏见仍有可能存在，例如，如果我们使用三个教师，每一个教师使用一种教学方法，在三者之间肯定会出现不同的效果。这时也许会由于学生的随机分布而引起了偏见，这就引出了混淆的概念。

避免混淆

由于变量的混淆也能引起资料的偏见。两种或更多的自变量或其他变量混合在一起，而它们的影响又无法分开，便会导致混淆。一个好的研究设计应减少变量的混淆或使这种混淆处于最小程度，这样各种影响就能够被分析、并且结论能被清楚地解释。在上述所说的3个教师的情况下，每个教师使用一种方法，“教师”和“教学方法”实际上被混淆了。如果一种教学方法班级的学生化学测验中的分数高于其他两种教学方法班级的学生，我们无法知道这种较好的成绩是由于教学方法还是由于教师的影响的缘故，由于每个教师只用了一种教学方法，所以我们无法区分出到底是教师还是教学方法的影响。

如果两种或更多的变量的影响不能被区别，那么我们说，这些变量被混淆了。

控制无关变量

尽管无关变量在研究设计中不是最主要关心的变量，但它们也许会对因变量发生作用。控制无关变量是为了能够辨别、平衡、缩小或减少它们的影响。当无关变量影响因变量的差异时，说明无关变量发生影响了。因此，控制无关

变量与控制因变量的差异可一起实现，这正如我们在前面所描述的那样。在化学教学的例子中，学生的能力水平被认为是一种无关变量，控制它的多种方法我们已经讨论过。一个好的研究设计能控制这些无关变量，而不是混淆它们与其他变量的作用或者忽视这些无关变量的作用。

检验假设时统计的精确性

在定量研究中分析过程一般涉及到统计程序的一些类型，因为定量研究包括对数字的分析和需要对假设进行统计检验。拥有适当的相当精确的资料能够证明假设是可信的，这一点非常重要。从统计学来说，精确性随着随机性或误差的减少而提高。与我们讨论的化学成绩测验的例子相同，当无关变量被控制时，随机差异就趋于减少。

当统计程序被用于检验假设时，因为总会有假设的，获得对资料的随机差异的估计是必要的。估计的精确性越高，对自变量的影响的分析就越灵敏。这就是为什么如果影响存在的话，分析将能更好地说明这一影响。无关变量影响着随机差异的估计，并可能使得统计检验对于实际存在的差异变得不灵敏。

研究设计应该为检验研究过程中的所有假设提供资料。有时可能涉及到大量而且复杂的假设，因此研究者应该仔细地核对并弄清该设计的哪一部分将为检验假设提供资料。

总结

本章讨论的思想适用于定量研究中的研究设计。定量研究从一开始就具有典型的研究设计的结构性，并且在研究的设计中很少有偏差。

除了提供研究问题的答案外，研究设计的一个基本目的就是控制差异。控制差异意味着能够解释事情发生的原因，至少能够指向较为肯定的结论的解释。我们也讨论了控制差异的4种方法，这些方法可以在具体的研究设计中组合运用。

我们界定了良好研究设计的4个特征，这4个特征绝不是彼此孤立的。例如，当我们进行设计以避免混淆时，我们可能控制无关变量。尽管紧接下来的研究中不可避免地会出现一些问题，但这些问题能够通过一个好的研究设计而避免发生或纠正。

当然，对于特定的研究来说，研究设计应该是明确而具体的。例如，本章所讨论的化学教学的例子就是一个实验——它很可能由行动研究所组成。特定的设计往往依赖于研究的特定目的和条件。在下面几章中，我们将讨论不同类

型的研究，每一类研究有其重点、一般设计、以及许多可能的变量。

有时我们说“选择”一个研究设计。确实，设计是我们选择的，但一项具体研究的变量、条件等是不可能囊括无余的。因此，寄希望于选择一个设计来完成获得完全合适的研究设计的任务的想法是不正确的。选择的设计还必须要转化成具体研究。

1 实验设计的含义

“实验”一词在目前的运用中有着似乎熟悉而广泛的含义，它如此广泛，以致如果有人试图用一种新的方法或程序看其效果是什么，我们亦可能称它是一项实验或一个实验过程。为了研制电灯泡，托马斯·爱迪生尝试了不同的灯丝材料以决定可以用哪种，他便是在做实验。

在教育研究中，我们用非常专门的描述来说明如何设计实验。然而也保留了广义上实验的基本概念，即尝试什么，控制一个或多个自变量以决定其效应，我们称所控制的自变量为“实验变量”。

实验是一种研究情境，在此情境中至少涉及一个自变量，即实验变量，它受到研究者的精心处理或控制。

广义而言，实验设计是为指导实验而预先策划的计划。狭义而言，实验设计是确定、安排在实验中建立某些变量所用的结构。设计包括自变量，其中需包含实验变量和可能存在的其他变量，如有机变量。既然要测量因变量（数据的分析），也就需标明在实验中的测量点。实验设计常以符号通过图表表示对这些变量的排列等等。这些图表和他们的排列、符号有一定的含义，这些将在本章中探讨。

我们来考虑一个例子。一位教育心理学家运用口头的、书面的、口头书面结合的3种形式来确定解决抽象数学问题的教学效果。实验参与者是大学新生，每个教学类型组都随机分配20名大学生。每一个体在学习实验室接受教学，并试着解决一系列数学问题。因变量是本系列中正确解决问题的得分。每一参与者单独接受实验，而非一组一组进行，整个活动约一小时。实验变量是教学类型，即变量的3种不同形式。

实验如图5.1所示，图的上半部分列出了整个实验。实验过程为实施适当

的教学并解决问题。实验期间收集对因变量的测量结果，收集齐全后常被记录下来。依据实验，数据可在实验进行时或结束后不久收集。本例的数据是教学后参与者正确解决问题的得分。

在图 5.1 虚线以下的部分，用符号表示实验。R 表示 3 个组的随机分配，由下标 G_s 表示。X 是用作介绍、使用实验变量的符号，其下标简单地表示变量的水平。图中用在 X 中的下标 V、W 和 C 表示教学类型——口头的、书面的和组合型。 O_s 表示收集数据（通常指进行的观察），该例中， O_s 的下标与组数一致。图中， O_1 表示从组 1 (G_1) 中收集的数据；该组进行的是口头教学 (X_v)。 O_s 是因变量的得分——这里指正确解题的得分。

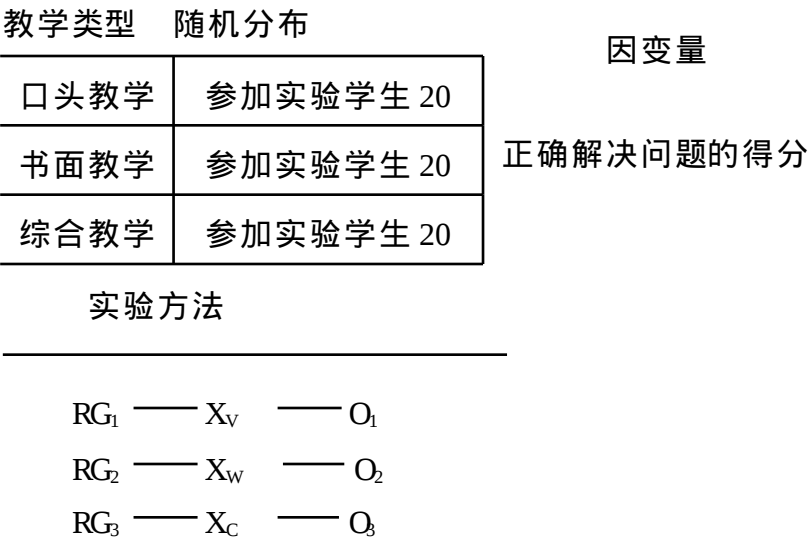


图 5.1 解决问题实验的实验设计

图 5.1 相对而言较简单，仅包括一个自变量——实验变量。数据仅在实验结束时收集。教学和解题的实验过程可能有些复杂，但设计为研究者提供了相当多的控制。分析得分时，如若研究者发现 G_3 的得分整体较 G_1 和 G_2 高得多（正确题解更多），我们便可以说教学的组合类型比其他两种类型在该类问题的解决中更有效。

为什么要增加实验设计的麻烦呢？因为它有益于研究者解释、理解实验所得数据。一般来说，实验设计的目的与研究设计的目的无甚差异：即使结果明了，并且实验设计可以加强控制。该例中，研究者之所以能得出不同教学类型的解题效果的结论，是因为设计使实验结构化，并将不同的效果显现出来，结果的类型与一定的自变量的效果有直接关系。

实验设计是实验中确定、排列变量所用的结构。

实验变量

前面讨论研究变量的类型时，一种主要的变量是自变量，它可能对因变量产生效应。自变量可采取多种形式，如有机变量、对研究的参与者实施的不同办法、程序之类的变量；实验变量是自变量但并非所有自变量均是实验变量。在前面的实验中，如果根据性别确定参与者，这样对 10 个男生和 10 个女生分别进行一种类型的教学。那么参与者的性别则确定为自变量。然而，这是一种有机变量，而非实验变量，教学类型才是实验变量。

可用作实验变量的自变量可多可少。从理论上说，实验变量可以有任意种有限水平。而在教育研究中，实验变量的水平相对比较少，比如在 2 到 5 个之间，很少超过 7 或 8 个。实验变量的水平有时被称为“实验处理”。下面举几个在教育研究中可能的实验变量的例子。

实验变量	可能的水平
1. 教学组织的类型	1. a. 组际教学 b. 自修教室
2. 获得概念所用教材类型	2. a. 口头 b. 图表
3. 实验用动物服用药量	3. a. 5 克 b. 10 克 c. 15 克
4. 任务持续进行所需时间	4. a. 5 分钟 b. 10 分钟 c. 20 分钟 d. 30 分钟
5. 历史教学的教学方法	5. a. 讲座 b. 小组讨论 c. 个体化教学
6. 一天中教学所用时段	6. a. 上午 b. 下午
7. 拼写教学中所用单词表长度	7. a. 10 个单词 b. 25 个单词 c. 40 个单词 d. 60 个单词 e. 100 个单词
8. 每天练习所用时间	8. a. 15 分钟 b. 30 分钟 c. 45 分钟 d. 60 分钟
9. 治疗类型	9. a. 只用药物治疗 b. 只锻炼 c. 混合使用药物治疗和锻炼

对于每一个实验变量来说，研究者可以调整、控制水平或进行实验处理。如第 7 个实验变量可以准备好单词表，提供给那些依据该表随机分成的 5 个组的学生使用。实验的目的无疑决定了单词表长度对拼写练习中学生表现的效应。

“被试”一词的使用

在一般的研究中，尤其在实验中，“被试”一词用来指参与实验的人。所以，教学一例中的被试是大学新生。如果实验使用 5 年级学生，他们就是被试。符号 S 表示被试。“被试”一词和 S 常在书面语中使用。

2 “好设计”实验的标准

介绍专门的实验设计之前，我们先来看一下“好设计”实验的准则。实质上，体现一个设计得好的研究的特征也应适用于实验设计。我们把这些标准列在下面，予以简单的解释。

1. 充分的实验控制——指对实验条件有足够的控制，以便研究者能解释结果。实验设计被高度结构化，如果实验变量产生效应，就可被观察到。亦指通过随机分组或把他们作为实验中的自变量来控制其他变量。

2. 不加入人为修饰——实验结果能否推广到非实验设定的情景，例如，一个教室，这一标准在教育研究中尤为重要。这即是说实验是在这样的背景下进行，以便其结论能应用于真正的教育领域。

3. 比较的基础——需要通过某种方式进行比较以确定是否有实验效果。在一些实验中，我们使用控制组——即不受到实验处理的组。教学实验中的控制组中常包含用传统方法施教的一组学生。在对动物进行的药物实验中，控制组应由那些没有服药的动物组成。当然，不是所有的实验都需要控制组。比较可以在两个或多个实验处理中进行，偶尔采用一些外在标准。

4. 通过数据获得充足的信息——数据必须足以检验实验假设，应有对假设作出判断的足够的精确度。

5. 非污染的数据——数据应充分反映实验效应。但不应被实验中出现的蹩脚的测量或误差所影响。来自不同组的个体不应相互影响而削弱实验效果，或导致实验效果不具代表性。

6. 无相关变量间的干扰——这一准则与充分的实验控制密切相关。可能存在其他正在操作的对因变量有影响的变量。如果这样，这些影响不应被错误地理解为实验效应。我们常常可通过实验设计分离、控制这些效应。

7. 代表性——研究者常希望将实验的结果推广到一些个体、条件、方法等中去。为了获得具有代表性的实验结果，研究者常采用某种随机方式，或者通过对实验被试的选择，或者通过对接受实验处理的被试的分配（如果使用控

制组，也运用上述方法的话)。

8. 省力原则——是指如果所有其他特征相同，我们更倾向于采用较简单的设计而非较复杂的。当然，一项设计必须为达到实验的目的尽可能地复杂，但复杂本身我们并不提倡。简单些的设计常常更易完成，解释起来也更方便些。

和其他任何类型的教育研究一样，实验易受技术上和方法上误差的影响。进行并充分完成一项合理的实验设计，需要缜密、仔细的计划，这样方能对误差进行最好的控制。实验设计同时还需要注意一些细节，计划应先于实验。

实验效度

良好设计的标准可概括为增加实验效度这一点。在第1章，我们讨论了教育研究的效度，实验效度的概念在本质与其是相同的。实验效度分为两类：内在效度和外在效度。

内在效度是指研究结果能被明确解释的程度。内在效度用于理解数据并从中得出结论，对实验处理是否能真正与因变量区分开来提出疑问。为回答该问题，研究者需确信能控制诸如无关变量并不会将产生的效应错误地作为实验处理效应。外在效度是指涉及实验结果的可推广度。实验可以推广到怎样程度的总体、变量、情境等。一般来说，内在效度越充分，结果推广度越大，研究越有用。

实验效度有两种类型，内在效度和外在效度。内在效度涉及解释结果所需的最小控制等，外在效度涉及结果的可推广的程度。

尽管实验设计的目的是希望实验的两种效度都高，但有些情况下，确保一种效度，就会削弱另一种效度。随着实验的控制越来越严格，在实验中可发生的和在自然教育条件下可发生的两者间的一致性就越小。例如，在教学技术的研究中，对实验的控制达到如此程度，以至实质上是人造的情境，只是实验变量在运转。这样大大提高了内在效度，但推广性就可能非常有限以至于结果无法在一个真正的教室情境下运用。这并不是说获得最大的控制就一定是不好的，实验的目的决定要求效度的程度。显然，一个结果无法解释的实验毫无用处，即使推广度可能很大。另一方面，实验后又发现其结果不能如实验目标中预期的那样得以推广，这样也不会令人满意。

内在效度涉及对无关变量的选择方法、测量方法等进行充分的控制。实验设计应能够发展，以便能使研究者充分地发现可能威胁内在效度的因素。毋庸

置疑，我们不可能在所有的实验中操作所有可能的因素，但研究者应了解有关变量以及与内在效度相关的可能出现的困难。这样，实验设计便能够充分地解释结果。

至于外在效度，它关心的是研究者所期望的结果能推广到的总体，但也可能包含对其他相关自变量的发现或对实验变量的修正。可能会有诸如班级大小、学校类型等研究者所希望推广的因素。例如，一项在郊区学校4年级学生中获得的实验结果是否对城区学校也适用？对初二年级呢？很可能不行，但这又需依据变量和实验的具体细节来判断。研究者也可能想推广到不同的测量差异的情境中去。例如，一个经过了事前测试的实验的结果，能够推广到没有事前测试的班级吗？外在效度就是考虑这类问题。

实验效度的干扰因素

实验设计应致力于提高实验效度，但实验效度并不仅仅依赖实验设计。实验细节对效度有影响，此外还有许多其他东西会干扰实验的效度，包括对内在的和外在的效度的影响。实质上，困难就在于确定这些因素是否对实验的效度有影响。对效度的威胁来自于对数据的模棱两可的解释。

坎贝尔和斯坦尼（1963）总结了对实验效度的影响因素，共有12个，影响内在效度的有8个，影响外在效度的有4个。表5.1列出并解释了这12个因素，同时举例说明了每一个因素是怎样发生的。所有的例子都假定已作了一些实验处理。

表 5.1 实验效度的干扰因素

干扰因素	举例
内在效度	
1．偶然事件——在实验进展过程中没有预料到的影响因变量的事件的发生	1．在相对较短的教学实验中，一组被试因学校停电而不能对其进行教学
2．成熟程度——时间在被试身上起的作用	2．在学习实验中，被试 50 分钟后因疲劳成绩下降
3．测验——前一次测验对随后另一次测验的影响	3．在一次以逻辑推理能力为因变量的实验中前测给被试者提供了有关后测的线索
4．测量手段——测量手段不统一会产生错误的结果	4．两个主考人对同一项数学实验进行后测所用的程序和方法不同
5．统计回归——挑选被试的误差，比如用极端分数进行回归，将对今后的测验产生不利的影响	5．在一项阅读教学的实验中，前测中阅读成绩差的组，较之于成绩一般、好的组进步大得多

续表

干扰因素	举例
6. 被试的选择差异——被试未能随机分配或挑选，而其中一个因素起了作用从而产生了组的不对等性	6. 一个教学实验的实验组本来就是一个高材生班，而控制组则本来是一个普通水平班
7. 被试的流失——非随机挑选的被试脱离实验，会产生不良影响	7. 在一项判断各类运动效果的健康实验中部分被试发现此项运动很难，便退出
8. 取样—成熟程度交互作用——由于取样不一带来的成熟程度的不一致	8. 在一项问题的实验中，选取初、高中教学班为被试，初中生比高中生更早地感到疲劳
外在效度	
1. 测验的交互作用——前测与实验处理发生作用，并导致结果不能推广到未经过前测的群体中	1. 一项体育表演实验中，前测为被试对实验处理提供了以某种方式作出反应的暗示
2. 抽样偏差和实验处理的交互作用——根据实验处理的需要而挑选现成班级进行实验，其结果不能推广	2. 进行对教学方法做实验处理的实验，选择低能力班级进行实验，其结果无法在能力各异的班上推广
3. 实验安排的副效应——被试知道他们要参加实验并对其具有新鲜感，也称为霍桑效应	3. 一项实习性的阅读教学实验产生的效果不会再以同样内容同样方式在正常补习阅读教学中产生
4. 多重处理干扰——一个被试受两项或两项以上的处理，（就像测量中设计的那样）会产生一种后遗症，导致不能推广到单独处理中	4. 在药物实验中，对同一动物依次给予4种不同的药物处理，从第2次到第4次服药的效果都摆脱不了服药可能有的长效影响

面对这些可能存在的不利影响，人们可能怀疑能否解释和推广这些实验结果。但这正好是良好研究设计的准则和实验的精心之所在。我们已讨论了通过一些方式加以控制，例如，随机抽样照顾到可能会出现的抽样偏差问题。在进行实验处理之前，随机分配提高了各组的对等性。在多数情况下无关变量能够通过诸如因素设计被控制为自变量。

本章中讨论的实验设计是一些一般性的设计，他们只适合于各自特定的实验情境，正因为如此，我们不应认为随便一个实验设计就能有效地防止各种因素对实验效度的干扰。如果我们对所应用的实验设计不加区别，测量方法本身就有问题，那么测量工具对效度就是一种威胁。

测验对那些仅含有一种收集资料的方法如仅施后测设计的实验来说，并不会构成威胁。在本章讨论的各种研究设计中涉及到了坎贝尔和斯坦尼（1963）提到的“真”实验的标准问题，也就是在实验处理中被试要进行随机的选择和安排。这样对被试的不同选择于实验研究设计来说并不是一种干扰。相反，它对准实验研究设计来说还是相当适合的，这在第6章中会详细论及。在进行实

验处理前的随机分配提高了各组的对等性。如果未经随机选择，就有可能出现极端数值，那么统计回归的运用就不算是一种干扰。但是当我们运用这些被试进行实验时，我们应能够识别并指出其对实验可能的威胁。例如，当统计回归的效果与实验效果相混淆时，就有必要从研究文献中估计统计回归的影响程度。

一个实验的周期越短，无关事件对其效度的干扰也就越小，对测量作仔细的分类和控制，可有效地减少测量手段带来的问题。如果实验被试没有离开，那么就不存在实验难以为继的问题。若在实验处理结束后仅施后测，那么在实验处理和测试之间出现两者交互作用的可能就不会太大，所以要通过计划和仔细地进行实验来减少对效度的干扰，许多无关变量可通过研究设计而成为控制变量，当然消除所有的干扰是不可能的，重要的是认识到这一点，并对结果作出相应的解释。

教育实验设计很少是尽善尽美的。通过实验设计，研究者试图获得充分的内在效度和外在效度。由于一种效度的提高可能会损及另一种，研究者必须设法保持其平衡，通过足够的控制而使结果具有可解释性，通过获得足够的事实从而使结果能够推广到意欲推广的情境中去。

实验效度在每一个具体的实验中都必须予以考虑。效度不会全有或全无。应该注意对效度的可能的威胁，通过实验设计和实验方式来消除这种干扰。

下面几部分将描述在教育研究中常用的一些很专门的设计。这些设计都以图表的形式进行了总体勾勒，并为每一个讨论的设计提供了一个例子。

3 仅施后测控制组设计

在讨论实验设计时有两个术语，前测和后测，常常在与收集数据有关的研究中用到。前测是指在实验处理之前对被试进行的测量或测验，后测是指在实验处理后进行的测量或测试。前测及后测要求一一确定被试，这样才能使前测和后测得分能够配对。不是所有的设计均需前测，但后测是决定实验处理的效果所必须的。

实验设计一般包含两个或更多的组。仅施后测控制组设计最简单的形式包含两个组：接受实验处理的组和控制组。被试在实验前被随机分为两组，实验组接受实验处理。在实验结束时，测量两个组并以此作为研究的因变量。这种测量最好在实验一结束就做，特别是当因变量可能随时间而改变时更需

如此。

仅施后测控制组设计对教育工作者来说是有设计，它不需要在很多情况下并不理想亦不实用的事前测试。被试只需根据所在组确定，其他自变量也可能在设计之中。两个组的设计可以表示如下：

$$\begin{array}{ccc} RG_1 & X & O_1 \\ RG_2 & \text{—} & O_2 \end{array}$$

图中，G_s 表示组，R 表示随机抽取或分至各组的人数。X 表示实验处理，短线“—”表示无实验处理。O_s 表示对相关变量的测量（测验、任务或观察），竖直放置的 O_s 表示何时进行。既然纵向上成一直线，它们便是发生在实验中相同的点上。这里，由于发生在实验处理之后，所以是后测。

仅施后测控制组设计还可以扩展到不只包含两个组，也就是说可以对两个或更多的组进行实验处理，增加组数至 3 个或更多。从总体中抽取的被试应随机分组，各种不同实验处理的效应应通过比较组间的行为表现来研究。从更广泛的意义上说，仅施后测控制组可图示如下：X_s 的下标标明了不同的实验处理。在该实验中处理次数为 K。实验中有一组，即控制组未受到实验处理。如果无需控制组，该设计为仅施后测随机组设计。

$$\begin{array}{ccc} RG_1 & X_1 & O_1 \\ RG_2 & X_2 & O_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ RG_k & X_k & O_k \\ RG_{k+1} & \text{—} & O_{k+1} \end{array}$$

下面举例说明仅施后测设计

例 5.1

一名 4 年级教师做一个有关辅助教学材料对阅读行为影响的实验。使用两种辅助教学材料，以及传统材料，他们构成 3 种独立（实验）变量水平。研究问题陈述如下：

一项有关不同类型的辅助教材对 4 年级学生阅读行为影响的研究。

4 年级的每 15 名学生被随机地分布在各个不同的自变量上。这样，有 45 名学生参与该实验。在每天的阅读教学中，每一组中使用辅助教材的 15 名学生花 20 分钟读这些教材。使用传统教材的组为控制组。8 周教学之后对学生进行阅读测试，这次测验的成绩是因变量。实验如图 5.2 所示。实验处理用 X₁ 和 X₂ 表示，短线“—”表示控制组。根据总体设计，本例中实验处理量 k = 2。

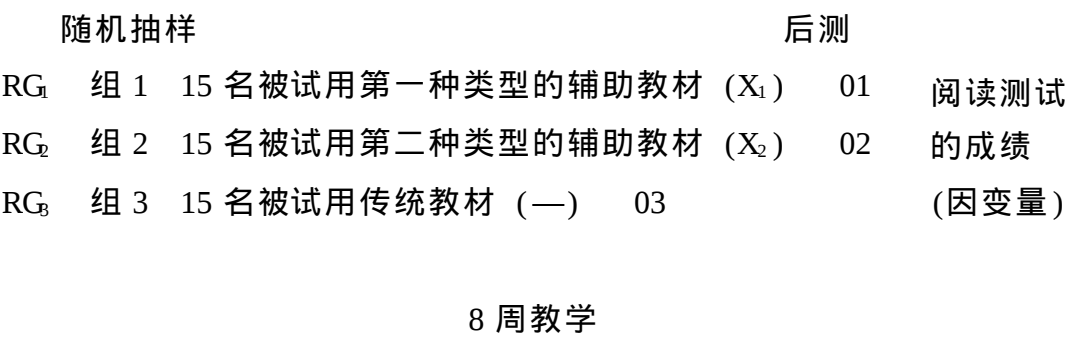


图 5.2 例 5.1 仅施后测控制组设计图，包含两个实验组和一个控制组

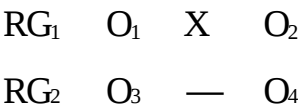
仅施后测控制组设计包括进行实验处理的所有组，以及一个控制组或比较组。被试只是在实验处理之后才接受测验。

4 前测—后测控制组设计

在实施实验处理之前加上测量的话，仅施后测控制组设计就被扩展为前测—后测控制组设计了。被试被随机地分配到两个或更多的组中去，先前只对估计与实验中假定的变量有关的被试进行测试，这可能是测量因变量的第 2 种测验形式。

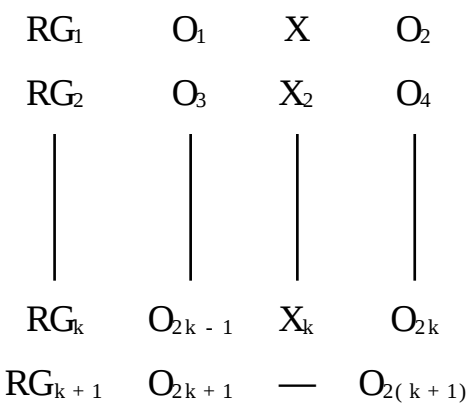
前测是要得出什么？前测得分可作为分析中的一项统计控制量。在一些实验中，尤其是在教学领域，要分析得分——后测与前测间得分的差异。

从最简单的形式来看，前测—后测控制组设计包含两个组——一组接受实验处理，另一组则不。图示如下：



现在仅施后测设计中的 O_s 增加了一倍，所以用奇数作下标的 O_s 表示前测，用偶数作下标的 O_s 表示后测。

前测—后测控制组设计可扩展到包含两个以上的组。总体上图示如下：



上述符号表示有 k 个接受实验处理的组和作为控制组的一个比较组。如果使用

两个或更多的实验处理，而不需要控制组，我们称该设计为前测—后测随机分组设计。

下面举例说明

例 5.2

一名研究者对几何概念的教学所用时间的长短对高中低年级学生有关空间关系的测验成绩的影响感兴趣。研究问题陈述如下：

有关立体几何概念教学的时间量对高中低年级学生在空间关系知识方面得分影响的研究

这一问题的陈述显然包含了一个实验的研究，因为教学时间量是研究者控制的变量。

我们采用两种大体相近的空间关系测验，一个作为前测，另一个作为后测。实验设置如下：随机抽取 40 名高中低年级学生，前提是所有人均在学几何课程。40 名学生分成 4 组，每组 10 名。实验处理实施如下：

- G₁ 接受 15 分钟为一单元的立体几何概念教学；
- G₂ 接受 2 个 15 分钟为一单元的立体几何概念教学；
- G₃ 接受 3 个 15 分钟为一单元的立体几何概念教学；
- G₄ (控制组)不接受立体几何概念教学。

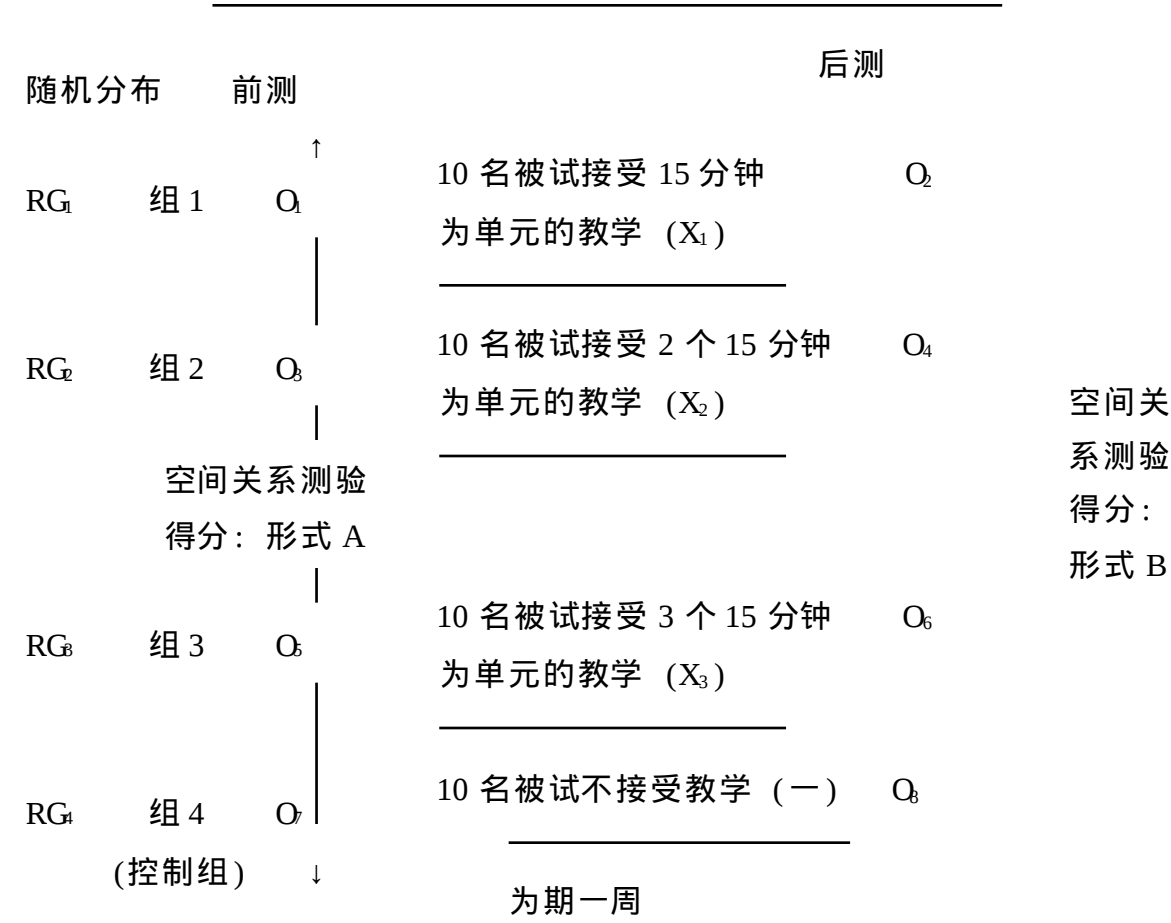


图 5.3 例 5.2 前测—后测控制组设计图，包含 3 个实验组和一个控制组

G_1 , G_2 , G_3 的成员在一导师辅导的情境下个别地接受教学 (必须注意, 对各组每一成员的教学应是一致的; 需要改变的仅是教学时间长短), 整个教学要花一周多的时间。教学前, 先对学生进行前测, 教学结束后不久, 对学生后测。实验如图 5.3 所示。

实验中因变量是前测与后测之间的得分。或者后测得分是因变量, 是对前测得分的调整。调整后得分应由统计方法产生。运用这一设计, 研究者可以确定, 对立体几何概念教学时间的不同, 与控制组相比, 是否影响空间关系测验的得分, 或者教学几乎没有影响。

前测—后测控制组设计包含的组数, 等于实验处理组数加上控制组。被试在实验处理前后都要接受测量。

5

所罗门四组设计

以最简单的形式把前测—后测控制组设计和仅施后测控制组设计组合起来可以得到一种新的设计, 这是所罗门 (Solomon, 1949) 提出来的, 该设计有 4 个组, 包含控制组、实验组各两组, 而实验组接受的实验处理是相同的。实验组和控制组中各有一组接受前测, 实验结束对 4 组都进行后测。每组被试都是随机分布的。

所罗门 4 组设计的图示如下:

RG_1	O_1	X	O_2
RG_2	O_3	—	O_4
RG_3	—	X	O_5
RG_4	—	—	O_6

因为是 4 组设计, 实验包含 4 个组, 只有一种实验处理。比较实验组和控制组后测得分可确定实验效果。只有一种实验处理, X 就无下标。组 1 和组 3 是实验组, 组 2 和组 4 是控制组 (图中没有 X)。

因为对一些组进行了前测而一些组不测, 所罗门 4 组设计的优点在于能使研究者检验前测带来的可能的效应。前测影响后测得分, 前测与实验处理相互作用都是可能的。也就是说, 实验处理的效应对前测过和没有前测过的组是不同的。因为在实际的课堂实践中前测不是规定要做的, 所以它经常是效度所要考虑的问题。下面举例说明所罗门 4 组实验设计。

例 5.3

教育心理学家在做有关观看解决问题的电影对年轻人逻辑推理测验成绩的影响的研究。研究问题陈述如下：

关于观看一场解决问题的电影对年轻人逻辑推理行为影响的研究

实验处理是看一场 30 分钟的电影。实验被试是在教育心理学班级注册的大学高年级学生。心理学家想进行前测和后测，这样至少可获得一些用于分析的分数。但还需考虑是，前测可能引起对电影一定的反应，作了前测的被试可能受到从电影中学点什么的暗示。这就是为什么要用所罗门 4 组设计而不是仅施后测设计，也不是前测—后测设计的缘故。

32 名学生分成 4 组，每组随机分得 8 人。设计见图 5.4G₁ 和 G₂ 组中的 16 名同学同时进行前测，第二天，实验组的 16 名学生看电影。之后，对所有 32 名学生进行后测。前测和后测平行进行但参加的是不同的逻辑推理测验。

心理学家现在能检验前测带来的可能效应和前测与实验处理相互间的作用。这正体现了所罗门设计较仅施后测和前测—后测控制组略胜一筹之处。然而，所罗门 4 组设计需要更多的组，相应地，需要更多的被试，这一点并不有利。设计可以扩展到包含更多的实验处理，但每增加一次处理，就另外需要增加两组——对一组进行前测，另一组则不测。如果实验变量包括两次处理，则需要 6 个组；如果有三次处理，则需 8 个组。

随机分组		前测		后测	
RG ₁	组 1	O ₁	8S _s 看电影 (X)	O ₂	
	(实验组 - 前测)	逻辑推理			
RG ₂	组 2	测验得分：O ₃	8S _s 不看电影 (—)	O ₄	
	(控制组 - 前测)	形式 A			逻辑推理
RG ₃	组 3		8S _s 看电影 (X)	O ₅	测验得分：
	(实验组 - 不前测)				形式 B
RG ₄	组 4		8S _s 不看电影 (—)	O ₆	
	(控制组 - 不前测)				

为期一天

图 5.4 例 5.3 所罗门 4 组设计，包含 2 个实验组和 2 个控制组

所罗门 4 组设计是仅施后测控制组设计和前测—后测控制组设计的组合。

6 因素设计

所罗门 4 组设计实际上属于因素设计的范围。

因素设计包括一系列广泛运用于教育研究的设计方式。因素设计的基本构造在于，每一自变量的所有水平与其他自变量的水平结合使用（技术上称全因素）。设计至少需要两个自变量，每一变量至少有两种层次。这种最低限度的因素设计称为一个二乘二（ 2×2 ）因素设计。从理论上讲，可以有任意数目的自变量，而且每一自变量可以有任意数目的水平。用数字来命名设计，如 2×4 ，阿拉伯数字的个数表示自变量数目，数字的值表示自变量的水平。自变量数目不必与其水平完全相同。一个（ $2 \times 3 \times 5$ ）因素有 3 个自变量，相应的变量水平为 2, 3, 5。符合该因素的例子就是：两种教学方法、3 级能力水平和 5 个年级。

在因素设计中，随着自变量和水平的增加，分组的数目也迅速增加。（ 2×2 ）因素有 4 个组。但若增加一个具有两个水平的自变量，就会使组数增加到（ $2 \times 2 \times 2$ ）个即 8 个（以 2^3 表示）。如果一个 2×2 因素，每个变量增加一个水平，则成了 3×3 因素，使组数增至 9 个。由于水平必须包含在所有的组合中，故而组的数目就是指定的因素设计中所有阿拉伯数字的连乘积。

在一项实验中，至少有一个自变量是实验变量。其余的自变量可能是有机变量，设计中加进有机变量的目的是为了便于控制。根据单元个数的变化，因素会迅速变得复杂起来。这里的单元是指变量水平的不同组合产生的组。为简化因素的图示，变量以字母表示，不同的层次则以数字表示。字母的下标表示不同单元。

与较简单的设计相比，因素设计的优点一般表现在两方面：因素设计只需一个设计，经济方便，并非要对每一自变量都提供不同的设计，且可以研究变量间的交互作用。就许多研究而言，研究变量间的交互作用非常重要。探讨交互作用的存在与否是研究的重要目的。

交互作用是对因变量的影响，这样自变量的影响便不能与另一水平保持一致。当两个自变量的联合影响与它们分开的（增加的）不等时，便出现了交互作用。这就意味着一自变量本身的影响与它和其他自变量水平混合时是不同

的。不同能力等级的学生从不同教学内容中受益不等，便是交互作用的实例。能力水平和教学内容为自变量。

交互作用最简单的类型是两个变量间的交互作用。有时我们称其为一次交互。交互作用中可能包括两个或两个以上的自变量。然而，自变量越多，交互作用越复杂，解释起来也就越困难。

实验中的交互作用是对因变量的影响，比如，一个自变量的变化对另一自变量的水平发生影响。

因素设计可以运用于不止一个自变量的调节和控制。由此，它常常被用来作为加强控制的设计，其中以相关因素作为自变量。从理论上说，因素设计可以扩展到包含任何有限数量的变量和水平。但是复杂的设计需要认真考虑，因为这样一项设计就其可提供的被试而言，并非经济可行。从实际看来，也可能无法解释涉及两个以上自变量的复杂的交互作用。例如，涉及 4 个自变量的交互作用——能力水平、被测的性别、教学方法以及教材类型时，结果就很难解释。

因素设计包含两个或两个以上的自变量，在单一设计中被称为因素。设计的单元取决于进行组合的自变量的水平。请看下面的例子：

例 5.4

一名教师想了解两种不同的美国历史教材的教学效果。一种图文并茂，另一种抽象且用详细的文字描述。参加实验的学生能力各异，且对学生进行了知识倾向测试。知识倾向作为控制变量；不同类型的教材对不同能力水平的效果也可能不一致。也就是说，教材类型和能力水平可能产生交互作用。所研究的问题可陈述如下，设计如图 5.5 所示。

		教材 (M)		事后测验
		M ₁	M ₂	
能力水平 (A)	A ₁	高能力被试者 20 名	高能力被试者 20 名	历史考分 (因变量)
	A ₂	中能力被试者 20 名	中能力被试者 20 名	
	A ₃	低能力被试者 20 名	低能力被试者 20 名	

——教学时间：一个学期——

图 5.5 例 5.4 2 × 3 因素设计

一项有关使用不同表现形式的教材对高、中、低能力水平的高中生美国历史学习成绩产生的效果的研究。

120 名学生根据能力随意分成三组：高、中、低。获得最高知识倾向测验分数的 40 名学生在“高”的组，其次的 40 名在“中等”组，其余的 40 名在“低”组。再从每一能力水平中抽取 20 名学生，随机指定使用两种教材中的一种。教学为期一学期。因变量是涵盖教材内容的一次普通历史测验的成绩。

图中因素设计的方案与前面的略有不同。组分布在不同的几个单位中，而非仅仅在几行上，自变量的分布亦是任意的。该例中，教材类型是实验处理，用 M（指材料）取代 X，所有 6 个组均在学期教学要结束时进行测试，O_s 不在图表中。

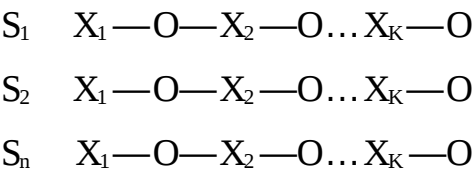
该例中，随机分配仅用于给学生分配不同的教学方式，显然，不能对学生的能力水平随机分组，因为这是个有机变量；能力水平也是主观确定的，这样 40 名学生在各部分都有分布。我们并不要求因素设计中的单元有相同的数目，但往往相同，因为通常说来，如果各单元所含数目相同，对数据的分析、结果的解释则更加简单。

7

重复测量设计

在一些实验中，被试要接受多次相同的测量。例如在学习实验中，被试常常完成一系列任务，如解决一系列问题，以了解学习是否已经发生了。重复测量也有前面已提到的设计的特点，例如，任何实验处理实施前都要进行前测。更为普遍的是，重复测量仅被当成后测。至于在一个因素设计中，其他的自变量可以在其中。实际上，许多重复测量设计是因素设计，外加重复测量的特点。

重复设计最简单的形式是对所有的被试进行实验处理。如果有 k 个实验处理和 n 个被试，则设计可如图表示如下：



S 表示特定的被试，所有被试接受相同方法的实验处理。O 在图中表示每一实验处理之后都对因变量进行的测量。请看下面的例子：

例 5.5

这里的例子比起最简单的设计可能要稍微复杂一些，其中包含除了重复测

量的实验变量外的两个自变量，这是正进行一项用大学新生作被试的学习实验，来解决一系列所谓 4 组概念获得的问题。被试在一学习实验室里独立解决问题。问题可用两种材料：图表的和文字的，这是一个自变量。有机变量为被试的性别，是作为另一自变量包括在其中。因变量是解决问题的得分的效果，由实验者确定。研究问题和相关假设陈述如下：

- 一项关于材料类型和性别对解决概念获得问题的影响的研究。
- H₁：使用图表教材的被试的平均成绩比那些使用文字教材的要好。
- H₂：男女性别对平均成绩无差异。
- H₃：被试的平均成绩随 4 项问题解决而增加。

实验中共有 40 名被试，男女生各 20 名，设计如图 5.6 把他们随机分在不同的材料组中，20 名用图表的，20 名用文字的。自变量——被试的性别和材料类型——与一个（2×2）因素设计相似，另外，有 4 种水平的实验变量问题，这样我们可以称之为一次重复测量在 4 种水平实验变量上的（2×2×4）因素设计，设计如图 5.6。

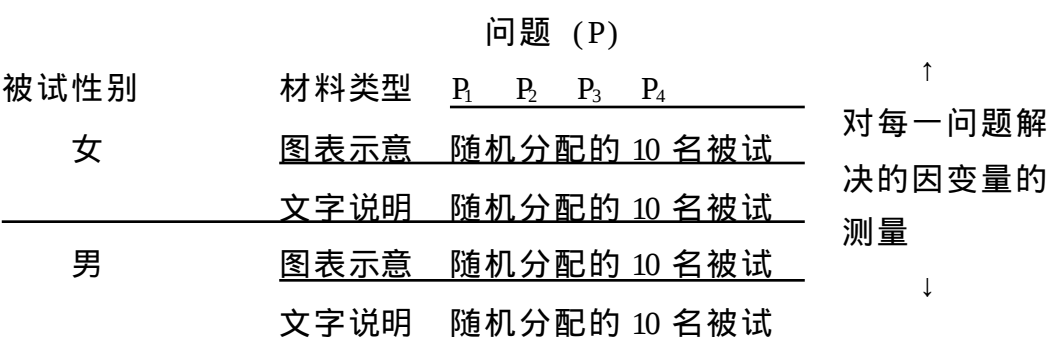


图 5.6 重复测量设计图

假定在这一例子中，用来解决问题的时间是因变量，那么时间越短，则成绩越优秀。在分析数据时，我们可以比较一下完成图表材料与文字材料的时间均值，因为共有 40 名被试，每人解决 4 个问题，因此共有 160 个因变量得分，80 个为图表材料分，另 80 个为文字材料得分，这样每组材料的均值就可算出来。同理，我们也可以通过比较男、女生的均值来判断性别对被试成绩的影响，因为男女被试各 20 名，每一类均值都可在各自 80 个观察值的基础上获得。

从上例中，要得出 4 个问题的平均值，每一个问题都有 40 个分值，每一被试都有一个分值。如果解决问题的平均时间从第 1 到第 4 呈下降趋势，那么就可证明学习已经发生，用重复测量设计获得的数据就可解释以上 3 种假设。

重复测量设计的一个特点是对被试测试产生大量的数据。该例中，虽然只有 40 名被试，但却有 160 个观察值或得分。重复测量设计还有一个特点就是所有观察结果并不是独立的。对不同被试的观察结果是互相独立的，但是对于同一个被试者的几个观察结果则是相互联系的，这是因为他们来自同一名被试。在分析数据时，必须考虑这一点。

重复测量设计是一种对有关因变量对同一被试进行一次以上测量的设计。

重复测量设计的缺点在于多次重复测量容易使人们熟悉这一特定的特征，在上例中，4 个问题被所有被试以同样的序列解决，如果解决问题的成绩有所提高，也可归因于问题变得相对简单。解决这一问题的途径是均衡设计，即打乱问题的排列顺序，对实验处理进行平衡。

8 时间延长设计

有时，在研究教育变量和行为科学的其他变量时，把握显示变量影响需要的时间可能是困难的。对某些变量，影响的持续时间是未知的。实验的设计可扩展为核实可能的后滞影响或核实影响持续的时间。这些可以通过额外的观察，即时间上延长来实现。如果我们以这样的方式扩展仅施后测控制组设计，则可能的设计图是这样的：

$$\begin{array}{ccccccc} RG_1 & X_1 & O_1 & - & O_2 & - & O_3 \\ RG_2 & - & O_4 & - & O_5 & - & O_6 \end{array}$$

在该设计中，对 G_1 进行了实验处理之后，观察两个组。然后，虽然没有进行额外的实验处理，但对两组进行规定的、固定的、连续的观察。时间间隔的长短依研究的变量而定。

这样延长了的设计对多重观察的影响就容易看出来来了，因为对组的测量不止一次。早先的观察可能对随后的观察产生影响。是否存在多重观察影响依所涉及的变量而定。另外，由于设计在时间上延伸了，它们特别易受对内在效度产生干扰的偶然事件和成熟因素的影响。

设计可以通过对实验组进行额外的观察得以延长。这种观察提供了有关实验变量可能的延迟影响和影响持续的信息。

请看下面的例子：

例 5.6

研究者正在研究 3 个不同的锻炼项目对男性成人锻炼后心率的影响。在一个简要而剧烈的练习后测量一分钟。该研究问题可陈述如下：

关于男性成人三种项目的锻炼对心率的影响以及影响持续时间的研究。

随机抽取 60 名 18 ~ 22 岁的男性，每 20 个人被分配到 1 种锻炼小组中去。被试在他们各自的锻炼项目中每天参加锻炼，持续 2 个星期。在 2 周的最后一天，测量他们的心率。收集心率数据时，被试先剧烈锻炼 5 分钟然后休息一分钟，接着测量心率。两星期的锻炼后，被试不再参加锻炼项目并被告诫不去独自参加锻炼。在锻炼项目结束后间隔 3 个星期，再次测量他们的心率。设计见图示 5.7。

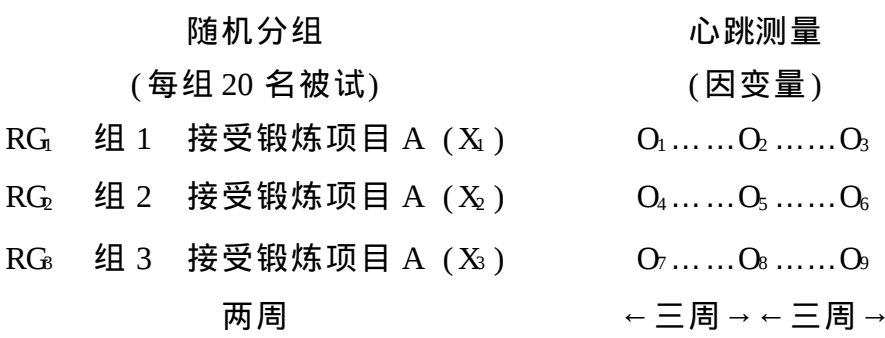


图 5.7 例 5.6 时间延长设计图

该例中似乎没有多重观察的影响。三周前的简单测量的心率并不对其后测量的心率产生影响。如果被试在锻炼项目结束后自己独自参加锻炼，将对内在效度产生干扰。在进行实验之前，为获得被试的意见，他们将按指示进行 8 个星期的实验，包括观察，这将是值得的也是必需的。

时间系列设计

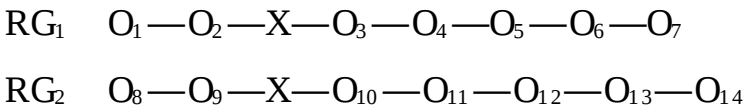
时间系列设计包含有重复测量，并且在两次测量之间插入一次实验处理。请看下面的例子：

例 5.7

一名大学教授给有 78 名学生的一个班教授“教育导论”课。这些学生被分为两个讨论组，在为期一学期的这门课上进行每周两小时的讨论，小组讨论是从第 3 周开始，此后每隔两周对学生进行一次前两周内容的测试，在整个学

期中共进行 7 次这样的测验，研究生被穿插在两组中参加讨论。

78 名学生被随机分为两组，每组 39 人，一组按传统的方法进行讨论，即只讨论已学习过的内容和可能提到的教育问题，这组为控制组。实验组除接受传统的方法外，在第 2 次和第 3 次考试间的讨论课上听取和讨论有关学校形势的录音，录音机的使用就是实验处理，这一设计可图示如下：



在上图中，两两相对的 O_s 表示测验同时进行，如 O₁ 和 O₈ 是相同的测试，字母的下标不同，表示测试来自不同的组别，O_s 间的水平线表示间隔为两周。

在这一例子中，人们的兴趣可能关注于 O₃ 和 O₁₀ 的不同，然而实验处理的迟滞影响也可能会延缓出现。O₁ 和 O₈，O₂ 和 O₉ 是在实验处理前测得的，且两组是随机分组，所以差别不应太大。

时间系列设计可扩展到多个组中，同时实验处理也可以多次插入，然而随着设计的扩展，偶然事件和成熟度对效度干扰的可能性也相应增加。时间系列设计在下一章中也将论及，因为他们在原始群组中也是适用的，而单个被试设计可看作是时间系列设计对个体的应用。

9

对实验结果的解释

如果一项实验有恰当的设计，研究就能得到有关实验的结果和结果性质的结论。然而，在解释结果时，我们要考虑数据（设计中的 O_s）的一些格局以及这些格局所表示的是怎样的实验处理。

对实验结果的解释可以通过两种方法来进行：一、给出某一实验结果（或 O_s 的一种格局），它的意义是什么？二、什么可以确定是否存在某种影响？下面将举出设计实例并作出简短解释。所用符号与前文一致。对观测结果 O_s 的解释，主要借助于对各观测结果的比较。虽然 O_s 包括各种观察，如前测或后测得分，在设计中也有通过计算平均值或某些组的测量总结得出的数据。符号“=”意为那些组的测量近于相等；“≠”意为本质上的不同（本质上的差异很可能通过推断统计得到确认）。在这些例子中，a，b，……等一系列结果是分别独立予以考虑的。

释例 5.1

一个 4 组前测—后测控制设计，它采用了 3 个不同的实验处理 X₁、X₂ 和 X₃ 以及一个控制处理。对所有组进行了前测和后测。设计图示如下：

RG₁ O₁—X₁—O₂

RG₂ O₃—X₂—O₄

RG₃ O₅—X₃—O₆

RG₄ O₇ O₈

a . 结果: $O_1 \neq O_2$, $O_3 \neq O_4$, $O_5 \neq O_6$, $O_2 = O_4$, 但 O_2 、 $O_4 \neq O_6$ 和 $O_1 = O_3 = O_5 = O_7 = O_8$ 。

解释: 所有实验处理都有影响。X₁ 和 X₂ 的影响相同, 但与 X₃ 不同。

b . 结果: $O_1 = O_3 = O_4 = O_5 = O_6 = O_7 = O_8$, 但 $O_1 \neq O_2$ 。

解释: X₂ 和 X₃ 不产生影响。仅 X₁ 产生影响。

c . 结果: $O_1 = O_3 = O_5 = O_7$ 和 $O_2 = O_4 = O_6 = O_8$, 但 O_1 、 O_3 、 O_5 、 $O_7 \neq O_2$ 、 O_4 、 O_6 、 O_8 。

解释: 没有实验处理影响, 但某些变化可能是成熟影响导致的前测与后测间的变化。

d . 应作怎样的比较来确定独立于实验的被试是否有变化?

