

科技部科普专项经费资助课题

科普评估理论初探 与案例指南

张义芳 主 编

科学技术文献出版社

Scientific and Technical Documents Publishing House

北 京

图书在版编目(CIP)数据

科普评估理论初探与案例指南 /张义芳主编. -北京:科学技术文献出版社, 2004.9

ISBN 7-5023-4762-3

I. 科... II. 张... III. 科学普及-评估-指南 IV. N4-62

中国版本图书馆 CIP 数据核字(2004)第 085234 号

出 版 者 科学技术文献出版社

地 址 北京市复兴路 15 号(中央电视台西侧)/100038

图书编务部电话 (010)68514027 (010)68537104(传真)

图书发行部电话 (010)68514035(传真) (010)68514009

邮 购 部 电 话 (010)68515381 (010)58882952

网 址 <http://www.stdph.com>

E-mail: stdph@istic.ac.cn

策 划 编 辑 科 文

责 任 编 辑 金 平

责 任 校 对 赵文珍

责 任 出 版 王芳妮

发 行 者 科学技术文献出版社发行 全国各地新华书店经销

印 刷 者 富华印刷包装有限公司

版 (印) 次 2004 年 9 月第 1 版第 1 次印刷

开 本 850×1168 32 开

字 数 256 千

印 张 10.5

印 数 1~2000 册

定 价 20.00 元

© 版权所有 违法必究

购买本社图书,凡字迹不清、缺页、倒页、脱页者,本社发行部负责调换。

内 容 简 介

科普评估是保证科普事业持续健康发展的一项基本制度。本书在系统研究国外科普评估实践经验的基础上,研究借鉴教育和科技领域的评估理论与方法,系统解答了有关科普评估中的基本问题,主要包括以下方面的内容:科普评估的概念与作用;科普评估的结构框架与理论模式;各类科普评估的指标框架;评估程序与方法以及对我国未来开展科普评估工作的相关认识与建议。书中附录部分编译了国外有代表性的各类科普评估报告,相信对我国今后的科普评估实践具有很好的借鉴价值。

本书的最大特色在于内容的系统性和实用性,希望科普管理者、科普研究者以及从事科普工作的一线人员均能从中发现所感兴趣的内容,也希望本书能通过推动科普评估实践为提高中国科普工作的成效发挥一定的作用。本书可以作为科普评估培训教材和各类科普评估人员的参考书。

科学技术文献出版社是国家科学技术部系统唯一一家中央级综合性科技出版机构,我们所有的努力都是为了使您增长知识和才干。

科普评估理论初探与案例指南

顾 问 张景安 张晓原 胡志坚 武夷山

策 划 李 普 李永葳 邱成利 王 艳

刘育新 王 宇 许佳军

主 编 张义芳

执 笔 张义芳 邓国胜 张 晶

序 言

过去半个世纪以来,科普事业对提高我国公民的科学文化素养做出了巨大贡献。特别是近年来,中国各级政府对科普给予了高度重视和大力支持,对科普事业的投入显著提高,《中华人民共和国科学技术普及法》的实施更为科普事业的进一步发展提供了法律保证。然而,不能否认的是,由于多方面的原因,我国的科普工作长期存在着总体效率不高、责任不强、创新不昌、实效不明等问题,这些问题直接影响着我国科普事业未来的健康发展。面对科教兴国战略的深入实施、国内科教管理改革以及先进国家科普管理制度等的影响,健全我国的科普管理机制、通过实施科普评估制度提高科普工作成效已成为一个十分现实而迫切需要解决的问题。

我国科普工作的确有建立评估制度的必要性和紧迫性。在我国以往的科普工作中,科普管理部门与执行机构着重做什么事,开展什么活动,却没能顾及一个很重要的问题,即行动的效果或结果是什么,也就是做得怎么样的问题。比如,我国在科普事业上的持续投资到底收到了多大的价值?重要的科普活动或项目的执行效果到底如何?有什么值得借鉴的经验教训?如果继续开展,有哪些地方需要改进?科技馆等是否切实履行了其科普使命而物有所值?科教频道和众多的科普栏目的传播效果怎样?这些为科普管理所必不可少的信息,只有通过评估才能获得。而且评估本身也发挥着棒的无形管理作用,它能促使科普机构提高其工作的责任、

效果、效率和社会影响力,有利于实现资源的优化配置。可以说,缺少评估,任何管理都是不健全的。

评估是保证科普事业持续健康发展的一项基本制度。要建立这一制度,需要政府制订相关的政策法规去推动,而在此之前,则需要研究人员首先解决科普评估模式、框架、指标、方法等技术工具问题。2003年中国科技信息研究所受科技部政策体改司的委托进行的两年期的科普评估理论与实践研究课题正是基于这一需要而确立的一项重要研究任务。

然而,做好这一课题面临很大的复杂性和挑战,因为科普评估研究在世界上还属于一个尚在探索的领域。在课题的开始阶段,我们花费了大量的时间,从期刊和网络上查找国外资料,但所得甚少。我们发现,美英等国家虽然开展科普评估实践已有多多年,但发表的研究文献却寥若晨星。在这种情况下,我们一方面采取了与国外的评估机构直接联系,获取它们的科普项目评估案例,另一方面,我们转向范围更广的教育评估,查找了大量教育评估文献,从中研究分析可供科普评估借鉴的内容。事实证明,我们采取的策略是有效的,因为从我们掌握的国外科普评估案例来看,它们的科普评估中确实借鉴了教育评估的很多理论和模式。

本课题的最大难点在于确定科普评估的指标框架。科普现象的复杂性、科普情景的多样性以及科普效果的难测性,使得指标框架的确立相当棘手,不可能建立一个放之四海而皆准的、适合各种情况的统一的指标框架。为此,我们采取了分类评估的方法,针对不同类型的科普评估建立不同的指标框架,比如,对科普项目评估,我们分科技展览项目、媒体项目和科技活动项目三个类型,分别确立了各自的指标框架。在这方面,我们没有现成的东西可以照搬,因为我们所看到的国外评估手册都不涉及指标的内容,我们所掌握的国外科普评估报告也都没有评估指标的明确说明,只有靠我们自己去演绎归纳。另外,对科普管理部门、执行机构的评

估,由于国内外基本上都未曾开展这方面的评估,也完全靠我们自己探索研究,这些都是我们格外花心思的地方。总之,对这一课题,我们投入了两年的时间,进行了大量认真细致的研究,但由于研究是探索性的,还需要接受未来实践的检验和修正。

在进行课题研究的过程中,我们越来越深刻地认识到,制度化的科普评估对确保我国科普事业的未来发展意义重大,应作为推动我国科普工作的一个重要组成部分,列入我国科普事业发展纲要。科普评估制度势在必行,如能真正落实,必将促进科普工作绩效的提升,加快科普组织的成长,开创我国科普事业的崭新局面。

第一章 科普评估 :概念、类型 及其在国外的的发展

20 世纪 70 年代以来 ,评估开始受到国际社会的普遍关注 ,并广泛应用于投资、教育和科技等领域 ,国外科普界也日益重视评估的问题。然而在我国 ,科普评估尚未普及 ,科普评估研究才刚刚开始 ,整个科普界对评估的定义、评估的类型、评估的作用等基本问题尚很陌生 ,因此 ,本章首先介绍一些有关评估的基本概念。

第一节 科普评估的概念与作用

一、科普评估的定义

评估近年来得到广泛应用 ,尤其是在投资领域 ,已经成为政府、金融机构或建设单位等投资主体进行项目投资决策的基本依据 ,然而这种含义的“评估”与我们所关心的科普评估完全属于不同的学科体系。前者是应用经济学的一个分支 ,以成本 - 效益为核心 ,以增加利润为目的。而后者属于社会学研究的范畴。因为其一 ,科普工作总体上具有社会公共项目的共同特点 :非营利性和追求社会效益。其二 ,科普属于教育的范畴 ,科普评估采用的是教育评估的研究方法。因此我们对科普评估的定义来源于教育评估的描述 ,即 ,科普评估是根据委托方的明确目的 ,由评估机构遵循

一定的原则、程序和指标,运用科学、公正和可行的方法,对科普组织、计划或活动进行价值评判,并提出改进的建议和措施的专业化活动,其目的在于提升科普教育的品质。

科普评估是评估在科普领域的应用。它具有一般意义的评估所具有的特点:

第一,时限性。科普评估是在一定的时间范围内完成(如一周或一个月),而不是贯穿于项目、计划或政策的始终。这也是评估区别于监测的主要特性。

第二,系统性、规范性和科学性。科普评估与以往工作总结最大的区别就在于,评估是有一套完整的理论和指标体系,规范、科学和有目的的评定活动。而中国传统的工作总结往往缺乏系统的定性和定量指标,随意性较大。

第三,学习性。科普评估不仅仅是对科普组织、计划或活动的相关性、绩效和成功度的简单评价,而是通过评价活动本身进行学习。甚至可以说,学习加责任就等于评估。事实上,通过评估过程进行学习的知识更具有权威性,特别是对一系列的项目、计划或政策进行的评估,从中所学习的经验更适合在更大的范围内推广。

二、科普评估与监测的区别

监测与评估是人们在实际工作或研究过程中经常混淆的两个概念。

监测是一种持续性的职能,它是指收集信息数据,了解工作是否按原计划进行,即检查工作是否按计划投入资源、开展活动、产出成果的过程。这种职能主要是为了给管理部门和主要利益相关群体及时提供项目、计划或政策的进展状况,如果发现项目的执行与原计划不符,就要采取措施纠正偏差,使其恢复到原计划设定好的轨道。其着眼点不在于项目、计划或政策的目标结果上。

评估与监测的主要区别在于:

第一,评估是一定时限内的工作,而监测是连续性的职能,它贯穿于计划、项目或政策过程的始终。

第二,评估是根据具体情况,有选择地进行。也就是说,并不是每一个项目、计划或政策都需要进行评估,而监测活动原则上需要对项目、计划或政策都执行。

第三,监测是针对单个项目进行,而评估可以针对一个或几个项目、计划或政策进行。如行业评估。

第四,监测只需要记录实际的数据,用于与计划数据对比;而评估不仅可以利用监测的数据,通常也需要进行一些项目、计划或政策以外的数据对比。例如,与行业标准规范对比,或与同类同规模的项目、计划或政策的数据对比。

第五,监测通常通过定期的报告、报表、考察等方式,向项目、计划或政策的执行人员了解投入、活动和产出情况;而评估主要通过阶段性的座谈、考察或问卷调查等方式向服务对象了解工作的效果、效率、影响和持续性。

三、科普评估的作用

评估作为一种管理工具,具有品评鉴定、激励、导向、诊断、改进发展等多种功能。

- 激励的功能:可促进机构努力向上,以求最佳绩效。
- 回馈的功能:即检讨反省。评估是科普过程的一系列反馈环。评估能促进组织“省思”和学习。
- 品评的功能:根据事实评判其绩效。
- 改进的功能:评估的主要目的在改进。通过评估,可以了解计划或项目的优点与缺失,进而形成改进的意见和建议,使未来计划或项目的实施更趋完善、更有成效。
- 品管的功能:评估可使产出维持一定水准,亦即对计划或项目的执行过程具有品质管制的作用。

- 诊断的功能 :通过收集资料和分析 ,可指出计划、项目、组织中的问题和困难 ,作为改进的依据。

在科普领域开展评估目前有两大主要意图 :一是“证明”(to prove) ;二是“改进”(to improve)。所谓“证明” ,就是依据收集的证据 ,判断科普组织、计划或项目的绩效 ,也就是要证明组织运作、计划或项目实施的“有效”或“无效”。所谓“改进” ,则强调在评估过程中 ,了解组织、计划或项目的优点与缺失 ,进而形成改进的意见和建议 ,使未来的组织工作、计划或项目的实施更趋完善、更有成效。我们认为 ,科普评估应特别突出改进发展的目的。正如美国著名的评估学者 Stufflebeam 所说 :“评估的目的不在证明什么 ,而在力求改进。”目前的许多教育评估似乎多偏重“证明” ,而非针对评估对象的短处 ,积极地辅助其改善 ,最终达到受评对象自我成长的理想目标。故评估的目的是什么 ? 是一种品质保证、绩效责任、认可的制度、改良的依据 ? 或作为机构发展的参考指标 ? 这些是我们进行具体评估前所应明确的。

另一方面 ,要真正有效发挥评估的上述作用 ,评估工作本身首先必须是高质量的。为保证评估的质量 ,美国教育评估标准委员会(Joint Committee on Standards for Education Evaluation)制定了 30 项后评估标准 ,用以评估其评估工作。这套教育评估标准同样适用于科普评估 ,它共有 30 项 ,分为四个组群 ,分别是效用的标准、可行的标准、正当的标准及精确的标准。

- 效用的标准 :即评估应当是有用的 ,应满足评估用户的信息需要(在科普领域 ,评估应能协助被评估对象确认其优点、存在的问题并提出改进的建议)。

- 可行的标准 :即评估应当是可行的。评估应运用可操作的、严谨的评估程序 ,所需的资源、人力和时间必须符合成本效益。

- 正当的标准 :即评估应当是合法和讲求伦理的。评估过程中应维护有关团体的权益 ,在出现利益冲突时能坦然处之 ,不妥

协,谨遵应有的道德规范。

- 精确的标准 :即评估应当是精确的。评估应建立在可靠的信息收集与分析基础之上,应清楚准确地描述评估对象的价值,显示评估计划、程序、结果等,并提供有效的研究结论与具体建议。

第二节 科普评估的类型

根据不同的标准,可以将评估分为多种类型。按照时间顺序,评估可分为预评估、形成性评估和总结性评估;按照评估者的来源,可分为自我评估、外部专家评估和参与式评估;按照评估内容,则至少可分为科普项目评估、科普计划评估和科普管理部门/执行机构的评估。

一、预评估、形成性评估和总结性评估

预评估是指科普项目或计划开始实施之前所进行的评估,也称事前评估。由于是在实施之前进行,因此实际上是对项目和计划可行性分析的评估。预评估的结果一方面可以决定项目或计划是否实施,以避免对低效科普项目或计划的大规模投入。另一方面预评估取得的数据可以作为基准线,在项目或计划完成后进行前后对比。

形成性评估是指在项目或计划开始后到完成前之间的任何一个时点进行的评估,也称事中评估。它的目的在于检查项目或计划的设计,或者评估实施过程中的重大变更及其影响,或诊断实施过程中的困难、问题,从而为项目或计划执行者提供改进的方向。

总结性评估是在项目或计划结束后进行的全面、系统评估,也称事后评估。总结性评估的目的是通过分析评估找出项目或计划的优点与缺失,并通过及时有效的信息反馈,为未来新项目提供经验教训。总结性评估一般更容易为科普机构理解和接受,也更受到重视。

预评估、形成性评估和总结性评估在评估时点、评估目的、评估主体上互有差异。在理想状况下,一次全面的项目或计划评估应当包括预评估、形成性评估和总结性评估三个阶段,但在实际评估工作中,由于经费和时间的限制以及评估目的的不同,经常有选择地进行。

二、自我评估、外部专家评估和参与式评估

根据评估者的来源,评估又可以分为自我评估、外部专家评估和参与式评估。自我评估即项目或计划的实施者进行的内部自我评估。专家评估是指项目或计划的实施机构聘请外部专家进行的评估。参与式评估是吸收受评对象(如科普项目执行者)和项目受益者(科普对象)参与评估实施过程的评估。在实际工作中,选择何种评估类型,将取决于评估的目的和具体的情况。

自我评估的优点首先在于评估者是项目或计划实施机构的成员,对组织内部结构与运行机制较为了解,对项目 and 计划实施的整个过程较为熟悉,对科普社区的社会经济背景较为了解;其次,自我评估的成本较低,在项目或计划实施的每一个阶段都可以进行评估,有利于及时发现问题、解决问题;最后,自我评估的结果或建议容易在执行过程中得以实现。这是因为一方面评估者本人就是项目或计划的执行人;另一方面他们对于评估过程中发现的问题有切身的体会。然而,自我评估也存在一些局限。由于评估者是实施机构的内部成员,因此评估结果往往缺乏客观公正性。而且,即使自我评估很客观,评估结果也不容易为有关方面完全相信。另外,由于自我评估者的日常工作较多,难以集中精力进行评估,因此自我评估较为粗略。

而专家评估则可以弥补自我评估的不足。专家评估往往较为客观公正,评估结果易于被有关方面所接受,专家的专业评估知识也可以增进评估的科学性与规范性;而且专家评估往往是集中一

段时间进行评估,效率较高。但是,专家评估也有其缺陷。首先,专家评估的成本较高,因而也不可能在项目或计划执行的每一阶段进行,更不可能对每一个项目都进行专家评估;其次,外部专家对项目或计划实施的整个过程、运行机制和背景并不十分清楚,而专家往往又需要在较短的时间内完成评估工作,因此有人戏称专家评估为“旅游式评估”;最后,专家评估的结果或建议并不容易在项目或计划的执行过程中得以真正落实。

可见,自我评估的长处正是专家评估的短处,而自我评估的短处正是专家评估的长处。有时,在条件允许的情况下,将两种评估方式结合起来,评估的效果可能更好。参与式评估正是在此基础上演绎发展出的新兴评估方式。

参与式评估是指外部评估者吸收受评对象和科普目标群体的成员共同组成评估小组,对科普项目、计划等进行评估。

参与式评估要求受评对象和科普目标群体人员参与评估的领域、评估指标的选择、收集和整理数据、分析结果、将评估的信息用于目标的实现等全过程。而不是传统的由专家或项目管理机构得出评估结论。需要说明的是,参与式评估绝不是为了简单地增加评估的人员,而是通过受评对象和目标群体成员的直接参与提高评估的质量。

参与式评估的最大特色是打破了传统的思想认识和评估方式,即“自上而下”,迷信上级和专家权威的评估。而是鼓励受评对象和目标群体成员参与评估过程,充分调动利害关系人的积极性。

一般而言,参与式评估不是一种自发的行为。因此,为了进行参与式评估,必须运用积极的、操作性强的方法保证利害关系人的参与。如果可能,还需要对他们进行参与式评估的培训。

三、项目评估、计划评估和管理部门/执行机构的评估

如前所述,科普评估是对科普组织、计划或活动,通过科学系

统地收集、整理和分析有关的数据资料,判断其绩效及缺失,并提出改进的行动建议。因此,从评估内容上说,科普评估应至少包括科普组织评估、科普计划评估和科普项目评估三种类型。

在这三种评估中,科普项目评估应用得最多。科普项目是指在一定时间和一项预算内,为达到预期目的而实施的科普活动或任务。科普项目评估就是运用社会学研究方法,对实施前、实施过程中或已完成的科普项目的设计、实施、效果、影响及持续性等进行判定及评价,它往往涉及以下多个纬度的一个或几个方面:①项目的需求;②项目的设计;③项目的实施;④项目的效果和影响。

国家、地方或机构层次的科普发展战略规划或计划是开展各项科普工作的基础,可以保证从长期发展的角度来推进各层次的科普工作,使科普资源得到最好的分配和利用。尽管这种重要性已经被普遍认知,但在我国的科普领域,各层次的科普工作很少有制定长期的、战略水平的规划的。凡事预则立,不预则废。科普事业是一项长期的事业,不仅需要有长期发展的战略目标和规划,还需要对规划实施情况进行跟踪,对执行效果进行评估。这方面,美国政府绩效成果法推行的政府绩效管理制度值得我们借鉴。

至于科普管理部门/执行机构的评估,应主要侧重于能力和绩效方面。我国的科普组织总体上还很薄弱,组织能力长期欠发达,同时缺乏绩效责任机制,严重制约着我国科普事业的发展,有必要通过开展制度化的评估促进科普组织的学习和成长。

第三节 科普评估在国外的兴起

科普评估在我国只是近年来才受到科普组织管理部门的关注,但在美英等发达国家,科普评估已经开展了一段历史,特别是20世纪90年代以来,由于这些国家逐步推行政府工作绩效评估制度,有力地促进了政府科普计划和项目的评估工作。

美国是世界上最发达的国家,也是科普事业最为成功的典范。美国政府的许多部门和机构,如国家科学基金会(NSF)、国家航空航天局(NASA)、能源部、商务部、内务部、农业部、教育部、国防部、卫生与公共服务部等,都负有开展科普工作的使命,其中以 NSF 和 NASA 的作用最为突出。

美国 NSF 的科普工作主要是通过其非正规科学教育计划实施的。NSF 有一整套严格的项目审批和评估制度。一方面,NSF 在其非正规科学教育(ISE)项目征集通告中明确规定:项目申请人所提供的项目申请书中必须包括严格全面的项目评估计划,即评估所用策略、方法、时间安排、预算(必须准确地反映内部评估和外部评估所需费用)以及评估者的姓名和资质证书等的详细说明,并且如果条件许可,应进行三个阶段的评估:项目启动前的预评估、项目进行中的形成性评估和项目完成后的总结性评估,总结性评估必须反映科普项目对参加者在科学知识、态度、兴趣和行为上的影响。评估计划必须与项目目标保持一致,尽可能采取实验项目的方法来获得预评估数据。另一方面 NSF 还专门设立了由外部专家组成的评审委员会,根据项目申请书、项目年报和项目终期报告等资料对项目绩效实行同行评议,实现对项目的外部监控。

美国国家航空航天局(NASA)的科普活动主要由其教育处负责组织实施。1991 年该处成立了评估部,承担全局教育项目的评估管理。NASA 的科普项目申请方式与国家科学基金会类似,在项目申请阶段就要求申请人提出有关项目绩效和影响力的评估计划。通过其成熟的项目评议体系来进行选择,开展广泛的内、外部评估来控制科普项目的绩效。同时 NASA 与外部评估机构建立长期合作,如:从 1998 年起与 Lesley 大学项目评估与研究组合作,对科普项目及其组织效率实施系列连续评估。另外,NASA 还特别重视教育评估系统的信息化工作,1994 年起开始创建教育评估信息系统,从项目参与者和管理者处收集信息,为评估及后续研究

提供信息基础。

在英国,大凡政府资助的重大科普项目和活动通常会委托评估公司进行评估,如1994年起举办的全国科技周(SetWeek)差不多每年都请英国评估协会公司进行效果评估。英国公众理解科学委员会还编写了关于如何评估科技周活动的小册子(So did it work),指导科普执行机构开展项目评估工作。

从发达国家的科普项目评估机制中,我们可以得出一些有益的启示:首先,科普评估机制的建立和完善有赖于政府、科普组织、独立第三方(外部评估机构)等多种力量的共同作用,其中政府的作用是关键。即使在美国这个“政府”很小的国家,政府对评估机制的形成也起到了至关重要的作用。1993年美国国会通过的“政府绩效成果法(GPRA)”,明确要求联邦各部门和机构对工作绩效承担责任。该法案为政府科普工作评估机制的建立和健全提供了强有力的法律保障,也促进了整个社会评估文化的形成。在该法案的要求下,各履行科普职责的联邦部门和机构都制定了详细、具体,且操作性较强的科普工作监督和评估细则,对科普项目制定了严格的评估要求,比如NSF要求科普项目执行机构要预留评估经费,将项目经费的5%~10%用于评估。对此,科普项目执行机构也予以积极认真的响应,严格按照资助部门的要求上交项目评估报告。此外,美英等国家评估业较为发达,有大量的独立的专业评估机构存在,很多科普项目的评估都是项目执行机构委托外部的专业评估机构完成的。外部评估机构作为独立第三方的广泛参与,保证了科普评估的相对公正性、客观性与科学性,成为政府有关部门开展科普工作评估的有力支撑。

然而应指出的是,虽然美英等国家开展科普评估已有多年,但有关科普评估的研究却还很少,科普评估研究落后于实践。为此,美国国家科学基金会等政府机构近年来大力支持对科普评估方法、指标的研究工作,以期科普评估做得更为科学和规范。

第二章 科普评估的结构框架 与评估模式

第一节 科普评估的结构框架

一、中国科普发展的现状、问题及产生问题的原因

从 1978 年到现在,中国科普工作取得了很大成绩。根据政府部门的统计,截至 2002 年,全国共建成各类科普场馆 425 座。每年出版的科普读物达七八千种。全国科技类报纸有 200 多份,科普杂志 300 多种。大部分省级以上电视台都设立了科教类频道,生产了一批群众喜闻乐见的科教影视作品。全国共建成科普信息服务类网站 100 多个。科技周等丰富多彩的群众性科普活动在全国各地区、各行业蓬勃开展。公民的科学素质有了明显提高。

政策法规是促成中国科普事业发展的最为重要的环境条件。近 10 年来,中国科普事业的政策法规环境获得了突破性改观。1994 年 12 月,中国政府发布了《关于加强科学技术普及工作的若干意见》,此后又相继制定了《2000—2005 年科学技术普及工作纲要》、《2001—2005 年中国青少年科学技术普及活动指导纲要》等科普政策文件。特别是 2002 年 6 月《中华人民共和国科学技术普及法》(以下简称《科普法》)的颁布实施,标志着中国科普事业进入了一个法制化发展的新阶段,这在世界上是绝无仅有的,显示了中

国政府对科普事业的高度重视。与此同时,政府依据《科普法》出台了鼓励科普事业发展的多项税收优惠政策。所有这些举措对中国科普事业的发展无疑起到重要的推动作用。

可以说,中国科普事业的发展已经迎来了前所未有的机遇和良好的发展环境,然而,中国科普事业目前还存在许多问题,这些问题有的是旧有的计划经济体制下沉淀下来的,有的是中国现有的经济社会发展水平所造成的,有的则是现行工作本身的问题。这些问题对中国科普事业的未来发展造成很大劣势,主要表现在:

1. 资金缺乏。虽然近年来中国政府用于科普事业发展的资金有了较大幅度的增加,但与科普事业发展的需求相比,仍然是杯水车薪。绝大部分的科普组织表示当前面临的问题仍是经费短缺。一些组织甚至表示由于经费太少,组织已到了难以为继的地步。虽然经费问题在发达国家的科普组织中也普遍存在,但从中国的情况看,经费问题似乎更为突出、更为紧迫。而且,与发达国家相比,中国科普组织的资金来源过于单一(基本上来自政府),即使科技馆这样的实体性科普组织也普遍缺乏市场营销和社会捐赠,严重制约着科普工作的正常开展。

2. 人才短缺。由于科普组织缺乏资金,员工物质福利待遇和社会地位低,致使许多优秀人才不愿从事科普工作,科普组织难以吸引或留住优秀的人才,甚至连基本够格的员工都不能维系,合格的兼职工作人员和志愿者也难以招募。

3. 组织能力长期欠发达。伴随缺钱、缺人而来的问题是,中国科普组织的能力长期以来欠发达,即缺乏有效管理已有资源的能力和发展的潜能。具体表现在组织没有一个合理的治理结构,不能有效调动员工的积极性,组织的公共关系等社会资本不足,组织缺乏持续发展的能力等等。

4. 效率低下、工作质量不高。一些科普组织不仅缺钱、缺人,组织能力不足,而且没有任何变革的迹象和发展的目标。一些科

普组织做一天和尚撞一天钟,不求任何发展,甚至只是徒有虚名,不开展任何实质意义的活动。所开展的大部分科普活动形式和手段陈旧,缺乏创新,科普内容偏重科技知识的传承,对科学方法、科学思想、科学精神的传播力度不大。

5. 违背科普使命。当前中国科普组织存在的另一个突出问题是,一些科普组织只是打着科普机构的牌子,却大肆从事营利性的非科普活动。例如,根据科技部与中国科协所做的调查,2000年我国共有29座省级科技馆,这些科技馆的展厅面积平均只占其建筑面积的不足25%,其余面积都没有用于展教活动,有的常年举办展销会等经营性活动。

6. 科普社会化程度还很不充分。中国科普组织在开发社会资源上明显不如发达国家,在科普教育与学校教育的整合上也鲜见有切实有效的行动。

造成以上问题的原因很多,主要有:

第一 科普工作的宏观管理体制和运行机制相对落后。我国的科普工作管理目前还沿袭着计划经济体制下的一些做法。政府对科普的支持主要表现为财政拨款和补贴,而不是以项目择优资助为导向的经费支持,其结果是科普组织缺乏追求卓越的竞争压力。与此同时,尽管科普主管部门努力建立全社会参与的科普运行机制,但这种机制的建立需要实质性的举措,比如在鼓励和支持研究人员从事科普方面我们还没有制定像美英国家那样的稳定资助机制和硬性的规定要求(据说已在考虑中)。2002年我国出台的鼓励科普事业发展的税收优惠政策应当说是一项重大的措施,但要得到切实的实施还需要制订与这些优惠政策相配套的认定标准和具体实施办法。可以说,未来我国能否促进科普事业有大的发展,在很大程度上取决于我们能不能尽快建立一套与科普事业发展需要相适应的管理体制和运行模式。

第二 科普组织缺乏先进的治理结构和提高效率的竞争机制。

由于我国的科普组织基本上作为社会公益性事业单位的形式存在的,故大都缺乏提高效率的市场竞争机制。社会公益性事业单位与企业的差别之一,就在于企业通过市场配置资源,优胜劣汰;而公益性事业单位常常具有垄断性,往往导致竞争机制明显不足,往往缺乏责任机制和发展的动力。中国政府已在关注建立非营利性科普机构和科普场馆的现代管理体制和运行机制问题。引入非营利机构模式的原因在于,一方面非营利机构较之事业单位更关注资金的使用效果,另一方面它的资金运转赢余不会落入个人口袋,而是用于机构的再发展。

第三,政府尚未建立科普评估制度。评估是政府对科普组织和项目实施绩效管理的必要手段。一直以来,我国的科普工作着重做什么事,开展什么活动,却忽视了一个很重要的问题,即行动的效果或结果是什么,也就是做得怎么样的问题。对政府来说,没有评估,就谈不上管理。评估还有助于促进科普组织努力向上,使有限的科普经费发挥最佳的效用,从而有利于中国科普事业的健康发展。

因此,要解决中国科普面临的问题,除了革新国家的科普管理和运行机制,改革科普组织的管理体制,当务之急是要在科普领域尽快建立科普评估制度。

二、科普评估的结构框架

(一)构建科普评估框架的指导思想

第一,本书构建科普评估框架的特点之一在于,该评估框架的建立是以问题为取向。科普评估的重要功能在于诊断科普工作中所发生的问题、改进工作缺失和指引未来的决策或行动,因此,科普评估框架应以了解问题为取向,即评估框架的建立要以解决中国科普存在的现实问题为出发点,尽可能通过该评估框架解决科

普工作的主要问题。比如我国现今的科普工作还缺乏战略规划、科普项目多有低水平重复、科普专业机构普遍缺乏资金和人才、组织能力长期欠发达、效率低下、缺乏创新的动力和活力等等。当然,我们必须清醒地认识到,评估只是解决问题的一种手段,它不可能解决所有的问题。

第二,系统性与灵活性的思想。即本科普评估框架的构建不是机械地拼凑,而是要尽可能具有合理的逻辑解释,同时要有充分的灵活性。一方面,各子模块能够合成一个科学的系统,另一方面,各子模块之间又是相对独立的,具有不同的功能。当评估经费、时间充足时可以进行系统的评估,当评估经费不充足、时间有限时可以进行个别模块的评估。

第三,全面性与可行性的思想。本评估框架一方面要尽可能解决中国科普事业发展过程中存在的问题,同时又必须清醒地认识到,评估只是解决问题的一种手段,它不可能解决所有的问题。

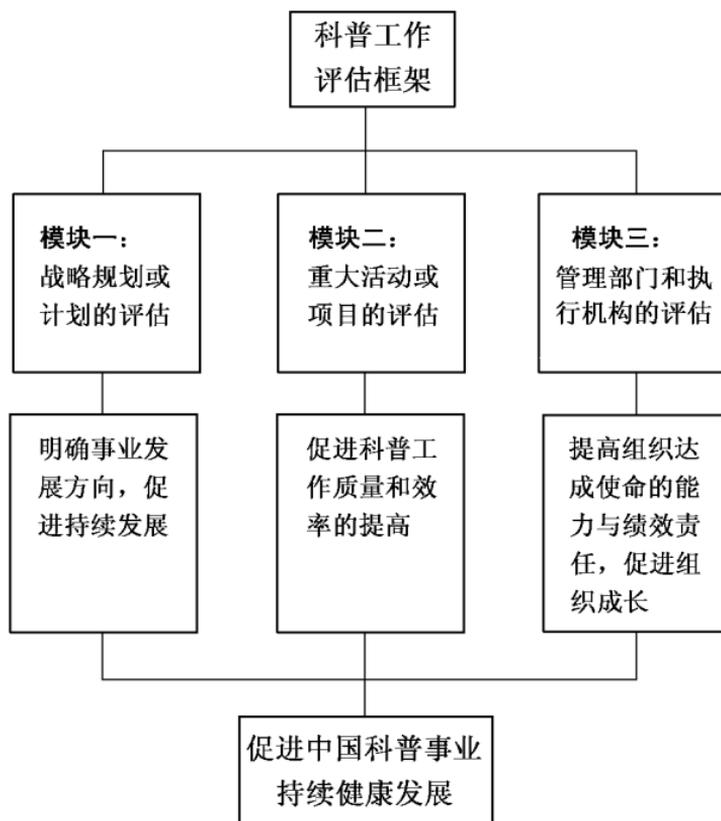
第四,评估框架的建立要有利于中国科普事业的健康、全面发展。

(二)科普评估的结构框架

中国科普事业目前存在的主要问题是缺乏战略规划、科普项目多有低水平重复、科普专业机构普遍缺乏资金和人才、组织能力长期欠发达、效率低下、缺乏创新的动力和活力等等。根据我国科普工作的现状,我们将科普评估框架设计成3个子模块:战略规划或计划的评估、重大活动或项目的评估、组织/管理能力的评估(见下图)。

第一个子模块是战略规划或计划的评估。所谓战略规划或计划可以是国家层次的,也可以是地方层次的,甚至是科普组织(如科技馆)层次的。该模块的目的在于通过评估明确各级科普工作的发展方向和发展战略,它解决的主要问题是未来一段时期内应

做什么(战略目标)以及怎样做(如何达成这些目标)的问题,以避免科普工作的盲目性和短期性,加强战略规划及政策的执行环境,加强战略规划与资源分配管理之间的联系,保证科普事业战略目标的实现。



第二个子模块是重大科普活动或项目的评估。该模块的目的在于通过评估促进日常科普工作的效率与效果,它试图解决的是中国科普机构由于缺乏竞争、缺乏绩效责任机制导致的效率效果低下的问题。近年来,中国政府加大了对科普事业的支持力度,科普机构开展的项目和活动因此越来越多。然而,一些科普机构在开展这些项目和活动时,缺乏现代项目管理的方法和技能,没有创新意识,项目实施的效果并不理想,有的甚至效率低下。其中一个

主要的原因在于缺乏项目评估。

第三个子模块是科普管理/执行机构的评估。该模块的目的在于通过对科普机构能力与绩效等的评估促进科普机构达成使命的能力的增强。它试图解决的是长期以来中国科普机构的能力欠发达的问题。

需要注意的是,以上3个模块解决的并不是一个层面的东西。有的重在治本,有的重在治标。这样构造的原因在于,评估只是解决问题的一种手段,有些东西并不是通过评估就能够解决。因此,在构建评估框架时,首先考虑的是通过某个子评估模块解决原因层面的东西,在评估无法解决原因层面的东西时,再考虑通过其他的子评估模块解决问题层面的东西。

总的说来,该评估框架兼顾了评估的系统性和相对独立性。一方面,各子模块能够合成一个科学的系统,每一个子模块都试图从某一角度解决我国科普某一层面的问题,另一方面,各子模块之间又是相对独立的,具有不同的功能。当评估经费、时间充足时可以进行系统的评估,当评估经费不充足、时间有限时,可以进行个别模块的评估。不过,对于中央及有条件的省市的科普管理部门,我们建议3个模块的评估尽可能都做。这是因为,战略规划或计划的评估虽然可以明确科普事业的发展方向,但不能促进科普组织的绩效;项目或活动的评估虽然可以促进科普组织效率的提高,但不能保证组织的发展方向、提高组织完成使命的能力;而组织能力评估虽然可以提高组织达成使命的能力,但不能保证组织正确的发展方向。惟有进行3方面的全方位评估,才能最终保证我国科普事业的持续健康发展。

第二节 科普评估的基本模式

一、科普评估可以借鉴教育评估的模式*

科普作为一种非正规教育,属于教育的范畴,因此,现行的教育评估理论和模式大体适用于科普评估。当然,科普教育与学校教育是有差别的,这种差别体现在评估模式上到底应有哪些细微改变,还有待我们未来在实践的基础上再进行深入研究。但可以肯定的是,科普评估应符合基本的教育评估理论和模式。

在教育领域,大规模的评估活动早在19世纪末就在英美等国家开始了。现代教育评估若从美国泰勒的“八年研究”(1932—1940年)算起,也已有60余年的光景。经过近半个多世纪的研究,教育评估理论体系从简单到逐步完善,评价模式从单一到多样,发生了一系列嬗变。美国评估专家古巴和林肯综合教育评估理论的发展,把教育评估划分成四个历史发展阶段:以“测量”为特征的第一代评估,以“描述”为特征的第二代评估,以“判断”为特征的第三代评估和以“建构”为特征的第四代评估。

第一代称为“测验和测量时期”,时间在19世纪末~20世纪30年代,其标志是“测量”理论的形成以及测验技术的大量实际运用。第一代评估的基本特点是:认为评估就是测量,评估者的工作就是测量技术员的工作——选择测量工具、组织测量、提供测量数据。因此,这一时期也被称为“测验”和“测量”时期。后来,随着教育测验运动的不断发展,人们认识到仅着眼于客观数字的测定并不能测定教育的全面效果,从而出现了对新评估理论的需要。

第二代称为“描述时期”,时间从20世纪30年代至50年代,

* 本部分取自参考文献[21]

它是随着泰勒“八年研究”而兴起的。这个时期的标志是泰勒目标评估模式的形成。泰勒认为,评估应该是一个过程,而不仅仅是一两个测验。评估过程中不仅要报告学生的成绩,更要描述教育结果与教育目标的一致程度,从而发现问题,改进课程教材和教育教学方案方法。泰勒的观点影响了整整一代人,在评估领域产生了巨大影响,形成了一个以“描述”为标志的评估时代。第二代评估的基本特点是:认为评估过程是将教育结果与预定教育目标相对照的过程,是根据预定教育目标对教育结果进行客观描述的过程;评估的关键是确定清晰的、可操作的行为目标;评估不等于“测验”,尽管“测验”可以作为评估的一部分。第二代评估和第一代评估相比,使评估走上了科学化的历程。

第三代称为“判断时期”,大约出现于1950—1970年,这一阶段主要是引入了价值判断的评估思想,“判断”是其主要标志。评估者不仅要运用测量手段去收集各种信息,还要根据一定的价值取向评判教育。在这一时期,一些评估专家对泰勒的目标评估模式提出疑问,认为,如果评估以目标为中心、依据,那么,目标的合理性又根据什么去判断?教育除了活动要达到预期的目标外,还会产生各种非预期的效应、效果,这些非预期的效果等要不要进行评估?在西方,有一种教育流派认为,教育乃是个人自我实现的过程(科普教育尤为如此),用统一的目标模式去统一个人的自由发展,去评估教育的结果,从根本上是不可以接受的。因此,他们对泰勒模式提出改良。这一时期教育评估理论空前繁荣,相继出现了各种模式,其中影响力较大的有CIPP模式和目标游离模式。第三代评估的基本特点是:把评估视为价值判断的过程,它走出了第二代评估“价值中立性”的误区,确认了价值判断是评估的本质。

所谓“第四代评估”是80年代于美国兴起的一种评估模式,也被称为建构模式。其创立者是古巴和林肯。第四代评估理论是在批判前三代理论的基础上建立起来的,是对传统评估理论的一种

反思。在他们看来以往的评估存在着严重的缺陷和问题,主要表现在:

(1)“管理主义倾向”。他们认为在传统的评估理论中,项目管理者与评估者之间的关系很少受到挑战。事实上,管理者总是通过提供资助控制着评估、决定评估的范围和任务,决定着评估的对象。这种关系在实际活动中常常造成四种不合理的后果:①管理者总是处在评估范围之外,他们的质量、过失无从察看;②管理者和评估者关系有失公平。评估者处于无权地位,尽管评估的问题、对象、收集资料的方法、如何公布结果等,通常都经管理者和评估者协商,但实际上,一旦发生分歧,最后决策权掌握在管理者的手里;③由于这种关系的存在,其他与评估有利益关系的人就更无法在评估中维护自己的利益、阐述自己的见解,也无法按评估建议,采取有效的改进活动;④这种关系还会导致一种有害的默契:管理者用各种方法保护自己不受损害,评估者则保证使用管理者认可的方法,以换取合同的兑现,保证评估的经济收入。

(2)“忽视价值的多元性”。这种多元性不仅表现在各利益集团之间,而且也深刻反映在个人的生活中,反映在人们对待许多具体的问题上。由于价值多元性的存在,人们常常要问,“这一评估是谁作的?”“为谁作的评估?”由此,受到伤害的被评估者就会采取不合作态度,“客观的”评估结果也就难以被具有其他价值标准的人接受。于是,希望通过评估来改进项目的期望也就落空了。

(3)“过分依赖科学范式”。评估者当然可以像其他社会科学工作者那样,积极采用和模仿自然科学中的实证探究方法。但是,他们就会面临这样的问题:①以“科学是纯粹价值中立”的说法逃避评估的道德职责,来寻求评估的客观性、中立性是有害无益的。因为“人类的所有活动,多少都具有价值的成分,即使是自然科学也不例外”,因此,价值问题是无法回避的。②过分依赖实证科学范式,使评估过于依赖定量数据而忽视对定性数据分析。③一直

以来,都是实证科学的范式在“强化和支持管理主义”。但是“数据资料自己不会说话”,在管理主义的支配下,这些数据资料却是评估者首先按照管理者规定的要求和范围,然后再用实证方法去提取的,这一点却常常为人们所忽视。

为了打破以往评估中错误的倾向,古巴和林肯提倡在评估中充分听取不同方面的意见,并把评估看作是一个由评估者不断协调各种价值标准间的分歧、缩短不同意见间的距离、最后形成公认的一致看法的过程,他们提出“全面参与”的观点,主张让参与评估的所有人员都有机会表达自己,所有人都是平等、合作的伙伴,把评估过程的控制特点与评估对象的伦理要求成功地协调了起来。可见,这种评估的范围是关注所有与评估有关的人的价值体系。

第四代评估的主要内容:认为教育评估是所有参与评估的人的心理建构的过程,评估描述的不应是事物真正、客观的状态,而应是参与评估者或团体关于评估对象的一种主观性的认识,评估的最终结果也不是事实发现意义上的对于事物的实态描述,而是参与评估及与评估相关利益的人或团体基于对对象的认识而整合成的一种共同的、一致的看法。

这种评估提倡的评估方法有两种:①应答性资料收集法:让评估对象与其他与评估有关的人自由表达自己的意见、所关心的东西和存在的问题,不受评估者的价值观影响;②建构主义方法:认为“现实”并不是纯“客观”的和“外在于人”的东西,它只不过是人们在与对象交互作用形成的一种心理“建构物”,这种“建构物”受认识者的社会、心理、文化因素的影响。

第四代评估的优点是:①提倡评估的民主,指出评估者和被评估者之间应形成一种和谐气氛;②提出价值差异的问题,指出评估中人们的价值标准各不相同;③提倡价值多元化,重视所有人的价值体系。其缺点在于:①它是一种不成熟的理论,只是提出了一些看法和研究思路,而没有固定的操作程序;②否认评估的客观性和

科学性。

纵观教育评估百年历史,教育评估理论体系从简单到完善,评估模式从单一到多样,每一代评估思想的提出都不是推翻重建的简单过程,而是完善、扩充的过程。而且这一过程永远不会停止。对科普评估来说,应密切关注教育评估理论的发展,并探讨这些理论对科普评估的适用性和借鉴意义。

二、对科普评估具有借鉴价值的教育评估模式*

评估模式是评估实施方法或过程的理论指引,长久以来一直是评估理论中不断努力的领域。数十年来,学者们提出了许多的评估模式来说明实际的做法,如泰勒模式、CIPP模式、目标游离模式、应答模式、鉴赏和批评模式、差距模式、全貌模式等等。其中泰勒模式和CIPP模式最为常见,应用得最多。这些模式对科普评估均有借鉴价值。

(一)泰勒模式

泰勒模式,又称为目标模式,是由泰勒(W. R. Tyler)等人在“八年研究”中提出来的。泰勒也因此被称为“教育评估之父”。泰勒模式是以目标为导向的,它认为评估要围绕教育计划和方案中所确定的目标开展工作,评估最终要看学习者行为的结果是否达到目标所规定的要求。这种模式的评估范围是计划和方案的目标,评估者的作用就是描述目标的实现程度。

其主要内容是:教育就是使人的行为方式发生变化和改进的过程,各种行为方式的变化就是教育目标,教育计划和方案目标的确定应该建立在分析学习者的需要、社会生活的需求、学科的发展、运用学习心理学和教育哲学等知识的基础之上。只有这样,教

* 本部分取自参考文献[22]

育目标才能反映一定时期人们的价值取向。而这种确定的目标又可以通过对学生的特殊成就来表示,教育评估就是通过对学生的特殊成就的考察来衡量教育目标实现程度的过程。

该模式具体操作程序是:①根据社会实际需要,拟定一般或具体目标;②把拟定的目标加以分解;③用行为术语界定分解的目标;④给出达到目标的要求;⑤确定满足客观性、有效性、可靠性三个要求的测量方法;⑥运用这些方法检查行为变化,收集行为表现的资料;⑦拿资料与行为目标进行比较,并根据检查结果,做出判断,说明做出判断的原因;⑧根据判断,对教育方案和计划中的目标进行修改,重复评估的过程。这种模式提倡的一般方法是定量处理法。因为,这种模式主张评估首先要把教育目标分解成许多可测量的行为目标,然后把学生的行为表现加以量化,通过比较来衡量学生的行为达到目标的程度。

泰勒模式的主要优点是:①强调以目标为导向,有利于检查目标的达成情况;②按照这种模式进行评估,结构紧密,可操作性强,容易理解和实施。其缺点在于:①没有对目标自身进行评估;②注重对预期性效果的评估,而忽视了对非预期性效果的评估;③重视结果评估,而忽视形成性评估;④评估标准来源于统一的目标,因此容易扼杀科普项目计划的创造性;⑤过于强调定量处理的方法,而忽视定性方法的应用。

(二)CIPP 模式

CIPP 模式是由斯塔弗尔比姆(L. D. Stufflebeam)1966 年提出的。他认为,完整的评估需要包括背景(Context)、输入(Input)、过程(Process)、结果(Produce)四种评估,以提供完整信息,改良方案。也就是说,该模式是一种全程评估,而且评估过程兼具逻辑顺序和循环特征。斯塔弗尔比姆认为,评估是一种有序性活动,首先应该确定计划和方案中所确定的目标是否合理与正确,然后再看

是否达到或在何种程度上达到目标所规定的要求。除此之外,评估还应该关注教育的过程,进而为计划方案的改进提供服务。对应于此,这种模式认为评估的范围应包括教育目标本身、教育设计、教育实施以及实施教育所产生的影响等方面。

该模式主要内容是:认为评估既要关注教育的过程以指导方案的实施和调整修改,又要用目标达到的程度来判别教育效果。要达到此目的,就应该对教育目标本身、教育计划、教育实施、教育影响分别进行背景评估、输入评估、过程评估和结果评估。

其基本程序是:①根据社会发展需要和评估对象等对教育目标本身进行价值判断,即背景评估(Context);②对教育方案、计划的可行性和合法性以及道德性进行评估,也就是对实现目标所需要的条件和可能获得的条件的评估,即输入评估(Input);③通过系统地搜集、整理、分析和综合大量的反馈信息资料,拿方案的实施过程和预定过程相比较,来探索教育方案和计划实施过程中的潜在问题,并寻求解决办法,以修改方案,即过程评估;④通过对方案实施结果的评估,取得大量的信息,并以此为依据来衡量完成目标的情况,即结果评估。这种模式认为评估除了要对学习者的行为表现进行量化处理之外,还要对评估的背景及相关情况进行描述和说明,因此,定量处理和定性描述两种方法的综合运用是这种模式评估方法的特点。

这种模式的优点是:①把目标引进评估领域,重视对目标本身的评估,弥补了泰勒模式的不足,使确立的目标更加符合社会发展的需要,切合实际;②强调评估贯穿活动的全部过程,对教育活动方案的实施条件和过程进行评估,及时反馈,为方案的修改提供依据,使目标能顺利达到。其缺点在于:①CIPP模式作为一种全程评估,其评估实施步骤复杂,耗费人力、物力和财力较大。因此,除非是要进行相当大的科普工作计划修正,否则不适合采用CIPP评估模式。其实,斯塔弗尔比姆本人并未坚持任何评估均需进行

全程评估,而主张评估者依情况采用适宜的评估种类。②CIPP模式的实施过程把评估者看作是决策者,这与实际评估活动中情况不符。

(三)目标游离模式(Goal Free Model)

