

现场研究中准实验设计的几个问题*

时 勘

中国科学院心理研究所

[提要] 本文从工业心理学角度,介绍了现场研究中准实验设计应注意的四个效度问题:统计结论效度、内部效度、构造概念效度和外部效度;接着,讨论了准实验常用的设计方法:1,一般性不可解释的设计,2,带有前测和后测的非处理控制组的设计,3,选择群体的设计,4,间断的时间-系列设计,并对这些方法的应用范围和需要避免的问题进行了探讨;最后,就进一步完善准实验设计方法提出了几点初步设想。

在现代心理学的许多研究方法中,准实验研究方法愈来愈得到广泛的应用。准实验设计是一种较难研究和执行的方法,从目前有关实验设计的研究文献所反映的情况来看,准实验设计虽然十分必要,但却不够成熟,而且在国内也介绍得较少。据此,本文在综述有关研究成果的基础上,尝试对现场研究中的准实验设计作初步探讨。

一、准实验设计的必要性

现场研究的实验设计一般可划分为两种类型:如果研究完全按照实验心理学的严密要求确定实验对象、反应、所用刺激水平及其组合,此类较为规范的实验设计被称之为真实实验设计(true-experimental design);然而,正如在临床心理学、社会心理学和教育心理学的研究中遇到的情况那样,在工业心理学研究中,特别是在组织管理心理学的现场研究中,实验者不可能把各种条件都控制到实验法所要求的程度。在此情况下,根据现场的实际状况,灵活应用实验法的基本原理而进行的实验设计,我们称之为准实验设计(quasi-experimental design)。

准实验虽然是在现场进行,而且不及真实实验控制得那样严密,但由于采用了准实验设计方法,使得它在解决问题上具有独特的功能,因而愈来愈被广泛地应用于心理学研究的不同领域,其主要原因在于:

第一,准实验设计方法与在现场研究中常用的自然观察和调查访谈在功能上有着明显的区别:后者一般只能回答“是什么”的问题,对事物发展也只能起到部分预测的作用,而准实验设计方法则要在尽可能控制条件的前提下,去回答“为什么”的问题,即揭示出实验的条件变化与结果变化的因果关系。这样,不仅能借助实验结论去预测结果,而且也能通过控制原因去产生某种结果。

①* 本文是在徐联仓教授的指导下完成的,林仲贤教授、陈永明教授、张武田副教授提出了宝贵意见,王新超同志提供了部分资料,特致谢忱。

第二，准实验设计方法拓宽了实验法的应用范围。一些过去被认为采用严格控制条件的实验不能进行研究的领域，采用灵活的准实验设计则可以达到研究的目的；此外，某些计划周密的真实验，在执行过程中因条件变化而夭折时，也可采用准实验设计的方式予以分析。

第三，在某些领域的研究中，如果设计得当，准实验将能得到比实验室实验更真实反映被试心理特点和行为规律的结果。

二、准实验设计的效度

前已述及，在现场研究过程中，由于准实验设计方法对各种因素的控制不可能达到标准实验方法所要求的水平，因而在解释实验的条件与结果的关系时，遇到了比真实验方法更大的困难。显然，解决这一问题的关键在于准实验的设计思想，即研究什么问题、控制什么条件，如何进行分析，怎样从结果中提出结论，所有这些内容中的核心是准实验设计的效度。换言之，如果明确了准实验设计的效度问题，研究者在实验的规划和执行过程中就有了正确的指导思想，这将有助于客观地解释实验结果。为此，有必要对准实验设计中最主要的几种效度及其相互关系进行探讨。

(一) 四种效度

实验设计的效度是指采用某种设计方法在获得实验的条件变化与结果变化的因果关系，并最终作出结论方面的有效性。准实验设计要求研究者能将实验法的相关知识与实际情况进行更灵巧的结合，就准实验的动态过程而言，最需要关注的是四个方面：实验资料的收集、统计与处理、实验变量间因果关系的确定，对因果关系的解释，研究结论的普遍性。针对这些问题可以用如下四种效度分别予以阐明：

1. 统计结论效度 (statistical conclusion validity)

主要考察在统计推理的基础上，能否得到有效的结论。统计结论效度也可视为对统计方法有效性的考察。这里，所选用统计方法的功能、测量和处理工具的信度、实验现场的不规则性、被试的异质性均会影响着此种效度。

2. 内部效度 (internal validity)