比较: O₇ 和 O₈, 若 $O_7 = O_8$, 则没有变化; 若 $O_7 \neq O_8$, 则有变化。(比较也可在 O₁、O₃、O₅ 和 O₈ 间进行, 因对被试进行随机分组, 故而 O₁、O₃、O₅ 和 O₈ 被认为是对等的。)

应该说明的是, 没有办法对该设计中前测的影响进行检验, 因为是对所有组进行前测, 而非前测比较组。如果有两个或更多的实验处理 (该例中为 3), 区别他们也很重要, 因为他们可能有不同的影响。该例中结果不会穷尽所有的格局, 但他们描述了可能的结果。

释例 5.2

研究者使用 3 组设计。包括两个不同的实验处理 X₁ 和 X₂ 以及控制处理。对各组不进行前测, 但对其进行两次后测, 一次在实验处理完成后不久, 另一次用来确定是否存在延迟的影响。设计如下:

RG₁ X₁—O₁— —O₂

RG₂ X₂—O₃— —O₄

RG₃ —O₅— —O₆

a . 结果: $O_1 = O_3$, 但 O_3 、 $O_1 \neq O_5$, 而 $O_2 = O_4 = O_6$ 。

解释: 立即有实验影响。实验处理 X₁ 与 X₂ 无差别。不存在长期影响。

b . 结果: $O_1 \neq O_3$, O_1 、 $O_3 \neq O_5$, 且 $O_2 \neq O_4$, O_2 、 $O_4 \neq O_6$, 但 $O_1 = O_2$, $O_3 = O_4$, $O_5 = O_6$ 。

解释: 实验既有短期影响也有长期影响。实验处理 X₁ 和 X₂ 影响有差别。影响有持续性, 即短期影响与长期影响无差别。

c . 结果: $O_1 = O_3 = O_5 = O_6$, 但 $O_2、O_4 \neq O_6$, 且 $O_2 \neq O_4$ 。

解释: 无短期实验影响。实验处理 X_1 和 X_2 的长期影响有差别。

d . 结果: $O_1 = O_3$ 但 $O_1、O_3 \neq O_5$, $O_1、O_3 \neq O_4$, 但 $O_2 = O_4$; $O_1、O_4 \neq O_6$ 。

解释: 有短期实验影响。而且 X_1 和 X_2 的短期实验影响相同。也有长期影响, 而且 X_1 和 X_2 的长期影响也是相同的。但长期影响与短期影响有差别。

e . 作什么比较才能断定是否存在长期影响, 并断定 X_1 和 X_2 的长期影响是否一致?

比较: 要有 O_2, O_4 和 O_6 间的比较。如果 $O_2、O_4 \neq O_6$, 说明有长期实验效果。然而如果 $O_2 \neq O_4$, 就说明 X_1 和 X_2 的效果是不同的。

f . 作出何种比较可以确定时间的推移是否影响被试中的因变量。

比较: 要有 O_5 和 O_6 之间的比较。这些观测结果都是控制组方面的, 是没有接受实验处理的。如果 $O_5 \neq O_6$, 那就说明时间的推移产生影响。如果 $O_5 = O_6$, 那就说明时间的推移并未产生任何影响。

g . 有什么办法可以确定三个组最初 (实验前) 是一样的? 这一点有重要意义吗? 有, 为什么? 没有, 为什么?

回答: 由于未进行前测, 所以无确定实验前的 O_s 一样。但有无这一点并不重要, 因为随机编组 (R) 时, 已经假定各组是等同的了。

释例 5.3

这个例子可能复杂一些, 因为包含了 6 个组。其中 3 个组进行了前测, 3 个组在实验刚刚完成时进行了后测 (这是一个改进了的所罗门 4 组设计, 在时间上作了延伸)。所有 6 组在实验完成一段时间之后, 都进行了后测。实验处理只有一项 (X), 同时使用了控制处理。实验设计如下:

RG ₁	$O_1 - X - - O_2 - - O_3$
RG ₂	$O_4 - X - - - - O_5$
RG ₃	$- - X - - O_6 - - O_7$
RG ₄	$- - X - - - - - O_8$
RG ₅	$O_9 - \quad O_{10} - O_{11}$
RG ₆	$- - - - - - - - - O_{12}$

a . 结果: $O_3 \neq O_5$, 及 $O_3、O_5 \neq O_{11}$

解释: 在前测组里, 存在长期实验影响 (这一结果与无前测组或短期影响无关), 但是, 这种实验影响对经过先前的后测和未经过先前的后测的两个组的影响是不一样的。

b . 结果: $O_2 = O_6$, $O_2、O_6 \neq O_{10}$, 但 $O_3 = O_5 = O_7 = O_8 = O_{11} = O_{12}$ 。

解释: 存在短期的实验影响。而这种短期实验影响对于前测组和未前测组是一样的。不存在长期影响。

c. 结果: $O_2 \neq O_6$, 但 $O_6 = O_{10}$, $O_2 = O_3 = O_5$, $O_6 = O_7 = O_8$, $O_{11} = O_{12}$ 。

解释: 实验处理与前测之间有一交互作用, 这一影响长期持续。注意, 这不是一个简单的前测影响, 因 $O_2 \neq O_{10}$ 。

d. 作出何种比较才能确定一个未经前测的控制组在一段长时期内是否有所变化?

比较: 要作 O_9 与 O_{12} 之间的比较。由于各组都是随机组合的, 如果第 6 组也参加了前测, 可以设想它的前测分数将与其他各组相似。由于前测早于实验处理, 因此 O_4 与 O_{12} 也可以进行比较, 如果 $O_9 \neq O_{12}$ 就说明未经前测的控制组在过去的长时间内有了变化。如果 $O_9 = O_{12}$, 则说明未经前测的控制组在过去的长时期内没有变化。

e. 作出何种比较, 才能判断出只经过一次后测的两个实验组有着长期的实验影响?

比较: 要看 O_5 , O_8 和 O_{12} 。如果 $O_5 = O_8$, 而 $O_5, O_8 \neq O_{12}$, 就说明存在着实验效应。而且这种实验影响未受前测的影响。如果 O_5, O_8 和 O_{12} 均不相等, 那么仍然存在实验影响, 也存在实验影响和前测的交互作用, 如果 $O_5 = O_8 = O_{12}$, 则说明不存在长期的实验影响。

f. 有哪 3 种不同于一般交互作用的影响, 可以用这种 6 组设计加以检验?

回答: (1) 实验处理影响 (包括长期的和短期的)。

(2) 前测影响。

(3) 前后测的影响。

在先前的设计中可以进行其他的比较, 当然, 也会有其他结果的格局。设计和解释结果的一个重要特点是, 必须存在着一个为了与确定可能的影响做对照的比较组。所以, 在第一例中, 因为对所有组进行了前测, 前测的影响无法确定。在第 3 个例子中, 可以确定前测影响, 但只能通过对作过前测和未作过前测的组进行比较。我们不能只通过考虑 G_1 , G_2 和 G_5 来确定前测的影响。

对实验结果的解释是个最基本的过程。设计在结构上帮助研究者进行想要的比较和对比, 以便确定有无影响。在任何一个特定的实验中, 了解变量和与研究有关的可能的结果, 也能帮助理解数据。

10 随机性和代表性

在本章描述的实验中, 涉及一些对被试的随机抽样和随机分组, 通过随机选择的样本, 可以代表所选的群体。如果对被试随机进行实验处理 (还包括控制处理, 如果有的话), 那么在实验之前, 被试组只会因为随机抽样的波动而

有所不同，因为随机分布实验组是等同的。

然而，如果有大量的可提供的被试被随机分布到实验处理中去，这些被试代表什么样的总体？假定研究者使用报名听教育起源课的 120 名学生来做实验，学生被随机分布到实验处理中去，但这并非是从某一更大的总体分布中随机抽取的。无须争论他们代表什么总体，因为他们是随机地选择其自身的学习课程的。这是一个有关外在效度的问题，只能由通过了解研究被试变量的研究者回答。120 名学生能代表在过去几年报名上教育起源课程的学生吗？他们总体上代表学教育的大学生吗？或者总体上他们是代表年轻人吗？可能不是后者，但对其他而言可能会成立。

如果总是可以在所研究的总体中进行随机选择，那当然很好。但事实并非如此。上例中，如果 120 名学生参加一项学习实验，结果想来是有很大的可推广度的。它们可能可以被推广至其他年龄组，这依实验条件而定。

在有些教育研究实验中，使用的是原始的组。这些组不是随机选取，这些成员也不是被随机分组。例如，当在实验中使用原始班级的学生，这种情况就会发生。使用这种组时，我们称之为“准实验”设计（该设计在下一章中讨论）。如果组别原先就不同于随机的性质时，那它就不仅会影响可推广度，而且也影响实验的内在效度，因为与研究的变量相关组可能存在原始的差异。

总之，在设计一项实验时，能够有随机的条件是最好的，但并非总是可能的。即使用了随机分组，被试也可能不是从一些大的总体中随机选取的，在逻辑上的推广还是有争议的。当然，任何实验结果的外在效度必须根据具体变量的背景和实验条件来确定。

总结

本章提供了一些实验研究中较常见的设计。实验研究区别于其他研究的不同之处在于变量的操作。实验设计为那些由研究者精心操作和控制变量的实验提供了实验的框架。可能会有这样的误解，以为设计的复杂性是一位更老练的实验者想得到的特性，其实一名真正成熟的实验者需要的仅是符合这项工作的实验设计——满足研究被试的需要，能充分地检验假设。不管数据的格局是什么，一个实验必须有明确的假设陈述，设计必须能检验这些假设，并为实验结果提供有意义的解释。

就这一点而言，读者应理解实验设计的本质所在和各种设计框架的逻辑。本章已介绍了有关良好的实验设计应具有的特征。一个构思精妙的设计并不能保证有效的结果，但是，一个不合理、不充分的设计必定会导致无法解释的结果和差异很小的结果。设计需先于实验考虑，并对其进行缜密的计划和应用。

没有什么实验后的处理、统计或别的什么方法能替代一个构思精妙的实验设计的作用。

第六章

准实验设计

1 效度问题

人们常常把实验研究称为“真”实验设计，这是因为它们具有随机安排被试（组）进行实验处理的特征。这样，实验对等组的获得及其在一定范围内的随机波动都归结于随机安排。然而，在进行教育研究时，被试的挑选和组合不可能总是随机的。在教育领域中，有许多被试是既定的，如由一个班级里的学生自然而完整地形成被试群体。当这种保持原样的组运用于一个实验中时，我们便称此实验为准实验研究。这种研究能够为教育作出有价值的贡献，但有一点很重要：研究者对实验结果的解释和推广应该特别慎重。

准实验研究是指一个实验运用原始群体，而不是随机地安排被试进行实验处理。

缺少随机组合本身潜在地影响着实验的效度——无论是内在还是外在效度。在第5章中，我们就被试的不同挑选方式对内在效度的威胁作了说明。假设在一个实验中，采用5年级2个现成的教学班，其因变量是自然科学的成绩，操作性地定义为一次自然科学测验的分数。这些班级在先前已经按能力进行了分班，一个班级属高材生班，另一个班属普通水平班。两个班级接受不同的实验处理。假如，一种实验效果产生在高材生班，就很难证明这个效果归因于实验处理。因为能力水平和实验处理是混淆不清的，我们就没有把握对其结果作出解释。

任何因素都可能对原始群组起作用，同时我们无法证明这些组是否属于较大群体的随机样本。随机的挑选或组合是一个过程，它或者做得到，或者做不到。对于准实验来说，它没有做到，因此存在着被试挑选的偏差损害实验结果的可推广性的可能。

一名采用不加处理的试验组的研究者要做些什么呢？为了实验的可推广

性，必须在同一的逻辑基础上对试验组的典型性加以论证。出于对内在效度的考虑，研究者必须尽力证实两个试验组间对等的程度，这需要对与所研究的变量有关的特征或变量加以认真考虑。例如，假设现在的班级运用于一个数学教学实验中，年级水平将很可能作为一个常量，或者作为一种需要控制的变量。研究者也想得到表明这些班级具有同等能力水平的证据，假如像 IQ 测验分数这样的实证资料可以提供的话，这会极有助于对等组的检验。实际上，这些资料有时可用于统计对照。即使采用实证性的资料，检验和确立对等组总不免带有某种基于有关实验条件和变量信息的主观判断性质。当我们解释实验结果时，必须对缺少随机性予以特别注意，因为，它可能被抵消的程度也决定了实验的内在效度的可信性。

当考虑到准实验研究的效度问题时，应该对它的缺陷有清楚的认识，并对实验组间的对等性进行确定，同时在逻辑的基础上对其可能的代表性和可推广性加以论证。

2 仅施后测、非对等控制组设计

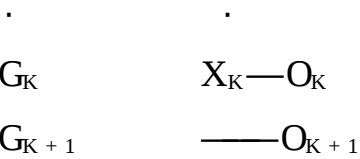
除了不是随机地将被试组成试验组外，许多准实验设计看起来非常像前章所讨论过的实验设计。当运用“非对等”这一术语时，它意味着随机意义上的非对等，但并不是说试验组间不可能在相关变量或特征上具有相似点。的确，就准实验设计来说，对其结果效度的可信性，很大程度上取决于保证实验组间的相似性。

使用前章中所引入的符号，非对等控制组仅施后测设计的最简单的形式可以图解如下：

$$\begin{array}{ll} G_1 & X_1-O_1 \\ G_2 & X_2-O_2 \end{array}$$

这种设计表示：一个试验组接受实验处理，同时另一个试验组作为控制组不接受实验处理。对实验组 G_1 完成实验处理不久，同时对两个试验组实施后测。这种设计可以推广为包括任意次的实验处理，对于 k 次实验处理，这种设计需要 $K + 1$ 个试验组：

$$\begin{array}{ll} G_1 & X_1-O_1 \\ G_2 & X_2-O_2 \\ \vdots & \vdots \end{array}$$



假如运用两次或多次实验处理而没有对控制组进行实验处理，那么这种设计就可以称作非对等多组仅施后测设计。

非对等控制组仅施后测设计包含与实验处理次数一样多的实验组，再加上一个控制组。使用现成的试验组进行，仅在实验处理实施后测量被试一次。

任何实验的效度依赖于特定的实验条件，采用非对等控制组仅施后测设计的实验效度一般是较低的。这一难点可能产生于实施实验处理前挑选被试的偏差和缺少前测。除非能获得可以提供有关各试验组相似程度这样的现成资料，否则不应该采用这种实验设计。这些资料虽然不能消除被试挑选上的偏差（假设它存在的话），但它们提供的信息可以避免对实验结果的错误解释。请看下面的例子：

例 6.1

一位初级中学的老师，教授初三年级 4 个班的自然科学课，要研究采用 3 种不同的新方法和传统方法（控制）对本学科实验部分的教学的影响。这位老师对 4 个班级分别采用一种不同的教学方法。因变量是对该课程中的实验内容实施期末考试的成绩。研究的问题可以陈述如下：

一项关于教学方法对初三年级自然科学考试成绩影响的实验研究。

实验设计的图解如图 6.3。

没有实施前测，但为了检查班级的相似性，查阅了其他材料，获得了以下信息：

- 1. 在班级中，男生和女生的比例大略相等。
- 2. 学生在先前初二年级时的自然科学成绩，1、2 和 4 班的大约相同，而 3 班学生成绩稍微高些，对于诸如数学等其他学科，初二年级时的学生成绩情况相同。
- 3. 尽管这所学校大部分不是按学生能力分班，但有一个英语优等班；因为受课程安排时间的限制，3 班的许多学生也在英语优等班里学习。

在这项研究里，教学时间和教师是常量。1 班和 4 班上午上课；另两个班下午上课（指上自然科学实验课）。

对于可能影响考试成绩的变量，1 班、2 班和 4 班看起来很相似，然而，3 班似乎是一个能力较强的班级，在解释实验结果时，这是不能不考虑的。既然任何一个班仅接受一项实验处理，那么实验处理和能力水平间可能存在着部分的混淆。

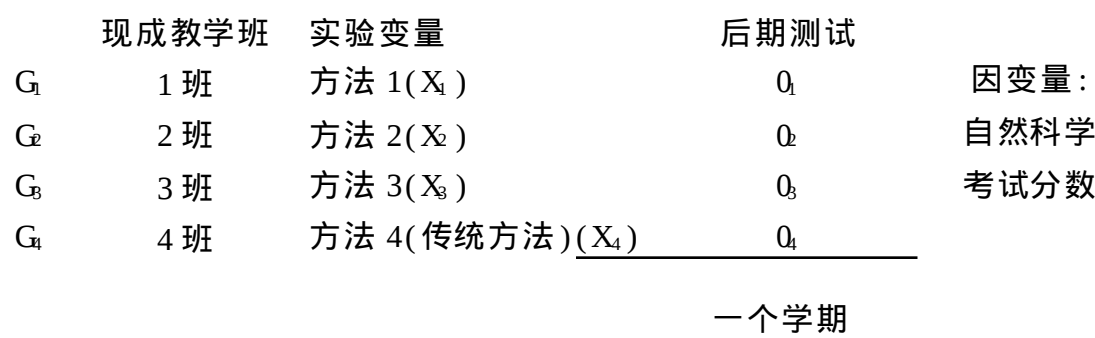


图 6.1 例 6.1 中非对等控制组仅施后测设计的图解，包括 3 个实验组和一个控制组

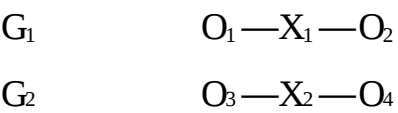
例 6.1 的实验结果及其解释 假设后期测试呈现出以下实验结果模式：O₁ = O₂，O₁ 和 O₂ 大于 O₄，而 O₃ 又大于 O₁ 和 O₂。

解释：方法 1 和方法 2 都比传统方法更有效，而且方法 1 和方法 2 效果相同。既然一班在上午教学，另外两班在下午教学，这些方法看来并不受一天中时间不同的影响。我们对于方法 3 无法得出确切的结论；实际上，它有可能不如传统方法有效，3 班后期测试的高分可能归结于学生的能力。

这个例子阐明了这样一个事实：根据实验结果的情况，准实验设计可以有多个解释。可以用方法 3 是有效的方法来解释 O₃ 的高分，也可以将 O₃ 的高分归因于实验组 G₃ 的高能力。假设测试结果 O₃ 比 O₁、O₂ 和 O₄ 都低，那么就可以毫无疑问地说：方法 3 不如其他方法有效，至少对能力较高的学生无效。1 班、2 班和 4 班的相似性允许我们相对地承认运用方法 1、方法 2 和传统方法得出的结论。

3 前测—后测非对等控制组设计

非对等控制组实验设计除了对被试也实施前测外，其他方面与非对等控制组仅施后测设计相似。这种实验设计的最简单形式仅需要两个试验组：一个实验组和一个控制组。假如不含有控制组，那么此设计称为前测—后测非对等多组设计。假如有 k 次实验处理，其一般形式可以图解如下：



$$\begin{array}{ccc} & & \cdot \\ G_K & & O_{2K-1} - X_K - O_{2K} \\ & & \cdot \\ G_{K+1} & & O_{2K+1} - O_{2K+2} \end{array}$$

前测的结论对核对各试验组的相似性是非常有帮助的，因为前测的分数是与因变量有密切联系的变量。这种前测是在进行实验前在同样的条件下对所有被试都进行的测试，其分数也可用于统计控制，在有些情况下，可能会产生增益分数。

前测—后测非对等控制组设计有助于核对试验组间的相似程度，其前测的分数可用于统计控制或产生增益分数。

请看下面的例子：

例 6.2

在 4 年级进行一项采用两种新的阅读方案的教学实验，研究各方案对学生阅读成绩可能产生的影响，这项设计的实验处理是新的阅读方案，传统的方案作为控制处理。同一学区的 30 个小学 4 年级班级参加实验，不进行随机挑选和组合。每种方案都有 10 名教师报名参加，即两种新方案和一种传统方案各由 10 名教师来施行。当然，每名教师只能施行一种方案。先用阅读成绩测验 A 卷对学生进行前测；然后用这 3 种教学方案教学 18 周；随后对学生采用阅读成绩测验 B 卷进行后测。这项设计可图解如图 6.2。

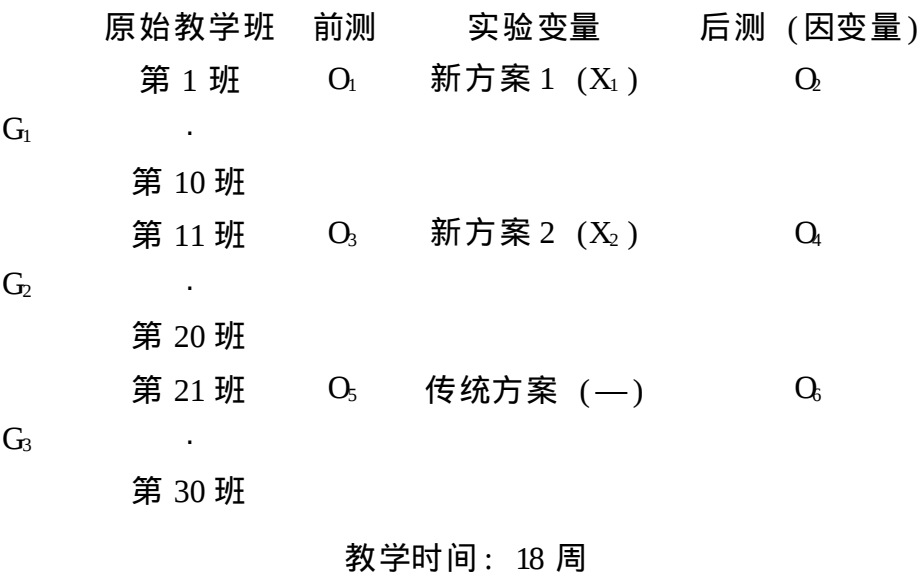


图 6.2 非对等控制组前测—后测设计例 6.2
图解，有两个实验组和一个控制组

事前测试分数有助于检查各班的相似性，但这并非是能够检查的惟一变量。由 30 位老师施行，且每位老师仅采用一种方案。在可能影响阅读成绩的

因素方面，每 10 个教师的构成实验组间是否相同呢？尽管各个教师间可能存在差异，但若从全体教师考虑，10 个教师一组的小组间则可能很相似，其中要检查的一个要素就是每个教师教学经历的长短。如果所有经验丰富的教师都集中到一个组里，那么在各组之间就会存在一种系统性差异。

如果 30 个班级所在的学校分别处于诸如社会经济条件差别很大的地区，那么这点也必须加以考虑。如果把一种方案限定在一种社会经济水平不同于其他地区的学校，实验将不会得到满意效果，因为这样会造成学校影响和阅读方案影响相混淆。理想的安排方法是：每 3 个社会经济条件相同的学校，各执行一种方案。某些学校将可能执行两个或 3 个方案，尽管这种安排或许会因不同班级学生间的相互交往而导致实验数据的“污染”。

图 6.2 中，虽然有 30 个班级，但仅有 3 次事前测验和 3 次事后测验。当我们分析实验结果时，首先要核对各组的结果，但对于像此例这样的大规模研究，通过进行如下所述的更为细致的比较，对实验结果进行分门别类的整理通常是有益而又必要的。

1. 假设一个方案（组）里的班级事前测试分数相似。比较这一方案内 10 个班级事后测验的分数，这些分数是彼此接近呢，还是差别很大？如果它们彼此相近，那么表明这个方案具有一致的效果；如果它们差别极大，那么固有的差异超过了方案的任何效果，或者说方案的效果对于各个班级是不一致的。

2. 假如前测的分数变化不一，根据前测分数，将这一方案内的班级分成两个或 3 个类型（高、中、低）。接着检查每一类型的后测分数，来确定这一方案内各类型间和同一种类型各个方案间增益分数是一致，还是不同。例如，假设对于前测中得分高的班级，接受新方案 1 实验处理的增益远大于新方案 2 和传统方案的增益。这是一种同一类别内的跨方案比较，同时它表明第一种方案对于那些起初就有较强阅读能力的学生来说是最有效的。

3. 如果这些班级前测分数相似，比较同一方案内各班级后测分数。假如它们大略也相同，那么诸如教师和学校等外部因素具有统一的影响；假如分数不相同，那么这些因素产生了不同的效果。

例 6.2 的实验结果及其解释 假设出现了以下结果： $O_1 = O_3 = O_5$ ，但 $O_2 \neq O_4$ ，而且 $O_2, O_4 \neq O_6$ ，但 O_4 大于 O_2 ， O_2 又大于 O_6 。

解释：根据前测结果，一开始小组间看起来十分相近，产生方案效果：两种新方案都比传统方案有效，而且第一种新方案最有效。

在许多非对等组的实验中，实验设计尽量把控制变量作为自变量包含其中（上述的社会经济水平如果能以这种方式包括在内，那它将是一个控制变量的例子）。从本质上来说，这种设计可以扩展为一种因素设计。如果随机性的某个方面得以考虑，这种实验设计的效度会大大地提高。在上述例子中，如果

30 个班级随机安排是不大方便的，但是将 10 人一组的教师分组建立在随机分配的基础上均等地进行，则是完全能办到的。这样的安排就可以使实验处理和控制处理中教师间的差异，得以折衷综合，趋于相等。当我们运用准实验设计的时候，一般都尽力让实验得到尽可能充分的控制。此外，要利用可搜集的信息来检查各组被试的相同性，对于实验结果的解释和推广应用，也应该根据这一信息的背景和实验条件来进行。

4 时间系列设计

时间系列设计包括一组准实验设计，这组设计涉及到对一个或多个原始小组的反复测量，并在至少一个实验组的两次测量之间进行一次实验处理。时间系列设计对那种过一段时间自然地周期性地对因变量进行测量的情景是有效的，如对一个班级的反复测试。观察测量应该是一致的，但因为存在多个因变量，要达到一致性可能是不容易的。

时间系列设计就是对一个或多个原始的被试组进行反复测量，并在至少一个组的两次测量之间插入实验处理。

单组时间系列设计

一个单组时间系列设计可以简单地图解如下：

$$G \qquad O_1-O_2-O_3-X-O_4-O_5$$

我们已经指出，系列设计没有对被试组进行随机分配。可以有任何可能次数的观察或测量，实验处理的插入，可随机地进行。观察测量可以与例行的考试合并进行，如在一个班级中每 4 个星期一次的测验。

时间系列设计的一个特征就是可能的结果模式有很多。这就产生了一个内在效度的问题，尤其是仅有一个被试组的实验，对于实验结果可能有其他解释，而未必是实验处理的效果。图 6.3 表示了 3 种可能出现的结果模式。坐标中横轴上的 O_s 代表测量，纵轴代表因变量的量度。

任何一个特定的实验和一种因变量，只能有一种结果模式。对于 A 模式的解释是：这项实验处理显然是有效的。直线的倾斜度直到第 5、6 次测量之间才趋于平缓。从表现上看，B 模式好像没有什么实验处理的效果。然而最后两次测量间出现的明显上升，则可能说明这一实验处理不是无效，而是迟效。

如果没有明显的外部事件造成这一变化，那么以实验处理效果来解释当然是合乎情理的。鉴于此因，预见到实验处理的介入与实验效果的出现之间的间隔时间，是很重要的。对于一定的变量，在 B 模式中的效果和它在 A 模式中的效果一样明确。应该指出，随着时间间隔的延长，无关干扰事件介入的可能性也会相应增加。

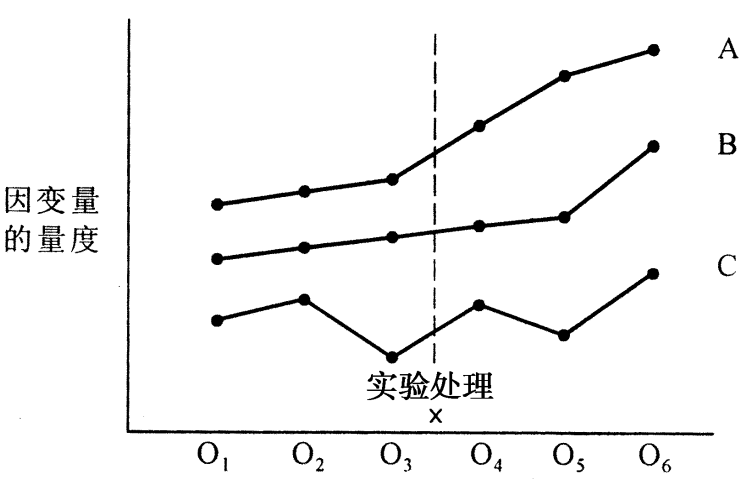


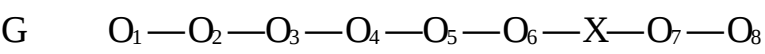
图 6.3 一个单组时间系列设计可能的结果模式

C 模式的不规则曲线，几乎排除了对实验处理作出结论的可能性。由于没有控制组，很难推断不受实验处理的结果模式会怎样。各测量点之间的波动可能表示有其他因素发挥了作用，而且其作用的强度超过了实验处理的效果。这时候，有必要对实验加强控制，直至它灵敏地反映出实验处理的效果来。然而，我们不能根据 C 模式的曲线得出实验处理无效的结论。

时间系列设计中的多次观测是颇有用途的，它不仅能使研究者据此确定实验处理的可能效果，而且在效果可能不出现时，使研究者避免作出草率的推论。思考一下 C 模式，如果只有 O₃ 和 O₄ 两个观测结果，那么研究者将作出结论：实验处理有效。其实两次观测结果的差异可能更应归结于其他因素。对于 B 模式，如果只观测 O₃ 和 O₄（甚至再加上 O₅），那么可能存在的迟效将会被忽略。因此，全面地考虑整个模式是很重要的。请看下面的例子：

例 6.3

一位理疗师在对 12 个人一组的病人实施一项为期 9 星期的康复计划。小组成员每天都要接受治疗，在每周末根据一份身体能力测试表进行测试。除了第 7 周（在随机的基础上决定的）期间施行一种实验治疗法之外，都采用一种传统类型的治疗法。此项设计可以图解如下：



假设此项实验的实验结果呈现如图 6.4 所示的模式，这些实验结果将怎样解释呢？有力的证据表明：实验治疗法要比传统治疗法更有效。前 6 周的身体

改善状况的模式非常一致，但整个第 7 周，曲线呈现出急剧上升趋势。在第 8 周期间，身体好转的状况又恢复到早期水平。这样，除非有某种其他因素影响，否则为何第 7 周期间体质测试成绩上升很快呢？这为解释实验效果提供了很好的事实。

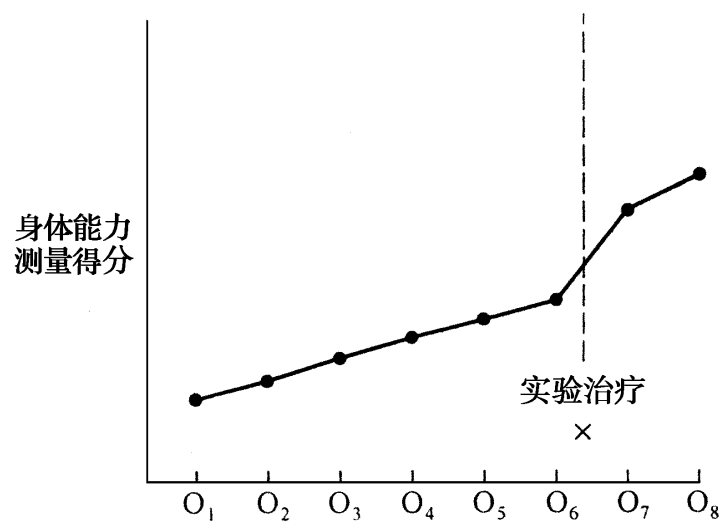


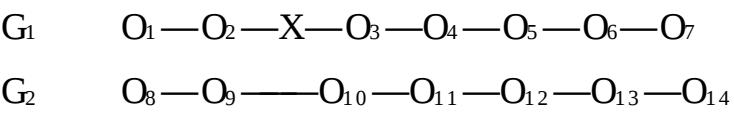
图 6.4 例 6.3 实验结果的模式，运用单组时间系列设计

在此实验中，保持观测的一致性是没有问题的，因为贯穿实验始终，都采用同一份体质成绩测试标准。我们来看另外一种情况。一位小学教师运用时间系列设计来检验拼写练习中个别练习与小组练习的效果，他以所教的班作试验组，每周这个班都在课外安排一定的时间来进行拼写练习，而且每星期五进行一次测验。平时的练习方式都是个别练习，但 6 周后的一个星期，采用小组练习的方式，并作为实验处理来进行。

这一实例中的一个重要问题，就是保持每周测验难度水平的一致。如果小组练习之后的测验比其他星期的测验容易，那么这一周班级的分数就会偏高，较容易的测试干扰了对实验结果的解释。当然，无论练习的方式如何，练习时间的长度应该是个常量。

多组时间系列设计

单组时间系列设计可以扩展，使其包括两个或两个以上的组。这种实验设计通常包括一个控制组，这种情况的设计例子可以图解如下：



此外，可以进行任意次数的观测，并且实验处理随机地插入一个实验组，各组测量的次数也要相等。

多组时间系列设计包含两个或两个以上的原始被试组，其中一个小组可以作为一个控制组，应至少对一个小组插入实验处理。

两个或两个以上小组得出的结果巩固了多组时间系列设计的地位，因为它为实验提供了比较，因而也就增强了实验的内在效度。例如，这种设计提供了一种检验伴随实验处理的外部事件是否起作用的可能性。假设在一个有控制组的实验设计中，对实验组实施实验处理之后，立刻进行观测所得的观测结果表明：两组都产生一个非常大的增长。既然在两组中都产生了这种增长，这就不能说是一种实验的效果（因为控制组没有接受实验处理），因此这很可能归因于同时影响两个组的某种外部因素。

在实验处理之前进行观测的结果可以用于检查小组间的相似性。对于任何多组准实验设计，小组间的相似性越大，从实验结果中得出的结论就越可靠。请看下面的例子：

例 6.4

一个老师教 1 年级 3 个班的代数，他决定进行一项研究，其研究问题可以陈述如下：

一项关于不同反馈类型对代数成绩影响的研究。

在这个学期中，这位老师进行了 5 次同等难度的一小时代数测验。尽管随着教学的进展，这些测验涵盖了不同的教学内容，但因为老师精心采用了难度水平大致相等的题目设计试卷，因而每次测验具有大体相同的难度。在第 2 次和第 3 次测验之间，教师给 1 班以正反馈（ X_1 ）；给 2 班以负反馈（ X_2 ），对第 3 班不做任何反馈（控制处理）。这项实验的图解如图 6.5。

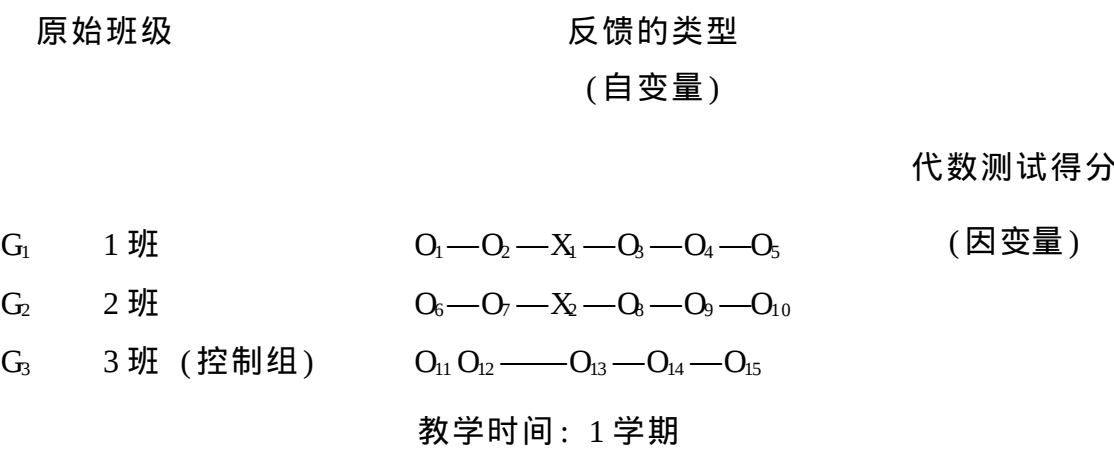


图 6.5 例 6.4 多组时间系列设计图解，包括两个实验组和一个控制组

这一实验设计不但可以使教师在两个实验组间做比较,而且也可以在实验组与控制组间进行比较。注意到实验处理仅在两次观测之间实施,在这一点上,我们可以考虑可能的实验结果模式和可以做的有关解释。因为有很多的观测值,解释结果就需要我们对之加以整理。下列格局我们可以独立地加以考虑。

例 6.4 结果模式 1 $O_1 = O_2 = O_5$, 同时 $O_3 = O_4$, 但 O_3 、 O_4 比 O_1 、 O_2 和 O_5 大; $O_6 = O_7 = O_9 = O_{10}$ 但 O_8 比 O_7 要小; 另外, $O_{11} = O_{12} = O_{13} = O_{14} = O_{15}$, 而且 $O_1 = O_6 = O_{11}$ 。

解释: 正反馈 (X_1) 提高了学习成绩, 其效果持续到第 4 次观测 O_4 ; 负反馈 (X_2) 降低了学习成绩, 但它仅产生了一个短暂的影响效果。因为第 3 组 G_3 的学习成绩高度一致, 看起来不像有任何外部因素在导致成绩发生变化。尽管没有随机地安排各被试组, 但依据代数考试成绩, 最初各班的考试分数是相等的, 所以各小组在实施实验处理前看起来是大体相似的。

例 6.4 结果模式 2 $O_1 = O_2$, $O_3 = O_4 = O_5$, 但 O_3 、 O_4 、 O_5 比 O_1 、 O_2 都大; $O_6 = O_7 = O_8$, 同时 $O_9 = O_{10}$, 但 O_9 、 O_{10} 又比 O_6 、 O_7 、 O_8 大; $O_{11} = O_{12} = O_{13}$; 但 O_{14} 、 O_{15} 要比 O_{11} 、 O_{12} 、 O_{13} 大, $O_1 = O_6 = O_{11}$, 而且 $O_4 = O_9 = O_{14}$ 。

解释: 因为这种模式中第 3 组 (G_3) 和第 2 组 (G_2) 相同, 那么负反馈 (X_2) 无效。正反馈 X_1 提高了学习成绩, 至少显示了有直接的效果。因为所有的班级在第 4 次测验中成绩都提高了, 所以很难对正反馈 (X_1) 的长效性作出任何推断。各个班级代数成绩的一致提高很可能归结于一种外部因素。无论什么因素引起了这种增长, 它都产生了一种持续到第 5 次测验的效果。既然各班开始测试的分数都相等, 因此各班在实验开始时看起来是相似的。

当然, 伴随大量的观察, 像上面的事例, 有可能出现大量不同的结果模式。假如这些模式都是无规则的, 要得出一般性结论往往是困难的; 如本例中, 若各班的观测结果始终波动不定, 就很难得出结论。另外, 如果先于引入实验处理前的各班测试得分存在差异, 那么就很可能存在一种选样的偏差。

时间系列设计的变形

前面的讨论集中在单组和多组时间系列设计——这是时间系列设计的基本形式。然而, 有一些实验设计的变型也可归于这种设计中。在时间系列设计中观测的次数依赖于所研究的变量, 但应该有足以建立结果模式的充分的观测值。其中一种变形是增加时间系列中观测的次数, 对于长期的实验或对于那些观测值可以紧密排序的实验, 观测的次数甚至可能达到 15 次或 20 次。假如实验时间被延长, 增加观测的次数确实会增加外部因素产生作用的可能性。

另一种变形是在时间系列中插入多于一次的实验处理。如果时间系列被延

长，这种变异是很有可能的。多次插入实验处理为这项实验效果（如果这一效果存在的话）的一致性，提供了一种核对的条件。有两种途径可以完成实验处理的多次插入：(1) 以随机的方式两次或两次以上将实验处理插入时间系列；(2) 一旦将实验处理插入，那么使它继续保持在实验的后续部分中。这两种方法可以图解如下：

1. 多次随机插入 X

G O₁—O₂—X—O₃—O₄—O₅—X—O₆—O₇—O₈

2. 连续插入 X

G O₁—O₂—O₃—X—O₄—X—O₅—X—O₆—X—O₇—X—O₈

这些方法中的任何一个都可用在上例代数班级的实验处理中。X₁ 和 X₂ 的两种强化可以多次随机地插入时间系列中；或者在它们第一次插入后，将它们持续地保持在这一学期的后续部分实验中。