这种模式诞生于20世纪60年代,是由美国教育家和心理学家斯克里文提出来的。严格地说,目标游离模式不是一种完善的评价模式,它没有完整的评估程序。因此,有人把它仅当做一种关于评估的思想原则。斯克里文断定:“对目标的考虑和评估是一个不必要的,而且是有害的步骤。”按他的观点来看,目标评估很容易使评估人受方案制定者所确定的目的的限制。因此,他建议把评估的重点由“方案想干什么”转移到“方案实际干了什么”上来。他认为,评估委托人不应把方案的目的、目标告诉评估人,而应当让评估人全面地收集关于方案实际结果的各种信息,不管这些结果是预期的还是非预期的,也不管这些结果是积极的还是消极的,这样才能使人们对方案做出正确的判断。

该模式主要内容是:评估要从检查方案的结果来判定其价值,而不应考虑目的或目标。为了降低评估活动中方案和计划制定者的主观意图的影响,不主张把预定的活动目的告诉评估者,以使评估者能够在没有偏见的情况下自由地肯定其优点,对“预期效果”和“意外结果”不加区别,从而检验教育的全部效果。

这种模式在评估时强调对观察到的结果的一切可能原因作认真的审查,以建立有根据的联系,并且还主张对原来方案的作用以及原来方案与方案结果相矛盾的方面加以分析和研究。

这种模式的优点是:对目标的完备性加以考虑,注重对教育活动非预期性效果的评估,扩大了对方案效果的关心范围。其缺点在于:如果评估组织中各个评估者具有不同的价值观念和价值标准,且评估对象较多时,会给评估的操作带来较多的困难。

(四) 应答模式(Responsive Model)

该模式由斯塔克(R. E. Stake)在 1973 年首先提出,后来由别人进一步发展。该模式认为教育价值在表现形式上有时是发散的、潜伏的,而且许多教育现象除了表现出为特定教育目标服务外,其自身也具有内在的价值。评估应当从关心评估活动所有人关注的现实和潜在问题出发,而不是以预定的目标或假设出发。评估要适合于一个复杂的多元的客观世界的现实并且满足各种层次、持各种观点人的需要。由此,这种评估模式所关注的范围就是各种人的需要。

该模式的主要内容是:强调采用非正式观察、访谈和定性描述分析的社会学研究方法,通过评估者与教育相关人员进行接触,了解他们的愿望,从中发现并选择出人们所关注的有价值的问题,然后把它同实际活动进行比较,对教育方案和计划做出修改,对大多数人的愿望做出应答,以使教育能满足各种人的需要。

其基本程序是:认为评估活动可以通过 12 个事件进行组织。这 12 个事件是:①识别方案的范围;②了解方案中所指定的活动;③确定评估的目的和人们的兴趣所在;④形成议题和问题的概念;⑤识别所需的资料;⑥选择观察者、判断者和评估工具;⑦观察指定的前提条件,了解它们之间的关联;⑧进行理论总结、准备描述性材料、开展个案研究;⑨确认或否定某些证据;⑩提高评估的可靠性,整理出供评估听取人使用的材料;⑪收集正式报告;⑫与方案评估的委托人,评估的听取人和方案的执行人进行交谈。这些事件在开展评估时可以彼此跨越,可以同时发生,也可以多次发生。这种模式提倡一种叫做“三角研究法”(triangulation)的评估方法。这种方法的含义是指进行评估时要求评估者不仅用不同的方法去研究同一问题,而且应该从不同的角度,让不同的人去分析评估同一现象、问题或方案,最终结果不受方法局限性的影响,从

而得出事物的真实规律。

这种模式的优点是：①从根本上否定用目标的形式体现评估的价值取向，肯定了教育价值的多样性、复杂性；②评估准则反映了与评估有关的各方面人员的需要，具有一定的民主性；③强调自然条件下的观察、访谈和描述性分析，避免评估信息遗漏，评估结果效度较高。其缺点在于：①评估结果的适用范围太小，可信度较低；②在评估过程中，要耗费很大的人力、物力和时间。

(五) 鉴赏和批评模式 (Connoisseurship and Criticism Model)

鉴赏和批评模式是由艾斯纳(Eisner)于1976年提出的。他认为真正的教育要追求合作、协商，在交往中共同发展，评估应该为增进和改善实践服务。由于教育是复杂的，所以不能按以往科学的评估方式，将“质”用“量”来表示，凡是不易量化的就应抛弃，评估应该把教育当做一个有机的整体，只有这样，人们才会看到教育现象的本来面貌。由此，这种评估模式关注的范围是教育过程这一整体。

该模式的主要内容是：认为评估应包括两部分：①教育鉴赏，主要是通过评估者凭自己的经验对教育中发生的事加以理解和意识，这种理解和意识是内部的无需表达出来的，就像艺术鉴赏家凭自己的经验对艺术作品进行欣赏一样；②教育批评，主要是评估者在鉴赏的基础上，对所评估事物加以“揭示”，这种“揭示”是外部的，需要通过言语和书面的形式进行表达，以使相关人也能够了解。

其基本程序是：主要对教育批评这部分进行规定，认为进行教育批评应该：①描述：通过多种方式描述教育情境的复杂特质，以帮助别人了解当时的情境；②解释：对发生在具体情境中的行为现象做出原因分析和判断，并理解其意义和含义；③评估：在描述和

解释的基础上凭自己的经验进行整体性的评估,并得出与教育活动的特点相一致的改进性结论。这种模式提倡一种定性的评估方法。这种方法不是测量预期的教育效果,而是强调运用观察、访问等方式对整个方案进行全面而深入的研究,以做出整体的理解。

这种模式的优点是:对教育现象进行观察,在整体上了解教育现象,比较灵活,从具体事实中得出对评估对象的总体看法。其不足之处:没有固定的操作程序,更多地依靠评估者个人的自身修养和对教育的认识。

总之,上述几种评估模式各有特色与局限,所适用的情景也不尽相同。从评估模式的应用情形来看,现阶段国外教育评估实践中,往往根据具体的评估任务,选择某一种模式(不一定是最新的),而且很少采用全真的模式,大多是采用模式的精义而已,也有采用两种或以上的模式加以整合设计的趋势。比如,尽管泰勒模式有这样或那样的不足,但由于其易于理解和执行,在现今的教育评估实践中,仍被广泛采用。在科普评估中,评估模式的采纳,在目前也应着眼于使用并整合现有的简单易行的评估理论模式。

第三章 科普评估的指标框架

建立科学合理的评估指标框架是科普评估的关键。评估指标框架是根据评估目标所选定的评估对象的一组属性。评估指标框架确定后,对评估对象的评估就转化为对这些可以直观、具体衡量的属性的考察。

第一节 建立评估指标框架的原则

评估指标框架是根据所提出的目的、任务和对信息资料的实际需要,在对所评估的对象进行初步理论分析的基础上建立起来的。经过初步的理论分析,可以确定需要进行评估研究的各种现象及其总体的内容、本质、属性和特征。在明确这些问题之后,需要根据评估研究的目的、任务和所要反映的现象总体的属性特征,有选择地规定一系列的标志,作为反映这些属性特征的工具,并据以形成各种指标和指标框架。构建评估指标框架应遵循如下原则:

第一,体现导向性,能引导科普事业的发展方向;所谓导向性是指评估指标框架的设立和使用,会引导受评科普机构重视所评估的内容,并朝评估标准的方向发展。

第二,突出重点,指标力求少而精。由于我国的科普经费不足,能够投入到评估的经费会非常有限,因此在设计评估指标时,要突出重点,在考虑指标体系的全面性的同时,要考虑指标体系的

精简性。

第三,评估应侧重于效果评估,指标的设置应抓住关键。背景指标、投入指标、过程指标、产出指标、效果影响指标是反映科普工作状况的重要指标,其中背景指标与投入指标是基础,效果影响指标是衡量科普工作优劣的关键。

第四,综合运用定性定量指标。理论上讲,为了能够科学客观地反映科普工作的质量特征,应该尽量选择定量指标。但鉴于科普的特殊性,科普工作的很多方面不能用定量指标进行描述,所以不可避免地要采用一定的定性指标。

第五,动静结合。科普事业处于不断发展之中,指标的设置应既立足于现状,也着眼于未来,体现时代发展特征,可根据实际情况适时修改指标体系,用新标准来评价,以适应科普事业发展不同阶段的目标要求。

第二节 建立评估指标框架的学理基础

一、科普评估与公众科学素养测量

科普事业的总体目标是提高广大公众的科学素养,因此要实施科普评估,首先必须明确科学素养的概念和内涵,否则,所确定的科普评估工具很可能与科普活动所要达到的实际目的不一致。

关于科学素养的概念,目前有许多定义和解释。有些定义是基于对概念本身的解析和研究,有些则是通过对一个具有科学素养的人的描述来界定。然而,所有这些定义或过于笼统概括,或内涵不够宽阔,不能作为指导科普效果测量和评价的理论依据。应当认识到,尽管科普事业的总体目标是提高公众的科学素养,但公众科学素养的养成更多地来自于学校教育,而不仅仅来自于科普教育,更不是某单一科普项目的结果。同时,科学素养具有复杂多

元的特征,它应该被看作一种多侧面的现象,而不是一个可以精确定义的单一结构,其表现是形形色色的,有认知的也有非认知的层面。为了科普评估目的,我们需要进一步界定科学素养在科普评估领域的操作性定义。

科学素养应当是知识技能、情感态度、行为表现的综合体,它是个体对科学活动、科学生活的感受、体验并内外化的复杂产物。目前,包括我国在内的世界主要国家都在开展公众科学素养调查,而且大都借鉴美国国际公众科学素养促进中心主任 J. D. Miller 教授提出的科学素养三纬度模型作为成人科学素养测量的基础。Miller 教授提出的科学素养测量纬度包括:对重要科学术语和概念(即科学知识)的了解、对科学原理和方法(即科学本质)的了解以及对科技的社会影响的了解。那么,在科普效果评估中是否可以完全采用 Miller 定义的纬度进行评估呢?答案是否定的。因为 Miller 的科学素养测量模型着重于了解公众(而且只是成人公众)具备基本科学素养的情况,而非具体一项科普计划或项目所获得的效果。而且,由于 Miller 的模型是为公众科学素养调查设计的,出于操作目的,目前只侧重于测量公众对一些基本的科学知识、原理和方法的掌握以及对科学技术发展、科技组织的态度和看法,内涵还有限,不能包容丰富多彩的科普活动的多样化效能,而且没有体现层级性。

目前,美英国家在进行科普效果评估时,基本上采用科学学习的效果测量模式,借鉴布鲁姆的教育目标分类学和社会学方法,按照认知域、态度域、行为域三个纬度,采用多种数据收集方法进行效果评估(与科学素养的 KAP 模式相似)。基本的内容框架大致如下:

认知域:衡量受众通过科普活动学到了什么,包括对科学知识、过程、方法以及科技对社会的影响等的理解和掌握。

态度域:了解科普活动是否增进了受众的科学意识、科学兴趣

和科学态度。

行为域 :了解科普活动促成受众产生哪些与科学有关的行为或行为倾向。

从逻辑上讲 ,认知、态度和行为包含了科普活动的三个层级效果。认知应当是科普活动达到的第一级效果 ,在认知的基础上如果受众的情感态度发生了变化 ,科普活动就产生了第二级效果 ,如果认知和态度的变化还外化为行为 ,科普活动就产生了第三级效果。而且 ,各层次还可以根据从低级到高级的水平进一步细分 ,比如认知域可进一步分为简单的知识记忆以及知识领会运用等较高级的思维能力。因此 ,就科普评估而言 ,认知、态度和行为目标分类较 Miller 的科学素养测量三维模型更能获得对科普活动的全面效果评价。

二、科普评估中对科普效果的认识

效果的测量和评价是科普活动评估的核心。科普效果是指科普活动对受众产生的认知、态度、行为上的影响 ,它有如下特征 :

(1)滞后性。科普活动对受众的影响因个人的心智、兴趣等而异。有的受众可能反应快一些 ,有的则慢一些 ;有的可能是连贯的、继起的 ,有的则可能是间断的、迟效的。时间的滞后性使科普活动的效果很多时候不能很快、很明显地显示出来 ,绝不会像“皮下注射理论”、“魔弹理论”所宣称的那样 ,立竿见影。

(2)层次性。科普活动的效果体现在认知、态度、行为三大方面 ,而根据布卢姆教育目标分类法 ,认知、态度情感等又根据从简单到复杂的水平而细分为多个层次。比如 ,在认知方面 ,不但包括科学知识的一般性学习 ,还包括对有关科学知识的应用、分析、综合等高层次思维能力 ,因此 ,我们在考察科普活动的认知效果时 ,不能只停留在低级层面 ,即衡量科普对象通过参加科普活动学到了哪些基本知识 ,更要重视知识习得之上的高层次认知能力的获

得,比如对有关知识的应用和分析能力。美国罗克曼研究与咨询公司 为儿童科普片“比尔·奈科学人”所做的评估中,在评价孩子们通过看节目学到了什么时,不但调查孩子们对节目中科学概念知识的掌握情况,而且还了解孩子们利用这些科学知识探究和解释日常生活中其他科学现象的能力,即所谓的高层次科学思维能力。这一点对科普评估的完整性很重要。

(3)适度性。实践证明,科普对受众的影响总体上往往既不像魔弹论那样夸大,也不像有限效果论那样消极。首先,从认知、态度和行为三大方面上看,一项科普活动即使在某一方面不发生效果,也会在另外方面发生效果。其次,科普活动对受众即使在微观、直接、具体、短期层面无效果,也会在宏观、间接、无形、长期层面发生效果。

由于科普效果的上述特点,因此在进行评估时应注意以下几点:第一,由于科普效果具有层次性和适度性,不同的科普活动所侧重达到的具体目标层次又不尽相同,因此,在评估前应首先对评估对象进行具体的研究,确定从哪些方面进行评估。评估人员要把最关键和最迫切需要了解的效果问题作为测定的重点,设立正式的测定目标,选定测定问题。第二,科普活动对受众的影响是潜移默化的、长期的,因此,科普评估既要有实时效果的评判,也要尽可能包括短期甚至长期影响的量度。目前,对科普项目进行短期影响评估已很普遍,但进行长期影响评估还较为困难,不过,国外已有一些长期性的科普活动兼顾进行了短期及长期影响的评估,如英国 Beta 技术公司对英国工程委员会“女性进入科学工程领域”活动的评估就是一例。第三,要真正理解什么是“完整的学习”,什么是“真正的效果”,脱离低级的表面评价。比如,在科普工作中,我们经常看到科普执行机构报告工作成效时用这样的内容来说明,如他们印刷了多少科普宣传册,组织了多少场科普报告会,办了多少展览,来参观的有多少人等等。诸如此类的内容是说

明不了科普效果的。拿举办科普报告会来说,即使安排了大量的听众听讲,但听众对所讲的内容不感兴趣,效果自然不能说好。可见,用以上内容作为指标来衡量、检测科普效果是片面的、错误的。在科普工作中,类似这种用不恰当的指标来片面描述科普效果的错误还相当常见。在正规的科普评估中,无论如何应脱离这样的简单做法。

第三节 战略规划或项目资助计划的评估

国家科学普及发展战略规划或项目资助计划是实现国家科普工作使命与目标的一系列决策和行动计划,事关国家科普工作发展战略、未来发展方向和科普工作的总体布局。

科普战略规划或项目资助计划的评估是指评估人员利用科学的方法,系统地收集相关信息,评估规划或计划的内容、实施过程、实施结果的一系列活动,目的在于提供选择、修正、持续或终止规划或计划所需的信息。它有助于提高政府的问责性、更明智地分配资源以及更有效地管理计划。

科普战略规划或计划的评估不仅指规划或计划实施绩效的评估,也包含规划或计划实施前与实施中的评估,即规划或计划评估可分成三大类:前期评估、中期评估和结果评估。前期评估指对拟制订的战略规划或计划方案的必要性、可行性及其定位、目标、任务、投入、组织管理等进行评估,目的在于促成最佳、最适当方案的选择,以期在规划或计划方案付诸实施前能够适时地修正内容与资源条件。中期评估则发生在规划或计划的执行阶段,旨在对规划或计划执行中的进展情况及存在的问题进行评估,进而提出哪些规划内容需要调整,哪些方面需要加强,以改善规划或计划的实施品质与策略,为规划或计划的后续安排和调整提供依据。结果评估则是对规划或计划的执行结果进行评估,目的在于确认规划

或计划目标的实现程度、执行效果与影响、组织管理的有效性、实施的经验教训等,为规划和计划的滚动实施、调整或终止提供依据。结果评估所关心的并非是规划或计划是否按原定计划执行,而是规划或计划执行后是否产生了预期的结果或影响。

总括起来,科普战略规划或项目资助计划的评估应至少包括目标评估、实施绩效评估和组织管理评估三个方面。围绕这三个方面,我们确定了如下评估指标框架:

- 战略规划或计划确立的目标是否符合国家科普事业的使命和愿景,是否因应了国内外环境的变化,具有战略性和前瞻性?
- 战略规划或计划确立的目标是否全面考虑了中国科普的现状和问题,具有宏观性(战略目标的涵盖面要广)和可操作性(具备执行规划或计划的条件和能力)?
- 规划或计划执行过程中都采取了什么举措?这些举措是否切实有效?
- 规划或计划执行中是否具备高效的组织管理和协调机制?
- 实施战略规划或计划取得的实际效果(实现目标的有效性)及影响(预期的或非预期的),社会有关方面的满足程度。
- 战略规划或计划中哪些目标未能实现,是何原因,应采取什么改进措施?是否需对战略规划或计划进行调整?

在美英等国家,政府为长期有效地推进科普工作,都实施了科普项目资助计划,如美国国家科学基金会的非正规科学教育计划、美国航空航天局空间科学办公室的教育/公众外延计划。这类计划的评估与战略规划的评估类同,通常包括计划预期目标的评估、预期目标达成评估以及计划的管理与运作机制的评估三个方面。而在实际评估工作中,尤其侧重于后两个方面,即预期目标达成评估和计划的管理与运作机制的评估。关于政府科普项目资助计划

的案例可参见本书附件一和附件二。

第四节 科普项目的评估

科普项目类型多样、形式繁杂,对各式各样的科普项目,我们不可能建立起一个普适、通用的科普项目评估指标框架。一种可行的方式是按照科普项目的类别,给出各类科普项目的参考性评估指标。这里,我们将科普项目分为3大类:展览项目、媒体项目以及城乡社区和青少年科普活动。围绕这3种类型,分别阐述各自评估中应考虑指标方面。

按照本书第二章确立的科普评估结构框架,这一小节原本是要论述“重大科普活动或项目的评估”的,但考虑到像科技周这样的大型科普活动,由于包含了大量的各式各样的科普项目或活动,其评估属项目集群评估,情景甚为复杂,国外对科技周这样的大型科普活动,迄今也未看到值得信服的评估模式,故本小节只探讨单一科普项目的评估,对集合多形式多项目的重大科普活动的评估暂不涉猎。

一、科技展览项目的评估

科技展览评估中所使用的词语,并不是博物馆固有的,原本是教育评估使用的词语。在现在所提倡的四种展览评估中,总结性评估开展得最早。在美国教育界,1965年通过的初等及中等教育法案中规定,受联邦政府资助的教育项目有义务进行评估。受此影响,博物馆展览项目资助机构也规定,受资助的展览必须进行评估。在申请展览项目资助时,博物馆要清楚说明展览的目标观众是谁,打算达到什么样的目标。此外,还要提出测定展览目标实现程度的评估计划。

迄今为止,欧美国家采用的科技展览评估方法还是20世纪

60年代后确立的,但近年来出现了一些新的研究,因此,展览评估研究者之间的方法和用语出现了差异。这里,我们介绍的是根据展览开发阶段划分的四种展览评估,并且我们按照四种展览评估出现的先后顺序,依次介绍这四种展览评估方法。展览评估最初主要是为了测定展览展出后的效果,即我们现在所说的总结性评估。之后,为了开发更有效的展览,评估逐渐渗透到展览的开发过程中,出现了形成性评估和预评估。

(一)各种类型的展览评估

1. 总结性评估

总结性评估是指展览制作出来并在展馆展示后所进行的评估,它有助于我们判定展览是否达到了预期的目标,了解观众通过参观学到了什么,获得了怎样的体验,其目的是衡量展览的展示效果与推广应用价值。总结性评估的调查通常包括:观察观众在展厅的活动(有时也称之为“跟踪”),对观众进行参观前后的问卷调查以及对工作人员和观众进行访谈。评估主要是评判展览质量和展览效果与影响两个方面,重在评判展览效果与影响。下面的指标框架只能作为一个笼统的参照,在实际评估中要注意根据展品类型、评估目的的不同予以取舍。比如,对非交互性的展品,“参与性”就可能不作为一个评价指标。

为了测定展览的展示效果以及调查展览是如何被利用的,参观人数当然是重要的指标之一,但还要通过观察法、调查问卷法和面对面访谈法等,对观众在展馆内怎样行动以及由此获得什么感触进行调查。与后面讲到的其他种类的评估相比,总结性评估是总括性最强的,需要对较多的观众进行周密的调查。其调查结果成为判断今后是否继续举办这样的展览,或者是否对展览进行必要改进的依据。

在进行展览评估时,要掌握观众的活动情况比较简单。通过

跟踪或定点观察等,既可以调查观众的行动路线和使用展览的时间,又可以明确展览中哪些部分没有被观众利用,或者观众利用的方式与博物馆原来的意图有什么不同。与此相比,测定观众观看展览的学习收获却是十分困难的。我们认为可以根据观众的行动或使用展览的方法等看得见的行动,推测观众学习的情况。不可否认,观众的行动与学习之间有一定的联系。但说到底这是推测,观众头脑中想的事情仅根据其行动显然难以了解清楚。因此,测定观众的学习收获,需要对观众进行访谈或问卷调查。最一般的方法是,在观众参观完展览之后,询问其对展览的第一印象。根据观众是否会提及展览所要传达的内容或主题,测定是否达到了展览效果。通过对比观众参观前后的情况,测定展览对观众产生的影响。与单纯在出口处调查观众“是否感受到了展览的意图”相比,比较展览入口和出口处的结果的方法,重点在于了解观众参观前后的“变化”。要重视“做不到的事情却做到了”的展览效果。但是,我们并不认为展览效果只是“带来新的认识”,使观众再次确认和回想起已知道的事情也是博物馆学习的一个重要方面。通过展览,观众会发生怎样的“变化”,可能会明确地显示出展览对观众的影响。然而,只将此作为展览的效果进行评估,似乎并不符合博物馆的情况。

如要进一步说一下对观众访谈的情况,就是上述两种方法都只测定了展览留下的短期影响。至于长期影响,我们认为必须事先留意测试不到的东西。在考虑博物馆展览对观众产生的影响时,并不仅限于短期的影响,既包括观众从博物馆回来后或者经过几年后再回忆起来的感受,又包括博物馆经历对观众今后的生涯产生的影响。在进行展览评估时,必须认识到展览对观众产生的影响既有短期的,也有长期的。

展览总结性评估指标框架

一、展览展示质量

1. 展览内容方面 科学性、教育性、趣味性与参与性(主要针对含交互性展品的展览)、易理解性和易学习性(比如对复杂的展览主题,是否采取了必要的阐释说明手段)。
2. 展览展品方面 展品的安全性、展品的可管理性(如是否耐用和易于维修及布展)、展品的可获得性(观看或操作一项展品是否需要排队,等候的时间会不会太长)。
3. 展览环境方面 展览环境是否舒适宜人、是否人多时显得拥挤,休息座位的提供、对展厅工作人员服务的满意度(是否安排了足够的场馆工作人员,工作人员的服务态度和水平)。

二、展览展示的效果与影响

1. 有多少观众参观了展览?
2. 观众都参观了哪些展品,驻留时间有多长?
3. 通过参观,观众有没有学习收获?有哪些收获,对科技的认识和态度是否发生了变化?展览对观众会产生什么长期影响?

2. 形成性评估(展览制作过程中的评估)

形成性评估是指在展览的开发和设计过程中所做的评估,其目的是发现展览的不足之处和可改进之处,以对其开发和设计过程进行改进或修改,完善和提高展品的价值。1976年,Screven在展览开发中引进了形成性评估的方法,并与总结性评估进行了比较和区别。他强调,展览评估不仅要测定展览展示后的效果,还要从展览的开发阶段开始,一边了解观众对展览方案的反馈一边制作展览,从而做出更有效的展览。总结性评估是在展览展示后进行的,而形成性评估是在展览的制作过程中进行的,是采用廉价材料制成的展览样品来调查观众的反应。

对展览策划人员来说,形成性评估是最为重要的评估,因为它吸收观众参加到展览开发过程之中,以观众的眼光判断展览需做哪些改进和完善。形成性评估可以反复多次,由参加测试的观众对展览组件的实物模型或仿造品进行试用,然后展览人员通过问卷调查以及非正式的、开放式访谈,获得他们对展览的意见和看法。借助形成性评估,策划人员可以清楚地决定如何调整和改进各展品,使其发挥预期的作用。形成性评估有助于策划人员判定展览的构思是否清晰(观众是否理解了?)、观众的动机和行为(观众是否想观看这一展品并清楚要做什么?),以及如果是交互性展品的话,展品的机械结构是否可行(是否按设想的方式工作?)。

形成性评估可以很简单,如询问观众某项展品是做什么的,也可以做得很复杂,如制作与实际交互性展品等同尺寸的展品原型,让观众试用。在形成性评估中,策划人员可以让观众以多种方式参与展品开发过程。譬如当策划人员对观众是否理解某一内容意见不一致时,观众可以帮助解决争议。观众有时甚至可以帮忙撰写展示说明。例如,在策划“被污黑的水域:油船泄漏概览”巡展时,展览开发人员对如何写好生物除污技术的说明一筹莫展,写了几稿都不满意。最后,一位观众在与一评估人员讨论时,提出了他的构思,其说明文句读起来朴实易懂,解决了困扰开发人员的一个问题。

特别是对于观众体验型的展览,由于形成性评估在实际调查观众如何使用展览的同时制作展览,这样在展览制成后,可以避免使用方法不清楚或使用不方便等失败情况。在解说词或图解等大致方案出来后,用纸等廉价材料制好,再向观众进行介绍,观察观众阅读时是否感到吃力或用语是否恰当,以决定最终的方案。总结性评估是对很多观众进行严密调查,而形成性评估只需要非常小规模的患者调查。例如,让10个观众朗读解说词的草稿,如果10个人都认为某个词语不易理解,就需考虑将这一词语更换为其

他的词语。在策展的起始阶段,重要的是小规模调查的积累。在展览方案进入最后阶段、接近最终方案时,只要增加调查的观众数量就可以了。在形成性评估中,一般采用观察法和面对面访谈法,先观察观众利用展览样品时的情形,在其利用后再进行访谈。作为调查的项目,主要有如下五项:

(1)吸引力。如果是以展览为媒介传达信息,首要的是,这一展览能吸引观众的注意。观众不看这一展览,就传递不了信息。所以,必须调查展览是否会吸引观众的注意力。

(2)保持力。虽说展览吸引了观众的注意,但是观众并非一定会仔细地参观展览。例如,在科学馆的展览中,有时会看到孩子们一个接一个摁展览装置的开关,这只能说明展览具有“吸引力”但不具有“保持力”。

(3)说明力。特别是在体验型的展览中,重要的是弄清楚观众会进行何种操作,这种方法是否是观众易理解的。虽然观众会接触其感兴趣的展览装置,但有时也会碰到如何打开开关、开关在什么地方等不知道如何操作展览装置的情况。传递信息的重要之处在于,是否向观众恰当地说明了展览装置的使用方法。

(4)传递力。展览都有预想传递的信息。传递力是调查展览向观众多大程度上传递了信息。

(5)感染力。感染力是调查展览是否有意思、观众是否喜欢这一展览或者观众在利用展览装置后,在情感上是否受到什么影响等。

3. 预评估(事前评估)

预评估是指在策展的最初阶段进行的评估。1986年,Screven将其作为展览评估的一部分使之系统化(有关预评估的详细内容,参见 Steven A. Griggs 所著“Front-end Evaluation: Three Case Studies”,British Museum,Department of Public Services,Office of Museum Programs,1982)。据说,在进行预评估的背景下,存在行

动科学或认知科学领域的影响。也就是说,展览也好,教育项目也好,都是学习者在通过某种经历进行学习时,发现新的经历与本身具有的知识或过去的经历之间的切合点。来参观展览的观众并不是等待灌输新信息的空器皿,而是根据自己的先验认识(知识、相关体验或错误认识等)去看展览的。如果观众是从其与已有认识的联系中开始学习的,那么策展人员应该尽量采用观众易进行联系的方式制作展览。因此,在开发一项展览时,应预先调查观众对展览主题以及展览方案的反应,为展览制作提供有益的信息。

通过预评估,我们能了解观众对一项拟议中的展览主题有的兴趣与感觉,知道什么,不知道什么,存在什么误解,对展览有何期待等等,展览策划人员由此能够更有效地设计展览。美国纽约科学馆的 Samuel Taylor 曾指导进行了一次大型微生物学展览的预评估。Taylor 发现,受调查的 5~11 岁儿童中,大多数的儿童不知道微生物究竟为何物,把它们说成是“快速运动的东西”、“电子之类的东西”、“能发声的东西”等等。受调查的成年人的回答虽然总体上好些,但也发现了某些错误。Taylor 由此建议,展览的构思(concept development)既要体现博物馆管理者的思想,又要兼顾普通公众的问题。“博物馆管理者可能知道什么重要,什么有意义,但我们必须从人们了解的、感兴趣的内容开始,找到展览阐释的切入点。”

预评估还有助于纠正展览策划人员的某些错误认识。例如,马里兰科学中心曾经计划搞一个“科学是什么”巡展。策划人员起初认为,普通公众对科学的看法深受好莱坞电影的影响:痴迷于研究的科学家穿着白大褂,在关着门的屋子里做些不可思议的实验。举办这一展览的目的就是要消除人们对科学的潜在的不良印象和误解。然而,在巴尔的摩进行的预评估访谈中,展览策划人员发现,人们对科学的看法要积极得多。这一结果促使策划人员决定彻底改变这项展览的中心主题。该展览最后更名为“实验!”集中

展示科学家用以解决问题的各种方法。

在展览的企划阶段,观众的态度也是要调查的内容之一,与观众的知识同等重要。有的展览主题可能令观众感到反感,这会影响到观众对展览的利用,所以,在展览企划初期调查观众对展览的态度,可以明确在展览时应该考虑些什么。而且,对观众态度进行调查,并不只是为了避免观众对主题产生厌恶感,也是为了架起展览内容与观众之间的桥梁,让观众积极地利用展览。

这样,在预评估中,对观众的访谈成为主要的调查手段,访谈大多是个别进行的,有时也进行焦点小组访谈。所谓焦点小组访谈,是指将若干相同背景的人聚集成一个小组(大约10名左右),然后对这一小组进行访谈,同时开展小组讨论。当希望了解某一特定人群的想法时,通常使用这种方法。只要根据设想的不同的观众群和展览内容,决定组成什么样的焦点访谈小组就可以了。例如儿童小组、教师小组等等。

三种展览评估(预评估、形成性评估和总结性评估)中,预评估最难以获取有用的信息。因为在展览开发的起始阶段,各种想法往往是模糊的,远未成形,这种情境下要确定恰当的评估内容就显得有些困难。策划人员最好事先多做些准备工作,弄清楚手头可以搞到哪些现成的信息。在一些发达国家,各种名目的民意测验和营销调研层出不穷,博物馆有时利用它们的调研结果。例如,奥克兰举办“加州的水资源”展览之前,调研人员首先查阅了由当地水管理机构和自来水公司进行的众多的公众调查,之后才拟定自己的预评估题目,调查的问题因有其他机构已有的调查作基础而减少了不少。

4. 修正性评估(Remedial Evaluation)

与前面介绍的总结性评估相同,修正性评估也是在展览制成后进行的评估。总结性评估并不是直接以修正展览为目的,而修正性评估的目的则是修正当前的展览。有的研究人员会把修正性

评估包含在总结性评估中考虑,有的研究人员则为了修正,通过展览模型进行试验,将修正性评估作为形成性评估来处理。目前修正性评估还不是固定的名称,但我们认为其目的是在展览方案确定后进行修正,与上述三种评估方式不同,所以单独作为一项进行说明。修正(Remedial)这个词语是从1990年开始使用的。修正性评估是在展览制成后先进行评估,明确问题点后对其进行修正,再利用模型进行试验,然后根据观众的反应来决定修正方案。在利用模型这一点上,修正性评估与形成性评估相似,但若说两者的不同之处,那就是修正性评估可以在实际的展览空间中进行试验。而形成性评估虽说制作了模型,但是不可能提供完整的展览空间,需要将解说词、体验型展览装置等展览要素分部分进行试验。虽说是重复试行制作出各种要素,但很多时候是不知道这些展览要素在展馆空间内将发挥怎样的作用。对完成后的展览进行修正需要很高的费用,而且非常不容易。因此,要对制成后的展览进行修正性评估,就必须事先做好这部分的预算。

(二)与展览评估相关的观众调查研究

这里,我们大致看一下与上述四种展览评估相关的观众调查研究。就实施展览评估和在展览开发中灵活运用展览评估结果而言,无论是观众调查还是观众研究都是不可或缺的。

1. 观众调查

每天会有各种各样的人来参观博物馆,而且都有各自的需求。所谓观众调查,是要了解哪些人来博物馆,哪些人没有来,以便了解博物馆的观众层。如果要对常设展览进行总结性评估,所调查的观众必须是能代表博物馆观众层的观众。观众调查是决定展览评估应调查哪些观众的基础,这对于展览评估是必不可少的。如果掌握了观众层,在组成焦点访谈小组时,就很容易判断应该听谁的意见。而且,如果弄清楚不来博物馆的人的特点,就可以将这些

人组成焦点访谈小组来听取他们的意见,以他们为对象策划展览项目,以扩大博物馆的观众范围。总之,观众调查既是博物馆展览评估的基础,又会启发博物馆未来的发展。

观众调查通常是大规模的问卷调查,但有时也会以个别访谈或焦点小组访谈的方式,了解人们对博物馆的期望等。观众的倾向是随着时间而变化的。所以,观众调查必须持续下去,以掌握最新的情况。

2. 观众研究

如果评估结果明确了不受欢迎或未被使用的展览,展览开发者就必须考虑应如何改善这一状况,这时就需要创造性。目前,对展览评估的关注,大多集中在研究如何办好眼前的展览,而人们对展览的利用,则追求更一般化的原理。一般认为,观众研究会对办好展览给予一些启示。例如,观众研究可以提供如下信息:一家人为什么会来博物馆?在馆内有哪些行动?在何种情况下孩子们通过展览学到了知识?观众研究会用与展览评估相似的方法收集信息,但会在更大的观众范围内进行更严密的调查。评估是针对某一特定的展览进行的,与此相比,研究是在更大的范围内引出可以应用的原则。

为了进行恰当的展览评估,为了在改善方案中灵活利用展览评估的结果,评估人员以及与展览评估相关的人员都必须充分了解观众研究的结果。评估与研究可以说是制作好的展览不可或缺的两个方面。

(三)博物馆开展展览评估的建议

博物馆进行展览评估,可以促使博物馆从观众的角度策划开发展览,这对展览的成功至关重要。然而,在现实中,很多博物馆受制于展览制作进度以及预算、人员的限制,在展览开发过程中并不进行评估,视评估为赘物。怎样促使这些博物馆把评估纳入他

们的展览开发过程？除了资助机构的要求外，可能的一个解决方法就是鼓励博物馆先从简单的评估做起。

美国布鲁克林儿童博物馆展览部的工作人员在制定新的展览开发过程时提出加入评估步骤，但博物馆负责人担心没有足够的人财资源去做评估，同时对评估的益处表示怀疑。为此，展览部人员须向博物馆负责人证明，评估不会过多地加重人员或预算的负担，评估有助于提高博物馆展品和计划的质量。他们在有关机构的资助下，启动了一个特别项目，通过聘请顾问、举办评估培训以及对一些展览项目进行评估尝试等措施，在博物馆推行评估做法。由于他们的努力，该博物馆目前对大多数的展览及计划都进行某种形式的评估。根据多年的实践，布鲁克林儿童博物馆总结出 8 条评估经验，我们选择其中较为重要的 4 条，以供人财资源同样不宽裕的我国博物馆学习借鉴。

美国布鲁克林儿童博物馆的评估经验

1. 从小小的方面做起

观众研究和展品评估可繁可简,博物馆不可能一开始就进行繁琐的这类工作。先从简单的预评估或总结性评估开始,比如弄清楚观众对展览主题知道些什么,是否明白某一小展品的说明性标签内容等。

2. 力求简单

展品评估不需要搞得很麻烦或很花钱。事实上,预评估和总结性评估很容易纳入展览开发过程,我们几乎没有什么理由不去做它。评估不一定是经过严格组织的很正式的过程,它可以搞得很简单,如让10位观众朗读展示说明,如发现他们在同一地方犯结巴,或者不能用自己的话解释其意思,那么这一展示说明应予以改动或重写。

3. 清楚自己最想了解哪些东西

通常,我们无力做到对一项展览的各个组件都进行测试评估,因此,确定哪些组件是最需要测试评估的就很重要。评估不过是又一种解决问题的方法而已。它能决定一样东西在概念或机械结构上是否可行,某一概念是否过于抽象,或者某一字词是否令人难以理解。这些都是具体的要评估的方面。

4. 获取专家的帮助

如果条件允许,最好聘请专家指导整个评估过程,因为他们清楚评估从哪儿开始做以及如何做,而且善于拟定要评估的问题(这是评估的难点)。如果负担不起聘请专家参与整个过程,可以让专家协助你制定评估计划,并教你去做。有时花上几个小时的咨询时间就会使你对评估的关键方面了然于胸。如果你连专家咨询也负担不起,可以阅读有关的文献。这方面的文章书籍已有不少,有些甚至提供你一步一步做评估的步骤。(推荐两本优秀的读本:Ross Loomis, *Museum Visitor Evaluation: New Tool for Management*; Sam Taylor, ed., *Try it! Improving Exhibits through Formative Evaluation*)

二、媒体项目的评估(以电视科技节目为例)

电视作为现代最有力的传播媒介,以其形象、生动、真实、感染力强、传播速度快和覆盖面大而成为有力的科普手段。在四大传媒中,电视拥有最广泛的受众群,人们平均花在看电视的时间高于其他媒体。电视科普有多种表现形式,主要包括科普专题或栏目(如中央电视台的科技博览、走进科学)、科普纪录片(如探索、国家地理杂志)、科技消息与新闻以及现场直播科普节目四大类。

在电视行业,对节目的评价主要从节目收视数量和收视质量两个纬度进行衡量。节目收视数量的评价指标是收视率。收视率是指某一特定时段内,收看某一节目的人数占电视人口的百分比。节目收视质量是指观众收看节目后的综合心理反映,国际上一般用观众的欣赏指数或观众满意度来反映观众的收视质量。

收视率和欣赏指数作为评价节目的两个基本指标,有着很大的不同。收视率所追求的是对收视状态的最基本了解,要解决的是最直接的表层现象描述。它只显示收看节目的人数多寡,但不能反映观众对节目的喜好和评价;欣赏指数所追求的是对受众的需求及满足程度的了解,要解决的是深层的心态和需求的解释,一般包括受众对节目的认知度、喜欢度和推荐度等多项评估指标。

对科普节目来说,由于其所担负的公益性科普教育的社会责任,单凭收视率大小评判节目的成败显然不可取(有的专家甚至反对使用收视率),而欣赏指数因能考核节目是否满足受众需求而显得更为重要。

不过,据我们了解,我国电视广播系统迄今尚未引入欣赏指数评价。再有,无论是收视率还是欣赏指数,目前还都不够完善,而且都是量化指标,缺乏对节目的定性分析思考。另外,科普节目具有教育特质,节目收视质量评价既要考虑节目本身的制作质量和吸引力,更要特别关注节目的科教效果,即观众通过收看节目实际

学到了什么,对科技的认识、态度和价值行为是否因此产生了变化,如此才能体现科普节目价值的全貌。基于这些原因,有必要提出一套专门针对科普节目的评估指标。我们认为,科普节目的评估指标应包括节目内在质量、收视效果和宣传外延工作三大方面。

(一)评价节目内在质量的指标

节目是传播内容的载体,是传播者与观众对话的平台。节目的内在质量会直接影响观众的收视效果。衡量节目内在质量的指标主要包括:节目选材是否适合目标观众的需要(適切性),内容是否够棒(如节目内容的科学性、节目的起承转合、内在的逻辑、语言的味道、故事的节奏)、表现形式和手段是否具有吸引力等等。这些通常由业内专家评议获得。

(二)评价节目收视效果的指标

根据传播效果的层级性,衡量节目效果的综合指标应包括如下几个层次:

1. 衡量节目作用于观众感觉、知觉层次的指标

衡量受传者对节目的感知程度的主要指标是收视率。不过,对科普节目评估而言,收视率指标的使用要格外慎重。这是因为,影响收视率的因素很多,除节目本身对观众的吸引力外,诸如目标观众规模、节目类型、节目所处时段、本频道前后播出节目的影响、其他频道同时段节目的影响、本频道观众亲和力、季节性的影响等都会对节目的收视产生影响。此外,收视率指标本身的测量信度和效度也值得推敲。信度是指工具(包括方法)本身的可信程度,效度是指调查结果与真实情况的符合程度。经常碰到这样的现象,在某一个城市有三家收视率调查公司同时为一家电视台提供收视资料,而它们之间的资料差异往往超过了统计误差范围。这表明收视率这一调查工具存有方法本身的不稳定性和方法之间弱

比照等的局限性。

在进行收视观众调查时,还应关注观众的性别年龄文化构成、收视动机等背景因素,对这些指标的分析能使我们获得更深层次的信息。

鉴于节目收视率自主性能较差,在进行节目收视率评价时,纵向和横向比较都有弊端。有专家认为,可以采用收视率评价动态基准值的办法。基准值的确定以过去一段时期同时段同类栏目的平均收视表现为依据,时间长度建议以一年较为合适。一般时间太短,没有代表性;时间太长,电视市场格局可能有很大变化,缺乏可比性。

2. 衡量节目作用于观众思维层次的指标

这类指标包括观众对节目内容理解了多少,学到了什么。

由于存在知沟等障碍,观众对节目内容往往做不到完全的理解。观众有误解信息的可能性,也会以各种方式曲解信息。换句话说,他们所理解的和节目原来要传达的内涵,可能会大不相同。好的科普节目应动用尽可能多的阐释手段促进观众完整正确地理解节目的内容。

衡量观众看科普节目后到底学到了什么,不仅要了解事实学习的情况,更重要的是了解观众所获得的高层次科学思维能力的情况。事实学习涉及节目所传播的科学概念和知识点,相比之下较易于衡量。而考察观众所获得的科学思维能力则相对困难得多,需要多种评估手段的综合使用、精心设计问题才能实现。美国罗克曼研究与咨询公司儿童科普系列片“比尔·奈科学人”所做的评估,在考察孩子们在观看“比尔·奈科学人”节目后学到了什么时,就将重点放在了解孩子进行探究和解释科学现象所具有的科学思维能力和技巧上,他们通过提出一些应用型的科学问题,来考察儿童的科学思维和解释能力。他们所设计的问题和提出问题的方式相当巧妙,对我们将来开展类似工作很有启示(可参见附件

三)。

3. 衡量节目作用于观众态度及行为层次的指标

科普节目不仅作用于观众的感知觉和思维层次,而且还可以进一步影响观众对科学的态度,甚至激发观众的某种行为或行为意愿,达到深层传播的效果。很多有成就的科学家走上科学研究的职业道路,据说很大程度上就是小时候受到一本科普书、一部科普电影或电视节目的激励。

态度指标主要是调查观众在观看节目之后,对科技的态度是否发生了积极的变化以及变化的程度有多大。

行为指标由完成的行为和行为意愿组成。比如,儿童在看了科普节目后,仿照节目的内容做小实验,这属于完成的行为。行为意愿同样重要,如孩子通过看科普节目,激发了对科技的兴趣,期望将来从事科技职业。

4. 衡量观众对节目欣赏程度的指标

该指标主要通过了解观众对节目的满意度获得。和收视率一样,满意度也受到观众背景变量的影响。英国有研究显示,满意度受观众年龄、性别和社会阶层的影响。年幼和年老观众倾向于给予节目较高的满意度评分,而中青年观众则很难讨好,他们不轻易对节目打高分。从性别来看,女性比男性慷慨,乐意给节目较高的评分。低下社会阶层的观众也比较容易满足,倾向于给节目较高的评分,而中上阶层则比较挑剔。但与收视率不同的是,满意度受节目播出时间的影响不大。一般来说,好的节目即使被安排在不太好的时段播出,只会影响其收视率,而不会大会影响其满意度。在进行观众满意度调查分析时,应注意这些因素。

(三)衡量节目宣传与外延工作的指标

宣传推广和面向学校和社区的外延工作是增加观众对节目的

认知度、扩大节目教育影响的重要手段。所谓外延,就是配合电视科普节目的播出深入学校和社区展开的相关活动(多半是非播出性质的活动),目的是要扩大节目的效果。美英等国家非常重视电视科普节目的外延工作。他们认为,像青少年科普节目,即使收视率甚佳,能达到的青少年也非常有限,而外延最能够达到学生,故主动接触学校社区十分重要。如果要外延达到学生群里面,就要透过跟学校的合作。比如,以电视科普节目内容为主线,设计学校辅导性科普读物,并利用网站提供每集节目的简单教案及教学指引,让有心利用电视科普节目进行教学的教师可立即运用于课堂。国外有影响的科普节目大都开展了相应的外延工作,并把外延作为科普节目的一个重要组成部分,这种做法很值得我们国内的电视媒体借鉴效仿,我们认为有必要将此作为一项衡量指标,以激励我国电视科普节目在深入学校社区上有所创新和作为。

三、科普活动项目的评估

除展览和传媒(如电视)两大科普方式外,科普工作还包括其他形式多样的科普活动,这些活动的类型发展得越来越丰富,从科学发明创造博览会、夏令营、科技俱乐部、科普讲座、科技知识竞赛到大学研究机构的见习研究等,不一而足。科普活动的规模也有大有小,期程也不尽相同,有的活动是一次性的努力,有些则属长期性的项目。

对这些科普活动项目,我们借鉴科协西部科普工程项目的实际评估经验,给出了基础条件和能力、组织实施过程、产出、效果与影响、满意度5类评估指标。这5类指标涵盖了项目背景、投入、产出和结果,是对活动项目的完整评估。在实际工作中,是否需要对项目进行从背景条件到结果的全程评估,应视活动的规模、性质等具体确定。比如,对小项目,可能进行效果评估就可以了。

科普活动项目的评估指标框架可以包括以下内容:

1. 基础条件和能力

项目实施的基础条件和能力评估,主要是为了客观评价项目实施的必要性和可行性。评估的内容包括:实施项目的适当性;为保证项目的实施,项目承担单位自身所能提供的组织协调能力、人员工作能力和经验、配套经费落实、物资和条件等方面的保障情况;项目计划所依据的政策支持、公众支持、专家智力支持、合作单位配合的基本情况。

2. 组织实施过程

项目实施方案的制定过程及评估:项目实施方案制定的依据,是否经过必要性和可行性分析,并根据专家评审意见进行过修改和完善;有没有根据项目实施过程中反馈的新问题及时修正项目实施方案;项目所制定的实施方案对于保证项目最终取得实效的客观评价。

项目实施过程的控制方式及情况:项目是否有计划地实施;项目的管理过程是否规范;项目经费是否做到专款专用;配套经费是否及时到位;是否对项目的管理和实施人员开展业务培训。项目总结和宣传情况:在项目执行过程中有没有注意收集整理相关资料和素材,并进行总结和分析;开展了哪些方面的项目宣传。

3. 产出

即通过项目的投入直接导致的成果,如开展活动次数,参与活动的人次,编写的科普材料发放数量等。

4. 效果与影响

即通过项目产出所达到的效果(直接结果)或影响(间接结果)。项目实施的效果应包括目标群体受益和项目承担单位受益等方面。

目标群体受益情况:通过参与科普活动,活动参加者是否增强了对科技的了解,他们的科技意识和科技态度发生了哪些积极的

变化,产生了什么行为影响。

项目承担单位受益情况:通过项目的实施,是否增强了项目执行机构开展科普工作的能力(如更新了科普手段,对项目实施人员进行培训,提高了项目管理人员和实施人员的项目管理水平等),以及积累了哪些项目管理方面的经验和今后申报项目以及多渠道筹集项目经费的经验。

5. 满意度

即有关方面对活动的满意程度。例如,活动参加者对活动内容与方式的满意程度、对承办机构的组织服务水平的满意程度;当地党政部门对活动的肯定和支持情况等。

第五节 科普管理部门/执行机构能力与绩效的评估

一、政府科普管理部门能力与绩效的评估

政府科普管理部门能力与绩效评估的目的,在于提升各级政府科普管理部门的能力和水平、提高政府工作的效率和效果以及确保国家在科普事业上的投资取得最大收益,故政府科普管理部门的能力与绩效评估应从管理机制和过程、能力条件建设、绩效结果三个纬度来衡量。同时为强化责任意识,施政绩效的评估应以结果为导向,并建立鼓励创新改良工作的机制。

1. 管理机制和过程

有无战略规划或远景目标、受资助项目的评审过程和方式、项目的实施是否受到监督、是否要求项目执行机构进行项目评估或总结、是否利用评估结果改进工作、是否注意到对新机会的有效识别和支持、有没有对历年的项目信息进行数据库管理并实现信息

共享、是否努力推广好的经验和做法、是否建立起鼓励创新改良本部门工作的机制和文化？

2. 能力条件建设

有没有稳定的预算？预算额多大？近年有没有增加？有没有从民间争取科普赞助经费的能力和经历？政府机构中有没有专门的部门负责科普工作？编制几人？他们投入科普工作的全时当量多少人年？其教育背景、工作经历、年龄结构是什么样的？对工作人员有没有进行相应的培训？

政府科普管理部门与民间科普专业机构、大众传媒、大学、研究机构、企业、其他可资利用的社会团体的关系如何？是否得到了它们的支持配合甚至建立了某些长期合作关系？

3. 绩效评估

目的是评判政府科普工作的效率和效果，特别是效果。效果的好坏主要通过评估政府科普工作目标的实现情况来衡量。

绩效评估是对政府科普管理部门工作过程和结果的评价，属于典型的事后评估。政府科普管理部门绩效的测定及评估比较复杂和困难。这是因为，一方面，政府科普工作的目标是软目标，表述笼统、抽象且难以量化为硬性指标，如提高公民的科技文化素质、激发青少年对科技的兴趣等，这为政府绩效的测定带来了一定的难度。另一方面，政府投入及产出的度量也使政府绩效难以确定。行政部门的产出不像企业的产业那样最终进入市场的交易体系，形成一个货币价格，其产出属“非市场产出”。至于组织的投入，行政部门的生产具有较强的劳动密集型特征，劳动投入是其关键投入，行政效率的高低主要取决于劳动的边际生产率，政府机构中工作人员劳动实际供给程度对政府绩效的高低具有重要作用，而劳动的投入往往难以量化。

由于以上原因，在实际的政府绩效结果评估中，通常只根据政

府科普工作目标的实现情况进行定性评价。

二、科普执行机构能力与绩效的评估

在我国,承担科普使命或活动的机构有很多,既有科技博物馆、青少年科技活动中心这样的专门科普机构,也有众多挂靠在大学、研究机构的科普教育基地。这些机构是我国科普事业的主要力量,其组织能力和工作绩效直接关系到整个国家科普工作的兴衰成败。开展对科普执行机构的能力绩效的评估,有助于增强我国科普机构的使命感,解决我国科普机构能力长期欠发达的问题,促进我国科普机构组织效率的提高和不断成长。这里,我们以科技馆为例,阐述科普执行机构评估需考虑的内容和方面。

科技馆实施评估的目的有三:一是评估观众的满足度,旨在了解科技馆的教育效果;二是考量科技馆内部运作的方式和效率,以了解其运作是否符合科技馆的使命和目标;三是衡量科技馆付出的人力、物力等代价是否符合服务的效果,以知悉科技馆的运作效益是否平衡。因此,科技馆评估的指标框架应包括:目标与使命、组织与运作、展览教育活动(包括外延教育活动)的开展情况、运作效益等方面。

1. 目标与使命

科技馆是否制定了战略规划和实现规划的战略?战略规划确立的目标是否考虑了现实的环境条件和未来的发展需要,具有可行性和前瞻性?战略规划是否得到切实有效的执行?