考察研究结论是否真正说明了统计中所暗示的因果关系，这涉及到研究者所揭示的因果关系是不是真正存在。在解释实验中各变量的关系时，研究者有时可能会把结果中某些变化的原因错误地归于实验条件（处理）的变化，而实际上，原因则是由第三种变量引起的。一般说来，准实验设计自身的周密与否往往决定着内部效度的高低。为了避免在解释因果关系上的随意性，研究者应当注意的因素是：被试的选择、成熟与补偿性平衡、施测时间、实验所用仪器和统计回归因素等。

3. 构造概念效度 (construct validity)

又称构思效度，指研究者所采用的理论或概念术语在解释实验业已揭示出的因果关系时是否有效。本效度被讨论的前提是因果关系已经确定，即实验结果确实是由处理（条件变化）带来的，内部效度已经得到了保证。这里要考察的是，不同的研究者可能会采用不同的理论或概念术语来解释这种因果关系，究竟哪一种解释是有效的或更为有效？因此，对概念的理解、方法的单一性、对实验条件的猜测、评价的理解、实验者的期待、实验程

序与处理的交互影响等因素都将影响着构造概念效度。

4. 外部效度 (external validity)

考察的是研究所获得的因果关系和结论在不同的个体、场所和时间等方面应用的有效性。外部效度强调实验所揭示的规律在客观条件发生变化时所具有的适用性。在准实验设计中,通过特定的被试样本、研究环境和时间所获得的因果关系,在新的外部条件下究竟具有多大的适用性,是一个值得谨慎处理的问题。为要提高外部效度,研究者应特别注意实验的处理与测定、选员、现场条件、历史等因素的交互作用。

(二) 效度之间的关系

第一,一种效度得以增长的某些途径和措施,可能会降低另一种效度。例如,采用严密的实验方法特别有助于提高内部效度,但因受多种条件的限制,可能会降低其应用范围,从而影响外部效度。这种效度之间的相互制约关系使我们考虑到,进行准实验设计时,应尽量避免这种相互干扰,并努力寻求将此种影响控制在最低限度的途径。尽管如此,一些交互作用仍然是不可避免的。不过,期望某一个实验满足所有效度的要求也是不现实的,甚至是不妥当的。

第二,内部效度在所有效度中居于最重要的位置,因为它考察实验所探讨的因果关系是否存在,是实施准实验设计的必要条件。

第三,除内部效度之外的其他效度,也可按其重要性排出顺序:构造概念效度、统计结论效度、外部效度。当然,这种次序也不是绝对的,要视具体的研究目的与内容而定。

(三) 准实验设计的标准

根据以上对准实验设计效度的探讨,可以认为,一个行之有效的准实验设计的标准是:

1. 它的实施程序应当是先后分明的;
2. 它能借助相应的统计推理方法明确而有说服力地解释潜在的因果关系;
3. 它应排除所有可能导致其它的因果关系的第三种变量;
4. 它将排除对因果关系在理论或概念术语方面模糊不清、易变的解释;
5. 它提供的因果关系及其结论应尽可能具备跨越人物、场所和时间等方面的概括性特征。

三、准实验设计的方法

这里,就准实验设计常用的几种方法作简要介绍和讨论。

(一) 一般性不可解释的设计 (generally uninterpretable designs)

此种设计在工业心理学中应用较广泛,但它多用于提出新假设,而在检验假设方面显得论据不足。主要有三种类型:

1. 只有后测的单组设计

又称为个案研究设计,设计模式如下:

$$X \qquad O_1$$

这里, X 代表实验的处理, O 代表观察, 1~n 分别代表处理或观察记录的顺序数。这种设计主要不能解决内部效度问题,其分效度也难以保证。

2. 带有前测和后测的单组设计

这是较常用的一种设计:

$O_1 \quad X \quad O_2$

也称为无控制组的前后实验,即被试事先接受前测(O_1),接受处理(X)后,再对被试进行后测(O_2)。采用此种设计要注意“历史”因素,即在两次测验之间,除处理之外,还发生过那些事件,是否会有影响。此外,统计回归因素也不应忽视,最后还要对一些时间因素的变化,如工作时间的改变、实验的起始和终止进行控制。

3. 只有后测的附加非等价组的设计

在现场研究中,当研究者尚未作好实验准备时,某个处理已被执行,由于缺乏前测资料,可增加一个非等价的控制组来保证实验的正常进行。如下所示:

$X \quad O$

O

这种设计存在的最明显问题是,缺乏实验前相同类型或等价组的资料,因此,组间的后测差异既可以归因于处理,也可以归因于非等价组之间的选择差别。这使得此种设计也被认为是不可解释的设计。不过,至少在三种情况下这种设计是可解释的。第一种情况是,控制组也可以接受可重复的、定量的施测,这样就可以让控制组也接受处理后测量其结果;第二种情况是,设计为分析型相关设计,亦可推理因果关系;第三种情况是,当获得的数据能予以确证,对某些产生了高交互作用的假设也可进行解释。