多次插入实验处理的变型可以归于时间系列设计中。

5 单个被试设计

教育研究中大多数实验研究涉及的是包含多个被试的群体，换言之，我们试图通过实验获得可用于群体而不是个体的实验结果。然而，对于有些实验情境，采用个体被试是合适而且是必要的——本质上讲，样本就是一个。在这些单个被试情境中，基本的试验方法是研究在实验条件和非实验条件下的个体。

单个被试研究有助于教师从事个别学生的研究（可能的行动研究）。以个体的方式对学生进行指导的辅导员，就可以使用单个被试设计。那些康复和理疗领域的研究者也可采用个别研究。一般来说，一个被试因为某种情况或问题参与一项研究，不存在随机挑选或分配的问题。因此，单个被试设计通常被认为是准实验设计。

单个被试设计一般要进行反复观测，有时要对因变量进行好几次观测。而且观测要高标准严控制，这样，观测的差异才不至于被解释为一种实验效果。我们必须对实验的条件加以细心描述，这样不但能加强实验结果的解释，而且有助于对实验结果的可推广性作出判断。

单个被试设计具有常被称为“单一变量规则”的特征。其意思是，在实验处理实施期间，仅有一个变量（即实验处理）改变。在传统处理（或称基线处理）和实验处理期间，所有其他条件——诸如时间长度、观测次数等，都保持不变。为了避免把某种其他效果误解为实验处理效果，有必要对实验结果作出

解释。

传统的处理或正常条件起作用的这段时间被称为基线，这段时间应该长得足以保证因变量得以稳定。假如一个因变量正波动不定时，实施了实验处理，那么就不可能判定因变量的变化是否归因于实验处理。

单个被试设计通常要进行反复观测，同时它还适用于“单一变量规则”——一次实验仅改变一个变量。

正如其他任何准实验设计一样，效度也是单个被试设计需要关注的一个问题。研究者必须对内在效度加以确定，然后才能对实施结果作出解释。同时要考虑到对实验观测结果（非实验效果）的其他解释，正如我们所希望的，对这些解释要去伪存真。面对多种的解释，有必要尽可能保持对实验的控制，同时要了解研究中可能起作用的其他变量的属性。外部效度依赖于研究工作和其他情景的相似性，这必须在一种逻辑的基础上加以论证。

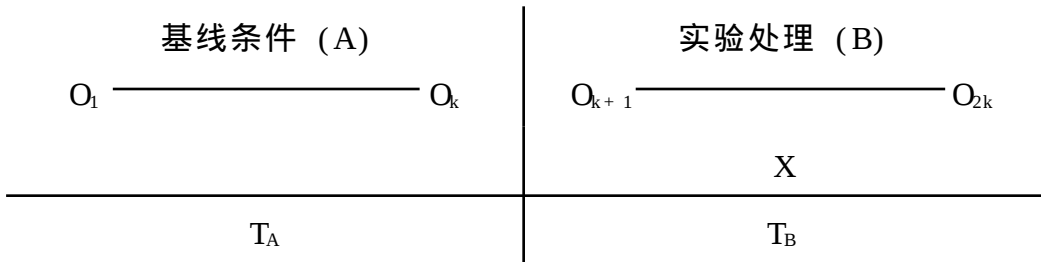
A—B 设计

我们采用一组相当独特的符号来表示单组被试设计，A 和 B 用来代表实验条件：A 表示基线条件，B 表示实验处理条件。既然实验使用个体被试，就不需要被试组符号。

A—B 设计是最简单的单个被试设计。一般来说，这种设计可以图解如下：

在这一设计中，研究者要在基线条件下观测单个的被试，直到因变量趋于稳定。接着将实验处理引入实验，再对被试进行相同次数的观测。图底部的 T_A 和 T_B 代表着实验设计中的时间段，并且 $T_A = T_B$ 。

对 A—B 设计实验结果的解释是建立在这样的假设基础上的，即要是实验处理没有引入的话，基线条件下的观测结果就不会发生变化。这种设计极易受其他变量的影响，这些变量可能与被试的经历和成熟有关，并可能被作为实验处理效果的成因来解释。这当然会成为对内部效度的一种威胁。既然两种条件之间产生的变化仅有一次，在某种意义上，就其内在效度而论，A—B 是最差的单组被试设计。



例 6.5

一位新老师陷入课堂管理的困境，而一位经验丰富的教师正在帮助这位新教师解决这一难题。这位经验丰富的教师对新教师进行了为期 4 周、每周两次的观测，采用的是一份教师表现观察记录表，诸如《有效研究的主要课堂观察表 (COKER)》，这段时间属于基线时间 (A)。8 次观测所得的资料组成了基线资料。在这 4 周期间，新教师的课堂表现非常稳定。

实验处理 (B) 是两位教师半个小时的磋商，其间经验丰富的老师讨论了新教师的课堂表现并尽量指导新教师的行为向改善课堂管理方向转变。这种磋商进行了 9 次，第一次是在实验处理条件 B 下的第一次观测之前，随后每进行一次观测紧接着一次磋商。就像条件 A 一样，条件 B 也实施 4 个星期，而且 8 次 B 条件下的观测也是在与 A 条件相同的情况 (同样的班级、同样长的时间、一天的同一时间等等) 下进行，惟一不同的是实验处理。这一研究的实验设计可以图解为图 6.6。实验资料是用 COKER 记录表观测到的。

例 6.5 结果模式 1 从 O_1 到 O_8 的观测值是稳定的，这使我们确信教师的目前行为不能改善课堂的管理。接着从 O_9 开始到 O_{14} ，观测值呈上升趋势，说明教师的行为改善了课堂管理，然后从 O_{14} 到 O_{16} 观测结果又趋于稳定。这一观测结果标示在图 6.6 中。

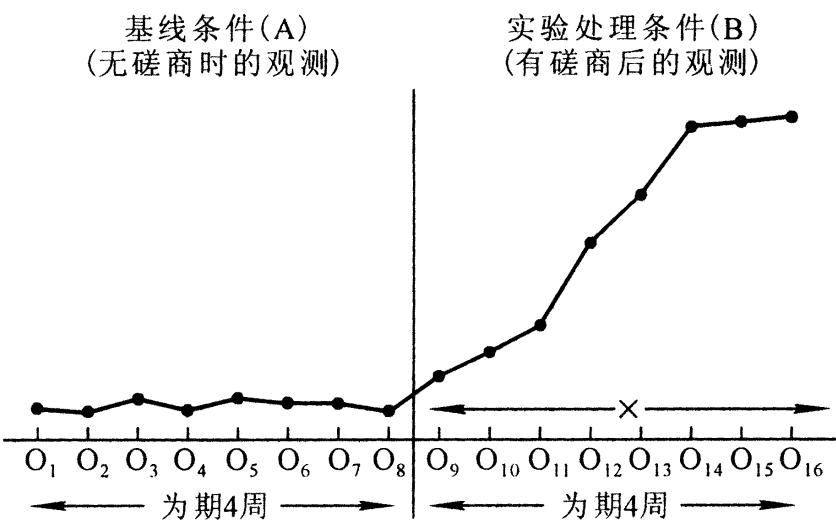


图 6.6 例 6.5 以结果模式 1 所做的单个被试 A—B 设计图

解释：这样一个结果模式，无可置疑地证明了实验处理取得了期望的效果。教师的课堂管理行为得以改善并处于一个稳定的状态。然而，这种实验结果也可能归因于新教师的自然成熟，加之由于时间相对较短，把这作为实验结果的另一种解释未必是正确的。

例 6.5 结果模式 2 从 O_1 到 O_5 之间，曲线表现出较大的波动，但 O_5 到 O_8 较稳定。从 O_9 到 O_{16} ，除了教师改善课堂管理的行为观测值稍高外，其曲线具

有和从 O_1 到 O_5 一样的波动模式。

解释：对于这些实验结果几乎不可能得到确切的解释。新教师的课堂表现非常不稳定，尽管实验处理似乎稍微改进了教师的表现，但在整个 4 周的时间内，教师课堂表现的稳定性并没有提高。显然，存在着诸如课堂条件或教师情感等其他变量的影响，并且其效果超过了实验的效果。

关于这项研究结果的推广性如何呢？假如结果模式 1 恰是我们用以推论的实验结果，那么这些结果将可以推广到其他新教师身上去，推广到那些具有和这项研究的老师特征相似和在类似的条件下教学的新教师。推广性必须经过详尽描述才能确立，这样才能证明推广情境与实验情境的相似性。因为这些教师要竭力解决的是一种即刻的、局部的问题，所以他们可能不太关心实验的推广应用性。

A—B—A 设计

A—B—A 设计是 A—B 设计的展开，它是紧接着实验处理这段时间又引入了一段时间的基线条件。此类设计因为实验处理在后期被取消，又可称为取消设计或删除设计。除了从基线条件到实验处理条件，再回到基线条件这一变化，其他特征——像持续时间的长短、观测的次数，都保持不变。与 A—B 设计相比，额外增加了基线条件的时间，实验结果模式得以扩展，因而实验的内部效度有所提高。对这一设计还需说明：每一基线条件或实验条件期间，所观测的次数相同；时间是个常量， $T_A = T_B$ 。请看下面的例子：

例 6.6

一教师有一学生，他的课堂行为极具消极性——表现为在课堂上连续而有破坏性地插嘴。这位教师对这一学生引起的破坏性情况进行了每周一次的记录。这种行为持续了 3 个星期基本不变，这段时间可作为基线条件。接着 3 星期，教师对学生进行每周两次的个别咨询，这些咨询作为实验处理，每周一和周四各进行半小时。3 个星期以后，这种咨询暂时停止。随后的另外 3 周，停止个别咨询恢复到原来的基线条件，教师继续搜集有关因变量的资料（整个星期中该学生造成破坏性情况的次数）。随后的 3 周基线条件与前一个基线条件相比较，没有其他明显变化，班级、所教学科等等都保持相同。这项研究图解如图 6.7。

例 6.6 结果模式 1 （符号的使用与第 5 章所介绍的一致，“ $=$ ”意思是大约相同）。 $O_1 = O_2 = O_3 = O_7 = O_8 = O_9$ ，同时 $O_4 = O_5 = O_6$ ，但是 O_1 、 O_2 、 O_3 、 O_7 、 O_8 、 O_9 都大于 O_4 、 O_5 、 O_6 。（注：因变量是学生在班级中造成破坏性情况的次数，因此低分是教师所期望的）。

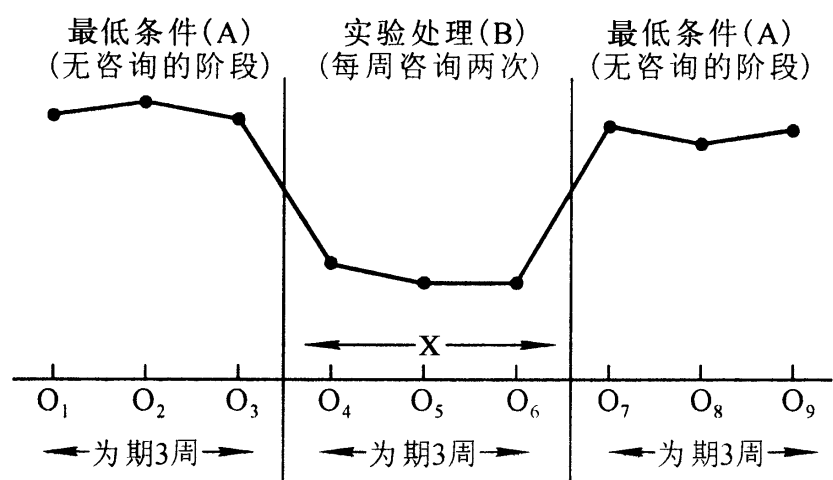


图 6.7 例 6.6 单个被试 A—B—A 设计，结果模式 1 的图解

解释：个别咨询取得了满意的效果，但这仅仅是一种短暂的效果。当这种咨询停止后，学生的课堂行为又恢复到老样子。好像无关变量并没有与实验处理一致的影响。

例 6.6 结果模式 2 $O_1 = O_2 = O_3$ ， $O_4 = O_5 = O_6$ ，同时， $O_7 = O_8 = O_9$ ，但 O_7 、 O_8 、 O_9 比 O_4 、 O_5 、 O_6 小，而 O_4 、 O_5 、 O_6 又比 O_1 、 O_2 、 O_3 小。

解释：这一结果模式产生了两种或两种以上的解释，因此，我们无法确定实验效果。或许存在一种实验处理效果，假如真有，它也是暂时的，但还有一种加速的持久效果，这一效果可能是难以解释的。有一种可能，就是某种与成熟相关联的变量在起作用。如果从 O_4 到 O_9 曲线呈现出持续的下降，而不单单是 O_6 到 O_7 之间的降低，那么这种解释的可能性会更大。

当然，可能有许多的实验结果模式，例如从 O_4 到 O_9 ，实验结果产生剧烈波动，那就不能做出有关实验效果的任何结论。实验处理可能在与干扰变量交互作用，或许学生的消极行为已经变得不稳定了，而这或许是，或许不是由实验处理所导致的。

A—B—A—B 设计

如果我们将 A—B—A 设计扩展，使之再包含一段实验处理期，那么它就成为 A—B—A—B 设计了。因为有了连续的观测值，而且因为基线条件和实验处理经历了两次循环，因此可以说，A—B—A—B 设计同 A—B 和 A—B—A 设计相比，提高了内在效度。如果两次循环的实验结果模式一致，那么可怀着极大的信心得出有关实验效果的结论。A—B—A—B 设计可图解如下：

基线条件 (A) $O_1 \text{ — } O_k$	实验处理 (B) $O_{k+1} \text{ — } O_{2k}$ X	基线条件 (A) $O_{2k+1} \text{ — } O_{3k}$	实验处理 (B) $O_{3k+1} \text{ — } O_{4k}$ X
T_A	T_B	T_A	T_B

与前面的设计一样，除了实验处理的引入，基线条件和实验处理的其他特征都是相似的。观测的次数相同， $T_A = T_B$ 等等，因而这一设计体现出单一变量的规则。请看下面的例子：

例 6.7

我们将把 A—B—A 设计中的例子延伸到 A—B—A—B 设计。这位老师决定在第 2 个基线条件期结束以后，继续对那位具有破坏性的学生进行第 2 次为期 3 周的个别咨询。这项研究图解为图 6.8。需要说明的是，这一设计可简单地看做是图 6.7 中的设计的延伸，其主要区别在于 A—B—A—B 设计的研究工作从 9 周延长到 12 周，同时从实验处理中获得另外的一组 3 次观测值。

例 6.7 结果模式 1 $O_1 = O_2 = O_3$ ， $O_4 = O_5 = O_6$ ， $O_7 = O_8 = O_9$ ， $O_{10} = O_{11} = O_{12}$ ； O_1 、 O_2 、 O_3 比其它任何观测值 O_s 都大； O_{10} 、 O_{11} 、 O_{12} 比其他任何观测值 O_s 都小； O_4 、 O_5 、 O_6 比 O_7 、 O_8 、 O_9 小，但又比 O_{10} 、 O_{11} 、 O_{12} 大。

基线条件 (A) (无咨询) $O_1 \text{ — } O_2 \text{ — } O_3$	实验处理 (B) (每周两次咨询) $O_4 \text{ — } O_5 \text{ — } O_6$ X	基线条件 (A) (无咨询) $O_7 \text{ — } O_8 \text{ — } O_9$	实验处理 (B) (每周两次咨询) $O_{10} \text{ — } O_{11} \text{ — } O_{12}$ X
三周	三周	三周	三周

图 6.8 例 6.7 单个被试 A—B—A—B 设计

解释：我们可以得出结论：实验处理产生了一种积极的效果。当实验处理停止时，学生的行为部分地恢复到原样，但当实验处理再开始时，积极的效果再次出现。3 周期间，学生在这种条件下的行为表现是稳定的。看起来不像有干扰变量伴随着实验处理一贯地产生影响。

例 6.7 实验结果模式 2 $O_1 = O_2 = O_3 = O_7 = O_8 = O_9$ ， O_4 、 O_5 、 O_6 相互不等，且不等于 O_1 ； O_{10} 、 O_{11} 、 O_{12} 相互不等，且不等于 O_1 或 O_4 、 O_5 、 O_6 。

解释：我们把这一结果模式的因变量分数在图上标示出来（如图 6.9）。实验处理期间，因变量反复无常的波动使我们无法得出关于个别咨询效果的任

何结论，无论是积极的，还是消极的。两个实验处理期的结果模式是不同的，因此，即使产生了咨询的效果，那它也是不一致的。大概我们惟一能得出的结论便是：个别咨询导致了学生行为变化或变得不稳定。这可能归因于实验处理和某种干扰变量的相互作用。因为两个基线期的学生行为都是同样稳定的，因而我们也可以说个别咨询没有产生长久的效果。

这些实验结果的推广性如何呢？这位老师可能将这一结果推广运用到其他年龄、品性与之相似，并且表现出相同类型的破坏性行为的学生。至于推广到其他老师、学生或其他问题行为上去，必须在逻辑上慎重考虑。为了结果的有效推广，还需要有在那种情境中进行个别咨询的知识，了解可能出现的效果。既然这是一个行动研究的例子，教师可能并不关心结果的推广性。要是出现了结果模式 1 的情况，教师就解决了这一问题。其他教师也许会对这一结果感兴趣，并在同一逻辑基础上将之推广到他们自己的情境中去。

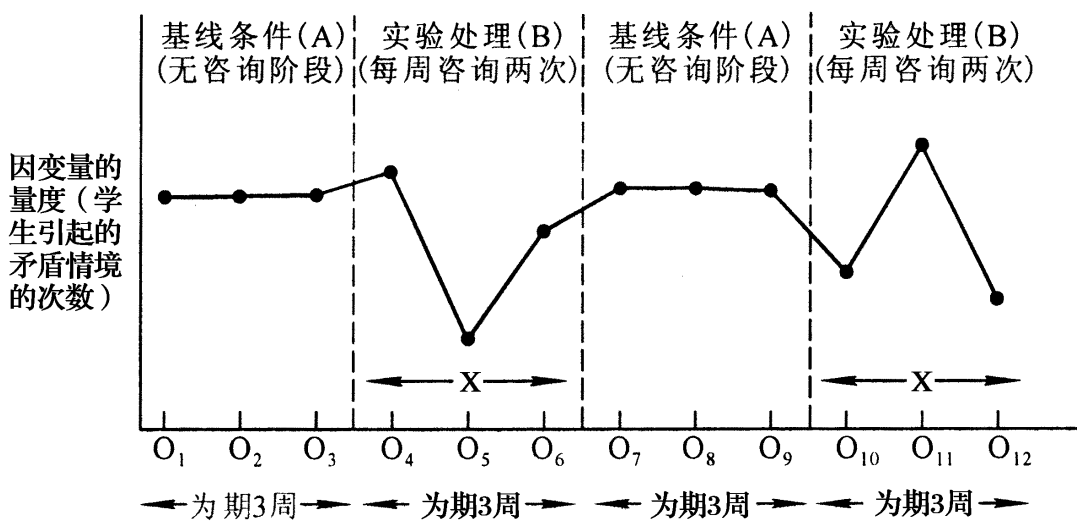


图 6.9 例 6.7 单个被试 A—B—A—B 设计结果模式 2 图解

多重基线设计

还有一种设计称为多重基线设计，它可以看做单个被试设计的修正。这种设计运用了 A—B 设计的逻辑，但并不限于一个被试、一种行为和一种情境，它们包含两个或两个以上的行为、情境和被试，或者这些项目的某种组合。因为这种设计基于不同的行为、情境和被试，所以它们包含多重基线。一般来说，一旦一种实验处理作用于一个被试，这种处理就要继续下去，因此，这种设计应用于那些实验处理开始后就不宜再撤除的研究。

多重基线设计行为交叉 在这种实验设计中，一个被试在同一种情境中的两类或两类以上的行为表现受到观测。等到各类行为基线都稳定了，才将实验处理运用到其中一类行为中，并保持一段特定的时间；接着再将实验处理以相

等的时间应用到第二类行为上，以此类推。如果实验处理实施之后，各类行为都产生了一致的变化，便为实验效果的存在提供了有力证据。

需要对这个设计作重点考虑的是行为表现的自变量。假如两类或两类以上的行为在其表现中是相互关连的，当我们对第一种行为进行实验处理时，它可能导致其余行为的变化。因此，这种设计最好运用于各类行为相互分散和独立的情境中。

多重基线设计被试交叉 这种设计采用两个或两个以上的被试，不过他们是作为单独被试分别参与到实验中的。多重基线产生于多个被试。在基线行为稳定以后，对一个被试施行实验处理；经过特定一段时间后，接着让第二个被试开始接受实验处理，以此类推。就实验处理来说，参与实验的各个被试相互独立是很关键的。只有这样，对一个被试实施实验处理才不会影响其他被试。如果一位教师采用这种设计，除非各个被试是从不同班级抽出来的，否则，要保持两个被试相互独立是很困难的。

多重基线设计情境交叉 这种设计的基线是同一被试的同一行为所处的不同情境。基线行为表现确立后，研究者将实验处理应用到一种情境中。经过特定一段时间后，这种处理再应用到第二种情境，以此类推。这一设计从本质上讲与多重基线设计行为交叉设计是相同的，只是变化着的情境代替了变化着的行为。正如前面所讨论的单个被试设计一样，多重基线设计在实验处理的施行期间，时间间隔和观测次数都是常数；至少在任何实验处理开始施行的时候，这两者是不变的。请看下面的例子：

例 6.8

这儿所举的例子是一个多重基线情境交叉设计。一位教师计划给一个学生对阅读、数学和社会学科这 3 门课进行个别化教学，因为这个学生在这 3 门功课学习上存在困难。个别化教学首先运用到阅读课，2 周后运用到数学课，接着社会学科也采用个别教学。当个别化教学用于一门学科以后，在整个研究过程中都继续使用。在这项研究中存在着 3 个因变量：学生的阅读、数学和社会学科的成绩。每隔两周，研究者对每一个变量观测 4 次。从操作上来说，因变量就是由教师所确立的教学目标的实现百分数。

这一实验设计图解如图 6.10 所示。需要说明一下，对于 3 门学科都有一个为时 2 周的起始基线期。数学和社会科学的基线期分别持续了另外的 2 周和 4 周，它们各自把个别化教学引入其中。整个研究历时 8 周。

例 6.8 实验结果的模式 既然有 3 个因变量，而每个因变量要观测 16 次，那么就要获得 48 个有关被试的分数。因为每个因变量各有一组 16 个分数，每个因变量都有一个模式，所以可能有很多结果模式。这里不是列出 48 个分数，而是给出一个可能的结果模式，如图 6.11。

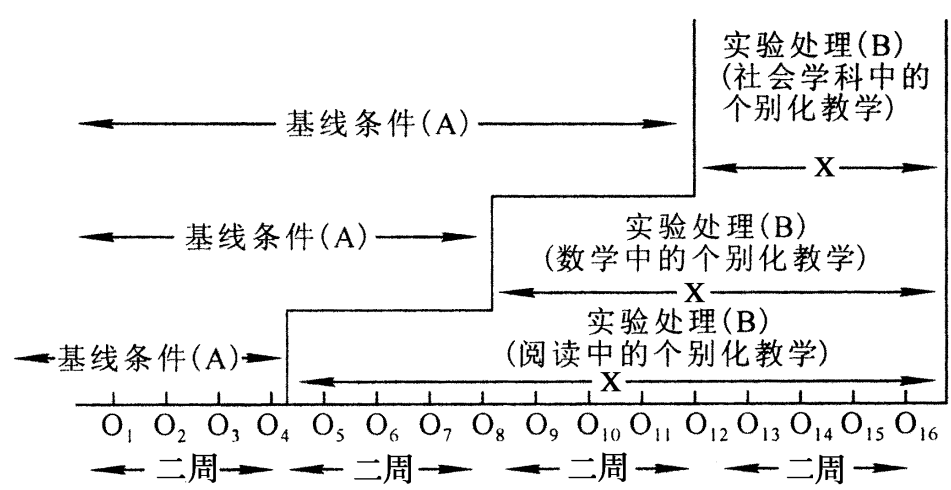


图 6.10 例 6.8 多重基线情境交叉设计图

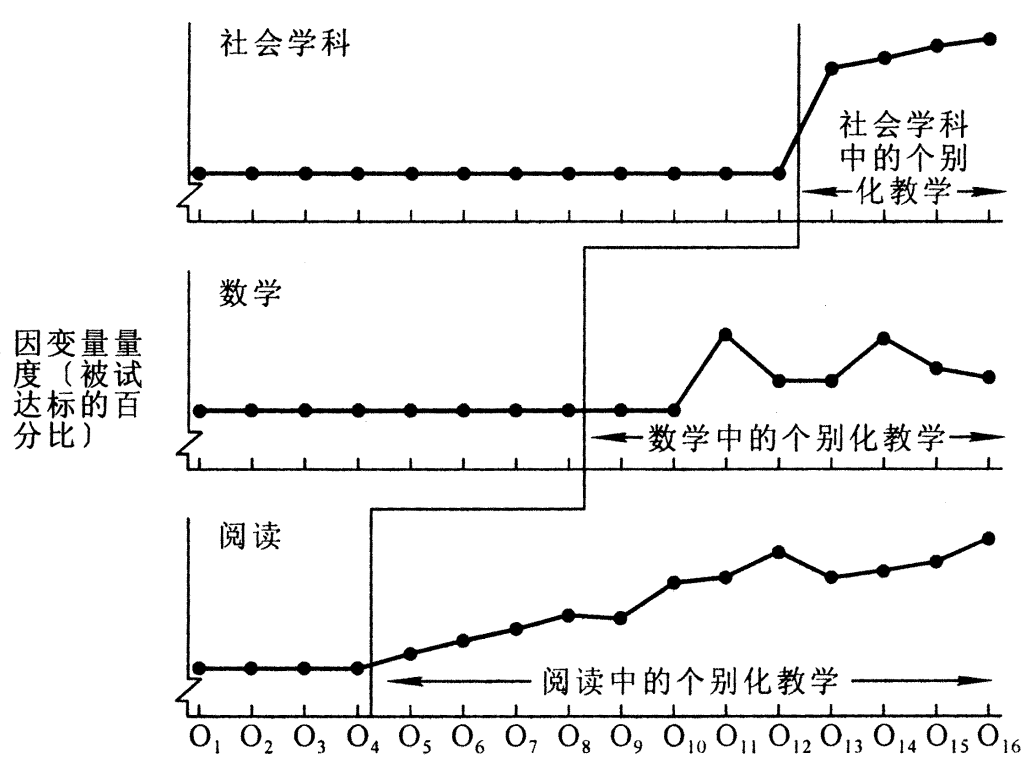


图 6.11 例 6.8 多重基线设计情境交叉结果模式图表

解释：实验处理（个别化教学）看起来产生了一种效果，但这一效果在各情境（被试各门学科）中并不稳定。在阅读课中，呈现出一种逐渐上升趋势，产生一种正效应。在数学课中，实验处理引入后，模式曲线极不规则，第11和第14个观测值的增加可能归因于一种外部变量；或许在这点上的教学目标，对被试来说简单而较容易达到。因此，不能总结说：这些增长归因于实验处理。对于社会学科来说，实验处理表现为一种迅速而持续的正面效果。概括地说，我们能得出结论：个别化教学对学生阅读和社会学科的成绩确实有效；这种效果对两门学科的作用都是正面的，但对阅读的作用是渐进的。如果说对数学也有效的话，个别化教学的结果可能是学生数学成绩的波动，也可能是数学

成绩有略微的提高。

多重基线设计一下子复杂起来了，而且实验结果的解释要求对结果进行大量的分类整理。依据所研究的变量，这些设计可以被改为包含取消条件（实验处理的撤除）的设计，这将使设计变得更加错综复杂。对于这样的设计，A—B—A 或 A—B—A—B 格式可能被采用。然而，因为其复杂性，这种设计一般仅仅在简单设计不能够完成一项特别研究时，才被使用。如果可以获得足够的内部效度，那么多重基线设计要比单基线设计会有更大的外在效度。这是因为在这种设计中包含了被试的多种行为、情境。

6 行动研究与准实验研究

在第 1 章中我们把行动研究定义为由地方教育工作者进行的研究，而没有把它作为一种独立的研究方法。行动研究可运用各种定量的或定性的研究方法，当行动研究涉及到实验时，它在性质上更接近准实验研究。教师、顾问、管理人员很少有随机地选择和安排被试以便进行实验处理的机会，更多的情况是，作为行动研究被试的学生已分到班中，或者学生作为研究被试是因为他们具有研究所需要的特征而不需要进行随机组合。例如，一个接受咨询服务的学生是因为他有咨询的必要，而一名教师用一种新的教学方法讲授历史课是因为他想了解新方法的效果，而班上的学生接受这一教学方法是因为他们是这个班里的成员，他们是因为某种原因而“自我选择”了这一班级，而不是随机安排的结果。

因为行动研究经常用于原始群体和个人，如果在研究中涉及到实验处理，这样的研究往往是准实验研究，大量的行动研究都是准实验研究。尤为值得一提的是单个被试研究设计，它是本章中其他研究设计的修改和延伸。单个被试设计对中小学的顾问来说是非常有用的，因为他们主要是同单一的个体进行接触，教师们旨在帮助个别学生的矫正教学和改进教学，也可以运用单个被试设计。当一个或多个教师对一个或多个班级进行教学法实验时，更多使用的设计类型是准实验设计中的仅施后测设计或前测—后测实验设计。

行动研究含有实验时，所使用的研究设计很可能是准实验设计。

总结

本章讨论了准实验研究，这种研究用于被试不是被随机地挑选和安排的设

计中。随机性的缺乏造成了确定研究效度的潜在问题。当我们使用原始的或自然形成的被试组时，有可能带来挑选的偏差，这样一来，小组间的相似性（或缺乏相似性）必须加以考虑。在单个被试设计中，被试的选择通常是因为某种问题或某种情景与这一被试有关，在单个被试设计中使用的被试几乎都没有随机挑选过。

尽管使用现成的小组或指定的被试可能会对效度造成威胁，但如果对研究的设计给以充分的注意，那么采用它们还是有效的。在教育研究中，不可能总是随机组合的现成实验组。然而，假如实验结果可以解释，那么使用现成的小组也能产生有价值的实验结果。

当我们使用两个基本点或两个基本点以上的现成被试组时，研究的信度依赖各个小组间相关变量相似的程度。相对于随机性来说，各组是非对等的，但它们之间的相似性是可以确定的。因此，适当地通过某种前测，掌握有关部门各实验组实验处理前的信息资料是很重要的。关于实验结果的可推广性要在同一逻辑的基础上加以论证。

当研究集中于被试个体而不是被试群体时，例如，在老师或辅导员对于个别学生进行实验时，我们采用单个被试设计。这种设计需要多次的观察和测量，基线条件和实验处理交叉进行。单个被试设计具有对施行实验处理的严格控制和对被试资料的广泛收集这些特征，然而，因为这些常常是在自然背景下进行的，所以干扰变量可能会产生影响。这种研究通常涉及大量的资料收集，同时对实验结果的分析可能需要大量的资料分类和整理。在图上标出这些实验资料对鉴别出显著的结果模式是有帮助的。

当含有实验设计的行动研究应用于学校时，本章论及的一些概念是相当有用的。行动研究通常在原始群体和个体中进行，该研究中作为被试的学生往往不是从群体中随机抽取，所接受的实验处理也不具有随机性。

本章和前一章介绍了实验研究和准实验研究，并且对此类研究中比较常用的设计进行了讨论。然而，为了适应具体的研究情况，可以对这些实验设计进行一些改动，例如，一个时间系列设计，我们采用随机安排的方式进行分组，并进行某种形式的反复观测，这样一来我们将形成一种“真”实验设计。重要的是使我们的设计适合实验的目标。只有对一项实验实行充分的控制，实验的结果才能得到令人信服的解释，并可按照我们的意图加以推广。

1 调查研究的范围和描述

前两章讨论了实验性的研究。在那两章里，一个研究方案就是一项实验，至少一个变量必须由研究者根据事先构想的计划予以控制。可是，实际教育情景中的许多变量是不易于操纵的。例如智力、性向和社会经济背景，不可能随意分布于不同个体，或进行实验控制。

一般来说，非实验的研究比实验研究更难控制。因此，对非实验研究结果的解释可能更少直接性而更多模糊性。但大量非实验研究是在自然条件下得出的结果，比在实验研究中控制一个或多个变量而得出的结果更具有普遍适用性。非实验研究通常是在自然情景下进行的，其中的许多变量同时发生作用。不过非实验研究也可精心设计，这不仅有助于研究的完成，还可以提高其结果的可解释性。恰当的方法论是由研究的问题及条件决定的。

调查研究可能是在教育研究中单独使用最广泛的一种研究方法。它涵盖了一个相当广泛的研究范围：从集中于探讨在自然情景下所发生的社会心理变量之关系的事后追溯研究的方法，到致力于确定现状与某些现象关系的调查的方法。调查过去常用作测量态度、意见或成就等任何自然情景中变量的数值。调查研究已有相当长的历史，它的发展在很大程度上被归因于社会学的贡献。由于调查研究在范围上极其广泛，列举一些足以表明它的特性的例子看来是有必要的。

调查研究是一种使用范围很广的研究，它包括对现状的研究，以及确定和解释社会的或心理的变量之间的关系的研究。

事后追溯研究

事后追溯意思是对已发生的事进行追溯。采用事后追溯研究，就是旨在追溯研究各种变量可能的关系和效果。研究者不可能对变量进行有意的控制。事后追溯研究历来有着不同的名称。不管把它叫做什么，事后追溯研究都可视为调查研究的一种类型。因为它强调非实验性的自然情景和数据资料收集的方式。从本质上讲，被试是“被调查的”。

事后追溯研究举例

在教育研究的文献中有着大量的事后追溯研究的例子，但是有些研究的报告太长而不可能在此作完整的介绍。下面介绍的是两项不太长的研究，以说明事后追溯研究的特性。

泰勒、弗莱和马罗亚麻（Taylor、Frye, and Maruyama, 1990）在一项题为“阅读学习与阅读能力提高”的研究报告中，对 195 名 5 年级和 6 年级的学生进行了研究。该项研究的目的在于调查学生在校和在家所花的阅读时间对于提高阅读成绩的效果。在研究开始之前先对学生进行一次阅读理解能力的测试，然后在 4 个月的期限内，从 1 月中旬到 5 月中旬，每天作阅读情况的记录。

阅读记录的信息随后进行整理，从 4 个方面，以每天所花的时间量进行统计：（1）在校规定的阅读时间（在阅读调查期限内），（2）在校业余阅读的时间，（3）在家规定的阅读时间，（4）在家业余阅读的时间。实质上，在这 4 个方面的阅读时间量就是研究中的自变量。然而，在分析的时候，自变量仅指在家阅读的时间量。因变量为在研究结束时进行的标准阅读理解测试的成绩。在研究开始前所测的阅读理解的成绩被用来作为分析、尤其是在统计处理时的控制变量。

这个研究报告是做得相当好的。它比较完备，它对过程的描述是明确和清晰的。研究的结论之一就是，“研究结果支持了传统的观点，即在阅读课进行真正的阅读，对于学生来说是非常有好处的。”

在一项对 1 388 名 4 年级学生的研究中，费恩和科克斯（Finn and Cox, 1992）对“四年级学生的参与与退缩倾向”进行了调查。主要研究问题如下：

1. 哪些特定行为使得教师据以判定某些学生是更积极地参与了课堂活动，而某些学生却具有更大的退缩倾向？

2. 参与程度在不同种族、不同性别和不同社会经济背景的儿童中是如何分配的？

3. 早期作业水平的差异是否使得一些儿童在 4 年级更多地倾向于参与课堂活动，而另一些儿童却较少参与？

该项研究中的 1 388 名被试是从田纳西州的 72 所学校中抽取出来的。

教师按照《学生参与问卷》——一种可以反映学生课堂参与程度的具有 25 个项目的调查表，来对学生个体进行分类。按照该问卷得分，学生被分为 3 组：缺乏参与者，被动参与者，积极参与者。这 3 类组别是研究中的一个主要自变量，另外，诸如种族，性别等作为附属自变量。

被调查学生从 1 年级到 3 年级的有关阅读与数学的成就测试的成绩已经获得。斯坦福成就测试和基本技能一级测试（由田纳西州所编制的课程参照测试）在每年春季也都要举行一次。上述这些测试的成绩作为研究中的因变量。在 1 年级末、2 年级末和 3 年级末都要对学生进行自我概念和动机调查，这些调查的得分可作为附属因变量。

对这些数据的分析是相当复杂的，而研究也会产生许多结果和结论。其中一项研究结果发现，参与组的学生在 3 个年级不同水平上的成绩均有高度的显著性差异。参与活动越积极，成绩水平越高。有兴趣的读者如需了解该项研究的更多的信息，可参阅有关报告。在这儿重要的一点是这些例子表明了事后追溯研究的性质。其变量是不能控制的；该变量在自然状态或学校情景中出现时可以测量。研究集中于确定变量之间的关系与效应。

在事后追溯研究中，当变量在自然情景中出现时，便可以对变量之间的各种关系进行研究。

其他调查研究

实际上，所有的成年人在这时或那时都会卷入一些调查之中，比如我们非常熟悉的民意调查。人们常在有关部门态度、观点、思想等方面接受调查，也常被要求去评价某项事物。不但一些较特定的群体，如某些职业群体，而且一般公众也都会常接受调查。

教育是调查研究中的一个重要领域。例如盖洛普民意测验中关于公立学校的公众态度的调查，这是一个全国范围的调查。地方学校也常做一些调查，这常称之为社区调查或学校调查。这些调查常集中于有关部门学校运转情况、社

区对于学校的看法，如何才能更好地运作等方面。学校调查也可设计用来估计将来的入学情况和学校系统的需求。学院或大学对它们的毕业生进行的调查，以获取对大学管理的看法和对整个课程的评价。教育专业人士也可能就一些教育问题进行调查，以表明他们的观点。评价性研究中也常包括一些对参与者或委托人的调查。因此，在许多教育研究情景中，调查研究可以说是一种相当合适的研究方法。

2 调查设计

一般来说，调查设计分纵向设计和横向设计两种。这些设计一般用于样本中，虽然它们也可以用在整个总体中（当调查运用于整个总体则称为人口普查）。区分它们的两个特征是：(1) 收集资料的地点，(2) 样本的性质。

纵向设计

纵向设计涉及随着时间推移收集资料的调查和在特定时间内及时收集资料的调查。一些纵向设计延续时间短，另一些则长，有可能是几年。但在任何情况下，这些数据都是在一定时间内分两次或多次收集的。

一种纵向设计是趋势研究。在此研究中，要对调查总体研究一段时间。要在该时间段内对不同人群进行随机抽样，而且，这些样本要不一样，但这些样本代表一般总体。趋势研究常常用于不同时间的态度研究。例如，在一个社区对学校态度变化的研究中，一般总体将可能是社区成员。每次测量态度时（也许是每年）都要从一般总体中随机选择样本。有的个体可能会不止一次地被选为样本。盖洛普对一场政治运动过程进行的调查，就是趋势研究的一个例子。

趋势研究是在不同时间内对研究总体的纵向研究，通常要从总体中取样，再测量随机样本。

趋势研究的一个变通形式便是群体研究，它也是一种纵向设计。在群体研究中，通常是研究特定的群体，而不是一般总体。通常是在一段时间的不同时刻进行随机抽样，而不是包括总体。趋势研究和群体研究的区别可以通过一个例子来表示。研究者对 A 研究地区的教师对待专业社团的态度感兴趣。在 15 年之中，每隔 3 年对该项态度调查一次。每一次都从教师总体中随机选择一组老师样本进行调查。总体中的成员可能和前一次不一样，至少部分不一样。但

在任何特定时候，它都是属于教师总体（在这种情况下，叫做一般总体）。这种方式的调查就是趋势研究的一个例子。

如果研究者对 A 研究地区从 1995 年以来的新教师对专业社团的态度感兴趣，这就牵涉到研究特定的人了。3 年之后，将会从这个群体中剩下的部分中随机取样，这部分人在 1998 年时会有 3 年教龄。尽管原来的一部分新教师可能已不从事教学工作，研究只包括 1995 年开始的新老师这个群体的态度。以这种方式进行的调查就是群体研究的一个例子。

群体研究是在一段时间内对特定群体进行的纵向研究。

在一些变化很快的群体中，群体的实际成员随着时间的推移会发生很大的变化。例如，在一所大学，每隔 4 年对本科生的态度进行调查，本科生的实际成员将会有很大的变化。但是，在这段时间中，每一个时段的本科生，仍然是研究的一般总体。

趋势研究和群体研究能够帮助研究者了解一段时间中的变化和过程。但是，每次收集数据时，都选择不同的随机样本。趋势是针对群体而言，而非针对个体。如果变化已经发生，研究者不能明确知道是哪些个体导致了这些变化。

纵向设计的另一个形式是“专门对象研究”。它是指在不同的时间，对一些个体样本进行数据收集。这些使用的个体样本叫做“专业对象”，它是在研究一开始就选定的。

专门对象研究是对同一样本进行两次或两次以上测量的纵向研究。样本可以代表一个特定总体或一般总体。

专门对象研究的一个优点是，不仅能够帮助研究者测出发生的变化，还能鉴别出变化是由哪些特定的个体导致的。专门对象研究还能提供关于变量之间时间次序的信息。如果研究想建立因果关系，这些信息就很重要，因为结果不可能先于原因。假设我们对大学教授对待管理中心提升模式的态度感兴趣，如果所有教授的态度都很好，这是由于提升，还是由于提升前的态度造成的呢？它对有没有提升会有一些影响吗？没有对发生的情况进行排序，就不可能建立因果关系（排序并非建立因果关系，它仅仅表明因果关系是否可能）。

专门对象研究有一些缺点，明显的一条是收集数据的过程中专门对象的耗损。因此，和其他纵向研究相比，专门对象研究延续时间短。它的另一个缺点是，对专门对象和研究者提出了要求，并要求研究者追踪并固定专门对象。如果专门对象所来自的集体变化大，最初的专门对象在以后的数据搜集中便不具

有代表性。专门对象研究对于短期的、静态的集体研究是可行的。例如，调查一学年中每季度学校董事会成员，就可以采用专门对象研究。另一个缺点是，专门对象会对某些变量越来越熟悉，这样一来，他们就会在回忆和反应上更加熟练。这种条件作用在其他方面也会起作用，使得专门对象变得疲劳、烦躁和粗心。

纵向设计通常是研究一段时间的变化和状态。一个特定的设计所需的时间和资料收集的数量，由研究的目的来决定。从取样的角度说，趋势研究是从整个总体中进行随机抽样，研究是从特定的总体中进行随机取样，而专门对象研究每次测量的是单一的随机样本。

横向设计

与纵向设计相对，横向设计是指对一个代表总体的随机样本，在一段时间内，进行一次性收集资料。横向设计不能测量个体的变化，因为它仅测一次。但是，在横向设计中，不同被调查的小组间的差异却代表了一个范围较大的总体的变化。请看下面的例子。

假定一名研究生正在对一个城市学校系统或地区学校系统的高中 1~3 年级学生的数学成绩进行调查。数学成绩在操作上定义为在综合标准数学测验中的成绩。我们选一个包括 1 年级、2 年级、3 年级学生的随机样本，每个个体都注明年级水平。另一种取样方法是从每一个年级水平进行随机取样，随后对样本进行测试，这样，研究者获得了所有 3 个年级的数据。

横向设计是指在某一时间从一个样本或者从一个以上代表两个或两个以上总体的样本中收集数据。

从两个或多个总体中同时选择样本，并进行与同一个研究问题相关的研究，就叫做“平行样本设计”。平行样本设计常常好像是横向设计，尽管它们也可能是纵向设计。在后一种情况下，它们将进行两次或多次数据收集，中间有一个时间间隔。

举例来说，在一项关于教师、学校领导和校董事会成员对待专业社团的态度 的研究中，采取了一个平行样本设计。这 3 种抽样被试所回答的态度调查表或问卷，如果不是相同的话，也是相似的。于是，不同样本的结果就可相互比较。

不同调查设计的特征概括于表 7.1 中。横向设计相对于纵向设计来说有一些逻辑上的缺点。与专门对象研究一样，数据的收集不必连续一段时间，对个

人进行跟踪发送信函也是不必要的。正是由于这些原因，横向设计比纵向设计在硕士论文和博士论文的研究中更为实用。但大多数纵向设计相对来说规模较大。纵向设计和横向设计的特点可以综合到一个复杂的设计中，这是针对从两个或多个亚总体（或总体）进行两次或多次取样而言。这样的话，可以把数据放在一个特定的时间内进行比较，也可以在不同的数据收集之间进行比较。

表 7.1 调查设计的特征

设计	研究的总体	如何取样
纵向设计		(两次或多次收集数据)
趋势研究	一般总体	每个数据收集时随机取样
群体研究	特定总体	每个数据收集时随机取样
专门对象研究	一般或特定总体 *	在整个收集数据的过程中，都用原始的随机样本
横向设计	一般或特定总体 还可能包括亚总体	从所有总体中同时随机取样

* 如果同时研究两个或多个总体，这就变成了平行样本设计。

3 调查研究方法论

进行调查的方法论包括一系列详细步骤，每一步都应当仔细计划。最初一步是阐述研究问题与开始形成调查设计。研究问题应当包括被研究变量的详细背景，无疑，它包括对文献的查阅。对调查变量必须作操作性解释，而且调查者应当具有有关的社会和心理变量相互关系的信息。这种信息对构建调查题目、选择测量手段，如测试，是有价值的。

如果决定进行抽样，而不是测量整个总体，下一步便是形成抽样的计划。需考虑多种多样的因素。必须确定抽样总体，以及抽样的单元。所抽样本必须能对总体或任一亚总体作有效推断。抽样程序可能是非常复杂的，而且获取样本需要大量的精力和资源。

虽然有些活动可以同时并行，但是，下一个主要步骤便是为收集资料做准备。由于调查包括访谈或问卷，这是一个重要步骤，因为需要设计调查的手段。当使用测验或目录编制收集资料时，诸如观察目录可能比其他的工具更可行。它对训练观察者和测试者，可能也是必要的。

无疑，在问卷题目设计的早期就考虑到确定问卷所产生的数据的具体类型是完全必要的。而且，考虑如何把数据进行列表、总结、分析也是必要的。用于分析数据的方法应予以明确。数据的定量分析可能被使用，或者可能采用描述或综合分析。调查所得到的数据必须能验证研究假设或回答研究计划中所提

出来的问题。特别是对于使用问卷和访谈法进行调查，有关部门答卷人的背景或人口统计的信息是重要的。因为在为了分析而划分变量时，它有利于分辨个体。例如，假如要对男、女的答案分别进行分析或要进行比较分析，那么，了解个体的性别便是重要的。调查随所需背景信息的数量而变。