科技馆是否切实履行了面向目标群体进行科普教育的使命?受众面是否足够宽泛?是否考虑了弱势群体的特殊需要?

2. 组织与运作

管理方面,科技馆是否建立起了合理有效的组织结构和管理运行机制?对展教项目或活动是否进行监督或评估,是否利用监

督信息和评估结果改进工作？是否建立起鼓励创新改良科技馆工作的机制和文化？

能力与条件建设方面：科技馆是否能够吸收并保持有合格能力和素质的员工(包括义工)，是否为员工(包括义工)提供了合适的培训，有没有努力创建一种学习型组织？科技馆的资金是否足以支持和保证其使命的达成，近年有没有增加？科技馆是否有较强的创收能力和筹款能力，资金来源是否多样化？科技馆设施的管理与运行是否满足业务开展的需要？

公关与宣传方面：科技馆与政府部门、所在社区、大众传媒、学校、研究机构、企业以及其他可资利用的社会团体的合作关系如何？是否得到了它们的支持配合甚至建立了某些长期伙伴关系？有没有与国内外同行开展经常性的交流与合作？是否利用各种营销手段(包括网站)使社会和广大公众了解科技馆？

3. 展览教育活动及效果

科技馆举办了哪些展览教育活动(包括面向学校和社区的外延教育活动)，展览教育活动的质量、效果和影响如何？

科技馆是否能识别新的机会，以创新的思想，扩大和完善展教活动，从而形成自己的特色和优势？

科技馆年参观人数，采取了哪些有效的措施吸引观众和回头客？

观众满意度评价，包括对科技馆展览、环境、服务、门票等的评价。

4. 运作效益

我国的科技馆基本上属社会公益性非营利教育服务机构，运作效益主要体现在社会效益上，但这并不意味着科技馆可以忽略市场效益，可以乱花国家和捐助人的钱。国外科技馆成功的例子说明，科技馆在保证其科普功能不缩水不走样的前提下，如能将市

场机制引入办馆中,运用营销手段,会最大程度地提高资金的使用效益,有助于解决科技馆经费窘迫的问题,最起码能减少“赤字”。

衡量科技馆运作效益的指标应包括:科技馆收支情况、科技馆单位面积展厅的年接待人数和成本、单个观众获益所需的资金、科技馆应保有的观众数、是否采取了积极有效的营销措施提高科技馆效益等。如果科技馆效益不佳,是什么原因造成的?

第四章 评估程序与评估方法

第一节 评估程序

通常,评估者在进行正式评估以前需要草拟一个评估的方案与步骤,这是进行评估工作的必要程序。然而,评估程序并不是固定不变的,而是根据具体情况可以灵活调整。一般而言,战略规划或计划的评估、重大活动或项目的评估以及科普管理/执行机构的评估的程序都不完全相同,但也有其共性的一面。本节着重介绍的是科普项目评估的程序或步骤。

评估程序通常包括以下几个步骤:第一步确定评估的目的。无论进行何种类型的评估,这都是评估者首先需要明确的问题;第二步确定评估的重点与关键性问题;第三步选择评估指标;第四步确定评估的方式;第五步编制执行计划;第六步处理与分析数据;第七步撰写评估报告;第八步交流评估结果并制定后续计划。

一、评估的目的

评估的目的是评估工作首先需要确定的问题。它有助于我们进一步思考所提议的评估工作。如果评估的目的明确,那么往往能够取得事半功倍的效果;反之,如果评估的目的不明确,那么评估就缺乏明确的方向,其结果不仅是浪费了资源,而且也起不到评估应有的作用。

表 4-1 提供了一个确定评估目的的基本思路。即在明确了

“谁想评估”、“他们为什么想进行评估”、“他们打算如何运用评估的结果”等问题后,便可以清晰地给出评估的目的。

表 4-1 评估目的的制定过程

谁想评估?	他们为什么想进行评估?	他们打算如何运用评估的结果?

评估的目的是:

表 4-2 列举了两种不同的情况,每种情况下评估的目的都不相同。例如,第一种情况的评估目的就是:“项目资助机构想了解项目是否达到了预期结果,然后决定这一类项目的资助价值”;第二种情况的评估目的是:“项目的管理人员希望通过评估了解项目的得失,然后决定项目管理是否需要变革”。

表 4-2 不同情况的评估目的

谁想评估?	他们为什么想进行评估?	他们打算如何运用评估的结果?
资助机构	了解项目是否达到了预期的结果	决定这一类项目的资助价值
项目管理人员	了解项目的得失	决定项目管理是否需要变革

二、确定评估的重点与关键问题

在明确了评估的目的之后,接下来的工作就是要确定评估的重点与关键性问题。通常,摆在评估者面前的问题非常多,且杂乱无章。这时,评估者需要冷静地思考哪些是需要重点评估的、哪些是一般性的、哪些是次要的,甚至可以忽略不计的。

评估者在确定评估的重点与关键问题时,需要考虑以下三个问题:

第一 根据评估的目的决定评估的重点。一方面,不同的评估目的关注的侧重点并不相同,因此,在选择评估的重点与关键问题时,首先必须依据的就是本次评估的目的;另一方面,评估的目的不仅清楚地描述了利益群体关注的重点,而且描述了如何运用评估的结果。它决定了评估所需要优先收集的信息。

第二 根据评估的经费与人力决定评估的重点与关键问题。评估是一项耗时耗力的工作,因此,评估的经费和人力也就决定了评估的范围、评估的时间,评估的重点内容。当经费与人力充足时,可能评估的内容就更多一些;而评估的经费与人力有限时,评估的内容就更少一些,评估的重点也就更集中一些。

第三,考虑不同的利益群体已经了解了什么,不了解什么,还需要了解哪些重要的信息。为了提高评估的效率,这也是评估者在选择评估的重点时需要回答的问题,否则可能出现事倍功半的结果。例如,评估工作所得出的结论都是相关利益群体已经了解或知道的信息。在明确了以上三个问题后,评估者便可以确定评估的重点与关键性问题。

三、选择评估指标

一旦评估的重点与关键性问题已经明确,那么接下来的任务就是选择具体的测度指标。指标是指反映总体现象的特定概念和具体数值。任何指标都是要说明一定社会总体现象的某种属性或特性。通过一个具体的统计指标,可以认识研究现象的某一特征,说明一个简单的事实。如果把若干有联系的指标结合在一起,就可以从多方面认识和说明一个比较复杂现象的许多特征及其规律性。

正因为如此,评估指标的选择也就变得较为复杂与困难。具体来说,每一个关键性问题几乎都可以列出许多测度指标,而评估者既可以采用一个指标来测度某一现象,也可以采用一系列指标

来反映某一现象。因此,在选择评估指标时,需要经过认真仔细的研究。一方面必须遵循构建评估指标体系的原则,另一方面也可以由评估者进行初步筛选,然后请一些同行专家进行评议,或采用其他科学的方法选择评估指标。

四、确定评估的方式

当以上所有问题都清晰以后,就可以确定评估的方式。通常,可供选择的评估方式或者说数据收集的方法包括文献法、问卷法、访谈法、观察法等等。需要注意的是,中国的科普经费较为缺乏,评估工作应尽量本着节俭的原则。因此,在选择评估方式时,应尽量选择成本低的方法。第一,文献法是评估方式中成本最低的方式。因此,如果能够通过文献法收集到信息(包括查阅有关文件、档案、年度报告、申请报告、监测数据和政府有关部门的统计资料等),就不必采用其他成本更高的方式,这样可以降低评估调查的时间、人力与资金成本;第二,在评估调查中,采用抽样调查的方法比普查有更高的效率。因此,在不必要的情况下,尽量避免进行普查;第三,在需要的情况下,可以采用定量与定性相结合的方法,采用文献法、访谈法、观察法等相结合的方式。

五、编制执行计划

第五步工作就是写出评估的执行计划,即构成整个评估过程的各个部分。包括:

(一)组建评估小组

众所周知,评估水平取决于评估人员的素质。组建评估小组,即决定谁加入评估小组,他的任务又是什么?而决定谁可以加入评估小组又取决于前面4个步骤的形成。

组建评估小组,最重要的一点是有一个好的评估协调人,他对

小组成员的工作起一个良好的协调作用,并对评估工作是否按计划进行负责。

评估小组的其他人员应该是评估项目领域的专家或行家。评估人员必须熟悉和掌握该项目领域的国内外现状和动向,最好是直接从事该领域研究或工作的、知情的、有远见的、懂评估的专家。只有这样才能做出科学的评估意见。

另外,评估小组最好吸收项目目标群体的代表参加,至少在关键的阶段,如计划和反馈阶段应征询项目目标群体人员的意见。

(二)分派任务

在确定了评估小组的成员之后,就可以将所有的评估任务分派给每一个成员。整个小组并不需要参与所有的活动,但是,编制计划必须列出需要做的每一项任务和活动及其责任人。表 4-3 给出了一个任务分派的例子。

表 4-3 小组成员及其职责

任务与活动	小组成员 A	小组成员 B	小组成员 C	小组成员 D	小组成员 E
①文献收集	#		#	#	
②评估内容 a	#	#			
③评估内容 b				#	#
④与目标群体访谈			#	#	
⑤与工作人员访谈		#			
⑥数据处理	#				#
⑦形成初步的评估结论和建议	#	#	#	#	#
⑧写评估报告	#		#		

注: # 代表所分配的任务,下同。

(三)日程安排

日程安排也是编制评估计划的重要环节。它应包括每一项任务完成的时间安排。表 4.4 给出了一个日程安排的例子。其中,两天的评估讨论会是编制整个评估计划的必要环节,甚至可以说是评估进程的起始点。然后可以制定完成评估报告的截止日期,这将为所有其他活动的完成时间提供参照。

表 4.4 评估的日程安排表

任务和活动	责任人	7月2日	8月2~3日	8月4日	8月15~18日	8月19日	8月20日	8月21~23日	8月31日	9月1日	9月21日	10月2日
①第一次评估设计讨论会(确定评估目的并决定是否聘请外部专家)	所有的利益群体代表	#										
②第二次评估设计讨论会	内部小组成员与外部专家		#									
③收集信息、文献观察 与受众访谈	外部专家与协调人 评估小组 外部专家与协调人			#	#	#						
④分析、小组会议 小组成员写报告 完成报告初稿	评估小组 评估小组 外部专家与协调人						#	#		#		
⑤发送报告初稿	协调人									#		
⑥反馈后完成报告	外部专家与协调人										#	
⑦追踪会议	所有的利益群体											#

通常,科普机构的自我评估与外部专家评估所需要花费的时间不太相同。一般情况下,科普机构的自我评估花费的时间较长,

而且评估的时间较为分散、断断续续,而外部专家评估所需花费的时间较短,而且评估的时间较为集中。在评估的过程中,科普机构内部的成员在花费部分时间进行评估工作的同时,通常还需要完成他们日常的工作职责。当然,也可能需要有专人集中时间来完成评估工作,时间一般在几天到两周之间。但不同类型的评估花费的时间有很大差异。例如项目评估所需花费的时间通常较短,而战略规划评估所需花费的时间通常较长。例如,整个评估的时间,有的需要几天,有的几个月(包括同所有利益群体回顾评估结果的追踪会议)。

(四)预算

考虑到评估的价值,因此,在一些大型项目或特别重要的项目中,需要在项目申请书中预算评估的费用。通常,进行自我评估的费用相对更少一些,而请外部专家进行评估的预算费用则高一些。一般外部专家评估的费用比内部人员的评估预算费用要高出40%左右,甚至更高。评估费用主要包括评估调查的差旅费、评估人员的劳务费和一些其他基本费用等(例如,调查问卷的制作费、评估调查时准备的小礼物等)。

六、处理与分析数据

在收集了数据信息后,就进入了关键的数据信息分析阶段,并依据分析结果写出初步的评估报告。

通常,在对数据信息进行分析之后,需要将一些分析结果反馈给提供信息的个人或团体。他们往往很有兴趣看到这些结果,并对此结果作出回应,而评估者在形成最终评估报告时应充分考虑到这些反馈的信息。

数据处理分析可以分为以下三个阶段。

(一) 数据处理并制表

数据处理是指把调查中记录的各种分散的信息,经过分类汇总,汇集成表示总体特征的信息的全过程。数据处理的方式包括手工汇总和计算机汇总两种形式。一般,有计算机的机构可以使用一些软件包,如 SPSS 等统计软件包处理数据并列交互表;没有计算机的机构则可以通过手工汇总处理简单的信息资料,并得出简单的交互表。

数据处理一般有以下步骤:对记录的信息进行手工检查校正、编码、录入、利用计算机对录入数据再次核对检查、制表等过程。

数据处理的最终反映形式是编制各种表格,即制表。制表又分为两个阶段,第一阶段是设计汇总表;第二阶段是出表。表 4-5 给出了一个频数交互表的例子。从表中可以看出,频数表可以较为清晰地反映干预前与干预后知晓率的变化,也即项目的效果。

表 4-5 预防艾滋病项目效果评估——知晓率变化频数表

指 标	项目实施前(%)	项目实施后(%)
艾滋病会通过母亲传染给婴儿	48.3	76.8
共享碗筷等日用品会传染艾滋病	37.8	64.7
和艾滋病人接吻会传染艾滋病	57.5	89.2
昆虫叮咬会传染艾滋病	15.6	57.3

(二) 对制表结果进行讨论与分析

在编制了各种交互表之后,评估小组就需要对这些处理的信息进行讨论与分析。对于一些定量定性指标,可以采用前后对比、有无对比等方法进行分析,从而作出评判和结论。

在制表之后,评估小组可以在此基础上形成初步的结论和建议。结论与建议可以分类列出。例如,何种信息来源导致了何种

结论,何种建议是根据什么结论得出的。这样可以增进评估结果的有用性。特别要注意的是,少而精的、具体的、有针对性的建议比混杂的、面面俱到的建议更有实质性价值。因此,在写评估报告时,应避免提出空洞的、不具有可操作性的结论。

(三)对评估的再评估

评估的一个重要目的之一是为了从评估过程中学习,并对评估工作本身进行学习与提高。因此,评估小组还应当对评估工作本身进行再评估。这既可以通过一般性的讨论,也可以通过一些规范的再评估格式进行。专栏 4-1 给出了一种再评估的格式可供参考。它用于收集每一个与评估有关的人员对评估的结果和恰当性的意见。

专栏 4-1 评估的再评估格式

请在每一栏的“是”、“否”中画圈,如果选择了“否”,请说明为什么。

1. 您认为,此次向工作人员介绍和描述的评估目的和评估方法:

- | | | |
|-------------------|---|---|
| ① 有帮助 | 是 | 否 |
| ② 使我们很快进入评估的角色 | 是 | 否 |
| ③ 使我们认为这是一次有价值的活动 | 是 | 否 |

为什么: _____

2. 您认为,本次评估的重点:

- | | | |
|-------------------|---|---|
| ① 我认为是的重点问题都被考虑到了 | 是 | 否 |
| ② 最重要的一些问题都被包括了 | 是 | 否 |

为什么: _____

3. 您认为,访谈或问卷:

- | | | |
|--------------|---|---|
| ① 所提的问题很恰当 | 是 | 否 |
| ② 我自己能够轻松地表达 | 是 | 否 |
| ③ 能以得体的方式进行 | 是 | 否 |

续表

为什么：_____

4. 您认为：

- | | | |
|---------------------|---|---|
| ① 被选择的评估项目具有代表性 | 是 | 否 |
| ② 评估人员很机敏,以致访谈很轻松 | 是 | 否 |
| ③ 被调查的人能够自由地说出自己的看法 | 是 | 否 |

为什么：_____

5. 您认为,评估的过程：

- | | | |
|------------|---|---|
| ① 是一个学习的经历 | 是 | 否 |
| ② 能激发人的热情 | 是 | 否 |
| ③ 我觉得自己很投入 | 是 | 否 |

为什么：_____

6. 您认为：

- | | | |
|------------------|---|---|
| ① 建议是恰当的,并且是有用的 | 是 | 否 |
| ② 建议表达了主要的考虑 | 是 | 否 |
| ③ 建议是依据发现的事实所得出的 | 是 | 否 |

为什么：_____

请您写出其他的您想表达的意见：

对评估结果的再评估有助于评估人员通过评估实践深切体会和掌握评估知识。这不仅有助于增进评估人员的评估能力,也有助于总结评估经验,发现评估中存在的问题,从而提高科普评估者的评估能力和评估工作的质量。

七、撰写评估报告

在完成以上六个步骤后,评估小组就需要编写评估报告,全

面、系统地反映评估目的、内容和结果。通常,评估报告由内容提要、报告正文和附录三部分组成。当然,评估报告的格式并不是固定不变的,评估小组可以根据具体情况确定评估报告的格式。

(一)内容提要

对于一份完整的、高质量的评估报告来说,1~2页的内容提要是非常必要的。特别是对那些没有时间通读全文的读者来说,评估小组提纲挈领地写出内容摘要有助于他们了解整个评估报告的核心内容。

内容提要通常包括评估的目的、评估的侧重点、评估的过程(评估的项目点、评估调查方法与基本数据等)、主要的发现与结论、政策建议等内容。

(二)正文

正文一般在数十页之间。通常,正文也包括三部分的内容,即绪论、主体内容、结论与建议。

1. 绪论

绪论主要介绍以下一些内容:

第一,背景。包括评估的目的、评估的范围等基本情况;

第二,评估小组的构成以及评估调查的方式(即收集数据信息的方法)。

2. 主体内容

评估报告的主体内容是对项目的适当性、效果、效率、影响、持续性和服务质量等各方面的基本评价。由于这部分是整个评估报告的主体,内容较多,因此常常被分成几个章节。

3. 结论与建议

这部分主要是阐述整个评估活动的结论以及依据评估结论所提出来的相关政策建议。通常还可以包括本次评估对未来或后续

项目的经验与教训、对评估活动的再评估等内容。

(三)附录

附录部分主要包括一些不宜放在正文但又有助于读者全面了解或理解正文的资料,例如数据和制表、调查问卷、访谈提纲、参考文献等方面的内容。

八、交流评估结果并制定后续计划

评估报告的完成并不意味着评估工作的结束。在完成评估报告之后,还需要进一步发挥评估报告应有的作用,交流评估的结果并制定后续计划。

(一)交流评估结果

评估报告完成后它又有什么作用呢?评估工作仅仅是为了完成一份评估报告吗?答案显然是否定的。然而,不幸的是,现实生活中却常常出现这种现象。一些机构在例行公事般完成评估报告后,便将评估报告束之高阁,变成了一堆废纸。这使得原本是一件具有非常积极意义的工作却转而成为了一种巨大的浪费。

这实际上又回到了第一步所思考的问题,即为什么进行评估?谁想进行评估?评估的结果用来做什么?

如果评估的结果不被相关的利益群体所分享,不能发挥其应有的作用,那么评估就毫无价值。因此,关键的问题是使得相关的利益群体分享评估的结果。这就需要评估小组通过各种方式在不同的利益群体之间交流评估的发现、结论和建议。

(二)制定后续计划

仅仅交流评估结果的信息,还并不足以发挥它的真正效益。因此,还需要制定后续的计划,使得评估的发现能产生影响,评估

的建议能得到贯彻落实。而根据我们的经验,目前中国的许多机构在进行评估时,往往忽略了这一最为核心的步骤。以至于评估不仅没有取得应有的效果,反而浪费了人力、物力。

后续计划的一个主要办法是召开后续会议,并使各相关利益群体的代表,例如资助机构代表、目标群体代表、项目负责人能尽量参加会议,通过会议使各利益群体尽可能对评估的结论和建议达成共识。在此基础上,将建议分解成具体的、可操作的任务或活动。有时,一项建议可能需要分解成几个活动。当然,没有达成共识的建议也就很难有后续的活动。

在明确了具体的、可操作性的活动之后,需要推选每一项活动的责任人,并估计每一项任务完成的日期。最后,还需要建立监测机制,使得活动按计划顺利进行。表 4-6 给出了一个基本的格式,可供科普机构制定后续计划时参考。

表 4-6 后续计划的格式

评估报告的建议	具体的活动	责任人	预计完成日期	实际完成日期

第二节 评估调查方法

在确定了评估的步骤或程序之后,接下来的问题就是具体的落实与操作。在所有的操作过程中,一个关键的问题就是如何收集相关的评估信息与数据,即采用何种方式进行评估调查。评估调查是指以评估为目的的社会调查方式,属于社会调查的一种特

殊的情况。通常,评估调查的方式有很多种,包括文献收集的方法、问卷调查方法、访谈法、观察法等等。各种调查方式具有不同的优缺点,适用于不同的评估对象、评估内容。因此,针对不同的评估调查应选择恰当的方式。例如,文献法适用于调查基本的背景情况,问卷法适用于知识调查,如目标群体对某一方面知识的知晓率,访谈法适用于态度调查,观察法适用于行为调查。在数据收集过程中,一般可采用2种或2种以上的调查方式,相互补充。可能的话,还可以比较不同方式的结果,以确保评估调查的信度与效度。

评估调查除了为预测或评估提供依据外,通常还具有其他一些功能。包括:了解目标群体的需求,以帮助评估人员或决策者分析所开展活动的目标是否正确或是否可以实现;在调查过程中,通过评估人员的作用,沟通科普机构与政府部门、其他科普机构、目标群体等各种利益相关群体之间的关系,促使各个方面的相互理解和相互帮助;在评估调查中,组织和动员项目、活动所在地的各种利益相关群体和组织参与评估与实施,根据调查所了解的情况,有针对性地提出措施与方案,以满足未满足的需求;根据评估调查所了解的信息资料进行综合分析,在此基础上提出必要的修改方案和建议,以供决策者参考。

一、文献法

文献法是评估调查中最常用的方法之一,甚至可以说几乎所有的评估调查都需要以文献法为基础和起始点。这是因为,无论采用问卷调查法、访谈法、观察法等,评估者事先都需要了解有关项目的背景材料。这样,在评估中才能做到有的放矢。

评估调查中,文献通常是指与调查对象有关的信息的一切书面文字材料。文献法,也叫二手资料调查法,就是通过收集有关的各种文献资料,选取其中对评估有用的信息的方法。

与问卷法、访谈法、观察法等调查方法相比,文献法最大的特点是把研究对象由直接方式转变为间接方式。评估调查人员不需要同目标群体等具体的人直接接触,而是查找各类文字材料。这种调查方法对评估者来说,既有优势的一面,也有其不足之处。

文献法的优点在于:

首先,文献法具有无反应性的优点。由于各种形式的文献研究都不需直接同评估对象打交道,而只是研究那些业已存在的文字材料、数据以及其他形式的信息材料,因此,在整个评估过程中,评估对象不会受评估者的影响而发生变化。虽然这种方法在收集资料过程中有可能受到评估者主观偏见的影响,但收集资料方法本身不会使正在收集的资料发生变化。文献法的无反应性优势对于评估工作具有特别重要的意义。这是因为,在采用其他调查方法进行评估时,特别是在基层进行评估调查时,被调查者有时对评估者带有偏见、成见或反感,不愿说出真实的情况或想法,甚至故意掩饰、说谎。出现这种状况的原因,一是评估调查时往往有当地项目管理人员陪同,被调查者担心说出真实情况可能会得罪当地项目管理人员;二是被调查者担心评估结果不好会影响进一步的资助。

其次,文献法具有费用低的优势。一般,文献法可以在较少人力和经费的情况下,获得比其他直接调查法更为全面的宏观信息。而一项大规模的评估调查往往成本较高,特别是采用综合对比法进行评估调查时成本更高。文献法的这一优势可以节约评估的成本,这也是文献法使用较为广泛的原因。

最后,文献法适合于进行纵向分析。由于问卷调查、访谈、观察等方法所研究的都是现时的情景,因而往往难以用来进行纵向对比和趋势分析。文献法在这方面具有特别的优势,而历史性的资料有利于评估者更好地作出比较和客观、公正的判断。

当然,文献法也有其缺点,这也是在进行文献法收集资料后

往往还需要进行其他形式的评估调查的原因。文献法的缺点表现在：

首先，许多文献的质量难以保证。文献资料毕竟不是第一手资料，甚至可能是多手资料，在这个过程中可能会出现一些各种各样的偏差，例如记录过程的错误、编码的错误、录入的错误等等。另外，一些历史性资料也可能带有个人的偏见或主观意图，从而影响到文献资料的准确性、客观性。

其次，文献法具有不完全性的特点，难以获得评估所必需的一些基本信息。例如，文献资料很少提供有关态度、感觉等方面的信息，也不可能提供目标群体个人的某些对评估极具价值的信息。特别是对于科普工作评估而言，这一点尤为突出。这是因为，科普活动在很大程度上就是以改变目标群体的态度、认知为目的。因此，文献法很难获得具体项目、活动中目标群体的态度、感觉、知识等方面的变化，而只能通过其他方法进行测度、评估。

最后，许多文献资料缺乏标准化的形式，难于编录和分析。这限制了评估者的定量分析。

二、问卷调查法

问卷法是现代社会调查中最流行的资料收集方法，甚至被美国著名社会学家巴比称为是“社会调查的根本性手段”。在前后对比法、有无对比法、综合对比法中都常常需要进行问卷调查。

(一)问卷的类型和结构

问卷调查法是一种以书面提问的方式调查社会信息的方法，它的形式是一份精心设计的问题表格，用以测量人们的特征、行为和态度等。问卷调查属于标准化调查，即要求所有被调查者按统一的格式回答所有同样的问题。

1. 问卷的类型

问卷依据填答或使用的方式不同,可分为两种主要的类型:即自填问卷和访问问卷。所谓自填问卷是指由被调查者本人填答的问卷,访问问卷则是由访问员根据被调查者的回答填写的问卷。这两种问卷既有联系,又有区别。自填问卷直接面对被调查者,而访问问卷则直接面对访问员。由于这种判别使得它们在具体的形式、设计的方法和要求等方面都存在一定的判别。一般,自填式问卷用于被调查者文化程度较高,对于一些敏感性问题的问卷也适于采用自填式问卷。

2. 问卷的结构

问卷一般包括引言、注释和问项三大部分。

引言

引言应包括调查目的、意义、主要内容、调查的组织者、选样的原则、调查结果的使用者、保密措施等。这样有利于争取被调查者的合作,消除他们的戒备心理。

注释

注释一般是指对填写问卷的具体要求,如填写说明,有时也包括对条款及措辞的进一步解释。例如,“请您勾出一个您认为最恰当的答案”等。注释可以列于条款之后,也可以集中列出。

问项

问项(Item)是问卷的主体,评估调查所需收集的信息正是通过问项来收集的。问项由问题和答案组成。问题可以是一个陈述,也可以是一个提问。这取决于评估调查的目的、方法和被调查者的情况。

问题又可以有两种形式,一种是封闭式问题,一种是开放式问题。封闭式问题由问卷设计人员提供对该问题的若干答案,由应答者在这些答案中进行选择;开放式问题则要求应答者提出自己对这个问题的回答。

封闭式问题是问卷设计中较多采用的形式,它的优势在于便

于计算机加工处理,它的缺点在于评估者事先划定了答案,有可能漏掉一些很重要的答案。而开放式问题在一定程度上可以克服封闭式问题的缺点,答案遗漏较少,但它可能出现答非所问的情况,而且为资料整理带来相当大的困难,不利于计算机处理资料。

(二)问卷设计的步骤与原则

1. 问卷设计的步骤

评估调查的问卷设计可以分为五个步骤:

第一步,确定评估的框架,列出需要评估的指标。这一阶段的主要任务是明确要评估哪些方面,每一方面需要测度哪些指标。

第二步,在明确了所有需要评估的指标后,研究者就可以为测定已经确定的指标编制直接或间接的问题。如果是封闭式问题,还需要给出问题的答案。

第三步,在所有需要测度的评估指标都“问题化”后,接下来的工作就是对问题进行排序、组合。一般排序的原则是,同类组合、先易后难、先次要问题后主要问题、先一般性问题后特殊性问题、先封闭问题后开放问题;

第四步,对问题答案进行编码,然后完善问卷的引言、注释。至此,整个问卷就初步完成。

第五步,在问卷初稿完成后,问卷设计的工作还远没有结束。由于问卷设计的质量直接关系到整个评估工作的好坏,因此,对问卷设计工作必须高度重视。在问卷初稿完成后,还必须对问卷进行试用和修改,甚至是多次反复修改。问卷只有通过试用才能发现问题所在,才能弥补缺陷和遗漏。因此,问卷设计者最好自己亲自参与问卷的试调查或者请相关领域的专家评议。

2. 问卷设计的原则

明确性原则

明确性原则是指问项中的问题以及回答方式的指导语句必须

清楚明白,不能含混不清。不能使被调查人产生不明其义、不得要领、产生歧义等现象。例如,您是否经常参观科技馆?这就属于含混不清的问题。因为很可能问卷设计者在心目中对“经常”这一词有一个标准,而应答者心目中的“经常”就可能与这个标准相去甚远。

单一性原则

所谓单一性原则是指不能要求被调查者用一个答案来回答两个或两个以上的问题。例如,您认为对农村劳动力的科普工作应增加实用技术培训而减少一般性科普知识宣传吗?答案“是”和“否”。这实际就违背了单一性原则。因为这个答案中的“否”是含义不清的。不同的被调查者选择“否”,可能会有不同的含义。如,有的人可能是认为不应增加实用技术培训,但可以减少一般性科普知识宣传,而有的人可能是认为既不应增加实用技术培训,也不应减少一般性科普知识的宣传。

肯定句原则

肯定句原则要求问题尽可能避免使用否定句。因为被调查往往会把句中的否定词漏掉,结果把否定句看成了肯定句,造成答案错乱。例如,您不知道避孕套可以预防艾滋病吗?答案“是”、“否”。这时,被调查者在回答“是”时,他实际的含义是他知道避孕套可以预防艾滋病。

非诱导性原则

诱导性是指由于条款的措词、内容等方面的原因使应答者有意无意地不得不选择某种答案。就是说,这种问项实际上是在诱导应答者选择某种答案。非诱导性原则要求问卷设计者在设计问卷时尽量避免诱导性的问题。例如,“多数人认为科普活动能显著提高人们的科技素养,您认为呢?”

另外,问卷设计中还应尽量避免使用假定性问题,应尽量提被调查者有能力回答的问题。

(三) 问卷调查的优缺点

问卷调查的最大优点在于问卷调查的原始资料很容易转换成数字,特别适于用计算机进行处理和作定量分析。因此,为了强调评估结果的可信性、准确性,评估者往往愿意采用问卷调查法;其次,与访谈法相比,问卷调查具有避免偏见、减少调查误差的优点。这是因为,访谈法中常常由于访问员的不同、访问员提问的方式不同以及被调查者回答问题的语言的不同等原因,产生出各种访问偏见,形成一定的误差;最后,与其他调查方法相比,问卷法具有较好的匿名性。特别是自填式问卷调查比较适合评估一些敏感性问题;第四,问卷法可以节约时间与人力,使问卷调查不受地理条件的限制,到达的空间范围十分宽广。

问卷法最大的缺点是回答率往往难以保证,尤其是邮寄式问卷调查。如果回收率过低,例如回收率在20%以下,可能会导致整个评估调查的失败;其次,不能保证填答问卷的环境和填答的质量。而访谈法则可以控制填答的环境。这一点可能最终会影响到问卷调查的信度和效度;最后,邮寄式问卷调查对被调查者的文化水平有一定的要求。如果目标群体是文化程度较低的弱势群体,例如贫困地区的妇女和儿童,可能就不宜采用邮寄式问卷调查,而适宜采用访谈法、参与式观察法等。

三、访谈法

访谈法又称访问调查法,就是调查人员主要通过与被调查者以口头交谈的方式了解社会信息的方法。因评估调查的目的、性质或对象的不同,访谈法可以分为多种类型。根据一次被访问的人数,可分为个别访谈与焦点小组访谈;根据对访问过程的控制程度,又可以分为结构式访谈、半结构式访谈和无结构式访谈。

(一)个别访谈与焦点小组访谈

个别访谈是指访谈对象是单个人的访谈,按方式可分为电话访谈和面谈。一般来说,选择这样访谈的对象要考虑其代表性、典型性。通常,对项目组织者或目标群体及一些重要信息提供者的个别访谈有助于深入了解有关项目的信息,特别是能更好地了解目标群体的态度、看法和感觉等问卷调查较难准确收集的信息。

焦点小组访谈就是与多人同时进行的访谈,它通过召集一小组特定被调查者开会讨论和交流等方式来进行资料收集。这是一种更省时、更高效的访谈方法,因此在科普评估调查中常常被使用。焦点小组一般由8~12人组成,在一名主持人的引导下对某一主题进行深入讨论。采用焦点小组访谈不仅能做到调查者和被调查者之间的交流,也能做到被调查者之间的交流。群体动力所提供的互动作用是焦点小组访谈成功的关键。使用焦点小组访谈的优点是节省时间,且由于是小团体访谈,易激发被调查者彼此间的互动,这种互动会产生比同样数量的人做单独访谈时所能提供的更多的信息。但焦点小组访谈的缺点则是小团体所谈论的内容并不能完全代表所有的情况,且由于多人讨论,内容难以控制,因此对主持人的要求较高。

(二)结构式访谈、半结构式访谈和无结构式访谈

1. 结构式访谈

结构式访谈又称标准化访谈。即按照评估者统一设计的访问表询问被调查者,并要求调查员以相同的提问方式、记录方式进行的访谈。它是一种对访问过程高度控制的调查,访谈的对象必须按照统一的标准和方法选取,一般采用概率抽样。采用这种访谈形式获取的资料,便于统计和比较分析,而且回收率较高。但这种访谈形式比较呆板,调查者难于临场发挥,被调查者的回答也缺乏

相应的弹性。这种调查形式适宜在调查者对被调查者的一般特点已有一定了解的情况下使用。

结构式访谈使用事先编制好的访问表,调查者按照访问表顺序提问,不能任意改变提问顺序。一般不作解释,如果作解释则要对解释的内容统一作出规定。调查者提问的证据与提问时的态度也须保持一致。

2. 半结构式访谈

半结构式访谈介于结构式访谈与无结构式访谈之间。它使用事先撰写的访谈提纲与主要问题,但具体如何提问则可以根据当时的情境灵活决定。半结构式访谈的提纲一般比较简单。它的优势在于既克服了结构式访谈束缚过多的缺点,同时又保持了结构式访谈便于统计分析的优点。由于这些特征与优势,因此半结构式访谈在科普项目评估中使用得最为普遍。

3. 无结构式访谈

无结构式访谈是一种自由的无控制的漫谈形式的访谈。它事先不预定问卷、表格和提问的标准程序,只给访问员一个大的调查题目,由调查者与被调查者就这个题目自由交谈,调查对象可以较为随便地谈出自己的意见和感受。调查者事先虽有一个大概的要点,但所提问题是在访问过程中边谈边形成的,随时提出的。

与结构式访谈相比,无结构式访谈最大的优点是弹性比较大,能充分发挥访问者的主观能动性、积极性;其次,由于不存在过多的提问限制,访问者可以顺着被调查者提及的问题,进行深入的调查。甚至可能由此得到比预想还要有意义的信息、资料。而这一点是结构式访谈所不具备的;最后,在无结构式访谈中,访问者能对问题作全面、深入的了解。然而,无结构式访谈也存在一些缺点。首先,无结构式访谈比较费时,从而使调查的规模受到很大的限制;其次,无结构式访谈无法进行定量分析,访问的结果往往取决于访问者的技巧、素质、经验,因此无结构式访谈对访问员的要

求较高。当调查员是一些经验不丰富或初次接触调查的非专业人员时,这一方法也不宜采用。

(三)访谈法的步骤与原则

1. 访谈法的步骤

以评估为目的的访谈法可以分为四个步骤:

第一步,访谈前的准备工作,如查看与评估任务有关的文献资料等;

第二步,明确任务。即评估者明确通过访谈需要了解哪些信息、资料。如果是采用结构式访谈,评估者还需要制定访谈的提纲,包括访谈的目的与要求、访谈的题目、访谈的纲目等;

第三步,进入访谈。进入访谈首先需要选择访问对象,最好能选择那些具有典型代表性的被访问者;其次向被访问者作自我介绍,说明目的,请求被访问者的配合,然后进入正式的访谈;

第四步,提问与记录。在进入正式访谈后,评估者就需要提出问题请被访问者回答,并记录对评估有用的信息资料。在结束访谈时要对被访问者的合作表示感谢。

2. 访谈法的原则

访谈调查过程中,需要遵守以下原则:

第一,客观公正的原则。评估者在访谈过程中要避免说出自己的价值判断和倾向性,以免被访问者受到诱导并附和评估者的观点。应尽量避免其他人介入访谈,尤其应避免项目管理人员介入访谈。这是因为,项目管理人员的观点往往会左右被访问者的观点,甚至使被访问者不敢说出真实的想法或事实;

第二,访问控制的原则。在访谈过程中,评估者需要把握访谈的进程,控制访谈的局面。当出现被访者不真诚合作的情况时,要设法设置某种情境,使对方有所反应,以迅速掌握访谈的主动;当对方偏离主题时,也要进行引导性提问,使他回到原来的主题。但

切忌粗鲁打断对方的谈话,以免对方难堪,从而产生抵触情绪。评估者控制访谈的主要手段包括提问和表情动作;

第三,言简意赅的原则。在访谈过程中,评估者所提问题需要言简意赅,不要不着边际。每个问题所涉及的范围不宜过宽,意思要明确。另外,在提问时,最好不要同时提出多个问题,以免被调查人在回答了其中一个问题后忘了其他问题。

(四)访谈法的优缺点

访谈法的优点在于能灵活机动地进行调查工作,调查环境与条款顺序可控。首先,它既可以提高被调查者答复的兴趣,可以达到很高的回收率,也可以限定某一特定的人员回答问题,增加回答问题的针对性;其次,访谈法可以使调查的内容、范围比较广泛。它既可以收集访谈提纲中的各种问题,同时也可以收集访谈提纲以外的资料;最后,它能适用于各种调查对象。不仅可以对文化程度高的人进行调查,也可以对文盲、半文盲人口进行调查。因此,在科普评估调查中,访谈法应用得比较广泛。

访谈法的缺点在于:首先,访谈法容易出现误差和偏见。它既受访问者的限制,也受被访问者的限制。对访问者来说,访问人员的素质、技巧、态度和经验直接影响到资料收集的结果。对被访谈者来说,其合作态度、回答能力的差异对提供材料的质量有很大的影响;其次,访谈法的匿名保证性较差,对一些敏感性问题难以获得真实性资料;最后,访谈法较为费时、费力,也费钱。

四、观察法

观察法是一种收集原始资料与信息的方法,也是评估调查中常用的方法。观察法既可以单独进行,也可以与其他调查方法配合使用,共同完成收集评估信息的任务。

(一)观察法的涵义与类型

观察法是直接感知与记录正在发生的一切同研究对象与研究目的有关的社会事实的一种调查方法。这种方法要求调查者使用眼睛注视特定的社会现象与环境因素,通过听觉等手段收集被调查者知识、行为与态度等信息。

由于研究目的不同、所需资料的性质不同,采取的观察形式也往往有很大的区别。从研究者能否对观察对象实行控制来划分,观察法可分为实验观察与实地观察;从研究者的观察程序来划分,可分为结构式观察和无结构式观察;从研究者在观察中扮演的角色划分,可分为参与观察和非参与观察。

1. 实验观察与实地观察

实验观察法是通过做社会实验的方式获取信息资料的方法。它是在人为设置的环境下进行社会观察。一般事前对观察对象经过一定程度的控制。实验观察通常是为了了解人们某些具体、细微的行为特征而进行的观察。它的优点在于能够深入、细致地分析事物的因果关系,找出事物成功与失败的缘由,并常常能够获得一些其他观察所不能发现的结论;它的缺点在于所耗时间、人力、财力较大,所获信息在很大程度上取决于能否有效控制实验过程。

实地观察是在自然的环境下进行的观察,这种观察不需要对观察场所和观察对象进行控制,而是深入到现实生活中对实际发生的现象进行的观察。它的优点在于观察结果的直观性、可靠性;它的缺点在于获得信息的偶然性和表面性,而且受时空等条件的限制。实地观察是科普评估中经常使用的方法。通常,评估者可以通过自己实地的观察了解许多重要的、真实的信息。

2. 结构式观察与无结构式观察

结构式观察是根据事先设计好的观察内容、观察项目,按照

制定好的观察表或观察卡，严格依据计划实施的观察。结构式观察往往事先对观察的范畴详细分类，对各项内容的观察和记录方法逐一规定。因此，结构式观察常将注意力集中到若干具体的、明确的、许多还是可以计数的行为和特征上。

无结构式观察是不预先设计观察内容、观察项目，也不专门探究某一行为，而是观察和记录在研究对象本身及其周围所发生的一切事件。无结构观察又可分为无结构参与观察和无结构非参与观察。

结构式观察的优点在于可进行定量分析，缺点在于缺乏弹性，了解的信息有限、不深入。无结构式观察与结构式观察正好相反。它的优点在于能发挥评估者的能动性，调查深入，信息量大；而缺点在于无法进行定量分析，评估结果受评估者个人因素的影响较大。

3. 参与观察与非参与观察

参与观察是指研究者参加到被研究对象的活动中进行观察的方法。这类观察又可以采取两种形式，一种是隐蔽观察，即参加到各种活动之中却不暴露自己的观察身份，目的是不影响被观察者的行为和语言；另一种是非隐蔽性的参与观察，这种观察是对被研究对象公开自己的身份，并通过对方的合作与支持尽可能充分掌握资料的方法。参与观察法的优点在于所获得的资料信息真实、准确；其次，评估者通过参与观察可以亲身体验项目影响的全过程；最后，评估者有可能真正了解最需要服务的弱势人群，了解未被满足的需求。参与观察法的缺点在于，所耗时间很长、经费较多，而且评估者要真正融入项目活动区也不是一件容易的事情。

非参与观察又叫局外观察，它是指研究者不介入被研究对象的活动中，而是作为局外人，通过观察来收集所需资料的方法。它的优点在于评估者的观察活动不会引起被观察者的各种反应，从

而保证评估结果的客观性 ;它的缺点在于无法体验和了解被观察者的心理活动和态度。

(二)观察法的步骤和原则

1. 观察的步骤

观察的全过程可以分为三个阶段 :

第一阶段是准备阶段。在准备阶段 ,评估者需要确定评估的任务和制定观察计划 ;

第二阶段是实施阶段。在实施阶段 ,评估者要进入观察场地 ,与观察对象交往 ,进行观察并做出记录 ;

第三阶段是资料处理阶段。在资料处理阶段 ,评估者需要整理和分析观察记录 ,最后提交评估报告。

2. 观察的原则

运用观察法进行评估调查与运用其他方法进行评估调查的目的都是为了收集到真实可信的资料 ,然后通过对资料的科学分析得到评估结论。然而 ,观察法的特殊之处在于它主要是通过眼观、手记。为此 ,在采用观察法时应注意以下原则与要求 :

第一 ,客观、公正。由于观察法主要通过评估者的观察获取资料 ,评估结果很容易受评估者个人立场、观点的影响 ,因此采用观察法必须防止先入为主的观念 ,保证调查的客观、公正。

第二 ,观察准确、全面。观察法依靠的是评估者的感知 ,因此观察结果出现偏差的可能性极大。这种偏差包括两个方面 ,一方面是评估者对观察的标志、范畴、量度不清晰所带来的偏差 ;另一方面是评估者可能只观察到现象的局部 ,却以偏概全所带来的偏差。这就要求评估者在采用观察法时需要特别注意观察的全面、准确。

第三 ,及时、详细记录观察的信息。通常 ,观察的结果很容易出现部分遗忘的现象 ,因此需要及时记录。如果记录不会引起被

观察者的注意力,也不会引起被观察者的不满或戒备,最好当场记录为宜。也可以采用录音等形式帮助记录。因为当场记录不容易丢失信息。如果当场记录会影响情境,引起被观察者的不满或戒备,就需要进行事后及时、详细的记录。

(三) 观察法的优缺点

观察法的优点在于:首先,它可以当时当地观察到现象或行为的发生,从而掌握第一手资料。观察不仅可以把握整个现场情况,而且还可以感受到情境和气氛,得到可靠的信息;其次,观察一般是在自然环境中进行,因此,它对评估对象的干扰较少,尤其是在非参与观察中,因而可以得到自然条件下真实的资料,这对评估者来说是极为难得的;最后,观察法特别适合研究无语言文字沟通的调查对象,例如对少数民族、聋哑人项目的评估;最后,观察法可以弥补其他方法的缺陷。在社会调查中,经常遇到调查对象不愿意接受访谈,拒绝回答,即使回答也是虚报或因记忆不清而产生较大的误差,而观察法则可以避免以上局限。

观察法的缺点在于:首先,大多数观察中,调查者对于环境因素难以控制。为得到真实可靠的信息,调查人员有时只能消极等待;其次,观察法只有较少的指标可以量化,而大量的观察资料只能采用定性的分析;最后,采用观察法时,一般成本较高,不可能进行大规模的样本观察。

第三节 评估分析方法

同其他领域的评估一样,科普评估中也存在定量与定性两种分析研究方法。定性与定量方法各有其适用范围,互有短长,将二者结合起来,可对科普现象作出客观全面的评估。

一、定量分析

定量分析也称量化分析,就是采用数学的方法,对评估对象可以量化部分从数量方面进行描述、分析和评价,取得数量化的分析结果。定量分析有一套完备的操作技术,包括抽样方法(如随机抽样、分层抽样、系统抽样、整群抽样)、资料收集方法(如问卷法、实验法)、统计方法(如描述性统计、推理性统计)以及模糊数学方法。定量分析的资料比较客观可靠,统计分析比较科学准确,具有较高的客观性和可靠性,缺点是难以研究科普过程中大量存在的不可量化的主观现象。