(二) 带有前、后测的非处理控制组设计 (untreated control group design with pretest and posttest)

与前一类设计相对应,本设计和后面将要介绍的二种设计均可归入一般性可解释的非等价控制组设计 (generally interpretable non-equivalent control group designs) 的范围。

带有前、后测的非处理控制组设计是一种广泛应用的设计方法,其主要设计模式是:

$O_1 \quad X \quad O_2$

$O_1 \quad O_2$

下面,我们来讨论一下这种基础性设计在控制组为非等价情况下的五种结果:

结果 1

这里所讨论的非等价情况是指接受实验处理组在前测中得分更高,而且在后测中,两组间出现了比前测更大的差异,如图 1 所示:

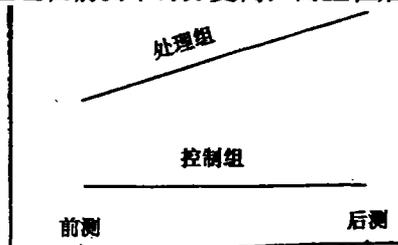


图 1 结果 1 示意图

从设计模式看,似乎控制了实验的所有因素,但此种结果至少存在三种对内部效度的

威胁：一，对比组中有一组成员在经验、疲劳或厌烦方面有更多的增长，如处理组更年轻、智力水平更高或更适于这种处理，即选择-成熟方面的威胁。如果采用匹配的方法，即把控制组得分高于平均数的被试与处理组低于平均数的被试相匹配，来达到前测中两组的平均分等同，但这样易使控制组的后测平均分向较低处回归，而处理组则向较高处回归；如果采用统计方法校正后测差异，其结果也只能反映处理和前测之间不清楚的结合效果。最好的分析策略是消除组间差异，计算出前测和后测的平均数，并把实验组别作为被试间因素，而把测验时间作为被试内因素进行双重分析，通过组别与时间的交互作用来说明处理的效果。二，测量的计分问题，如果前测分过高或过低，都会使处理后的变化难以分析。三，某种偶发事件只影响对比组中的一个组，这种不平衡也会造成对内部效度的威胁。

结果 2

该模式主要说明两组被试的分数都有增长，但增长率有差异（图 2）。这种情况也难以避免选择-成熟因素的交互作用。因为组间的增长差异同样也会反映在组内，如果组内和组间增长趋势是一致的，这将会危及对处理效果的认同。

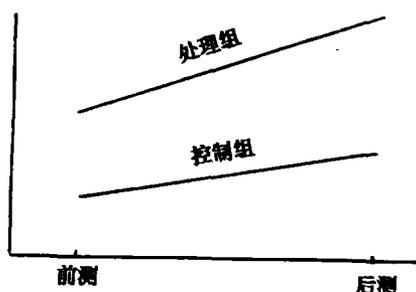


图 2 结果 2 示意图

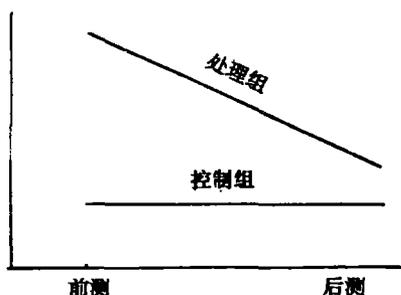


图 3 结果 3 示意图

结果 3

结果 3（图 3）与前二种结果有很大的区别：前测中在分数上具有较大优势的处理组，在后测中却与控制组明显地缩小了差别。这说明，开初的差异是没有意义的：可能两组之间本来就没有大的差异；也可能控制组也接受了处理。

结果 4

受试者是从较低的起点开始接受处理的，得到如下结果：

这种结果可能是天花板效应或历史因素的影响。

但是，此种结果比结果 1 更具说服力：因为处理组的前测分数本来就较低，如果能测定出操作较差组的稳定模型，就可以排除人们对处理效果在回归影响方面的怀疑。总之，稳定的、较低的前测分数比易变的、不规则的降低更能说明问题。