最初的问卷或访谈应该“尝试一下”，也就是说要有先行的测试或接触。这样一来的尝试对象应与具备研究的对象相同。尝试的目的是纠正模糊、混乱或准备不充分的题目。尝试的反馈对完善工具有很大的用处。尝试一般在5~10名个体中进行，偶尔也有超过20人的。

当测量工具被认为是令人满意的时候，收集资料便可开始了。研究者应按照抽样计划来收集资料。如果使用访谈法，应当提供一些对访谈者进行系统检查的方法。这一点可以通过两个访谈者访问同一个体来进行检查。这种一致性检验被称为“可靠性验证”。对于有些类型的访谈，用可靠性验证对特定访谈者的一致性检验也是合适的。这可以通过对访谈录音和访谈记下两次相互独立的回答来完成。

资料，即对调查的反映，针对结果开放型问题，需要列表、综合。需要把答案归类，而且分类系统必须按有利于列表和综合的目的来构建。这一系统建立在答案的内容分析或先导分析之上。资料的转化被称为编码。如果来自于问卷或访谈的信息能够直接输入计算机，这一任务会变得很容易。

资料分析有赖于资料的性质，在一定程度上，把答案量化、进行统计分析是合适的。而且也可做定量描述。分析最终必须通过验证假设和回答调查所要研究的问题的形式来体现。如果推断是对总体作出的，那么，这种分析就应当是为总体提供的。许多孤立的分析通常是在单一的调查所得资料的基础上进行的，而且孤立分析与不同类型的分析一样，可能都是适用的。例如，对从实际信息的频率组成的题中得来的资料的分析与对从态度量表中得来的资料的分析不同。前者可能牵涉比例，而后者更可能牵涉频率或方式。那些来自于教师访谈的，有关他们对学生中社会互动的了解材料可能通过描述进行综合。进行调查的最后一步是准备汇报调查结果和结论。

图7.1总结了进行调查的方法的步骤。图的左边包括了几个主要步骤，右边则展示了在每一步骤下的活动。在有些情况下，这些活动重叠成两步。对于某一特定的调查而言，并非所有的活动都是必要的。如访谈人员的培训，仅仅对与访谈相关的研究来说，才是必要的。

成功地完成一项调查不是一件简单的事：有几种可能的误区和问题可能会对调查造成威胁，不能为各个步骤保证足够的时间和资源，如样本的数量，是一个普遍的问题。抽样的程序可能会被破坏，或者不可能有足够的资源对题目作出充分的测试和矫正。

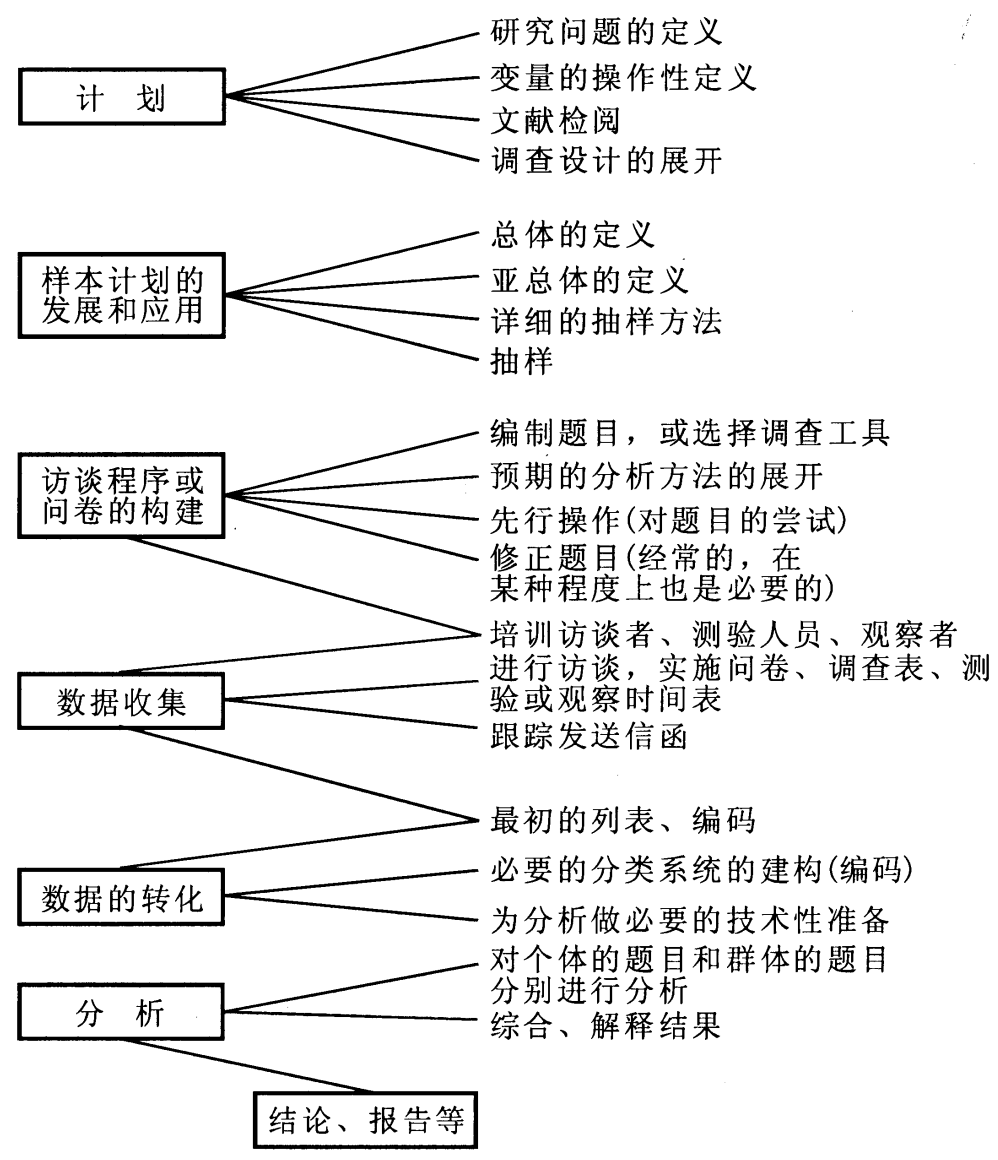


图 7.1 进行调查的步骤流程图

访谈或问卷所构建的题目可能是贫乏的，并会产生无用的资料。在跟踪发送信函上的失败，是一个非常明显但又很普遍的问题。当问卷返回的时候，集中资料和列表这一程序的不充分常常是调查无效和令人迷惑的原因。不考虑不回答的人可能使结果偏向一方，并导致无根据的结论。最后，如果研究者的报告是产生于孤立的、没有综合的单独分析，从调查中获得最大的信息量是不可能的。细心计划对于防止调查的严重困难还需走一段长路。

4 问卷调查

在所有的调查中，能把它们明显区分开来的特征之一是资料收集的方法，当然，邮寄问卷就是一种使用得相当普遍的搜集数据的方法。问卷被用来进行从地方到全国范围的各种调查，诸如对学校系统甚至是单个学校进行的社区调

查。问卷的长度及复杂性也会变化。但是，无论情况如何，牵涉问卷的调查都需要一系列连续的活动。这一过程早期一些活动可能是同时进行的，或与后面的活动相重叠。进行问卷研究的不少精力都集中在制定良好的题目上，以及使答卷人完成问卷上。但是，所有的活动都是重要的，而且要求注意细节。所有与研究相关的问题或假设都要清楚地界定。这将促进问卷题目的制定，防止包含无用的题目和遗漏必要的题目。

题目的制定

问卷的题目制定是一个简明的过程。但是，如不仔细注重细节，题目可能不能很好地排列在一起，不能为研究提供必要的信息。在讨论题目的形式前，我们先考虑一下题目制定的准则，它们是：

1. 除了少数几个要求提供背景或统计信息的题目外，其余题目要与研究的问题、假设直接相关。

2. 题目要清楚、不含糊，使用术语要使答卷人明白。避免使用模糊的、技术性的术语及行话。

3. 在一个题目中只准包含一个问题。违反这条准则的例子是：“你赞同对学生进行竞争性较低的测验和实施教师等级制度吗？”

4. 防止使用导向性问题。这些问题隐含着假设或期望的结果，这些题目常常暗含一种具有偏向性的回答。一个导向性问题的例子是：“你赞同在学校放松纪律，甚至认为纪律会损坏青少年的道德发展吗？”

5. 避免那些会对答卷人带来社会或职业压力的问题。总之，不要问那些让答卷人感到不满的问题。如问一个老师“你对维持良好的课堂学习气氛感到困难吗？”就是违背这一准则的例子。

6. 避免问那些个人的或微妙的问题。这些题目包括关于确定的收入及卷入违法活动的程度等。

7. 所提问题应是答卷人能够提供信息的问题。所有的题目都要与答卷人的信息背景相适应。

8. 要能使答卷人读懂题目。在可能的情况下，要用柔和的词语而不是生硬的词语。如调查教师时，应使用“纠正”或“纠正性行为”，而不是用“惩罚”。

9. 短一些的题目总比长一些的要好，简单一些的题目总比复杂一些的要好。宁可用两个或更多的短一些的题目，而不要用一个详细复杂的题目。

10. 当收集定量的信息时，要求答出明确的数量（如行动频率）而不是平均数。例如，可以问“在前两个星期，你帮助你的孩子做了几次功课？”而不

能问“在一个月內，你帮助你的孩子做功课的次数平均是多少？”

11. 题目的选择答案应当是可以穷尽的，选项应具有排他性。对有些题目，为避免强迫答卷人做不愿做的回答提供一种中立或中庸答案是必要的，例如“没有明确的态度”或“不确定”等。

12. 尽可能避免使用否定性题目和双重否定的题目。使用否定性题目的例子如：“下列教学策略你不用哪一种？”

这些准则可能有点太泛，但实际上它们是简明的。尽量使答卷人完成问卷的任务变得容易些，以使他们能够有效地完成，而且不致产生含混。

拟定题目是一项基本的工作。题目越简单越好

题目的格式

用作问卷的题目一般有两种类型：(1) 选择回答型或被动选择型题目，它要求答卷人从两个或更多的选项中选择一个；(2) 结果开放型题目，它要求答卷人自定答案。

选择回答型题目增强了调查对象回答的一致性。把数据列成表通常比较简明、耗时少。选择回答型题目可能存在限制答卷者回答广度的缺点，但是，假如选择回答型题目能够提供所需数据，这就不是一个问题了。主要是由于许多选择回答型题目需要涉及研究的主题、问题或假设，所以，充分地制定选择回答型题目通常需要更多的时间和精力。但是，这里所花费的时间是值得的，何况，在对回答内容作列表统计和解释时所节省的时间，足以弥补先前的损失。

在结果开放型题目的回答过程中，一些在回答选择回答型题目时没有出现的情感和信息可能会暴露出来，所以，允许答卷人做更自由的发挥。结果开放型题目的缺点是，在不同的答卷者之间，回答的时间长短和内容是不一致的。问题与答案都容易受错误解释的影响。在任何一份问卷中，对结果开放型题目的回答更比对选择回答型题目的回答难以列表、综合。在某种程度上，选择回答型题目应当用于取得相关信息的研究中。这些题目不需要花费答卷人太多的精力，因为这种题目不需要组织答案，而且在较短时间内便可作出。

题目格式的两种类型是，从两个或多个选项中选择一个的选择回答型和答卷者自定答案的结果开放型。

当然，结果开放型题目更具有弹性，而且比选择回答型题目的结构性差。但是，题目类型可以从没有结构到完全结构化之间灵活地变化，我们来看这样

一个题目，先从非结构性的开始：

你对数学尖子班怎样看？

一个相对的部分性结构化的题目是：

你对大学里与高级职业有关的数学尖子班的有效性怎样看？

结构性的题目则是：

你认为目前的数学尖子班应当是：

- a. 继续下去不要修改
- b. 继续下去但要修改
- c. 不要继续

在这一例子中，数学尖子班这一术语指的是一项被访谈对象所熟悉的特定班。

通常用在问卷调查中的选择回答型形式是李克特量表。李克特量表是一种带有顺序测量量度数字的量表。在一系列有关部门的答案中，一个答案对应一个数字点。答案以在点上打“ ”或在代表该点的字母上划圈的方式给出。当总计结果时，给这些点指派从1到5或从0到4的数值，然后把有关同一问题或主题的所有题目的得分总计起来。

有任意可能多组的李克特答案。一组答案的最重要的特征是答案与题目结合。下面便是两个这样的答案组的例子：

非常赞同	非常满意
赞同	满意
不确定	不确定
不赞同	不满意
很不赞同	很不满意

在后面讨论社区调查的时候，我们会给出有关李克特量表题目的例子。

先行性测试 在问卷的最后形成定稿之前，问卷的题目应在一小组人中做一个测验操作。这是问卷的试测，其间，那些依靠简单的检查不能明确的问卷的不足可能会凸显出来。该小组成员不必是预期反应者中的任意一个样本，但该小组成员应对研究的变量比较熟悉，并能对题目作出有效的判断。一个班的学生，也可能是研究生，常常就是一个有效的试验操作小组。试验操作的结果

应能鉴别出误解、含糊，及无用或不充分的题目。通过先行性测试，可能会觉得需要增加题目，数据统计上的操作困难可能被发现。完成问卷的指导语中的一些问题也可能出现。在试验操作结果的基础上，为问卷的最后形成作些调整是必要的。

除了排除模糊性和使指导语清晰以外，试验操作还可以避免没有提供信息或只是提供了极少信息的结果，下面一个题目是从 25 个回答者中所获得的实验操作结果。有下划线的数字为不同组别的回答频度。

请选择适当的一组（最接近整年限的），标出你的教龄：

0 ~ 10 21 11 ~ 20 3 21 ~ 30 1 超过 30 0

调查的结果只提供了极少信息，因为可选择的组类不能对回答者作出区别。可供的选择项可以改变一下，就会有这样的结果：

0 ~ 2 7 3 ~ 5 5 6 ~ 8 4 9 ~ 11 5 超过 11 4

现在选择的组类便提供了许多关于选择者教龄的信息。

题目的试验操作提供了鉴别令人迷惑、模糊的术语的机会，并获得有关可能结果的信息。

前言

前言在任何问卷调查中都是重要部分，它是向个体介绍问卷，并鼓励他们作答的工具。明确或隐含表示的报酬，都可在前言中提到。文字应直截了当，阐明调查的目的和潜在价值，并说明个体作答的重要性。前言中不要有任何能引起个体对调查目的或内容有所怀疑的内容。研究人员应向个体保证，他对所有回答都是保密的。

谁在前言上签名也很重要。如果前言上署上了与回答人有某些关系的人的姓名，回答率就会提高。例如，送给指导顾问关于指导机构的问卷，在前言上，可以签上管理该机构的大学领导人的名字。如果一名大学生发出一份问卷，在问卷的前言中只写明数据收集的目的，那他只能得到有限的或失望的回答。

问卷的格式

总的来说，调查问卷应具备吸引人和易懂两个特点。虽然彩色排版对答卷人的动机和所产生的影响还没有被证实，但还是可以使用的。问卷不应太长，

以免使回答成为一件艰辛的任务，且又花去了回答人大量的时间。其内容应在逻辑上相互联系，以便能激起答卷人回答的兴趣。

完成问卷的说明应是简明扼要的，对于复杂、困难的内容，可举个例子加以说明。如果有多种格式的内容，将格式相同的排在一起，内容的逻辑联系应优先于格式考虑。即，在完成问卷的过程中，回答者的思想是有逻辑和层次的，而不是“跳跃式”的从一个概念到另一个概念。

前言应指明回答的步骤。问卷的前部分应与所探讨的问题紧密相连。大部分问卷包括答卷人的背景和社会信息（如地位、职位、职称等），这一部分应出现在问卷的后一部分。不然，从前言到内容便被隔断了。这部分内容应当是选择回答型格式，且所给选择的部分应当是互为排斥、互不相容的。

所有开放型题目都应出现在问卷的后部，从某种程度上讲，这并不会打乱问卷的逻辑顺序。有时，一份问卷涉及几个概念，开放型题目则可以分散在整个问卷中。无论如何，问卷都不能以一组要求大段书写的开放型题目作为开头。如果某些内容对于答卷人来说比较困难，这些题目就应放在问卷靠后的位置。

问卷的排列不能给人以拥挤的感觉，各种题目也要易于回答。题目与页码应是标有数码的。问题和它的选择项应在同一页上，而且选择项要适合于该问题。如还需什么特殊说明，比如要求回答得清楚、明白，以便机器批阅，这些说明就要清楚地强调指出。最后，在问卷末尾，这份问卷交还给谁的姓名、住址要写清楚，提醒答卷人将问卷装入贴邮票的信封，密封后写上地址寄出。

问卷格式应能吸引人，并且一目了然，题目按逻辑次序排出。问题的回答应方便、没有混淆。

提高回答率的方法

问卷研究中一个持久的问题就是可能出现很高的不回答率。调查研究的有效性直接有赖于回答率和回答的质量。回答率是指答卷人送回问卷的比率，回答质量依数据的完成情况而定。既然回答者可能不代表要调查的总体，所以，正是不回答导致了偏见的可能出现。

不回答的可能性是问卷研究的一大问题，因为不回答可能导致数据的偏差。

无疑，回答率越高，调查越成功。许多调查也允许出现一些不回答的情

况。但是，什么是可以接受的回答率呢？制卷人在最低回答率上的意见不一。回答率根据调查的总体也可以有所变化。但是，一般来说，调查一类专业人群，最低回答率被认为是 70%。当调查一般民众时，因为不回答的原因可能是缺乏兴趣，所以允许出现更高的不回答率。比如一项关于学校工作的调查研究，对教师进行调查的典型回答率在 70% ~ 90% 之间，学生父母的回答率为 60% 以下，毕业生的回答率为 60% ~ 70% 之间。

实际上，如果回答人认为回答问卷的花费（时间和精力之类）相对于报酬来说是低的，他就会回答。迪尔曼（Dillman, 1978, P.13）提出，研究者可能提供的报酬有：

1. 被另外一个人认为是积极的；
2. 向回答者表示感谢；
3. 和回答者讨论调查的重要性。

当然，有时也会用金钱作为报酬，但由于资金问题，这常常是不可行的。通过后期分析金钱小费对于回答率的影响，认为一定的小费能够提高回答率，最高可提高大约 20%，小费越多，影响力越大。即使仅 1 美元的小费，也能导致提高约 20% 的回答率。附邮的小费比公开的小费效果更好。不管是对一般公众还是对专业人群的调查，都支持了这一结论；不管问卷的内容特征如何，结论也是一致的。即使在发送跟踪信函时，小费也能产生积极的效果。

我们有理由认为，一张富有吸引力的、看上去很专业化的问卷能提高回答率。实际上，有一项调查研究表明，问卷的外观有 80% 以上的重要性。虽然外观的作用并不一定都如此重要，但专业化的、大方的外观总是有利于提高回收率。

提高回答率的又一个建议是，邮寄问卷前告诉答卷人将与他联系。可以用明信片，或电话联系。在美国一项全国性的护士许可证的调查中，抽取了如下 3 组：（1）没有发给预约信的控制组；（2）收到了个人预约信的实验组；（3）收到了非个人预约信的实验组。后者没有写明内部地址和“亲爱的朋友”等类问候。结果控制组的回答率是 51%，实验组的回答率分别是 63.3% 和 64.2%。受到非个人预约性的实验组回答率最高。显然，预约信比单单邮寄问卷提高了回答率。预约与答卷人联系是一件额外任务，但是，如果预计回答率不高的话，还是值得做的，况且它还可控制样本大小。当然，调查对象的特性会对回答率产生影响。如果调查专业人群，并且内容和他们的专业有关，这时候的回答率要比调查一般民众要高。

有一项调查使用了 3 种时间暗示条件：（1）30 分钟；（2）15 分钟；（3）无时间暗示，要求答卷人在现场完成一份需 28 分钟才能做完的问卷。他们发现，回答率最高的暗示条件是 30 分钟那一组。因此，时间暗示提高了回答率，

但时间暗示需要和完成问卷的时间相匹配。由此可以推断，答卷人不会被一个不真实的时间暗示所欺骗。尽管减少完成问卷的时间可能提高回答率，但任何时间暗示都应是现实的。

跟踪发送信函 跟踪发送信函对于几乎所有问卷调查都是必须的。跟踪发送的信函应在前言所定最后期限过后几天送到答卷人手中。跟踪发送信函应事先有计划；在有些情况下，两次或两次以上跟踪发送信函是应该的。如果经济上允许，把一份问卷的副本，连同一张贴好邮票、写好自己的地址的信封，一起放在跟踪发送的信函中。有时，也可使用一些有提示的明信片。但如果是这样，它就应该在最初的邮寄完成后 10 天内发出。当然，我们不可能和明信片一起邮寄另一张问卷，这是一个缺点。如果问卷就在手边，而不是必须从最初的信件中寻找，个体更有可能答卷。

跟踪发送信函对于问卷调查，被认为是必要的，而且应事先做好计划。

有两种跟踪发送信函的方法：(1) 紧接着给那些没有回答的人寄一封信或明信片；(2) 给每个人都发一个空白问卷的信封。前者更受欢迎，因为它比较便宜，而且排除了从同一个人那儿收到两张已做好的问卷的可能。后一种方法只有当不能鉴别出不答卷人时才使用。如果使用后一种方法，那么就应该告诉答卷人，如果他们回答了第一次问卷，就不要做第二次了。

电话、电报或特殊的邮件也可以用作跟踪发送信函，但是，由于它们在时间和花费上都比较贵，因此，使用不是很广泛。可以连续不断发送信函，但是，随着跟踪发送信函的发出，不回答率应降低。如果预期的回答率是低的或没有一般那么高，连续不断地发送信函可能也不会起什么作用。

总之，为了提高回答率，建议研究者设计一张富有吸引力、简洁的问卷，而且应附有信息丰富的前言，提供适时的跟踪发送信函。如果要给出时间暗示，她们应当是现实的。如果把金钱的刺激作为一种选择的话，这是会有所帮助的。根据实证研究的结果，用来提高回答率的跟踪发送信函，金钱刺激等方法，都是很有效的。

鉴别不回答的原因

较高的不回答率所带来的困难是，数据可能具有偏差性。决不可以简单地假定，回答者是当初问卷调查对象的随机抽样。另外，采取增大样本数目的方法也不能抵消偏差，这仅是增加了数据的量而已。

如果产生数据偏差，他们就不能代表研究的群体。下面的例子是美国对一个州的数学教师所做的调查，正好反映了这一点。几乎所有规模较大学校系统的管理中心都作出了回答，但由于某种原因，那些较小学校系统的管理中心并没有送回问卷。因此，返回的问卷数据表明，在全州的大多数学校中，数学教师的总数已经超出了平均需要数。所以，不回答的结果就导致了一个数据差距，它明显地扭曲了现实。我们很难保证不回答是在调查群体中随机分布的。

人口统计的信息对于鉴别由不回答而造成的可能的取样偏向会是很有益的。而且，问卷中包括一些人口统计信息的题目，确实可以帮助确定不回答的对象。研究者应根据与研究变量有关部门的回答者的特征鉴别出不回答的原因。是不是某个亚群体——例如，男人、初级中学教师、郊区教师、小型学校系统的管理人员有一个较高的不回答率呢？通常，不知道谁交回了问卷的调查方法，是不会令人满意的。我们可以通过决定人口统计变量种类，事先鉴别不回答的对象，例如与小学教师相对的中学教师。

鉴别出具有较高不回答率的亚群体，并不能揭示答卷人的情感，但是，它却能鉴别出不回答群体。如果不回答是与答卷人对问卷的某种情感类型相联系——即不答卷人不喜欢它们，回答的样本就是有偏见的。因为回答与不回答和研究的变量有关。

为了获得有关不返回问卷者的特征和他们的原因的信息，与不答卷的样本进行访谈是可能的。虽然一个 25 人左右的样本通常已经足够，但检查不答卷样本总会投入额外的精力。根据经验，如果不回答率超过了 20%，就有必要与不回答的人核对原因。

在一些研究中，证明不答卷人在他们的情感或不感兴趣的问题上保持中立是可能的。如果这一假设能站得住脚，在解释数据的时候便可以应用它，而且在核对不答卷之前，不回答率较高是可以忍受的。这一假设在调查一个总的公众总体时，比调查一个专业总体可能更为有效。

不回答这一现象是不可忽视的，研究者必须知道不回答的原因，确定不回答的原因是很重要的。

不完全的和可能不真实的回答

尽管可能要求问卷回答者对所有题目进行回答，但常常有一些回答者会省掉一道或几道题目。被省掉的题目可能是该题目没被理解，也可能是回答者对题目没有把握。如果整套问卷只有少数几道题目被省略，并且对于这些省略似

乎没有什么任何统一的模式，那么省略几道题将不会带来什么不便。但是，如果省略频繁出现，那就值得对问卷作一些核实。

当要求回答者对某个人或某件事作评价时，作消极评价比作积极评价更倾向于频繁地省掉一些题目。因此，如果一些评价试题频繁地被省略，这可能要加以注意。开放型题目常被省略，可能仅仅由于回答者不想花精力回答，也许有某些别的原因。对此，可以对题目的回答进行内部核对。一种办法是把问卷分为两组，其中一组为被回答的题目，另一组为被省略的题目。接着按两组分别对相对题目的回答进行分析。如果得到的结论不一致，那么省掉题目就与回答者对这些题目的回答方式有关。

处理这些省略，没有任何始终如一的既定的方法。有可能是由于题目难于理解。调查人员应当对于省略掉的内容进行检查。如看上去值得注意，那就要找到省略题目的模式或者与其他变量的可能关系。例如，如果男性完成问卷都不回答某一道题（性别作为人口统计信息的一个题目），那么这个结论将值得注意。

省略掉题目是一件麻烦事情，如果有一些值得注意的省略现象出现，那么就有必要找到一些可以解释其原因的模式。省略实质上是不回答问题，它可能使结论出现偏差。

当就敏感性问题进行调查时，回答者可能会有不诚实的动机。例如吸毒是一个敏感问题的例子。要求获得有关部门非法行为或危害社会的行为的信息总是希望渺茫，这类信息许多只能靠自我报告而获得。以识别有关吸毒学生调查中的不诚实和粗心的回答者的方法为例，可以有4种方法：

1. 不可能的回答——回答者表示使用某种不存在的药物。
2. 药物使用的频率和范围难以置信。
3. 答案组合不可能——运用许多题目来进行题目之间的一致性检测。
4. 诚实性题目——要求回答者在回答题目时表现出的诚实性的程度。

制定一个捎带主观色彩的推测来剔除不诚实回答者的数据。例如，用以上4种方法中的任何一种皆可以识别出不诚实的回答者，那么他的数据就应被排除。或者，如果回答者在诚实性题目中表现出不诚实，那么这足以反映出他的回答是不诚实的。当然，抛弃掉数据会令人不快，但是，把错误的数据去掉而使一组数据减少，这与包含有错误信息的一组数据相比，则是一种改进。

有可能对大多数教育研究中的调查而言，极端地认为回答者会不诚实地完成调查是毫无道理的。这种情况，只有在对敏感性问题上作出回答会对回答者造成不利影响时才会出现。例如，涉及违反规定的行为的问题，诸如考试作弊

或非法行为。当陈述这些问题时，应当对不诚实和粗心的回答进行某种核对。

对问卷回答进行“诚实性检测”有许多方法。当在调查中要表述一些敏感性问题时，应当使用上述一种或多种方法。

举例

在这里，我们来看一些问卷题目的例子。第一个例子是一个对俄亥俄州私立学院中的咨询中心人员进行的假设性调查。研究问题的陈述、假设、以及相关题目已经在表 7.2 列出。在实际研究中，这些假设会包括更多的题目。同时，这里给出的假设，也仅是实际研究的大量假设中的少数例子。

表 7.2 研究问题、假设和问卷题目举例

研究问题：
对俄亥俄州私立学院咨询中心人员的业务准备与活动的研究。
假设（例）：
1. 顾问的基本作用是对学生的学业问题进行一对一的咨询。
2. (a) 大部分顾问（30 %或更多）在社会服务领域或专业领域拥有硕士学位。
(b) 大部分顾问在学院这一级工作经验不到 3 年。
3. 该中心工作量每周花在中心活动的平均时数超过 50（在全职基础上按比例算）。

这些假设隐含着中心功能，专业准备和工作量等问题。涉及到问卷调查，有时候提出一些相关的研究问题要比提出特定的假设更为方便。和以上 3 个假设相关的问题之例可以是：

- 1. 咨询中心的顾问之专业功能是什么？
- 2. 中心顾问已经获得的最高学位是什么？在学院水平上，顾问的经验如何分配？
- 3. 中心顾问的工作量怎样？工作量在专业功能之间如何分配？相关题目 (为每个假设或问题举 2 个题目的例子)。

假设 1 - 1 请确证你所理解的你们咨询中心的基本专业的功能是：

- a. 对与学业无关的问题实行集体咨询
- b. 对与学业无关的问题实行个别咨询
- c. 对与学业有关的问题实行集体咨询
- d. 对与学业有关的问题实行个别咨询
- e. 无以上情况

- 2 请给下面的功能排序，1 为最主要，5 为最次要

a. 集体咨询，b. 个别咨询，c. 为咨询做准备，d. 咨询后跟踪发送信函，
e. 准备与咨询有关的常规试卷

假设 2: - 1 确证最高学历:

——学士，——学术博士，——在专业领域，包括社会服务领域的硕士，
——博士

假设 3: - 1 指出每周花在咨询者身上的全部时数——直接和咨询者的联系:

——少于 10，——10 ~ 20，——21 ~ 30，——31 ~ 40，——41 ~ 50，
——多于 50

- 2 指出在一个具有典型意义的星期，你在活动场所从事专业活动所花的时数

社区调查 学区往往会发现社区调查——学区内居民对学校看法的调查是有益的。这种调查不仅提供了有用的信息，而且起到了协调公共关系的作用。当人们被问及他们对学校的看法时，他们常会饶有兴趣地抱有积极态度。社区调查可在整个学区的学校甚至私立的学校进行。整个社区都该被调查，或者调查仅限于在这些学校的学生的家长。不管怎样，问卷的页数要有限制，一般不要超过 16 开大小的两张纸。例如:

图 7.2 包含了一个社区调查的 15 个题目。这个调查问卷共有 35 个题目，其中 32 个的选择答案用李克特量表的 5 级量度表示，其余的 3 个项目为:

你是如何获得关于学校的信息的 (有 3 个选择答案，选出合适的项目)?

你有孩子在读 ABC 公立学校吗?

最后，是一个结果开放型的题目:

如果愿意，请你提出对学校的一些其他看法 (答题地方不够，可另附答案纸)

对这一问卷的一个特别说明是，对问题持中立态度有两种情况，一是被调查者对此项目不够了解；二是持不置可否的态度。如果有必要区分这两种情况，那么另一个选择项“不知道”可作为一个选择项填入表内空格。

委托调查 表 7.3 包括了从一个委托调查的第二部分中抽取出来的 7 个项目，这一部分是关于 AEL 的产品和服务的概况。这 7 个方面是:

表 7.3 AEL 的产品和服务概况调查表

1. 总的满意情况
2. 知识或技能
3. 信息分享
4. 在计划中的使用情况
5. 在实施中的使用情况
6. 指导所带来的好处
7. 实施时在次级客户中的使用情况

表 7.3 中的题目都是关于在计划中的使用情况。某一方面的题目在问卷中集中为一组，这样一来有利于回答者在完成问卷时思路的连贯性。这部分题目的指导语在表 7.3 中也同时给出。问卷前 3 个方面的 22 道题在这儿就省略了。

提示：下面是有关 ABC 社区学校的特征、活动及政策等项目的描述，每个项目都有从“极赞成”到“极不赞成”等 5 个答案供你选择。其中“中立”的答案表示你对该项目不了解或不置可否。

项 目	极赞成	赞成	中点	不赞成	极不赞成
1. 大学预备课程					
2. 在初等学校的学生应学习一些基本技能					
3. 近年来改进课程的努力取得了成功					
4. 7~12 年级学生的课外活动对他们来说是合适的					
5. 家政学是课程的重要组成部分					
6. 音乐是课程的重要组成部分					
7. 体育是课程的重要组成部分					
8. 职业教育是课程的重要组成部分					
9. 赞同对安尼中学职业教育比例的调整					
10. 赞成教育委员会 91 年秋实施的 4 年制高中 (9~12)、					
初中 (6~8) 以及 5 年级制小学的学制计划					
11. 对 ABC 教育委员会的工作基本满意					
12. 对 ABC 公立学校的中心管理基本满意					
13. 学校领导在管理学校上是成功的					
14. 高中的指导部门对学生的指导能满足学生的需要					
15. 初中的指导部门对学生的指导能满足学生的需要					

图 7.2 社区调查题目举例

有关 AEL 产品与服务情况的问卷

请就 AEL 的产品与服务的一般情况，而不是某个方案或项目的具体内容的情况来回答下列题目。这里产品或材料指的是任何印刷信息（小册子、业务通讯、工作手册等）和录像磁带。服务则涉及与 AEL 工作人员的联系，例如，

研讨班、电话交谈、技术指导与咨询等。请仔细阅读本说明和五点量表的等级。然后请回答这部分的所有题目。

1	2	3	4	5	圈选一项				
很赞同	赞同	不确定	不赞同	很不赞同					
23. 我曾在对方案/ 活动作计划时使用过 AEL 材料					1	2	3	4	5
24. AEL 使我接触到一些对我作工作计划有帮助的人们					1	2	3	4	5
25. 在对方案/ 活动作计划时，我认为 AEL 是我的第一手信息。					1	2	3	4	5
26. 在对方案/ 活动作计划时，我曾接受过 AEL 的服务。					1	2	3	4	5
27. 为了在作教育方案的计划时用得上，我积累了一些 AEL 的材料。					1	2	3	4	5
28. 我曾把从 AEL 活动中获得的技能运用于自己的方案/ 活动的计划之中。					1	2	3	4	5
29. AEL 使我接触到了一些对我作工作计划有用的其他材料。					1	2	3	4	5

5 访谈调查

对于进行人数较多的大范围调查来说，问卷调查的所需费用不大，但它也有一些缺点，诸如有不回答或粗心地回答现象。访谈是一种进行调查的又一有效方法，使用这一方法，相对于问卷调查来说，有如下几方面的优点：

- 1. 如果访谈的进行确定下来，就不会存在不回答的问题；
- 2. 如果必要的话，访谈可提供一种向深层探索的机会，以及陈述、理清题目的机会；
- 3. 调查的完成可以做到标准化；
- 4. 由于开放型题目获得回答，调查显得更加成功；
- 5. 访谈有助于避免题目回答的遗漏；
- 6. 有时不通过访谈，一些有关个人的数据就无法取得。

这里与上面 6 点相关的是，从那些有受教育缺陷的成人那里收集数据，需要进行访谈。因为即使题目明明白白地写出来，他们也可能由于缺乏回答的动

机而拒绝作答。

然而，访谈在时间和精力上所花的代价是昂贵的。近年来，电话访谈得到了大面积地使用，并作为面对面访谈的替代形式。电话访谈可以显著地降低调查的费用，同时它与面对面访谈相比，还有其他一些优点，这在以后要谈到。如果采取面对面的形式，访谈必须有计划，因为这涉及到两个人之间的互动。即使采用电话访谈，除非访谈非常简短，预先做好计划总是非常好的。

使用访谈比使用问卷有诸多好处，但在时间和精力上的代价高昂。

访谈题目

访谈题目像问卷题目一样，分为选择回答型与结构开放型两种格式，而且它们在结构程度上亦该变化。然而，非结构性题更可能用在访谈中，原因是非结构性题目给回答人留下了大量可解释的余地。在访谈中，这种可解释性能够控制、掌握，在问卷中，则是不可能的。一般来说，在应用访谈法进行调查的过程中，每一个访谈对象都要被问及同样的一组问题。为了使不同的访谈对象配合工作，措辞可能要做轻微地变动。例如，假如学生、教师、校长都是同一研究的访谈对象，各种各样的题目格式都可用在同一访谈中，而且访谈者还要控制格式的转换，以免发生混淆。

无论访谈项目是要求作出结果开放型的回答还是选择回答，它们都应当完全以问题的形式，用富有意义的、清楚的术语向访谈对象陈述。而且术语对所有的访谈对象都应有一致的含义。访谈题目应对访谈对象有充分的意义。有时，可供任意选择的措辞或调查内容会随同访谈题目一同给出，但这样做需小心从事。先来看下面这一结果开放型访谈题目：

你最喜欢设在这个社区的学校的哪一方面？

如果访谈对象犹豫不决，就可以给他提供一项可供选择的调查项目，比如：

是否对诸如学校的设备、教学质量、时间安排、管理等类问题都有兴趣。

如果使用了这一选择性的调查项目，那些犹豫不决的人所回答的内容就会和那些没有得到这一选择性说法的人的回答有所不同。至少那些犹豫不决的人通过提示而获得了更多的框架。一种较好的办法是为所有的访谈对象提供同样

数量的框架。比如：

你认为下列各项中哪些项目是本社区的学校对学生产生强烈影响的因素？

a. 教学质量；b. 设备；c. 时间安排；d. 中心办公室管理；e. 学校（建筑）管理

在提供上述选择性的结构性项目外，也可以请访谈对象排列其他项目。

盖洛普民意测验的题目

以美国为例，最为广泛的教育调查之一是一年一度的大众对公立学校的态度调查，它被称为盖洛普民意测验。阿莱姆、罗斯和盖洛普（Elam、Rose and Gallup, 1992）曾公布了第 24 年度测验的结果。这次测验是一次家庭内部访谈，它是针对美国的所有地区、所有类型的社区做出的，访谈在 1992 年 4 月 23 日至 5 月 14 日之间进行。全国性的样本由 1 306 位成人组成。

第 24 年度测验的例题如下：

你认为学前课程是否有助于帮助那些低收入和贫穷家庭的儿童在他们在校的青少年时期表现更好呢？选择项为 4 种：非常、有一些、不多、一点也不。

测验结果给出了各选择项回答所占的百分比，这些百分比数据是按照全国性样本中的 3 类被试——没有在校儿童的成人、公立学校的儿童家长、非公立学校的儿童家长分别统计的。紧接着这一问题之后是一个有关部门是否愿意为学前课程交纳更多税金的问题。与前一个问题一样，对这个问题的回答也作同样的分类统计。

有一个关于提倡种族/种性宽容的问题是：

在你看来，本社区的公立学校是否正在采取必要的步骤提倡在不同种性或种族背景的学生之间的相互理解和宽容呢？

选择项为 3 种：是的、没有、不知道。在这里，报告中的百分比如前述的问题一样，是对回答做同样的分类统计。然而，在紧接着的一个有关部门为促进理解与宽容而增加课时总量的后续问题中，对回答者则是按种族和年龄而分类统计的。

定期的民意调查都有一个关于学年长度的问题。在第 24 年度调查中的问

题是：

在一些国家，学生一年在校时间为 240 天，而美国只有 180 天。如果把你所在社区的公立学校的学年长度增加 30 天，而达到一年 210 天或 10 个月左右，你的看法如何？你对这种观点是赞成还是反对？

按照全国性总样本给出了回答的百分比，同时也给出了对于同一问题从 1982 年以来所有回答的百分比数据。

盖洛普民意调查还涉及到其他一些问题，并且提供了大量的相关信息。自从 1974 年以来，每年的盖洛普民意调查都要求美国人对公立学校按照从 A 到 F 进行等级排列。因此，除了提供当前的信息之外，对于一些问题，调查结果的表现形式是跨年度的。盖洛普民意测验可作适当的创造性的推广，对于进行社区调查也可提供资料和参考。

访谈的进行

任何通过访谈来收集资料的过程，都须对访谈的进行有所准备。为了进行确定的访谈，须对访谈人员进行训练，而且这些访谈过程必须“标准化”，以便回答者能够把访谈作为一件尽可能连贯、一致的事情来接受。由于访谈需要时间，所以，除非调查非常有限，要使用大量访谈人员。需要对他们进行训练，使之获得有关部门调查的特定知识。当带有问卷的时候，在访谈前应该预先测试，而且题目要修正到满意为止。需要训练访谈人员，使他们在一系列访谈过程之中，不同访谈对象之间都能做到一以贯之。

为了安排访谈时间，需要确定可能的访谈对象与访谈者彼此都方便的时间。访谈者应当有一张弹性时间表，以便对访谈对象来说随时都很方便、可行。例如，假如那些调查在白天不可行，把访谈集中在晚上和周末就是必要的。

在访谈时间安排好或访谈工作开始后，取得访谈对象的配合是有必要的。一封告知访谈对象研究情况的预约信在取得访谈对象配合方面是有效的。这封信不仅是通知性的，而且还能再次确认访谈对象，特别是那些当接纳一陌生人时为个人安全而担忧的人。应当让访谈对象了解研究的目的和他们贡献的重要性。访谈对象不应受到访谈与其后的数据使用的威胁。使访谈对象了解访谈，并对访谈感到适意，便是对加强合作做了不少工作。

因为访谈是一项社会工作，所以，访谈人员和回答者建立良好的关系是重要的。方法应是实事求是、有效的，态度友好但不能“好得像朋友似的”。要

确保让回答者接受信任性信息，而且回答者不应受到访谈问题威胁。如果回答者有问题，访谈人员必须知道值得深入探索的方法及允许阐述的程度。除非回答人直接对题目做了陈述及话题本身是搜求信息的一部分，否则不应游离主题。

应用在访谈中的资料收集方法应当为访谈者有效把握，以使访谈的进行不会受到干扰。一台录音机即可保留整个的对话，但是，访谈者在使用前应征得访谈对象的同意。如果访谈录音不实用，或行不通，就必须运用速记。结构性问题可能只要求打一个表示选项的“ ”，而对非结构性问题，则必须做简短而概括所有的主要论点的记录。数据的工作应当尽可能做得不露声色，不致引起访谈对象的怀疑。比如，假如访谈对象所做的回答简短，访谈者就不应写得很多。

访谈应准备好框架提纲，在友好和实事求是的气氛中进行，以便获得有效的信息。如果可能，应对所做的回答的精确性予以核实。

错误的潜在原因

尽管访谈员会尽量与访谈对象的情感和观念相协调，但访谈题目本身却不能保证对此种情感相协调。个体必须自愿地正确地作出充分的口头回答，困难在于，假如个体不具备足够的知识来回答问题，或是不愿意泄露有关部门的信息，就会发生问题。有一种回答不精确或不正确的倾向被称为“回答效应(response effect)”，如果存在这种倾向，那就是实际回答与正确回答之间的差异。访谈者必须能够识别理解错误和理解困难的不同，当场作出决定，进行深入地了解。

收集访谈资料时的潜在错误的原因，其中之一是，回答者在回答中的某种倾向，实质上产生了一种“回答效应”。回答者可能缺乏回答的动机，因为受到了访谈者的威胁，或者把他或她自身理解为名人来回答，提供的是社会上普遍接受的或专业上具有优先性的答案，而不涉及任何真实的情感。还没有一种方法论上的技巧能保证资料的精确性，但是，强调回答的真实性和建立初始核对是可能的。访谈人员要注意不可暗示，有无优先应予考虑的回答，有争议性的问题应力求避免，直到适合的背景和信赖的关系建立为止。在访谈的过程中，访谈者可能想知道回答者说的是否是真话，从而设置相关的问题来核对各种回答的一致性。换言之，访谈包含了要提出内容相同而形式与遣词造句及出现、地点不同的各种问题。

另一个可能的错误原因是，访谈人员对访谈对象的消极影响。消极的例子比如，如果访谈人员在访谈中心神不宁、居高临下地对访谈对象讲话，或者对访谈对象抱有成见就不能和访谈对象建立和睦的关系。访谈人员应尽可能地与访谈对象相互配合，这样才可能获得理想的答案。比如，如果访谈对象会因访谈人员是异性而对回答问题造成障碍，那就得确保访谈人员和访谈对象是同性。

第三个可能的错误原因与完成访谈的方法有关。在各种访谈之间可能不一致。采访时间太长，就会使回答者不堪劳累而感到厌倦、疲惫。访谈的最长时间应视回答者的特性及访谈内容的趣味程度而定。专业化的回答者的访谈时间可以稍长于普通回答者。一般来说，访谈过程不宜有其他人在场。采访的地点，应对双方都适宜、舒服而又方便。

访谈中可能的错误原因有很多，比如访谈对象的回答效应，访谈人员的消极影响，以及进行访谈时不一致、不适宜的方法。

对分叉性题目的评论

在问卷调查和访谈调查中都可能会包含有一些“分叉性题目”(branching items)。这类题目常要求依据对某一题目的回答而让回答者跳过一些题目。例如在一个地方性调查中有这样一个问题：

你赞成延长学年时间吗？

回答的选择项是：赞成、反对、不知道。如果回答是“赞成”，那么便有如下一个附加问题：

你赞成把学年延长 20 ~ 30 天吗？

如果对前一问题的回答是“反对”或“不知道”，则后一问题对该调查对象就是不合适的。如果一个正在进行的调查，分叉性题目很少引起困惑，那是因为被调查者能够跳到下一个适当的题目。因此，在问卷调查中，应清楚地指明要跳过哪些题目进行回答。例如：

如果选择“反对”或“不知道”，请从第 16 题开始回答。

即使这样，也不能确定是否所有的回答者能按照说明跳过适当的题目进行回答。这个例子也说明，访谈调查需要更好地进行资料收集的质量控制。

其他调查

许多教育调查的资料收集是通过问卷和访谈之外的其他方式来进行的。在这些调查中，资料的收集被高度“标准化”，而且为那些经常使用的已出版的测验和量表所控制。我们常常利用像班级这样一个完整的群体，从可能被称为“被迫观众”的人中收集资料。每年在每一个学校系统所进行的标准化测验，便是这样一个调查的例子。这一调查，基本上是为了测查学生的成绩水平。

在调查中的测量可采用任何形式。标准化成就测验以前已提到过。另外可以使用兴趣和态度量表。对于地方或区域调查的量表和测验都是具有结构性的。

6 调查结果的分析和评论

调查到的资料或数据可以使用适当的统计方法进行分析。在许多调查中，至少有部分的调查结果是以描述性的方式予以报告出来的。通常要提供对某个题目各选择项的回答的百分比或比例数字。前面我们讨论过的盖洛普民意测验的调查结果就是以百分比的形式报道的。这些百分比是按照回答者所属的不同类型，诸如年龄、种族、是否有孩子在校上学等，作分类统计的。

因为每次调查常包括许多的题目，因此最终的数据较繁多，这就需要一些数据分析方法。让我们回顾一下关于阿巴拉契亚教育实验室（AEL）的委托调查。调查的一部分内容曾列在表 7.3 中。表中的 7 个题目（23~29）构成了关于在计划中使用情况的分量表。尽管这里的每个题目是作为总量表的一部分进行测量的。但把每个题目的得分加起来作为分量表的得分也是可以的。得分高低是在最低 7 分到最高 35 分之间。得分越高，则反映了 AEL 的产品或服务对于其客户的计划的影响也就越大。从问卷的人口统计变量部分可看出，回答者可分为 3 类：一般雇员、雇主、受雇的专业人员。这些人口统计变量鉴别出调查对象的不同类型。实质上，这些人口统计变量可作为分析分量表的得分，即分析因变量的一个自变量。这样就可以在这些不同组别之间进行比较，正如在实验研究和准实验研究中对实验组和控制组进行比较一样。让我们考虑一下这项调查可能的结果与对它的解释的一些假设性模式。