对比法是科普评估活动中最常用的定量分析方法之一。通常,当我们通过问卷调查、访谈等方式得到了相关指标的调查数据后,往往还不能判断这一指标值的高低、好坏,还需要有一个参照对象,通过比较各类数值才能作出判断,得出评估结论。对比法就是通过比较发现差异与成效的方法。对比法有很多类型,其中主要有:前后对比法、有无对比法和综合对比法等。

1. 前后对比法

前后对比法即将项目实施前的情况与项目完成后的情况进行对比,以评估项目效果的方法。其公式为:

$$P = I_2 - I_1$$

其中, P 代表项目效果, I_2 代表项目完成后的情况(即项目组后测值), I_1 代表项目实施前的情况(即项目组前测值)。

例如,为了解科技活动对幼儿素质的影响,天津幼儿师范学校课题组在1998—2000年间选择北京、天津和广东省的35家幼儿园作为测试实验基地,评估幼儿园科技活动对提高幼儿观察力、想象力、主动性和好奇心的影响。该课题组采用了前后对比法。首先,在幼儿园还未开展科技活动前,对幼儿观察力、想象力、主动性和好奇心进行测试,了解幼儿实验前的素质水平。然后,在开展

相应的科技活动后,再对幼儿进行测试,通过比较科技活动前与科技活动后幼儿心理素质的变化评估科技活动对提高幼儿素质的效果。

前后对比法看上去非常简单,但在实际操作时,也需要注意一些问题。

第一,缺乏前测值。科普机构由于缺乏资金,因此在实施项目前往往没有进行前测,而在项目完成后,为了评估项目的效果才进行后测。此时由于缺乏对比的基线,无法进行判断,从而为评估工作带来了困难和麻烦。这也是今后科普机构评估中可能遇到的一个问题。因此,这就需要督促科普组织在进行一些重要的大型科普活动中注重前测或基线调查工作。

第二,前后对比法本身存在的局限。由于影响项目结果的因素往往较为复杂,除项目实施本身会影响结果外,常常还有许多项目以外的因素也会影响到项目完成的结果。也就是说后测与前测值之差可能并不完全归因于项目实施的结果。例如,电视科普系列片的播放都需要较长的时间期,假如B地的学生2002年开始看该系列片,在看该系列片之前的相关科技知识水平得分为4分(前测值,10分量表)。看节目一年后,2003年的相关科技知识水平得分为7分(后测值,10分量表),因此,根据前后对比法,可以得出评估结论,该系列片的播放使得B地学生的相关科技知识水平上升了3个分值(后测值减去前测值)。然而,B地学生的相关科技知识水平的提高可能还与学生的其他课外科技学习经历有关(如看科普读物,参观科技馆等),因此,从2002年到2003年B地学生相关科技知识水平的提高并不一定完全是看该电视科普系列片的结果。而要解决这一个问题,则可以采用有无对比法。

2. 有无对比法

有无对比法是选定一个与项目组近似的但没有实施项目的对照组(即控制组),通过项目组实施项目的结果与没有实施项目的

控制组结果进行对比,以评估项目效果的方法。其公式为:

$$P = I_2 - C_2$$

其中, P 代表项目效果, I_2 代表项目完成后的情况(项目组后测值), C_2 代表控制组同期的情况(控制组后测值)。

仍以某电视科普系列片的评估为例。如果影响学生相关科技知识水平的因素除了该系列片之外,还可能还有其他多种因素,那么要评估该地学生观看该电视科普系列片的实际效果,采用前后对比法可能会有较大误差,因此需要进行有无对比。首先选择一个控制组,即其他条件与 B 地相近,但没有播放该系列片的 C 地。假定 B 地自 2002 年播放该节目后,2003 年看节目的学生的相关科技知识得分为 7 分,而控制组虽然没有观看该节目,但由于一年来学生的科技认知水平的提高,2003 年学生的相关科技知识得分为 5 分,因此,可以认为, B 地播放该电视科普系列片的实际效果使当地看节目的学生的科技知识水平上升了 2 个分值。

可见,采用有无对比法和采用前后对比法进行评估,结果有时相差很大。因此,在评估时,要特别慎重选择评估的方法。在项目的结果一般只受项目因素影响,而很少受到其他非项目的外部因素影响的情况下,可以选择前后对比方法;在项目的结果除受项目因素影响外,还受许多非项目的外部因素影响的情况下,适宜于采用有无对比方法。

由于有无对比法需要同时对试验组和控制组进行测度,往往会增加评估的费用与评估的时间,特别是当控制组与试验组距离较远时。因此,在评估经费有限,预定的评估时间较短时,适宜于采用前后对比的方法。另外,由于有无对比法不需要前测值(前提条件是假定试验组和控制组的基线值相同),因此,在缺乏前测值的情况下,可以采用有无对比的方法。

然而,有无对比方法也有一个较大的局限,即它假定试验组和控制组的基线值相同,而实际生活中,这一假定条件往往并不存

在。尤其是当试验组和控制组基线值相差较大时,采用有无对比方法有较大的误差。这时,则可以采用另一种对比法,即综合对比法。

3. 综合对比法

综合对比法是通过比较项目组前后测之差与控制组前后测之差以评估项目效果的方法。其公式为:

$$P = (I_2 - I_1) - (C_2 - C_1)$$

其中, P 代表项目效果, I_2 代表项目组后测值, I_1 代表项目组前测值, C_2 代表控制组后测值, C_1 代表控制组前测值。

仍以某电视科普系列片的评估为例。如果试验组 B 地和没有播放该节目的控制组 C 地起始条件并不相同。B 地学生 2002 年的相关科技知识水平为 4 分,而 C 地学生 2002 年的相关科技知识水平为 3.5 分,因此,根据综合对比法的计算:

$$P = (I_2 - I_1) - (C_2 - C_1)$$

$$P = (7 - 4) - (5 - 3.5)$$

$$P = 1.5$$

即 B 地播放某电视科普系列片的真正效果是 1 年来使 B 地学生的相关科技知识水平提高了 1.5 个分值。

需要注意的是,虽然综合对比法相对准确,但由于综合对比法既需要项目实施组与控制组的前测值,也需要项目实施组与控制组的后测值,对评估经费与评估时间的要求更高。这对于科普组织来说无疑是较大的负担。因此,在科普项目的实际评估过程中,综合对比法并未见使用。但为了使评估尽量准确,一些重大科普项目的评估常常是前后对比法与有无对比法并用,例如美国科普系列片“比尔·奈科学人”的评估就同时采用了这两种对比方法。

二、定性分析

定性分析是相对于定量分析的一种分析研究方法。所谓定性

分析 就是采用非数量化方法 ,对评估对象的不可量化的因素从性质的角度进行描述、分析和评价 ,做出定性的结论。定性分析能够处理科普过程中那些难以用量化方法描述的因素 ,从而从总体上掌握研究对象性质的基本情况或确定引起某一现象变化的原因及过程。

科普评估强调运用定性分析方法的原因在于 :第一 ,科普活动的目标是促成人的学习 ,它是以人为主体的 ,而人的心理活动和行为极为复杂 ,无法完全以自然科学的定量实证方法予以解析。第二 ,科普活动的因果关系不像自然科学那样有明确的界定。第三 ,定量方法来自于稳定性、一致性高的自然科学领域 ,而对变动性大又存在异质性的科普领域则不很适用。第四 ,科普活动除了要增进人们的科技知识 ,更重要的是要提高人们的科技兴趣、态度、价值行为上的素养 ,这些非智力因素很难进行定量测量 ,大多需要采用定性的方法。第五 ,科普评估的目的决定了定性分析的必要性。科普评估主要是为了促进科普工作的改良和发展。一般而言 ,定量分析对于划分名次等级、排队鉴定来说 ,具有可比性强的优点 ,假如评估的目的不是进行相互比较的话 ,定性的分析结果更能给评估对象有价值的反馈。

定性分析最鲜明的特点是 :①对特定问题的研究能达到相当的深度 ,能抓住问题的本质 ;②信息更真实、生动和相近 ,尤其是人们主观性的信息(如兴趣、动机、满意度等) ;③能发现和界定科普过程中复杂模糊的问题和现象。当然 ,定性分析也有其局限性 ,最大的局限性是主观性较强 ,对评估者的要求相当高 ,所以其效度受到一定的影响。为此 ,评估者在进行评估的过程中 ,对一些定性的评估问题 ,很多时候会借助数学模糊论的方法 ,将其数量化 ,以求得到较为科学客观的评估结论。最常见的是采用语义差异法 ,将定性指标量化并赋值。例如 ,将某一定性指标的测量按 5 分量表分成 1~5 个档次(好、较好、一般、较差、差) ,并对每个档次内容所

反映的指标的趋向程度提出明确、具体的要求,每档根据指标内容的趋向程度给5~1分的分数,即第一档(好)对应指标分数为5分,第二档(较好)为4分,第三档(一般)为3分,第四档(较差)为2分,第五档(差)为1分等。这样,在对被调查者进行调查后就会得到一组量化的测量结果。这种分数在方法学中称为“虚拟量”,而借助虚拟量的测量,称为间接测量。由于虚拟量乃人工量化的结果,并无严格的等距可测性和理论定位可测性,由此获得的结果还是存在一定的信度和效度问题。

提高定性分析可靠性和有效性的另一方面是采用三角互证法,即在研究某一问题时使用不同来源的资料或用多种方法收集资料,而后比较不同来源的信息,以确定它们是否相互证实。

定性方法与定量方法在功能和分析方法上存在差异,它们在数据收集方法上也各有倚重。另外,定性分析的数据抽样也不同于定量分析中的数据抽样。定量分析的样本要能全面准确地代表总体,而定性分析的抽样重在典型,重在有分析的余地,能提供很多深层信息,说明很多问题。常见的定性抽样方法有:极端抽样、普通抽样、配额抽样、滚雪球式抽样。

定量分析与定性分析的区别是显而易见的,但它们之间的关系并非矛盾对立、水火不容,而应该是互为补充、互相支持的。在科普评估中,常常是两种方法结合使用,主要表现为以下几种方式:①对定量资料给出定性的解释,比如在进行电视科普节目收视率调查时,评估人员可以进一步了解观众的收视动机和行为,由此获得对节目收视状况更丰富、更深入的了解;②对某些定性资料进行量化分析。比如,在调查博物馆观众对展览的教育性、趣味性等的看法时,可以让观众按照语义差异量表打分,获得一种数量化的评估结果;③将定性分析作为形成假设的途径,进而通过定量实验来检验这种假设;④同时收集定性与定量资料,两种方法互补使用。

因此,正确地运用定量分析与定性分析,以多元的方法进行科普评估,应该成为我们的基本原则。

第四节 评估调查分析中应注意的问题

在评估的过程中,评估人员应秉持评估的意义。评估是应用研究,评估的最终目的是结果能得以“应用”;评估是一个过程,它是科普过程的一系列反馈环;评估有助于决策的科学化;评估有助于促进组织“省思”和学习。

科普评估人员必须走出评估的误区:

- 评估不是统计,不是给项目或机构评判等级优劣。评估的主要目的不是证明,而在改进。
- 评估没有所谓“正确的”方法可循。采用何种评估方法完全取决于你评估的目的是什么、评估结果意欲何为以及所能获得的评估资源等因素,要视具体的情况酌定。
- 设计评估的过程中,艺术性与科学性同样重要,要富于创造性!
- 尽可能全面获取调查数据,但不一定需要大样本量。

科普评估的目的决定了其指标数据收集要尽可能全面,这样才能够得出具有普遍意义、相对准确的结论。对于“全面”通常存在一种误解,认为大样本是使结果具有代表性和准确的最好办法,其实并非如此。较大的样本的确能减小抽样误差(即样本的推论统计和实际母体群统计之间的误差),然而相比其他两种影响抽样结果的因素:抽样偏差和响应偏差,抽样误差是影响全面性最小的因素。抽样偏差是指由于所选样本群中有些样本受某种因素的影响,不能提供样本信息而产生的统计误差。而响应偏差是指回答不能反映回答者的真实态度和行为,如错误理解问题或刻意隐瞒事实时所造成的误差。抽样偏差和响应偏差更能够影响结果的准

确性。因此,在资源(经费、时间等)一定的情况下,应优先考虑尽量减小抽样偏差和响应偏差,而不是选择更多的样本空间。

要减小抽样偏差,需要从时间、地点多方面仔细构建样本空间,尽量不遗漏样本单元。在这方面,国外一些科普项目的评估做得很好。比如英国 Education Extra 基金会为“地铁中的科学”所作的评估,评估小组采用观察法来考察地铁乘客对车厢内卡通科普招贴画的兴趣。他们通过精心设计观察时间段,构建了比较全面的样本空间,选择工作日上下班时间定位普通的地铁上班族,选择放学时间来了解儿童的反应,还考虑到周末乘客构成的改变而选择周末时间进行观察。美国儿童科普系列片“比尔·奈科学人”的评估中,为获得“该节目是否改变了孩子们的科学态度和科学意识?”的有关数据,评估人员除了直接从孩子那儿获得数据外,还将教师、家长都纳入了样本空间。要减小响应偏差则需要在数据收集方法的使用上下功夫,精心设计问题防止误解和误导、选择训练有素的访谈者都是有效的途径。

主要结论与建议

科普评估是一个复杂的系统工程,我们所给出的评估模式和指标框架是科普评估科学化进程中的初步尝试,还需要经过未来评估实践的检验。以下是我们在进行科普评估研究过程中所形成的一些结论和建议,希望能够启发思路,从而推动我国科普评估制度化和科学化的进程。

一、科普评估应淡化鉴定、评比目的,突出改进发展作用

传统的项目评估仅仅关注项目是否达到了预定的目标或是否产生了预期的效果,以此判断项目是成功还是失败。然而,近年来,国外评估界提出了更先进的评估理念,就是不仅要关注项目是否产生了效果,而且要分析这些效果是如何产生的,诊断项目还有哪些不足的地方,为什么存在这些不足,并提出相应的改进措施。对科普管理者来说,这样的评估无疑有价值得多,因为它能帮助项目管理和资助机构掌握已开展科普项目的经验教训,将其引介给其他科普项目执行机构,使未来的科普工作日臻完善,这对确保科普事业的持续发展是非常重要的。在此,我们有必要强调,科普评估是一种经了解、评绩效、明得失、找原因、寻改进、再发展的循环过程,科普评估能完成的最大贡献是确定科普工作需要改进的方面,提高科普工作的效果和质量。正如评估专家斯塔弗尔比姆所说,“评估最重要的意图不是证明,而是改进”。科普评估应为科普事业的可持续发展服务。

由此引伸一步,评估的目的若是为了诊断得失以谋求改进的话,就应淡化评估的鉴定、分等功能,在评估时不应有评比打分,且将评估结果的优劣等第作为未来提供资助的依据。这种矛盾情形最易引起评估者进退失据,甚至引发评估者与受评者的争端。

二、提倡互动、参与式评估

所谓互动、参与式评估,就是吸收受评对象(如科普项目执行者)、利害相关人参与评估过程。以往的评估基本是自上而下的,评估者完全以一种权威的方式出现在评估对象面前,所有的主动权都掌握在评估者手里,受评对象只有被动地接受评估,没有任何的发言权。这样的评估由于忽视了受评对象的主体性,忽视了人的价值观的多元性,其结果往往难以为受评对象所接受,最终导致通过评估改进缺失的期望落空。

在科普领域提倡参与式评估,就是要打破传统的“自上而下”的评估方式,吸纳受评的科普机构、科普受众参与到评估中来,其好处体现在:①受评机构参与评估有助于扩大评估设计的焦点及范围,降低评估中不实际、不公平等问题,提高评估工作的质量。这是因为,受评机构作为科普项目的执行者,对科普项目的情况了解全面而又深透。他们除了了解项目的现状和存在的问题外,还知道这种状况是如何形成的,了解这些问题的解决受到哪些条件的制约。这就是说,他们不仅了解存在的问题,而且了解问题的根源与症结所在。这些信息对于形成科学正确的评估指标和结论是必不可少的。②让科普项目的受众(比如青少年)参与评估,可以通过他们的切身体验、从科普受众的角度反映科普项目的效果和问题,评价科普活动是否满足了他们的需要。③科普项目五花八门,外部评估专家在评估一个项目时,对项目的情况不可能很了解,常常以容易研究的问题为主,只利用自己熟知的评估工具、根据一种共同的模式来评估不同的科普项目,具体科普项目的特色

较难在预定的指标体系中得到反映,由此极容易形成片面的评估结果。而要克服这一缺陷,也需要尽可能地发动受评对象、目标受众参加评估活动。④吸收受评对象参与评估,有助于评估结果的回馈利用。从以往实施的专家评估的情况看,受评对象对外部专家评估或担心抵触,或敷衍塞责,总认为自己工作的得失会被挖出来,从而对自身的未来发展造成影响。在这种情绪下,专家评估的结论和建议很多时候不能得到他们的真正认可和接受。我们开展科普评估的最终目的是为了改进工作,推动科普事业的健康发展。科普评估的这一目的能否达到,在很大程度上取决于各种科普执行机构是否理解评估活动,是否接受评估结论,能否在他们中间产生积极的心理效应。科普评估如果没有科普执行机构的积极参与,是很难达到预定目的的。

三、科普评估应定性、定量结合

评估一词代表了品质与价值,它可有质与量两种向度,故应从定性和定量两方面共同把握。从我们所建立的科普评估指标框架可以看出,科普评估采取的是定性和定量结合的方法。因为绝大部分的科普项目都是发生在复杂的社会环境中,是无法严格控制 and 限制的多因素环境,而且科普的对象是人,科普评估涉及许多“人”的方面,存在着大量不可量化的主观现象。人的需要、人的学习兴趣、个体的经验和主观认识、个体态度、价值观念等的变化都是无法进行准确量化的。另外我们前面也提到,实施科普评估的主要目的是寻找不足,谋求改进,这一作用也只有借助定性分析才能实现。当然,定性评估有其局限性,最大的不足就是缺乏客观性。因此,在具体的评估中,对那些能比较科学地量化的指标,还是尽量用定量的方法去评价,但决不能为了追求所谓的客观性、可比性,而将许多在目前尚无法量化的被评估属性硬是量化。

事实上,研究美国和英国的科普项目评估实践可以发现,国外

科普项目的评估多是综合地运用定性和定量两种方法,他们主张采用“三角互证法”来研究复杂的科普现象,即用不同的方法来研究同一现象,以此来增加结果的可靠性。一种典型的评估计划是:以定性研究方法如焦点小组座谈作为开端,在座谈中发掘用于问卷调查的问题,问卷调查后进行深度访谈来进一步澄清和深化问卷调查的某些结果。定性和定量方法结合使用的另一个原因在于,当不同的方法所获得的数据之间不一致或发生分歧时,它提供了一个对数据收集和分析过程进行省思的机会。

四、大力推进科普评估制度建设,开创我国科普管理新局面

在科普领域建立评估制度有助于科普管理部门加强对科普工作质量的宏观管理,促使科普机构追求卓越,注重工作成效,不断改善科普工作的质量。对科普管理部门而言,科普评估是一项重大的管理制度,并有其政策目的。对科普机构而言,是不曾经历的巨大改变。不论这一制度的推进方式、未来发展方向以及科普机构对此的了解与情绪等问题,在推行这一制度前,都需要有所谋划。为求评估制度得以顺利推行,我国的科普主管部门应于实施前完成一系列的准备工作,并在实施中辅以必要的扶持和管理。结合我国科技和教育评估管理的经验,我们认为近期应从以下几大方面着手:

第一,将科普评估工作列入我国科普事业发展纲要,尽快建立健全的评估管理体制并形成工作网络,使评估管理制度化。建议成立一个跨部门的国家科普监督评估办公室,作为政府行使科普评估职能的专门机构,负责制定并落实评估政策、评估规范、评估的跟踪机制、后设评估(对评估的再评估)机制等;研究、制定评估工作计划,制订评估指南,为各类评估的开展提供指导和技术支持,推动科普评估理论和方法的研究,促进学术交流和评估工作人员培训。

第二,制定有关科普评估的政策和条例,开展好科普评估的教育宣传活动。建立完善科普评估的相关政策条例,是对科普评估进行有效管理的需要,是实现政府宏观指导与管理的行为体现。应尽快制定科普评估的发展规划、实施步骤和措施,制定颁布《科普评估管理办法》、《科普评估管理办法的实施细则》等管理规范条例,对科普评估者行为、所承担的义务、责任作出明确的规定。

同时,应做好科普评估的宣传推广工作。有鉴于国内还缺乏评估的社会文化,整个科普界对于科普评估的理念与目的、内容与范围、方法与程序、评估效用等问题,还没有基本的认识,故推行科普评估会面临很大的挑战。科普管理部门应加强对科普机构评估观念的宣传与沟通,使科普机构具有正确的评估观念,赢得他们对评估的普遍重视和支持,以期在各方共识下,科普评估工作得以顺利有效地开展。

第三,科普评估制度的建立应采取逐步推进的方式,由小范围到大范围逐步实施。先由某些重大的科普项目或条件比较好的科普机构进行试点,再逐步普及到较大范围,最终使评估与机构拨款和项目资助挂钩。在推行外部专家评估、访视评估的同时,要鼓励科普机构开展自我评估,逐步形成自我评估与外部评估相结合的评估制度。每3~5年对科普基地进行一轮考察评估,切实解决科普工作中存在的问题。在推进的过程中,可以采取“规划—执行—检查—修正行动”(PDCA)的过程模式,边做边修正,使科普评估管理日臻完善。要建立评估结果的反馈答复和跟踪机制,监督评估报告中提出的建议的落实状况,采取措施,保障评估结果能得到切实有效的利用。同时,要重视科普评估的数据库和网站建设,对完成的评估报告以及评估中积累的经验教训进行制度化储存,为今后的评估实践及评估研究提供借鉴经验。

第四,着力发展和培养科普评估专业机构和人才,逐步建立评估机构人才库。评估是一项专业化很高的工作,所需的知识与方

法十分广泛,如果由缺乏科普评估训练的人员或自认为是专家的外行人员深入参与科普评估工作,很可能导致错误的评估结果。为保证评估的质量,应使评估人员和评估机构处于一种开放的、竞争的制度环境中,并形成评估专家的管理机制,促进专家的公正性和责任感。从教育领域的评估实践来看,评估人员大多以教育专家和研究者为主,这些评估人员虽然具有社会声望与地位,但与专业评估仍有差距,以致评估的公正性受到质疑。而另一方面,如果由非科普领域的其他领域的评估人员进行科普评估,也必然会因领域的隔阂不能有效地做好科普评估。因此,在科普领域,一方面要加强对科普专业评估人员的培养培训,另一方面要探讨发展独立的科普评估专业机构的可能性。

总之,在科普领域引入评估机制势在必行。它有利于提高科普执行机构的责任心和工作成效,有利于实现科普资源的优化配置,有利于提高科普管理的科学性。引入这一制度,需要政府科普主管部门的积极推动以及科普机构的重视和协同努力。

参 考 文 献

- 1 The 2002 User-Friendly Handbook for Project Evaluation. NSF
- 2 FOOTPRINTS : Strategies for Non-Traditional Program Evaluation. NSF
- 3 User-Friendly Handbook for Mixed Method Evaluations. NSF
- 4 W.K. Kellogg Foundation Evaluation Handbook
- 5 A Guide to Programme Evaluation. Evaluation Associates Ltd. Dec. 1997
- 6 John Falk & Lynn Dierking. Evaluation : A Checklist
- 7 The Program Evaluation Standards. <http://www.eval.org/>
- 8 So did it work ? Evaluating public understanding of science events. COPUS , UK
- 9 COPUS : The Committee on the Public Understanding of Science , An Evaluation of Schemes and Projects. Evaluation Associates Ltd. Dec. 1997
- 10 Innovating and Evaluating Science Education : NSF Evaluation Forums , 1992-1994.
- 11 http://www.nsf.gov/pubs/1995/nsf95162/nsf_ef.pdf
- 12 NASA Education Evaluation History
http://ehb2.gsfc.nasa.gov/edcats/on_line_reports/History98v95.ppt
- 13 Implementing the Office of Space Education / Public Outreach Strategy
http://spacescience.nasa.gov/education/resources/evaluation/OSS_EPO_Task_Force_Report.pdf
- 14 Participatory Evaluation. The Evaluation Exchange , Volume I , No. 3 / 4 , Fall 1995
- 15 Jessica Luke. What is Participatory Evaluation
- 16 Robert E. Stake. Program Evaluation : Particularity Responsive Evaluation.
- 17 Guba and Lincoln. Fourth Generation Evaluation

- 18 Simon Zadek. Beyond Fourth Generation Evaluation
- 19 Stuart Naylor and Brenda Keogh. Science on the Underground : an initial evaluation. *Public Understanding of Science* , 1999. 8 , 102-122
- 20 UNDP. 面向结果的监督与评估. 北京 科学出版社 , 1999
- 21 李雁冰. 质性课程评价 , 从理论到实践(一). 上海教育 , 2001(11)
- 22 张辉华. 西方主要教育评价模式述评. *福建高教研究* , 2002(3)
- 23 王景英 , 梁红梅. 后现代主义对教育评价研究的启示
- 24 教育目标的分类. [http ://jky.zjnu.net.cn](http://jky.zjnu.net.cn)
- 25 苏孝同. 对中专教育评估的思考
- 26 林之达 , 郭晴. 论传播效果的层级性
- 27 Courtney Abrams and Lynn D. Dierking. Summative Evaluation for SciTech Clubs for Girls. December 1996
- 28 Kathleen McLean. Planning for People in Museum Exhibitions. ASTC
- 29 Cynthia A. Char. Evaluation of the “Science in the Stacks” : A Museum-Library Collaboration to Creat Traveling Science Exhibits for Libraries.
- 30 John H. Falk. On Beyond Now : Strategies for Assessing the Long Term Impact of Museum Experiences
- 31 Randi Korn. Studying Your Visitors : Where to Begin
- 32 Lynn Dierking and Wendy Pollock. Questioning Assumptions : An Introduction to Front-end Studies in Museum , 1998 , ASTC
- 33 Marianna Adams. How to Use Focus Groups for Evaluation.
- 34 井岛真知. 博物馆の评环境を考える
- 35 U. S. science and technology museums. *IEEE Spectrum*
- 36 Kimberly Zern. Handheld Computers and Learning in the Informal Museum Setting Evaluation Report
- 37 National Aeronautics and Space Administration. Office of Space Science : Education /Public Outreach Interim Evaluation Report , Oct. 2002
- 38 NASA /Gulf Coast Education Initiative Formative Evaluation Report
- 39 SRI International. Globe Year 6 Evaluation Report. Oct. 2002
- 40 STEPS 2001 Report

-
- 41 Dierking and Abrams. Summative Evaluation for SciTech Clubs for Girls
 - 42 Evaluation of the Intel Computer Clubhouse Network. Center for Children & Technology
 - 43 Learning from Television : A Research Review. Center for Children & Technology
 - 44 Evaluating Science WWW Resources
 - 45 Mark St. John. The Turing Test : A New Approach to Evaluating Investments in Educational Capacity and Infrastructure. Assessing the Impact of the Exploratorium 's Institute for Inquiry
 - 46 The Exploratorium 's Institute for Inquiry. Four Cornerstone Claims : A Summary of Evaluation Findings
 - 47 Hammrich & Klein. An Interim Evaluation Report : Head Start on Science and Communication

附录 国外科普评估报告实例

附件一 美国国家科学基金会非正规 科学教育计划评估报告(1998)

中译者注——美国国家科学基金会非正规科学教育计划是美国政府实施的最重要的科普计划,2001年度的预算已达约5600万美元。美国COSMOS公司受国家科学基金会的委托,对该计划实施十年来的效果进行评估。经过认真调研,COSMOS公司于1998年初向基金会提交了最终评估报告。

评估报告认为,美国国家科学基金会在非正规科学教育计划上的投资是成功的,发挥了有效的杠杆作用。该计划正在实现其确立的各项目标。评估报告还坦率地指出了该计划在政策、实践和管理方面的某些不足及可改进之处。

需要说明的是,本报告的任何意见和结论均为COSMOS公司评估小组的一家之见,不一定是国家科学基金会官方的观点、意见和政策。

1. 什么是非正规科学教育?

人们是从哪里了解科学和数学的?对科学、数学、工程和技术(SMET)教育中的知识概念的接触了解也许来自于交谈、课堂教学、电视节目、特定的事件等等。人们的了解不会仅经由某单一渠道——也决不会仅限于教室和实验室的范围。很多人是在正规学习环境之外、在个人内在兴趣的驱使下了解科学的。事实上,人们普遍能记得住的早期科学经历很可能是那些非正规科学活动——参观动物园或科学展览、观看电视科学节目、聆听科学家谈自己的工作、在家内后院探索自然,甚至在厨房做“实验”等等。

国家科学基金会(以下简称 NSF)的一个主要目标是要“促进有益于社会的新知识的发现、综合、传播和利用”。实现这一目标的一个策略就是“使教育富于发现之乐趣,并注重教育与科学探究的联系”。

“非正规科学教育”是自愿、自主和终身性教育。这种学习为进一步的科学活动和学习提供了体验基础和动力。NSF 的非正规科学教育(ISE)计划对一些项目提供资助,在这些项目中,“学习主要是由内在兴趣、好奇心以及探究、操纵、幻想、完成任务和社会交往等欲望驱使的。非正规学习可以是线性或非线性的,它通常由学生自定进度,并注重视觉或实体效果”。

非正规科学学习者——各种年龄、兴趣和背景的人——是从学校之外的场所了解科学的,其教材和活动最初并不是为学校使用或作为课程的一部分而开发的。非正规科学教育活动的参加者可自主地与科学打交道,而不是把活动作为规定的学业的一部分。参加非正规科学活动有几大好处:更好地了解科技学科中的概念、主题、过程和思想;增进对这些学科的职业机会的了解;提高对科学、数学及其应用的认识和了解。

面向青少年或普通公众的电视科学系列片和科学节目,科学和数学题材的电影,在科学及自然历史博物馆、科技中心、水族馆、自然中心、生物园、植物园、动物园和图书馆举办的展览或教育活动以及在社区和青少年活动中心开展的教育计划和活动,这些都属于非正规科学教育的组成部分。人们在这些环境中积累的科学经历交互作用、不断强化,同时受到家庭、学校、同伴和社会的影响。

很多从事科学相关职业的人声称,他们最先对 SMET 萌发兴趣归功于非正规而不是正规的科学接触。他们认为,博物馆和科学中心是激发他们幼年科学兴趣的最重要的因素。

2. NSF 在促进非正规科学教育中起了什么作用？

非正规科学教育(ISE)计划是 NSF 小学、中学和非正规教育处实施的多项计划之一,旨在提供高质量的科学、数学、工程和技术(SMET)教育学习机会。除 ISE 计划外,NSF 小学、中学和非正规教育处还负责师资强化、教材开发、总统数学和科学教学优秀奖以及高级技术教育四项计划(又增设了家长参与教育计划——译者注)。这些计划旨在解决幼儿园前直至 12 年级学生的需要,利用正规(教室)教育环境促进高质量的 SMET 教育。而 ISE 计划的独特之处在于,它鼓励前述支持非正规科学教育的各组织、机构与高教界、工业界、中小学校、联邦机构、州/地方政府以及其他科学工程组织建立合作与伙伴关系。它着眼于发展有助于整个教育体制的改革的能力与资源。

ISE 计划的一切方面都是为了激发并保持人们对数学和科学的终身兴趣。该计划力求做到：

- 使更多的青少年,特别是未被充分代表的青少年(如少数民族、女孩、残疾人)以及社会服务水平低下地区(如农村)的青少年领略 SMET 的奇妙,在校内、校外都进行 SMET 活动；
- 促进非正规教育与正规教育间的联系,以期实现更加优秀、更具创造性的各种学习环境下的 SMET 教育；
- 激励家长及其他成年人更有效地支持改善正规及非正规 SMET 教育的质量,并使之惠及更多的人,鼓励家长支持孩子在家里及其他场所进行科学、数学活动；
- 把非正规科学教育计划和活动传递到目前尚未获得或绝少获得这种机会的较大的地区(如农村地区和市中心贫民区)；
- 通过提高儿童及成年人的科技素养丰富其生活质量。科学素养的提高能使人们知晓 SMET 在其日常生活中的影响,激励他们进一步谋求在这些领域的经历,有助于他们对具有社会意义

的有关政策问题作出有信息依据的、负责任的决定。

按照 ISE 计划制订的有关目标,如下方面的项目也受到支持:

- 将最新的研究成果应用于 SMET 教育;
- 通过电子联网、技术援助和业务发展等活动,促进加强非正规科学教育基础设施;
- 进行非正规教育方法研究,以确定用于激励公众了解科学、数学和技术问题的创新性教育方法的效果;
- 把新内容引入现有的非正规教育活动,以提高其受众的科学、数学和技术兴趣和素养。

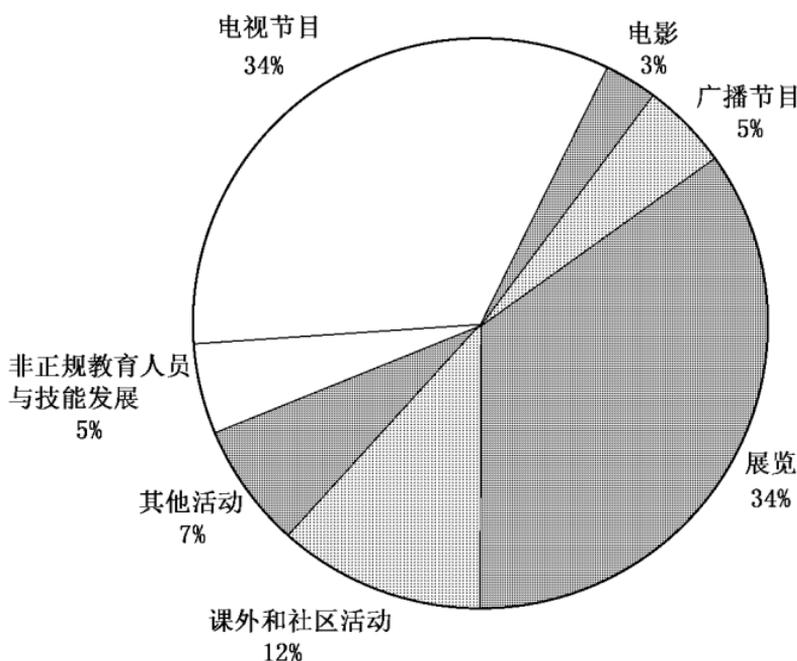
ISE 计划对符合其目标的活动提供资助。在该计划所资助的众多项目中,“神奇学校汽车”可谓集各种科学知识之大成者。这是一部面向少年儿童的富于创新性的电视教育系列片,凝聚了众多领域的人士或组织的共同努力。为扩大这部电视系列片的影响,使其不止限于电视观众,该项目以系列片为桥,与博物馆、青少年组织、公共广播公司(PBS)和一个商业实体(学生节目制作公司)建立联系,由此形成的合作派生出教师与家长指南、到儿童博物馆和科技中心的巡回展、图书馆博览会的教学展示,向 100 多家博物馆分发活动箱以及通过国家机构传递青少年外延活动材料等一系列产品和活动。

任何组织——公共或私营、营利或非营利机构——均可提交开展非正规科学教育活动的项目建议。ISE 计划鼓励为赢得大量受众、改进传统方法而尝试开展各种新颖、新兴的非正规科学教育活动。这类活动包括利用商业广播、新的博物馆交互展览方法、面向特殊受众群体的通俗出版物、新开发的电视和交互型学习媒体、家庭学习资源以及基于家庭和社区的成年教育项目等。

如图/表 1 所示,1984—1994 财年,ISE 计划资金大部分投向制作电视节目和办展览,其中支持办展览的资金比例略高。其余的资金用于支持诸如课外及社区活动、非正规教育人员与技能发

展、无线电广播节目、联系学校的活动、电影等。

图/表 1 ISE 计划经费按活动领域的分配 1984—1994 财年



3. 什么是“非正规科学教育(ISE)”计划评估?

此次 ISE 计划总结性评估是要探明 ISE 计划 1984—1994 财年间资助的项目产生的影响。总结性评估与形成性评估是有区别的。形成性评估是在计划执行阶段进行的评估,以期改进计划的执行,而总结性评估是通过调查计划的累加效果来评价其长期影响。国家科学基金会编写的《科学、数学、工程和技术教育项目评估手册》将总结性评估定义为“通过评估给出一项干预所产生的价值的判断结论,并提出此项干预是继续执行、或修订或终止的建议”。

本次评估的目的,是考查 1984—1994 年特别是 1990—1994 年间的 ISE 计划的影响。以往国家科学基金会资助的评估调研侧重于评价单个项目的目标,而不是调查 ISE 计划的长期影响。

此外,我们对以往的非正规科学教育项目评估进行研究后发现,评估中缺少与 ISE 计划非常相关的评估目标的数据。例如,一项评估调查考查了科技中心为学校和老师提供的科学教学支持。在另一项评估调查中,ISE 计划只是一短期评估试点调研的对象,用来帮助国家科学基金会解决困难的评估问题。此外,无论是国家科学基金会资助的还是非国家科学基金会资助的非正规科学教育项目评估,都没有关注与社区组织的协作问题(ISE 计划资助的 Ghostwriter 项目除外),也没有关注受评项目对非正规科学教育体系的影响。

因此, COSMOS 公司在进行本次评估设计时采取了以下假定:非正规科学教育不一定是一系列只具有短期影响的干预措施,相反地,它是能长期影响科学学习和体验情境的手段。本次 ISE 计划评估的重点是该计划的较长期影响,评估的依据是 6 大评估目标——4 个 ISE 计划的目标(图/表 2 中的目标 A-D)和两个附加目标(E-F)。这些目标被用来评估国家科学基金会的整个 ISE 计划。

图/表 2 指导 ISE 计划评估的六个目标

指导此次 ISE 计划评估的目标包括国家科学基金会 ISE 计划的 4 个目标(A-D)和两个附加目标(E-F)(两个附加目标被国家科学基金会确定为本次调查评估的目标,尽管它们不在 ISE 计划的原定目标之列):

A. 使更多的青少年,特别是未被充分代表的青少年(如少数民族、女孩、残疾人)以及社会服务水平低下地区(如农村)的青少年领略 SMET 的奇妙,在校内、校外都进行 SMET 活动。

B. 促进非正规教育与正规教育间的联系。

C. 激励家长及其他成人成为明达的支持者,使他们支持改善正规及非正规环境下的科学、数学和技术教育,让家长支持他们的孩子在家里及其他地方进行科学和数学活动。

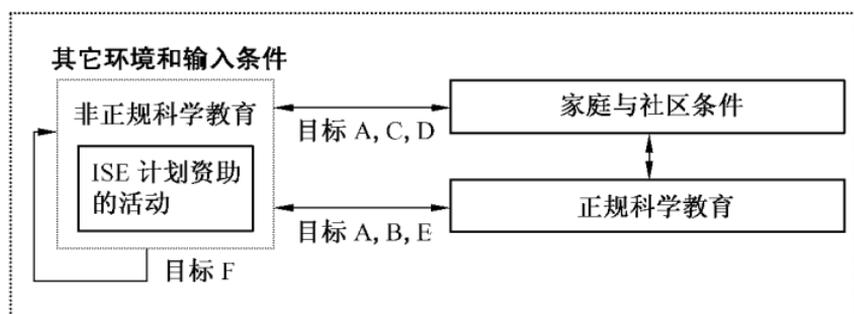
D. 以丰富人们的生活水平为目的,提高儿童及成人的科学素养,使他们更加了解科学、数学和技术对其日常生活的影响,从而能在今后继续寻求科学和数学体验,对具有社会意义的科学政策问题能作出有信息依据的、负责任的决定。

E. 促进正规和非正规教育界各机构、个人之间建立联系,加强合作。

F. 对非正规科学教育体系产生广泛和深远的影响。

我们从一种宽阔的非正规科学教育视角来设计本次评估的方法。因为科学素养的提高不可能是单个 ISE 项目的结果。因此,评估调研不是采用典型的线性输入-输出方法,而是承认 ISE 计划受家庭、社区和正规教育的影响——也就是说,受非正规科学教育活动所处情境的影响。考虑到非正规科学教育的情境性——要在总和而不是逐个项目的基础上考虑非正规科学教育,我们的评估框架(见图/表 3)纳入了六个评估目标(A-F)。这样设计的目的是提供一种能解释和比较评估结果的比较逻辑——将 ISE 计划资助的活动与非国家科学基金会资助的活动进行结果比较。同样,将 ISE 计划项目人员的答复与从事科学职业的人员的答复进行比较。设计中也考虑了国家科学基金会“费用分摊”要求,考虑了 ISE 项目还同时得到其他来源的资助,如联邦机构、州和地方的机构、公司或者基金会的资助。

图/表 3 情境条件和国家科学基金会
非正规科学教育计划的评估目标



由于多家机构共同资助的活动很难解析清楚,因此,我们选择某些未被 NSF 资助的非正规科学教育项目和活动作为对照组进行研究。有关对照组问题,应注意以下三个问题:

首先,对照组中的项目应具有与 NSF 资助的项目相似又不完全相同的特征。

其次,如果发生以下情况,对照组项目可能已经受到 ISE 计

划指导原则的影响 :①对照组项目的执行机构以前向 ISE 计划提交过项目申请书 ,并经过 ISE 计划主管人员的评审 ,该项目的执行机构将 ISE 计划项目评审人的意见吸收到其后来的项目申请中 ,最终得到 NSF 之外的其他机构的资助 ,甚至得到 NSF 的资助 ;②对照组项目的执行机构以前曾经得到过 ISE 计划的资助 ,并使用 ISE 计划项目评审人的意见来报告一个非 NSF 资助的项目的进展 ;③在我们开始收集评估调查数据后 ,对照组项目有可能又获得了 ISE 的资助。由于在识别非 NSF 资助的非正规科学教育项目上存在困难 ,所以我们请 ISE 计划的主管人员和顾问来审查我们最终的项目选择结果。

在可能的范围内 ,我们同时收集 NSF 资助的 ISE 计划项目和非 NSF 资助的非正规科学教育项目的调查数据。

NSF 资助的 ISE 计划项目 :调查数据来自于 :

- ISE 计划的档案资料数据 ;
- 对 1984—1994 年间受 ISE 计划资助的 210 名 (总计为 277 名)项目负责人的调查 ;
- 对受 ISE 计划资助的 64 名 (总计为 179 名)机构负责人的调查 ;
- 对 15 家机构的实地考察。在 ISE 计划资助的 347 个项目中 ,这 15 家机构获得了 52 个资助项目(许多机构获得了多个资助项目) ;
- 在 22 家机构进行的专题小组座谈会 ,邀请有关的非正规科学教育活动人员参加。

非 NSF 资助的非正规科学教育项目 :调查数据来自于 :

- 从美国科学促进会(AAAS)会员名录和美国全国科学教师协会(NSTA)数据库中随机抽取 262 个会员进行调查 ;将调查结果汇总并归入“从事科学相关职业的人员”调查类别 ;

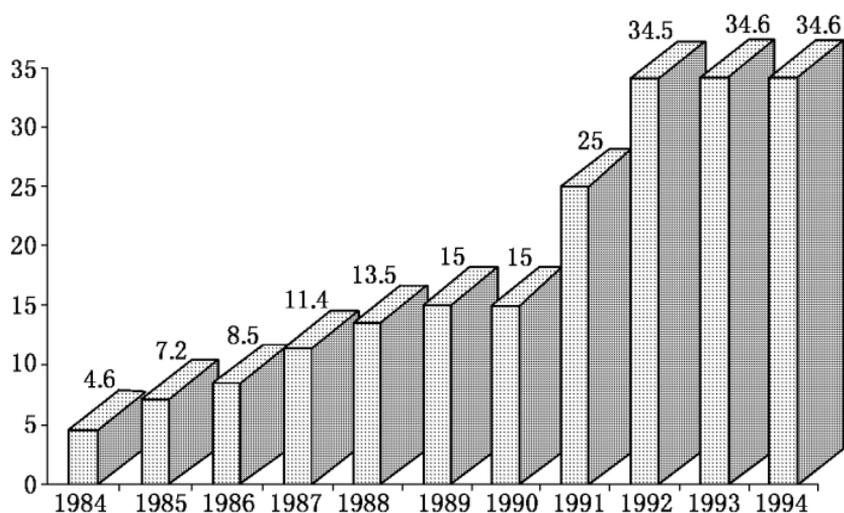
- 对获得非 NSF 的 ISE 计划资助的 6 家代表机构进行实地考察；
- 在六家机构进行的专题小组座谈会，邀请有关的非正规科学教育活动人员参加。

受资助的项目被分为 4 类：媒体类、科学中心和博物馆类、专业培训类以及社区类。

4. NSF 在 ISE 计划的投资怎样？

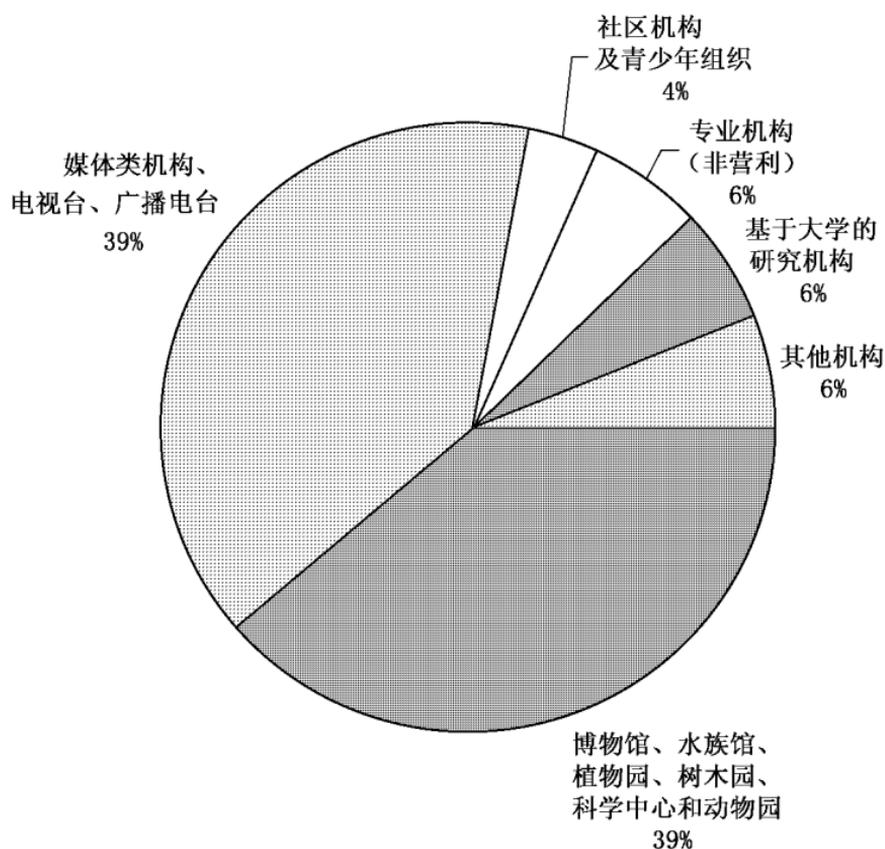
ISE 计划资助决策以事先确立的 4 大计划目标为指导。如图/表 4 所示，ISE 计划实施十年来，年资助额稳步增长，两次大的增长发生在 1990—1993 财年（应为 1990—1992 财年——译者注），两次增额共计 1850 万美元（应为 1950 万美元——译者注）。

图/表 4 NSF 非正规科学教育计划经费(单位：百万美元)



图/表 5 显示了 ISE 计划拨给各类机构和组织的资金比例。

图/表 5 ISE 经费按执行机构的分配 1984—1994 财年



作为多部儿童电视科学系列片的制作机构,儿童电视节目制作公司(CTW)自 1984 年以来共获得 ISE 计划 31 506 743 美元的资助,比其他任何一家机构都多。图/表 6 列出了获 ISE 计划资助最多的 25 家机构,从中可以看出,ISE 计划十年提供的资助中,36% 流向像 CTW 这样的传媒机构,40% 流向博物馆和科技中心。

5. 如何通过杠杆作用扩大 ISE 投资?