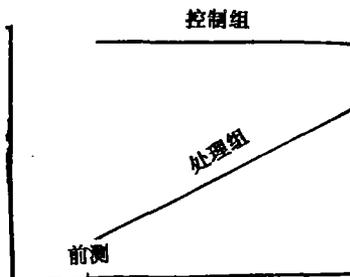


图 4 结果 4 示意图

结果 5

这一结果表明（图 5），两组得分从前测到后测在位置上发生了明显的转换，该模式的重要性不在于交叉，而是平均分数的差异转换。

这里，前测得分较低的处理组超过了前测得分较高的控制组。该结果能明确解释处理组达到高水平的原因，但要说明为什么超过控制组，则要应用回归技术，也在排除选择、成熟、年龄等因素的影响。

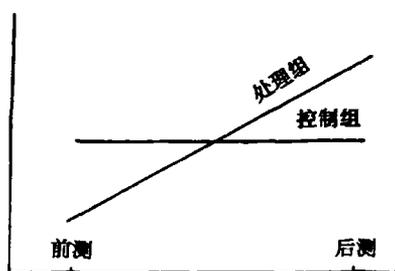


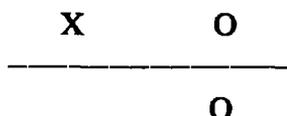
图 5 结果 5 示意图

(三) 选择群体的设计 (selection cohort designs)

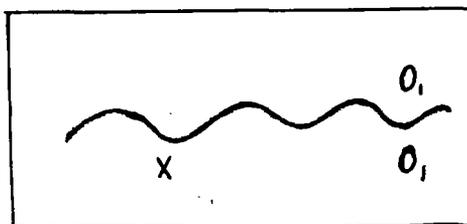
选择群体的设计在于寻找除了处理因素之外，尽可能少量差异的群体，以减少人为因素造成的选择性问题。如用学校中的不同年级、家庭中的兄弟姐妹来作为对比群体。这一方面可将经历过某种活动或具备某种经验的群体视为处理组；另一方面可找一些与之背景变量相似的群体作为控制组，从而增加他们的可比较性。下面，简要介绍该设计的三种主要形式：

1. 无前测的选择群体设计

该模式用于说明，对比组之间产生差异的原因在于其中一组没有接受处理。



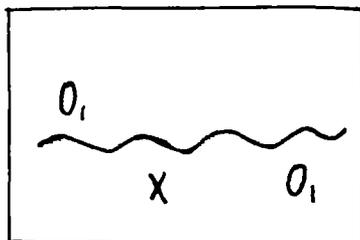
Minton (1975) 曾采用此法研究了儿童观看芝麻街节目 (Sesame Street) 后在社会行为上的差异。但她的实验设计 (见下面的模式) 忽略了兄长 (见下面的模式) 忽略了兄长 (见下面的模式) 忽略了兄长 (见下面的模式) 在出生次序、成长环境上与处理组的差异。因此，此类研究最好在同一环境的不同学习年限的体中进行。



2. 带前测的选择群体设计

为了避免选择-成熟因素 (如年龄、工龄)，需要安排一个前测控制组，使之与后测处理组在成熟水平上达到精确的相同：

Ball 和 Bogatz 在 1970 年曾采用这一方法。他们根据被试观看节目内容量分别将前测控制组和后测处理组划分为 4 种水平、8 个组，其结果充分排除了对内部效度的威胁。



(四) 间断的时间-系列设计 (interrupted time-series design)

在现场研究中，有时因引入某个处理会中断原有的数据模式，如果需要对处理前后的因果变化进行测定，可采用间断的时间-系列设计方法。该模式是 Campbell 和 Stanly

(1963) 以及 Riecken (1974) 等人提出的。主要有三种类型:

1. 具有非等价、非处理的控制组的间断时间-系列

该设计的基本模式是:

$$\begin{array}{cccccccccccc} O_1 & O_2 & O_3 & O_4 & O_5 & X & O_6 & O_7 & O_8 & O_9 & O_{10} \\ \hline O_1 & O_2 & O_3 & O_4 & O_5 & & O_6 & O_7 & O_8 & O_9 & O_{10} \end{array}$$

Lawler 和 Hackman (1969) 曾就上述模式进行过工作时间与出勤率之间关系的研究, 结果说明, 在现场研究中, 相同事件对所有人都会发生影响, 这种设计很难证实处理组受到了更多的影响。

2. 具有非等价因变量的间断时间-系列

上一类型设计对内部效度的最大威胁是历史事件, 我们可以通过收集因变量的时间-系列数据和其它变量数据来控制之, 其模式是:

$$\begin{array}{cccccccccccc} O_{1A} & O_{2A} & O_{3A} & O_{4A} & O_{5A} & X & O_{6A} & O_{7A} & O_{8A} & O_{9A} & O_{10A} \\ O_{1B} & O_{2B} & O_{3B} & O_{4B} & O_{5B} & & O_{6B} & O_{7B} & O_{8B} & O_{9B} & O_{10B} \end{array}$$