假设举例 1

有 3 种职业角色类型：教师、中心办事员、高等教育的教授。假设这 3 组人员在计划中的使用情况的分量表上的平均得分分别为 28.6、29.2 和 17.3。分析表明，第三组的得分明显地低于前两组的得分，而前两组差别不大。我们可以由此得出结论：AEL 的产品和服务对于地方性学校的计划的影响，要比对学院或大学教学的影响大得多；对教师与中心办事员则具有同等程度的影响。导致这一结果的原因也许很多。大学教授也许在制定计划时没有接受帮助的需要；或许是 AEL 的产品和服务更适合于地方性学校，对大学则不然，或许还有另外的原因，或者是这些原因共同作用的结果。

假设举例 2

假设在计划中的使用情况的分量表上的平均得分按照回答者所工作的州：肯塔基、弗吉尼亚、田纳西和西弗吉尼亚，分别予以计算出来，那么工作所在的州便成为了自变量，有 4 种类型；而在分量表上的平均得分仍然是因变量。假设这 4 种平均得分很接近，均约为 25，那么对这个调查结果的解释就会是：在这 4 个地区，AEL 的产品和服务对于计划的影响是基本一致的。看来 AEL 对计划的作用似乎并不受地区上的邻近所影响——AEL 所在地西弗吉尼亚，与其他更偏远的一些州在这方面并没有什么不同。

假设举例 3

从个别题目中得到的调查结果也可以提供丰富的信息。假设对每个题目回答的平均数和标准差都计算了出来，但这些数据也只是对测试量表的位置和变量的“粗略”估计而已，因为单个题目只能是在一个有序的量表中得到的测试。得到的结果如下：

第 23 题，平均分 = 4.2 标准差 = 0.4

第 24 题，平均分 = 2.6 标准差 = 1.3

这一结果表明，AEL 材料被运用于计划中，与 AEL 提供帮助接触其他对计划有用的人员相比，前者得到大量的肯定，也产生了较小的标准差。然而，第 24 题的回答有较大的标准差，则表明了在这一点上的不稳定性。很显然，一些被试确实发现 AEL 帮助接触到了对计划有用的其他人员，而另一些被试却相反。

借助于人口统计变量、不同题目以及在各种影响项目上的分量表的得分，从诸如 AEL 委托调查中可以有各种不同的分析。研究者必须对这些变量进行分类以便决断哪些分析是最有用的，从而满足调查的目的。这里需要提醒读者的是，以上例子均为假设，不要把他们看成是现实调查的结果。

偶尔，也可以运用其他调查中的一组或几组题目。大规模量表调查中的题目通常具有很好的结构性和很高的可信度。例如，要做一个地方性学校调查，就可以借用盖洛普民意测验中的题目。调查结束后，可以通过数据库的比较加以分析，或者把当地的调查结果与全国范围的结果进行对比。

当在调查中使用到成就测验或态度量表时，通常要用平均数、标准分数和测量范围来描述分数的分配。如果使用很多题目来测量一个单一的概念或特征，就要结合这些题目的分数以概括出一个总体分数。变量之间的关系是可以被检验的，这种情况在事后追溯研究中常被使用。

调查结果常以表格的形式进行总结，这些表格包括一般的描述性信息。重要的是分析和随后的结果报告要与研究问题相符合。在设计调查表、进行访谈，或为收集调查数据选择工具时，对分析和报告的计划应在调查之初就做好。

调查结果常以描述性的方式进行报告，但任何适当的分析和随后的报告都要切合于调查的目的。

总结

调查研究无疑是教育研究中使用得最为广泛的非实验的方法。它被用于许多不同的情况以调查许多不同的研究问题。事后追溯研究也包括在这一章里，因为它在性质上是非实验的，并且在所使用的方法上与调查研究所使用的方法相一致。

我们常常把问卷和访谈与调查研究联系起来，之所以这样也许是因为它们的广泛使用。分发问卷、邮寄问卷法比访谈法使用得更多，是因为进行访谈有时间和精力要求。然而，电话访谈被发现其使用率越来越高。

问卷调查已经遭受了一些外在的批评，这部分是由于问卷常常制定得较差，也由于它们容易产生太多不回答的影响。一个设计良好的问卷调查要求有从研究问题的确定到数据的分析、解释等若干步骤。注意细节，使用强化回答的方法，尤其是跟踪发送信函，将会对克服问卷调查中的问题起到很大的作用。

访谈与问卷相比，有这样一个优点：如果准许访谈，而且访谈人员经过足够的训练，那就不会有任何错误或无用的数据。如果必要，访谈还可以提供进一步深入阐述的机会。但访谈代价高昂，而且需要训练访谈人员。由不同的访谈人员进行的访谈必须是一致的，有时我们称之为相互可靠性评价，而且在与同一访谈人员进行的若干访谈中，也应如此。后者被称为内部可靠性评价，即当访谈人与他或她进行两次或更多次的访谈时，访谈的程度应是一致的。先行性研究应在问卷或访谈的题目确定前进行。

我们讨论了调查设计，它基本上有纵向设计和横向设计两种类型。在纵向调查中，数据的收集要在一段时间的两个或更多的时间点上对总体进行研究。依据研究总体，纵向调查可以变化为趋势设计、群体设计和专门对象设计。横向设计常常涉及两个或更多的总体或亚总体，在同一时间点上进行数据收集。

成功地完成一项调查不是一件简单的事，有几种可能的误区和问题可能给调查造成影响。不能为各个步骤提供足够的时间和资源是一个普遍性的问题。抽样的方法可能会被破坏，或者说不可能有足够的资源对题目进行充分的测试和矫正。访谈或问卷所制定的题目可能是贫乏的，并会产生无用的数据。在跟踪发送信函上的失败，是一个非常明显但又很普通的困难。集中数据和列表这一方法的不充分常常是调查无效和令人迷惑的原因。不考虑不回答的人可能会使结果偏向一方，并导致无根据的结论。最后，如果研究结果的报告是分割的，缺乏综合的分析，从调查中获取最大信息是不可能的。尽管细心计划不能保证调查的成功，到目标还需走一段长路，但它对于一项成功的调查来说也是重要的。

1 定性研究的概念、意义和理论基础

在西方实证的教育研究传统中，量化的研究长期以来占了主要的地位。20世纪80年代以来，教育研究中的定性研究方法逐步在教育学术研究中得到广泛应用并成为教育研究中的一类重要方法。虽然定性研究反映出对定量研究的反思，但是依然体现出实证研究的精神和本质。本章主要介绍定性研究的概念和研究设计，使读者有一个基本的了解。

定性研究的基本特征

在教育研究中，定性研究是常见的。例如，通过对一所学校发展史的研究以发现这所学校成功的特点；比如通过对教育文献的比较，发现不同时期教育政策的变化；再比如，我们访问学生以了解学生的一些想法，都是定性研究。又如，著名课程学者杰克（Jackson, 1968）深入学生，和学生教师一起上课和交谈，写出了教育名著《教室中的生活》，提出了“隐蔽课程”的概念，这也是定性研究的成果。如果单纯依靠量的研究可能并不能发现“隐蔽课程”这一重要主题。

定性研究是指研究者运用历史回顾、文献分析、访问、观察、参与经验等方法获得教育研究的资料，并用非量化的手段对其进行分析、获得研究结论的方法。

实际上，在定性研究中，研究者本身是研究工具。定性研究有以下几个方面的基本特征。读者对以下几个特征的了解，会对定性研究的概念有进一步的理解。

定性研究是整体性的 定性研究者以整体的观点进行研究，他们不将研究

的场所、人们或团体缩减为变量来处理，而是将现场所有的人、事物看做一个整体。研究者运用这样的观点，采取多方面搜集资料的方法，以发现所有的研究角度和资料的脉络，描绘一幅教育现象的整体的图画。

定性研究者试图尽可能完全地描写研究问题所涉及的一个现象、一个场所、一个群体或一个文化，这样描写可能包括研究对象的历史、宗教、政治、经济和环境，整体的观点能促使研究者透视其当前的教育现象或文化情景。整体的观念可以随时引导研究者认识教育现象相互依赖的性质，其研究发现可能比部分的总和要大得多。

定性研究是“自然式的探究” 定性研究强调在自然的情境作自然式探究，而不人为地控制任何变量。定性研究者把教育研究看做是在自然情境中展开的。对事实、文献的整理也好，与被研究者交谈，观察某种教育现象和参与某种教育经验的过程也好，其基础都要求在一种自然的情境中展开。定性研究要求研究者在自然的情境中进行探索。

定性研究是“描述”的 定性研究的资料大多是文字形式的描述性（descriptive）资料，这些资料被称为软性的资料（soft data）。其内容包括：文本资料、现场记录、访谈记录、官方文件、私人文件、备忘录、照片、图表、录影带等，这些资料提供有关场所和客观现象的厚实的描述（thick description）。研究者搜集这些描述性资料时，注意了场所中发生的每个细节，假定场所中没有一件事情是琐碎或不重要的，并假定每件事情都可能是一个线索，以更广泛地了解所研究的现象。

定性研究注重过程 定性研究者关心的不仅是事情的结果，而且关心现象的过程。他们关心的是：人们活动的方式是什么？如何应用某些名词和标记？某些概念何以被看做是常识？活动或事件如何自然地发展？例如，在研究教师期望对学生的影响时，量化的研究着力显示了前测和后测结果之间的变化，定性研究则注意描述教师期望如何通过日常座位的安排、学生的发言机会、奖惩的分配和权力的赋予等师生间的互动过程影响学生的学习。而且，定性研究者在研究的过程中不断发现研究焦点。他们在进行研究之前，并不假定他们已经充分地知道研究的重点所在。

定性研究是“弹性的” 定性研究者在进行研究时是弹性的（flexible）和临场性的。研究的问题不是由操作定义的变量来界定，而是在现场进行研究时逐渐呈现而成的。研究者进入研究现场之后，才逐渐澄清和判断研究的有关意义，并在搜集资料的过程中发展和确定研究问题，而不是持着特定的待答问题或待考验的假说进行研究。定性研究计划、时间表、访谈议程、研究工具、或研究设备等，也都是在研究现场观察和探究时逐渐产生和确定的。

定性研究的探究程序是开放的和见机行事的，需要参照在现场情境所搜集

的事实持续地定义研究的方向。因此，定性研究者应是有弹性，能将新的情境看做是机会而不是困难。在研究过程中如果有不能预期的事件发生，必须愿意改变研究的程序或计划。定性研究方法不像其他研究取向能被确定和标准化，在定性研究中，有些研究的指引可以依循，但绝不是规则，方法可以服务研究者，但研究者不是程序和技巧的奴隶，定性研究者就像是一位艺术家。

定性研究是“人性化的” 研究者探究人世现象的方法会影响其如何看他们自己，当定性研究者探究某个社会人群时，是亲自地去了解他们，去经验他们在社会中日常生活。量的研究取向中常失去的人性特质，诸如：美丽、痛苦、信仰、挫折等，在定性研究中则能得到体现。定性研究者常能体察到人们的内在生活、道德努力以及人们的成功和失败。

定性研究者看世上每件事情都好像它们是第一次发生，没有一件事情是理所当然的。所有的场所和人们是相似的、也是独特的。他们是相似的，因为在任何场所和人群中，都能发现某些一般的社会过程；而他们又是独特的，因为透过每个场所和每个人，都能从中探究某种社会生活和人性的内涵和意义。就如人类学家虽然经常观察研究的是某一个场所或甚至一个人，但经过对这个场所或个人的风俗、人情、语言、仪式、或民俗知识等的深刻了解，人类学家对于基本的人性也能产生全面的洞察。

定性研究是“非判断的” 对定性研究者而言，所有的观点都是有价值的，研究者不是要寻求真理或标准，而是仔细地了解人们的观点，在定性研究中，所有的人都被视为相等的，少年罪犯的观点和法官的观点一样的重要。

研究者需有意识地放弃个人对任何特定对象或场所的价值判断，以避免对他们所观察的现象作出不适当或不需要的价值判断。但是所有人都是社会环境的产物，每个人的信念、偏见和喜好受社会化的影响很深，因此研究者也不可能是完全中立的。然而，定性研究者努力公平地看待其他文化或人群，而避免明显的偏见，是非常必要的。

为什么要选择定性研究方法

实际上，由于教育研究的对象的特殊性，量化的研究所反映的变量与变量之间的关系仍不足以全面地反映教育研究的特性。而且，在教育研究过程中，研究者往往不仅会碰到量的资料，也会有大量质的资料。因此，教育研究中不仅需要量的方法，而且也需要质的方法。具体说来，定性研究在以下几个方面相对量的研究具有自己的特点。

第一，量的方法主要依赖标准式测量方法，在有限和固定设计范畴中研究教育经验和现象，定性研究方法，使得研究者更仔细地研究所选择的论题，并

且不受先前决定的分析范畴所限制，实地（field study）研究更具深度、开放性，且详尽周密。

第二，量的研究方法之优点，在于它能够测量许多人对一些有限问题的反应，并运用资料的比较和统计方法，使得在具有普遍性的意义上发现教育研究的结果。这些结果往往用简洁和经济的方式呈现。相反地，定性研究方法会产生关于较少数人的丰富又详尽的资料，可增进我们对所研究个案和情境的深入而具体的了解。正因为如此，定性研究也会减低其结论普遍性的意义。

第三，量的研究的效度，取决于研究工具的结构。研究者首先要确定该工具所实际测量的是研究目标。而且，测量工具必须以合适的标准化的程序来施行测量。因此，量的方法的研究重点在于测量工具——调查问题、测验项目或其他测量工具等。在定性研究中，研究者即是工具（the researcher is the instrument）。这是定性研究的一个基本的观点，也是进行定性研究特色和难点所在。定性研究之效度，关键在于进行资料分析、实地工作的技巧、能力和执行其工作严谨性。另一方面，这一人类工具的独特性是建基于弹性、洞察能力等品质之上的，这些都足以弥补量的研究过程中的不足。

在定性研究中，研究者可更仔细地研究所选择的论题，并且不受先前决定的分析范畴的限制，在研究中更具深度和开放性，且详尽周密。定性研究方法会产生关于较少数人的丰富又详尽的资料，可增进我们对所研究个案和情境的深入而具体的了解。正因为如此，定性研究也会减低其结论的普遍性意义。

从上述分析可知，定性研究有一定优点，也有一些局限性。由于质和量的方法具有不同的优缺点，它们是可相互补充而非排斥的研究策略。质的和量的资料，可以在不同研究或者同一研究中得到体现，构成了教育研究的丰富而多元的内涵。

定性研究的理论基础

定性研究方法的重要来源，是与量的研究不同的理论传统和导向。教育研究及整个社会和行为科学，由于以不同的核心问题为焦点，而演变成不同的学说。研究方法的差异性，对于不同研究的学术传统，具有特别的涵义。同时，研究者如何研究世界，决定了研究者获得了什么有关于这个世界的知识。

因此，在讨论定性研究方法问题时，我们必须强调理论与方法的连锁（theory-method linkage），这也是本章最关注的内容之一。理论——方法连锁

之概念，意指理论筑基于方法之上。定性研究方法的重要来源，与有量的研究不同的理论传统和导向。定性研究，把研究者带进一个现实经验世界，以致结果和发现皆建基于这一经验世界。定性研究的理论传统主要包括以下几种。

民族志学 民族志学这个研究取向与众不同的，是该取向从文化观点来阐释和应用其发现结果。民族志学者应用了参与观察法和密集的实地工作研究过程。这种方法聚焦于惯常的例行公事，以及日常生活中的琐事。要求学者将其自身对现实的信念存而不论，来研究日常生活的现实。在一项方案或组织中去阐明日常生活中早已被视为理所当然的现实，会成为了解、改变和建立人们日常环境为基础的现实的一种力量。民族方法学的研究大量应用了质化的方法。

现象学 现象学的研究聚焦于下述问题：“对于这些人而言，其现象经验结构和本质是什么？”被经验的现象，可能是一种情绪（寂寞、嫉妒和愤怒等）；或可能是一种关系和教育事件；也可能是一项课程方案或一种教育文化。然而，现象学研究着重于描述人们所经验的事项，以及他们如何去经验这些事项。研究者可应用一般的现象学观点，以阐述教育经验背后的本质问题。因此，现象学的研究方法着重于研究者的个人经验和洞识。

具体说来，现象学的研究方法有以下几个方面的特点。1. 现象学鼓励研究者以较高的独立性来分析经验。2. 现象学强调对经验的结构作明确的描述。现象学呈现经验结构的精髓。3. 将人的本质保留于经验之中。

发现法 (heuristics) 是现象学研究法的一种，在较大的现象学架构之内，发现研究法有两个重点的元素。首先，研究者必须对所研究的现象具有个人经验和热切的兴趣。第二，研究中的其他协同研究者必须分享其经验该现象的强度。发现法并不探究偶然的经验，它聚焦于研究者深刻的人类经验；它结合了人类经验及其强度，以产生对现象本质的理解。因此，发现法所关注的是意义，而非测量；关注本质，而非表象；关注质，而非量；关注经验，而非行为。发现法说明且强调研究者是定性研究中最主要的工具；而且它挑战了传统科学主张研究者须客观和独立的看法。

符号互动论 另一项与定性研究关系密切的重要理论，是符号互动论 (symbolic interactionism)，符号互动论者所询问的问题为：“如何会出现共通的符号和理解，并为人际的互动赋予意义？”

符号互动论是社会心理学的学派，强调意义 (meaning) 和解释 (interpretation) 的重要性。把意义和解释作为基本的人类历程，以拒斥行为主义和机械性的刺激—反应心理学。人们通过彼此间的互动，创造彼此分享的意义，而这些意义则成为他们的现实。布鲁姆 (H. Blumer) 说明三项符号互动论的基本原理。1. 人类行动的基础是在事物对他们所具有的意义。2. 事物的意义呈现于一个人与其他人的社会互动中。3. 事物的意义，在个人应付所面临之事

物时所使用的解释历程（interpretative process）中被处理，并通过该解释历程而加以修正。

上述前提使研究者倾向于定性研究，以之作为了解人们如何知觉、理解并解释其世界的惟一真实的方法。只有通过开放的、自然的研究与归纳分析，来与人们保持密切的接触和直接的互动，符号互动论者才能了解被研究者的符号世界。

诠释学 诠释学（hermeneutics）是一个对定性研究有所指引的理论学派。诠释学所询问的问题是：“什么是人类行动得以发生、作品得以产生的条件，如何能对其意义进行诠释？”

诠释学的哲学，主要是关于对诠释理解或意义的研究，并特别注意脉络关系和原始目的。诠释学一词，意指希腊学者用来诠释轶闻传说、故事，和其来龙去脉的技术。为了理解并诠释一篇原文，诠释者必须要知道作者想要表达的是什么，理解其所指的意义，以及将该文件置于历史和文化的情境脉络中来审视。

2 定性研究设计

定性研究设计中的几个重要概念

读者在了解一般的定性研究设计程序以前，了解定性研究中一些概念是必要的。

研究问题 目的（purpose）是研究的出发点，有关设计、测量、分析及报告方面的决定均取决于目的。因此，研究过程的第一步是清楚研究目的。研究目的往往和研究类型有关。基础性研究的目的是为了获得知识，发现真理，从而对教育基础理论有所贡献；应用性研究是为了解决某一实际问题。总结性评价研究为了决定某一方案之效果，而形成性评价研究是为了改进某一教育方案。

然而，定性研究方法的重要来源，是与量的研究不同的理论传统和导向。第一种取向是“从问题到现场”，是指研究者在进入研究现场之前，对于想要研究的问题多少有些一般的概念或想法，这个问题可能来自研究者个人的兴趣、某个教育理论或其他研究尚未发现的主题。然后参照这个问题的想法，研究者选择他认为适合的场所。带着这个预拟的研究问题进入现场，但是对于研究问题和现场的适合性保持开放的态度，预期到他预拟的问题或初始的概念可

能并不适合该现场，而可能需要另外寻找研究现场。

第二种取向则是“从现场到问题”，是指研究者不先界定一个研究问题，再去找一个适当的研究场所，而是在日常的生活或工作中参与经验，在某一个研究场所中发展学术研究的兴趣和问题。

在定性研究过程中，研究问题和研究场所之间的顺序关系可分为两种取向。第一种取向是“从问题到现场”，第二种取向则是“从现场到问题”。

换言之，上述的两种取向，前者是研究者较倾向于对某个问题感兴趣，想了解这个问题是怎么一回事，而去寻找和选择研究场所；后者则较倾向于先对某个或某类场所感兴趣，想了解那儿究竟发生了什么事，然后在现场中发展研究的问题。然而，研究者不论是从研究问题出发，或是从研究场所出发，在研究进行的过程中，研究问题和场所总是需要相互切磋，发展彼此的适合性。

立意抽样 立意抽样也许最能说明量化研究和定性研究在方法上的差异。一般说来，定性研究集中深入地研究精心选择、数量较小的样本，有时甚至只有一个个案（即 $N = 1$ ）。量化研究方法一般选用随机选择的、数量较大的样本。两者不但抽样之技巧不同，抽样方式的逻辑基础也是不同的。随机抽样的逻辑，取决于选择一个随机性而又在统计上具代表性的样本。样本产生的结果会具有超越样本之外的有效类推。随机抽样目的是为了实现研究结果之类推性，使研究结果具有普遍性的意义。

立意抽样（purposeful sampling）的逻辑，在于选择信息丰富的个案作深度的研究。这些样本中含有大量对研究目的至关重要问题的具体资料。例如，一项研究的目的是为了研究一个教育方案在社会低阶层中之影响。如果集中深入地了解经过精心挑选的一小部分贫困家庭之需要、兴趣和鼓励措施，其收获会多于数量大而有统计代表性方案的标准化资料所提供的信息。因此，立意抽样之目的就是选择资料丰富的个案，以说明研究所关注的问题。立意抽样规模和个案数取决于研究之目的。表 8.1 罗列了立意抽样的一些类型。

定性研究集中深入地研究精心选择、数量较小的样本，有时甚至只有一个个案（即 $N = 1$ ）。定量研究方法一般选用随机选择的、数量较大的样本。两者不但抽样之技巧不同，抽样方式的逻辑基础也是不同的。

最后，就像研究的其他方面一样，样本规模是否合适也取决于同行的评论

与默契。我们在充分强调深度立意抽样之长处时，我们应该注意其不足之处，并谨慎地不从立意抽样样本中做过分的类推，这对减少人们对小量样本的顾虑是十分有益的。

表 8.1 立意抽样的几种类型

类 型	目 的
极端个案抽样	从研究现象的异常的表现中了解情况，如杰出的学生或者特别落后的学生等。
典型个案抽样	找出的样本能有力地说明现象。
最大变异个案抽样	记录为适应不同条件而产生的独特或不同的变异。
同质个案抽样	集中于同类对象的研究。优点是集中、减少变异，简化分析。
分层立意抽样	说明具体研究团体之特性；便于比较
关键个案抽样	可做逻辑类推。因为如果这个个案是这样，其他的案例也会如此，所以资料有最大限度的应用性。
标准抽样	选择所有符合某一标准的个案，如所有在某学校中父母离异的儿童。
滚动抽样	透过不断认识资料提供者，了解到研究个案的情况，滚动确定适合于研究的个案。
验证性个案抽样	阐明和深化最初之分析，寻找例子进行验证。
否证性个案抽样	寻找例外，检验变异。

试验性研究 有的研究者为了发展有关的研究问题或改进资料搜集的计划，而先进入一个场所，进行试验性研究 (pilot study)。

试验性研究可能比最后的资料搜集计划较为广泛，并且可能包含研究的实质内容和方法上的论题。在研究的内容方面，试验性研究的资料能提供研究者产生对有关研究主题的洞识。而在研究方法方面，试验性研究能提供有关现场问题和探究技巧的资料。试验性研究的报告虽然可能是备忘录的形成，但显示了研究者寻找到的研究方向或现场技巧，是很有价值的。对于初次用质的方法来进行研究的教育学者来说，在进行正式研究前设计一个试验性研究是必要的。

在研究的内容方面，试验性研究的资料能提供研究者产生对有关研究主题的洞识。而在研究方法方面，试验性研究能提供有关现场问题和探究技巧的资料。

定性研究的一般过程

定性研究的一般过程是：

第一，确定研究目的。

第二，确定研究的重心或计划。这个提议将更明确地指引研究者寻找有关资料的方向。

第三，分析的单位，即确定“个案”（case）是什么，如果这个个案是指某个人，个人即是研究的主要单位。例如，早期儿童的同伴关系等。如果再收集许多这些个人的个案资料，即是有关童年影响或同伴关系的“多重个案研究”（multi-case studies）。个案所指的也可能是某个事件、某个组织形态、某个小团体或某个地理区域。实际上，生活中几乎任何主题都可能被选作个案，个案决定了搜集和分析资料的范围。如选择一个学校可以作为一个个案研究其课程实施情况。

第四，实施现场研究，通过观察访问等方法收集资料。

第五，分析资料。即研究资料和研究目的之间的逻辑联结。

第六，解释研究的发现，撰写研究报告。

与量的研究方法不同，在定性研究过程中，从上述的发现研究问题等研究程序，并不是那么步骤分明、循序渐进的，并不是一条直线的行进方式，而是一个周而复始的循环过程。大多数的社会科学研究依循直线式的探索模式，这种模式首先确定一个研究问题，接着形成假设、作有关研究名词的操作定义，然后设计研究的工具、搜集资料、分析资料、作成结论，最后报告研究结果。其间的每个步骤都相当清楚，研究者通常都知道他们在探寻些什么。定性研究者趋向于一个周而复始的探索模式。研究者选择研究计划后，会一再地重复探索的过程：提出问题、搜集资料、作记录、分析资料、提出问题、搜集资料、作记录、分析资料……

定性研究者趋向于一个周而复始的探索模式。研究者选择研究方案后，会一再地重复探索的过程：提出问题、搜集资料、作记录、分析资料、提出问题、搜集资料、作记录、分析资料……

明确了这一点后，下面的关于研究的问题也许是可以参考的。

1. 本研究的主要目的和重心是什么？

2. 研究的广度与深度如何？
3. 要分析的单元是什么（个人，团体或方案）？
4. 使用何种抽样策略？收集什么资料？
5. 运用何种控制的方法（自然研究，实验设计，准实验设计）？
6. 采用何种分析取向（归纳，演绎，内容分析，统计分析）？
7. 如何描述结果的效度和信度（多元资料来源，多重方法或者多个研究人员）？
8. 何时开始研究？其顺序及阶段安排如何？
9. 如何处理道德论题及其他实际问题（保密事项，获得进入场所的许可，接触有关人员和档案的途径，如何接受技能训练，是否有耐心和毅力等）？
10. 可利用的资源是什么？

3 定性研究中的伦理问题

在定性研究过程中，特别是在搜集资料的过程中，经常触及伦理上的课题。例如：进入现场的暗中策略、尊重研究对象（如儿童）的特质等。在强调研究者与研究对象互动的定性研究过程中，伦理是一项需要认真思考的课题，因此本节进一步综合讨论定性研究的伦理课题。

定性研究者必须运用自己的常识判断，依照自身的价值观及对人性的看法，并以一颗敏感和设身处地的心态去作实践的抉择，在研究情境中逐步发展具体的伦理细节。

柏格登（Bogdan, 1983）在指导学生作现场工作的课程中，给予初学现场工作的学生们下列几个简明的伦理建议：

1. 非常小心地收放现场记录，确定不要将现场记录留在任何人可能拿到的地方。
2. 不要和其他人讨论或谈到研究内容或研究中的任何人。
3. 在现场记录和最后的研究报告中，使用研究对象或场所的假名。
4. 不要将所发现的任何资料，告诉可能用这些资料使研究对象困窘或受伤害的人们。
5. 和研究对象有个清楚的协议，让研究对象知道他们能期望从研究中获得什么，和需要完成什么义务。

然而，在这份简明的指引之外，定性研究者随时随地都可能触及无法预期

的伦理情境。有待研究者作反省与决定。

质的教育研究者面对的通常是学校校长、教师、学生、家长、或职员，当研究者进入一所学校或一个班级观察与访谈时，校长或教职员等对研究者最常说的戒心是：“他或她是想来评估我们办学与教学的好坏吗？”这种防卫心理是人之常情，研究者应能预期、理解、而予以包容的。

在教育的研究情境中，研究对象受到的伤害通常不会是身体上的，而是属于心理上或社会性的伤害，在这种可能发生的情况下，研究者在进行研究之初，就需要设想周密，与有关当局或人员有个正式的协议，伊里可森（Erickson, 1986, p.141）就举出这样的例子：

伊里可森和其他研究人员在一个都市学校系统中，针对西班牙双语教师们进行一年的观察和录影。这些教师们尚未获得正式的教师执照，每年仍须和学校当局重新签订教师合约。在这种情况下，伊里可森在研究开始之前，即先与学校校长和教育行政监督们磋商，达到一个正式的协议，其内容主要是：行政主管们不能要求观看研究者的观察记录，不能要求研究者谈论被研究的教师，也不能要求观看录影带而据以评鉴教师。录影带必须先让教师们看过，而且若没有经教师的同意，研究者之外的任何人都不能观看录影带。

这个例子显示了一个设想严谨的定性研究中，研究者如何努力试图兼顾科学研究和伦理责任。定性研究者必须运用自己的常识判断，依照自身的价值观及对人性的看法，并以一颗敏感和设身处地的心态去作实践的抉择，在研究情境中逐步发展具体的伦理细节。这样的伦理观念本质上也正反映了定性研究之根本特点，如关注具体的情境、磋商的过程、日常生活的脉络和常识性的了解等。

从本质上说，研究者所搜集的各种资料是否能呈现事实的真相，分析和解释资料时是否呈现了对象的本义，也都涉及研究的伦理课题。保护研究对象的最重要的伦理责任应是尽力地搜集资料和仔细地分析资料。避免误解或扭曲研究对象的观点。

总结

定性研究是指研究者运用历史回顾、文献分析、访问、观察、经验参与等方法获得教育研究的资料，并用非量化的手段对其进行分析、获得研究结论的方法。

在定性研究中，研究者可以更仔细地研究所选择的论题，并且不受先前决

定的分析范畴的限制，在研究中更具深度和开放性，且详尽周密。定性研究方法会产生关于较少数人的丰富又详尽的资料，可增进我们对所研究个案和情境的深入而具体的了解。正因为如此，定性研究也会减低其结论的普遍性意义。

在定性研究过程中，研究问题和研究场所之间的顺序关系可分为两种取向。第一种取向是“从问题到现场”，第二种取向则是“从现场到问题”。

定性研究者必须运用自己的常识判断，依照自身的价值观及对人性的看法，并以一颗敏感和设身处地的心态去作实践的抉择，在研究情境中逐步发展具体的伦理细节。

1 文献和文献研究法

文献研究法也是教育科学研究中运用最基本的方法之一，是专门对人类历史长河中所收集的文献进行分析研究的方法，因其不直接参与和接触具体活动，故称非接触性研究方法。

随着历史的发展，文献一词的涵义得以扩展。朱熹曾注：“文，典籍也；献，贤也。”明确指出了文献即是对贤者及其思想学说的典籍记录。后来在我国的《文献著录总则》中，文献一词的涵义扩大了，它包含了记录有知识的一切载体，即把人类知识用文字、图形、符号、声频和视频等手段记录下来的所有资料，既包括如图书、报刊、学位论文、档案、科研报告等书面印刷品，也包括像文物、影片、录音录像带、幻灯片等实物形态的各种材料。

文献是记录已有知识的一切载体，是把人类知识用文字、图形、符号、声频和视频等手段记录下来的所有资料。

文献研究法就是对文献进行查阅、分析、整理并力图找寻事物本质属性的一种研究方法。它有广义和狭义之分。广义的文献研究法既包括定性研究，又包括定量研究。狭义的文献研究法仅仅指定性研究。本书从广义的角度来理解和阐述。

文献研究法是对文献进行分析研究的方法，有广义和狭义之分。

文献研究法不等同于历史研究法。在历史研究中，必须运用文献研究法来研究历史文献。但是，文献研究法绝不限于历史研究领域，它既可作为一种单独的研究方法运用于其他学科，同时也可作为其他研究方法的基础。它通过对文献资料进行理论阐释和比较分析，帮助研究者发现事物的内在联系，找寻社会现象产生的规律性。

2 文献研究法的定性分析和定量分析

文献研究法主要有非结构式定性分析方法和结构式定量分析方法，它们各自从不同侧面对文献中所包含的信息进行加工和整理。

非结构式定性分析

通常说来，文献是对有关事物性质、功能和特征等方面的描述，这些定性描述较少展现研究主题内的变量关系，因此研究者往往倾向于应用逻辑推理探索事物之间的逻辑关系，而不是它们的数量关系。对文献进行定性分析研究成了教育研究者最常用的方法之一。

教育研究者常常倾向于对文献进行定性分析，应用逻辑推理探索事物之间的逻辑关系，而不是其数量关系。

定性分析一般是对文献中所包含的信息进行分类，选取典型的例证加以重新组织，并在定性描述的基础上得出结论。文献的定性分析在辨别过去的趋势并用该信息去预测与此相关的未来模式方面，具有特别的价值。与定量分析相比，定性研究有其显明的特点：

首先，它不太注重文献资料的数量特征和完整程度，虽然它不完全排除在研究的过程中进行一些简单必要的数量分析，但更注重的是对文献的性质作出分析研究，目的在于探索事物的特殊性和规律性。所以，只要该文献中包含着所研究课题的重要思想和内容，便满足了研究的要求。

其次，文献的定性研究的一个重要方面是个人文献研究，包括对信件、日记的分析研究，注重的是文献的个案，允许从事研究的人员选择足以证明某一观点和问题的例子。例如，美国著名社会学家托马斯和兹那尼茨基在1918年作了一次经典的个人文献研究。他们的文献资料是在美国的波兰移民和他们在波兰的亲属之间互通的大量个人信件。通过分析研究，他们第一次提出了对农民信件的分类型，并得出结论说所有的农民信件都是一种基本类型即问候信的变异。

第三，从某种意义上说，文献的定性分析类似于观察研究，因为许多文献

具有叙述的自然性、亲切性和发自内心的感觉等优点，所以在研究时比较灵活随意，研究过程的规范程度不太高。一般说来，教育研究者在样本抽取时，可选择一个小样本或个案进行研究，尽量避免那种只能从表面形式进行研究的大样本。而且，所选择的小样本或个案是根据研究者自己的特殊兴趣主观确定的，而不是随机抽取的。

文献的定性研究有不太注重文献资料的数量特征和完整程度，注重个人文献研究及根据研究者自身兴趣和课题要求选择小样本或个案等特点。

但是，我们不能不意识到，文献的定性研究难免具有主观性。尽管许多社会学家认为，托马斯和兹那尼茨基的定性个案研究方法为开展关于生活的各个方面，尤其是人们内心精神生活的透彻研究提供了极好的机会。但也有许多倾向定量分析或统计研究的社会学家明确提出，援引这些例子也不能构成科学的证明，这种研究方法是十分主观的，特别是因为调查人员可能对人们的生活史作了有意识的选择，以便达到他们自己的目的。由于定性分析存在着以上明显不足，因此我们在文献研究时不能仅仅满足于对文献的性质、特征和属性作出描述和分析，揭示其逻辑关系，还必须对文献中各个因素间的数量关系作出研究，即进行结构式的定量分析。

定性分析难免主观性，所以定量研究就成为一种必要的补充。

结构式定量分析

定量分析又叫内容分析，是对明显的文献内容作客观而有系统的量化并加以描述的一种研究方法。定量方法有助于确定一个过去时代的事实。现在，它已成为教育研究领域内一种重要的文献分析方法。特别是随着电子计算机的普遍应用，文献的定量研究越来越广泛。可以说，定量分析帮助提供对过去学校真实情况的描述。

1989年，提兹在研究澳大利亚公立高中时，收集了维多利亚州1946~1985年间每一所学校的入学人数。然后，在报告这些数据时，他采用了图表百分比的形式说明这些年的升学率，使用了百分比、比例等描述统计方法。同年波尔曼和马格在研究教师受教育的经济回报时，采用了一些更为复杂的定量手段。

定量分析又叫内容分析，是对文献内容作客观而有系统的量化并加以描述的一种研究方法。

定量分析的实质是将用言语表示的文献转换成用数量表示的资料，有助于使用正式的假设、科学地抽取的大型样本及计算机等现代统计技术对文献作出分析研究。在教育研究中，对各类教育文献资料，如论文、调研报告、会议文献、专题论著及各种教科书、教学计划、教案、课堂实录、学生练习等作出定量分析，从不同角度作多方面研究。

定量分析主要用于趋势分析、比较分析和意向分析等诸多方面。

趋势分析 定量分析可利用同一对象在不同时期内容资料量化结果的比较，分析某种现象或思想观点的发展过程、演变规律及今后趋势。例如，通过收集某一时期我国教育学期刊有关课程政策的文献，作出内容分析，可以发现我国课程政策的变化过程及趋势。

比较分析 定量分析可通过对同一中心问题但对象或来源不同的样本量化结果的对比，对不同地区、学校、团体和个人的教育思想、教学效果及工作方式等进行比较。例如，比较不同国家和地区各类教育问题的异同，比较两个学术流派学术观点的异同，比较不同学校、不同教师教学方式和教学效果的异同，比较两种不同教学计划的差异，比较两个不同学校学生学习成绩、学习态度、思想品德状况的差别等等。

意向分析 定量分析可通过某一对象在不同问题或不同的场合所显示出来的文献资料进行内容分析，研究该对象的意向。例如，分析某优秀教师的工作经验特点，分析不同类型教师、学生的心理状态、潜在动机和价值观念，分析某一时期不同对象在教育问题上的关注焦点，分析某教育家的教学目的、动机等。

定量分析的主要用途有趋势分析、比较分析和意向分析等。

定量分析具有明显性、客观性、系统性和量化等方面的特点，正因为它拥有对大量文献进行系统结构分析的优点，因而可弥补定性研究缺乏系统和确切的不足。

文献研究的步骤

从本质上说，任何研究没有惟一的、确定不变的方法，文献研究也是如此，因为各个研究在研究中所使用的方法不尽相同，有些人对文献的探索往往要等到所有资料被穷尽后才开始组织和解释，另一些研究者则从一个或多个方

面不断接近所研究的问题。况且，每一种研究方法还会出现许多变化。

尽管如此，为了增加对文献研究过程的理解和把握，我们还是从方法论的角度把其分成四个相互衔接的环节和实施步骤，但必须牢记这些步骤之间可能会有很多重叠，而且在研究中这 4 步常常紧紧地连在一起。

文献研究可分解为 4 个相互衔接、紧密结合的环节和实施步骤，即确定研究问题并拟定研究计划、收集和评价文献资料、综合分析文献内容和形成结论。

确定研究问题并拟定研究计划

研究问题的确定及研究计划的拟定，是任何研究的第一步。通常，研究者必须首先明确其研究目的，例如，两位美国学者在 1989 年确立的一个关于中国教育的研究题目是：“革命的反思：中国 1985 年教育改革的思考”。他们在报告的第一段便清楚地说明了研究目的：

本文的目的是阐发对作为变化了的形势结果的教育改革的看法。本文对改革发生的背景进行了详细的探讨；分析了导致教育改革的因素；总结了这次改革的关键特征；批判性地分析了它们对学校系统、高等教育和教师的影响。

虽然文章没有列出所研究的问题，但我们从其目的的陈述中可知其研究的问题至少包括：

- 1. 1985 年中国教育改革的动因是什么？
- 2. 这次教育改革的主要内容是什么？
- 3. 这次改革的教育背景是什么？
- 4. 这次改革对学校系统的影响有哪些？

在研究问题确定以后，同样重要的是针对所要研究问题拟定研究计划。研究计划是研究工作的蓝图，它指导着研究者按一定步骤在各个层次上秩序井然地展开研究。文献研究工作计划主要包括研究的目的和意义、研究的主要内容和阶段、收集文献的途径和方法、研究工作的进一步安排和时间分配、研究人员的具体分工、研究经费的预算、研究成果的形式等。

收集和评价文献资料

文献研究法是以文献资料为研究对象的一种研究方法，因此文献的收集和

评价是文献研究过程中一个重要的环节。全面、准确、迅速地收集真实可靠的文献，是决定文献研究质量的关键。

收集文献并不仅仅是收集似乎与研究问题相关的能够得到的文件，文献研究的基本原则是：只要有可能，尽量使用第一手资料。

第一手资料是指原来的或该事件或经历的首次记录，第二手资料则是至少一次被援引的关于该事件的叙述。例如，在研究 19 世纪英国教师和社区成员之间的关系时，当时的教师所写的与社区居民共同生活的记录是第一手资料，而其他人对该问题所写的报告属第二手资料。同样，约翰·杜威的著作是第一手资料，而由他的学生所作的解释则为第二手资料。

第一手资料是第一次关于要研究的事件或经历的描述，第二手资料则是对事件或经历至少处理过一次的资料。

在对文献资料进行收集的过程中，还必须作出批评性评价，这一评价包括两个方面，其一是外在批评，其二是内在批评。

外在批评 文献研究的外在批评在于评价文献的有效性，它是确定文献效度的工具，所要回答的问题是文献是否真实、可靠。

在确定文献的效度时涉及到许多方面的因素，有时任何一个方面都有可能导致文献无效。从记录文献的角度看，作者在事件发生背景中的地位非常重要。作者是否处在能够写出有效文献的位置上？作者是现场观察者吗？时间和地点等因素是否与已知内容相符？以第一手资料形式出现的文献是否为代笔人所写？

文献研究中的外在批评在于评价文献的有效性，包括文献在哪里产生？什么时候产生？由谁记录的？

内在批评 文献研究中的内在批评在于评价文献内容的意义和可信程度，强调的是文献内容本身。有时，内外批评之间会有交叉。

内在批评在检验文献内容、确定文献真实性程度时，作者是一个重要因素。例如，作者是否有什么偏见？一个对现行教育政策持反对意见的人可能比赞同者更多地强调不利因素。同时，对作者的写作风格和技巧进行适当分析也很重要，作者写作时大量引用了他那个时代已有的文献吗？作品是比喻性的还是真实事件的记录？

文献研究中的内在批评在于评价文献内容的意义、精确度和可信程度。

内在批评和外在批评在收集文献资料过程中非常重要，它们从不同角度确定材料来源的可信度和可利用性。如果材料来源不真实，就不能被运用。即使内容真实，但与研究问题无关，也是无用的。

综合分析文献内容

研究者从图书馆、档案馆、资料室等地方获取大量资料，并对这些资料作出批评性评价以后，便开始进入到综合分析阶段，这是一个对已经收集并经过评价的文献内容进行整理、描述、分析，并从中得出研究结论的过程，具体分为非结构式定性分析和结构式定量分析两种。

文献研究的第三步是指对已经收集并经过评价的文献内容进行整理、描述分析并得到结论的过程。

形成结论

文献研究方法的最后一个步骤是作出与研究问题相关的结论，这在很大程度上依靠对从文献中得来的信息的逻辑分析。

在最后一个步骤中，结论形式、解释是非常重要的，对最通常的解释应举例说明。如果存在其他解释的可能性，至少也应该提到。对于所有解释，作者尽可能保持客观。

总结

文献研究是对已结束和已发生事情的处理，某种意义上可以说是一种间接的调查研究，因其非接触性和间接性故呈现许多优点，有其独特的价值。

突破时间和空间的限制 每个人的亲身实践和经验总是受到时间和空间的限制，我们无法亲身经验前人的生活，不可能直接观察、访问前人的思想和活动；即使是同时代人，因经费、时间等因素的限制，也无法对研究对象作亲身考察。文献研究便可突破时间和空间的限制，使研究者对那些他们无法或难以亲自接近的研究对象进行研究。例如，在研究孔子的教育思想和学说时，可依

靠记载其生平和思想活动的文献《论语》。假如不用文献研究法，这项研究是无法进行的。

简便易行，费用较低 与实地调查法、访谈法等直接接触性方法相比，文献研究法具有方便、自由、费用低等优点。只要查找到文献，随时随地都能进行，不受研究对象、研究场地和研究情境等因素的限制。

真实性较强，可靠性较大 通常，人们总是习惯于日记、自传、信件、遗书等个人文献中自然地流露出真实的心理和行为动机，因此，与访谈和问卷调查相比，文献的坦白程度相对较高，真实性较强。同时，文献研究的研究结果可靠性较大，因文献研究是从现存的文献资料着手的，故不需要研究对象的直接介入，不会因研究对象的各种反应而影响研究结果的可靠性。另一方面，文献研究法可用较大的样本，从而可在一定程度上保证研究结果的可靠性。

文献研究法具有突破时间、空间限制，简便易行、费用较低及真实性较强、可靠性较大等优点。

尽管文献研究法有以上优点，但也存在着一些不足：

首先，文献本身存在一定的不完善性。例如，在许多教育文献中，作者往往出于特殊目的和意图夸大或掩盖部分事实，从而使文献记载出现偏差。此外，文献的保存常具有选择性，凡符合社会需要和当事人利益的文献，保存的时间便长久些；凡知名人士所写的文献，可得到较好保护。而且，许多文献尤其是日记、信件等个人文献，不是为研究目的而写的，时常会包含一些研究者所不熟悉的知识。

其次，收集文献难度较大。在许多情况下，一些事件根本没有文献记录。此外，政府机关的文件和档案因涉及公务机密，不宜公开外借；至于个人文献，除公开发表以外，需经当事人许可方可获得和使用，否则会因侵犯隐私而触犯法律。因此，在进行文献研究时，很难把所需文献找齐找全，容易造成文献不足的遗憾。

第三，文献抽样缺乏代表性。因文献资料大多以文字记载的形式保存的，所以能否留下文献资料，在很大程度上取决于文化程度的高低，受教育程度高的人写文献的可能性相对较大，故若仅仅依据现存文献来了解和分析人们所处生活状况、生活情形及思想观念，可能只能了解社会中某一阶层的情况，其文献抽样未必具有代表性。

文献研究法具有文献不完善性、收集文献难度大、文献抽样缺乏代表性等不足。

像其他类型的教育调查一样，文献研究有其独特的特点。它研究的是过去，文献研究在于探索和分析资料而不是产生资料。过去的变化已经产生，人们无法操纵它们。

文献研究法覆盖教育问题研究的许多领域，诸如一般教育史、教育专题史、教育法史、学制史等。通常说来，文献研究一般具有连续性特征，许多重要教育问题的研究首先要依据文献资料来开展，比如课程改革常以过去的哲学、观念、发展和课程情况作为依据，文献研究对于确定过去的形势并用它理解现在问题的意义都是必要的，因此，在文献研究基础上产生的解释能帮助我们确定研究现在教育问题的行动路线。