在本次评估的所有 ISE 计划资助的项目中,值得注意的是,ISE 计划资助仅构成项目总费用的一部分——这是 NSF“费用分

担”或“资金扩大”原则所规定要满足的条件。NSF确实把自己看

图/表 6 获 NSF 非正规科学教育计划资助最多的 25 家机构

1984—1994 财年

机 构	总资助额	获准项目数
儿童电视节目制作公司(纽约)	31 504 743	9
旧金山探索馆	6 896 034	11
富兰克林学会	4 992 302	11
国家公共无线电广播公司	4 679 663	4
学生节目制作公司	4 660 647	1
芝加哥菲尔德自然历史博物馆	4 405 938	4
纽约科学馆	4 121 628	8
波士顿科学博物馆	3 886 634	8
波士顿公共电视台 WGBH	3 767 035	13
俄勒冈科学工业博物馆	3 430 481	4
芝加哥科学工业博物馆	3 429 628	3
QED 传播公司	3 266 044	5
明尼苏达科学博物馆	3 061 183	5
西雅图公共电视台 KCTS	2 977 824	3
太平洋科学中心(西雅图)	2 665 545	3
内布拉斯加公共电视台	2 653 000	4
美国科促会	2 542 541	5
康奈尔大学	2 449 017	5
科技中心协会	2 447 382	11
女孩公司	2 399 081	3
棱柱制作公司	2 289 267	4
纽约公共电视台 WNET	2 167 536	5
波士顿儿童博物馆	2 167 536	5
哥伦比亚科学工业中心	1 875 938	3
夏洛特科学博物馆	1 846 756	3

作项目开发的“催化剂”。它的投资作用是,通过提供手段、制定计划和资金支持,帮助其他组织从事科学教育和学习的前沿活动,以此加强和提高国家在 SMET 领域的能力。

要实现 SMET 领域的杰出地位,重要的是要在商业、工业、研究人员、学校和非正规科学教育机构之间建立合作关系,这在 NSF 的工作目标中已体现出来。很多博物馆和科学中心或是作为项目的活动场所,或是作为项目的协作单位,其数量之多反映出 NSF 对合作的强调。有效的伙伴关系能把社会中的最优秀的思想结合到一起,并保证了财、人、物资源的供给。

本次被评估的获 NSF 资助的机构都以 ISE 计划所提供的创始资金为基础,从州与地方政府、其他基金会、当地工商公司以及联邦其他机构争取额外的资源支持。扩大的资金部分通常用于开发更多的教学材料或将外延活动扩大到广泛的受众群体。此外,大多数的项目在计划阶段就开始争取从其他渠道获得必要的资助,以支持项目中 ISE 计划未资助的部分活动。一些接受调查的机构报告说,在获得 NSF 的 ISE 计划资助后,其他资助机构不再对项目进行立项评议,转而依靠 NSF 的同行评议过程——这表明民间资助机构对 NSF 管理模式的信任。

ISE 计划资助的项目的负责人称,尽管争取其他渠道的额外资助有一定的困难,但如果 ISE 计划不提供部分资助作催化剂,通过杠杆作用扩大资助或费用共担这样的举动就很难有结果。

ISE 计划十年来的数据表明,ISE 计划的“资金扩大”举措已得到国家的认可及额外的资助。例如,在芝加哥举办的“非洲展”,全部费用的 73% 是从洛克菲勒基金会、菲尔德博物馆会(芝加哥)、国家人文科学捐赠基金和乔伊斯基基金会(芝加哥)筹集来的。

同样,基于社区的“科学保姆”项目从一家民间投资机构获得了其 63% 的预算。而且,“资金扩大”举措为开展外延和传播活动来扩展科学学习提供了额外的资源与实物支持,这类活动包括向

教师及家长分发活动指南或与技术相关的活动场所建立联系等。

洛厄尔天文台曾一度无力满足公众对参观活动的要求。在 ISE 计划的资助下,该天文台增加了面向儿童和家庭的多项活动,并利用从其他渠道争取的追加资金建立了参观者中心。

6. ISE 计划有望实现其目标吗?

本次评估表明,ISE 计划正在不同程度地实现其制定的各项目目标。该结论是根据实地访问、焦点小组座谈、现有的评估和调查等所获得的支持性证据得出的。

目标 A:使更多的青少年领略科学、数学和技术的奇妙,在校内、外都进行这类活动,尤其针对那些来自未被充分代表的或社会服务水平低下的群体(如少数民族、残疾人和妇女)。

很多 ISE 计划项目受到公众的喜爱。从我们掌握的展览参加人数记录和电视收视率来看,ISE 计划资助的展览和电视节目吸引了大量的青少年。例如,“音乐是怎样产生的”巡回展向儿童和成年人展示声音和音乐的物理原理,1988—1995 年间在 18 个博物馆进行了展出,吸引了 220 万名观众前来参观。该展览在费城富兰克林学会科学博物馆展出的 78 天里,吸引了 16 万名观众,尤其受到在校学生的欢迎。设在亚利桑那州弗拉格斯塔夫的洛厄尔天文台举办的“天文学家的工具”展也收到了同样的效果,每年吸引近 8 万名参观者,而举办此展前,天文台的年参观人数为 6000 人。洛厄尔天文台是一家民间纯经营性的行星研究机构。“天文学家的工具”展是该天文台的天文学家及项目成员设计建造的一组交互型展览,在非正规科学教育环境中教授天文学研究的历史和新概念。

ISE 计划媒体类项目包括在公共电视台播放的一些深受欢迎

的儿童系列片。例如,由学生节目制作公司(Scholastic Productions)摄制的“神奇学校汽车”电视系列动画片描写了女教师带着她的一班学生从事一系列动手型科学探索活动。这部系列片在公共广播公司(PBS)的300多家电视台播出,每周至少播一集,覆盖度达到全美96%的家庭。1995年进行的尼尔森收视率统计结果显示,该系列片在2~5岁少儿节目中收视率居第一位。“阅读彩虹”是为5~8岁儿童制作的电视系列片,每集表现一本好看的书的某一主题,由于每天都在播出,同样拥有大量的观众。

尽管有上述这些可观的数字,要评估青少年对科学的热情仍很困难——除非根据纯粹的受众数据作出推断。因此,有关青少年科学热情的数据及由此得出的结论大都得自于评估小组实地参观中进行的访谈、观察、文献综述和焦点小组座谈调查。例如,将参加波士顿科学博物馆“检验理论”展的青少年(参观者在该馆的各活动站可进行从提出假设、开展实验、收集数据到检验自己提出的科学过程理论等一系列活动)作为评估小组选择的焦点小组座谈的参与者,这些学生在参加活动的两个月后仍旧接受有关其实验问题的询问。当被问及是否喜欢学校的科学课程时,他们很干脆地说“不喜欢”。当被问及他们在参观展览时做了些什么事,他们马上回答说,“我们进行了科学探究”。这些青少年认为展览展示的是“有趣的科学”,与他们在课堂上学的科学不一样。不仅如此,参加者都能回想起他们两个月前做的各种实验。

评估班子在儿童电视节目制作公司(CTW)现场参观期间做出的正式评估结果显示,在观看了由该中心摄制的“3-2-1接触”科技系列片后,观众对科学的热情有所提高,这表现在如下方面:对观众的科学兴趣测试结果表明,看片后比看片前的科学兴趣要强;观众普遍不再认为科学枯燥乏味;对本节目表现的科技事实信息有了更好的了解;观众有兴趣参加后续的活动。

就争取更多的少数民族、妇女和残疾青少年成为ISE活动受

众这一目标而言,最直接的证据来自专门为此目的而开展的 ISE 活动。这些活动在显示它们如何努力接触其目标受众方面做出了很大努力。例如,在新墨西哥州阿尔伯克基市的“探索”新兴科学中心 1994 年吸引了 7 万多名参观者,其中 47% 是拉美裔美国人(占当地人口的 35%),14% 是印第安人(占当地人口的 11%),3% 是黑人(占当地人口的 1%)。参观“天文学家的工具”展的学生中,50% 是印第安人,20% 是拉美裔人,20% 是高加索民族后裔,10% 是黑人。同样,参与 SERIES(供学校或社区环境下使用的动手型课程)的工作人员报告说,在某一活动点参加活动的 200 名青少年中,50% 是拉美裔人,30% 是高加索民族后裔,20% 是黑人。

目标 B:促进正规科学教育与非正规科学教育之间的联系。

目标 E:促进正规和非正规教育界各机构、个人之间建立联系,加强合作。

这种联系的形式之多、数量之大以及它们在激发青少年的科学热情、帮助成年人及子女更好地了解科学方面的作用令人叹为观止。这些联系包括媒体、大学、教育机构、专业团体和社区组织之间的合作。有了这样的联系,项目就增添了多学科的特征,就有了资源保证,一些价格昂贵的材料也因此能买得起了。

ISE 计划项目在建立与正规教育的合作方面收效也很显著。由 ISE 计划项目开发的外延材料和活动已促使非正规科学教育与正规科学课程合二为一。教师们通过参与 ISE 计划项目,如担任项目咨询小组的成员、参加项目发展与评估小组、编写 ISE 课程或宣传材料等,得以将这些宣传材料和活动作为富有创意的教学资源加以利用。教师们还利用博物馆举办的实验班所开发的教材拓展学生野外考察学到的概念。“神奇的学校汽车”和“阅读彩虹”电视系列片的录像带及配套宣传材料也被教师们用于课堂教学中。

“科学保姆”是基督教女青年会(YWCA)开展的ISE计划项目,旨在鼓励女性青少年参与科学活动。该组织主办的保姆培训班包括训练保姆们如何与少儿一道利用非正规科学活动。

ISE计划项目开发的科教产品和材料,特别是动手型活动,同样提供了与专业发展活动的联系。例如,SERIES项目提供课程培训,供救世军(一慈善机构——译者注)、公共住房局、男童子军、营火少女团、钓鱼与游戏委员会和国家环境教育协会使用;“神奇学校汽车”和“阅读彩虹”两部电视系列片的材料被国家师资培训学院引用;“科学嘉年华会联合体”为新兴科学中心提供展品及示范节目。

目标C:激励家长及其他成人成为明达的支持者,使他们支持改善正规及非正规环境下的科学、数学和技术教育,支持使这种教育为更多的人能够接受,让家长支持他们的孩子在家里及其他地方进行科学和数学活动。

目标D:以丰富人们的生活水平为目的,提高儿童及成年人的科学素养,使他们更加了解科学、数学和技术对其日常生活的影响,从而能在今后继续寻求科学和数学体验,对具有社会意义的科学政策问题作出有信息依据的、负责任的决定。

本次评估中发现,很多ISE活动鼓励成年人参加和全家参与。“天文学家的工具”展专门设立了供全家活动的展室,这使得“洛厄尔天文台之友”成员人数在5年内由200人增加到1000人以上,也使该天文台得到的预算拨款增加了约10万美元。像4-H农村青年会这样的地方公民组织因采用SERIES课程,促使农业季节工人扭转了对其子女参加4-H组织的消极态度,因为这些家长们现已懂得科学对其子女的教育和未来前途的重要性。SERIES节目在帮助“改变”问题少年方面也获得了成功。该系列片项目主任举例说,4个城市的四年级学生采用SERIES教材后,

出勤率和表现都变好了。

“检验理论”展在设法激发家长与孩子间交互作用方面颇为用心。该展览的工作人员经常听到家长和孩子就展览所提的问题进行热烈的讨论。参加“阅读彩虹”焦点小组座谈调查的家长称,在看完“阅读彩虹”电视系列片后,他们不再惧怕科学,不再担心帮孩子参与学校科学活动。一名年轻的参加者评价说,该节目把科学与实际生活和日常事物联系起来。另一个调查目标小组的5名儿童也同意这一说法。

目标 F 要对非正规科学教育体系产生广泛和深远的影响。

关于 ISE 计划第 6 个目标的实现情况可从其他 5 个目标的总的实现上找到证据,然而,这样的推断须在下述前提下才是合理的,那就是通过实地访问及现有的评价对 NSF 资助和非 NSF 资助的 ISE 活动所作的比较证明,两者在活动的水平与类型、目标和结果上没有明显差异。

事实并非如此。NSF 资助和非 NSF 资助的 ISE 活动最明显的不同是在项目总体规划及策略上。与非 NSF 资助的项目不同的是,NSF 的 ISE 计划为其项目制订了更加全面和系统的过程。NSF 推荐的过程——评述和修改项目建议、采用咨询小组、形成性评估、总结性评估和宣传推广——在所调查的 6 家非 NSF 资助的 ISE 项目中都没有全部见到。非 NSF 资助的项目的宣传推广活动大都局限于通过散发传单或现有的信息源作的广告。另外,非 NSF 资助的项目缺乏 NSF 资助的项目所具备的两大重要因素——资金扩大和跨学科性。只有一个非 NSF 资助的项目——“活跃的青少年”兼具上述两大特点。该项目把活动的直接重点放在以往获得科普场馆服务较少的目标群体上,这在所调查的非 NSF 资助的项目中也是唯一的一个。

此外,与 NSF 资助的项目相比,非 NSF 资助的项目所开展的

活动类型较少——非 NSF 资助的项目的投资似乎更有限。不过，很多非 NSF 资助的项目确实开展了与其他机构的合作，这可能是由这类项目在资源上的限制促成的。

ISE 计划资助的项目似乎方向正确，即努力实现 ISE 计划制订的 6 大目标。NSF 重视博物馆和科学中心作为 ISE 项目活动场所或协作单位的作用，同时 NSF 认识到这类机构在激发儿童科学兴趣方面的重要性。从事科学相关职业的一些人士的评论就证明了这种正面影响。

7. 人们对 ISE 计划资助的活动有什么看法？

为弄清楚科学工作者、获 ISE 计划资助的机构以及 ISE 计划项目/活动的负责人对 ISE 计划方方面面的看法，我们针对上述三类人员进行了 3 次不同的调查，提出了一些问题。从调查的数据可以看出 ISE 计划是如何影响个人及专业人员的科学经历，如何促使他们参加科学活动的。

我们对 ISE 项目负责人提出了一些问题，以确定其项目在多大程度上让青少年和成人参与非正规科学教育，以及其他方面是怎样实现 ISE 计划目标的。教师、科学家、工程师、化学家和研究人员应我们的要求报告了他们目前的工作以及他们本人在青少年时期的科学经历（如曾激发其科学和数学兴趣的难忘的活动和经历）。把他们的答复与 ISE 项目策略对比和评估后发现，ISE 计划发展的方向是正确的。向有关机构负责人所提的问题旨在获得这些机构的看法，以确定与其他资助源相比，ISE 计划资助的影响。

总的来看，我们收到的答复展现出 ISE 计划策略的积极形象：它促进非正规与正规科学教育的联系，吸引青少年和成人参加非正规科学活动，重视对项目的评估和增进机构间的合作。因此，根据“专家们”的意见，ISE 计划正在努力实现其目标。下面将就受调查者有关 ISE 计划各目标领域的答复作些扼要概述。

- 与正规科学教育的联系更多

在要求机构负责人比较 NSF 与其他资助源的影响时,他们回答说,与非 NSF 资助的计划相比,NSF 资助的计划更加注重公众参与,重视评估以及与其他机构的联系——充分显示出 NSF 的 ISE 计划对非正规科学教育的影响(参见图/表 7)。

平均而言,73%的受调查机构负责人强烈感到,NSF 的 ISE 资助促进了非正规科学教育各个方面的发展,仅有 43%的受调查者认为非 NSF 的 ISE 资助也做到了这一点。

图/表 7 NSF 与非 NSF 的 ISE 资助对非正规科学教育系统的影响
(受资助机构负责人的答复)

非正规科学教育方面	表示“非常同意”的机构负责人			
	NSF 的 ISE 资助的影响		非 NSF 的 ISE 资助的影响	
	人数 ¹	%	人数 ¹	%
增加了公众的参与	50	83	22	56
采用评估以改进计划或项目	47	83	16	40
与其他机构进行了合作或联系	47	78	19	48
达到了以前很少服务的群体	45	76	21	51
包括了新的让公众认识的科学领域	42	73	15	38
采用了新技术	42	72	17	43
实现了与非正规科学教育的密切配合或联系	41	70	17	41
实现了与正规科学教育的密切配合或联系	37	66	16	37
增进了公众的支持	37	63	13	33
利用了技术咨询专家来策划产品、展览或活动	36	63	15	38

注 1. 允许有多种答案。

- 吸引了更多的青少年及成年人参加

ISE 项目负责人和从事科学相关职业的人一致认为,在促进青少年和成年人接受非正规科学教育方面,有两种活动效果非常好:到学校访问或外延活动以及与某些组织(如女童子军)的计划协调。从事科学相关职业的人认为,科学夏令营报告会(非 NSF 的 ISE 活动)在这方面也很成功,虽然很少举办让青少年和成年人共同参加的科学营活动。

如图/表 8 所示,从事科学相关职业的人员报告说,参观博物馆是他们童年时代所经历的最难忘的科学活动。这种活动激发了他们早年对科学的兴趣,加强了与他们在校正规科学课程的联系,在他们的头脑中植入了科学思想,这种思想在今天对他们仍然起作用(参见图/表 9 和图/表 10)。

图/表 8 童年时代最难忘的非正规教育活动
(从事科学相关职业的人员的答复)

非正规科学教育活动的类型	从事科学相关职业的人 (受调查人数 254 人) (%)
参观天文馆、水族馆或动物园	92.9
参观科学博物馆或自然历史博物馆	85
有与科学相关的业余爱好或科学玩具	74
出于兴趣读科学书籍或杂志	72.8
收看电视或收听广播的科学节目	65
参观植物园	58.3
参观科学博览会	41.7
参观社区主办的青少年科学活动	21.3

图/表 9 青少年时代获知的、现仍在发挥影响的科学思想的获得渠道

非正规科学教育活动的类型	从事科学相关职业的人 (受调查人数 244 人) (%)
参观科学中心、博物馆或展览会	51.6
收看或收听传媒的科学节目(电视、电影、无线电广播)	44.7
参加社区主办的青少年科学活动	11.5

图/表 10 首次与学校发生联系的早期的非正规科学教育活动
(从事科学相关职业的人的答复)

非正规科学活动的类型	从事科学相关职业的人 (受调查人数 250 人) (%)
在科学中心、博物馆或展览会上看到的東西	50.0
收看电视或收听广播中的科学节目	44.8
参加社区主办的青少年科学活动期间从事的活动	13.6

因此，NSF 的 ISE 计划瞄准科学中心和博物馆(该计划的一个显著特点)，把目标集中于这类传统的非正规科学教育场所。如科学工作者所称，科学中心和博物馆激发了他们的科学兴趣。用他们自己的话来说：

“在参观了加利福尼亚的一家博物馆之后，我变得对鲨鱼、恐龙和矿石异常感兴趣，我开始跑图书馆，以便对此了解得更多。”

“在自然中心的直接观察使我第一次萌发了强烈的兴趣和好奇心，我开始有了想要“知道”和“探究”答案的渴求和兴趣。”

“通过参观博物馆、动物园、天文馆和水族馆，使我了解到万物之间是如何相互联系的。”

虽然这些人认为 ,博物馆和科学中心激发了他们对科学的最初兴趣 ,但他们却把参观这类场馆列在他们过去六个月来最经常从事的活动的第三位。而实际上 ,这一偏差反而更说明了他们对童年经历回忆的可信性 ,表明是真正的记忆中事情 ,而不是在表达其现时对博物馆的支持呼声。

NSF 的影响显然是重大的。例如 ,从事科学相关职业的人谈及他们目前的非正规科学教育活动和 NSF 所推进的活动间的相关性。结果表明 ,二类所确定的活动排序相同 ,都把“鼓励朋友参加与科学有关的活动”列在了首位。

“芝加哥的众多博物馆 ,特别是科学工业博物馆点燃了观众的科学兴趣 ,并使之保持下去 ,这种兴趣在小学、中学或高等教育中从没有得到培养。学校中的科学是枯燥乏味的 ,而博物馆中的科学则充满了发现和乐趣。现在的情况仍是这样。”

——一位从事科学相关职业的人士

• 更加重视评估

如图 /表 7 所示 ,NSF 资助的机构的负责人称 ,与非 NSF 资助相比 ,NSF 资助重视评估。以下是大量的受访者就此发表的一些典型言论 :

“NSF 的工作人员确实认真地采用了一套评估体系 ,对已有的和新的计划活动、标识、出版物、展品等逐项进行评估。NSF 资助事实上扩展了教育和传播的每一领域 ,促使和指导我们完成其他重要的设计 ,如标识总设计等。其他非 NSF 资助源虽然也扩大了我们所提供的教育机会 ,但在‘深度或广度’或程度上远不如 NSF 项目。”

机构的负责人还指出 ,与其他非 NSF 资助源相比 ,NSF 加强了它在发表和传播项目评估结果方面的工作。

• 更多的合作

如前所说 ,博物馆和科学中心在激发人们早期对科学萌发兴

趣方面起着非常重要的作用。获 NSF 的 ISE 资助的各机构的负责人称 ,ISE 资助促进了他们与博物馆或科学中心的合作。与此相比 ,由非 NSF 资助促成的各种合作中 ,与博物馆的合作却排在了第四位(参见图 /表 11)。

图 /表 11 NSF 资助和非 NSF 资助对增进合作的影响比较
(受资助机构负责人的答复)

受访人数 179 人

合作者类型	认为增进了合作的回答者			
	NSF 的影响		非 NSF 的影响	
	数量 ¹	%	数量 ¹	%
博物馆	40	69	28	58
小学	35	64	33	71
基于社区的科学或数学教育活动	35	63	22	51
个别研究人员	35	63	16	35
私营工商企业	33	59	30	66
学院或大学	29	54	22	49
青少年组织	29	50	30	62
中学	25	46	26	57
社会服务机构	22	39	16	35
电视台或广播电台	21	38	17	38
日托中心	8	14	12	25
宗教组织	7	12	6	13

注 1 :允许有多种答案。

NSF 的 ISE 项目与博物馆的较广泛的合作非常符合这项计划的目标 :使非正规科学教育深入到广大的各种受众中去 ,建立非正规科学教育与正规科学教育间的联系 ,激发青少年的科学热情。受调查者对所调查问题的回答反映了 NSF 的 ISE 项目具有的这种倾向 :

“ISE 计划增强了我们与其他博物馆之间采用某种方式建立的连带关系。如果没有这项计划的支持,我们是不会这样做的。”

“在博物馆大家庭中,动物园和水族馆往往被看作‘后妈生的孩子’,即不是真正的博物馆!也就是说,它们不是真正的科普场所。NSF 的资助使我们得以证明,这种观点是错误的。”

“我们现在和本州的科学中心建立了非常密切的关系,这种关系使我们获得了很多其他的项目。”

8. 政策、实践和计划方面有哪些启示?

根据报告中所列举的证据,我们从对 ISE 计划的总结性评估中得出的主要结论是:该计划在逐步实现预期目标,而且对目标群体产生着影响。如果我们重新评估和改进这些目标,该计划就会发挥更大的作用。为清楚起见,我们把调查结果和建议归纳为以下三大方面:政策影响,非正规科学教育实践和 ISE 计划管理。

• 政策影响

尽管本次评估证实 ISE 项目正在逐步达到预期目标,但在数据收集和评估过程中,我们获得了下述重要发现:

ISE 项目正在逐步实现目标,但 ISE 计划的重点内容因过于广泛而分散,没有反映出科学、数学、工程和技术(SMET)在 NSF 工作中较显著的优先性。

如图/表 12 所示,对 ISE 计划在 1990—1996 年间所资助的项目进行分析后发现,侧重下述两个领域的项目很少:①数学——仅占总项目的 5.4%,需要对其给以更高的重视,因为数学对所有科学学科都具有关键的作用。②计算机科学、工程和技术——总计仅占总项目的 5.1%,也未能反映出整个国家在科学、数学、工程和技术教育中对这些领域所给予的重视。再者,在当今社会,数学和计算机科学是通往其他科学和学科的大门,需要我们给以更多

重视。

有趣的是,这种不足并不是 ISE 计划在实现其目标上的一个失败,而是说明了这些目标本身存在一个根本性问题。已确立的 ISE 计划的目标(六个评估目标中的四个)都未直接涉及受资助项目的课程内容。由于该计划的目标是过程目标,它们不能确保 ISE 计划项目与 NSF 所要求的任何优先内容相一致。

图/表 12 ISE 项目按主题领域的分布 1990—1996 财年

n = 310

主题大类	数量
科学	277
数学	17
计算机科学、工程和技术 ²	16

1. 主题领域依 EHR 数据库提供。为核实项目的主题分类,我们阅读了所有项目的文摘,必要时对分类作了改动。所有项目的主题分类经核实之后,ISE 计划项目官员又对其进行了审核,以确保数据均准确性。
2. EHR 数据库未区分这三个学科领域。

并不奇怪,在进行评估的这段时期,ISE 计划年度项目申请通报和指导原则并没有清楚地阐明它对申请项目所期望的优先内容,因此,为了今后减小由此产生的缺口,在向项目申请者提供的信息中,NSF 应指明对项目所期望的方向和内容,从而 NSF 可以促进那些能更好、更稳妥地实现其广泛的 SMET 教育目标的项目申请和拨款。NSF 项目也应该更多地重视传统学科领域(如数学和计算机科学)以及新兴学科领域(如材料科学)。这些改进将会促进 ISE 计划在将来更有效地发挥作用。

- 非正规科学教育实践

ISE 计划应重新审订已建立的过程和内容目标,使它们能更直接地与 NSF 目前的 SMET 教育内容重点相一致。事实表明,ISE 计划已经成功地加强了几种值得继续重视和强调的非正规教育实践。很多创新性的 ISE 项目所开展的活动为科学及相关领域扩展或开辟了新的前景。参与者能够进行一些科学程序或以亲自动手的方式“感受”科学,也能够了解科学和数学在日常生活中的影响。我们所调查的 ISE 实践产生了一些高质量的外延宣传材料和活动,这有助于把对所经历的现象的认识和了解传播到家庭、社区或学校中去。很多 ISE 项目编写了教师指南,请教师任项目咨询专家,并且制作了供课堂使用的教材,如书籍、磁带、录像带、软件、招贴画和成套教具。这种外延策略鼓励人们进一步从事某些具体课题,激励人们参加相关的活动,而且鼓励家庭巩固和扩展他们参加非正规科学教育活动时学到的概念。

外延策略和活动也被用来鼓励那些未能充分代表的群体的参与,然而,这种努力成功与否尚且未知,因为很少有正式数据表明其获得了成功。那些并非专门用于鼓励未被充分代表群体参加的项目——尽管它们可能已吸引了大量的青少年和成年人,似乎未能达到大多数的少数民族或社会服务水平低下的群体。

这项计划应强调某些专门的实践,这会有助于确保未来项目包含 NSF 认为和其科学、数学、工程和技术优先领域非常相关且必不可少的实践活动。

根据受众年龄范围评估项目的非正规科学教育实践较为容易。在这方面,ISE 项目做得很好,支持了 ISE 计划倡导的科学面向所有年龄的每一个人的思想。很多项目,特别是在科学中心开展的项目,都推出了适合多代人参加的活动。即使是以青少年为主要目标受众的项目,所开展的活动同样鼓励成年人参加。在开发同时适用于所有年龄的人的活动的过程中,ISE 项目鼓励更多

的家庭参与,激发公众对非正规科学教育场馆的兴趣,由此增强了整个 ISE 计划的潜能。

除因项目而异的策略以外,ISE 计划现行的管理策略同样有助于加强整个计划的实施。例如,NSF 计划官员在提交项目申请之前先对初步的 ISE 项目申请进行评审。他们采用严格的项目申请评审标准,且力求包括诸如过程性评估、咨询委员会和合作等内容,这些因素反过来会增加项目的资源供给,保证项目的多学科重点,瞄准未被充分代表群体且增加其机会,它们成为决定资助与否的依据。正是因为经过认真设计、深入研究,并且经证明具有较高的水准,通过 NSF 的严格评审的项目才会有较高的成功概率。该计划在能力建设方面的另一个特点是它在项目的初始发展阶段给予执行机构的支持,鼓励执行机构在该阶段过程中进行过程性评估。

• ISE 计划管理

ISE 计划应为今后的申请者更清楚地范定它认为必要或具有高度优先的非正规科学实践。NSF 甚至可以提供某些做法的实例或建议,它们最适合于完成该计划的各个单项目标,而不是集中于整体目标的实现。NSF 也希望能够确定某些达到未被充分代表群体的实践,找到能更好地量测这类项目效果的方法。项目的质量控制是该计划极力鼓励的另一事项,可通过广泛利用咨询委员会和其他的技术评审委员会来实现。几乎每一个被调查的项目都以各种方式采用了专家咨询委员会。通常情况下,会成立多个咨询小组,既有机构内的,也有机构外的。很多情况下,不同的咨询小组处理项目的不同的工作,如教材等。咨询委员会的其他活动培养了促进项目实施的合作。有时通过“平权小组”等手段,咨询的作用是要帮助提高未被充分服务群体接受非正规教育的机会。

与 ISE 计划管理有关的另一个重要调查结果是评估问题。

尽管过程性评估已成为几乎每个项目的组成部分——如果没有 NSF 的鼓励,就不会进行这种评估——总结性评估则只到现在才被制订和执行,这两种评估今后应得到更大的支持。具体说来,评估中需要增加成年人和儿童参加或观察非正规科学教育活动的人数以及他们参加这类活动后所产生的短期和长期结果。再者,有关少数民族和其他未被充分代表的群体的数据需要更清楚,以便更好地评估项目在这方面的影响。

调查发现的另一个不足是 ISE 项目提交的最终报告的内容。最终报告往往仅论述项目管理方法或其他做法,却没有指出 SMET 教育中的问题和结果。报告中找不到有关参加者的数据,形成性评估所获得的教训,由此而带来的活动或产品在质量上的改进,以及为确保在资助期之后继续进行项目活动所需采取的对策。ISE 计划最终报告指南应该强调包含结果,而不仅是过程数据。

总之,ISE 计划应继续秉承现在采用的很多好的作法,如采用咨询和技术委员会控制项目质量,鼓励对项目进行评估,特别是总结性评估,重视少数民族和未被充分服务群体的作用。即使 ISE 计划仅对一项大活动的发展或生产阶段提供资助,也应探索进行跟踪调查的方法以测定项目后期的结果。最后,NSF 应更详细地规定对 ISE 计划项目最终报告的要求,报告中要有表明项目是否达到目标的有关数据。计划官员也应该开发将最终报告存档、采用报告数据判定未来的项目竞争和回答政策质询的方法。

9. ISE 计划投资值得吗？

美国的未来要求我们对人民、制度和思想进行投资，而科学正是这种投资必不可少的组成部分，是可带来非凡红利的、无止境的可持续资源。

——克林顿总统及戈尔副总统，摘自《科学与国家利益》

表征 NSF 在 ISE 计划上的投资成功与否的最有力的证据或许就是它提高了人们接触非正规科学教育的机会这一显而易见的结果。也就是说，有越来越多的人能越来越方便地接触科学。通过提供涵盖多种学科的各种活动，达到不同类型的受众群体，以及实现与其他组织的有益的合作，ISE 项目已使人们对科学及相关领域的态度有了积极改变。

ISE 资金已被很好地用于引导项目加强与各种机构的合作和更正规地采用评估以提供形成性和总结性反馈意见。在如下所有方面——博物馆和科学中心的作用，合作的重要性以及评估的重要性——NSF 资助的 ISE 活动都表现出与非 NSF 资助的活动的不同。

对 NSF 资助的 ISE 项目的现场调查表明，很多项目将资源用于开展丰富多彩的主题和活动，进一步延伸了 NSF 的投资和受众范围。NSF 资助的展览项目加入了教学用具、可以带回家的资料或电影等内容，而广播、电视和电影等媒介类项目中同样包含了提供书籍、因特网服务、CD-ROM、培训材料和手册。多样化活动产生了两个重要效果：活动逾越了家庭、学校和非正规科学教育场馆等不同机构间的界限；提供不同类型的学习和适合不同年龄受众参加的各色活动，也就是说让项目“有适合每一个人参加的活动”。

很多项目还具有跨学科性，这种方式有助于强化年轻人对日

常生活中科学相关性的理解,比如:有一个孩子在看完一集“生活中的数”之后说:“数学很重要,它无所不在。”这种方式也有助于参加者认识到,其他学科是与科学相关的,是进行动手型学习的机会。同时,这种方式与 NSF 教育与人力资源局的其他计划及 2000 年目标计划所进行的教育体制工作相辅相成,其中,NSF 的 2000 年目标谈到要在数学、科学、历史和英语等传统学科中实行综合性的学习。

在 NSF 资助的 ISE 项目中,随处可见的另一个主题精神是它们重视以尽可能多的方式包含多种群体,如焦点小组座谈的参与者、电影电视人物、展览画宣传对象和项目参加者,这也是推动 ISE 计划获得成功的因素之一。例如:“阅读彩虹”系列剧塑造了一个黑人男子形象,“神奇的学校汽车”系列剧塑造了一个富有创造性的女教师以及一群不同种族的青少年。这些角色打破了关于对科学有兴趣的人的定型,并证明科学贯穿于我们的日常生活之中。通过丰富人们对科学的看法,让学生们进行交互性、自主性科学探究活动,这些项目使青少年和成年人能构想他们自己从事科学工作的情景。SERIES 项目特别提到了这样一件事,据参加焦点小组座谈的青少年称,当那些参加了 SERIES 项目的年轻女学生被要求画一幅科学家的画像时,她们都画的是她们自己。

在推进合作的过程中,ISE 项目扩大了资源,融入了其他合作组织的多种多样的教育材料及观点。这使科学跨越了更多的界限,拥有了更多的观众。NSF 资助的活动中包括很多展览联合体,它们自然而然促进合作,发扬各种项目活动和学科中所体现的发展和多样化的精神。最后要说的是,NSF 通过总结性评估和散发教材活动,保证了项目的教材及其经验教训留存下来,对非正规科学教育界和未来的项目发挥指导作用。在这方面 NSF 的战略获得了成功。例如:NSF 资助的展览联合体向有关组织散发《可汲取的教训》书籍。一些新的或最近完成的项目已采用或计划采

用总结性评估。然而,在很多 ISE 项目中,这种评估是不完整的,或只受到一定程度的重视。因此,该计划在这方面可能需要进一步的监督和发展,只有这样,才能确保项目进行总结性评估,确保结果是完整的,更有益于整个 ISE 计划的实施。

附件二* 美国国家科学基金会非正规科学 教育计划访评报告(2001) 评估执行机构 :访评委员会(COV)

目录

引言

A. ISE 计划工作流程与管理的廉正效率评估

1. ISE 计划采用价值评议程序的效果
2. ISE 计划采用 NSF 价值评议标准的情况
3. 项目评审人员的选择
4. 经评审产生的受资助项目集

B. 成果评估 :ISE 计划投资的产出与结果

5. 战略目标 :人力方面
6. 战略目标 :创新思想
7. 战略目标 :工具方面
8. 重点领域
9. COV 认为 ISE 计划需要改进的方面
10. COV 对 ISE 计划实现本身目标的绩效评价(非 GPRA 成效评价)

* 附件一和附件二都是针对美国国家科学基金会非正规科学教育计划做的评估报告,但附件一的评估报告是评估机构按照该计划本身确立的目标进行的评估,而附件二则是访评委员会按照 NSF 依美国《政府绩效成果法》确立的基金会工作评估目标所做的评估。两个评估报告的视角不同,有不同的参照价值,在此我们一并译出。

11. 对 COV 评估过程、形式和核心问题的反馈意见

C. ISE 计划官员向访评委员会提出的补充问题

D. ISE 计划官员对访评委员会评估报告的反馈意见

引 言

评估所采用的资料

非正规科学教育(ISE)计划人员将其资助项目按领域分为三类:①博物馆类项目;②媒体类项目;③青少年和社区类项目。根据拨款类型又分为:标准拨款、滚动拨款、规划拨款、会议拨款以及探索性研究小额拨款(SGER)等。

在评估中,COV 研究了 ISE 项目官员提供的全套资料,包括:

- 专门为 COV 编撰的有关 ISE 计划管理的详细数据,包括各年度资助项目的统计、资助项目按计划领域的分布,资助项目主持人的情况、否决和撤销的项目、追加和补充拨款项目以及项目申请受理(“驻留”)时间的统计。

- ISE 计划的评估报告,COV,1998。
- NSF 工作人员及 COV 成员指南,2001。
- NSF 小学、中学与非正规教育战略规划(1995—2000)。
- 小学、中学与非正规教育年度报告,1998、1999、2000。
- 近年的小学、中学与非正规教育计划项目征集与指南册。
- NSF 非正规科学教育计划评估报告,COSMOS 公司,1998。

- ISE 计划效果评议报告,1999。
- Fastlane 通讯评审指南。
- 90 份项目申请报告的样本。这 90 份申请报告是 ISE 计划的工作人员从 1998、1999 和 2000 年的所有申请报告中随机抽取

出来的,其中 45 份最终获得资助(三类项目中每类 15 项,平均每年每类项目 5 项),45 份被否决或撤销(三类项目中每类 15 项,平均每年每类项目 5 项),并且包含了各种资助类型(标准资助、滚动资助、规划资助、会议资助以及探索性研究小额资助)。

- 由 ISE 计划各子领域通过 Fastlane 提交的全套计划总结。
- 2000—2001 年举办的一系列 ISE 外延会议的简短报告。

COV 会议过程

NSF 教育与人力资源部的一些工作人员,包括 NSF 助理主任办公室的 Stutsman 博士以及教育与人力资源部中小学与非正规科学教育处的 Kahle 博士和 Vanski 女士,欢迎并接待了 COV 成员,强调了 COV 为教育与人力资源部所做工作的重要性。教育与人力资源部研究、评估与交流处的 Johnson 博士与 COV 探讨了政府绩效成果法的相关问题。Johnson 博士指出,像以往的 COV 一样,我们需从以下四个因素入手,评价 ISE 计划工作流程和管理的效果和效率。这四个因素是:

- 价值评议程序的采用
- NSF 价值评议标准的采用(学术价值及广泛的社会影响)
- 项目评审人员的选择
- 通过评审产生的资助项目集

Johnson 博士后来又专门提到政府绩效成果法最近的修改以及 NSF 在其 2001 年绩效计划中最新提出的 ISE 计划绩效目标。她特别指出,中小学与非正规教育处以及非正规科学教育计划的官员希望 COV 能根据 ISE 计划 1998—2000 年间的执行结果,从总体上评价 ISE 计划是否在下述一项或多项指标上取得重要成就,以此判定 ISE 计划是否成功。这些指标是:

有关“人力”的成效指标

- 是否提高了美国幼儿园-12 年级的学生以及所有年龄的公

民的数学、科学和技术技能,使得他们在技术社会中具有竞争力。

- 是否为公众提供了获得和了解科学工程领域研究与教育的途径。

有关“创新思想”的成效指标

- 是否形成了有助于把发现与创新、学习和社会进步联系起来的合作伙伴关系。

- 研究和教育过程是否相互协同。

有关“工具”的成效指标

- 是否充分利用因特网形成了网络连接,使得科学、数学、工程和技术信息能为广大公众所获得。

NSF 中小学与非正规教育处负责非正规科学教育计划的 Deman 先生向 COV 阐述了 ISE 计划的主要目标,这些目标是:

- 增进广大公众对科学、数学和技术的了解;

- 使越来越多的青少年——特别是科技教育上未被充分代表的群体(如少数民族、女孩和身障者)以及教育低下地区(如农村社区)的青少年——感受科学、数学和技术(SMT)的奇妙,并促使他们在校内和校外进行这样的活动;

- 把非正规科学教育计划和活动推广到目前还没有或很少开展此类活动的较大的地区,如农村地区和城市内的贫民住宅区;

- 通过提高儿童及成人的科技素养,使其了解科学、数学和技术对其日常生活的重要性,从而丰富他们的生活质量,促使他们追求在这些领域获得更多的体验,帮助他们在掌握正确信息的基础上,对具有社会意义的政策问题做出负责任的决策。

Deman 先生还指出,就特点来说,ISE 项目力求做到:

- 提供除正规课堂环境之外的其他丰富的、有激励性的科学学习机会,由此提高各种年龄、兴趣和背景的人对科学、数学及其应用的欣赏和了解;

- 发生在多种多样的环境中,如博物馆、动物园、植物园和社

区组织,并且利用多种传媒手段,如广播、电影、展览、交互性技术以及印刷等;

- 促进非正规和正规教育之间的联系,从而推动和改进各种学习环境下的创造性的 SMT 教育;

- 激励家长和其他的成人成为 SMT 教育的有力支持者,从而推动正规和非正规科学教育提高质量和普及性;

- ISE 项目或者服务大量的受众或者能在国家或地区的范围产生重大影响;

- 促进具有类似目标的机构之间形成合作,特别是在项目沟通非正规和正规教育界的情况下。

Demam 先生还指出,ISE 计划办另外提出了 5 个“非正规科学教育补充问题”供 COV 考虑。NSF 中小学与校外教育处的项目官员 James 女士对提供给 COV 的信息类型做了额外的解释。

在准备分成小组开会时,COV 决定各小组要评估规定其做的所有 30 个项目。COV 解释说,采取这种广评方法而不是精评少数项目,会有助于 COV 更好地回答 ISE 计划提出的有关补充性问题,即未来采取什么样的优化方法设计 ISE 项目集。

在整个评估过程中,当 COV 要求提供额外的材料时,ISE 计划的工作人员均给予了迅速和有效的回应,并且积极参加了 COV 在 2001 年 4 月 3 日为中小学和校外教育处 ISE 等计划的官员召开的非正式反馈会议。

A. ISE 计划工作流程与管理的廉正效率评估

1. ISE 计划采用价值评议程序的效果

a. 总体设计

从总体看,ISE 计划价值评议的设计有很强的效力。从我们

评议的项目卷宗看,ISE 项目文件资料完备,评审人员的评审意见详尽、有思想性,而且是立足于帮助项目主持人成功地实施项目,或者当项目申请被否决或驳回时,会帮助项目主持人重新提交更有说服力的申请书。ISE 项目官员的评价意见同样是建设性的,看得出是经过了认真的思考。ISE 计划评审过程的效果应当作为其他价值评议过程的范例。

显然,ISE 项目官员所扮演的角色对于 ISE 价值评议设计过程非常重要。有关卷宗的证据表明,ISE 项目官员的工作是卓有成效的。

b. ISE 计划项目评审过程的效果

总体上看,ISE 计划项目评审过程是杰出的,其成效要归功于 ISE 官员的工作能力。由于 ISE 官员能够将评审小组的意见通过晤谈的方式转达给项目申请者,使项目设计与 ISE 计划目标更趋一致,从而使 NSF 的资助能取得最佳效果。看上去获资助的项目都有很高的质量,由于均响应 ISE 计划项目征集书的要求(RFP)而与 ISE 计划的目标一致。当专家小组评审意见不一致时,项目官员积极地收集了澄清信息和补充性的评审意见。对于被否决的项目,评审过程的结果也是可靠的、适当的而且文件证据齐备。

我们建议,在新项目官员上任培训中,应特别强调,项目官员要做的一项重要工作是:通过与项目申请人进行晤谈,有效地将评审小组的意见传达给项目申请人。

从项目申请的类型考虑,青少年与社区分委员会提到项目评审过程存在以下差异:

正式的专家小组评审:ISE 计划在为征集的项目组织正式的专家评审时,评审过程是严格的,而且文件证据齐备。评审小组包括了各种观点和专长的专家。评审小组受到 NSF 征集项目书、ISE 项目官员和 NSF 评审标准的适宜引导。项目官员记载的专

家评审意见概要都是对专家小组所写意见的极好的提炼。ISE 项目官员与项目申请负责人的通信联系和晤谈也都有文件记载,是支持项目发展的财富。不过,有些明显的例外情况:在项目资助批准之后,很少有证据表明 ISE 项目官员与项目主持人再有什么联系或对项目实施了监督(例如项目官员的实地考察)。当然,我们发现也有例外,例如,在项目官员对项目资助增加了规定性条款或进行了实地考察的情况下,后续工作做得很好,而且有完好的记载。

预申请:ISE 项目预申请的评审由 NSF 内部两名或两名以上项目官员进行。预申请的管理高效,做到了及时顺畅地向项目申请人提供信息。

探索性研究小额资助(SGER)和会议资助的申请:SGER 和会议经费的评审过程不像预申请评审那么严格。根据 NSF 的规定,这类申请仅需一名项目官员评审即可。不过,我们查看的这类申请的卷宗过于简明,申请人和项目官员之间的交流并没有完全记录下来。尽管我们承认项目官员的工作时间已很紧张,但我们还是敦请 ISE 计划再增加一名项目官员来做 SGER 项目申请的评审。增加一人评审有助于提高有价值申请项目推荐的可信度。

c. 效率、决策时间

COV 审查了 ISE 项目申请的递交日期、评审日期、项目官员与申请人的交流讨论日期以及项目官员进行项目分析的日期。总体上看,1998—2000 年的三年期里,ISE 计划的管理效率提高了,从专家小组评审到做出资助决定的时间缩短了。有很多例子表明,ISE 项目官员对项目申请的答复是迅速、高效和认真的,申请否决通知在评审的两个月内即发给了申请人,而且这种高效是一种规范,而不是例外。

尽管我们对 ISE 计划在缩短决策时间方面做出的成功努力

给予肯定,但这一过程尚有需改进之处。某些问题出在 ISE 项目官员提出项目建议之后。对许多受资助者来说,从推荐到收到正式通知,其间的时间滞后严重,给他们带来的负担实际上加大了项目的相关成本。

我们还注意到,因一位 ISE 项目官员(现已离开中小学与非正规教育处)工作不认真,造成了这样的问题。例如,由于这位官员的原因,经青少年和社区分委员会评审否决的 15 个项目申请中,有 3 个过了好长时间才收到否决通知,有一项获批准的申请也拖延了很长时间。而同样是 5 万美元的计划项目的评审,另一位项目官员只用很短的时间就完成了。为避免此类情况再次发生,我们建议,ISE 计划在对新来的官员进行培训时,应向他们明确提出效率标准,并建立持续监督的程序。

d. 项目官员提出资助项目建议所具有的文件证据的完备性

ISE 计划的官员在文件记载方面一贯做得很好。多数卷宗的文件记载可以称得上模范,其中一些尤为出色。当评审过程中出现异常情况时,项目官员会将相关信息记录下来。例如,如果一个即将上任的项目官员曾参加过专家评审小组,另一位项目官员就会把将如何处理这件事的过程记录下来。有些项目官员把电话联系的情况也记录了下来,这一点我们建议所有人都效仿。

e. 与 ISE 计划项目征集、通告和指南中规定的优先领域、标准的一致性

我们发现,ISE 计划的项目征集、通告和指南非常一致,都明确阐述了该计划的优先领域和评审标准。价值评议程序始终贯穿计划的优先领域和评审标准。同样重要的是,项目官员的所有口头谈话也始终恪守 ISE 计划的评审标准和目标。

由于项目官员在价值评议过程中起着至关重要的作用,COV

担心,一名超负荷工作的项目官员可能没有时间与资助者进行早期的协商,并在项目进行过程中进行不断的监督。

2. ISE计划采用NSF价值评议标准的情况

a. 绩效目标:评审人员执行价值评议标准的情况

(1) 评审人员在评审中是否充分考虑了“学术价值”这一标准?

COV发现,所有评审人员都认真考虑了“学术价值”这一(首要的)标准。

(2) 评审人员在评审中是否充分考虑了“更广泛的社会影响”这一标准?