Ross 等人 (1970) 所做的关于英国呼吸测醉惩罚条例效果的研究就是根据上述模式进行的。研究结果说明, 这种设计解决了内部效度问题, 但却影响到构造概念效度和外部效度。

3. 转换重复的间断时间-系列

本设计类型的基本构想是, 假如有两个不等价的样本组, 在不同时间轮换接受处理: 当一个组接受处理时, 另一个组为控制组, 其基本模式如下:

$$\begin{array}{cccccccccccc} O_1 & O_2 & O_3 & & O_4 & O_5 & O_6 & O_7 & O_8 & X & O_9 & O_{10} & O_{11} \\ \hline O_1 & O_2 & O_3 & X & O_4 & O_5 & O_6 & O_7 & O_8 & & O_9 & O_{10} & O_{11} \end{array}$$

Parker (1971) 等人所做的有关电视引入与图书流通率的关系的研究就是应用该模式的成功范例。此类设计既控制了内部效度, 又保证了构造概念效度, 由于两个组在不同时间里至少共接受了两次处理, 外部效度也得到了保证。不过, 此类设计还得注意某些历史因素对内部效度的威胁。

四. 几点初步设想

本文所述情况表明, 准实验设计是现场研究中切实可行、且颇具发展前景的方法之一, 它扩大了实验心理学方法的应用范围。在应用准实验设计方法时, 应特别注意效度问题, 并根据现场的实际情况选择相应的设计方法。我们也应当看到, 由于准实验设计方法尚不够完善和成熟, 给研究者解释实验变化的各种关系带来了困难。从准实验研究现状看, 某些实验结果的解释, 其不确定性, 不合理性甚至错误定论等情况并不少见, 对此必须进一步采取对策, 否则, 将耽误准实验设计方法的发展。为此, 笔者提出几点不成熟的设想, 以求教于心理学前辈与同行:

第一, 加强实验心理学在现场研究等应用领域的方法学探讨。这里包含二层意思: 一方面应将实验心理学方法的最新成果迅速转化至各应用研究领域, 例如, 认知心理学有关

人类信息加工的理论与方法可用于指导人机系统专家决策模型的探讨；另一方面，应当从整体布局的角度，形成真实验与准实验的研究系列，以适应科学技术和社会发展所提出的新要求。

第二，加强新的理论、观点和方法在准实验设计中的应用。例如，为了提高内部效度，从目标与手段入手，剖析作为处理本身的有效性。如果要建立专家诊断决策模型，可应用多级模糊综合评判的方法，以便更精确地揭示之。当然，在采用某种新方法之前，事先要确定“质的规定性”，这样，新的理论、观点和方法才会“成为科学发展的强大工具”。

第三，加强电子计算机等现代新技术在现场研究中的应用。前已述及，准实验设计面临的^{最大}难题是对处于复杂背景中各种变化及结果的解释。随着微电子技术的发展，电子计算机已在心理学研究各个领域得到了广泛的应用。例如，国际上一些专用于社会科学研究的软件包，如 SPSS, OSIRIS, AIDA 等，用于分析现场研究的结果，能从多侧面、多层次的角度，更深入地揭示出现行分析方法难以涉及的问题，增强实验结果的可解释性。再如，新近引入我国心理学研究中的栅格分析方法 (repertory grid method) 就是计算机技术与心理学研究方法巧妙结合的范例之一，此法在我国管理者跨文化研究中已取得初步进展。笔者认为，此法在准实验设计的某些课题研究中，如解决认知复杂度、组织发展、职务分析、职业指导、培训与效果评估，是会更有助益的。这些正是我们在今后研究中试图探究的问题。

参考文献

- [1] 林仲贤、朱滢、焦书兰主编，《实验心理学》，科学出版社，1987年第1版，第1~13页。
- [2] 徐联仓、卢盛忠编著，《管理心理学》，科学出版社，1986年8月第1版，第153~179页，198~199页。
- [3] 邵阳，《明日的管理者研究-中国部分》，硕士学位论文，中国科学院心理研究所印制，1988年。
- [4] Campbell, D. T., & Stanley, J. C., In N. I. Gage (Ed), Handbook of reseach on teaching chicago: Rand McNally, 1963.
- [5] Smith, J. M., In Graduate Management Research. London, 1987.
- [6] Dunnette, M. D., Handbook of Industrial and organizational psychology, Rand McNally, chileago, 1975, 223-326.