从确定研究问题到形成结论，文献研究构成一个相对完整的整体，它是重构过去发生过的事件和解释事件意义的系统过程。

文献研究是重构过去发生过的事件和解释事件意义的系统过程，能为研究者研究教育问题提供观察的角度，能使决策者从中获益。

在任何时代，都有一些教育问题摆在研究者面前。文献研究能为研究者研究教育问题提供观察的角度。在制定教育政策时，任何决策者都可从文献研究中获益。

1 比较研究法的发展历史与作用

古罗马著名学者塔西陀曾说：“要想认识自己，就要把自己同别人进行比较。”比较是认识事物的基础，是人类认识、区别和确定事物异同关系的最常用的思维方法。比较研究法现已被广泛运用于科学研究的各个领域。在教育科学研究中，比较研究是一种重要的研究方法。

比较研究法的历史

比较研究的最初运用可追溯到古希腊亚里士多德所著的《雅典政制》。该书对 158 个城邦政制宪法进行了比较。19 世纪以后，比较研究逐渐成为教育研究中的一种重要方法。其发展的历史脉络大致如下。

萌芽阶段 古希腊色诺芬（Xeuophon，公元前 430 ~ 前 355）关于希腊与波斯教育的介绍。古罗马西赛罗（Cicero，公元前 143 ~ 前 106）在《论演说家》一书中记叙了希腊与罗马的文化教育状况。意大利旅行家马可·波罗（Marco Polo，1254 ~ 1324）在其珍贵的《马可·波罗游记》中记叙了东方教育的情况。突尼斯的伊本·卡尔（Ibn Khaldun，1332 ~ 1406）非常欣赏东方伊斯兰文化和教育的价值，对教育比较的要素表示关注，对比较的方法提出了某些见解。1647 年，瑞典的约翰·洛克森纽斯（Johan Loccenius）已开始对一些国家的教育历史进行比较。此后 100 余年中，英国的威廉·佩蒂（Willan Petty，1623 ~ 1687），法国的狄德罗（D.Diderot，1713 ~ 1784）、孔多塞（Condorcet，1743 ~ 1794）也对教育的比较研究方法的形成作出一定的贡献。

总而言之，19 世纪以前，少数杰出人物在教育领域内运用比较研究法，虽然是可贵的，但又是自觉的。从严格意义上讲，他们的活动不能被看做科学意义上的教育比较研究。所以，我们把它称作比较研究的“萌芽阶段”。

形成阶段 19 世纪，研究者通过实地考察，搜集资料并进行了简单的类

比，以借鉴他人的长处改善自身的教育。典型的人物有法国的朱利安（Marc-Antoine Jullien，1775 ~ 1848），美国的霍拉斯·曼（Horace Mann，1796 ~ 1859），英国的马修·阿诺德（Matthew Arnold，1822 ~ 1888）。这个时期，比较研究法虽有发展，但仍以描述为主，是一种带有功利目的的教育借鉴活动，还不能解释教育现象的种种内在原因，比较分析的方法也未成为对研究者们起主导作用的方法。

发展阶段 20 世纪前叶，由于一些国家显露出在人才培养上的不足之处，教育难以适应社会发展的需要，因而不少国家开始研究影响教育发展的各种因素，探讨社会与教育之间的关系。例如，康德尔（Isaac Kandel，1881 ~ 1969）认为比较研究应从历史的纵向着手，并结合整个国家的其他背景。汉斯（Nicholas Hans，1888 ~ 1969），施奈德（Friedrich Schneider，1881 ~ 1974），马林森（Vernon Mallinson）等坚持认为对各国的教育进行比较研究应着重分析研究影响教育发展的各种因素。

成熟阶段 60 年代后，比较研究运用社会科学的，准自然科学的手段，用定性定量资料深入分析教育结构，确定各种因素在教育发展中所起的作用，寻求更精密、更精确的方法。这个阶段最杰出的代表是乔治·贝雷迪（George Bereday，1920 ~ ）。他在对历史法、因素分析法进行研究、吸收、批判的基础上创建了四阶段比较研究法，从而使比较法进一步具体化、科学化。这四阶段分别是描述、解释、并列、比较（详见下文）。

比较研究是教育研究的一种重要方法，其发展大约经历了萌芽阶段、形成阶段、发展阶段、成熟阶段。

比较研究法的作用

作为一种思维方法，比较研究贯穿于教育研究的全过程。无论在科学实验的过程中，或是在理论研究中，比较研究都是不可或缺的基本方法。在教育科学研究过程中，比较研究具有重大的作用，主要表现在如下几个方面。

比较研究可以帮助人们更好地认识事物的本质，把握教育的普通规律。

我们常说：“不怕不识货，只怕货比货”。我们认识一个事物常借助于与其他事物比较来实现。因为只有比较，才有鉴别；只有鉴别，才有认识。在实际生活与工作中，比较是认识事物本质的最基本的也是最重要的方法之一。

教育科学研究是一个复杂的认识过程，比较研究可以帮助我们更好地认识教育的客观规律。就普及义务教育而言，由于二战后国际竞争日益加剧，科技

发展异常迅速，知识经济年代已经来临，因而义务教育的年限必须加长。英国为 11 年，美国是 8~12 年，法国 10 年，日本 9 年。发展中国家也加快实施义务教育的步伐，例如，印度为 5 年，委内瑞拉为 9 年。通过比较，我们不难发现推行普及义务教育，并不是什么人的主观愿望，而是生产力与科技进一步发展提出的客观要求。

比较研究法能使人们更好地认识本国、本地的教育状况。

“不识庐山真面目，只缘身在此山中”。不借助他人，有时我们很难认清自己。在教育研究中，我们通过跨国比较、跨文化比较、跨学科比较、跨地区比较，可以找出一些共同的问题，发现哪一些是本国、本地的特殊问题，因而能更好地认清本国的教育状况，有效地推动教育的发展。

比较研究能帮助人们获得新的发现。

唐代医学家孙思邈在行医中发现：富人常得脚气病，穷人常得夜盲症。他比较了富人与穷人的饮食：富人多吃荤腥油腻、精米白饭；穷人吃素食粗粮。从而推断脚气病可能是因为缺少米糠之类东西引起的。再如“差生”的成因问题，上海市实验小学通过对“优生”与“差生”在课堂中举手发言的情况进行比较研究，从而得出“差生”的成因是缺少教育机会引起的。

比较研究为教育政策的制定提供依据。

教育政策的制定是一个复杂的过程，涉及到诸多因素。为使决策科学、合理就必须对各种因素有一个全面的认识，这就需要将某一问题与其相关的事物进行比较，从比较中分析优劣得失、长处弱点，从而使政策的制定符合教育规律与教育实际。例如，我国要对高校的招生制度改革就必须对各国的高校招生制度进行比较，然后制定相应的政策。再如中小学课程改革政策的制定也应建立在广泛的比较研究基础上。我国目前制定的三级课程政策就是在比较了中外课程设计的基础上作出的。

比较研究有四大作用：有助于认识事物的本质和教育的普遍规律；有助于更好地认识本国、本地的教育状况；有助于获得新的发现；有助于教育政策的制定。

2 比较研究法的概念及种类

比较研究法的概念

比较研究法在教育科学研究中广泛运用而且具有极高的价值，但是，什么叫比较研究法呢？《牛津高级英汉双解辞典》解释说：比较研究法就是对物与物之间和人与人之间的相似性或相异程度的研究与判断的方法。我国吴文侃、杨汉青主编的《比较教育学》认为“比较法是根据一定的标准，对不同国家或地区的教育制度或实践进行比较研究，找出各国教育的特殊规律和普遍规律的方法。”很显然，这个定义仅适用于“比较教育”这个学科领域，所以必须对它进行另外限定。我们认为教育中的比较研究法是根据一定的标准，对两个或两个以上有联系的事物进行考察，寻找其异同，探求教育之普遍规律与特殊规律的方法。对此，我们可以作如下理解。

1. 同一和差异是在各种事物之间普遍存在的一种客观联系，这种同一与差异是比较研究的基础。
2. 在教育科学研究中，比较是一种有目的有计划的认识活动，比较哪一些对象以及比较对象的哪些方面都是根据研究的实际需要而决定的。
3. 比较的对象可以是两个或多个国家或地区，也可以是同一国内的不同地区，甚至是两个学校或班级，也可以是教育实践……。总之，比较的对象涵盖教育的所有领域。
4. 比较的过程追求“同中求异”、“异中求同”。即要发现表面上相同或相似的事物有什么本质上的区别，或表面上似乎有明显差别的事物在本质上有什么相同之处。
5. 比较不可能十全十美。比较研究总是有选择地对某些事物的某些方面或一个方面进行考察，而暂时有条件地撇开其他方面。

比较研究是根据一定的标准，对两个或两个以上有联系的事物进行考察，寻找其异同，探求教育之普遍规律与特殊规律的方法。

吴文侃、杨汉青：《比较教育学》，人民教育出版社 1989 年第 1 版，第 19 页。

比较研究法的种类

根据不同的标准，我们可以把比较研究法分成如下几类。

按属性的数量，可分为单向比较和综合比较。

单项比较是按事物的一种属性所作的比较。综合比较是按事物的所有（或多种）属性进行的比较，单项比较是综合比较的基础。但只有综合比较才能达到真正把握事物本质的目的。因为在科学研究中，需要对事物的多种属性加以考察，只有通过这样的比较，尤其是将外部属性与内部属性一起比较才能把握事物的本质和规律。

按时空的区别，可分为横向比较与纵向比较。

横向比较就是对空间上同时并存的事物的既定形态进行比较。如教育实验中的实验组与对照组的比较、同一时间各国教育制度的比较等都属于横比。纵向比较即时间上的比较，就是比较同一事物在不同时期的形态，从而认识事物的发展变化过程，揭示事物的发展规律。在教育科学研究中，对一些比较复杂的问题，往往既要进行纵比，也要进行横比，这样才能比较全面地把握事物的本质及发展规律。

按目标的指向，可分成求同比较和求异比较。

求同比较是寻求不同事物的共同点以寻求事物发展的共同规律。求异比较是比较两个事物的不同属性，从而说明两个事物的不同，以发现事物发生发展的特殊性。通过对事物的“求同”、“求异”分析比较，可以使我们更好地认识事物发展的多样性与统一性。

按比较的性质，可分成定性比较与定量比较。

任何事物都是质与量的统一，所以在科学研究过程中既要把握事物的质，也要把握事物的量。这里所指的定性比较就是通过事物间的本质属性的比较来确定事物的性质。定量比较是对事物属性进行量的分析以准确地制定事物的变化。定性分析与定量分析各有长处，在教育科学研究中应追求两者的统一，而不能盲目追求量化，教育毕竟是一个不同于工人制造产品的活动，很多东西并非能够量化。但也不能一点数量观念都没有，而应做到心中有“数”，并让数字来讲话。

比较研究大约可分作单项比较与综合比较，横向比较与纵向比较，求同比较与求异比较，定性比较与定量比较四大类。

3 比较研究法的运用

运用比较研究法的步骤

比较研究法的运用没有一个固定的模式，但仍然有一个基本的操作步骤或者说是实施程序。下面着重介绍美国贝雷迪的四阶段操作程序和我国教育研究中常用的比较研究法的实施步骤。

美国比较教育学家贝雷迪在《教育中的比较方法》（“Comparative Method in Education” Holt Rinehart and Winston, Inc, 1964）把比较研究法的实施分成四个阶段：描述、解释、并列、比较。

描述

比较研究从详细描述比较的对象国的教育开始，就是说对各国的教育制度和现状尽可能周密、完整、客观地描述出来。为此，必须收集相关的资料文献。为了从文献中得到更加确切的知识，有必要对研究之对象国进行实地考察。

解释

第二步是解释。在完成对所比较研究的各国教育加以详尽而客观的描述之后，就要对所了解的教育情况进行解释，即说明这些教育现状所具有的意义，以便不仅了解事物是怎样的（How），而且了解事物为什么那样（why）。也就是说，以社会学、政治学、经济学、人文学、历史学、心理学、哲学等学科知识为基础，把所描述的教育现状与社会的一般现象联系起来进行思考，并说明这些教育现状所具有的多方面的意义，这就是解释阶段的目的。

并列

比较研究的第三步是并列。从严格意义上讲，比较研究从并列阶段才开始。

在这个阶段，首先把前一阶段里已描述并解释过的教育事物进行分类整理，并按可以比较的形式排列起来；然后确定比较的格局，并且设立比较的标准；最后进一步分析资料，提出比较分析的假设。

比较

比较研究的最后步骤就是比较阶段，在比较阶段里，要对并列阶段提出的假设按照“同时比较”来证明正确与否，然后作出一定的结论。

贝雷迪把比较研究分成四个阶段：描述、解释、并列、比较。

就我国的教育研究实践来看，运用比较研究法一般按如下步骤进行。

确定比较的问题

比较研究首先要明确比较什么，这是比较的前提。这一环节包括以下细节：(1) 选定比较的主题。如“教师性别对小学生个性的影响”，“经济发达地区与贫困地区民办学校之模式的比较研究”等。(2) 确定比较的内容。也就是确定比较的项目。如“优等生与差生学习机会的比较”这个主题，我们可以从“被老师提问的机会”、“参与教育活动的机会”、“家长提供的机会”、“同伴交往的机会”、“参加各种竞赛的机会”等方面加以比较。(3) 确定比较的范围。要明确是班内的比较、校内比较或是跨校学区甚至是跨国比较。

制定比较的标准

没有标准就无法进行比较。比较的标准可据实际情况制定，但要求明确化、具体化，即具有可操作性。如“发达国家与发展中国家教育经费之比较”这个主题，在确定比较的标准时以总经费作标准显然不妥当。但以人均教育投入，经费占国民生产总值的比重，经费占公共经费支出的比重为比较标准，那就不难分析出各国对教育的重视程度了。

搜集资料并加以分类、解释

要通过各种途径尽可能多地收集相关的各种资料，并对资料进行鉴别，保证资料的权威性和客观性。然后，对各种资料按比较的指标进行归类、并列。最后，对这些归类好的资料作出解释，即赋予资料以现实意义，为下一步的比较分析奠定基础。

比较分析

这是比较研究的最重要的一步，在这个阶段要对收集到的材料逐项按一定的标准进行比较，并分析其之所以产生差异的原因，而且要尽可能地进行评价。比较时应以客观事实为基础，对所有的材料进行全面的客观的分析。

结论

通过对材料的分析比较得出结论，并对所得的结论进行理论和实践的论证。

以上就是比较研究的一般步骤，这些步骤是相互联系、环环相扣的。确定比较的问题是运用比较研究法的前提；制定比较的标准是运用比较研究法的依据；材料的分类与解释是运用研究法的基础；比较分析是运用比较研究法的重心；得出结论是运用比较研究法的目的。比较研究不是为了比较而比较，而在于探索教育规律并得出符合客观实际的结论。

比较研究的步骤一般可分为五步：确定比较的问题；确定比较的标准；收集和整理资料；比较分析；结论。

运用比较研究法的条件与规则

比较研究法广泛运用于教育科研实践，但并不是任何时候都能运用，只有在如下条件符合时才能运用比较研究法。

同一性

所谓同一性，是指进行比较研究的对象必须是同一范畴、同一标准、同一类事物，否则就不可以比较。正如马克思所说：正是因为倍尔西阿尼是一位歌唱家而且人们把她同其他歌唱家相比较；人们根据他们的耳朵的正常组织和音乐修养做了评比，所以他们能够认识倍尔西阿尼的无比性。倍尔西阿尼的歌唱不能与青蛙的鸣叫相比，虽然这里也可能有比较，但只是人与一般青蛙之间的比较，而不是倍尔西阿尼与某只惟一的青蛙之间的比较。只有在第一种情况下才谈得上个人与个人之间的比较，在第二种情况下，只是他们的种族特性或类特性的比较。

双（多）边性

比较只有在两个事物或两个事物以上才可能发生。换言之，比较的对象必须要在两个以上。当然，比较研究还要求从不同的角度对两个被比较的对象进行分析比较。

可比性

可比性是指被比较的对象之间具有一定的内在联系，具有某些本质上而不是表面上的共性。为了保证可比性，必须注意概念的统一。例如，英国的 Public school 与美国的 Public school 就是不同的概念。前者是贵族学校，而后者则是贫民学校，不具备可比性。

运用比较研究法必须满足三个条件：同一性，双（多）边性、可比性。

在满足以上条件的前提下，比较研究就可以开展了。但具体运用应注意如下几点。

资料的可靠性与解释的客观性

供比较研究的资料必须具有权威性、真实性、最好是第一手资料。而且要具有代表性，能反映普遍的情况。最后要求资料具有典型性，能反映事物的本质。这就要求采样的科学化，资料归类合并的程序化，资料收集人员要有扎实的教育理论修养和深厚的基础理论素质。

在解释资料时根据当时当地的客观实际，运用科学的理论加以全面的分析，并保证解释的客观性而不带有个人的偏见。

全方位多角度进行比较

制约教育发展的因素甚多，而且任何事物都不是孤立地存在的，而是与其他事物密切联系的，所以应坚持全方位多角度的比较。在比较时，任意选择个别条件，片面地进行比较，或者割裂事物之间的联系，孤立地进行比较，都是不正确的。

比较事物的本质

事物不仅有现象的异同，更有本质的异同。比较研究不能仅抓住表象而忽视本质，否则就难以准确地认识事物。“假如一个人能看出当即显而易见之异，譬如，能识别一支笔和一头骆驼，我们不会说这个人有了了不起的聪明，同样，另一方面，一个人能比较两个近似的東西，如橡树和槐树，或寺院与教室，而知其相似，我们也不能说他有很高的比较能力。我们所要求的，是要能看出异中之同和同中之异”。就是说我们要透过现象抓住事物的本质。要进行本质的比较就要通过大量的、典型的材料分析其内在关系，从历史的、社会的、经济、社会风俗等角度进行探讨。

运用比较研究法的规则：资料的可靠性与解释的客观性；全方位多角度进行比较；比较事物的本质。

4 藏、汉儿童数学思维能力发展差异性的比较研究案例

为了帮助我们对比较研究方法有更具体的理解，下面以《藏、汉儿童数学思维能力发展差异性的比较研究》为案例加以说明。

黑格尔：《小逻辑》，商务印书馆，1980年，第253页。

确定比较的问题

教育受儿童身心发展规律的制约,所以必须了解儿童身心发展的特点。而我国是一个多民族国家,在我国的少数民族中,藏族人口最多,居住面广,有悠久的历史 and 独特的文化,在国际国内有很大影响。藏族教育是我国少数民族教育的重点,同时也是我国教育的难点,而对藏族儿童和藏区其他民族儿童思维发展的特点与规律的研究是解决藏族教育这一难题的关键。所以该案例选择藏、汉儿童进行比较研究。但是为什么要选择数学思维能力作为比较的问题呢?这是因为数学是人的“思维体操”,以其特有内容、形式和方法概括显示出人的思维水平。

选好了比较的主题,接下来要确定比较的范围。该案例选择了9~15岁七个年龄组的藏、汉在校儿童。其中藏族儿童选自甘肃省甘南藏族较集中的碌曲县和夏河县的民族中小学,汉族被试选自碌曲县、夏河县的普通中小学。各年龄组藏、汉被试各20名,男女均各半,按数学成绩上、中、下三个等级分层随机取样(上30%,中40%,下30%),并尽可能使年龄组与年级相对应。

确定比较标准

比较的指标有三个:数学思维能力;文化背景;个性特征。数学思维能力分为比较能力;分类能力;概括能力;运算能力;问题解决能力等五个标准。文化背景分为父母亲的职业;文化水平;对子女学习的关心程度;对子女学习的指导程度;对子女学业的期待程度;对数学的态度;家庭学习环境;学校教育环境;教师对学生的期待水平等标准。个性特征分为数学学习动机;兴趣;态度等标准。

收集和整理资料

资料的收集通过两个途径来进行。一是应用国际通用的瑞文标准推理测验(Raven Standard Progressive Matrices,简称R·SPM),另外编写数学思维能力测验题目测试被试。各组测试题按照试题范围的大小、抽象程度的高低赋以不同的权重。二是问卷和访谈、文献分析(查阅学生档案、作业等)、实地观察收集相关资料。

拥有了大量资料以后,把它们加以整理。剔除那些无效的样本,然后根据不同的标准进行分类处理,并对数据进行统计,最后汇总,必要时要制成表或图。

比较分析

根据以上资料该研究从如下几个方面进行分析：藏、汉儿童思维能力发展趋势和水平的差异性；不同年龄组藏、汉儿童达到数学思维各级水平的百分比的比较；藏、汉儿童文化背景与学习动力的比较；藏、汉儿童数学思维能力与文化背景和数学学习动力因素的列联相关比较。

通过以上比较发现藏、汉儿童的数学思维能力有明显区别，原因何在？这需要进行进一步分析。本案例从教育水平的差异；语言的影响；个体意识倾向性的差异几个方面进行分析。

结论

根据以上比较分析得出结论：藏、汉儿童的数学思维能力发展存在显著差异，这种差异与儿童的家庭、学校环境以及语言等文化背景因素和某些个性特征有关，而与民族性无关。

总结

比较研究是根据一定的标准，对两个或两个以上有联系的事物进行考察，寻找其异同，探求教育之普遍规律与特殊规律的方法。其发展大约经历了萌芽阶段、形成阶段、发展阶段、成熟阶段。

比较研究有助于认识事物的本质和教育的普遍规律，有助于更好地认识本国、本地的教育状况；有助于获得新的发现；有助于教育政策的制定。

第十一章

实地研究的方法

1 现场研究

实地研究中，研究者主要是在研究现场应用谈话观察等方法获得资料的。在本章中，我们首先介绍现场研究的一般方法，而后对实地研究的几种主要获得资料的方法进行讨论。

现场对实地研究者来说意义是巨大的。首先，透过直接观察等活动，研究者能够更深入地理解教育现象的完整情境脉络。其次，第一手经历能使一个研究者思想开放，不断适应新发现，并且采取归纳的方法。实地工作的第三个优点是，研究者有机会接触到人们日常不太注意的事情。为了让某人在访谈中提供有关信息，他们必须相当清楚地报告所期望的信息。所有的教育体系都包含了例行公事的程序，教师和学生可能会认为这些例行公事是理所当然的事情，而无法觉察到重要的和细微的差别。只有对于一个没有进入这些例行公事中的观察者来说，这些重要的、细微的差别才是十分明显的。

进入研究现场

进入研究现场是需要勤劳和耐心和技能的工作，并不是一件简单的事。一般说来，研究者需要和现场的人们磋商以进入现场，逐渐获得他们的信任和合作，并且慢慢地搜集适合其研究兴趣的资料。如果进入某个现场碰到困难，研究者可以继续尝试下去，试着继续和现场的人们磋商。并没有一定的规定，决定什么时候研究者该放弃某一个场所，而改换另一个场所。然而，为了避免盲目地尝试，研究者在进入某个现场时，需要进行初步的评估和运用一些进入现场的策略。

研究者在接近和进入某个研究现场时，基本上宜先评估和了解研究者自己和现场的关系、现场的性质、以及现场人们的属性。进入研究场所时，有三个

基本的参考原则：一要坚持，持续地寻求进入现场：二是有弹性，如果最初的想法行不通，尝试不同的计划或新的方式；三要有创意，例如送研究对象小礼物或卡片。

进入研究现场有两种基本的策略：一种是暗中的策略，当研究者无法通过商议进入现场，或该场所是对局外人封闭的场所，研究者即在暗中进行；另一种是公开的策略，即获得进入现场观察的允许，这种方式较少产生研究伦理上的问题，大多数定性研究者倾向于采用这种公开的策略。在教育研究中，一般以用公开的策略为多。

进入公众的场所研究时，对于那些只随意接触的人们，研究者不需对他们介绍自己是研究人员和解释研究目的，而对于那些将有持续关系的人们，研究者应该对他们介绍自己，尤其在人们开始怀疑其研究的意图之前即要表明自己的身份。

但是，也有学者认为，实地研究在某种方式和程度上常常是暗中的、秘密的（Lofland & Lofland, 1984, p. 24），其原因是：

1. 许多研究者通常是因个人的兴趣先进入某个现场看看，而后才决定正式进行参与观察研究，在这种情况下，除非研究者预先告知现场的人们他可能在这里进行研究，否则在进入现场的初期，即不可能完全避免某些无意的欺骗。而且，研究者最初使用暗中进入的策略，常只是为了无干扰地进入某个现场，等到他与研究对象建立了信任的关系，即可能告知其研究的意图和兴趣。

2. 在很多社会行为的研究中，研究者不希望研究对象知道他正在进行研究或他所感兴趣的行为是什么，以免研究对象呈现的行为受到影响，例如在研究犯罪的亚文化时，或是为了获得真实的资料时，暗中的策略即是需要的。

3. 有的研究者虽然使用公开进入现场的策略，也并非明白告知现场每个人其研究的目的，研究对象并不一定能同样地了解研究者所使用的所有名词，因此研究者和研究对象对于研究目的的概念可能是不同的。

虽然实地研究常都含有某种程度的秘密性，但在基本的研究取向上，多数实地研究者是倾向于采用公开的策略。我们在本书中立场也是倾向于公开的研究取向。

进入研究现场有两种基本的策略：一种是暗中的策略，当研究者无法通过商议进入现场，或该场所是对局外人封闭的场所，研究者即在暗中进行；另一种是公开的策略，即获得进入现场观察的允许，这种方式较少产生研究伦理上的问题，大多数定性研究者倾向于采用这种公开的策略。

如何进行现场研究

在进入现场以后,如何进行现场研究。下面提供实地工作的一些指导原则。

1. 做实地札记时应是描述性的。
2. 从不同的维度收集各种不同的信息。
3. 通过多种不同资料进行交叉验证 (观察、访谈、文件记录、录音以及照片), 在研究中可以使用多种方法。
4. 采用摘录, 在参与者自己的语言描述中, 去捕捉参与者对其经验的想法。
5. 选择主要的信息提供员, 并谨慎小心地使用他们。从他们所提供的观点中提炼出精华的智慧, 但是同时应当切记他们的观点取向是有限的。
6. 注意实地工作的各个不同的阶段。在进入阶段与参与者建立信任和密切的关系。在实地工作成为例行常规的中间阶段中, 保持清醒、训练有素。当实地工作接近尾声时, 重点放在总结出有用的综合描述。在实地工作的各个阶段中都要认真仔细地做好详细的实地札记。
7. 尽可能完全参与研究的全方案, 对过程有一个完整的体验。
8. 你的实地札记和评鉴报告中要包括你自己的体验、想法。

首先, 实地札记是描述性的。札记应当标明日期, 而且应当记录下下列基本信息。如: 观察进行的地点、在场的人员、自然环境如何、产生了何种社会互动, 以及进行了什么活动。

实地札记也要包括观察者自己的感受、对体验的反应, 及其在已经发生的事件中对于观察者的重要性的意义的思考。感受和反应应当在体验之时, 也就是观察者在实地之时, 就记录下来。这些感受的性质和强烈程度也应加以记录。在自然式研究中, 观察者自己的体验是资料中至关重要的一部分。接近方案和参与者的目的, 是使观察者有机会去体验这一具体场合, 从而了解这一场合中的情况。如果观察者没有在实地札记中记录场合情况的资讯, 那么到场合中去的就失去了意义。

如何进行个案研究

个案研究在实地研究设计中有特别意义。从某种程度上说来, 实地研究就是一种个案研究, 因为研究最佳答案可以从个案分析中获得。个案可以是个人、研究计划、机构或团体。实地分析的个案研究法 (case studies), 是一个收集、组织和分析资料的特殊方式, 目的在于获取对每个研究个案的综合的、系

统的和深度的信息。因此，个案分析的起点，是要确信每个个案的资料尽可能的完整。

从某种程度上说来，实地研究就是一种个案研究，因为研究最佳答案可以从个案分析中获得。

对单一方案的实地研究可以是一个个案研究。在单一方案的个案研究中，研究者可以对几个参与者进行个案研究。假如选用这样的方法，分析可以从对个别个案的研究开始。然后，对个别个案所进行的比较分析，将成为研究资料的一个主要部分。

个案记录内容

一旦原始的个案资料已积累到相当程度以后，研究者便可以写出个案记录了。这个个案记录将大量的个案资料结集和组织成一个综合性的和主要的资源库。个案记录包括所有将会被用来做最后个案分析和个案研究的主要信息，并将信息加以编辑，剔除多余的信息，使其各部分相互连结，并将个案记录加以整理，以便按时间顺序或按主题对资料进行分类。个案记录必须是完整且便于管理的，它应该包括进行后续分析所需的所有资料。

2 参与观察

观察是科学研究的一种基本的方法，也是教育研究中重要的收集资料的方法。

什么是参与观察

在教育研究中，观察法一般分为两种，一种称为实验观察，另一种就是参与观察。实验观察是指对观察的情境作实验控制，然后观察其结果。参与观察也称为自然观察，它是指就研究对象在自然的状态下研究者参与某一情境中去对研究对象进行观察。在定性教育研究中，参与观察是研究的主要方法之一。

参与观察也称为自然观察，它是指就研究对象在自然的状态下研究者参与某一情境中去对研究对象进行观察。在实地研究中，参与观察是主要方法之一。

参与观察源于人类学家的现场工作，对人类学的现场工作者而言，参与观察是经年累月地住在当地社区，将自己融入社区人们的生活中，尽量精通人们的语言，并维持一个专业者的距离，以此观察人们的日常生活和活动，了解人们的基本信念和期望，并有系统地作成资料记录。

例如，一位幼儿教育学者曾经进行过一个参与观察的研究。幼儿的发展与辅导是幼儿教育界最基本的课题之一，但人们在思考如何辅导幼儿的各种行为发展时，非常缺乏在自然情境中详细地描述这些行为的资料，这些资料对于制定辅导的原则或策略是重要的。传统的发展理论观点只将社会化看做是孩子长成大人的过程，而没有考虑到孩子成长过程的日常生活事件，研究者经常脱离了幼儿生活的情境脉络，而以大人的观点去解释幼儿的行为。

我们大人如欲真正了解幼儿自身对其自我或周围事物的看法，就必须亲身进入幼儿的生活世界，作个参与者和观察者，这位学者自身即实地进入一所托儿所，运用参与观察法和录像的辅助，搜集幼儿交朋友和发展友谊之实际过程的资料。这种亲临现场作个参与者和观察者的方式，是实地研究者搜集描述性资料的主要方法。

观察活动中研究者参与的程度

区分观察策略的主要区别，是有关观察者参与程度问题。这不仅仅是参与和不参与之间的选择。参与的程度，是完全进入场所作为一个完全参与者，到完全脱离这一场所成为旁观者的连续体。这两极端之间的连续向度中具有极大的变异。

研究者在现场观察时，由介入或参与他们所观察的人们和活动的程度不同，得出区分不同的参与观察的类型。

参与程度会因时而异。在一些实例中，研究者可能在开始之时以一个观察者的身份出现，然后逐渐成为研究历程中的参与者。在其他的实例中，研究者可能开始就成为一个完全参与者，以便体验一下开始进入方案中的感受，然后在研究阶段逐渐减少参与，直到最后从一个旁观者的立场扮演临时观察者的角色。

实际上，每个人如果不是其他文化的参与观察者，至少也是他自己文化的参与观察者，但一般人常只是无系统或不完全地观看事物。现场研究者绝不只是在那儿被动地观看人们，而需要与人们有紧密和长期的接触。从参与观察所获得的初步资料，可以提供研究者发展访谈问题或其他更特定研究工具

的基本线索，参与观察也能进一步检验和评估访谈和其他方法所获得的资料。

研究者在现场观察时，介入或参与他们所观察的人们和活动的程度有高低的不同，区分成不同的参与观察的类型。这些类型中，一个极端是研究者完全不参与现场的活动，作个完全的观察者，另一个极端则是研究者完全参与现场活动，研究者和研究对象的行为没有什么明显的不同。而一般的现场研究者乃是介于这两个极端之间，有的趋向于参与现场情境作观察，但并不影响现场主要活动，有的趋向于以一种现场局内人的角色来观察现场活动（参与者即观察者）。

同样地在参与和观察之间的连续体上，斯潘帝尼（Spradley, 1980）将观察者的参与方式区分为下列五种：

1. 完全不介入所研究的人们的活动，例如，坐在电视机前观看系列的电视节目、儿童卡通、或广告片，并分析其主题，观察者即完全不介入。

2. 介入程度低：观察者出现于现场，并不积极与人们互动，只是找一个观察地点做观察记录，或与人们有些交谈，在公共场所观察都是从这个程度的参与开始，场所的人们不认识观察者，也不知道他在做什么。

3. 中度的参与：参与者寻求维持现场的局内人和局外人之间的平衡，亦即维持参与和观察之间的平衡。

4. 主动积极的参与：观察者做现场人们所做的事，主动积极地参与人们的活动，更充分地知道人们的行为规则。

5. 完全的参与、介入程度最高：在研究现场，观察者已经是日常的参与者，研究自己的日常生活或工作的现场。教师进行的行动研究就包括一个完全参与的观察活动。

如何进行参与观察

研究者可以从以下几个方面注意参与观察的基本要求。1. 参与观察者的活动须有明显的自觉，记住场所中的活动，提高注意的层次。2. 参与观察者需要作较广角度的观察，日常的参与者较受限于个人的目的。3. 日常的参与者通常局限于个人的主观经验和目的，参与观察者则同时经验局内人和局外人的角度，一方面入乎情境中，经验局内人的情感和经验，一方面出乎情境之外，观察现场中的人们和活动。4. 参与观察者需要有自我反省的能力，学习以自己为研究工具，增进自己对情境的敏感度，以搜集丰富的资料。5. 参与观察者需要作仔细的记录，记录客观的观察和个人主观的感觉。

参与观察较大地依赖研究者本身的观察技能。从观察的角度来看，可以分为二种基本的情况：全面的观察和焦点的观察。

从观察的角度来看，可以分为二种基本的情况：全面的观察和焦点的观察。

全面观察 全面的观察者观察活动并没有特定的焦点，全面注意场所中的研究对象的的活动。他们往往参照下列三种问题的取向，进一步观察发现某些较特定的资料：

1. 什么……（地方、行动、事件、感情等）？
2. 你可不可以仔细地描述……（物体、时间、目标等）？
3. 你可不可以告诉我有关所有的……（人们、活动等）？

全面观察往往是观察初期的一个阶段。

焦点的观察 研究者在现场逐渐找到观察的焦点后，即增加焦点的观察。研究者在现场选择观察焦点的标准包括：

1. 研究者个人的兴趣：初步分析观察资料所呈现的各个领域，研究者问自己哪些领域看起来是有趣的，也许自己以前所作的研究或阅读将对那些有趣的事有作用。

2. 被研究者人的建议：有时正在被观察研究的人们也会提出他们觉得重要的事情，研究者可参照他们的建议，选择进一步观察的焦点，有时注意听人们经常重复的话题，也可能是观察的焦点，如果研究者已经和现场的人们建立密切的关系，也可直接问他们：“你认为什么是最重要的部分？”

3. 理论的兴趣：初步分析观察资料所呈现的各个领域，有些会涉及研究者理论背景。例如，一个具有社会学理论背景的研究者会对学校的某些社会组织有兴趣，开始观察一个三年级的班级，并辨认出许多有关社会组织的领域（有关孩子的、教师的、或群体的）。

4. 符合社会的需要。初步分析观察资料所呈现的某个领域，如果是社会上人们关心的问题或文化环境中的主要问题，也可能是观察的焦点。

全面参与观察的过程就像是一个漏斗上部，要将看到的所有事物容纳进去，这种观察将持续进行于整个研究过程。焦点的观察逐渐缩小观察的范围，选择的观察更集中于某些特定的焦点。这时，研究者开始在现场寻找特定类别之间的不同，经常问自己一个对照式的问题：“这一类和那一类之间有什么不同？”如发现了一两处不同，试着继续去寻找更多的不同之处。

3 访谈

什么是访谈方法

在定性研究中，访谈（interview）通常是两个人（有时包括更多人）之间有目的的谈话，由其中一个人（研究者）引导，搜集对方（研究对象）的语言资料，以此了解研究对象如何解释他们的世界。在实地研究中，访谈可能有两种可用的方式，一是作为搜集资料的主要策略，二是配合参与观察、文件分析或其他研究技巧，作为搜集资料的辅助方式。

上一节所阐述的参与观察，是在自然的情境直接看到人们所做的和所说的。如果研究者希望了解人们的内心观点，除了观察他们正在进行的活动或事件，同时也需要访问他们，才得以了解他们的信念、动机、判断、价值、态度和情绪。参与观察通常相当费时和费力，而且因为观察者无法直接观察现场过去发生的事，也不能强迫进入所有的场所和私人的情境，因此参与观察并不完全适用于所有的研究情境。而谈话的方法则能补充参与观察的不足。

访谈是实地研究中一种获得资料的重要方法。在实地研究中，访谈的方式基本上也分为非正式的和较正式的两种取向。而访谈的结果较大地依赖研究者的技能。

访谈的方式

在日常生活中，人们谈话的方式有很多种，有很随意的、也有很正式的，在实地研究中，访谈的方式基本上也分有非正式的和较正式的两种取向。

非正式访谈 非正式访谈就像是日常自然的谈话，是一种随意的、自由的、开放式的、非指示性的和非结构性的谈话方式。非正式访谈不限制于问题的特定形式或次序，主要让报道人引导谈话的方向，研究者的角色是显得感兴趣的，以表情和简单的语言反应和鼓励报道人继续说，或参照报道人谈话中呈现的话题提出问题，让报道人觉得研究者很愿意从他那儿学到一些事情。

通过报道人自己陈述的话，研究者可在确定研究兴趣的焦点之前，学习什么是对报道人重要的，发现报道人对他们的生活、经验和处境的各种想法。在

研究进行初期，实地研究者常透过许多这样非正式的谈话搜集资料，并用来建立和维持与报道人的关系。

正式访谈 正式的访谈主要是一种结构式的访谈。研究者一般预告准备访问的问题，按照这些问题对报道人进行访问。在作正式的访谈之前，研究者必须较详细制定计划，使访谈依计划进行。

访谈的一般做法

访问往往和参与观察相联系。通过参与观察，研究者能注意到哪些人最值得进行访问，这些人通常即是拥有最多第一手资料的主要报道人。研究者也能透过参与观察检验主要报道人的谈话资料是否符合其实际的行动。

一般而言，研究者在访谈之初须先向报道人澄清访谈的动机和目的，研究报告中将使用虚拟的人名和地名，将让报道人检视研究报告的初稿（如有的话），以及告知研究结果是否将出版和在何处出版等。研究者通常要进行访谈之后，才能知道要与报道人作多少次的访谈。既然不能预告明确地知道要与报道人作多少次访谈，要避免一开始即要求报道人承诺花很多时间作访谈，等到作了一次或者几次访谈之后，再较直接地和报道人讨论进一步的访谈时间表。

访谈的次数和时间参照研究者和报道人各自的时间表而定，对于成人来说，每次访谈通常需要两个小时。如果时间太短，可能无法探索许多主题，如果时间太长，则可能让研究者和报道人都觉得很累。为了保持访谈的流程，应该试着每个星期都和报道人会面，至于访谈的地点，应试着找一个不受干扰的场所，例如在报道人的家里或研究者的私人办公室，让报道人觉得轻松自在。

访谈者和报道人的关系大多数是单方面的，经由这种关系，访谈者有机会进行研究以获得学位或出版成书，而报道人可获得的，除了满足于有人认为他们的生活和观点是重要的，并没有很多实质的回报，但他们必须付出很多时间和精力接受访谈。由于这种单方面的关系，访谈者必须努力维持报道人的访谈动机，维持关系的最好方式是将报道人视为一个人而不仅是资料的来源。

在进行访谈的期间，访谈者和报道人的关系可能趋向于紧张，报道人可能厌倦于回答问题或开始觉得访谈是他们生活中的负担，而访谈者可能开始对报道人不愿回答问题或跳开特定的问题，而感到不耐烦，结果两人都对访谈感到厌烦。访谈者应该试着敏感地察觉到报道人的情绪低潮，当觉得有些事不对劲时，可对报道人表达关切以缓和紧张的气氛，有时也可暂时停止访谈，休息一段时间，或一起作些社交或娱乐活动，藉以增进双方的友谊关系。

访谈中的注意事项

访问者的态度 首先，访谈者应该经常察觉到他的话语和表情如何影响到报道人，访谈者对报道人必须是同情的和友善的，但不应是讨好式的逢迎。再者，保持沉默也是一个有效的访谈策略，在谈话过程中报道人如有短暂的静默，访谈者不需要急于再问问题，而应让报道人在回答之前再消化、理解、想一下问题；报道人谈论过一个主题后，访谈者保持短暂的停顿或沉默，也可鼓励报道人再继续说，提供更多资料，而不会觉得有压力。

如果不能确知报道人的意思，可请报道人再澄清或重复描述，让他们觉得访谈者很有兴趣更了解他们，例如对报道人说：“你这样说的意思是……？”“我不完全懂你的意思，请你再解释一下”，并以此检验报道人所提供资料的一致性。访谈者也可以复述报道人所说的话，让他们知道访谈者所听到的，再确认访谈者是否真正了解他的谈话，并有机会再解释或澄清自己谈过的事情。

访谈中的问问题 访谈者的角色不仅是要获得答案，也是要学习问什么问题 and 如何问问题。研究者在现场的参与观察也总是伴随着访谈和发问。这些问题可分为三类。

1. 描述的问题 (descriptive question)：这类问题是请报道人描述一般的情况，最常用在所有的访谈过程中，例如，问报道人：“请你描述一下你参加的会议？”；“你能不能告诉我你在办公室做些什么事？”或是请报道人带领到场所四处看看，并说明场所的用途和活动，例如，问报道人：“你能不能带我四处看看，并介绍一下？”也可提出一个假设的情境方式，请报道人描述自己的经验或看法，例如，问报道人：“如果家长向你抱怨孩子在学校没学到什么，你会怎样说？”

2. 有结构的问题 (structural question)：这类问题是要发现报道人的文化知识领域，了解报道人如何组织他们的经验，例如：“你在假期中捕捉到哪些不同种类的鱼？”；“转入你的学校需要经过哪些阶段？”结构式的问题常被重复提出，例如报道人说出了六种活动，研究者可能再问：“你能不能想到其他种类的活动？”

3. 对照的问题 (contrast question)：这类特定问题是用来发现报道人在生活中如何区分物体或事件的意义，一个典型的对照式问题是：“什么是 X 和 Y 之间的不同？”

描述的问题常在研究初期用来辨认报道人的生活方面，列举出他们生活中重要的事件、经验、地方和人物，有助于确定研究范围和方向。描述的问题也能促使报道人提到某些特定的经验，而刺激引发较特定的结构的或对照的问

题，然后再产生更多描述的问题，接着再一次产生特定的细节问题，就这样来回于描述的问题和特定式问题之间，直到研究者已获得一个清楚的概念架构。描述的问题可对研究范围形成一个广泛的了解，而结构的或对照的问题则可精练和扩展这个了解。

4. 访谈的记录

访谈如果是搜集资料的主要方式，研究者在访谈期间最好作仔细的访谈日记，在日记上列出每次访谈讨论到的主题，这可帮助研究者知道哪些话题已经被谈过。日记也可替代如参与现场观察记录中的“观察者的评注”，在日记中可记下资料呈现的主题、解释和可以此了解报道人语言意义的表情和姿势，这些记录有助于引导进一步的访谈。

在访谈日记中还可记下研究者在访谈情境和报道人的谈话，有时在访谈休息时或和报道人非正式的接触中，研究者也能听到一些在报道人生活中重要的事情。

实地研究常强调研究者是搜集和分析资料的一种工具，这个比喻意指研究者实施定性研究时，必须广泛地运用自己的经验、想象、智慧和情感。例如，报道人谈到他与同学的相处关系时，研究者亦反省他自己的家庭经验以及他与父母的关系，这种经验配对的过程可以帮助研究者了解报道人谈话的意思。

实地研究的访谈者不是像机械人似的资料搜集者，在访谈过程中，对于报道人谈话音调的转变保持敏感是很重要的，因为这些转变是一个人的态度和情感的重要线索，尤其当报道人的音调转变成悲伤、沮丧、或哽咽时，报道人可能是想谈这个主题以宣泄过去的悲伤经验，也可能是因触及个人内在的秘密而觉得有压力，这时研究者须学习敏感地区分其中的不同，体恤报道人音调中内蕴的态度和情感。

4 搜集资料的其他形式

搜集文件

就实地研究而言，文件（document）的主要用途是检验和发现研究的线索。实地研究强调丰富地描述场所和人群的现象，因此需运用多重的资料来源，如参照人类学研究的“多重工具取向”，搜集资料的方式除了参与观察、主要报道人的访谈、生活史的搜集、结构式访谈之外，还包括一般社会学研究运用的问卷（questionnaire）、投射技术（projective techniques）、其他心理研究工

具和现场工作的技术设备。研究者可参照个人的研究取向，选择使用各种不同的搜集资料方式，以增进研究结果的可信度。

问卷

问卷的使用接近结构式的访谈，但问卷在本质上不同于访谈，访谈有一个互动的本质，而问卷则缺乏，因为其在研究者和研究对象之间有所隔离。然而，问卷可用来搜集较多人的资料。通常在研究的较后期，教育研究者对现场教育现象有大致的认识后，参照其对现场知识以发展适用于该现场情况的问卷，以调查现场参与者的信念、态度、看法和对人、事物的价值观感。例如，华尔柯（Wolcott, 1984, p. 10）进行一位小学校长社会生活的实地研究时，在学年结束前设计了一份问卷，请学校的教师们填答他们对这位校长的感觉，这份问卷调查的结果呈现教师们对校长的多重印象。

投射技术

选择使用一般的心理研究工具或测验，可以补充和增进现场工作，这里就以投射技术（projective techniques）为例，说明其在实地研究上的应用。投射技术可使人们产生有关文化和心理反应的资料，这些反应通常呈现个人的需要、害怕、爱好和一般的世界观。例如，罗夏克墨渍测验（Rorschach Inkblot Test）是一个古典的投射技术，心理学家或精神病学家让病人解释一系列墨渍图片，再根据病人的反应资料作论断。人类学家信亦运用罗夏克墨渍测验、主题统觉测验（Thematic Apperception Test）、画人测验（Draw-a-Person Test）、或语句完成法等投射技术，引发人们的人格反应和价值态度，以探究人们的文化特质。

照相机

照相机是视觉的伸展和记录，能充分地掌握现场的视觉现象，帮助研究者更密切地感觉和认识现场，尤其某些现场细节太多或太含糊，较难用语言文字描述，即可用相机直接留存视觉的意象记录。照相机可用来拍摄人们的活动、事件和场所。所拍得的照片或幻灯片可用作访谈的参照资料，让现场人们自己解释其中的意义。在分析和撰写报告时，照片或幻灯片则能呈现研究者可能不记得或不注意的细节，也能用来呈现和解释研究发现。必须注意的是，在实地研究中，相片不是答案，而是探究的工具。还要注意的，照相机常常反应了拍摄者的观点。