COV发现,总体上看,评审小组在如何充分、有效地考虑“更广泛的影响”这条标准(第二条)方面很不一致。有些评审小组的每位评审人员都考虑到这一标准,而有些评审小组中只有一人甚至没有人考虑这一标准。这个问题由于以下两个因素而变得更加严重:

第一,有些评审人员的评审意见是将两条评审标准融在一起讨论得出的,没有明确区别结果有多少是根据某一条标准得出的。例如,在针对一项申请的评审中,评审人员虽没有明确提及第二条标准,但提到一些与“更广泛的社会影响”有关的问题。

第二,一些评审人员似乎没有认识到,对未充分服务群体的影响是NSF确定第二条标准的一个重要因素。

COV发现,在查看的所有卷宗中,约有四分之三的卷宗的评审小组至少有一位专家考虑了这条标准。我们估计,不到一半的专家在评审中直接提及了这条标准。不过,有些证据表明,在报告的过程中对第二条标准的重视提高了。我们认为,第二条标准对于实现ISE的目标是必不可少的,因此,我们对这条标准没有得

到一致的重视表示关注。COV 青少年和社区项目分委员会特别关注了所查看的卷宗中评审人员是如何考虑第二条标准的。

b. 绩效目标：项目官员执行价值评议标准的情况

(1) 项目官员在评审中是否充分考虑了“学术价值”这一标准？

COV 发现，所有项目官员都认真考虑了“学术价值”这个(首要)标准。

(2) 项目官员在评审中是否充分考虑了“较广泛的社会影响”这一标准？

项目官员在评审中对两条标准都考虑了，他们往往尽力弄清楚专家评审小组的成员对第二条标准未直接提及或重视不够的情况。

c. COV 对 NSF 价值评议体系的关切

COV 评估过程中掌握的信息没有清楚地展现 NSF 项目官员是如何向专家评审小组介绍价值评议标准的。这一点值得注意，因为在评审人员考虑“较广泛的社会影响”这一标准时，搞不清楚他们脑子里是只有所达到的受众数量，还是同时考虑了未充分代表的群体或者教育欠发达地区。

根据我们的评估结果，我们建议，在对评审小组成员的指导时，ISE 项目官员应认真地向他们解释第二条标准的含义，特别要求评审人员对每一条标准都要给与切实的重视，而且要明确规定，评审人员只能将那些按两条标准评价均靠前的项目申请排在前面。我们特别建议，在评审过程的一开始就将第二条标准的全文提交给评审专家审读。

3. 项目评审人员的选择

a. 采用足够数量的评审专家 以期取得平衡的评审结果

我们发现 ,各个 ISE 项目评审小组的成员人数合适 ,足以取得平衡的评审意见。

b. 选拔具有一定专长 /资格的评审人员

我们发现 ,评审人员的组成很均衡 ,他们具有教学和社区教育领域多样化的专业知识和专长。COV 认为 ,在某些情况下 ,专家评审小组中如能增加一些科研人员 ,评审小组的构成会更合理 ,但应注意的是 ,参加 ISE 评审小组的科研人员必须同时具备 ISE 领域的专门知识。总体上看 ,从评审人员提出的非常好的评审意见可以明显地感到 ,ISE 评审人员确实具备了足够的专门知识和应有的资格。例如 ,评审人员指出在一项申请中存在的问题 :项目可能会对目标受众中的某些人群形成文化冒犯或不合。

c. 评审人员的选拔考虑了地区、机构类型和未充分代表的群体的平衡

首先 ,我们对 ISE 计划官员在平衡我们 COV 专家构成方面的卓越表现表示赞赏。COV 成员构成的平衡与我们以前作为 ISE 计划评审人员的经历是一样的。

我们对项目卷宗的评议显示 ,ISE 评审人员在组成上实现了性别、机构规模和类型(博物馆、大学、媒体、社区或青少年组织)以及地理分布等的平衡。尽管我们根据自己以往作为专家评审小组成员的经历知道 ,ISE 项目专家评审小组包括来自不同种族的成员(但没有包括残障者) ,但从卷宗中的信息我们无法判断评审专家属于哪个种族、是否残障或其语言方面的背景。COV 认识到 ,关于评审专家的个人资料的收集和提供可能涉及到法律方面的问

题,因此,我们建议 ISE 项目官员应要求评审专家们自己提交一段描述其个人背景和专业知识等情况的文字。这种信息会大大有助于专家小组的评审过程,因为它可以确保和记载下评审人员的多样化和组成的平衡,进一步促进不同观点的交流,这种交流使评审过程富有价值且有利于推出好的项目。

d. 一旦出现利益冲突,NSF 官员应能认清并妥善处理,且作充分的文字记录,以说明所采取的措施是正确的

单从项目卷宗提供的证据,很难评价 NSF 官员对利益冲突问题是否关注了。不过,根据我们自己以往参加项目评审小组的经历,我们知道 NSF 的官员是非常关注这些问题的。

4. 经评审产生的受资助项目集

a. 总体的科学或工程性

获资助项目的科学质量非常高,表现出对科学内容适当性的严格关注。不符合科学/工程标准的那些申请都或被撤销或被否决了。

b. 资助范围、规模和期限的适当性

我们评价后认为,ISE 计划资助的项目在范围、规模和期限上是合适的。不过,要使项目的总结性报告更有效地指导 ISE 领域,许多 ISE 项目的资助期需延长。这样说是由于我们注意到,NSF 的项目征集书有这样一种倾向,即越来越重视各项目总结性报告中有关项目获得的成效,而多数项目的资助期太短,项目负责人不能在此期限内提供符合 NSF 项目征集书的有说服力的总结性信息。因此,如果今后的项目征集书继续有这种期许,我们认为,应延长个别项目的资助期。

c. 对新机会的有效识别和支持

从近年 ISE 计划为支持农村和小社区的项目申请所做的努力来看, ISE 计划对新机会的支持是值得赞赏的。而且从长期的新机会看, 某些新机会(如电子网络的创新应用等)也在得到 ISE 计划的资助。不过, COV 发现, 当考虑 ISE 计划对近期的新机会的支持时, 事情就变得较复杂了。我们在后面讨论在资助项目集中平衡考虑对高风险和创新项目的资助时将再涉及这个话题。

d. 通过资助新的项目申请人等措施, 适度保持体系的开放性

当评价 ISE 计划在保持资助体系开放性方面的表现时, COV 各分委员会的看法不尽一致。博物馆分委员会认为, ISE 计划在这方面基本上是成功的, 他们提到了 ISE 计划对新机会开放的一些例子。媒体分委员会认为, ISE 项目官员利用“规划资助”保证了资助体系对新的项目申请人的开放。他们提到, 在媒体领域, 较大的全国性媒体机构因拥有先进的设备和设施, 能提出大型的媒体节目项目计划, 因而能更经常地获得资助。但青少年和社区分委员会在 ISE 计划对青少年和社区项目中, 几乎看不到对新的项目申请人资助的证据, 他们指出目前的资助分配很大程度上向那些长期获得 NSF 资助的机构倾斜。

e. 有什么证据表明, 申请者在项目申请中注重研究与教育的结合

为使我们的评议方式适合非正规教育领域, COV 决定从以下三个方面界定“研究”: ①项目利用和依靠已有的研究和评估基础的程度; ②ISE 项目/活动结合新的科学研究内容的程度; ③有关人们学习方式的新知识在非正规科学教育方法和战略中的应用程度。我们从上述各个方面对项目卷宗进行了评估。

- 有很多但不是所有的 ISE 项目成功地实现了科学与教育的结合。有些项目申请书中引用的研究和证据基础非常弱。尽管这可能部分地是由于规定的篇幅有限,但 COV 担心,这些 ISE 项目不能很好地学习利用已有的知识基础,最终也不能为新的知识基础做出贡献。下面我们会更多地探讨这个问题,因为它可能与缺少对项目的成效评估有关,而且关系到如何应对这样的挑战:即重视并投入更多的资源,在整个 ISE 领域构建一个可靠的研究和评估基础。

- 我们注意到,很少有证据表明新的科学研究已被融入到 ISE 计划和活动的内容之中。

- 尽管 ISE 项目基本上都是建立在充分了解人类发展和学生各方面的发展的基础之上,但我们注意到,这些项目很少利用了有关人们学习方式的最新研究成果。这样讲的同时,我们承认,只是在我们评估的这三年期的后期,国家才开始注意将认知和学习与教育计划和教育战略联系起来。

f. 有什么证据表明,来自未充分代表的群体的项目申请量增加了

从提供的 ISE 项目卷宗看,COV 分委员会不能充分评价来自未充分代表的群体的申请比例随时间的变化情况。但我们确实注意到,在递交申请的项目主持人中,绝大多数是白人,而且主要是男性。我们发现,只有一项来自少数民族服务组织的申请(一项规划申请)获得了资助。不过,COV 博物馆分委员会发现,ISE 计划资助的各个博物馆项目都包括了对如何达到和服务于未充分代表的群体的探讨。

g. 项目的平衡

以下我们要谈的是 ISE 计划为获得平衡的项目集所应考虑

的各种要素。

• 高风险。COV 注意到,对“高风险”资助的定义有很多。在传统的研究中,那些有效地提出高风险、高收益创新思想的申请通常能获得资助,因为它们产生的积极成果会大大丰富人类的知识结构。但在 ISE 领域,由经验丰富的机构或项目主持人提出的确有创新性的项目有可能不被认为是“高风险”的(由于他们以往的业绩记录),而同样的项目如果由经验不多的机构或项目主持人提出就可能被认为是高风险的。最后我们发现“高风险”的含义是随着时间变化的,就在不久前,整个 ISE 计划还被认为是一个“高风险的”举措。

按照这样的定义,COV 媒体分委员会几乎没有发现 ISE 计划资助高风险媒体项目的证据。这在某种程度上是可以理解的,因为与传统的研究领域相比,非正规科学教育领域更难判断项目是否能取得“高收益”。COV 博物馆以及青少年与社区分委员会也很少发现获资助的高风险或创新性项目的例子。他们注意到,问题不是出在 ISE 计划对好的高风险项目申请做出了不应有的驳回,而是由于申请者中很少是高风险项目申请者或者被公认具有高风险创新思想的执行者。

鉴于非正规科学教育的性质,COV 认为,对于创新思想超出传统 ISE 活动边界的项目申请应慎重考虑,而对新的项目主持人或机构提出的项目申请应大胆资助,因为这是拓展 ISE 领域、培养一批有能力从事这项重要工作的专家的一个重要途径。在这方面,我们担心 ISE 计划是不是在发出一种含糊的信息,因而使新的申请者却步。也就是说,目前的申请程序要求申请者证明他们已经具备启动一个 ISE 项目的能力(包括实施经验)。如果是这样的话,那些原本有前途但目前尚缺乏经验的新的项目主持人(或者正在考虑全新项目的项目主持人)可能会过低地估计其成功的把握,以至于放弃努力。

我们建议,ISE 计划继续执行其战略,鼓励、培植更加广泛和多样化的潜在申请者更多地参加到 ISE 计划中来。目前,ISE 计划为鼓励较小社区的参与而做的工作会增加“高风险”项目的数量。为进一步做好上述工作,我们建议该计划集合非正规科学教育领域的意见,对非正规教育情景中的创新和高风险做出更好的界定。

- 多学科性。关于 ISE 项目集是否很好地兼顾了多学科活动,各分委员会的评价结果不一样。媒体和博物馆分委员会认为,ISE 计划在这方面做得不错。而青少年和社区分委员会虽然看到 ISE 计划有许多将科学和素养结合起来的项目,但他们发现很少有项目是探究多种科学学科的,或注重学科间的传统界限在发生怎样的变化。

- 创新性。我们这里说的“创新性项目”是指针对以前不曾达到的新人群或重要人群而尝试的探索性创新思想,或是新的创新思想、策略或模式。经评判,我们认为 ISE 资助项目集兼顾了这方面的平衡。例如,一个探讨非正规科学教育中性别差异的项目(项目号 ESI 98-15021)采用创新的方法研究男女在教育上的平等和参与问题。它也是一个促进项目交互滋长的例子,因为性别差异问题在圣何塞儿童博物馆的一个项目以及家庭科学学习项目中也涉及到了。

B. 成果评估 :ISE 计划投资的产出与结果

5. 战略目标 :人力方面

目标 造就一支具有国际竞争力和全球参与能力的多样化的劳动力队伍,其中包括科学家、工程师和受过良好教育的公民。

非正规科学项目的目标是要培养有科学素养的公众,找到培

养有科学素养的公众的方法,所以该计划应明确地支持这个战略目标。

2001 财年的绩效目标 这一期间公布的成果如能证明 ISE 计划总体上实现了以下一个或多个指标,该计划就被认为实现了其在人力方面的战略目标。

a. 提高了美国幼儿园-12 年级学生的数学、科学和技术能力

COV 认为,利用该成功度指标来评估 ISE 计划的成功是不恰当的。然而,获资助的 ISE 项目大都通过各种形式与正规教育发生密切联系,例如为学校提供课程指南、教师培训活动和材料以及符合国家科学教育标准的辅助性学习材料。

b. 提高了所有年龄公民的数学、科学和技术能力,使他们在技术社会中有竞争力

COV 认为,在这一指标上,ISE 计划成果显著,取得了成功。例如,一些 ISE 项目,如蒙特什尔科学博物馆的 TEAMS 巡回展览协作项目和 Rutgers 大学的荒野公园科学教育项目,通过研究如何实现非正规环境中的家庭学习,并通过让家长参与孩子的科学和数学教育来提高所有年龄的公民的 SMT 技能。一些面向少儿的 ISE 项目也成功实现了这一指标,ZOOM(WGBH 教育基金会,媒体项目)和 BUSYTOWN(俄勒冈科学与工业博物馆,博物馆项目)是其中的两个典型例子。提高了各年龄段公民数学、科学和技术能力的其他 ISE 项目还包括:

- ESI99-01923(空间科学探索节目,媒体项目,Soundprint 媒体中心制作)对该节目进行的焦点群体研究评估表明:

- 听众把从该节目获知的信息与自己的生活体验联系起来;
- 教师发现该系列广播节目不仅适于课堂使用,还适于学生独自使用;

一听众反映,公共广播必须起的一个作用是要向公众提供通过其他媒体不能获得的各种话题的深度信息。

- ESI99-01975(智能无限广播系列节目,媒体项目,Lechtenstein 创新思想媒体公司制作):该节目目前正在全国 150 个公共广播电台播出。

c. 体现美国社会多样性的科学、技术和教育工作者队伍

COV 认为,采用该成功度指标评估 ISE 计划是不恰当的,不过,该计划资助的一些项目与这一指标有关。例如,在下面“创新思想”这一绩效战略目标中提到的天文学习项目(太平洋天文学会开展的一个青少年与社区项目),由于它在多语社区中开展工作,编写了双语材料而且针对的是少数民族集中学区,因此它也是实现 NSF 第二条标准所述原则的例证。此外,新英格兰水族馆的“大海的声音”项目也是针对多样化学生并鼓励多样化社区参与科学事业的一个范例。通过该项目,学生和教师能探究科学概念,了解各种各样的科学职业,而这在过去并不属于幼儿园-12 年级学生应有的经历。该项目为如何使视障学生和成人体验动手性科学树立了一个全国样板。该项目每年通过教师资源中心为城市学区的 15000 名教师提供课程材料,这些材料在该项目资助期过后仍在使用。该项目留下的财富是 4 页的“先进组织者”名录以及分发给 1 万多名教育者的、用以指导学生在家里、校内及校外进行海洋科学探索的活动指南。

d. 促进世界最杰出的科学和工程专业人员参与进来

COV 认为,不应用这个成功度指标去评估 ISE 计划的成功。不过,该计划资助的一些项目与这一指标有关。例如,它资助 WGBH 教育基金会主办的世界科学节目制片人大会(媒体项目)将世界优秀的科学节目制作者召集在一起,交流专业知识和经

验。

e. 公众能有机会接触科学工程领域的研究和教育，享受其福祉

COV 认为，ISE 计划已成功地证明在这一指标方面取得了重大成果。我们引用各种项目来说明 ISE 项目是怎样让公众有机会接触科学工程领域的研究和教育并享受其福祉的。

- 根据“比尔·奈科学人”电视节目的评估结果(KCTS 电视台制作 媒体项目)，在观看了该节目后：

- 94% 的受访学生利用科学工具包做过至少一次实验；

- 在收到科学工具包后，孩子们更多地提出进行科学活动，家长也知道从哪儿能找到科学信息和资源了。

- 美国微生物学协会制作的关于微生物知识的四集电视系列节目(项目号 ESI96-14737)每周有 160 万个家庭收看。到 2000 年 6 月，该项目出版的动手活动出版物“Meet the microbe”已在全国发行了 2 万本，另有 17000 项活动被人们从 [www. microbe-world.org](http://www.microbe-world.org) 网站上下载。

- WGBH 教育基金会制作的“ZOOM”电视节目(项目号 ESI 96-14743)包含以下评估结果：“通过收看 ZOOM 节目，孩子们在学习科学内容和过程，他们对 ZOOM 节目中的一些具体科学概念的了解发生了很大改变。”

- 史密森博物院研究院制作的“加拉帕高斯龟岛”立体电影(项目号 ESI 97-05298)是公众了解科学发现过程的范例——在拍摄过程中发现了 15 种新的脊椎和非脊椎动物，它是 NSF 媒体项目直接带来科学进步的例子。

6. 战略目标：创新思想

使科学和工程前沿领域的发现与学习、创新和社会服务联系

起来。2001 财年绩效目标 此期间报道的 ISE 计划成果如能从总体上证明 ISE 计划在以下一个或多个指标方面取得显著成绩 ,那么该计划就是成功的。

以下关于“创新思想”这一战略目标的绩效衡量指标与 NSF 中小学与校外教育处年度报告以及 ISE 项目征集指南中提到的目标几乎没有共同之处。我们查看 ISE 项目卷宗获得的证据非常支持“ISE 计划已成功地实现了该计划自身确定的预定目标”的判断。这里 ,我们强烈建议 ,应采用 NSF 中小学与校外教育处年度报告中阐述的 ISE 目标 ,作为 COV 将来评价“创新思想”这一战略目标的衡量标准。

a. 一个扎实且不断增长的基础知识基础 ,有助于增进所有科学与工程领域的进步

COV 认为 ,不应采用这一成功度指标去评估 ISE 计划的成功。不过 ,我们下面关于构建 ISE 领域知识基础的评估意见与这一指标有关。

b. 推进科学、工程和技术前沿的发现

COV 认为 ,采用这一成功度指标评价 ISE 计划是不恰当的。

c. 形成有助于将发现与创新、学习和社会进步联系起来的“合作伙伴关系”

COV 认为 ,ISE 计划成功地证明在这一指标方面取得了显著成果。有很多 ISE 项目可以作为 ISE 注重科学发现并努力建立教育、研究和科学家的协同伙伴关系的例证 ,这里我们特别选择“天文学项目联盟”项目 ,作为 ISE 计划在这方面工作的突出例子。该项目通过记录科学记录、创办学生研究刊物以及合作研究天体物理现象等各种各样的活动 ,将业余和专业天文学家、在校学生、教师及学生的家人联系起来。该项目不但使以前从未合作过

的群体间形成合作伙伴关系,而且它所创造的联盟体系产生了比资助期所确立的目标更多的新联盟。该 ISE 项目开创了公众科学协作的新创新思想,这种新创新思想正以各种形式在全国重现。

d. 研究和教育过程互为促进

COV 认为,ISE 计划成功地证明在这一指标方面取得了显著成果。我们发现,在取得这种成果方面,某些探索研究小额拨款项目(SGER)和研究活动追加科普拨款项目特别突出,因为这些项目使研究人员可以向公众宣传其研究工作。“南达科他矿物学院的南极化石”项目(项目号 ESI 98-15231,博物馆类项目)就是 ISE 计划努力实现这一“创新思想”目标的极好例子。其他一些项目也证明实现了研究和教育过程的相互促进,如 NSF 资助科学技术中心协会举办的“科学博物馆中的数学活动”(项目号 ESI 99-06982,研究探讨了科学中心如何鼓励观众参与数学学习)、旧金山探索馆的“记忆”展览(项目号 ESI 99-80498)和科学与工业博物馆的“遗传学解码”展览(项目号 ESI 96-14298),后两个项目在展览开发过程中都很好地融入了研究内容。

7. 战略目标:工具方面

提供最先进的、能广泛获取的信息基础和可以共享的研究和教育工具。2001 财年的成效目标:此期间报道的 ISE 计划成果如能从总体上证明 ISE 计划在以下一个或多个指标方面取得显著成绩,那么该计划就是成功的。

a. 有助于促进发现的共享平台、设施、设备和数据库

COV 认为,用这一成功度指标评估 ISE 计划的成功是不恰当的。

b. 有助于提高科工劳动力生产率和成效的共享平台、设施、设备和数据库

COV 认为 ,用这一成功度指标评估 ISE 计划的成功是不恰当的。

c. 充分利用互联网实现信息网络化和连接 ,使所有的公众都能了解科学、数学、工程和技术(SMET)信息

COV 认为 ,ISE 计划成功地证明在这一指标方面取得了显著成果。我们查看的项目卷宗显示 ,一些 ISE 项目成功创建了可广泛访问的网站 ,而且试图利用因特网的其他功能达到广大公众。例如 ,旧金山探索馆的“电子指南”项目(项目号 ESI 99-01985)正在探索新的方式 ,利用互联网使更多的公众获得信息。“提高微生物知识素养合作组织”(媒体项目 ,项目号 ESI 96-14737)是 ISE 计划利用互联网向所有公众提供科学、数学、工程、技术信息的极好例子。该项目创建的站点 www.microbeworld.org 第一个月就被点击 689620 次 ,访问人数达到 29779 人。该站点建立的头一年就被点击 300 多万次。

d. 有助于科学工程资源有效利用的信息分析与政策分析

COV 认为 ,用这一成功度指标来评估 ISE 计划的成功是不恰当的。但我们注意到 ,某些 ISE 项目产生的效果和影响与这一指标有关。例如 ,教育发展中心开展的“Access by Design”(项目号 ISE 96-27091)创造的工具有助于科学家、社区组织、技术供方、最终用户以及公共政策领导者了解并采取有效措施获得对技术的平等利用。该项目出版了 9 本小册子(供社区组织在制定利用技术为社区服务的计划时使用)、1 份研究报告(总结了技术利用的障碍及行动建议)、4 篇论文以及 1 份综合性的资源和信息指南。这些材料以及为编写这些材料召集起来的各个小组已经促成多家基

金会和社区服务组织继续开展合作并调整其做法,比如罗得岛印第安人委员会、全国城市联盟和本顿基金会。它们已经对 CEOS、Radcliffe 公共政策学院以及加州大学洛杉矶分校公共政策学院的公共政策对话产生了影响。

8. 重点领域

在以下各方面,判断 ISE 计划投资和可产生的成果能否在将来也表现出同样显著的绩效。

a. 战略成果:人力方面

ISE 计划设立的目的并不是致力于幼儿园-12 年级的系统活动、教育人员培养或教师的专业发展、近期的劳动力需求(先进技术教育计划)或者通过部族学院或创新伙伴关系计划(PFI)扩大各方的参与。

b. 战略成果:创新思想

我们在上面考虑了 ISE 计划的资助项目集合如何保持高风险、多学科或创新性研究的平衡的问题。ISE 计划设立的目的并不是针对 NSF 在信息技术研究、纳米科学和工程、环境中的生物多样性或数学方面的基础研究、功能基因组或认知神经科学等领域的具体研究计划。

c. 战略成果:工具方面

ISE 计划设立的目的并不是 NSF 要投资发展每秒万亿次的计算系统、重大研究仪器计划、科学工程信息/报告/数据库或新型科学数据库等设施 and 工具。

9. COV 认为 ISE 计划需要改进的方面

我们对 ISE 计划官员在项目评估和管理方面的出色表现给予赞扬。NSF 小学、中学和非正规教育处年度报告中描述的 ISE 计划的突出成就给我们留下了非常深刻的印象。我们建议 ISE 计划将这一报告公开出版,以使公众了解 ISE 计划对提高美国人的科技素养所做出的巨大贡献。

与此同时,针对 ISE 计划,我们提出下面一些问题,并就某些问题给出解决的建议。

- 我们认为,就任务要求来说,ISE 计划官员人数显然不够。其后果是,项目官员无法投入足够的时间,与受资助者进行项目初期的必要协商以及对进行中的项目实施连续的监督。
- 我们认为,ISE 计划的项目评审工作做得很好,这很大程度上应归功于项目官员的出色表现。然而这种出色表现不是必然的,应通过对新的项目官员实行上岗培训以及不断地进行工作勘漏继续保持和发扬这一业绩。
- 尽管 ISE 计划的操作符合 NSF 的有关规定,即所有的探索性研究小额拨款(SGER)申请都只由一名项目官员负责评审,但我们还是敦请 ISE 计划再多委派一名项目官员评审 SGER 申请,从而提高对这些申请提出可资助项目建议的可靠性。
- 我们认为,ISE 计划应制定新的程序,以确保评审人员更认真地重视“社会影响”这一价值标准。要实现这一改变,需要给评审专家更明确、更有力的指示,包括确立“唯学术水平和社会影响两条判据都位居前列的申请才能排在前面”这一新的评审标准。
- 由于很难断定评审专家的种族背景,我们建议项目官员应让评审专家提交一份介绍其背景和专长的个人简介。

- ISE 计划官员一旦提出资助项目建议 ,应注意缩短通知滞后的时间。
- 我们认为 ,ISE 领域缺少一种“发展模式” ,因此建议 ISE 计划办制定一项目征集书(RFP) ,以促进 NSF 在 ISE 领域的投资所获得的知识共享。这种知识共享对于 ISE 领域的能力建设是非常必要的。我们建议 ,将 SGER 资助用于进一步鼓励未被充分代表的社区的能力建设 ,具体详见下文的 C.1、C.2. b 和 C.2. c。

10. COV 对 ISE 计划实现本身目标的绩效评价(非 GPRA 成效评价)

正如本报告其他部分提到的 ,ISE 计划应对以下方面给与更多的重视 ,提供更多的资源支持 :

- 加强针对科学、数学和技术领域未被充分代表的以及未得到充分服务的人群和社区的科学教育外伸活动 ,加大这方面的支持力度 ;
- 加强评估计划的制定 ,并重视形成性评估和总结性评估结果的应用 ;
- 为第一次申请项目者(尤其是来自未充分代表人群的项目申请者)提供帮助。

11. 对 COV 评估过程、形式和核心问题的反馈意见

1. 由于下面的几个因素 ,限制了我们评判 ISE 计划战略成效目标成就的相关信息进行深入的评价 :

- 为让公众和决策者了解非正规教育在促进科学事业健康发展上的显著贡献 ,关键的一点是 ISE 计划必须按《政府绩效成果法》的要求证明其成效。然而 ,COV 评价模板的许多问题和指标并不适合 ISE 计划 ,而且我们担心 ,如果我们不回

答这些问题,会威胁 NSF 今后获得这一重要方面的持续支持。我们建议 NSF 重申 COV 评价模板中的问题和指标,或对其加以解释,使之与我们所评价的非正规教育事业的目标相一致。

- COV 评价会的规定时程不允许我们对 ISE 计划的成果和影响作出应有的系统评价和分析。要进行系统的评价,需要对项目最终报告、总结性评估报告以及成果进行认真和广泛的研究。我们之中前些年参加过 COV 访评的人发现,由于对 ISE 计划增加了 GPRA 模式评估以及不仅评价过程而且要评价成果的要求,访评的工作量较前些年增加了 1 倍多。访评委员会有足够的时间查阅卷宗,回答目录中 A 部分的问题,但我们需要更多的时间查阅项目最终报告和成果,以便系统地回答 B 部分的问题。我们建议将 COV 的评审时间从两天延长到三天。
- 可持续性和长期影响是 ISE 项目成果的重要方面,基于此,我们怀疑通常的三年项目期限不足以评估项目成果。

2. 我们发现,“效率,通知的时间”这一标准很难评估,因为卷宗中的文件资料显示,各种具体工作的联系时间有时是矛盾的或不清楚的。例如,有一个项目卷宗的表 1 中将“回复时间”列为“越快越好”,而在另一项目卷宗中,记载的项目申请的否决日期比申请递交的时间还要早。尽管偶尔的打字错误在所难免,但我们建议在卷宗的左边增设工作进程记录表,这样,这类错误就比较容易避免了。

3. 我们发现,按照 ISE 计划的三大资助领域将 COV 分成三个分委员会很有效,因而建议以后继续采用这个方法。根据我们的经验,我们还提请 ISE 计划办要特别注意保证向每个分委员会提交的项目卷宗是按资助领域正确分类的。

C. ISE 计划官员向访评委员会 提出的补充问题

1. 我们是否收集到 ISE 项目适当且充分的数据,使我们可以评估这些项目在多大程度上实现了其预期的目标和成果? 我们要在项目的最终报告和总结性评估报告中报告项目的哪些成果?

目前多数的 ISE 项目记载了其成果,而且许多项目提供了证明其成功的轶事式数据。但我们查阅的项目中只有少数几个项目提供了高质量的成果评估数据(以间接或直接的成功度指标的形式)。我们强烈建议,ISE 计划应要求项目申请者在其项目申请书中明确提出项目的预期成果,并给出测度这些成果是否实现的一套指标。这些指标不应只是基础数据的收集(如项目服务的人数)。为此,专家小组和 ISE 计划官员在评审项目时应将项目申请者提出的项目成果评价计划作为关键的考虑因素。

我们知道,对 ISE 计划和 ISE 领域来说,履行这样的建议很具挑战性。为应对这一挑战,首先需承认成果测度方面的困难,对于与受众没有直接或持续互动的媒体和博物馆项目来说尤其困难。第二个挑战是要承认,我们几乎没有适当的收集关键数据的方法,我们需要找到创造性的新的数据收集方法,其中包括收集有关基础设施和能力变化的“间接指标”(例如,我们不能确定一个全国播放的儿童电视科学节目与学生学习成绩提高之间的直接联系,但我们可以确定一些间接指标,如收视率、与对照组相比实验组的孩子在认知方面的提高以及对教员的培训等体现能力建设的数据)。第三个挑战是纵向数据的获取——在项目资助期结束时几乎不可能评价出一个项目的真正影响,因为真正的影响尚未表现出来。

基于这些挑战,我们还建议,ISE 计划要开发新的评估工具,

比如开发 ISE 项目评估变量测度框架和规范,用于指导项目负责人制定项目评估计划。评估变量的系统使用会有助于今后的 ISE 研究,有助于整合从各个 ISE 项目获得的知识。尽管 ISE 计划不能要求每个项目都利用评估变量测度框架中的所有要素,但该测度框架将提供一个结构性框架,有助于获得可部分比较的数据集。

我们还建议,ISE 计划广泛传播项目的中期报告,这些中期报告叙述了项目进展情况,并包含项目机构在执行项目时所获得的一些新知识,这样的信息有助于加强 ISE 领域的知识体系。

2. 在指引非正规科学教育领域的发展方向方面 ISE 计划应如何发挥更加积极的作用?例如,是否应鼓励提交更多旨在评估非正规科学教育的效果和影响的应用研究项目?

通过支持和资助优秀的展览、电影和电视节目,将非正规教育和学校课程相联系以及要求进行项目评估等,ISE 计划已经对非正规科学教育领域产生了重大影响。在指引非正规教育领域发展方向方面,ISE 计划也发挥着重要作用,而且我们相信,它会继续支持对非正规学习影响的研究。除此之外,我们对 ISE 计划提出以下几条建议。

a. 提出明晰的计划愿景

1998 年度的 COV 报告曾提出过这样的建议,现在我们仍认为 ISE 计划需要有清晰的愿景。

b. 加强研究和评估能力

ISE 领域作为一门学科,缺少一种将从项目中获得的知识基础用于指导未来工作的“发展模式”。我们建议 ISE 计划办利用其在资助、召集会议和出版方面的能力去建立 ISE 的研究基础。我们还建议从如下两个行动战略的角度来看待这个问题:①开发测度 ISE 成果的新方法;②促进影响该领域发展的新知识的传播和利用。

- 建立一种机制,通过它可以“挖掘”NSF 资助的项目的形成

性评估和总结性评估数据,编撰非正规科学教育方面的文献、知识、研究和事实依据。我们特别建议 ISE 计划发布一项目征集书(RFP),以促成由 ISE 计划资助的项目所获得的知识形成知识库,并鼓励由 ISE“生产者”和评估专家合作共同提出项目申请。

- 为了解决非正规科学教育评估特别是成果和影响评估的难题,ISE 计划应支持举办相关的会议和征集论文,并创立一个相关工作组。这方面可以借鉴综合社区计划评估圆桌会议的模式,它是由阿斯本研究所资助并实施的。

- 像“探索与新发现校外中心”(ASCEND)计划指南中要求的那样,ISE 计划应要求那些旨在促进个人学习和技能掌握的项目至少实现参加者成果(收益)方面的某些目标。对于那些致力于基础能力建设的项目,应要求有明晰的逻辑模式和变化模式,并确定一套中期成果或衡量基准。

- 帮助项目执行机构制定项目评估计划。理想的情况是,网罗一批 ISE 计划评估咨询专家,可以通过电子邮件和电话与他们联系,获得他们的帮助。由于评估专家人数有限,可以只向那些已获得了有利的评审结果且处于商议阶段的项目提供评估专家资源。

- 与其他类似领域(如社区发展和青少年成长)的有造诣的研究和评估人员建立联系。寻找擅长进行影响评估的评估人员,并与之合作。

c. 培植非正规科学教育事业

ISE 计划自 1984 年实施以来,发展了很多新知识,我们认为,ISE 计划应期许实现这些知识的共享,并为之创造基础设施条件。为了培植 ISE 事业,我们建议 ISE 计划应采取以下措施:

- 举办项目负责人参加的会议。

- 在 NSF 的网站建立专门的非正规科学教育网页,网页内容可以涉及 ISE 的各个方面,比如研究与评估的设计和结果,样

板项目介绍以及非正规科学教育的理论和教育方法。

- 资助出版非正规科学教育领域的论文。
- 组织或共同资助重要的会议演讲,这些会议不局限于科学教育领域,例如可以是青少年组织、少数民族服务机构以及素质教育等领域的年会。

d. 借助迅速发展的课外活动势头

由于地方和全国性的民间基金会的大力资助,课外活动计划获得飞速发展,而五年前这方面的计划还微不足道。然而,从事这些计划的人员并没有挖掘利用已被开发出来的优秀的科学、数学和技术资源,几乎不了解非正规科学教育的方法和潜力,而且课外活动通常也没有很好地利用其周围社区、学校和科研机构的资源。课外活动计划为 ISE 计划提供了很多机会,ISE 计划可以借助学校课外活动来试验、演示、实施和评价非正规教育方法,可以采用与学校科学教育标准一致的学习内容和技能,可以提高学生的学习成绩,缩小成绩差距。为利用课外计划带来的机会,我们建议 ISE 计划与从事课外活动的主要机构和专家建立合作伙伴关系,制定未来五年的合作战略计划,并支持相关的合作项目。

e. 促进公众了解和欣赏非正规科学教育

ISE 计划应继续努力提高公众对校外学习的欣赏,进一步打破学校、社区和科学与文化机构之间在科学教育上的界限:

让评估人员评估这些不同机构间建立的协同关系是否可以提高科学素养,如果是的话,是怎样提高的,程度如何;

与传播专家合作,研究如何推进非正规教育过程以及如何向公众传播生活在这个技术复杂化的时代所需的科学知识。

f. 要始终关注科学内容的教育

ISE 计划三大类型的项目都必须像重视科学过程和技能教育那样重视科学内容的教育,青少年与社区类的项目尤其要如此。这样做不仅仅是为了帮助孩子们的科学课学习和缩小成绩差距,

而且也是为了满足普通公众了解艾滋病研究与预防、克隆、药物试验直到环境政策等各种科学问题的需要。

g. 扩大青少年与社区类项目

我们建议,ISE 计划扩大对青少年与社区机构提出的项目的支持(ISE 计划资助的青少年与社区类项目的范围还很有限,而且 1999 年和 2000 年度社区类项目的资助额出现大幅度下降)。

3. ISE 计划应如何确定其资助内容以充分利用万维网等带来的新机遇? ISE 计划的指导方针是否照顾到该领域的所有活动?

在对这一问题提出建议之前,我们想强调,网络应被看作一种手段而不是一个目的,而且不是向所有目标群体的传播都可以依赖网络进行,因为在科学方面未被充分代表的社区也是最不可能接触网络的。另一方面,NSF 应利用其地位确保所资助建立的网页和技术工作都能使残障人和贫穷的社区充分享用到。

网络是一种重要的新兴非正规科学教育管道,在项目征集中应明确网络是达到非正规教育人群的一种潜在的独立手段。在不到十年的时间里网络已经达到 25% 的美国公众,比以前所有的技术创新都要快。尽管网络技术面临着诸如“数字鸿沟”等挑战,但它确实是一个重要的有待开发的机遇。由于网络技术的潜力如此巨大,ISE 需要决定是否抓住这个可能带来丰硕成果的机遇,研究人们在网络环境下怎样学习科学和数学。COV 认为,ISE 计划可以通过资助对“如何利用因特网最有效地支持学习”的研究和评价,在这方面发挥重要作用。

我们还注意到了以下需要或机遇:

- 需要寻找实现项目网站再利用的途径,比如可以通过 NSF 作为入口网站,使项目网站变成较为永久的连接;
- 需要教会儿童和普通公众更好地利用网络;
- 发展网络广播、远程学习、人际电子互动以及虚拟参观研究场所的机会。

4. 2001年,ISE计划增加了1000万美元的资助,用于将受资助的范围扩大到较小社区的非正规教育机构。在大城市市区人口不足300万的情况下,COV认为ISE计划需采取哪些策略以确保ISE项目对社区产生影响?

首先我们认为,在ISE计划的一揽子资助中,博物馆类或青少年与社区类最有可能做到这一点,因为它们的战略定位就是提高当地社区居民的科学素养。也就是说,实现这一目标的真正机会是资助在一区域内有较大影响的地方性项目,而不是支持全国性的工作“播种”到一个地区。

我们赞扬ISE计划官员近年来开展的这种深入到社区的活动,并鼓励其继续下去。为鼓励较小社区参与ISE计划,我们还提出了如下一些想法,这些想法在很大程度上都要依靠建立合作伙伴关系:

- 小社区的潜在申请人与经验较丰富的项目负责人结成合作伙伴;
- 鼓励有助于增进较小机构开发和管理项目能力的项目申请,包括小机构间的协作项目;
- 请全国性的机构和机构领导人帮助ISE确定扩大受资助人员范围的策略;
- 制定一项“能力建设”的资助计划,为有经验的项目负责人等提供最高5000美元的资助,让他们作为咨询人员帮助从未获得过ISE资助的较小社区了解“如何使一项可行的项目变成NSF值得资助的项目”。我们认为目前的ISE工作人员配备不足以实施该计划。必须建立“向新的潜在受资助者传播机会”的新模式。例如,可以发布一项目征集书,鼓励有关组织(如美国科技中心协会等)利用它们在全国的影响力组织召开地方争取NSF资助的研讨会;
- 制订一项旨在提高小社区的社区机构(CBO)能力的计划,

使小社区的 CBO 有能力参与非正规数学、科学和技术教育的工作。这样的一项计划应有助于提高小社区的 CBO 争取联邦和非联邦资助的能力。同样,我们还建议 ISE 制订一项旨在提高区域性和全国性机构以及大城市社区机构的能力的计划,以形成一种能成功地影响小城市地区的计划机制。

• 面向社区,其形式可以是:

- 通过已有的地区服务合作伙伴扩大计划资助范围,如可以通过农村教育体制改革计划向农村社区延伸;
- 与美国农业部及其在全国的 4-H 组织合作开展政府部门间计划;
- 支持全国性组织为未得到充分服务的小社区提供工作援助。

5. 根据 COV 采用的 GPRA 报告模式得出的评价结果,您对 ISE 官员应在哪些方面改进计划过程和管理有何建议?

ISE 计划的项目申请管理是卓有成效的,该计划对文献记录和利益冲突程序的重视堪称该领域的典范。在力求使受资助项目获得最大产出和成果方面,ISE 官员做的非常出色,但这种卓越表现在 COV 访评报告的 B 部分并没有得到明显反映,因为 COV 访评报告遵从 NSF 的 GPRA 评价模板的要求,而 NSF 的 GPRA 评价模板又是基于 NSF 浓厚的学科研究文化设计的。NSF 的 GPRA 测度侧重于人、创新思想和工具三方面的最终成果,而 ISE 计划的侧重点是基础设施建设。这种不一致使 ISE 计划的 COV 评价人员处于两难境地,对 ISE 计划不得不采用用来测度研究计划的“人、工具、创新思想”评价标准,而 ISE 计划本来应采用文化改变、组织结构和过程改进等的评价标准来测度。

为能从 COV 获得有价值的建议,NSF 应将 COV 评价模板中 B 部分进行调整,使 COV 评估标准和评估指导原则与 ISE 计划项目征集手册中的产出和结果目标相一致。另外,明智的是应确保

这些“调整过的”的指标反映在：①项目申请指南中；②专家评审小组的职责中；③外部评估人员的要求中；④提交最终报告和产品的指导原则中。

此外，关于 ISE 计划过程 and 管理的改进，我们还提出如下一些建议：

- 增加人员费和差旅费，从而使计划官员能够对小社区和首次申请项目者给予更多支持。从我们评议的项目卷宗可明显看出，计划官员的作用对项目结果有特别大的助益；

- 加强项目成果的报告，这将有利于 ISE 领域的发展；
- 对提供给 ISE 评审专家的大量资源和信息提供分类以及优先排序的更多说明，会有助于评审专家的工作；

- 对(获资助的和被否决的)项目主持人进行调查，了解他们对 ISE 计划过程的反馈意见，这对 ISE 计划改善评审过程和管理也会有所帮助。

D. ISE 计划官员对访评委员会评估报告的 反馈意见

首先，ISE 计划官员对 COV 作出的全面的评价报告和得出的许多正面的结论表示感谢。我们还要感谢他们认真提出的建议以及对 ISE 计划官员所提问题的答复。

本报告分为三个部分：计划的运作和管理、ISE 计划官员提出的补充性问题以及对今后 COV 会议评议过程的建议。在我们分析 COV 报告的时候，有些部分涉及的是相关的意见。在这样的情况下，我们将相关的意见合在一起答复。

计划的运作与管理

- COV 认为，就任务要求来说，ISE 计划官员人数显然不够。其

后果是项目官员无法投入足够的时间,与受资助者进行项目初期的必要协商以及对进行中的项目实施连续的监督。COV 还认为,ISE 计划项目评审工作的成功很大程度上应归功于项目官员的出色表现。然而这种出色表现不是必然的,应通过对新的项目官员实行上岗培训以及不断地进行工作勘漏继续保持和发扬这一业绩。

答复:由于上述原因,NSF 的中小学与非正规教育处(ESIE)经常考察工作人员的工作量问题。在进行本次 COV 访评中,ISE 计划的工作人员出现了配备不足的情况。一名项目官员(PO)离开了 NSF,还有一名在休病假。而且 ASCEND 计划增加的工作加重了“科学素养部”的工作负担。最近的一些人事变动会使工作量问题得到一定的缓解。“科学素养部”已经为博物馆项目招进了一名全职项目官员和一名兼职项目官员,也为媒体项目增加了一名全职项目官员。另外,还有一名有着丰富的学校和非正规教育经验的人被派到科学素养部。

今年,NSF 教育与人力资源学部(EHR)和中小学与非正规教育处(ESIE)都为新来的 PO 提供培训。新的 ISE 官员将参加这些课程,并侧重于 COV 强调的方面。此外,ESIE 为每一名新 PO 指定一名有经验的 PO 作指导者。像 ESIE 的所有计划一样,ISE 计划也有一个对资助决定和项目运作进行经常性监督的过程,该过程由 ESIE 的处长和副处长负责,它有助于确保 ISE 计划的质量。

• COV 建议,ISE 计划应制定新的程序,以确保评审人员更认真地重视“社会影响”这一价值标准。它建议给评审专家的评审说明中应详细地解释第二条标准(社会影响标准)的含义,更突出地强调要重视每一条标准,并明确规定只有学术水平和社会影响都获高分的申请才能排在前面。COV 尤其建议,在评审开始时就将第二条标准的全文给评审专家看。

答复:ISE 计划的官员对 COV 建议中提出的问题非常敏感。

他们已经在向评审小组作简要介绍时对 NSF 的两条评审标准(“学术价值”和“社会影响”)给予了更多的重视。ISE 官员将继续对程序进行反复核查,以确保在召开评审小组会议时或在会前给评审专家的书面材料中对评审专家给予更明确的指导。

- COV 对卷宗的审查表明,ISE 计划在项目评审人员的组成上兼顾了不同的性别、机构规模和类型(博物馆、大学、媒体方面、社区和青年组织)以及地理分布。尽管 COV 成员从自己以往作为评审小组成员的切身经历知道,ISE 计划的专家小组包括了不同种族的成员,但从 ISE 项目卷宗中我们看不到项目评审人员属于什么种族或是不是残障人。COV 了解,收集和公布评审专家小组成员的统计资料可能涉及到法律方面的问题,因此它建议 ISE 官员应要求项目评审小组成员自己递交一份描述其背景和专业背景等有关情况的材料。这种信息将对项目评审专家小组的评审过程提供很大帮助,它可以确保和记录项目评审专家的多样化和平衡组成,并进一步促进观点的交流,这种交流会使得评审专家小组的评审过程富有价值且有利于产生好的项目。

答复:ISE 官员一直都在努力实现评审人员的平衡和广泛的代表性,涉及到性别、机构的类型和规模、地理位置、种族的多样性和残障人等问题。ISE 计划要求评审人员提供最近的简历,并要求他们在自愿的基础上提供个人信息。专家评审小组成员的选择需经 ESIE 处长批准,ESIE 处长同样会注重上述的多样性。尽管在每一卷宗的表 1 中提供了一些信息(学科、地理代表性、机构隶属关系、性别),但并没有提供其他信息(种族、民族)。ESIE 正在建立正式的旨在系统收集每个专家评审小组中评审人员特征信息的程序,这些信息以后会提供给 COV。还有一点应该注意,那就是 NSF 采用的学科方面的数据不能充分满足 ISE 计划的专业要求(如展览设计、大型电影的制作)。我们会继续努力寻找机会,要求 NSF 为 ISE 计划扩大专业划分。

• COV 建议 ,在 PO 提出资助建议后 ,应尽量缩短通知滞后的时间。COV 审查了申请递交日期、评审日期、通信和商议的日期以及 PO 对评审进行分析的日期。从总体上看 ,近三年来效率不断提高 ,从专家小组评审到批准一项资助之间的时间不断缩短。尽管 ISE 计划在缩短决策时间方面做出的成功努力得到了赞扬 ,但该过程还尚待改进。有证据表明 ,有些问题发生在 ISE 的 PO 将其建议提交上去之后。

答复 :ISE 官员非常重视及时作出资助决定。因为 ESIE 处理的申请量很大以及其他许多必须完成的任务 ,所以缩短滞后时间有时很困难。ESIE 正在制订提高申请处理过程效率同时保持资助决策统一的战略。一旦卷宗上交 ,ESIE 就不能再控制资助的及时性。

• COV 发现 ,“效率 ,通知的时间”这一标准很难评估 ,因为卷宗中的文件资料显示 ,各种具体工作的联系时间有时是矛盾的或不清楚的。COV 建议在卷宗的左边增设工作进程记录表 ,这样 ,这类错误就比较容易避免了。

答复 :效率和通知时间一直是个挑战 ,但它们对 ISE 官员极为重要。ESIE 处目前采用日常工作记录单来记录整个部门的卷宗处理进展情况。在项目拨款之后该日常工作记录单就会中止。在卷宗中建立一个工作记录栏可能会有助于说明 ISE 和 ESIE 的项目申请的经过 ,因而我们会考虑采用它。不过 ,由于 NSF 正在朝电子卷宗的方向发展 ,所以工作记录栏可能就没有必要了。

• COV 注意到 ISE 官员在文献记录方面一直做得很出色。多数卷宗都有很好的记录。当过程中出现异常情况时 ,项目官员记录了相关的信息。有些 PO 记载了电话联系的情况 ,这一点我们建议所有的工作人员都应效仿。

答复 :ISE 官员赞同良好的文献记录对于有效的项目监管是必不可少的。所有的 PO 都已认识到 :①用日志记录与项目主持

人的电话交谈或其他交流方式的重要性 ;并②将这类记录放到拨款项目卷宗中。随着电子邮件应用的增加 ,文献记录会更容易 ,可将复制件放入卷宗中作为有关项目的决策和项目重要进展的记录。NSF 的 Fastlane 正在增加将与项目主持人交流的文献记录添加到电子卷宗中的功能。一旦这个功能可以使用 ,ISE 官员就会熟练掌握并系统地经常地使用它。

• COV 认为 ,ISE 领域缺少一种“发展模式” ,因而建议 ISE 计划提出一项目征集书 ,目的是促进 NSF 在 ISE 领域的投资所获得的知识共享。这种共享对于打造这一领域的实力是非常必要的。COV 建议 ISE 计划利用其在资助、召集会议和出版方面的能力去建立 ISE 的研究基础。它还建议从如下两个行动战略的角度来看待这个问题 :①开发测度 ISE 成果的新方法 ;②促进影响该领域发展的新知识的传播和利用。

答复 :ISE 官员同意 COV 提出的该领域缺少一种发展模式观点。如前所述 ,ISE 计划正在探索测度成果的新方法 ,并且鼓励提出有助于增进 ISE 领域实力的项目申请。这些项目申请中有些是作为会议得到资助的 ,并且已出版或即将出版介绍最佳实践和最新研究成果的出版物。2001 财年 ,ESIE 资助建立了一个新的学习与教学中心(CLT) ,即非正规学习与学校中心 ,它将致力于研究正规和非正规环境下科学学习的交叉。

另外 ,ISE 计划还为增强该领域的实力采取了一些措施 ,包括举办研讨会以及通过共享最佳实践帮助机构开发较好的项目。在 2001 年夏季进行计划指南评议的时候 ,ISE 已开始考虑采取哪些有效的措施 ,支持以在从事非正规科学教育的个人和机构间传播知识为目的的项目。

• COV 建议 ,ISE 计划应更加重视并提供更多的资源用于 :①对在科学、数学和技术领域未被充分代表的和未得到足够服务的群体和社区提供支持 ,并重视支持与它们有关的外伸活动 ;②对首次

申请项目者(尤其是来自未充分代表的群体的首次申请项目者)提供帮助。③评估计划的制定以及将形成性评估和总结性评估的结果用于个人和基础设施的能力建设。此外,COV 还建议将 SGER 资助用于进一步鼓励未被充分代表的社区的能力建设。

答复:对有望获得资助的项目申请者,ISE 计划官员在他们提出想法和申请时会向他们提供帮助,主要是在预申请阶段。预申请为 ISE 计划官员提供了围绕好的项目申请的基本要素和项目设计问题对项目申请者给予指导的机会。最近 ISE 计划又采取了一些措施来帮助未被充分代表的和未得到足够服务的群体和社区。例如,去年在全国范围内召开了一系列专题研讨会。ISE 计划官员目前正在探讨还可以采取哪些方式来为未充分代表的目标群体提供服务。至于 SGER 资助,NSF 对其使用有严格的规定。这一资助机制不可能满足 COV 提出的要求。ISE 计划官员将继续探讨可提高未充分代表的社区科学素养的能力建设策略。

ISE 项目中必须包括形成性评估和总结性评估两方面的计划。尽管这些有助于证明计划的影响,但很显然,评估非正规科学教育活动目前还面临困难,需要开发新的评估指标和方法,在评估项目在发展和加强教育基础设施上的效果时尤其如此。去年,ISE 计划已经得到了一家合同评估机构的支持,为其评议项目评估并开始着手开发能表现计划成功的指标。在未来几年里,ISE 计划将继续加强在计划和项目两个层次的评估能力。

• COV 认为,鉴于非正规科学教育的独特性,对于想法超出传统 ISE 活动范围的申请应慎重考虑,但应积极寻求新的项目申请人和机构提出的申请,因为这些是拓展 ISE 领域和培养一批 ISE 专家的重要途径。COV 建议,ISE 计划应继续执行其鼓励和培养更广泛更多样化的潜在申请者参与的战略。目前 ISE 计划为鼓励小社区的参与而做的工作可以提高该计划“高风险”项目的数量。为进一步推动这一工作,COV 建议 ISE 计划努力增加这方面的投

入,更好地界定非正规教育中的创新和高风险。

答复:ISE 计划官员同意 COV 提出的应大胆资助新项目申请者和机构的观点。ISE 计划将不断地审视和修订其指导原则,以鼓励新的项目申请人和机构,促进更广泛的参与。ISE 计划还将通过外伸工作和向专业机构的宣传征集新的项目申请人和机构参加该计划。

- COV 赞扬了 ISE 项目官员在项目评审和管理方面的出色表现。它尤其被 ESIE 年度报告中记载的成绩所打动,并且建议 ISE 计划考虑将该报告改编出版,以便帮助公众了解 ISE 计划对提高美国人的科技素养做出的巨大贡献。

答复:ISE 计划官员将考虑怎样证明计划所取得的成绩以便于向公众宣传并为公众所了解。ISE 计划认为,让公众了解 NSF 和 ISE 对科学素养的贡献非常重要。值得一提的是,ISE 计划已要求获资助的项目在所有出版物、媒体、网站和展览宣传中都要注明该项目是得到 NSF 资助的。

- 鉴于可持续性和长期影响是 ISE 项目成果的重要方面,COV 怀疑,通常的三年项目期限不足以评估项目成果。

答复:ISE 计划并没有硬性规定三年的项目期限。三年期限通常是由项目主持人自己定的,他们认为三年的期限足以完成一个项目。ISE 计划官员鼓励项目主持人延长项目期,前提条件是期限的延长是合适而又审慎的,包括了进行总结性评估的时间。此外,ISE 计划还在积极探索各种证明项目影响的评估方法,使之有利于促进更广泛的计划评估。

- COV 建议 ISE 计划再增加一名项目官员负责 SGER 申请的评审。尽管 COV 知道 ISE 计划的操作符合 NSF 的规定,即所有的 SGER 申请都只由一名 PO 负责,但 COV 还是要敦促 ISE 计划再多委派一名 PO 来负责 SGER 申请的评审,从而提高 SGER 申请提出可资助项目建议的可靠性。

答复 :NSF 的政策没有要求 SGER 申请须进行外部价值评议 ,而且没有规定应有几份评审意见。ISE 计划官员理解并非常支持 COV 增加一份评审意见的建议理由。为了使对 SGER 申请的资助决定更具信服力 ,ISE 将增加一名 PO 做 SGER 申请的评审。

ISE 计划官员提出的补充问题

- COV 强烈建议 ,ISE 计划应要求每一份项目中明确提出预期的成果和一套可用于测度这些成果的实现程度的可信的指标。这些指标不应仅仅是基础数据的收集(如项目服务的人数)。为此 ,专家小组和 ISE 计划官员在评审项目时应将项目申请者提出的项目成果评价计划作为关键的考虑因素。

答复 :ISE 计划官员将审议 ISE 计划的指导原则 ,并考虑在其中加入有关上述建议的具体要求。通过 NSF 教育与人力资源部研究、评估和交流处(REC) ,ISE 目前正与两家机构合作确定 ISE 项目的成果与产出指标。ISE 计划官员也在探讨项目数据收集方法 ,使之有助于建立一个较大的数据库 ,能用来评估整个 ISE 计划的影响。最近召集的一次 ISE 计划未来发展专家研讨会已建议 ,在 ISE 计划的指导原则中提出评价项目影响的指标 ,这是促进该领域评估能力发展的一个手段。ISE 计划打算在未来的几年中执行这一策略。

- COV 建议 ,ISE 计划要开发新的评估工具 ,比如开发 ISE 项目评估变量测度框架和规范 ,用于指导项目负责人制定项目评估计划。尽管 ISE 计划不能要求每个项目都利用评估变量测度框架中的所有要素 ,但该测度框架将提供一个结构性框架 ,有助于获得可部分比较的数据集。

答复 :ISE 计划官员目前正在与 REC 和一家外部评估机构合作 ,共同研究 ISE 项目评估变量和规范问题。目前 ISE 官员还不能肯定上述工作会产生 COV 建议的测度框架 ,但他们会考虑这

一评估工具问题。

- COV 建议 ,ISE 计划广泛传播项目的中期报告 ,这些中期报告描述了项目进展情况 ,并包含项目机构在执行项目时所获得的一些新知识 ,这样的信息有助于加强 ISE 领域的知识体系。

答复 :ISE 官员认为 ,公布项目主持人提交的中期报告或年度报告可能不妥。项目中期报告的目的是让 PO 了解项目的进展情况 ,PO 可以以此作为决定项目是否继续获得资助的基础。形成性评估从本质上讲是为了发现项目在设计或执行过程中的问题。因此 ,它可能会对项目在特定时间造成相当不好的印象。而且 ,中期报告还可能包含项目的一些私有信息。

不过 ,ISE 官员完全赞同需要建设和传播 ISE 知识体系 ,以推动这一领域的发展。ISE 计划鼓励项目主持人到专业会议上做演讲 ,在期刊上发表文章以及在网站上公布项目总结性评估报告 ,并希望以此促成 ISE 领域共享研究成果。但 ISE 计划也认识到 ,这些活动太专门化而且太分散 ,因而不会有很好的效果。目前 ,中小学与非正规教育处正在和教育与人力资源局的其他部门合作 ,共同探讨实现 COV 所建议的目标的策略。这类策略可能包括以网站为基础进行项目成果的传播 ,支持项目主持人或机构撰写论著 ,公开发表总结性评估结果和案例研究报告等。

- 1998 年度的 COV 报告曾建议 ISE 计划需要有一个愿景 ,本年度的 COV 仍认为 ISE 计划需要有一个清晰的愿景。

答复 :ISE 计划官员认为 ,非正规科学教育领域出现的新机遇和新变化要求该计划必须制定新的愿景。2001 年 9 月 6~7 日 ,ISE 计划官员与博物馆、媒体和社区非正规教育方面的专家就本领域的发展方向进行了研讨。ISE 计划官员将利用这次研讨会的结果并通过与项目价值评议人员等的意见交流 ,明确 ISE 计划的未来发展方向 ,使之与 NSF、EHR 和 ESIE 的目标相一致。

- ISE 计划自 1984 年实施以来 ,发展了很多新知识 ,COV 建议

ISE 计划应努力实现这些知识的共享,并为之创造基础设施条件。为了培植 ISE 事业,COV 建议 ISE 计划应采取以下措施:①举办项目负责人参加的会议;②在 NSF 的网站建立非正规科学教育网页,网页内容可以涉及 ISE 的各个方面,比如研究与评估的设计和结果、样板项目介绍以及非正规科学教育的理论和教育方法等等;③资助出版和传播非正规科学教育论文;④组织或共同资助重要的会议演讲,这些会议不局限于科学教育领域,例如可以是青少年组织、少数民族服务机构以及素质教育等领域的年会。

答复:尽管参加会议演讲的数量受到旅费的限制,但 ISE 计划的确组织或主办讨论会。ISE 计划官员正在制定旨在鼓励和促进非正规科学教育领域新知识共享的策略。正如前面提到的,网站将在 ISE 计划的知识传播战略中发挥重要作用。由于 ISE 计划官员人数很少,他们的主要工作是评审项目申请并提出资助建议,因此他们在这方面能做的事情是有限的。为此,ISE 计划今后将探索运用合同资助、专项拨款以及支持项目主持人进行新知识传播的机制。

• COV 建议 ISE 计划应与从事课外活动的主要机构和专家建立合作伙伴关系,制定未来五年的合作战略计划,并支持相关的合作项目。

答复:NSF 的课外探索与新发现中心计划隶属于科学素养部,该计划支持的是课外、周末和暑期活动项目。ASCEND 的经费来自于依据 H1-B 签证法提供给 NSF 的经费。如果依据最新的 H1-B 法可以获得学生活动经费的话,ISE 官员将重新考虑这一计划的方针思想,并根据去年提交的项目申请的经验强化该计划的指导原则。ISE 官员一定会考虑 COV 提出的与从事课外活动的主要机构和专家建立合作伙伴关系的建议。

• 为促进 ISE 计划继续努力提高社会各方面对校外学习的欣赏,进一步打破学校、社区和科学与文化机构之间在科学教育上的界

限,COV 建议 ISE 计划应让评估人员评估这些不同机构间建立的协同关系是否可以提高科学素养,如果是的话,是怎样提高的,程度如何?应与传播专家合作,研究如何推进非正规教育过程以及如何向公众传播在这个技术复杂化的世纪生存所需的科学知识。

答复:ISE 计划官员同意,对在正规环境和非正规环境中的学习都还有很多东西需要了解。计划官员正在考虑采取哪些方式鼓励项目申请人申请跨越正规和非正规学习界限的项目以及申请研究正规和非正规教育机构之间协同关系的项目。

• ISE 计划三大类型的项目都必须像重视科学过程和技能教育那样重视科学内容的教育,青少年与社区类的项目尤其要如此。这样做不仅仅是为了帮助孩子们的科学课学习和缩小成绩差距,而且也是为了满足普通公众了解艾滋病研究与预防、克隆、药物试验直到环境政策等各种科学问题的需要。

答复:ISE 计划官员同意科学内容对于 ISE 项目非常重要,而且一直都在努力确保申请项目的科学内容在项目官员给出资助建议之前得到恰当的评审。青少年与社区类项目往往比较注重科学过程和技能的培养,因为参与者介入其中的时间比较长。但这种对科学过程和技能培养的重视并不是说就无需适当重视内容问题。为此,ISE 计划官员将审议 ISE 计划指导原则并对其进行必要的修改,以便更明确地表明重视 SMET 内容,并研究如何强化非正规科学教育机构与高校等具有丰富 SMET 内容的机构之间建立合作伙伴和协作关系的要求。

• COV 认为,ISE 计划应资助对“如何利用因特网最有效地支持学习”的研究和评价,这对非正规教育领域会大有帮助。COV 还注意到了以下需要或机遇:①需要寻找实现项目网站再利用的途径,比如可以通过 NSF 作为入口网站,使项目网站变成较为永久的连接;②需要教会儿童和普通公众更好地利用网络;③发展网络广播、远程学习、人际电子互动以及虚拟参观研究场所的机会。

答复 :ISE 计划办收到越来越多的利用网络的项目申请,有的项目包括一个网络成分,有的项目则以网络活动为核心。在这类项目申请中,很多申请令 ISE 计划官员心存疑虑,因为它们没有很好地界定其目标受众,也没有提出用来评估项目对目标受众影响的适当的评估计划。这类项目中最终赢得资助的是那些包括许多互动机会的项目,比如市民—科学家项目以及包含研究场所虚拟参观内容的网络广播项目。ISE 计划将继续鼓励有关“如何利用因特网最有效地支持学习”方面的创新性项目,同时寻找机会支持培养该领域提出此类申请和项目的的能力,并支持开发证明这类项目效果的评估方法。

• COV 赞扬 ISE 计划官员近年来开展的深入到社区的活动,并鼓励其继续下去。为鼓励较小社区参与 ISE 计划,COV 还提出了如下一些想法,这些想法在很大程度上都要依靠建立合作伙伴关系:①让小社区的潜在申请人与经验较丰富的项目负责人结成合作伙伴;②鼓励有助于增进小机构开发和管理项目能力的的项目申请,包括小机构间的协作项目;③请全国性机构和机构的领导人帮助 ISE 确定扩大受资助人员范围的策略;④制定一项“能力建设”资助计划。

答复 :ISE 官员同意这项建议。事实上,上述的一些工作有些已付诸实施。即将修订的 ISE 计划指导原则将鼓励能力建设活动,包括鼓励有经验的机构与缺乏经验的机构建立合作伙伴关系。

对未来 COV 会议评价过程的建议

• 为让公众和决策者了解非正规教育在促进科学事业健康发展上的显著贡献,关键的一点是 ISE 计划必须按《政府绩效成果法》(GERA)的要求证明其成效。然而,COV 评价模板的许多问题和指标并不适合 ISE 计划,而且我们担心,如果我们不回答这些问题,会威胁 NSF 今后获得这一重要方面的持续支持。COV 建议

NSF 重申 COV 评价模板中的问题和指标 ,或对其加以解释 ,使之与所评价的非正规教育事业的目标相一致。

答复 :由于 GPRA 模板是用来评判 NSF 所有计划工作的影响的 ,它不可能只针对一项计划进行修改。在 GPRA 框架下 ,NSF 也在鼓励各个计划修改 COV 报告模板 ,以便将特定计划的成功度指标包括进去。不过 ,ISE 计划官员完全同意 COV 对 GPRA 绩效报告的担心。中小学与非正规教育处(ESIE)已经将 COV 的这一担心向教育与人力资源部(EHR)作了汇报 ,教育与人力资源部又将其转达给了负责在 NSF 实施 GPRA 的有关人士。

- COV 评价会的规定时程不允许 COV 人员对 ISE 计划的成果和影响作出应有的系统评价和分析。要进行系统的评价 ,需要对项目最终报告、总结性评估报告以及成果进行认真和广泛的研究。前些年参加过 COV 访评的专家发现 ,由于对 ISE 计划新增加了 GPRA 模板评估以及既要评过程也要评结果的要求 ,访评的工作量较前些年增加了一倍多。COV 成员建议将 COV 评价会的时间从两天延长到三天。

答复 :ISE 计划的官员对 COV 在规定的两天时间内所做的出色工作表示赞赏。我们同意 COV 的意见 ,三天可能会更好。此外 ,ISE 计划官员将设法把有助于 COV 评价工作的所有信息综合起来并尽早发给他们。

- COV 发现 ,根据 ISE 计划的三大资助领域将 COV 分成三个分委员会很有效 ,因而建议以后的 COV 会议继续采用这一方法。

答复 :ISE 计划官员也认为分委员会的组织形式很有效 ,建议以后继续采用。目前 ,向 ISE 计划递交的项目申请大致分为三大类 ,即 媒体类、博物馆类和青少年 /社区类 ,COV 分委员会是围绕这三类进行组织的。将来 ,针对某些特殊类型的申请项目(如网络项目) ,ISE 计划有可能会考虑增加其他的分委员会。

附件三 “比尔·奈科学人”电视 系列片评估报告(1996)

评估机构 :罗克曼研究与咨询公司

“比尔·奈科学人”是一个拥有广泛少儿观看者的儿童科教电视系列片。它专为 8~10 岁的儿童设计。该系列片通常在星期一到星期五的傍晚在公共广播公司(PBS)电视台播映,并且每周末在商业电视台上播映。该系列片还开展了外延教育工作,包括编写有助于教学的教师指南以及为在家里和学校观看此节目的孩子提供有关该节目的简报和材料。

罗克曼公司是位于旧金山市的一个独立的研究和评估小组,受公共广播公司西雅图电视台(KCTS)委托进行“比尔·奈科学人”电视系列片及其教育外延工作的评估。评估内容包括探究该系列片对儿童的影响。不论是该系列片的投资人还是 KCTS 都关注该系列片对儿童科学学习成效、态度变化以及对女孩和少数民族儿童的影响。

评估背景和方法

此项评估调研在三个地区进行,以体现全美儿童观看者的多样性。在每个地区,我们都从乡村、郊区和城市征募学校观看者和家中观看者作为此次评估调查的受试者。在人口密集的地区,那些不曾观看该系列片的课堂作为对照班级,也作为评估的调查对象参与进来。对照学校是按照观看该系列片的学校的学生人数对等选择的。学校的教师和在家中观看该节目的儿童每月会收到一份节目时间表,每 2~4 周会收到一些奖励,如科学玩具、明信片或电话等,因为我们希望在为期五个月的评估调研中能保持一个高

收视率。

我们为学校观看者和家中观看者提供了多集的“比尔·奈”节目录像带。这些录像带作为一套基本节目,供他们随时或按时观看,而且遇到雨天天气,会使那些原本不看该节目的孩子们有机会观看。我们要求校内观看者和家中观看者至少看完录像带上的12集节目。我们还鼓励他们每周至少看一次“比尔·奈”的其他集的节目。

我们此次评估调查的学生(包括看节目的学生以及未看节目的对照班级的学生)都在1995—1996年初冬时节完成了看前问卷调查表,在1996年春末完成了第二轮问卷调查表。在该项调研开始和结束的时候,我们对一些学生及项目研究人员进行了访谈。这些学生有的是在家看节目的,有的是在学校看节目的,还有的是对照班级的学生。参加访谈的学生是经过挑选的,以便对女孩和少数民族学生能过采样。看节目的班级的教师和未看节目的对照班级的教师也填写了问卷调查表,并参加了访谈,我们对作为调研对象的在家看节目的孩子的家长也进行了问卷调查和访谈。

样本

我们选择在三个地区进行该项调研:加利福尼亚州首府萨克拉曼多、宾州费城以及印第安纳州的印第安纳波利斯和布卢明顿。在每个地区,我们从市区、郊区和乡村征募观看者,包括学校观看者、家中观看者和对照班级。我们从总共1350名学校儿童中收集评估材料,包括观看组的大约800人和对照班级组的550人。我们的数据包括了22所学校的45个班级。

我们从作为调研对象的观看班级和对照班级中选取了三分之一的问卷调查表作为此次调研分析的学生样本。在所选取的样本中,我们可以分析大约220份可用的观看组调查表和大约140份可用的对照学生组调查表。在调研开始前和结束后,我们对大约

140 名观看节目的学生和 100 名未观看节目的学生进行了访谈和动手型活动测评。

我们也选取了家中观看者,以代表“比尔·奈”节目观看者的多样性。这些家中观看者位于所有的居民区里。我们从作为调研对象的学校、其他学校、女童子军、课外俱乐部中挑选并招收了大约 375 名家中观看者。我们也招收了一些在家受教育的孩子以及以前曾索取过“比尔·奈”节目材料的儿童作为调研对象。在调研开始和结束时,我们对 331 名儿童进行了测评,并对大约 90 名家中观看者开展了访谈和动手型活动测评。

最终用来进行数据分析的总样本包括大约 540 名家中观看者和学校观看者,以及 140 名未看节目的对照组儿童。在调研开始和结束的时候,我们对这些儿童进行了平均 400 多次访谈。

此次评估还增加了一项调研内容,即在课外托管机构观看此节目的目标年龄儿童进行调研。罗克曼公司在一些托管中心放置了电视录像机,要求“比尔·奈”节目的放映成为这些托管中心的一个固定的(虽然不是主要的)备选活动。中心的工作人员监督收视情况并负责维护设备。在调研开始和结束的时候,我们从大约 50 名课外观看者那里采集了信息。

问卷设计

我们研究了大量的适合儿童发展的科学概念。我们问的问题多侧重于测评孩子们的科学知识和批判性思维能力,以判定观看“比尔·奈”系列片在多大程度上促进了孩子们的科学学习。另外一些问题侧重于测评孩子对科学和科学家的态度、科学意识。还有一些问题是针对儿童可能有的科学行为。知识、态度和行为是本次调研的三个主要点。

我们运用多种评价形式获得“比尔·奈”系列片对儿童的成效和影响。我们既向那些在学校或在家中观看该节目的儿童、也向

那些没有观看节目的对照组儿童问问题。我们还从教师和家长那里采集信息。

考虑到目标观看者是从3年级直到5年级的学生,其能力高低不同,我们为三年级学生设计了简化的测评表,为4~5年级学生设计了较长的测评表。我们在当年初冬采集了一次信息,次年春末又采集了一次。在进行校内观看者调研的时候,我们把书面调查表发给了观看班级组和对照班级组学生。对于家中观看者,有时是个别测评的,有时则与在学校、课外场所和娱乐小组的观看组一同被测评。绩效测评主要对个人,偶而以成对的学生为对象,视具体场景定。

提请读者注意的问题

我们下面谈到的评估调研并不是一组严格受控的活动。在家中和在学校,我们都有一群自愿的受试者,然而,在本调研期间,“比尔·奈科学人”只是他们的生活中的一小部分。教师仍然必须专注于阅读、写作和算术教学。去年冬季(1995—1996年),东部和中西部地区遭受了一连串的气象灾害,在费城和印第安纳两地,也经常是大雪天气,学校不得不停课。因此,教师经常会改变他们的教学计划,集中于补基础课,科学并不总是能纳入基础课教学之中。家长们也说,孩子在傍晚看“比尔·奈科学人”节目必须和其他活动(如踢足球、上音乐课、做家务和家庭作业)平衡考虑。

从实验的意义上说,本次调研的对照班级并不是界线分明的、未被污染的对照组。事实上,在进行调研之前、之中和之后,对照班级的许多学生是看过“比尔·奈”节目的。事实上,我们在各调研地区招募的家中观看者中,很多是来自对照班级。此外,在采用“比尔·奈”节目教学的学校中,节目在班级之间互相传播。我们多次听说,看“比尔·奈”节目的班级的教师同对照班级共享他们的录像带。他们显然觉得,他们不想独占有益的教学材料,而不顾及这

种举动将会影响本次评估调研的结果。我们还发现,在该调研项目开始之前,已有 60%~80% 的儿童熟知“比尔·奈”系列片。少数民族儿童最不熟悉该系列片,只有 60% 少数民族儿童说他们过去曾看过该节目。因此,我们决定从调研分析中删去大约 20% 的对照组学生,因为他们经常观看“比尔·奈”节目。

在我们所采用的评估设计中,“比尔·奈”节目观看者和非观看者之间的对比只是我们能用来研究该系列片影响的若干种对比中的一种,也是可以用来显示变化的多种关系的一种。由于对照班级的样本受到大量污染,对照班级的学生从个人到整个班级都可能成了该节目的观看者,因此,没有观看节目的学生和观看节目的学生之间的对比的有效性明显降低了。

在调研开始时,观看节目的班级和未观看节目的班级的测评结果差别非常小,这有力地说明我们选择的是相同的受试群体。在调研结束时,对照班级组和观看班级组之间出现了很多小的差别,这暗示有两种不同的假设:对照学校不再是一个好的对照组;或者观看节目对儿童并没有产生影响。我们认为前者才是事实。

此次评估设计中还涉及到第二种对比,即:针对此次调研,组织曾观看过该系列片的学生更集中地观看“比尔·奈”节目,比较他们在集中观看前后的差别。事实上,比较一下孩子们观看“比尔·奈”节目前后的回答,就可以发现,他们对“比尔·奈”节目的认识和理解前后有着许多的差别。因此,本报告的重点放在研究分析观看者在观看此节目前后的差别,而不是研究分析该节目的观看者和非观看者之间的差别。

另外,我们的初步分析表明,除了对某些科学知识问题的回答略有不同之外,不同年级的学生对所测评问题的回答没有什么差异。因此,在分析中,我们将各年级组统合在一起,忽略了年级间的微小差别。

“比尔·奈科学人”的观看效果

本报告首先提两个问题,这两个问题体现了我们的调研目标:

- 孩子们观看“比尔·奈”节目后学到了什么?
- 这些孩子是否因观看了该节目而改变了他们对科学的态度、他们的科学意识和科学行为?

这两个指导性的问题集中于探究该系列片对各种环境下的儿童所产生的影响。此外,我们还为“比尔·奈”系列片今后的发展策略提出建议的方向,以期继续扩大和加强该节目对儿童、教师和家庭的影响。

孩子们观看“比尔·奈”后学到了什么?

- 与观看节目之前相比,儿童在观看“比尔·奈”节目后,能够对一些科学概念提出更完整和更复杂的解释。

- 与观看系列片之前相比,儿童在观看“比尔·奈”节目后,能够对一些科学概念提供更加全面和更加复杂的解释。不过,这些孩子尽管掌握了基本的科学概念和科学过程,但他们往往不能将科学原理应用到其他情景中。

- 看过该节目的孩子虽然不总是能充分清楚地说出正确的科学概念,但是他们不再像以往那样经常回答错误。

- 观看该节目的孩子确实学到了一些科学事实和科学概念,通常是因为这样的信息在节目中反复出现的结果。学生能够基于他们从所观看的节目中获知的事实回答问题。

- 对于那些涉及儿童所熟知的具体关联物的问题(例如骑自行车的物理原理),我们发现学生对这类问题的认识在观看节目前后有了明显变化。

- 我们的问题中带有具体关联物很多是在“比尔·奈”系列节目中出现的。学生对这些问题回答结果似乎与具体集的节目的

收看有关。

- 在动手型活动测评中,那些定期观看“比尔·奈”节目的学生对科学概念能给出更好的解释和扩充。

- 一开始,女孩的知识基础总是不如男孩,但是在观看该系列片之后,她们往往赶了上来,与男孩的差距缩小了。从科学知识和科学思维的提高来看,女孩和男孩进步一样大,甚至进步更快。

- 从科学知识和科学思维的提高来看,少数民族学生和多数民族学生的进步一样大,甚至进步得更快。不过,刚开始的时候,这些少数民族学生掌握的科学知识也是比多数民族学生少,以后则开始赶了上来。

评估策略和构想

我们通过各种各样的途径收集儿童对科学概念的了解的信息。我们设计了4组问题和一个动手型活动,以确定儿童观看该系列片可能带来的科学思维和科学知识方面的变化。

- 现象演示:调查人员演示一个实验,然后请儿童选择与实验的基本概念有关的答案。

- 提出假设:搜集有关科学知识以及科学概念间关系的信息,比如,对另一个星球存在生物提出生物学和地质学方面的解释。

- 确定动物的属性和特征,以探究儿童的事实学习以及传统的科学思维能力。

- 提出一些应用性的科学问题让孩子们解答,以此探究儿童的高层次思维和科学解释技能。

- 进行一些访谈和动手型活动,以评价儿童探究和解释科学概念的能力。

从以上可以看出,我们的大多数问题不是侧重于了解儿童的事实学习,而是集中在了解儿童进行探究和解释科学现象时所具

有的批判性思维和高层次思维技巧上。孩子们在观看此系列片时,的确发生了事实学习,这往往是由于事实信息在该系列片的一集节目中反复出现,有时甚至在反映类似科学概念主题的两集节目中反复出现。学生能利用从节目中获知的事实回答问题。但是,在我们能够证实的该系列片所产生的影响中,大部分涉及的是孩子们的思维技巧以及他们对科学现象进行解释或者确定解释的能力。

与观看前相比,观看“比尔·奈”的儿童在观看后能够对一些科学概念提出更完整和更复杂的解释,且明显地胜过了那些没有观看该系列片的儿童。然而该结果不具有一致性和普遍性,因为对于有些问题,当学校观看者完成得不理想时,家中观看者却取得好成绩;有时又会发生相反的结果;有时双方的成绩相同。我们提问的许多问题取自于“比尔·奈”节目中的演示和实验以及该系列片的教师指南,许多结果似乎与特定集的收看有关。

虽然总体上观看节目的儿童比未观看节目的儿童做得更好,但有时少数民族儿童和多数民族儿童做得一样好,甚至更好。有时他们则不如多数民族儿童。至于性别方面,调查结果也比较复杂:在观看节目之前,女孩回答问题的能力往往不如男孩,但是到调研结束时,女孩开始赶上了男孩。

以下是我们对调研结果的讨论,并呈现了一些所收集的数据。

演示问题:了解基本的科学概念

在观看了两个(最多六个)科学演示之后,孩子们被要求对节目中演示的科学现象选择一项或多项解释,并在该演示实验和其他科学概念之间建立联系。可选择的答案范围较广,从重述与科学原理的明显的联系到抽象的联系。此外,每个问项中均包含两个错误的答案。通过统计正确答案和错误答案的数目,从而计算出孩子们回答这些演示问题的得分。

概要

总的说来,在回答这些科学演示所依据的科学原理上,家中观看者和学校观看者的能力较之观看前发生了很大变化。对于针对科学演示所提的各个问题,对照组则没有显示出明显变化。有关帕努利原理(飞行原理)的演示所提的问题,家中观看者和学校观看者都回答得很好。该原理出现于“鸟类”和“飞翔”两集节目中,学校观看者和家中观看者都被鼓励观看了这两集节目。

对从气球中排出空气这一演示所依据的科学原理,只有学校观看者增进了了解。这项演示出现在该系列片“运动力”一集的节目中,而我们提供给学生的录像带中并没有包含这集节目。

与学校观看者相反,在针对概念演示所提的问题上,家中观看者往往获得显著的进步。家中观看者得分好的问题往往是“比尔·奈”节目中实验和演示的直接复制(例如海洋哺乳动物利用一种脂肪物质来保持体温);大多数的这些问题取自于我们鼓励儿童观看的那些集节目。总的说来,在家观看该系列片的儿童看的节目集数也往往更多些,很可能曾看过至少一次有关的实验演示。

总体上,与男孩相比,女孩在演示问题上得分往往较低。对于一些问题,女孩和男孩都取得了进步,但大多数情况下,女孩的成绩赶不上男孩。

以下介绍并讨论我们针对演示提的问题及有关结果。每个问题项侧重于某一科学概念,这一科学概念在该系列片的某一集或多集节目中曾解释过。

概念 帕努利原理(飞行原理)

当你对着一张纸吹气时,看看会发生什么情况。这或许有助于我们解释(圈选你认为尽可能多的回答):

- a) 飞机是如何飞行的 +
- b) 当你对着一张纸吹气时,发生了什么 +
- c) 纸有多么结实 -
- d) 气压 +
- e) 如何改变温度 -
- f) 鸟类是如何飞翔的 +

与观看前相比,家中观看者和学校观看者在回答有关帕努利原理的问题上都获得了显著进步。在观看过程中,该概念被演示了两遍,这两次演示出现在我们鼓励儿童观看的“飞翔”和“鸟类”节目中。我们发现,无论是多数民族观看者还是少数民族观看者(在家中和在校内),无论是男孩还是女孩,这一进步都是明显的。

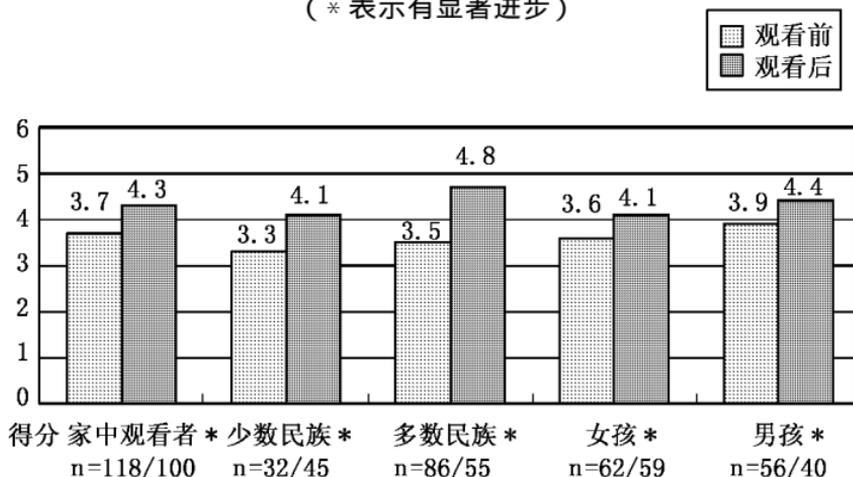
按6分量表,家中观看者的得分从观看前的3.7增加到观看后的4.3,其中男孩(3.9~4.4)和女孩(3.6~4.1)进步都很明显。但男孩观看前后的得分均高于女孩。

学校观看者的得分从3.3显著地增加到3.9。所有年级的学生都有进步,但基本上是少数民族学生(从3.3增加到4.0),多数民族学生没有看出有任何进步。在学校观看“比尔·奈”节目的学生中,男孩的得分没有提高,女孩则进步明显,她们的得分从3.0增加到4.0,赶上了男孩(男孩观看前的得分为3.8)。

与观看前相比,观看班级在观看后更多地把这个演示与“飞机是如何飞行的”联系在一起,其人数比例从开始时的31%增加到结束时的51%。而且,他们对气压这一概念有了更好的了解。

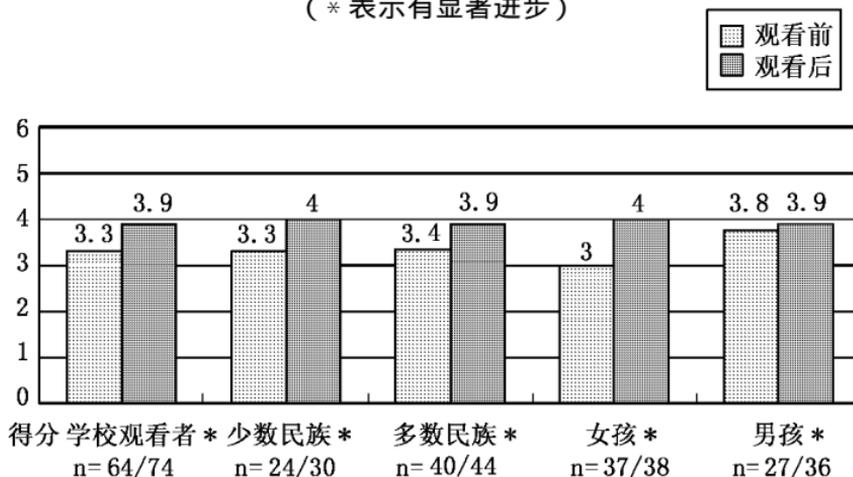
帕努利原理 :家庭观看者

(* 表示有显著进步)



帕努利原理 :学校观看者

(* 表示有显著进步)



概念 :气压 /推力

我们给一只气球充气,然后放手,让里面的空气排出,看看会发生什么。这可能有助于我们解释:

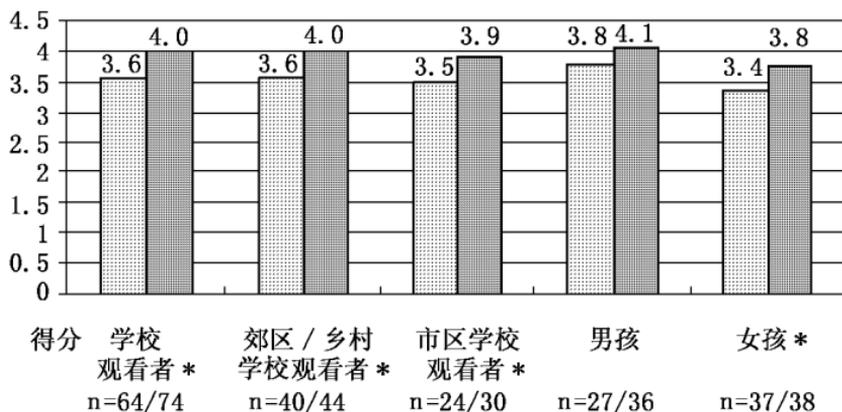
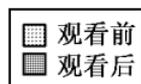
- a) 当空气从气球中排出时会发生什么 +
- b) 火箭是如何工作的 +
- c) 鱿鱼是如何游动的 +
- d) 什么是气压 +
- e) 肌肉为什么能起作用 -
- f) 种子是如何从树上掉下来的 -

学校观看者在观看了节目后对这一问题的回答有了显著进步(按6分量表,得分从开始时的3.6增加到4.0),很多学生知道了气压这一基本概念。在这一问题上,多数民族学生进步明显(得分从3.6增加到4.0)。虽然少数民族学生也有进步,但是他们的得分提高得不多。

在学校观看节目的学生当中,男孩在这个问题上没有取得显著的改善(3.8 ~ 4.1);而女孩进步则大些(3.4 ~ 3.8),与男孩的差距缩小了。尽管很少有观看者能够将气压原理运用到动物界,例如鱿鱼是如何游动的,但很明显,女孩在观看系列片之前更不可能有这样的想法,而在观看之后她们的理解与男孩没有什么差别了。

气压演示 :学校观看者

(* 表示有显著进步)



概念 绝缘体

我们的一只手涂上猪油或起酥油,另一只手不涂。然后将双手放入冷水中。没有涂猪油的手会感到更冷。这可能有助于我们解释(圈选尽可能多的正确答案):

- a) 为什么猪油会使我们对东西的感觉困难了 +
- b) 海豹是如何保持体温的 +
- c) 如何使你的房子凉快些 +
- d) 毛毯为什么能使你保暖 +
- e) 为什么油和水不能混合 -
- f) 泥浆是如何形成的 -

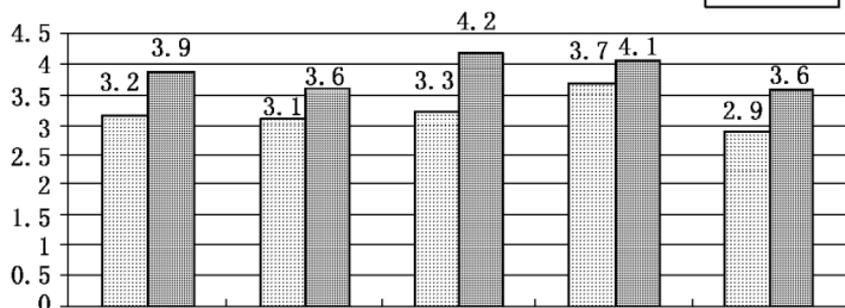
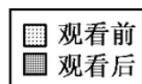
我们发现,家中观看者对这一问题的回答有实质性的增加。按6分量表,家中观看者的得分从观看前的3.2上升到观看后的3.9。其中少数民族观看者和多数民族观看者进步都很明显,他们的得分分别从3.1增加到3.6、从3.3增加到4.2。但两者的差距并没有缩小。

在家看节目的孩子中,男孩和女孩都进步明显(分别从 3.7 增加到 4.1 以及从 2.9 增加到 3.6)。然而,女孩在观看节目后的得分仍然明显地低于男孩。

学校观看者在观看“比尔·奈”节目后,在这一问项上的得分并无明显提高。

绝缘体问题测试:家庭观看者

(* 表示有显著进步)



得分家中观看者 * 少数民族 * 多数民族 * 男孩 女孩 *
 n=122/131 n=29/32 n=93/59 n=33/43 n=73/72

概念 板块构造学说

我们把一个盒子装满沙子,然后将其侧面向一块推挤,看看会发生什么。这可能有助于我们解释(圈选尽可能多的正确答案):

- 当地震发生时,如何形成了山脉 +
- 如何做猫砂盒 -
- 当大陆发生迁移时,有时会发生什么 +
- 沙子如何变成岩石 -
- 能量是如何从一种形式转换为另一种形式的 +
- 将沙子放入盒子里,然后推挤盒子,会发生了什么情况 +

对这一问题,只有家中观看者进步较大,他们的得分从 2.9 增加到 3.2,其中多数民族儿童明显强过少数民族儿童。主要进步

体现在孩子们认识到 ,这是板块构造学说和地震的一个例子。

还有两个演示问题 ,一个是关于排水的概念 ,另一个是关于向心力和离心力的概念。对这两个问题的回答 ,观看者没有产生任何可确认的变化 ,所以在这里不作讨论。

假设问题 科学知识以及科学概念间的关系

为进一步了解儿童的科学知识以及理解和解释科学概念的能力 ,我们请四、五年级的学生回答一个问题的三种情况 ,这样为学生们提供了一种开放式回答问题的机会。这一问题是 :如果你遇到一位外星人 ,你会告诉它我们陆地(或者水或者生物)的一些什么事情?对这一问题的回答使我们得以掌握学生对概念性知识以及概念间关系的了解情况。

我们根据学生们回答的科学性或非科学性以及回答的基础性或复杂性来计算得分。例如 ,当被问到有关生物的问题时 ,“生物会生长”的回答将作为具有科学性和基础性来记分。如果回答说 :“大多数生物的生长都需要水” ,那么这个回答作为具有科学性和复杂性来计分。如果回答说 :“一些生物很好玩儿” ,那么这可能被作为非科学性的回答来计分。

概要

我们得到的确凿证据表明 ,孩子们不仅逐渐地能够讲出地球上水和生物的一些事实 ,而且懂得他们所说的这些地球要素之间有着较为复杂的联系。该结果似乎与孩子们观看有关节目的机会多少有关。“比尔·奈”系列片中关于水和生物的节目比关于地球的节目更多 ,观看的儿童也更广泛。所讨论的假设问题如下 :

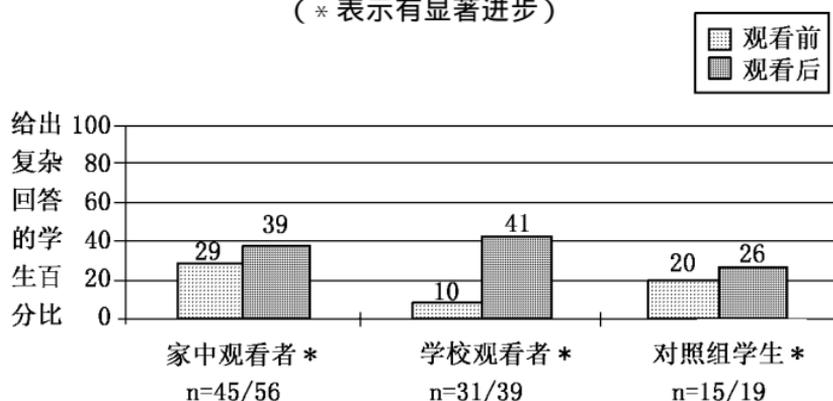
水 我们让孩子们讲出有关地球上水的一两件事情。家中观看者对该问题的回答在复杂性上有显著的增加。在观看节目之前 ,只有 31%的家中观看者给出了复杂的回答 ,而到调研结束时 ,

53%的家中观看者给出了复杂的回答。至于学校观看者或对照组学生,对该问题的回答在观看前和观看后并没有显著的差别。

生物:我们让孩子们谈谈地球上的生物。学校观看者和家中观看者对这一问题的回答都有了明显进步。在调研开始之前,29%的家中观看者给出了复杂的回答,而到调研结束时,39%的孩子们给出了复杂的回答。至于学校观看者,起先只有10%的学生给出了有关生物的复杂阐述,而到调研结束时,41%的学校观看者给出了复杂的回答。与之相比,对照组对该问题的回答前后没有显著的改变。

如果你遇到一个外星人,你会如何向他讲述地球上生物的事情

(*表示有显著进步)



陆地和大陆:当我们让孩子们向外星人解释陆地和大陆时,家中观看者、学校观看者和对照组的学生都没有产生可衡量的变化。这可能是因为我们向学校和家庭提供的录像带中含有4集有关水的节目和若干集有关动物的节目,但只有1集是有关陆地的节目,因而我们不大可能看到这方面的进步。

动物的属性问题:事实学习和传统的科学思维

有关动物的主题在“比尔·奈”系列片中经常出现。为了解孩

孩子们从中学到了有关动物特征的哪些知识,我们向孩子们提出了关于某些动物独特特征的问题。我们从他们所回答的动物特征的准确度和回答的复杂性两个方面对其打分。例如,对于动物,我们让孩子们解释是什么特征使得哺乳动物称之为哺乳动物,或者是什么特征使得鱼类称之为鱼类,或者是什么特征使得鸟类称之为鸟类。然后,我们根据他们列举的某类动物的种种属性,按正确答案的多少对其结果打分。我们期望看到的回答是惯常使用的科学概念和词汇,而不是笼统的回答。比如,如果一个孩子回答哺乳动物是温血动物,她就得1分。如果一个孩子答出哺乳动物是温血动物,并且呼吸空气,那么他就得2分。如果一个孩子回答哺乳动物是“可怕的”,他就不得分。

概要

在观看节目之前进行的问卷调查中,儿童对“什么特征使得鸟类称之为鸟类?”和“什么特征使得鱼类称之为鱼类?”的回答得分很高。大多数儿童(至少85%)能够恰当地回答这个问题。尽管观看节目后他们又有所进步,但是孩子们在观看节目之前的得分已是如此之高,因此在调研结束时不可能出现大的进步。

然而,对于“什么使得哺乳动物称之为哺乳动物?”的问题,孩子们在观看之前的得分则低得多。因此,我们发现家中观看者和学校观看者对这个问题的回答均取得了显著进步。对照组学生对这一问题的回答没有发生显著变化。

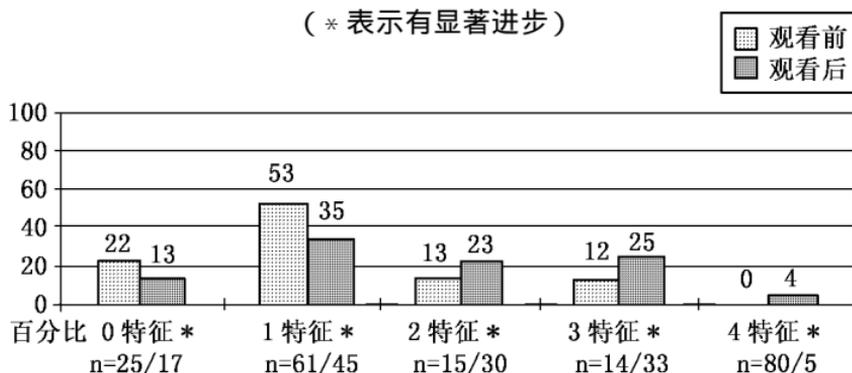
哺乳动物:观看者对有关哺乳动物属性的了解增加了,可能是由于节目中有各种被感官知的现象。哺乳动物是家庭和学校观看者最经常观看的节目之一。这集节目不断地重述哺乳动物的属性:温血、有毛发、给它们的动物宝宝喂奶、生子(在整集节目中出现了约10次)。此外,我们也让儿童观看“海洋哺乳动物”一集的节目,在这集节目中“比尔·奈”又重复讲述了“哺乳动物”一集中讲

过的许多知识。相对于其他两种动物的节目——“鸟类”和“鱼类”，哺乳动物节目中事实的重复显然多得多。

至于家中观看者，回答正确的百分比从 78% 增加到 87%。此外，在调研结束时，家中观看者回答正确的百分比越来越高。答出两个或更多属性的儿童的百分比从 25% 明显地增加 52%。至于学校观看者，给出正确答案的百分比从 77% 增加到 93%，给出两个或更多正确答案的儿童从 22% 明显地增加到 45%。

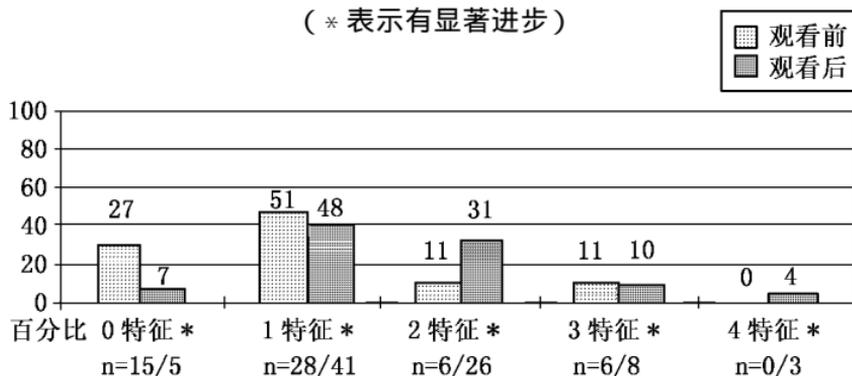
哺乳动物的特征：家庭观看者

(* 表示有显著进步)



哺乳动物的特征：学校观看者

(* 表示有显著进步)



应用性问题解答：高层次思维和科学解释

我们也给孩子们准备了一些应用性科学问题，并提供了多项

可供选择的答案,以有助于他们解释有关的现象。供选择的答案既包括不可能的答案也有复杂的答案,我们要求孩子们选出所有的正确答案。比如,一个问题是问骑自行车下坡时如何骑得更快,选项包括“使道路潮湿、快骑、换挡、俯下身以便风不会阻挡你”等。

概要

对于那些具有为儿童所熟知的参照物的问题(例如骑自行车的物理原理),我们常常发现,孩子们的回答在观看前后有了显著变化。

对于每一个较具体的问题,到调研结束时,观看组进步明显。有时家中观看者进步最大,有时学校观看者提高最多。

女孩一向进步明显。通常在调研开始时,女孩对各个问题的得分都低于男孩,但在观看节目之后,她们进步很大,最后得分已与男孩不相上下。

以下是我们收集的有关科学解释问题的调研结果。我们提出的每一问题都针对某一科学概念(主要的科学概念),这样的概念在该系列片的某集或多集节目中曾阐释过。

概念:重力

当你骑自行车时,为什么下坡容易上坡难?