录像机

录影设备能捕捉不被注意或容易被遗忘的连续性细节,它就像是忠实且有耐心的观察者,能长时期正确且仔细地记录人们的行动,特别适用于微观地分析人们的沟通和互动过程,例如:表情、手势、口语的沟通和视觉的形态等,拍摄的影片还可暂停或重复地分析。然而,录像机也和前述的照相机一样,可能受制于使用者的观点,并且可能干扰现场和改变现场人们的日常行为。

录像机拍摄的现场影片是很直接的访谈资料。实地研究学者常使用这种影片作为刺激,让当事者看着影片中自己的行为举动,描述他的感觉、解释他在影片中的行为举动,并回忆当时他的心里在想些什么。亦有教育人类学家(Tobin, Wu, & Davidson, 1989, pp. 4 ~ 9)使用录影片比较研究三个文化(日本、中国和美国)系统中的学前教育,他们到这三个国家拍摄幼儿学校典型的一天活动,再利用这些影片作为访谈的媒介,分别让三个国家的幼儿教师、行政者、家长、幼儿、和儿童发展专家观看自己国家和他国幼儿学校的活动影片,从不同的观点谈论同一活动事件,因而引发影片中局内人和局外人之间、幼教实际工作者之间、以及美国人、中国人和日本人之间的对话。有的教育研究者使用录像机拍摄教师的教学行为和事件,再让教师观看影片中自己的教学实况,以激发其回忆和叙述自己的方法。

实地研究的描述性资料除了研究者产生的上述现场观察记录和访谈记录之外,研究者还可搜集现场原有的或现场人们产生的文件资料,以及运用问卷、测验和仪器等其他方式搜集资料。

总结

实地研究(Ethnographic research),也有译有人种志研究等,是为特定情境中对教育过程以及现象提供科学描述的研究,是定性研究方法中发展最快、最有代表性的一种方法。60年代兴起于美国,80年代以后普及开来。这种研究的方法和手段已日趋成熟,形成了明显特色。

本章详细介绍了现场研究、参与观察、访谈和搜集资料的其他方式。读者可在具体的教育研究中有选择地加以运用。

第十二章

定性研究资料的分析

1 通过分析发现主题

在定性研究的过程中，与获得资料同样重要的一个步骤是分析资料。分析资料的质量对于获得研究的结果是十分重要的。分析资料的实质是发现研究的主题和结果。这也是定性分析的目的。

生产性分析和无生产性分析

定性研究的分析有两种基本形态：生产性分析和无生产性分析。无生产性分析指在收集完资料后再进行分析。这样，虽然研究步骤比较清楚，但在研究途中可能失去未经分析组织的资料，而且很难知道他们是否已经搜集了某一主题的资料。一旦他们在分析资料时再发现问题，又可能很难再回到现场搜集资料，以补足资料的漏洞和缺失。因此，这种无生产性的分析方式并不是主要的分析的模式。生产性分析是指资料的搜集和分析同时持续地进行，这样比较容易发现主题，并提高研究的效度。

研究者应用分析的方法也有不同的做法。有的研究者喜欢在作最后密集的资料分析之前，让自己暂时和研究资料分隔一段时间，即暂时将研究分析的工作放置一旁，以整理和理清自己的思绪。有的研究者在完成现场工作和结束资料的搜集工作后，即开始密集地分析资料，因为其间等的时间愈久，如果想回去请报道人检验资料中的要点，将可能会愈困难。有些研究者在资料分析过程中，甚至在资料分析后并写成研究报告初稿时，仍然和报道人保持联系，或请报道人阅读报告的初稿并提供回馈和建议，作为研究效度的检验。

定性研究的分析有两种基本形态：生产性分析和无生产性分析。生产性分析是指资料的搜集和分析同时持续地进行，无生产性分析指在收集完资料后再进行分析。研究者必须思考如何以有意义的方式处理现场呈现的大量资料。从资料中发现有关研究场所或人们的主题和重要概念，这样的过程可能引发某种解释，并可能引发进一步的探究问题。

发现主题和概念

定性研究的资料分析过程就像穿过森林发现一条路径，是对于研究者思考能力的一种测试。研究者必须思考如何以有意义的方式处理现场呈现的大量资料。研究者一件一件地检视资料，这些资料可能是十几页的现场活动描述记录，也可能是报道人讲的一句话，研究者须运用选择性的知觉、批判性的思考、或常识性的判断，持续地比较、对照和评价资料，从资料中发现有关研究场所或人们的主题和重要概念，这样的过程可能引发某种解释，并可能引发进一步的探究问题。

定性分析的几种理论

在资料搜集的过程中，持续地从资料中发现主题和发展概念，通常是一个很困难的过程，并没有简单的定规可依循，但研究者可参考下列理论去检查和分析资料。

分析的归纳 分析的归纳（analytic induction）是搜集和分析资料的一个取向，以及发展和检验理论的一种方式，其进行的程序是：

1. 在研究的初期，发展某个特定事实或现象的概略定义和解释。
2. 将这些初步发展的定义和解释引入搜集到的资料中。
3. 在资料中遇到不适合已形成定义和解释的新事例时，即修改定义和解释。
4. 主动寻找可能不适合已形成公式之定义和解释的新事例。
5. 找寻其他大多数资料反面的事例，并分析探寻每个反面事例的意义，反复地定义现象和形成解释，直到建立一个较为普遍的关系。

因此，简单地说，分析的归纳是依靠提炼资料而作成归纳，而不是依靠计算某个事实的次数而形成通论。

敏感的概念 研究者在定性研究中发现的主题方法也是一种寻找“敏感的概念”方法。这种理论强调在定性研究中注重发现和提炼研究中重要的“敏感的概念”。实证科学中的理论须有效地和经验世界结合才是有价值的，理论的目的是在于产生经验世界的分析架构。敏感的概念是用来警告研究者注意经验世界的一般特征。研究者需要仔细地检验现象的独特性和其与环境其他现象的关联性以发现“敏感的概念”。经由持续的检验，概念即被测试、改进和例证。

实地理论 实地理论（grounded theory）的基本观念是从现场参与者的观点看他们的生活世界，分析归纳他们真实生活的主题和概念，以了解他们组织其生活世界的观念架构。实地理论的发展取向是一种“持续比较法”，其目的在于从资料中归纳性地发展和建构理论。

在资料搜集的过程中，研究者可参考分析的归纳、敏感的概念和实地理论等理论去检查和分析资料。

2 定性资料的编码

我们从一些研究资料中择取该种类别的资料实例，说明每种编码类别的特性。

脉络编码（**context codes**），意指研究场所、主题、或研究对象的一般资料，这些资料显示出研究的较大脉络。

情境定义编码（**definition of the situation codes**），意指研究对象如何定义场所或特定主题，研究对象对于场所或研究产生的一般整体观点和看法，他们希望完成什么，他们如何定义他们所做的，什么是对他们重要的。

研究对象所持的观点，意指所有或有些研究对象对于现场特定方面的思考方式，其中包括共享的规则、标准和一般的观点。

研究对象对于人们和物体的思考方式的编码，意指研究对象对于彼此、局外人和形成他们的世界之物体的了解。如一位师范院校学生试教实习的书面报告，其中描述她对于儿童的了解和感想。

过程编码（**process codes**），意指事件随时间改变的顺序，或是从一种状态到另一种状态的变化。有关个人、团体、组织或活动随时间的改变，这种改变至少形成两部分的顺序，其编码名称常是时期、阶段、步骤、生涯、年代或顺

序的关键点（如：转换点、过渡期）。学生生活史研究即经常使用过程编码显示研究对象的重要生活阶段。

活动编码（**activity codes**），意指经常发生的行为种类。这些活动的行为可能是不正式的。如编码名称可以是：“学校作业”、“开玩笑”、“在学校的早操”、“午餐”、或“班级旅行”等。

事件编码（**event codes**），意指发生在研究场所或研究对象生活中的特定活动，这些活动并不常发生或只发生过一次。如研究对象经常注意和谈论的事件。

策略编码（**strategy codes**），意指人们完成各种事情的方法、方式和技巧。例如，教师运用策略进行教学或控制学生的行为，学生运用策略通过考试或交朋友，或是校长运用策略解聘老师或开放新职位等。这种编码须注意区分其策略的认定，是出于研究对象自己的判断或是研究者的判断。

关系和社会结构编码（**relationship and social structure codes**），意指人们之间通常呈现的行为类型，这些类型并不是组织表上所正式定义的；或是指较正式定义之社会角色和关系。如“友谊”、“师生关系”和“学生间的友谊”。

方法编码（**methods codes**），意指研究的程序、问题、喜悦和困境等。本书的研究程度方法之各章节标题即是方法的编码，在记录中的观察者评注或备忘录亦可被编码为“方法”。

分析资料可以从编码入手。编码就是对已有资料进行归类 and 整理。根据资料类别的形态和资料的种类。

3 分析资料的过程

资料的分析工作是将资料分类、筛选、建构和再建构成如上节所述的各种类别，这样的分析工作过程是辩证的，其中须处理的问题，例如：一个主题或类型是可辨认的吗？某个顺序或过程是明显的吗？如何确定概念之间的关系？

资料的分析是一种理智性与技术性相结合的工作，其分类概念有时来自于有关研究问题的文献、现有理论概念、现场工作中呈现的预感、或可能来自于常识性的怀疑和想象力。其实际工作方式则又涉及剪辑和建立档案的技巧。不同的研究者可能对照研究的目的、资料的数量、个人的喜好和习惯、以及可资利用的人力物力和时间等因素，选择运用下列的编码分析方式和步骤，也可能

自行创造另一种资料分析的工作方式。

资料的分析工作是将资料分类、筛选、建构的过程，不仅依赖研究者研究的技能和策略，而且依赖研究者的思考和洞察力。

发展编码类别

在搜集资料的过程中，研究者持续地阅读和检查每份资料，列出资料中呈现的每一个主题、概念、解释、类型或主张，并尽量明确地写下自己的想法。在这一过程中，研究者应能开始意识到哪些资料将可能适合某一种类别，一个编码的类别可能是一个重复的谈话主题，也可能是一个叙述性的主张，就将它们都列写出来，随着新的资料的增加，这些初步形成的类别将可能不断地被扩张、合并、调整和修订。

资料的编码

在每份资料首页上端编写可代表资料原出处的标题代号（例如：“观3”代表第三次现场观察记录；“访3”代表第三次的访谈记录；“文3”代表编号第三号的文件资料）。并且分别将每份资料编写上页数，以便在分析资料时可以对照检验。每份资料编上标题和页数后，每份资料至少拷贝成三份，一份用来作编码的试验，一份用来作最后的编码和剪辑，另一份则完整地留存为原始资料。

分类整理资料

分类整理资料是将已经按照每个类别编码的资料单位组合在一起，这是一个机械式的操作过程，其操作方式是剪下相同类别的现场记录、访谈记录、文件资料和其他资料片断，将它们放在已标明该类别名称的一个资料夹、大信封、袋子、或盒子里，以这种方式建立分类整理资料的档案系统，将会更充分地了解资料的脉络。

研究者在分类整理资料时，将会进一步比较有关主题、概念和主张等不同的资料片断，因而改进和提炼原有的想法。研究者可能发现有些含混的主题有待清楚地说明，也可能发现有些观念和主张并不非常适合资料的分析架构，而

应该准备舍弃这些观念和主张，并发展新的观念和主张以配合资料的架构。

在定性研究中，决定需要多少资料以支持一个结论和解释，并没有可依循的定规，这总是有待研究者自己的判断，有时候最好的洞察可能来自于少量的资料。有些定性研究者在写研究报告时，会提供有关统计数字证据。定性研究者可能提供对其结论和解释合理支持，但从不真正地证明这些结论和解释。

4 定性研究报告的写作

报告的内容要点

本节介绍和阐述定性研究报告的内容要点、撰写过程以及呈现形式。一般定性研究报告呈现的内容要点是：引言或介绍、研究的目的或主题、相关文献、研究的方法或策略（包括：研究场所、与研究对象的关系、资料的搜集、研究者的心境和资料的分析方式）、现象的描述和分析、讨论和解释、以及结论或涵义。

这些要点在报告中呈现的顺序、分量或内容并没有一定标准化，研究者在撰写研究报告时，可能参照其研究主题和研究报告的性质组织其报告的内容。一般发表在期刊上或出版成专书的定性研究报告，为了帮助读者了解研究的过程和评估研究发现的可信度，其研究者在报告中通常会告诉读者下列要点或回答下列问题：

引言 研究的背景和其中涉及的基本概念是什么？谁赞助这个研究？研究报告的呈现架构是什么？

研究的目的或主题 研究者想了解什么？想让读者们了解什么？研究的目的的是什么？原本研究目的是什么？它如何因时间而改变？报告中的每个部分如何关联到研究目的？所研究的文化现象的主题是什么？探究的焦点是什么？

相关文献 哪些观点、模式、理论、或学说和这个研究主题有关？研究的先前概念架构或理论观点是什么？这个研究和哪些学术领域有所关联？过去有关此研究主题的研究发现是什么？此研究主题和过去有关研究比较，有何相同或独特之处？

研究的方法或策略 研究场所：研究场所的物理情境是怎样的？研究场所的社会情境是怎样的？如何和为何选择这个特定的社会场所？如何进入研究场所？

与研究对象的关系：研究者在现场采取什么角色？如何采取这个角色？如

何与报道人接触和获得他们的合作？谁是主要报道人？主要报道人的特征？哪些报道人较容易接近？或较难接近？和各种报道人相处时，碰到些什么问题？如何处理报道人的怀疑或不信任？

资料的搜集在现场花了多少时间？整个研究进行了多少时间？搜集资料的主要方式或辅助方式是什么？采用哪一类型的参与观察、访谈、和文件搜集方式？如何决定何时观察什么或何时访问谁？对于要观察或访谈的内容，研究者有多明确的概念？在何时或以什么方式作观察或访谈的记录？是否使用录音机或其他仪器设备？什么事情影响到搜集资料的工作？现场工作有任何特别的困难或问题吗？如何检核观察记录或报道人的叙述？

研究者的心境：在研究的不同阶段，研究者个人对于现场和人们有何感觉？研究者个人如何去了解研究场所或报道人？在研究过程中，研究者个人的兴趣和价值观为何？是否经验到什么情绪上的压力？这些主观的想法和情绪如何影响到研究？

资料的分析：如何整理和分类所搜集的资料？使用哪一种归档和分析方式？在研究之前，研究者已经有哪种分析的观点？在观察和分析过程中，这些观点是否有所转变？搜集和分析资料的过程有多大程度的重叠？分析资料时，研究者经验到什么重要的困难或帮助？在分析过程中，什么是研究者个人的主观经验？报道人是否检阅资料的分析初稿？检阅后说了些什么？

描述和分析 现场和报道人的实际情况或现象是怎么样？正在描述的现象主题是什么和为什么？报道人持有或运用着哪些内在观点？报道人使用哪些知识和行为的类别？报道人如何建构他们的行为？哪些观察实例可呈现这些实况或类别？哪些报道人的谈话可呈现这些实况或类别？

讨论和解释 研究呈现了怎么样的社会、文化或教育意义？什么观点用来解释描述的现象和其意义？描述的现象和文献上哪些理论或学说有所关联？

结论或涵义 摘述整个研究过程和研究发现是什么？回顾整个研究过程，研究发现所呈现的主要意义是什么？研究发现对于有关的现象或理论，蕴含着什么意义？有了这次的研究经验，如果重新再作这个研究，将可能有什么不同？对于未来其他研究者，在研究领域或方法上的建议是什么？

以上所列举的乃是一般定性研究报告的内容要点，这些要点在报告中呈现的顺序、位置、分量、或内容并没有标准化，有的报告将上述要点分门别类地呈现，有的则将其中某些要点融合在一起，例如：将研究主题、研究方法和文献背景皆放在引言中说明；资料的描述和分析解释交互呈现在报告中；另外，随着定性研究的逐渐普遍，有的研究者可能假定读者对定性研究方法已有所了解，因此在报告中仅简要地说明研究的方法和步骤，或将研究方法的说明写成很长的注脚或放在附录中。

定性研究报告的撰写过程要先考虑所预期的读者，在资料搜集和分析之时，报告的起草工作即可持续地进行，写作的过程能有助于澄清思想和组织资料。研究者在开始撰写报告时，可先发展一份暂时的纲要或目录，再对照其中的项目发展写作的方向。描述和分析解释的部分是定性研究报告之核心，在报告中描述和分析常交互呈现，并用具体的例子沟通抽象的概念。报告的初稿写成后，须加以仔细地编辑和修订，让现场报道人检验报告是否呈现现场的真实现象。研究报告写完之后，需再作最后的修订，让专业人员评论此研究报告是否有意义的、完整的考虑到不同的观点、展现有意义的足够证据并以生动活泼的方式写成。

总结

定性研究的分析有两种基本形态：生产性分析和无生产性分析。生产性分析是资料的搜集和分析同时持续地进行，无生产性分析指在收集完资料后再进行分析。研究者必须思考如何以有意义的方式处理现场呈现的大量资料。从资料中发现有关研究场所或人们的主题和重要概念，这样的过程可能引发某种解释，并可能引发进一步的探究问题。

第十三章

两个案例

1 一个个案研究的实例

读者在阅读了前几章的内容后，可能就会看到，其实定性研究也是并不陌生的，我们平时做的许多研究都有定性研究的因素。下面我们具体分析两个定性研究的案例。第一个是一个个案研究的实例，其基本的研究设计是对一个学生进行长期的跟踪研究，以了解一种课程方案的实效。第二个实例是调查学生数学学习的投入情况，先进行课堂观察然后跟进访谈。

我们先介绍个案研究实例。其基本的研究设计是对一个学生进行长期的跟踪研究，以了解一种课程方案的实效。这一研究案例在设计上有几个特点：第一，用典型抽样方法确定研究对象；第二，用跟踪调查的方法对研究对象进行了几年的研究；第三，用多元的资料来源（对学生、父母等调查）及多种研究方法（观察、访谈等）增强研究的效度等。

背景

对迈克来说，坐在高中的教室里实在是件无比困难的事。一些课程他根本就跟不上。但是，在数学课上，他总是第一个做完考试卷的学生。他解释说：“我喜爱数学，而且每次都能在十分钟左右就做完试卷，可是我却学不好其他的科目”。

他在大学二年级时第一次听说“以经验为基础的职业教育”（Experience-Based Career Education: EBCE）。迈克坦承：“我确实是只为逃避我不喜欢的课程才去参加聚会的”。

但是听过 EBCE 的解释后，迈克很快被这个观念说服了。他不但喜欢“在工作中学习”的想法，而且认为这个方案可以让他以自己的速度来工作。没有学分、无需教师的这一想法，也深深地吸引了他。

迈克把一些描述性材料带回家给父母看。他们于是也陪着迈克去 EBCE 的

学习中心参加了一个晚班，以对方案有更深入的了解。现在，经过两年的方案训练，迈克已经是个高年级的学生了。他的父母也想让他弟弟参加这一方案。

去年，EBCE 的早期测验证实了迈克的学校经验相当参差不齐——有些学科学得好、有些学科学得差。他的阅读和语言成绩比他学校里低年级学生的平均成绩还要差，但在学习技能上他却显出高于平均水准的能力，而且证明了他在数学方面的超凡才华。

上学年初，在迈克开始进入 EBCE 方案中时，EBCE 工作人员认为迈克过于好动、谦卑、缺乏自信心、对健康和身体外表漠不关心。他们还判断迈克有严重的写作障碍。后来 EBCE 负责迈克学习的主任为他设计了一个学习计划，以建立他的沟通能力（包括在写作与人际关系两方面），同时鼓励他探索多项职业计划。这次，他把当地的动物园当做资源，访谈动物园的经理，并对阿拉斯加的棕熊进行了仔细研究。迈克还加入自愿参加的探索者童军俱乐部（Explorer Souting Club），定期帮助动物园做些事情。迈克说：“我确实喜欢和熊在一起。它们真顽皮。你们知道吗？当它们靠在栏杆上摩擦茸毛时，就像在拉小提琴。”对动物园学习计划的评鉴，是这一年中迈克最后完成的工作，从中可以看出他有很大的进步。负责他学习的主任对迈克的评论是：“你现在完成学习计划的比速度快。我认为你现在比过去花更多的时间来想把工作做得更好。”

迈克在生活技巧发展（life skills development）这一方面入门很慢。与其他的同伴一样，在脱离了通常来说严格的学校生活之后，他经验了一个所谓的“自由冲突”阶段。迈克意图摆脱他学习方案中较“学术”的一面。起初，迈克很少查问自己的任务，也经常不告知工作人员他正在做什么。到年底，他在这两方面都有了长足的进展。

依靠着每周的日记写作，迈克在文字沟通上获得了显著的进步，无论是在观点和感受的表达上、还是篇章结构上，都是成绩斐然。迈克的行为发生了有趣的变化。“我过去总是长时间看电视，从不读书。”第二年开始后，迈克说：“我去年读了两本书，今年夏天又看完了八本。现在我已完全投入到书中，而不再看电视了。”迈克喜爱的阅读材料是科幻读物。迈克对作业的态度也有了改变。

“以前在学校呆了六个小时之后，我不愿坐下来写作业。但是在 EBCE 方案中，我不是坐在教室里，所以我不在乎回到家里写日记，和做一些学习计划所要求的事情。”

迈克的个人发展也在不断变化。很多变化归功于他的指导教师。他的指导教师是个小学教师，她告诉迈克在工作场合穿上清洗过的干净衣服之重要性，当他的衣着改善之后，无论是指导教师还是方案工作人员都对迈克鼓励有加。

指导教师还告诉迈克她对他说的一切非常感兴趣。让他说话时放慢速度便于她听清楚。

在参加 EBCE 方案的过程中，迈克的出勤率提高了。那一年，迈克只缺席过 6 天。这比学习计划中其他人的缺勤率还要低，他们一年的平均缺勤率是 12.3 天。迈克的这一出勤率比他在高中时要好得多。

这一学习计划的开始，源自于工作人员意识到了它的必要性，因为迈克喜欢其他的同学，但好像缺少与同伴和工作人员的社交活动。学年初的报告显示他看起来依赖性很强、谦卑，是个不成熟的交谈者。透过这些观察，迈克的学习主任同他协商了一下学习目标，以及如何采取相应的活动，才能提高他的沟通技巧，同时有助于他解决一些人际关系的问题。学年结束时，迈克的沟通技巧有了明显的改善。他说：“我现在敢当众讲话了。”

迈克未完成的学习计划与他自己的经验和兴趣有关。在此之前，他已从加拿大移居波兰地区十年了，但是仍时常回去看望亲戚。这一学习计划是关于一个功能性公民区的移民法律和规定。同时，它还将有助于迈克提高他的文法和拼字程度。因为学生们可以选择在高年级完成从低年级开始的学习计划，所以迈克有机会在这年来完成这一学习计划。

谈起这一年，迈克说：“我变得比我原来想的要好得多。”他在 EBCE 的新经验中最喜欢的事情，是以自己的速度进行学习一项工作，及有更高的自由度。

学年结束时，从迈克的考试成绩可以看出，他的阅读和语言技巧都有了明显的改善。在他本就超出一般人的数学和学习技巧方面，则只有微幅的进步。

学年初和学年末都举行了态度测验，这些测验显示迈克在增强自信心、对于社会中各种角色的理解、对与自己不同背景、不同观点的人的理解、接受变化等诸多方面，都有了显著的提升。

迈克的志向没有改变。他仍想大学毕业后从事电脑程序设计工作。“刚刚开始这一学年时，我的确不太懂电脑。现在我觉得已经懂了很多，而且希望了解得更多，这样我才能够以此作为生活的工具。”

(迈克在 EBCE 第二年的描述被省略了。这一个案研究是从第一学年的描述中选出的。)

迈克对于 EBCE 的看法

迈克在报告中写到他在 EBCE 的经验，特别是在各个学习阶段的经验，使他的基本技巧得到了全面的提升。他感到他在每个学习场所都可以做他自己想做的事情。根据迈克所说，这些经验促使他选择自己喜爱的领域作为职业，也

使他开始关注其教育学到的东西，“比我在高中的四年所学到的还要多。”他甚至没有要求偿还他的旅行费用，因为他说他太喜欢这个方案了。

迈克父母的看法

当迈克第一次向他父母谈及这一项方案时，他们关心的是方案包含哪些内容，是否是个好方案，是否有教育意义。当三月份访谈迈克的父母时，他们感到 EBCE 方案已使迈克更加成熟，而且使迈克知道他未来将走的路。

迈克的父母说 EBCE 的工作人员把迈克各方面的学习成果都及时通知他们，迈克也常常把自己在 EBCE 方案中的活动告诉他们。而在高中时，迈克惟一告诉他们的是摄影。迈克的职业计划从他开始参与 EBCE 方案至今没有变化，他的父母并没给他施加任何影响。但 EBCE 方案已帮助他排列了机械操作员和汽车驾驶这两种职业的可能性。

自从迈克开始 EBCE 方案，他的父母发现迈克更加成熟了，也更加可靠和积极向上，他也开始思考和关心未来。他的写作水准有了长足的进步，阅读量也逐渐增加。

迈克的父母觉得 EBCE 方案在各方面都使迈克受益匪浅。他们对 EBCE 方案中的各个方面都给予高度的评价。

测验迈克的进步情况

尽管迈克在参加方案的第一年中，即在基本技巧综合测验各方面都有所进步，但是他第二年的成绩却相对地下降了。特别明显的是迈克数学和学习技巧成绩的下降。

虽然迈克在整整两年中的学习态度成绩一直呈上升趋势，但也在第二年参加方案时有所下降。迈克接受事后测验的成绩也在社区资源、成人、学习和工作规模方面，明显低于 EBCE 的平均成绩。

迈克两年中逐步提高了工作能力、自信心、沟通能力、完成任务的能力，以及心理社会成熟度的可信赖性。他在工作能力、完成任务能力、完成任务的社会责任感等方面的成绩，明显高于 EBCE 事后测验的平均成绩，只有在接受变化这一项上比平均成绩低。

迈克于毕业后接受访谈时，他每周在一家餐厅里工作六个下午，他前一年曾在那家餐厅兼职过。他希望能够在那里工作一年左右，一直做到厨师，然后到商学院进修一年的电脑课程。

2 一个课堂观察的实例

第二个实例是调查学生数学学习投入情况。这一研究的设计特点是：第一，先进行课堂观察然后跟进访谈，使观察为访谈提供基础，使访谈更深入获得信息。第二，在抽样的方法上，尽量体现量的研究的一些特点使之具有适当的代表性。第三，把定性研究和量的研究相结合，适当地用统计数据说明问题，并以此作为编制问卷的基础。

研究目的

了解小学生数学课堂学习的情况。研究者用三周的时间在数学课堂教学中对学生进行观察，了解我国学生投入模式以及他们是如何学习数学的，在课堂观察的基础上对学生进行谈话，了解学生投入数学学习有什么样的特点。具体地说，就是了解我国学生在数学课堂中做些什么和想些什么。

研究方法

在这项预研究中，研究者使用了课堂观察和课后访问的方法。

课堂观察 笔者首先在四所不同层次的学校各选择一个班级中听一次课，了解数学教学的基本情况和学生在课堂教学过程中的行为，并归纳出学生在数学教学过程中的基本行为表现（共10种行为表现）。在此基础上，选择一所学校中的一个班级进行两周听课活动，进一步跟踪研究学生在数学课堂学习中的行为。为集中起见，研究者在听课中选择一排学生（8名学生）记录他们的行为。

访问学生和教师 使用非结构化的问题在课后访问有关的教师和学生，主要了解学生在课堂数学教学和学习时的主要感受、想法和使用的学习方法，了解学生在行为过程中的情感和认知投入的情况。

研究步骤 本研究的研究步骤是：选择四所学校的一个班级听课，并分析听课的结果（花一周时间），课后访问学生和教师；选择四所学校中的一个班级连续听两周课，在课后访谈学生和教师；整理研究结果。

研究结果

观察发现,我国学生投入于数学学习过程的模式并不是单一的,学生首先在行为上投入于数学教学中。在同样的课程和教学条件下,不同的学生可以产生不同的行为。其次,在同样的课程和教学条件下,学生的情感体验是不同的。即使是具有大致相同行为的学生,其情感体验也有不同。再次,与学生的行为和情感体验相联系,学生在学习过程中产生的学习策略也是有所不同的,其中所反映的认知水平也有不同。行为、情感和认知三个方面构成了学生学习数学中的基本的模式。这一点,和国外最近的研究是一致的。另一方面,我们在研究中发现,中国学生在这三个方面的表现和国外的学生有很大的不同,这就使研究具有一定的中国化的特点。在观察中还发现,学生在课堂中参与的一个重要内容是人际关系的互动,主要是师生之间的互动。由于研究范围和时间等因素的限制,我们没有研究学生社会互动中投入的情况。

我国学生投入的基本特点

当我们以西方的理论框架深入学校研究在我国数学教学中学生投入的问题时,就会发现,我国学生在学习上有自己的一些特点。这些特点包括了我国社会性的特点和学校教学背景特点等等。

第一,被研究者年龄上的特点。虽然五年级的学生在行为和心理上的学习的方式已经形成,但是由于被研究者年龄较小,一般以中学生为研究对象的问卷很难应用于小学生。这一特点有两方面的表现:第一,小学生对项目的反映相对比较综合,一些国外对中学生的问卷项目难以在小学生中分化。因此,问卷的变量不宜太多和太细。第二,小学生对较长的问题往往理解上有一定的难度

第二,我国的社会特点。本研究是在上海进行的。在我国的大中城市中,家长普遍“望子成龙”或“望女成凤”,希望自己的孩子读书好,能考进好学校。这一点和我国数学学者关于儒家文化下的数学教育诊断是符合的。因此,长期以来,考试在我国的社会中占据十分重要的地位。对于学生来说,这一特点体现在两个方面。第一,我国学生数学学习的负担依然相对较重,学生一般在课外还花相当一部分时间学习数学。因此,课外学习时间是一个必须调查研究的变量,这一点和国外有所不同。第二,学生学习数学的目的,在很大程度上是为取得一个好成绩,以便升入理想的学校。学生在考试和成绩方面体验相对比西方学生多。学生对考试成绩的心理压力(或动力)比国外学生强烈。学生在学习的历程中不断地获得“成功”与否的体验。我们的

研究必须反映学生在这方面的情感体验。

第三,教学模式背景的特点。在目前正在进行的素质教育的改革中,一些新的数学课程和教学方法受到人们的关注。但是,我国的教学模式仍然具有以教师为主导的特点。我国教师一般教学方法比较细致,讲得多、练习多、管得也比较多,对学生的要求比较具体。另一方面,我国传统社会有“师道尊严”的传统,在一部分小学生心目中教师是至高无上的并且总是正确的。因此,长期以来,我国一部分学生形成了依赖教师的认知特点。对教师的依赖问题是学生学习策略的一个重要性的问题。因此,认知投入除了深层次策略和浅层次策略两个方面以外,还存在着思维的独立性或学习的独立性问题。研究还发现,我国学生行为投入不可能脱离教学模式的基本框架,是在特定教学环境中的投入。学生在教学中的行为表现和国外学生也有所不同,对教学过程学生行为的具体分析将是有意义的。

第四,家长的特点。在上海等城市中,家长大多对孩子要求和期望较高、管得较多。家长几乎比较一致地认为,只有孩子以后考进理想的大学才是出路。因此,学生在学习过程中对家长的依赖相对而言比较大。由于研究时间的限制,这方面情况不列于本研究的范围。

学生在数学课上的行为表现 根据第一周对学生的观察,数学课上学生出现的行为大致有以下十项:回答老师的提问(回答)、问教师问题(提问)、看书自学、和同学讨论与上课有关的问题(讨论1)、和同学讨论与上课无关的问题(讨论2)、做习题、听教师讲课(听课)、做与上课无关的小动作(包括眼睛看窗子外无所事事)(做小动作)、作教师指定的小实验(操作)、其他(包括准备工作等)。由于讨论1和讨论2在后继研究记录中无法很清楚区分,二者合并为“讨论”一项。其中,回答和提问用次数统计,其他用时间统计。为方便计,回答和提问在计算时间比例时忽略不计。

为了更具体地记录学生的行为,研究者选择8名学生(一排学生)进行两周的跟踪调查,表13.1是记录的结果。

表 13.1 学生数学课堂行为记录

学生	回答 次数	提问 次数	看书 小时	讨论 小时	做习题 小时	听课 小时	小动作 小时	操作 小时	准备工作, 布置作业等
学生 1			.3	.5	3	2.5	.2		.1
学生 2	8		.3	.2	3	2.7			.1
学生 3	7	1	.3	.2	3	2.7			.1
学生 4			.3	.2	3	2.7			.1
学生 5	5		.3	.2	3	2.7	.3		.1
学生 6	1		.3	.5	3	2.7			.1
学生 7	2		.3	.2	3	2.7	.3		.1
学生 8			.3	.2	2.5	3	.5		.1

该学生在听课时约有 .2 小时做小动作, 故不计算入听课时间, 下同。

从上述表中我们可以看出，第一，学生在课堂中的主要活动是听课和作业，占的时间很多，而其他活动相对占用的时间较少。第二，学生在数学课中的课堂作业占据了最多的时间。第三，从集体讨论的情况来看，学生被提问比学生自己提问多得多。从上述表中我们可以看到，上述行为可以归纳为二个大类：听课或者参与讨论和作课堂作业。如下表 13.2：

表 13.2 学生数学课堂行为二大类

回答	提问	看书	讨论	听课	做小动作	操作	
		准备工作等	操作	准备			作课堂作业
第一类活动				听课或者参与讨论		第二类活动	课堂作业

因此，在数学教学中，学生主要的行为表现分为参加集体教学活动积极努力程度和学生在解决数学问题中钻研的精神两个基本的方面。学生行为投入要反映这两个基本方面的质量。

总的说来，在同一个课堂教学中，学生对课程内容的反应是不同的和投入程度是不一样的。在行为投入方面大致可以分为在数学课上的表现和课外时间投入二个方面。在数学课上，学生的表现是各不相同的，这些不同可以概括为以下二个方面。

第一，在听课和师生讨论中的专心和努力程度（简称专心）不同。学生在投入上述活动的时候，有的同学听课比较专心，有的学生思想开小差；有的学生相对而言比较尽力，有的学生则比较随意。在上数学课的过程中，学生之间的注意的持久性并不一致，不同学生在课堂中积极学习时间并不相同。在课堂上参加讨论的情况也有不同，有的学生在课上基本上是听众，基本上不发言，而有的学生积极参与讨论，在课上很活跃，有的学生有问题喜欢问，有的学生则不问。

第二，学生在课堂练习中解决数学问题的钻研表现（简称钻研）也有所不同，有的学生做习题比较认真，有的学生做习题马马虎虎；有的学生遇到难题就不做了，有的学生虽然开始做不出，但能持续努力，直到做出为止；有的学生做习题兴趣很大，有些则不然。由于作业在数学课上占最多的时间，反映学生在课堂作业中的学习质量是必要的。

另一方面，调查发现，作为课堂数学学习的延续，学生一般在每天课外还要花一定的时间学习数学（包括完成数学家庭作业等），另外不少学生在周六和周日花一定时间学习数学。学生在课外学习时间上有较大的区别，教师反映，大约有 1/ 3 的学生在课外参加过数学提高班的学习活动，有一半以上学生都参加过家庭教师的补习活动，大部分的学生家长额外给予学生布置了作业。除了学校布置的作业以外，学生花在数学课外学习的时间大约每周 2 ~ 3 小时。

加上其他学科的学习负担，学生的学习负担还是比较重。

基于上述发现，我们把行为投入定于以下二个方面：一是学生在数学课上的表现，二是学生在课外的时间投入。学生在数学课上表现包括了二方面变量：第一，专心，主要反映学生在课堂集体教学过程中的专心与努力程度，如听课专心程度和参与讨论等；第二，钻研，主要反映在做课堂习题时的钻研情况；学生的时间投入主要指课外学生数学学习所花的时间。学生的时间投入包含了每天作业时间和每周额外学习时间二个变量。后继的研究已经证明了这两个时间变量有不同的意义。这几个方面大致反映了学生在数学课堂中行为的基本框架。这为编制调查行为投入的问卷提供了一个依据。

学生在数学投入教学活动中的情感体验

从上述的描述中，我们可以看到，从表面看，数学教学似乎是齐步走。而其实学生的行为表现有很大的不同，而且其背后学生的情感体验更有很大不同。从表1中，我们可以看到学生2和学生3行为表现差不多，数学期中考试的分数也差不多，但两位学生对数学的感受却是不同的。学生3对数学学习有一定兴趣，对数学知识很好奇，在学习的过程中喜欢一些思考性的问题，而学生2对数学本身并不太有兴趣，但她对自己要求比较高，取得好的数学成绩有一种成功感觉。

在听课结束后，研究者询问了学生对数学课上的情感感受，综合学生的反映有这样几种：第一，对学习内容和过程感到有趣。第二，虽然谈不上对学习有趣的感受，但完成学习任务或者取得好的成绩感觉到愉快和满足，第三，对考试和测验的焦虑，对考试成绩很担心，第四，对数学学习活动的厌倦。这几个方面和国外的研究是一致的，但是并没有国外的研究中出现的“愤怒”体验。然而，在这四个变量中，我们研究的学生表现和特点与国外的研究还是有所不同的。

第一种，乐趣感。有乐趣感的学生认为数学学习内容和数学课很有趣。如应用问题、巧妙算法和美丽的几何图形都吸引了他们。他们在学习的过程中不断地产生好奇心，而好奇心又不断地得到满足。例如，有学生说，我觉得上课中有的数学问题很有趣；有的学生说，我很喜欢解决应用问题，解应用题可以把数学知识和自己平时生活联系起来；有的学生说，爸爸妈妈总和我一起做数学思考游戏，我觉得很有趣；有的学生说，我比较擅长几何，所以一直对几何问题很感兴趣。有的学生说，在几门课中，我最喜欢数学课，觉得上数学课时时间过得很快。有的学生在上课以后说，数学课老师今天讲课很有趣，所以我觉得很开心；数学课堂上有一些讨论，同学们在一起讨论很愉快。

第二种，成功感。许多学生认为，虽然数学对他们说来并不十分有趣，但是他们努力学习主要是为了取得好的成绩，这一类学生认为好的数学成绩总是给予他们愉快的体验。有的学生说，爸爸妈妈对我学习数学要求比较高，我这一段数学成绩不错，我觉得很高兴。有的学生说，我的数学成绩有进步了，我很高兴。有的学生说，最近一次考试我的成绩在班级名列前茅，老师也表扬了我，我太高兴了。

第三种，焦虑感。有一些学生对数学考试和测验存在焦虑。由于各种不同的原因，这些学生害怕数学测验和考试，在测验的过程中表现出来不同程度的紧张，这种情感体验也影响了学生平时的学习。特别是在不理解数学知识或做不出数学问题等等情况下。这些学生有的很希望能学好数学，有一定的上进心。有的学生说，爸爸妈妈对我的要求很高，希望我的数学成绩好一些，我的成绩总是不理想，上课有的问题不会做，很着急。有的学生在上课以后说，我每次数学测验总是很紧张，以致影响测验的成绩，有一些本来能想出来的问题也想不出来了。有的学生说，我的数学成绩不好，我看到数学课很害怕，看到数学教师也不敢多说话。有的学生说，我每次测验很粗心，计算结果总是有错误，我对考试成绩很担心。

第四种，厌倦感。也有一些学生对数学学习活动感到厌倦。这一类学生表现为在学习数学的过程中无精打采，有的学生讨厌做数学题，有的学生感到学习数学很疲劳，对学习新的知识厌倦。这些学生得过且过，并无很大的上进心。例如，有的学生在课后说，我觉得数学课上练习太多，内容很没劲。有的学生说，我觉得数学除了对付考试以外并无太大的用处，所以我一做数学就感觉到没有什么劲头。有的学生说，在几门课中，我最不喜欢数学，学习数学使我觉得有些疲劳。有的学生说，我学习数学是为了完成任务，我对学习数学没有什么兴趣。

在研究的过程中，我们还进一步发现以下特点，1. 情感体验为“有趣”的那些学生中，在学习显示一种兴奋，喜欢思考和表达自己的观点。2. 对数学学习“厌倦”的学生中大多数是数学成绩不太好，从而产生对数学的恐惧。3. 虽然对同一个学生来说，学生对数学学习的体验并不是一成不变的，特别是随着学习内容而变化而改变，但在一定时期，学生在数学课堂学习中，学习基本的情感体现具有相对的稳定性。4. 对数学学习而言，学生上述的情感体验并不是只有一种情感体验。如觉得数学有趣的同学在一定程度上也会产生焦虑的感觉。5. 在一定的时间内，学生虽然可能有不同的情感体验，但是总是有一种主要的情感体验。

学生在数学课堂学习中的认知投入

我国学生在数学学习中的认知水平是否有差异呢？在表 1 中提到的学生 2 和学生 3 在认知方法上就有较大的区别，学生 3 学得比较活，比较注意知识之间的联系，对课外有趣的问题也会花时间去钻研，而学生 2 则学得比较死板，注重用反复练习的方法提高成绩。因此，这两位同学认知投入的策略和水平不同，但他们的成绩却是大致相同的。在小学五年级的小学生中，学生的数学学习中的认知水平方面已有一些不同，这些不同往往和学生所采取的学习策略有联系。在小学五年级学生中，学习的方式已经初步形成。并且，这种学习的方式也是在数学教学的过程中形成的，这种方式将会影响到以后的学习，是重要的学习素质，具有深远的影响。学生的不同的认知投入类型往往和教师的教学有关，也与其在教学过程中情感体验有关。综合起来，学生在学习数学过程中产生了不同的学习方法，这些方法往往和认知水平有关。这些方法大致有以下几种：

第一种，记忆的方法。这种方法的基本特征为背公式或方法。这些同学认为，记忆是学习数学的最有效的方法，背出基本的公式和方法是最重要的。

第二种，练习的方法。这种方法的基本特征为反复练习，在解题练习中学数学。学生的基本想法是，对于数学学习来说最重要的是练习，只有反复练习才能学好数学，没有练习就没有数学。这些同学通过反复和练习记忆概念公式以形成技能，他们认为数学成绩是练出来的。

第三种，测验的方法。这种方法的基本特征是把学习数学的过程作为一个准备测验（或者考试），进行测验和事后订正的过程。在他们看来，没有测验，就记不住知识。这些学生的基本想法是，学习是通过一个又一个测验来形成的，准备测验，进行测验以及事后订正等过程可以使學生掌握数学，形成技能。

第四种，审题的方法。这种方法的基本特征是把审题和分析作为学习数学的关键。这类学生的基本想法是，在小学数学中，应用问题是最重要的，学好应用问题关键是看清问题（包括条件和问题），并把这些资料和学过的数学知识联系起来，在这个过程中分析和综合是重要的。

第五种，归纳的方法。这种方法的基本特征是归纳出数学单元知识的要点，在此基础上加以整理、理解和记忆。这类学生认为，在老师讲课或者自己看书的基础上，都要归纳出若干要点，对这些要点进行理解和记忆是重要的。

第六种，依赖家长的方法。这种方法的基本特征是有问题找家长。学生的家长成了学生学习数学的陪读，由家长归纳出若干要点，并解答难题，学生的

一有问题，就找家长。离开家长的陪读，学生无可适从。

第七种，依赖老师。这种方法的基本特征是完全听从老师的安排，教师的教学方法就是学生的学习方法，教师的教学进度是学生的学习进度。这些学生认为，学习数学的最好方法是听从老师的安排，老师教什么，我学会什么；老师怎么教，我怎么学。

第八种，联系和理解的策略。这一类方法的基本的特点是把学过的知识和新知识联系起来。这类同学认为，学习数学要想一想这些数学知识在生活中有多大的用处，或者联想到以前学过的知识。这类同学学习比较活，思路一般比较广。

第九种，对学习方法进行总结的策略，其基本特点是经常对学习方法和解决问题的方法进行总结，对自己的学习过程进行一些反思。例如，这些学生如果用一种方法解题并无法求出答案时，会用另外的方法来尝试。在读数学课本时，经常思考其中有哪些是要切实掌握的要点，而不是简单地把课本看完。在学完数学章节时，总是回头看看哪些内容我掌握得不够。

第十种，自学的方法。这类学生认为，学习最好根据自己的实际情况和进度进行，自学是最好的学习方法。

对上述十种数学学习的策略可以归纳为三个大类，我们进一步提出了三种不同的问题：第一，学生是不是注重用联系和理解的方法学习数学？第二，学生是不是依赖于死记硬背的方法学习数学？第三，学生在学习策略方面是否具有独立性？前二种学习策略和 Marton 等（1976）提出的两种学习取向是基本一致的。后一种依赖的策略，是和我国的学生学习特点相联系的。对上述方法归纳如表 13.3：

表 13.3 学生认知策略的三个方面

浅层次策略		深层次策略		依赖的策略	
记忆的方法	练习的方法	审题的方法	归纳自学的方法	依赖家长方法	依赖老师
测验的方法		联系和理解的方法	对学习方		
		法进行总结			

实际上，学生这三种策略表现体现了三种认知的特点，反映了学生认知投入的基本情况。浅层次的策略、深层次的策略和依赖教师（或者家长）的策略，实际上表现了三种不同认知水平。浅层次的策略表现了死记硬背的和机械的认知水平；深层次的策略具有理解和反思性的认知水平。依赖的策略，依赖

自学的方法既可以是深层次的，有时也可能是浅层次的，归纳中没有计入自学的方法。

教师或者家长（在课堂教学的情境下，主要体现为依赖教师）策略，实际上体现了学生认知策略仍不成熟，从而认知水平并不稳定。当教师使用较好的教学策略时，学生的学习策略就比较好，当教师使用不够好的教学策略时，学生的学习策略也不够好。后继的相关研究中我们发现，依赖的策略更多地与深层次的策略有关。我们在研究中发现，这种情况在我国的小学生中特别普遍。表13.4归纳了这三种水平：

表 13.4 三种认知水平

认知投入变量	策略特征	认知水平
浅层次	死记硬背和机械的方法	较低层次的认知水平
深层次	独立理解、联系和反思	较高层次的认知水平
依赖	以教学策略为学习策略	认知水平不稳定

对学生在数学教学中投入模式的考察

研究的结果表明，学生的数学课堂投入模式（简称投入模式）具有多方面的特点。其一，学生具有各自不同的认知、情感和行为投入的方式和程度；其二，不同的学生可能具有不同的组合，形成了不同的数学学习模式，至少有以下几种主要的组合，在预试中投入模式的研究为进一步的量化研究提供了一个基本的线索。

一是高认知投入，高情感投入和高行为投入。这一类学生是我们所期望的，但这一类学生在调查的学生中实际上人数不太多，这是优秀学生的典型情况。也是数学教育所想要达到的目标和效果。

二是低认知投入，低情感投入和低行为投入。这是非投入的典型情况，也是相当一部分“差生”在数学学习上的表现，而在实际数学过程中，教师对“差生”的看法往往集中于他们的行为，如上课思想开小差，做习题不用心等。但是并没有关心“差生”的情感体验和认知策略。

三是低认知投入，低情感投入和高行为投入。预研究中学生的这种投入模式给我们以深刻的印象。一些学习成绩较差的学生属于这种情况。他们在数学上花费的时间很多，做的习题也不少，总而言之，是投入多，收效少。一些数学考试成绩较好的学生也属于这种情况，他们虽然考试成绩良好，但对数学学习并无兴趣可言，死记硬背成风，题海战术盛行。从这种学习情况中可以反思我们的数学教学。这种数学学习模式是否和“高分低能”现象有本质联系，还需要进一步的研究。

利用学生投入的概念，可进一步具体确定学生数学学习中投入模式，诊断数学中的问题，并初步预测数学学力。例如，同样是数学成绩不够理想，可能原因是不同的。同样学习成绩较好的学生，投入的模式也是不相同的。下面的双向表提供了一个工具。在表 13.5 中提到的学生 2 和学生 3 的学习模式如下：

表 13.5 两个学生学习算术应用题的投入模式

学生投入	行为投入			情感投入				认知投入			
	变量	专心	钻研	时间	乐趣感	成功感	焦虑感	厌倦感	高层次	低层次	依赖
学生 3											
学生 2											

可见，虽然这两个学生行为表现和学习成绩差不多，但情感投入和认知投入却是不相同的，学习模式是不同的。那么，他们的学业素质是否不同呢？目前学校考试成绩是否反映了数学学习的质的结果呢？这还需要进一步的研究。

通过上述观察和访问，我们对学生在数学教学过程中投入有了一个初步的认识。同时，我们可以对学生投入的模式作以下几个方面的思考。

学生在数学教学过程中的投入构成了基本的课堂学习的方式 我们在研究中发现，数学课程实施的过程，每一个学生以自己的方式作出反应，这种反应就是学生投入于教学过程的模式。学生在教学过程中并不是简单的接受者，而是一个积极投入其中的参与者。没有学生的这种投入，教学是不会发生的，学生在不同的方面以不同的形式投入数学学习过程之中，包括在认知、情感和行为各方面的投入，产生不同层次的数学学习（包括浅层次和深层次学习等）。

因此，同样在一个数学教学的教室中，学生的数学学习是不一样的。研究还进一步发现，数学教学并不是像表面上看到的课堂里面齐步走的样子，而是存在着一个个别学生的层面。从课程实施的角度来看，存在着一个个别学生的层面发生的过程。在同一个教室中，由于学生投入的模式不同，发生着不同的数学教学。如果没有这个个别学生层面的数学教学，不可能有学习的结果。这个过程证实了 Gage（1985）关于“学生中介反映”的论断。因此，对此研究是十分必要的。有关学生投入的研究实际上为研究这样一个教学层面提供了一个基本的线索。

学生投入模式是学生重要的学习素质 我们在研究中体会，数学教学不仅仅帮助学生形成知识和改变行为，而且每时每刻创造着学生对数学学习不同的情感体验和认知的方式，从而形成学生对数学学习活动的态度和 learning 的方式。因此，学生的行为、认知和情感的投入本身不仅是重要的学习变量，而且也是重要的学习的产物。它们是重要的数学学业素质，也是素质教育的重要内容。

学习不仅仅是学生行为的改变，而且是学生的情感体验和认知方式形成的过程。这三个方面形成了学生学习的基本取向，这种取向和态度是学生终身受用的。

在研究的过程中，我们同样注意到，在数学教学过程中的研究往往是教师的“教”的问题，对学生的“学”相对研究得比较少。数学学习评价比较关注的是学习者的知识方面的结果，对学生使用什么学习策略，达到什么思维的层次，似乎关心不多。这些都影响到数学素质教育的实现，不利于学生数学素质的形成。学生怎样投入于数学学习，甚至比学习何种数学知识更为重要。另一个方面还需要进一步研究的是，投入的模式和学习结果的模式是否有重要的联系，这是素质教育理论和实践两方面都需要研究的问题。

学生投入和教学策略改革 影响学生投入模式的因素很多，包括学生的目的取向、知识基础以及个性特点等。另一方面，教学模式和教学组织对学生投入有很大的影响，实施小班教学实验的学生投入的情况一般好于大班的教学模式中的学生。因此，改革课程和教学是提高学生投入的基本途径，学生投入是设计和实施教学的基础。

正当我们研究学生投入问题时，上海某报报道了这样一件小事：某小学五年级的学生因为在解方程中漏写了一个“解”字，被教师罚抄一千遍，一直抄到晚上十一点。这样的数学教学，令人扼腕，令人深思。在我们看来，这样的面目可憎的数学教学虽然促使学生的“行为投入”，但却又是造成学生认知和情感上不投入的重要原因。我们在预研究中发现，从学生投入的角度，可以从很多方面反思目前的一些教学策略，例如：

教学过程中的学生兴奋点。在研究的过程中，我们也注意到，在教学过程中，教师比较注意教学中有重点，但是没有注意学生的兴奋点，大多数的学生在一节课中甚至没有什么时间特别兴奋。这样，教学的重点仅仅是教师的重点，而不是学生关注的重点。我国学生的情感投入相对较弱，和课程设计没有注重学生的兴奋点只注重教学重点有关。因此，从学生的角度出发，合理考虑教学过程上的兴奋点并使之和教学重点的结合，是一个值得考虑的问题。

长效教学和学生可持续发展的问题。有的被访问的教师认为，需要进一步研究学生投入与学习结果之间的关系，这种关系是比较复杂的。学习结果有即时的效果，也有长期的效果。素质教育所要求的是一种长期的效果，或者说，我们需要的是一种长效的教学和课程设计，这就需要研究学生投入的问题。

数学学习的诊断问题。相同的知识错误，造成的原因可能是不同的，同样是学生不投入，其模式也可能是不同的。学生投入的方法为诊断数学学习上困

难的学生提供了一个基本的线索。

总之，这次预研究的结果使研究者初步了解了我国学生在数学教学过程中投入的基本情况，为建立问卷提供了一个基础。这一研究也说明了需要对国外的问卷作出一个修改，以符合我国学生的实际情况。

第十四章

行 动 研 究

1 行动研究的理论背景：观念的变化

严格地说，行动研究并不是一种独立的研究方法，而是一种教育研究活动，是一种教师和教育管理人员密切结合本职工作综合运用各种有效的研究方法，以直接推动教育工作的改进为目的的教育研究活动。这与 20 世纪 60 年代以后逐渐形成的研究型教师的理念有关。教师们比历史上任何时期都强烈而自觉地意识到，处于日益复杂的社会中的日益复杂的教育实践，要求他们为自己的实践活动寻求一种理性的基础，凭借个人的狭隘经验就能搞好教育的时代已经一去不复返了。“教师即研究者”、“教师即反思的实践者”的观念，正成为时尚，行动研究正是在这样的背景下兴起的。由于行动研究在国际上很盛行，国内也越来越关注，所以这里予以专章介绍。

行动研究并不是一种独立的研究方法，而是一种教育研究活动，是一种教师和教育管理人员密切结合本职工作综合运用各种有效的研究方法，以直接推动教育工作的改进为目的的教育研究活动。

20 世纪 70 年代以来，“教师即研究者 (the teacher as researcher)”、“反思的实践者 (reflective practioner)”、“反思性教学 (reflective teaching)”、“行动研究 (action research)”等等概念越来越多地出现于教育理论界的各种文献当中。这些表述形式道出了一个共同的声音：作为实践者的教师，这个与一切专业理论研究者一样有着自己特定的知识和思想，有着理解能力、认识能力和思考能力的人，应该结束长期以来的消极被动的“教书匠”形象，而代之以积极、主动的新形象。

近代以来，为适应社会经济、文化发展对人的素质之要求不断增长的需要，教育在社会生活中的地位不断被加强，其间重要的举措之一即是，制度化的专门教育机构——学校产生并逐渐完善。学校的制度化为有组织地、大量地人才培养提供了极大的便利，然而其潜在的消极力量也逐渐显露，终于在 20

世纪成为人们不得不认真关注的事实：伴随学校制度化而来的课堂组织的一致化、课程的系统化和教师职业的官僚化使得教育中技术化的观念出现并很快成为这一领域的“正统观念”，直到今天，技术的观点不但仍然是广大教师中普遍的一种观点，而且也是多数教育研究者理解教育的普遍视角。

在技术的观点下，教育被认为只是一个传授系统（delivery system），教师能够关心而且必须关心的只是，对于给定的教育目的和教育内容，有许多可能的完成途径，研究者们和教师们的共同任务是确认在这些可能的途径之中哪一些具有相对更大的效用，以便在实践中予以选择和运用。——这种“纯”技术的教育使得教师们的视野被狭隘地限定于科目内容及其分发方式上，他既不能完全而深透地了解这些科目内容背后的目的，也不能完全而深透地了解他的教学活动在学生身上产生的实际效果。

事实上，教育发生于具有极大复杂性的社会情境之中，它既有特定的社会历史背景，又指向我们理想的某种未来，从结果上看，教育也是一种产生社会后果的“社会活动”，而不单单是个人发展的问题。教育的这些现实特性使得教育实践不可能被简化成简单的技术控制过程，教师也不可能仅仅是一个技术操作工人。教育本身以及教育所处环境的复杂性使得从事教育的人必须基于他们对实践的深思而做出许多决定，他必须对自己所处的环境、对自己行为的目标及其可能的后果进行审慎的思考和判断。为了确实地提高教育的效率，教师必须学会做出实践中的“决策”（decision-making），必须成为一个“决策者”（decision-maker），而不可能仅仅是某种抽象指令的简单的执行者。

在技术的观点之下，教育结果往往被简单地理解成教育目的的现实化，而教育目的则被想当然地理解成业已充分地包含于既定的教育内容当中，教师对教育目的分享也就相应地被简化成对教育内容的理解，一个教师只要对他要教授的教育内容有了相当的理解，似乎就完成了对教育理想的分享，从而就成为国家或社会思想的执行者了。然而这种被物化的“教育目的”与活生生的教育实践的差距使得教师实际上被拒于真正的教育目的之外，这种拒绝教师实质性地参与教育目的的做法导致了一个可笑的局面：作为“人类灵魂工程师”的教师，自己却缺少一个“灵魂”。教师“灵魂”的缺失产生的第一个教育上的悲剧就是：学校一日甚似一日地开始变成各种各样缺少主题的知识与技能的大拼盘，学习活动也一日甚似一日地开始变成各种各样不明方向的记忆与复述，——当人们为教育这部列车做了种种修补加工、使它的速度越来越快的时候，却突然发现：终点站不见了。

传统的教育研究与教育实践显然难以解决这些难题。由于缺少当事人的参

与和过多地脱离了现实背景，传统的教育研究对教育实践而言越来越显得“不可靠”、“不真实”也“不民主”；而另一方面，20世纪50年代以来教育的发展也改变了人们对教育本身的看法，人们发现教给学生一些固定的知识已经不再重要，重要的是要让教学过程成为一个“问题解决”的过程，要让学习成为一种探究，传统的教学模式显然也不能满足这一需要。这样，改革，无论对教育研究还是对教师，都成为一种必需。

正如哈贝马斯 (J. Habermas) 所说，科学概括出来的研究知识并不能直接地驱使社会实践，还必须有一个“启蒙过程”以使某一情境中的参与者能够对自己的情境有一真正的理解，并做出明智而谨慎的决定。“心理分析学家没有权力为行动提供什么建议：当病人要行动时，他得自己作决定。”这种观点在教育上代表了一种对理论研究与教师的新理解：无论是教育政策，还是理论研究，要想真正地让实践中的教师们走上某种轨道，单靠外在地指手画脚是没有作用的，对作为人的教师的启蒙——不管以什么样的方式——必须成为这一工作的重要的一部分。

这意味着，在传统上对教师的“专业性” (professionalism) 的界定——对学科内容的掌握、必要的教学技能技巧之外，教师还必需拥有一种“扩展的专业性” (extended professionalism)，它的内容可能包括：(1) 把对自己教学的质疑和探讨作为进一步发展的基础；(2) 有研究自己教学的信念和技能；(3) 有在实践中对理论进行质疑和检验的意向；(4) 有准备接受其他的教师来观察他的教学，并就此直接或间接地与他进行坦率而真诚的讨论。总而言之，扩展的专业特征就是：有能力通过系统的自我研究、通过研究别的教师和通过在课堂研究中对有关理论的检验实现专业上的自我发展。

这样，一种新的教师形象就出现了：理论观念的倡导者和教师之间必须建立一种平等的关系，对教育与教学所做的各种理论上的说明必须使自己立于教师的检验之下，而教师也应该成为“科学共同体”中的重要一员。

教师之从事教育研究，在初期只是一些教育理论工作者的倡议，其意旨或多或少地都与缩小理论与实践之间的差距有关；但在今天，一位教师从事研

Carr and Kemmis, *Becoming Critical*, Deakin University Press, 1986, p.158 ~ 159 .

Stenhouse, L., *An Introduction to Curriculum Research and Development*, Heinemann Educational Books Ltd., 1975, chapter 10 .

倡议教师参与教育研究并不是一个全新的观念，早在20世纪早期，美国的白金汉 (Buckingham, 1926) 就在他的一本著作中提出了“教师即研究工作者” (the teacher as research worker) 的倡议，同一时期杜威 (J. Dewey, 1929) 也提出过“从事研究的教师” (researching teacher) 的概念，并指出，只有直接参与教育过程的教师积极参与了课程编制过程，课程编制才会有效。在30年代泰勒 (R. Tyler) 等主持的“八年研究”中，教师已经被作为研究群体的一部分。

究，已经不再完全是出于缩小理论与实践之差距的考虑，而是有了一些重要的现实背景了。

1. 20 世纪 70 年代以来，由于教育越来越被理解为一种会产生社会效应而不仅仅是促进个人发展的活动，人们认为教师有责任审慎地对待他们的教育实践，有责任对他们的行为进行反思，并有责任判断和超越他们实践的局限，这种责任兑现的方式之一，就是教师直接参与教育的研究。

2. 日益频繁而且日见深刻的教育改革使得新的教育思想、新的课程计划、新的教学方法和设施不断出现在教师们的面前，这些不熟悉的实践既要求教师知识结构上的更新，也要求教师情感与技能上的适应，而且，改革往往需要教师对这些“新生事物”做出评价与讨论，这种压力也成为许多教师从事研究的动因。

3. 最近几十年、特别是 80 年代以来，教师的“专业化”问题成为一个颇为教育界关心的话题，这个话题论争的结果之一，是使相当一部分人相信：教师的专业特性尽管不是全部、但至少在很大程度上乃是表现于一种“理智”上的特征，表现于一种对教育情境、教育过程和教育结果的深透理解与把握上，这种理解与把握的最有力而可靠的源泉当然就是“研究”。

4. 教育政策的改革带来了局势的变化，在教育上有着自由主义传统的国家自不必说，就是在一些原来是教育决策中央集权的国家，近年的改革也使教师和学校在课程等问题上有了更多的自主权和责任。这些自主权为许多教师从事研究提供了重要的现实基础。

教师从事教育研究，可有多种方式，例如，他可以像专业研究人员一样，确立一个课题，进行实验或调查，收集数据，最后分析数据，产生研究结果。但由于一般教师教学时间紧张，很难像专业研究人员一样有比较充裕的时间从事研究，加之一般教师对教育理论的背景了解不多，在研究的视野、分析的方法等方面可能会受到比较大的限制，所以，这样的研究方式在实践中往往并不可行。一般教师便于从事的研究，是从实践中来、到实践中去、在实践过程中进行的研究，以参与和合作为特征的“行动研究”，正是这样的研究方式。

2 行动研究的缘起与特征

行动研究的缘起与发展

尽管在许多介绍教育研究方法的著作中,“行动研究”被放到与“调查法”、“实验法”这类技术性的方法并列的地位,但是它与“调查法”、“实验法”却有着非常重要的不同之处:“行动研究只是一种进行研究工作的方式,而非一种方法,盖此种研究方法是在强调,由实际工作的人员在实际的情境中进行研究,并将研究结果在同一个情境中来应用,至于研究的设计与进行,仍须采用其他各种研究方法。”它的思想背景要比其他“研究方法”复杂得多。

“行动研究”作为一个术语出现,最初始自美国的柯利尔(J. Collier),他在1933年至1945年担任美国印第安人事务局局长期间,为探讨改善印第安人与非印第安人之关系的方案,让局外人士参与到研究过程中来,与他和他的同事合作,他称这种方式为“行动研究”。

20世纪30年代后期以后,许多社会学家和教育学家开始对行动研究发生兴趣,一些学者对行动研究的思想作了研究和阐发。早期行动研究得到比较系统的阐述得益于美国社会心理学家勒温(K. Lewin),勒温在他40年代的著作中比较系统地阐发了行动研究作为一种问题解决策略的目的、方法和步骤,在他看来,行动研究包括:(1)分析问题、搜集事实;(2)制定行动方案计划,执行它们,然后搜集更多的事实并予以评价;(3)这一整个行动循环的重复,螺旋的重复。50年代,行动研究的思想被广泛地介绍到教育领域,“教育行动研究”曾经在美国风行一时,柯雷(M. S. Corey)是这一时期教育行动研究的重要倡导者之一。

70年代以后,教育界关于教育理论与教育实践的关系、研究者与教师的关系等方面进行了经久不衰的争论和研讨,作为这场论争的结果之一,行动研

如贾馥茗、杨深坑:《教育研究法的探讨与应用》,(台湾)师大书苑公司,1988年版。

吴清明:《教育的科学研究方法》,转引自贾馥茗、杨深坑:《教育研究法的探讨与应用》,(台湾)师大书苑公司,1988年版,第105页。

cf. Kemmis, S., *Action Research*, in M. Hammersley (ed), *Education Research: Current Issues*, Paul Chapman Publishing Int., London, 1993.

Hammersley, M., *On the Teacher as Researcher*, in M. Hammersley (ed): *Education Research: Current Issues*, Paul Chapman Publishing Int., London, 1993.

究得到了很大的发展。到今天,在教育领域,行动研究已经成为一项声势浩大的国际性运动:澳大利亚,欧洲的英国、德国、奥地利、西班牙,北美,中南美,非洲,亚洲的印度、尼泊尔、菲律宾、斯里兰卡、泰国等地,都开展了各种形式的“行动研究”,理论上的研究也如火如荼,不但每年都有大量的行动研究的文献出现,而且还出现了行动研究的专利,许多杂志也出版了行动研究的专号。在我国,台湾省的教育学者 70 年代以来对行动研究也作了许多介绍与研究,大陆近年来也作过一些介绍,甚至已经作过一些实践中的尝试。

行动研究在其发展过程中、尤其是在 70 年代以来的发展中,受到了各种思想的影响。在英国和澳大利亚,斯腾浩斯 (L.Stenhouse) 70 年代中期提出的“教师即研究者 (the teacher as researcher)”的思想为行动研究的发展提供了一种重要的理论背景:而在美国,施瓦布 (J.J.Schwab) 70 年代提出课程探究的“实践模式”的思想、舍恩 (D.Schon) 80 年代提出的“反思的实践者 (reflective practioner)”思想,也为行动研究的发展注入了新的血液;中南美、非洲和亚洲一些地区的行动研究的发展则很大程度上得益于弗莱雷 (P.Freire) 的思想;80 年代以来,受德国哈贝马斯 (Herbemas) 提出的“批判的社会科学”思想的启发,以英国的卡尔 (F.Carr) 和凯米斯 (S.Kemmis) 等人为主要代表,提出“批判的教育科学”的设想,也作为一股力量,促进了行动研究的发展;同时,最近十几年来关于教师的“个人知识”的研究,如加拿大的康内利 (F.M.Connelly) 和柯兰迪宁 (D.J.Clandinin) 的研究,也在一定程度上加深了人们对行动研究的理解。

行动研究的基本特征

因为存在着如此众多的理论背景,在今天要为“行动研究”作一个明确的

Torsten Husen and T. Vevilie Postlethwaite (eds.) *The International Encyclopaedia of Education* (2nd edition), 1994, “Action Research”.

cf. Stenhouse, L., *An Introduction to Curriculum Research and Development*, Heinemann Educational Books Ltd., 1975.

cf. Schwab, J. J., *The Practical 1 - 4, in Science, Curriculum, and Liberal Education: Selected Essays*, The University of Chicago Press, 1978.

cf. Schon, D., *The Reflective Practioner*, New York, 1983.

cf. Carr, F., *For Education*, 1995; Carr and Kemmis, *Becoming Critical*, Deakin University Press, 1986.

cf. Connelly and Clandinin, *Teachers as Curriculum Planners: Narratives of Experience*. Teachers College Press, 1988.

界定已经很困难。行动研究无论是在实践上，还是在理论上，都已经形成了一个庞大的“家族”，很难用一个定义把这个“家族”中各个成员的具体旨趣、范围与途径都概括出来。不过，它们在一些特征上还是一致的。在目的上，“行动研究意在帮助实践工作者省察他们自己的教育理论与他们自己的日复一日的教育实践之间的联系；它意在将研究行动整合进教育背景，以使研究能在实践的改善中起直接而迅捷的作用；并且它意图通过帮助实践工作者成为研究者，克服研究者和实践工作者之间的距离。”

我们可以从下面三个方面来理解行动研究的基本特征：

为行动而研究 (research for action) 传统上，“教育研究”的旨趣是为了获取“真理”，这种旨趣假定关于教育真理的知识能够通过教育实践工作者很好地再现于教育实践之中，而教育研究的任务则是直接为这类知识的增加作贡献。虽然直到今天，这种研究观还是许多研究者甚至许多实践工作者所持的立场，然而最近几十年来对教师的研究却越来越清晰地发现：这种实质上是把教师视为一个简单“中转站”的观念对教育理论与实践的关系可能是估计得过于简单了，教师这个与专业理论研究者一样有着理解能力、认识能力和创造能力，一样有着知识与思想的“人”，在教育过程中的作用绝不仅止是某个“专家”理论的简单执行者。正如哈贝马斯 (J.Habermas) 所说，科学概括出来的研究知识并不能直接地驱使社会实践，还必须有一个“启蒙过程”以使某一情境中的参与者能够对自己的情境有一真正的理解，并做出明智而谨慎的决定。

行动研究本身就包含了这个“启蒙过程”，实践者不但直接参与了研究过程，而且在这个过程中，他是“科学共同体”中平等的一员，而不是某种“权威”教诲的聆听者。行动研究的根本旨趣不是为了理论上的产出，而是为了实践本身的改进。如施密斯 (D.L.Smith) 所说的那样，行动研究的精义在于：它是这样一种革新的过程，这个过程的目的在于某个人或某团体自己的、而不是其他人的实践之改善。因为“改善”是一个难有终结的目标，所以“为行动而研究”的旨趣要求行动研究是一个不间断的螺旋、循复的过程。

对行动的研究 (research of action) 尽管传统的教育研究也未尝不包括对行动的研究，然而行动研究“对行动的研究”却要现实得多。如果说行动研究“为行动而研究”的旨趣暗示了这种研究方式是“以实践为中心”的话，那

Torsten Husen and T.Vevilie Postlethwaite (eds.) *The International Encyclopaedia of Education* (2nd edition), 1994, "Action Research".

Carr and Kemmis, *Becoming Critical*, Deakin University Press, 1986, p.158 ~ 159.

David L. Smith and Terence J.Lovat, *Curriculum: Action on Reflection*, Social Science Press, 1990, p. 173.

么,“对行动的研究”则表明行动研究是一种“以问题为中心”的研究形式。“问题”(行动中值得研究的对象)在不同的理论流派看来可能并不一样,例如,在有些人看来,行动中值得研究的是这样一些问题,这些问题可以通过数据的收集、分析,通过某种技术的创制或应用而获得解决;而在另一些人看来,实践中最重要的问题是如何让教师了解他的行动“意味着什么(已经或可能在学生身上产生什么样的实际影响)”,这样,问题可能并不是固定的,而是随着研究的深入而不断变化的;另外的观点可能会认为行动研究应该关注的问题是“权力”、“平等”、“控制”等等相对更为宽泛、更政治化的问题。但不管是哪一种理解,问题的发现与界定都是行动研究的起点。

因为特定环境中的实践者所面临的问题总是特定的,所以行动研究中作为研究对象的样本往往也是特定的,而不必具有普遍的代表性。这个特点也决定了行动研究应该是有弹性的,而不是僵硬地遵循某一个严格的程序。它不但要求参与研究的教师掌握一定的研究技能,更重要的是,它要求教师有对实践问题的敏感能力、有适时调节研究方法或侧重点的应变能力。

在行动中研究 (research in action) 行动研究既不是在实验室里进行的研究,更不是在图书馆中进行的研究。行动研究的环境就是教师工作于其中的实际环境,从事研究的人员就是将要应用研究结果的人,研究结果的应用者也就是研究结果的产出者(至少是其中之一)。这双重身份整合在同一主体的身上,使得行动研究过程实际上成为教师的一个“学习过程”。正如柯雷(M. S. Corey)在总结自己的行动研究经验之后所说的那样,“行动研究是学习的一种途径”。教师在行动研究过程中通过不间断地对自己教学行为的直接或间接的观察与反思,通过与专业研究人员或其他合作者的交流,不断地加深对自己、对自己实践的理解,并在这种理解的基础上提高自己。这样,行动研究就超越了传统上对“研究”功能的界定——真理知识的增加,而成为“人的发展”的一个过程。正因如此,近年来,行动研究作为一种“专业发展 (professional development)”途径的作用越来越受到人们的重视。

3 行动研究的基本过程

尽管行动研究的不同理论背景使得行动研究有了许多模式,每一种模式由

Corey *Action Research to Improve School Practice*, Teachers College Press, New York, 1953 .

Grundy *Action Research as Professional Development*, Occasional Paper of “ *Innovative Links Between Universities and Schools for Teacher Professional Development*”, Australia, 1995 .

于理论上的假设不同,关注的问题也不一样,在实施行动研究的具体步骤上也有一些差异。但在基本的操作过程方面,行动研究基本上遵循了勒温(K.Lewin)确立的一些基本思想。这些思想包括:行动研究的起点应该是对问题的“勘察”——界定与分析;行动研究应该包含有对计划及其实施情况的评价,并在这种评价的基础上加以改进;从总体上,行动研究的进程是一个螺旋循环的过程。

对行动研究过程比较早期的论述大都提出了一些可操作化的描述,如柯雷在50年代初期为行动研究确定了五个相连续的步骤:(1)明确问题;(2)确立解决这个问题的行动之目标与过程;(3)按设计进行行动,并对行动做记录,收集证据以确认目标实现到什么程度;(4)对有关材料进行整理,概括出关于行动与目标之间关系的一些一般原则;(5)在实践情境中进一步检验这些原则。我国台湾省的学者也描述了一个非常操作化的行动研究步骤:(1)发现问题;(2)分析问题;(3)拟定计划;(4)搜集资料;(5)批判与修正;(6)试行与检验;(7)提出报告。

后期对行动研究过程的论述则不再如前期一样可操作,更多的倒是强调行动研究作为一种特殊的“研究形式”所应该具备的某种“精神”或“基本思想”。温特(R.Winter)在他的一本比较有影响的著作《从经验中学:行动研究的原则与实践》中,认为行动研究的关键过程可以用三个词表达出来,那就是:“观察”、“反思”和“运用”。凯米斯和麦克塔格特(Kemmis and McTaggart)对行动研究过程的描述更多地继承了勒温的传统:计划、行动(实施计划)、系统地观察、反思,然后重新计划、进一步应用、再观察、反思。格伦迪(S.Grundy)对行动研究过程的描述如凯米斯等的观点基本上相同,只是她认为,在“计划”之前,应该先有对问题的“勘察”。

应该讲,整齐划一的行动研究模式是既没有必要,也不可能的,因为整齐划一的模式本身就与行动研究的主旨不符。行动研究自然可以在遵循其基本精神的基础上有多种多样的形式,不过,作为行动研究的一个完整单元来讲,无论是哪一种对行动研究过程的理解,“计划”、“实施”与“反思”三个环节总

Grundy *Three Modes of Action Research, Curriculum Perspectives*, Vol 2, No.3, 1982 .

Corey, M.S., *Action Research to Improve School Practice*, Teachers College Press, New York, 1953 .

贾馥茗、杨深坑:《教育研究法的探讨与应用》,(台湾)师大书苑公司,1988年版。

Winter *Learning from Experience: Principles and Practice in Action Research*. The Falmer Press, 1989 .

Torsten Husen and T.Vevilie Postlethwaite (eds.) *The International Encyclopaedia of Education* (2nd edition) ., 1994, “Action Research” .

Grundy, S., *Action Research as Professional Development*, Occasional Paper of “Innovative Links Between Universities and Schools for Teacher Professional Development”, Australia, 1995 .

是不可少的。

计划

“计划”阶段主要完成的是：明确问题、分析问题、制定计划。

在教师的教育教学实践中，总会遇到一些需要解决的问题或难题（problem），比如，班上的学生对某门课的学习积极性可能不高，班上的课堂纪律可能不好，某种教学方法可能对某个班的学生不能产生预想的效果，想知道某种新的教学方法在这个班上应用会有什么样的效果等等。要解决这些问题，首先要对问题本身进行确认，要尽可能地明确这个问题的种类、范围、性质、形成过程及可能影响。比如说：

（1）我们遇到的是一个什么问题？是教学方法问题，还是学生的心理问题？是课堂纪律问题，还是学习动机问题？等等。

（2）这个问题是普遍的（几乎所有的学校、所有的班级都有这个问题，这个问题是这个年龄阶段的学生普遍具有的等），还是特殊的（只有我这个班上的学生有这个问题，只有这个门课有这个问题，只有某一个学生小团体有这个问题等）？

（3）这个问题的原因可能有哪些？主要是受社会大环境的影响，还是主要受学校学风的影响，或者只受这个班的班风的影响？是因为教学内容的原因，还是因为教学方法的原因，或者是因为学生心理方面的原因？等等。

（4）这个问题的存在对课堂教学的效果有什么可能的影响？会不会严重地影响到教学的效率？或者只是改进之后会比较明显地提高教学效果？等等。

对问题做了界定与分析之后，接下来要考虑的问题是如何解决这一问题，提出一个总体的计划。总体计划一般应该包含以下的一些内容：

（1）计划实施后预期达到的目标。目标的陈述要尽量可见可行，从小做起。比如要提高学生对某门学科的学习兴趣，应对“学习兴趣”这样一个比较模糊的概念进行更详细的分析，分解为一些可操作、可监测的目标，比如注意持续时间、课堂参与程度、提出疑难的数量、课后自学时间等等。

（2）对课堂教学试图改变的因素。比如为了提高学生对某门学科的兴趣，在充分分析问题形成原因与过程的基础上，可能考虑到要改变教学内容的呈现形式，可能考虑到要改变教师的语言风格，也可能考虑到要改变学生座次的安排等等。一般而言，为了便于分析研究结果，一次改变的因素不宜太多。

（3）行动的步骤与时间安排。行动的步骤设计是行动研究中非常重要的一个环节，为能适应没有预料到的效果和以前未曾认识到的制约因素，行动研究的总计划是足够灵活的，具有暂时性和尝试性。因此在研究的进程及采取行动

的大致时间安排上也要反映出这个特点。

(4) 本研究涉及到的人。课堂教学小而言之存在于整个学校生活中，大而言之存在于整个社会生活中，课堂行动研究不可避免地要涉及到一些可能与本研究有密切关系的人，比如校长、其他教师、家长、学生群体等等，行动研究计划应该考虑到如何处理与这些人的关系，如何与他们进行交流、获得必要的信息，如何尽量减少他们可能对本研究造成的不良影响，以及如何使他们配合本研究的正常进行。

行动研究的计划不是在实验室里进行的，它得以实施的背景就是课堂教学的活生生的实践背景，因此，制定行动研究计划时有必要充分考虑到现实中一些可能的有利与不利的因素，以期对计划得以实施的现实性有一个客观的认识。需要考虑的问题如下：

- (1) 教师本人有无从事研究工作的经验？研究能力如何？
- (2) 所遇问题在以往的研究或实践中已有什么样的解决方案？它们对本研究有什么参考价值？
- (3) 有哪一些可资商讨、交流或合作的人？
- (4) 计划实施可能会受到哪些人或事的影响？
- (5) 在多大的时间跨度上（几个月，一学期，一学年）可以实施这一计划？
- (6) 在所处的实际情境中，有哪些物力、人力方面的限制？

实施

实施阶段主要包括两个方面：行动及对行动的观察。

把计划付诸行动是行动研究的核心步骤。行动研究的根本目的就是要解决实践（行动）中的问题，改善实践（行动）的质量，因此，行动研究中的行动与其他研究方法（如实验法）中的行动相比，具有更大的情境性和实践性，它是在不脱离正常教学秩序的前提下进行的，因此，在行动研究的实施阶段，按计划行动需要注意到这样一个问题：由于教学实践受到诸多现实因素的影响，因而一方面教师应当尽量严格地按照原定的计划进行实践，而另一方面，教师也应当充分考虑到现实因素的变化，能够保证行动研究的计划在实施过程中具有适当的弹性，可以根据实际情况做出必要的调整。为了保证研究结果分析的客观性，对计划的调整应当做一记录，并说明之所以调整的原由：如果在实施计划的过程中发现原先制定计划时对现实条件的分析不充分甚至有误，最好返回头来修订计划，以保证整个研究过程能够有序地进行，并使研究的出发点与结果尽量好地一致起来。

在实施计划的过程中，另一个重要的任务是对行动情况进行观察和记录，收集有关资料，以便及时地对计划实施情况有一个大致的了解，并最终对本研究的过程和结果作出比较全面、深透的分析。根据课堂教学的现实条件，除了不可少的文献调查等方法外，经常用到的收集资料的方法主要有：

(1) 观察。具体而言，教师可以邀请自己的同事或者相关领域的专业研究者到自己的课堂中来，帮助对课堂教学情况进行观察和记录。同事之间相互帮助观察或邀请专家参与观察。一方面因为大家有比较共同的经验基础和兴趣点，可能更容易发现问题；另一方面也易于促进交流与合作，而且由别人来观察也比较客观。这种观察方法可能存在的不良效果是，另外一个观察者在课堂中的出现可能会影响课堂气氛和学生的心理定向。教师也可以委托一个或几个学生，请他们对课堂教学情况进行观察和记录，这样的观察与记录相对上一种方法可以减少课堂教学受到的干扰，但是由于学生对课堂教学情况的理解可能与教师有很大不同，记录的结果可能会出现偏差，而且请学生记录课堂教学情况，也会影响这部分学生的课堂学习。教师当然也可以自己进行观察，但有必要提醒的是，教师自己的观察要尽量保持客观，不要让先入为主的主观愿望左右了观察结果。最好用这几种观察方法同时进行，相互参照，效果会更好一些。如果有条件能够对课堂教学情况进行录音、录像，对行动研究的资料收集、结果分析将有很好的帮助。

(2) 访谈与问卷。访谈是获得对课堂教学之了解的重要手段，这种方法获得的资料可能未必能够进行严格的统计分析，但是却往往能够获得对于课堂教学问题的背景、成因、过程及影响的比较深刻的理解。访谈的对象视不同的问题可能包括一般社会人士、有关专家、学生家长、学校领导、同事、学生。为了获得比较充分的材料，在正式访谈以前拟定一个访谈提纲是必要的。可以在正式访谈之前，先选定一两个对象进行试验性访谈，待访谈结束以后，根据访谈的情况修订访谈提纲，然后再进行比较大面积的访谈。因为行动研究的过程总在某个幅度之内进行着弹性变化，访谈的主题、对象、人数等等可能会根据实际情况有一些变化。问卷是获得比较客观的、数量化的材料的重要手段，如果没有特别的困难，在行动研究中运用问卷调查是很必要的。无论是在问题分析阶段，还是在实施过程中，或者是在结果分析时期，都可以运用问卷获得有关的信息。

(3) 个人资料收集与记录。收集与记录个人资料也是获得对课堂教学之理解的重要方法。个人资料主要有两大类：学生的个人资料和教师本人的个人资料。学生的个人资料可能包括学生的家庭背景、简历、以往成绩表现、兴趣爱好、个性特点等等，视不同的问题，可能会设立新的内容项目。行动研究是以改善实践为宗旨的研究，因此鼓励教学实践中的重要主体——学生参与到行动

研究过程中来，并把自己在这一过程中的一些体验表述出来，不但对行动研究的进行具有积极的意义，而且也可以收集到许多有价值的资料，有利于全面地分析研究结果，以便为以后的进一步研究工作打好基础。

教师自己的个人资料也是行动研究的重要材料。加拿大教育学家康内利和柯兰迪宁通过对“教师个人实践知识”的研究发现，教师通过分析自己的实践经验、个人历史，可以获得对于自己所从事的教学实践的清晰认识，并在这一认识的基础上提高自己的实践能力。教师有意识地记录自己从事教学工作的所见、所闻、所感，并适时地分析和总结这些材料，往往能够获得对于自己实践的比较全面的认识。因此，我们建议教师在从事行动研究的过程中记录和积累自己的“教历”，主要内容可以包括：自己对某堂课或某个单元的设计，实施的情况，自己对实际教学过程中出现的一些问题的观察与分析，通过分析提出的可能解决方案，这些方案在下一堂课或几堂课中的实施情况等等。这样比较系统地记录教师个人对教学实践的设计、观察与思考，一段时间后进行分析与总结，不但能使教师对自己的实践有一个比较清晰的了解，而且也能在行动研究结果的分析提供真实的、体系化的材料。

反思

行动研究的具体展开尽管可能在实践中表现出不同的形式，但是这些具体展开过程是不是可靠，却是每一个从事行动研究的教师都应该考虑的问题。“反思”主要任务在经过一段时间的试验，收集了相关数据之后，对原先的“分析”、“计划”和“实施”进行必要的调整。调整需要建立在对行动研究的评价的基础上，对行动研究的评价，大致可以从以下七个方面进行：

(1) 问题界定是否明确？行动研究起点是教师在实践情境中遇到的问题，因此，关于这个问题的种类、范围、性质、形成过程及可能影响，必须予以明确界定。

(2) 行动的操作定义是否清楚？尽管行动研究与传统上的教育研究相比，具有更大的情境性，但是与研究有关的重要概念，必须要清楚地赋予操作性的定义，把概念的含义清楚地用可操作的行为表示出来，从而尽量减少主观臆断给研究带来的不良影响。

(3) 研究计划是否周详？计划要考虑各方面的因素，详细制定实施步骤，按部就班地予以完成。这一方面为计划实施后的总结与反思提供方便，另一方面也使整个研究过程增强可控性，减少随意性。

(4) 研究者是否按计划执行？有了周详的计划，接下来就要检查研究者是否按计划进行研究。一般来讲，计划一旦确定，就需要按部就班地予以实施，

如果因为对现实的估计不足需要变动，要把这些变动作为计划的修订，而不能随意进行变化，以免影响对结果的分析。

(5) 资料收集与记录是否详尽无误？因为行动研究一般规模不大且情境性很强，资料收集和记录要尽可能详尽无误，只有这样才能为解决问题提供有效的参考。行动研究是以解决问题为出发点，修改甚至假造证据是没有必要的。

(6) 研究的信度与效度如何？行动研究所收集的资料，不但要尽量准确无误，而且必须与研究目的等有关，否则将影响研究的“效度”，另一方面，如果研究涉及数次测量，测量的结果应该有很高的一致性，否则研究的“信度”就值得怀疑。行动研究虽然是针对实践中的情境性问题的，但不能因为这一特点忽视研究所必要的信度与效度，不能想当然地对待研究结果。

(7) 资料的分析与解释是否慎重恰当？行动研究所获得的资料，多是与某个特殊情境问题有关的资料，因此对于这些资料的分析 and 解释应该十分慎重，特别应该注意不要轻率地作出推论，应当力求现实地、恰当地分析与解释。

4 行动研究对我国中小学教师的意义

20 世纪 80 年代以来，我国中小学校开展了大量的教育改革与实验，许多由国内外教育理论家提出的教育思想、教育理论直至教学方法或技术在实践中进行了实验，同时，基于我国国情乃至基于当地甚至本校实情的各种创造性探索也不断涌现，呈现出生机勃勃的景象。在一些教育科研和实践水平都比较先进的地区，如上海，在“走群众性教育科学研究道路”的思想指导之下，十几年来，普教系统开展了大量的教育研究。这些研究成果丰硕，许多新的教育思路、教育教学模式出现并逐渐发展成熟，有些思路与模式甚至已经产生了全国性影响。

在这些骄人的成绩背后，我们也不能不看到我国教师们所从事的教育研究工作的一些问题。首先是研究的目的，尽管我国教师所从事的大部分研究确是始自实践中的问题，然而不在少数的这类研究的目的却是为了“理论”上的建树——这个目的本身并无可咎，但是它带来的后果却是颇值深虑的：“为改革而改革”的“改革”、为创制一种特别的“××教育”而进行的“研究”已经成为近年我国教育界的一个景观，实践者的这种理论旨趣使得“教育研究”在实践者手中突然急功近利起来，而且此“功”此“利”又偏偏不是现实中的实践，而是“理论”。这种倾向在对“理论”与“学问”有着某种特别嗜好的现

代中国也许颇多可解之处，不过它却着实削弱了这些研究的应有价值：因为目的（至少是最终目的）是为了“理论”而不是实践本身，所以一旦“成果”出来，就成为一种僵固的、似乎无须再改进的东西，只要专等别人来运用就行了，相应的，产生这种“理论”的那个实践也当然地就成了一个“模型”，只要等着别人来参观与学习就行了，也不可以再做什么变动，如此情势，不但与“实践改进”的意旨大相径庭，甚至可能已经成为“实践改进”的绊脚石；更何况，此种研究对“理论”的建树究竟有多大的意义也还值得怀疑。

其次是研究的形式问题，教师的研究即使再系统，它在根本上也只是意味着教师对自己实践的一种“省察”与“反思”，它的最大的现实意义在于它可以让教师“理解”（广义的一种“学习”）在他的实践中有着内在联系的各种要素的含义，从而使他的实践更具有有一种“理性”的特征。这意味着，教师的研究工作将具有与专业理论研究者很不同的一个特征，那就是一种始自现实的（而不是止于逻辑上的）“反思”或“评价”，这也意味着，教师的研究工作将不可能有一个凝固的“成果”，而必将是一个不间断的“过程”。我国教师从事的相当一部分研究在形式上还是循着传统上专业理论研究者的那一套，确定“课题”——寻找理论上的依据或做一两个调查——“课题”解决（研究也相应结束）。这个形式的潜在危险是，它最终可能脱离实践，成为游离于实践之外的、教师们教书之余的另一项活动内容。以团体（学校或教研室）为单位的研究工作一旦采用了这种研究形式，还将产生的另一个危险就是，由于缺少基于具体实践的“反思”与“评价”，研究的结果最后可能会变成一种武断的命令，成为这个团体中每个成员都必须遵守的某种规章，“为真理”而进行的“研究”带来的不再是“解放”，反倒成了附加的束缚。

正是因为我国教师所从事的研究工作中存在着这样一些已在或潜在的问题，我们觉得，引鉴、吸收行动研究的一些思想，无论对于我国教师的研究工作本身，还是对于我国教育实践和教育理论的发展，都还是很有价值的。

总结

自1997年始，受华东师范大学教育学系的委托，我们与上海市一所小学进行合作研究，其间，我们尝试按照“行动研究”的精神，以对方小学的教师（以志愿从事研究为前提，组成“教育研究志愿者组合”）为主体进行了一些行动研究的尝试。一段时间下来，取得了许多成绩，也有不少教训。以下把我们感触最深的几点总结一下，或许对其他愿意尝试进行行动研究的学校和教师有所帮助。

1. 关于“合作”问题。行动研究非常强调合作，这个合作可以是中小学

教师与专业研究人员的合作，也可以是中小学教师之间的合作。就我们的经验，在我国进行行动研究，至少现阶段，最好能有专业研究人员的参与，当然前提是以中小学教师为主体，专业研究人员不能“包办”。绝大多数我国中小学教师工作负担都很重，很少有时间阅读教育书籍、报刊，这不但大大地限制了他们分析问题的视野，而且在正确对待教育理论、运用理论语言等方面，也有一些欠缺，如果有专业研究人员的参与，这个欠缺不至于对研究造成太大的负面影响。

2. 关于研究的“成果”问题。行动研究的本义，是强调行动的改善，并不以理论的产出为根本目的，所以它强调行动研究的过程应是循环往复的过程。在我们的尝试中，我们强烈地感觉到：无论是有关学校领导，还是一般教师，往往一开始就把“研究成果”盯在一些外在的东西（如专著、文章等）上，而不太在意研究内在的一些价值。就我们的切身体验而言，转变“成果”观念，不但是进行行动研究的必要保障，而且能够产生比专著、文章更有实质性价值的效果：关注“行动”的改善、强调“研究”过程中的交流与研讨，不但会大大提高教师对研究的兴趣，而且，它对教师素质的提高——行动研究的培训价值——也能起到很好的效果，而这后一点，对我国中小学教师来讲，意义尤为重大。

3. 关于资料的收集。在我们的行动研究尝试中，感触颇深的一点是：我们的中小学教师在教育教学中表现出了非常突出的实践智慧，令人惊叹；但另一方面，很少有教师能够把自己的实践记录下来，作为分析、总结、反思的资料。我们在前期的试验中忽视了对教师进行资料记录、收集的必要帮助，结果给此后的总结、评价带来了很大的困难。后来我们帮助教师们制定了每一步行动需要记录的资料清单，并在与教师的研讨和交流中建议他们把一些有分析价值的事件记录成“案例”，这为我们以后的研究提供了很大方便。

后 记

本书为教育部“高等师范教育面向 21 世纪教学内容和课程体系改革计划”重点课题《公共师范课程建设与教学改革的理论与实践研究》的成果之一。教育部师范司组织专家组对本书进行了评审，评审专家对本书给予了充分肯定，同时对本书提出了积极的修改意见，现在的书稿是根据专家组的意见修改而成的。专家组成员如下：

组长：裴娣娜教授

成员：叶澜教授、朱小曼教授、林奇青处长、徐辉教授、缪小春教授
借此机会，谨对专家组的辛勤劳动表示衷心的感谢，并对师范司领导和高等教育出版社对本课题的完成和出版给予的支持，表示衷心的感谢。

孔企平博士为本书撰写了第 8、12、13 章，王建军博士撰写了第 14 章，韩晓燕副教授撰写了第 9 章，傅建明副教授撰写了第 10 章。

袁振国

1999 年 11 月 14 日