(可以圈选你认为的多个正确答案)

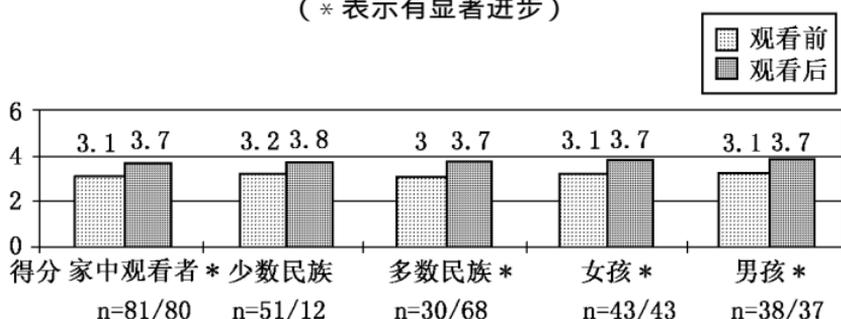
- a) 因为向上困难,向下容易 -
- b) 由于重力的作用 +
- c) 因为当你在高处时,难于做事 -
- d) 因为坡是斜的 +
- e) 因为地球拉着你 -
- f) 因为下坡时风更大 +

为描绘重力的概念,我们提出了这个骑自行车上下坡的问题。家中观看者和学校班级观看者在观看节目前对这一问题的回答通常过于简单,而在看完节目后,这一情况大大减少了,孩子们对重力的概念有了较复杂的理解。在家看过节目的学生比未看节目的学生显然更能发现导致这一现象的原因是重力。

我们发现,家中观看者的得分从开始时的 3.1 显著增加到 3.7 (按 6 分量表),这一进步很大程度上归功于多数民族观看者,他们的得分从 3.0 提高到 3.7;少数民族观看者的得分从 3.2 提高到 3.8,但是这一变化不具备统计上的意义。男孩观看者和女孩观看者都取得显著进步,得分分别从 3.1 增加到 3.7 以及从 3.2 增加到 3.7。

重力问题测试 家庭观看者

(* 表示有显著进步)



概念 运动/空气动力学

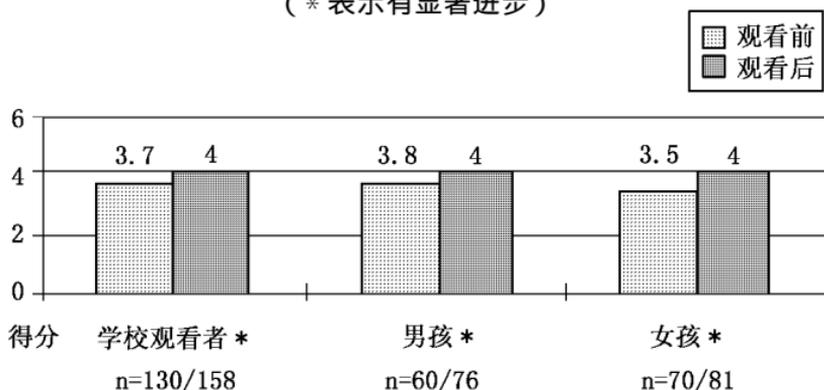
解释你如何能更快地骑车下坡。

- 俯下身去,以减小风的阻力
- 弄湿路面,使道路光滑
- 踩车更快
- 换挡
- 直立,以便风推你前进
- (使身体的重心)偏向自行车一侧

对于这一问题,只有学校观看者取得了显著进步,他们的得分从3.7提高到4.0(按6分量表)。学校观看者中,少数民族观看者进步超过了多数民族观看者(后者没有什么变化),他们的得分分别从3.4增加到4.3以及从3.8增加到3.9。男孩没有取得显著进步(观看前为3.8,观看后为4.0),而女孩进步明显(从3.5增加到4.0),并且赶上了男孩。

运动/空气动力学问题测试:学校观看者

(*表示有显著进步)



其他两个问题一个是有关江河对环境的价值,另一个是有关湖泊中的船艇失踪。所有观看者对这两个问题的回答在观看前后都没有任何实质变化。调研初的调查中学生们得分接近最大值,调研末的调查中学生们得分达到了最大值。

访谈和动手型活动问题:探究和解释科学概念的能力

我们也对儿童进行了看前和看后访谈,请他们探究和解释某些科学现象和一般的科学概念。调研人员将用于科学探究的材料分发给学生,供他们做进一步的观察和研究。

方法

我们设计这种一对一或一对二的访谈,是为了判断学生是如

何将科学看作为对现象的探究的。我们还想研究孩子们如何根据其观察所得的证据和已有知识来提出假设,以及他们如何与日常生活中的科学概念联系起来。我们为儿童提供科学材料,请他们模拟“比尔·奈”(或针对对照组儿童的其他科学节目)在电视上会怎样使用它们。要做好这件事,学生们必须进行观察、探究和调研。

通过一些访谈,我们了解孩子们对科学中的“主要概念”的思考,这种思考很可能因观看“比尔·奈”节目而得以增强。在两个不同的问题中,我们通过让孩子们依据各种特征对动物进行分类,以探寻他们对科学概念的一般性理解,并探寻他们对动物共有特征的更复杂的了解。我们也探寻孩子们如何利用现有知识,提高对地球结构和演化过程的认识。

概要

调研初始,约 387 名儿童参加了访谈,在调研末,462 名儿童参加了访谈。参加访谈的孩子积极参加了我们安排的动手型活动,通过看节目,对我们所提问题的回答变得复杂和全面了。孩子们的这种进步很多来自于强化的探究活动。参加的孩子摆弄并观察物体,提出假设,并进行尝试。调研者由此可以看出儿童解释科学概念的能力。

总的调研结果充分肯定了“比尔·奈”系列片的影响。对于每个动手型活动的测评,家中观看者或学校观看者都认为,动手型活动提高了他们主动探究、观察和生成假设的效果。我们发现,家中观看者有时候做得比学校观看者好,有时候则差些。对某些问题的回答,家中观看者和学校观看者都取得了非常突出的进步。

在试图搞懂一件物体可能用来做什么以及如何用来演示一个科学概念时,观看者很可能经历了各种各样的科学过程。看过节目的孩子采用较有效的方法(利用动物显而易见的特征)来对动物

分类,他们在回答动物分类和有关地球的特征的问题时,同样利用了较多的科学事实作为他们回答问题的依据。

探究和解释问题的评分

我们从以下四个访谈问题中选择两个问题问孩子,包括看节目的孩子和未看节目的孩子。这些问题是:

问题 1 材料/弹跳问题

“比尔·奈”节目(或其他儿童科学节目)正在寻找能在电视上进行科学展示和表演的东西。最近我们找到一些我们认为比较有趣的东西,但我们拿不准该如何在电视中展示它们。让我给你看一看这些小球,它们看上去很相像,但是如同我们所发现的那样,它们彼此十分不同。(让每个球在硬地面上弹两次,以显示其不同。然后把球递给孩子们,让他们玩。)(提示:一个球在硬地面上能弹起来,而另一个则不能。)

首先让我们看看我们会发现什么情况。

你认为这些适于在电视上表演吗?为什么?

如何使用它们?

对此你还有什么想说或想做的吗?

假设你就是节目主持人,你会说些或者做些什么?

问题 2 陀螺/潜望镜

“比尔·奈”节目(或其他少儿科学节目)正在寻找能在电视上进行科学展示和表演的东西。最近我们找到一些我们认为比较有趣的东西,但我们拿不准该如何在电视中展示它们。给儿童三件物品中的一件,让他们表演如何使用。例如,调研人员给儿童看一个陀螺,使陀螺旋转,然后提问:

你认为它适于在电视上表演吗?为什么?

如何使用它?

对此你还有什么想说或想做的吗?

假设你就是节目主持人,你该说些或者做些什么?

问题 3 动物分类

我们为孩子们准备了一套动物图像卡片,卡片上有人、猫、蝙蝠、鲸、鲨鱼、鱼、龙虾、鸭、蝴蝶、蚂蚁和蜘蛛等。我们告诉他们,一些科学家基于动物的相似性和差异性,将动物分成纲、目和科等不同等级的群组。在这一活动中,我们希望孩子们充当科学家的角色,按照卡片上各个动物的特征的异同,把这些动物分成3个或更多的群组,并让他们列举尽可能多的理由,以说明自己为什么那样分。然后,我们分析他们的思路,并询问他们是否可以用其他方式进行分类。最后,我们假想一种动物,让孩子们为其归类,并解释其理由。

问题 4 地球科学

我们为孩子们准备了一个地球仪,告诉他们一些科学家在研究地球,以弄清楚有关地震、河溪流、水循环、湿地以及地球季节、气候、大气和海洋学等方面的情况。孩子们知道有关这些事情的情况吗?如果让他们开展一个科学项目的话,他们愿意谈论哪些?调查孩子们运用的哪些基本概念是来自“比尔·奈”节目中的主要概念。

我们采用基准条规法(grounded rubrics)计算孩子们在这些问题上的得分。对于这些探究和解释问题,我们留意学生们是否做了观察、比较和实验,以及是否根据他们的观察得出解释。

我们给孩子们的最后总分是根据他们在以下三单元的各自得分计算的:探究活动、解释的质量以及与观看“比尔·奈”节目的联系。

例如,在探究单元,我们的评分是根据孩子们在进行多种观测、与其他事物进行联系与比较以及提出和完成实验方面的见解和能力得出的。比如,在有关材料的问题中,如果一个儿童在地上拍球,她将在探究部分得1分;但如果她在不同的地面上拍球,她

将得 2 分。

在解释单元 ,如果孩子们提出了证据 ,或者提出了一项实验 ,或者进行了一项实验 ,他们将得分。例如 ,在潜望镜活动中 ,如果学生说出潜望镜是利用了镜子反射的原理 ,他将得 1 分。如果他解释了镜子的位置并进行了实验 ,他将得 3 分。

在与“比尔·奈”节目的联系单元中 ,如果孩子们能够将有关的现象与“比尔·奈”系列片中的有关节目联系起来 ,他们就会得分。例如 ,如果孩子提到陀螺持续旋转 ,正像“比尔·奈”系列片中“使物体旋转”那集节目中介绍的物体一样 ,她将得 1 分。

对于分类问题(问题 3) ,我们根据孩子们用来进行分类的动物属性的类别来计算得分。例如 ,在这个问题上得分低的儿童 ,可能是根据动物的腿的数量将动物分类的。将动物正确归类(如哺乳类、鱼类、昆虫和鸟类)的孩子将得到最高分 4 分。对于地球科学问题(问题 4) ,我们的评分要看孩子们是否利用了其已有知识或者“比尔·奈”节目中有相关的事实信息。得低分的孩子的回答不具有科学性 ,例如“我们有四季” ,或者“地震在加利福尼亚发生”。而回答说“地球一年绕太阳转一圈 ,它歪着身子转 ,有时它距离太阳近些”的孩子得分会较高。

以下是我们针对学生探究和解释科学概念的能力所进行的访谈和动手型活动的调研结果。调研中提的每一问题都针对某一科学概念 ,这样的概念在“比尔·奈”系列片的某一集或很多集节目中曾阐释过。

概念 :材料 / 弹跳问题

在这一概念问题上 ,家中观看者取得的进步最明显。虽然家中观看者和学校观看者开始时的综合得分基本相同 ,分别为 5.6 和 5.7 ,但到后来 ,家中观看者的综合得分增加到 7.1(15 分量表)。在三单元的得分中 ,家中观看者在探究现象和与“比尔·奈”

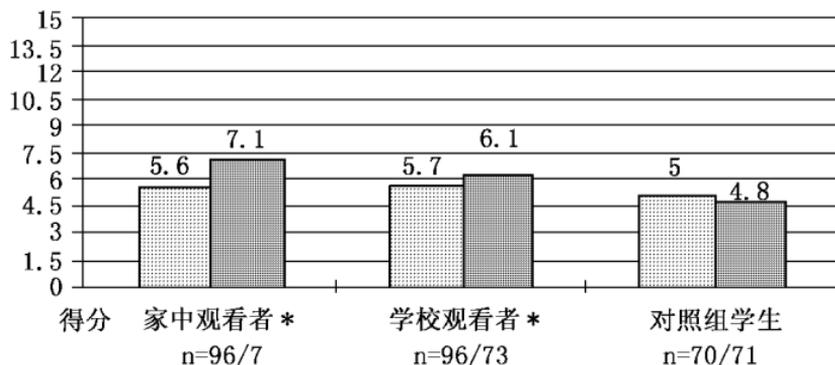
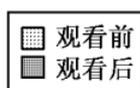
节目发生联系两个单元上进步最大。

男孩家中观看者和女孩家中观看者都明显地提高了他们的得分。女孩的得分从开始时的 5.3 提高到 6.0 ,男孩的得分从 6.00 提高到 7.69。

学校观看者也提高了他们的得分(到 6.1 分) ,但不具备统计上的显著意义。对照组的得分则没有显著的提高。

材料 /弹跳问题测试

(* 表示有显著进步)



概念 运动 / 重力 / 旋转陀螺

在这一问题上 ,每组学生(看节目的和不看节目的 ,家中观看者和学校观看者)的得分都有统计上的显著提高 ,不过 ,学校观看组和家庭观看组的学生进步最大。学校观看者的综合得分从 4.6 提高到 8.1(所有三个分析单元里的得分都有显著增加) ,家中观看者的得分从 4.6 提高到 8.1(所有三个分析单元里的得分也都有显著增加)。这些学生的进步主要体现在他们在观看后进行了更多的观察和比较 ,并提出了更多的实验。在观看后的测评中 ,孩子们常常谈到诸如“重力”和“力”的概念 ,并经常提及系列片中“重力”一集的节目。

对照组的综合得分从 3.6 提高到 5.0。与观看前所进行的测评结果相比,未观看节目的学生在后期已能够进行一些观察和比较。

概念 光线 / 光学 / 潜望镜

我们发现所有三组儿童的得分都有显著的增加,而家中观看者和学校观看者的收益比对照组大。家中观看者的综合得分翻了一番,从 3.4 提高到 7.0,在所有三个分析单元都显示出进步。学校观看者的综合得分也翻了一番多,从 3.5 提高到 8.0,在所有三个分析单元都显示出进步。对照组的得分提高了大约 70%,从 3.4 增加到 5.8。

对于学校观看者,其进步多体现在看过节目后,学生们能更多地观察事物并把该事物与其他事物进行比较,能进行更多的实验。在调研末,在描述潜望镜时,孩子们谈到了镜片角度和镜片。一个孩子说到,“镜片会反射。你看一个镜片,这个镜片会让你看到另一个镜片的图像。”

概念 波

学校观看者进步明显,其综合得分从 5.4 增加到 7.7。进步部分地体现在学生们能更多地进行观察和探究,并提出了更多的实验。家中观看者看后得分与看前相比没有明显提高。

概念 动物分类

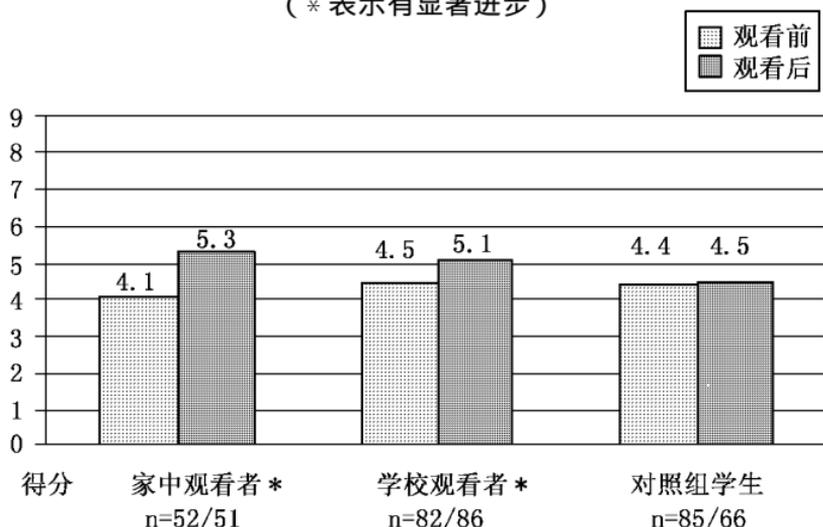
在这项测评中,家中观看者和学校观看者都取得了显著进步。按 9 分量表,家中观看者的得分从 4.1 提高到 5.3,在所有分析指标上都取得了改善。学校观看者的得分从 4.5 提高 5.1。对照组则没有显著的进步。家中观看者在这项测评中表现特别出色。

学校观看者组的得分增加主要归功于那些少数民族学生,他

们追上了他们的多数民族同伴。少数民族学生的得分从 3.6 提高到 5.3。在家观看节目的女孩的综合得分从 4.4 提高到 5.6,在学校观看节目的女孩的综合得分从 4.2 提高到 5.1。在家观看节目的男孩的得分从 3.9 提高到 4.9,在学校观看节目的男孩的得分从 4.9 提高到 5.3。

动物分类测试

(* 表示有显著进步)



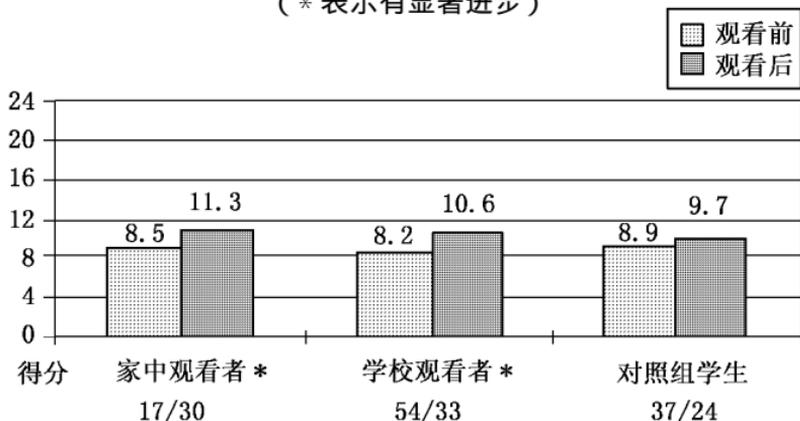
概念 地球科学

当我们给孩子们一个地球仪并让他们谈些有关地球的事情时,孩子们常常以社会和政治问题以及地球科学方面的信息回答我们。孩子们在地球科学方面的知识大都来自于社会研究话题(即 地球有七大洲,有不同的国家,有不同的季节)。他们中很少有人能说清楚构成季节和环境的概念。

我们只在费城地区得到了关于这一问题的可靠的、一致的数据,它表明“比尔·奈”观看者探究和解释科学概念的能力有了提高。学校观看者取得了显著进步(按 24 分量表,从 8.2 提高到 10.6),家中观看者也进步明显(从 8.5 到 11.3)。

地球知识测试 :费城地区

(* 表示有显著进步)



看节目的孩子对科学有何行为反应？

虽然我们没有直接调查孩子们与科学相关的行为,但我们确实询问过儿童:如果给他们机会,他们可能选择的行为的种类。我们让孩子们选择他们愿意从事或研究的各种非科学类和科学类活动,以及选择他们喜爱的书籍或电视节目,我们从中研究他们的行为偏好。我们也询问他们在过去的两个月是否进行了任何科学相关活动,以及他们愿意或不愿意从事科学工作的原因。

我们发现,在我们的调研过程中,家中观看者和学校观看者的行为偏好没有发生显著变化。不过,他们的回答大都反映出他们对科学的积极态度和对从事科学活动的兴趣。这些学生的家长和教师报告说,孩子对科学和科学活动有浓厚的兴趣,通过参与调研,他们对科学和科学活动的兴趣增强了。大约一半的儿童报告说,他们愿意搞科学活动,例如研究海洋,而不是在海洋中游泳。这一比例在我们整个的调研过程中没有发生变化(也可能是因为在很多的调查问题中,非科学类选项对孩子的吸引力比我们最初认为的要大)。

我们注意到以下倾向:

- 孩子们报告说,他们对科学以及对进行科学活动有浓厚的兴趣,这与家长和教师们反映的情况一样。
- 与未观看节目的学生相比,看“比尔·奈”节目的学生更倾向于选择在暑期进行科学活动而不是其他非科学类活动。
- 男孩和女孩想在电视上作个人公开露面的话,“比尔·奈”节目是最普遍的选择。

当我们询问学生是否在家进行过什么很酷的科学活动时,我们发现,超过三分之二的孩子称他们在家中进行科学活动。大部分家庭科学活动是老师布置的家庭科学作业以及孩子们要在家继续做的学校发起的科学活动。在家看该系列片的学生的家长也说,他们的孩子在家做过大量的实验和科学活动(参见以下的父母和教师部分,可以获得有关这方面的更详尽的信息)。

根据学生问卷的调查结果,学生们看前和看后测评都显示,在我们让他们选择喜欢的科普书时,他们选择最多的书是《在家制作科学物品集》,其次是另一本科普书《海洋中的生物》。

在学年末,当被问到他们最愿意看哪个电视节目时,有更多的男孩和女孩选择了“比尔·奈”节目,数量超过了其他节目(包括“卡门圣地牙哥”、“代笔人”、“许愿骨”、“家务事”、“贝克曼世界”)。男孩的首选是“比尔·奈”节目(49%),其次是“家务事”(20%)。女孩的首选也是“比尔·奈”,其次是“代笔人”(21%)。

学校观看者也选择“比尔·奈”作为他们最想看的节目(从看前的18%上升到看后的44%)。另一方面,对照组(未观看节目的学生)的首选则是“家务事”(在调研初及调研末均为50%)。

在学校看过“比尔·奈”节目的孩子更愿意选择科学活动作为他们在夏季中的首选活动。他们常常选择找化石、收集石头和计算机控制的机器人之类的活动,而不是运动、艺术和戏剧。

看“比尔·奈”节目是否改变了孩子们的 科学态度和科学意识？

我们设计了一组问题,用以调查儿童对科学和科学家的态度和信任以及学生们对他们周围的世界利用科学的感受。一些问题针对儿童的认知定型。我们问孩子能在哪里开展科学、什么类型的职业使用科学、科学家具有什么类型的特征等。

- 我们发现,学生对科学的态度没有因为观看“比尔·奈”系列片节目而发生多少变化。当然,起初就对科学持积极态度的学生在看过节目后不可能发生什么变化。

- 孩子们似乎正在摒弃传统的有关科学和科学家的成见。

- 性别差异似乎由来已久。男孩通常将科学家画成男孩或男人模样,而女孩则倾向于将科学家画成女孩模样,几乎没有女孩把科学家画成一位妇女科学家。

- 看过“比尔·奈”系列片节目的孩子越来越多地将科学家画成活跃的、忙碌的人,但不是整天呆在实验室里。

- 看过“比尔·奈”系列片的少数民族学生对各种职业利用科学的情况有了更多的认识。

总的说来,没有什么指标显示儿童在科学态度方面发生了显著变化。当然,儿童对待科学的态度已经是非常积极的,不可能因观看“比尔·奈”再有什么改善。

当我们让孩子们回答“我想学习科学”或“我不想学习科学”这一问题时,绝大多数的孩子(90%以上,包括家中观看者、学校观看者以及对照组学生)回答说想要学习科学,最普遍的理由是“它很酷、很有趣,而且充满乐趣”,其次是“想了解生活中的物理常识”。

科学与工作

由于观看了“比尔·奈”系列片,一些学生认识到科学在各种职业和工作中所起的作用,他们的视野变宽了。我们问孩子们,科学知识对于各种职业是否是必要的。少数民族学生对这类问题的回答在看过“比尔·奈”系列片后取得了明显的进步(得分从看前的3.7提高到看后的4.6),他们对各行各业需要科学的认识提高了。多数民族学生对这类问题的回答在看后也取得了明显进步,得分从看前的4.5上升到看后的5.2。由于少数民族学生看前的得分较低,因此尽管看后有较大的绝对增长,但还是没能赶上多数民族学生。

使用科学的人

下列哪些人的工作需要科学?(可以圈选多个答案)

- a) 午餐菜单设计者
- b) 理科教师
- c) 校长
- d) 校医室的护士
- e) 看护植物与树木的人
- f) 兽医
- g) 邮差

我们也问孩子们一系列有关使用科学的人的问题。女孩比男孩更多地认为科学家要外出工作(60%对52%)。有此看法的女孩从观看节目前的49%提高到观看后的60%。

画一位科学家测评

- 在看前和看后的测评中,几乎所有的男孩(95%)画的都

是男性科学家。

- 在看前和看后的测评中,大多数女孩(67%)画的是女性科学家,并且所画的科学家往往是个女孩;很少的女生(不到15%)把科学家画成成人妇女模样。

- 与看前测评相比,看后测评时把科学家画成成人模样的孩子减少了。

让儿童画一幅科学家工作的图画,我们希望由此发现,在经常观看“比尔·奈”节目之后,孩子们对科学家的认识有什么变化。对孩子们所画的科学家的研究表明,孩子心目中的科学家有着许多刻板化的东西,如实验室大褂、试管、胡子和眼镜。

我们要求儿童画一个正在进行科学研究的科学家,然后给这个人取个名字,并为图画注上一个标题。作为看前测评的一部分,所有参与调研的儿童都画了一幅画。在看过节目后,所有儿童还要再画一幅画,作为看后评估的一部分。我们要求他们为画中人物取名,并且给他们的图画加个标题,以便我们可以确定画中的科学家是男性还是女性,同时更好地理解画中的科学家正在做些什么。

我们然后从两方面分析孩子们画的画:画了刻板化的东西(如实验室工作服、眼镜、实验室设备、科学产品)以及用了刻板化标题(例如“尤里卡”)的孩子得1分。我们还分析了画中的其他东西,包括科学家的性别、活动场所是在户内还是在户外以及科学家是在主动工作还是被动工作等。我们也分析了画中的主人公,特别要看孩子们的画中人是否是他(或她)自己,还是另一名儿童、一个成年人、比尔·奈、贝克曼、一个怪人或“科学狂人”。

根据我们所使用的分析判据,在图画的内容或性质方面,我们没有发现显著的变化。正如以前的调研者已经发现的那样,孩子们还是画有关科学家和科学活动的那些老套的东西,例如试管、标尺或烧杯、在实验室(而不是在现实世界中)进行科学研究的科学

家,而且所画的科学家主要是男性。尽管如此,我们还是注意到一些倾向,虽然这些倾向从调研开始到调研结束没有统计意义上的差别,但是它们显示出了某些变化。

例如,在看前和看后的测评中,当被要求画一位正在进行研究的科学家时,几乎所有的男孩(95%)画的都是男性科学家。尽管大多数女孩(67%)画的是女性科学家,但所画的科学家往往是女孩,很少的女生(不到15%)把科学家画成成人妇女模样。

在看后测评中,孩子们所画的主人公也发生了相当大的变化,看过“比尔·奈”节目的孩子中,有大约20%画的是比尔·奈本人。比尔·奈通常被画成节目中的样子,在实验室里或在做实验,如在摆弄一个陀螺仪或拿着一个飞盘。画比尔·奈的男孩比女孩多。

有意思的是,与看前测评相比,在看后测评中,孩子们画的科学家中大人模样的少了,更多出现的人物则是孩子。在调研开始时,25%的孩子的画中人是正在进行科学研究的大人,在调研结束时这一数字减少到16%。同样,有43%的孩子起初画的是正在进行科学研究的小孩,在调研结束时这一数字增加到52%。看来孩子们开始认为他们自己或其他的小孩子也是能够进行科学研究的。

教师们对“比尔·奈”节目的看法

- 每月在教室观看节目约5次。
- 教师播放节目时通常没有一个明确的教学计划,也没有围绕节目设立课程。
- 教师们报告说,他们的学生对科学很有兴趣,非常愿意开展科学活动。
- 教师们称,“比尔·奈”系列片的许多节目配合了他们的课程需要,可以作为有用的教学工具。
- 在学校需要和播放日程之间的协调,是在学校中更有效和

更广泛应用该节目的一项障碍。

背景情况

45名教师(作为调研对象)参与了我们的学校调研,其中27名是观看组的教师,另外18名则属于对照组的教师。这些教师来自萨克拉曼多、费城、印第安纳波利斯和布卢明顿;所在学校有的是在市区,有的是在郊区、乡村和内城区,都是三、四、五年级的教师。在上述三个地区,我们都分别招募了9名观看组的教师和6名对照组的教师。

在调研开始和结束时,我们分别请教师完成了一个问卷调查表。观看组教师中,77%的教师完成了前期问卷调查表,70%的教师完成了后期调查表。对照组教师中,83%的教师完成了前期间卷调查表,67%的教师填写了后期间卷调查表。

观看组和对照组的大多数教师(84%和83%)不具备符合小学基本教育要求以上的科学背景,如主修或辅修过科学或科学教育课程。不到一半的观看组教师(47%)和对照组教师(45%)在过去的三年里曾接受过科学方面的教职员工培训。

几乎所有的教师(94%的观看组教师和89%的对照组教师)在家观看科学节目,如“国家地理”、“发现”频道和“新星”节目。约四分之三(74%)的观看组教师和一半(55%)的对照组教师报告说,他们的教室里有生物,例如豚鼠、蜥蜴及其他观赏动物。三分之二的观看组教师和55%的对照组教师报告说,他们的学校办科学展览。

科学课程

观看组的教师报告说,他们每周用于科学教学的时间约为4.1小时,对照组教师花费的时间略少些,约为每周3.6小时。观看组和对照组教师讲授的科学内容从天文学、动物到水循环和气象,范

围很广。

几乎所有的教师(95%的观看组教师和90%的对照组教师)教的是整装式的科学课程,极少数的教师教的是分门别类的科学课程。当我们让教师们选择最能描述其科学课程的话时,我们得到了不同的回答。

	观看班级	对照班级
我主要使用教科书	11	11
我使用含有动手型活动项目的教科书系列	44	50
我使用动手型科学探究课程	28	11
我主要使用动手型科学工具包	17	22
我主要使用多媒体项目	17	22

教师们也报告说,他们的科学课中融入了多种多样的活动,观看组和对照组在这点上没有什么差别。科学课活动包括演示(观看组51%,对照组49%)、小组项目(观看组48%,对照组51%)、动手活动(观看组50%,对照组50%)和科学展览(观看组56%,对照组44%)等。

我们也询问教师在科学课中使用各种教学器材的情况。几乎所有的教师(94%)报告说,他们在上科学课时使用过电视和录像:四分之一的教师经常使用电视和录像,约70%的教师偶尔使用它们。一半的教师经常使用教科书,而三分之一的教师偶尔使用它们。很少有教师使用计算机技术教科学课:37%的教师从不使用计算机技术,(49%)的教师偶尔使用。教师使用的其他材料包括实验室作业单和科学杂志。他们也把举办客座讲座和开展动手型科学项目纳入他们的科学课中。

观看模式

在参与本调研之前,约四分之一的教师以前在上科学课时曾

使用过“比尔·奈”节目。然而,这种使用是无规律的,时间也不长。

在受调查的教师当中,没有人参加过针对“比尔·奈”节目的教师培训。只有5%的教师曾索要过“比尔·奈”节目的教师指南或节目时间表。

在我们的调研期间,观看组的教师平均每月在班上播放5次“比尔·奈”节目。45%的教师报告说,他们每周播放一次,25%的教师每周播放两次或两次以上,25%的教师每月播放两三次,另外5%的教师每月播放一次。

虽然学校观看者观看节目的频次可能比家中观看者要低,但是每月5次的频次可能更适应小学班级的科学课程。

不同地区的观看情况也有很大的差异,美国中西部和东部地区的下雪天气打乱了课堂的教学计划。由于雪天的影响,教师们有可能放弃科学课,以补上丢下的数学和阅读课。加大激励或许能提高观看次数,但教师们放映节目的频次似乎接近正常值。

我们请观看组的教师说明他们在课堂上是如何使用“比尔·奈”的,并为他们提供了以下选择:“我们观看比尔·奈节目”,“我利用比尔·奈节目讲解科学概念”,“我们所做的实验或活动取自比尔·奈节目”。虽然一些教师不能确切说出他们如何使用该节目,但大多数的教师(78%)告知了我们一些情况。教师们通常为学生放映该节目。放映期间,教师要花费约20%的时间为学生讲解科学概念,用16%的时间组织学生进行“比尔·奈”节目中的实验或活动。

我们提供的“比尔·奈”录像带中有一集讲消化知识的节目。节目中提到坚果等食品。我们用火点燃这些食品,让孩子们观察糖类、非糖物或产生能量的食品的燃烧情况。孩子们对实验瞠目结舌。

我们也询问教师,他们在课堂上采用什么样的科学活动来充分发挥“比尔·奈”节目的教学作用。教师们称,他们开展了多种科学活动,包括动手实验(50%)、演示(50%)、小组项目(49%)和科

学展览(50%)。

在该调研中,教师们采取了多种方式将“比尔·奈”节目用于教学(教师可以选择多项):

利用“比尔·奈”节目的教师

课堂上讨论“比尔·奈”录像片断	85%
利用“比尔·奈”节目讲解科学概念	80%
看后做“比尔·奈”节目上的实验	70%
有时间时看节目	45%
用“比尔·奈”节目作为对学生的奖赏	40%
利用“比尔·奈”教师工具包中的实验	35%
采用“比尔·奈在家跟我做”活动工具包	15%

教师报道说,他们的学生非常喜欢看“比尔·奈”节目。按1~10分量表,他们对学生的观看兴趣打了8.5分。

科学兴趣、科学知识和科学技能

根据教师们的报告,他们所教班的总体科学技能水平并没有显著的提高。当我们让他们按1~10的量表为其所教班的科学技能水平打分时,1代表“完全没有”,而10代表“非常高”。教师对他们所教班的打分为5.5。在调研结束时,我们又一次问了这一问题,此时打分略提高至6.6,但不足以构成统计上的显著差别。同样,教师们对他们所教班的总体科学知识打分在测评前为5.8,在测评后为6.8,测评前后的差别同样称不上明显。

在我们的调研过程中,教师们对他们所教班的总体科学兴趣的打分也基本未变。根据教师们所说,在测评前,他们的班级就对科学有相当浓厚的兴趣,得分为7.9。而在测评后,这个数字略上升到8.2。同样,这个差别在统计学上并不构成显著性差异。

教师们还认为,他们的班级特别愿意进行科学活动(按1~10的量表,得分9.2)。由于看前的得分已经很高,因此学生们在看后的得分不可能再有多大的提高。在看后的问卷中,教师们给其学生的打分为9.3分,几乎与看前相同。

教师对教科学课的兴趣和满意度

总的说来,教师们对讲科学课非常感兴趣。按1~10的量表,1代表“完全没有兴趣”,10代表“十分感兴趣”,他们对自己的兴趣评分为8.9。

在调研过程中,教师们对讲科学课的满意度变化明显。在调研初,教师们为他们的满意度打分7.4(按1~10的量表),在调研末,他们的满意度提高到8.4,增加明显。

在调研前后,教师们花在科学教学的时数没有变化。观看组的教师报告说,他们每周用于科学教学的时间约为4.1小时。对照组的教师报告说,他们每周用在科学教学上的时间约为3.6小时。

“比尔·奈”节目作为教学工具的效果

教师们认为,“比尔·奈”节目在向学生们讲授科学概念上非常有效。按1~10的量表,1代表“完全无效”,10代表“非常有效”,教师们给“比尔·奈”节目的评分为8.5。超过四分之三的教师给“比尔·奈”节目打了满分10分。

就与学校课程的配合而言,63%的教师认为,“比尔·奈”系列片节目的主题只是“有时”配合了他们的课程;25%的教师认为,“比尔·奈”节目的主题很好地配合了他们的课程;13%的教师认为,节目很少配合他们的课程,没有教师认为该系列片一点都没有配合他们的课程。

当被问到“比尔·奈”系列片中是否有些节目可以作为特别有

效的教学工具时,教师们给出了各式各样的回答。他们共提到了多达 25 集的“比尔·奈”节目,这些节目对他们讲授某一具体的科学题目或科学单元很有助益:

“消化”这集节目对我们学习人体和青蛙组织很有帮助。

“岩石和土壤”这集节目非常配合我们的课程。

只有三个教师(13%)反映,“比尔·奈”的有些节目的教学效果不是很好,如脑、地震、旋转物体和波四集节目。他们认为这些集的节目过于高深,他们的学生理解不了。

针对如何使“比尔·奈”节目成为更有效的教学工具,教师们建议了许多方法。很多的建议集中在要使该系列片更方便地为教师使用。教师们询问是否可以在授课时间放映节目,是否可以有他们自己的一套节目以及是否可以预先得到一份电视播放时间表,这样他们会知道什么时间播放哪些集的节目。

我喜欢在年初制定我的科学单元教学计划,如果能预先知道“比尔·奈”具体节目的播放时间表,我就可以将“比尔·奈”节目纳入到我的教学计划中。

教师们希望“比尔·奈”节目能更多地基于学校的科学课程,他们要求节目能更密切地配合他们的课本。他们也希望节目的播放时间安排能在整体上配合他们的科学课程的范围和单元次序。一名教师认为“比尔·奈”节目中的实验不容易在课堂上做。

我再说一下,如果“比尔·奈”节目能以学生的科学课程为基础,它会成为非常有价值的教学工具。因为它能演示课堂上演示不了的概念实例。

一些教师还表示,他们非常渴望该系列片能开发更多的外延材料(如课程相关读物),他们希望得到更多的现有外延材料:简讯和活动工具包。一些教师希望给学校更多的宣传读物。一名教师要求有一个教师工具包,其中不仅包括教师指南,也要包括一份录像带。

当我们问教师们,在本次调研过程中,他们所教的班级是否有与“比尔·奈”节目相关的具体事情时,教师们的回答积极踊跃。一些教师报告了该系列片如何增加了学生们的科学知识:

我用“比尔·奈”的一个节目录像为学生们讲解和复习“热”单元的知识。学生们在测评中做的非常好,显示出他们对传导、对流和辐射的概念有了极好的理解。

我吃惊于他们学到的专业词汇,如地壳构造板块等等。

其他的教师提到,他们的学生是多么地喜欢该系列片。

当我提及比尔·奈的名字,班上的大多数学生都表现出对他的热爱。

每当班上的学生观看“比尔·奈”节目时,他们就齐唱“比尔·奈”的片头曲。

观看组的教师介绍说,由于观看了“比尔·奈”节目,学生们进行了许多课程外和课堂外活动。孩子们在家里常常同家人一起观看“比尔·奈”节目。他们向老师讲节目上的事情,把在家里做的实验带到班上展示,并且通过“比尔·奈”活动取得额外学分。其他的学生浏览“比尔·奈”节目网站,一些学生利用他们从“比尔·奈”中学到的知识开展科学展览项目。

总的说来,教师们对“比尔·奈”作为教学工具的作用感到非常满意,愿意将其作为课程的一部分继续使用。85%的受试教师计划明年在他们的课堂上继续使用“比尔·奈”节目。15%的教师说不能确定是否还会使用。教师们也渴望获得有关“比尔·奈”系列片的其他现有材料,并索要有关的教师简报、光盘和书籍。当我们问及想向节目制作者传递什么信息时,他们的回答显示了对节目的高度赞扬和热爱。

我尽量多看“比尔·奈”节目。我认为该节目是迄今为止最优秀的小学生电视科学节目。

我喜欢你们正在做的事情。比尔用幽默来抓住学生们的注意力,这种方

式令人振奋。当我在“试一试”栏目中看到一个女学生作为角色模范做科学实验时,我不禁为之感动。出色的工作。继续下去。

我觉得“比尔·奈”是一种极好的讲授科学的方式。请制作出更多的录像带。

家长们对“比尔·奈”节目的看法

- 在家看“比尔·奈”的孩子平均每月看 10 集。
- 61%的家长称,在看过“比尔·奈”节目后,他们的孩子对科学的兴趣提高了。

• 根据家长们的反映,“比尔·奈”节目影响了他们的孩子对科学活动的兴趣和参与(95%的家长报告说,在过去数月的调研过程中,他们的孩子进行了科学探究、科学实验或科学活动)。

• “比尔·奈”节目促进了家庭互动。92%的家长报告说,他们和孩子一起看过至少一次“比尔·奈”节目,92%的家长说,他们的孩子同他们讨论过这一节目。

背景情况

我们的家中观看者包括 351 名儿童,他们分别来自萨克拉曼多、费城和印第安纳地区的市区、乡村和郊区。这些儿童按所在地、年级和性别的分布情况如下:

	市区	乡村	郊区	三年级	四年级	五年级	男孩	女孩
萨克拉曼多 总数 :141	48	43	50	42	47	42	71	60
费城 总数 :113	34	30	49	31	45	37	42	71
印第安纳 总数 :97	44	7	46	11	36	50	51	46

我们请在家看节目的孩子(家中观看者)提交月收看时间表,这样我们便可以掌握他们观看过“比尔·奈”的哪些节目,以及他们多长时间观看一次。我们也请这些孩子的家长在十二月完成一份看前问卷调查表,并在第二年的五月调研结束时完成看后问卷调查表。我们的问卷向家长提出了多种问题,涉及他们的孩子的看电视习惯以及对科学的态度和行为。对问卷调查获得的数据进行分析后,我们获得了以下结果。

在家里看电视的孩子的收看习惯

根据家长们的答复,85%的孩子是通过公共广播公司电视台收看“比尔·奈”节目的。约10%观看者没有明确说明他们是从哪家电视台看的。另外1%的观看者看的是自己购买的“比尔·奈”节目录像带。

在家看“比尔·奈”节目的孩子中,约57%的孩子看过其他科学节目,如“贝克曼世界”(33%)、“发现频道”(26%)、“国家地理”(19%)、“新星”(11%)、“牛顿的苹果”(8%)和“自然频道”(5%)。

我们询问家长,他们的孩子是否经常收看下列的任何节目:“发现频道”、“贝克曼世界”、“牛顿的苹果”、“神奇的校车”和“代笔人”。根据家长们的答复,一半的孩子常看“发现频道”、“魔术校车”和“代笔人”。三分之一的孩子常看“贝克曼世界”,约6%的孩子常看“牛顿的苹果”。

“比尔·奈”节目收看情况

根据我们的调查,在家看“比尔·奈”的孩子平均每月看10集节目。十二月份的平均观看量最低(一个月7集);次年四月份的平均观看量最高(一个月11集)。十二月份观看次数低主要是由于我们的调研是从这一月份开始,而且正好赶上圣诞节假期。

家长们指出,有很多情况妨碍了他们的孩子经常性地收看“比

尔·奈”节目。选择最多的一种情况是节目时间表冲突(70%) ,其次是时间不够(29%)、“比尔·奈”节目时间的变动(13%)、个人原因(10%)、孩子或家里人想看其他节目(10%)、缺乏观看兴趣(6%)、缺少电视机、录像机或有线电视(5%)。

我们要求在家看节目的孩子每周看 1~2 集的“比尔·奈”节目,我们还送给他们 12 集节目的录像带。根据这些孩子提交的观看时间表,12 集“比尔·奈”节目的观看情况如下:

节目名称	%
消化	61
水循环	61
哺乳动物	60
鸟类	59
波	58
地震	55
河溪	51
鱼类	50
旋转物体	50
湿地	49
飞行	49
海洋哺乳动物	48

孩子们对科学的兴趣

根据孩子家长的介绍,孩子们似乎有点偏爱学科学。他们喜欢学校的科学课;当我们让家长们选择自己的孩子喜欢的两个科目时,家长们选择了科学,其次为数学、阅读、写作,最后是社会学科。从学生问卷调查来看,学生们也报告说他们对科学有相当大的兴趣。

尽管孩子们原本就喜欢科学,但是家长报告说,在看过“比尔·奈”节目后,他们的孩子对科学的兴趣更大了。61%的家长报告说,他们的孩子对科学的兴趣有了一定程度的提高;41%的家长称,他们的孩子对科学的兴趣有了相当程度的提高;20%的家长称,他们的孩子对科学的兴趣有了明显的提高。

看过节目之后,我的孩子开始看更多的有关哺乳动物、爬行动物、恐龙和宇宙外空的读物。他还用他的零花钱买这样的书。

我儿子重新看了“比尔·奈”系列片中有关帕努利原理和飞行动物(如猫头鹰)的那些集节目,之后他为学校科学集市制作了一件展品。

我女儿对哺乳动物非常好奇。她将这集节目的录像带拿到学校,与她的教师讨论,并说服她的老师放给全班的同学看。

我们也请家长们对他们的孩子看“比尔·奈”节目的兴趣打分,按1~10分的量表,1分代表“毫无兴趣”,10分代表“非常感兴趣”。51%的家长的打分在8分以上。四分之一的家长打了10分,说他们的孩子对观看“比尔·奈”节目非常感兴趣。只有五分之一的家长对他们的孩子的兴趣打分为1、2、3或4分。

我太激动了,我的儿子可以参与这个节目了!他把观看你们送的录像带作为乐事。他真的喜爱那些录像带。他喜欢看它们,胜过其他的动画片、录像片或者电影。

星期六早晨,艾瑞克会早早起床,就是为了看“比尔·奈”节目。

根据在家看节目的孩子的家长反映,“比尔·奈”节目对孩子从事科学活动的兴趣以及做什么样的科学活动产生了影响。95%的家长报告说,在过去数月的调研过程中,他们的孩子进行了科学探究、科学实验或科学活动,孩子们参与科学活动的兴趣也提高了。

家长反映的情况 孩子们的科学兴趣和进行的科学活动

科学探究		科学活动		科学兴趣	
次数	百分比	类型	百分比	类型	百分比
5 次以上	23 %	实验	55 %	实验	55 %
3 ~ 4 次	21 %	制作模型	25 %	博物馆	41 %
很少	41 %	科学家庭作业	8 %	电视节目	41 %
1 次	10 %	混合物体	6 %	其他	56 %
无 / 没回答	3 %	其他	20 %		

家长对科学和“比尔·奈”节目的兴趣

在我们调研的数月里,92%的家长同他们的孩子一起观看过至少一次“比尔·奈”节目。四分之一的家长报告说,他们陪同孩子看节目达10次以上,四分之一的家长超过5次,20%的家长陪同孩子看过3~4次,而其余四分之一的家长看过1~2次。

大多数家长也对科学感兴趣,或者至少对他们的孩子的科学兴趣表示欣赏。将近三分之二(63%)的家长报告说,他们曾与孩子一起进行过科学相关活动。这些活动十分多样,包括做科学实验(17%),参观科学博物馆(14%),制作模型(11%),完成科学集市项目(10%),去动物园或宠物商店(6%),培植庭园(5%)和参加户外活动(4%)。

我们对考古学特别感兴趣了,正在计划参加一些考古挖掘活动。

我让我女儿参加了印第安纳大学教育学院组织的科学考察活动。我们参观了三个博物馆,那里有很多科学展品。

“比尔·奈”与家庭互动

据家长们反映,孩子们与他们讨论过有关“比尔·奈”节目的各种各样的事情。35%的家长报告说,他们的孩子曾与他们探讨某

些“比尔·奈”节目的主题。22%的家长称,他们的孩子提到过“比尔·奈”节目中的具体内容。例如,他们演唱主题歌时会说“科学法则”或“科学是有趣的”。16%的家长报告说,他们的孩子讲过节目的具体情节。约13%的家长说,他们的孩子提到过他们在节目中看过的实验。其余10%的家长说,他们同孩子一起讨论过科学概念,如植物如何生长以及地球怎样自转等。

她常常问我有关科学的问题,并且试图难倒我。

我的两个孩子(6岁和8岁)和我讨论我们看过的所有“比尔·奈”节目,而且要写出有关的情节——目的是看看我们能记住多少有关节目的内容。

每当科尔比观看“比尔·奈”节目时,他总是把学到的东西讲给家里人听。

他总是唱那首主题歌。他甚至让我边做饭边唱。

我们每天观看两集的“比尔·奈”节目——最好的例子是我5岁的女儿也开始每天观看这个节目。在1996年“暴雪”期间,我们到外面玩,她从一个非常高的雪堤上滑下,还高喊着:“看呀,妈咪!”“它是我的科学滑梯!”就像比尔·奈讲他的发明时说的话一样。

总体上,家长对“比尔·奈”节目非常满意,这些节目激发了他们的孩子对科学的兴趣。家长不仅希望他们的孩子继续观看节目,而且他们自己也想与孩子一起观看。当被问到今年夏天是否准备与他们的孩子一起观看“比尔·奈”节目时,几乎所有的家长(92%)说他们将这样做。

我们一直观看“比尔·奈”节目,我对该节目的效果感到特别高兴。它令人惊奇。我现在在家自己教孩子学习,我们不再缺乏可做的科学活动了。谢谢你。不仅孩子们喜欢和盼望“比尔·奈”节目,妈妈和爸爸也是如此。

感谢你们制作了如此有趣的节目,我确信我的儿子能从中学到很多。你们做了一件了不起的工作。继续做好它。我们能为这个节目继续做下去做些什么吗?

课外托管机构：“比尔·奈”节目在 其他场所的使用

理论依据与问题

孩子是“比尔·奈”节目的主要观看者，要调查该系列片对儿童的影响以及该系列片产生影响的潜力，有必要去孩子呆的地方去调查。尽管孩子们大都是在家里和教室里看节目的，但也有很多孩子每天要在课外托管机构呆上好几个小时。我们决定为市区和郊区的一些托管机构提供“比尔·奈”节目录像带，使观看“比尔·奈”节目成为这些托管机构备选的活动项目，并调查“比尔·奈”节目对那里的孩子产生的影响。我们的调研选在费城进行。

在这些托管机构进行的调查表明，使用教育娱乐电视系列片既有局限性，又存在机遇。首先，在课外场所，放映“比尔·奈”节目有来自有组织的非正式活动的竞争，特别是在天气暖和的日子更是如此。其次，这些托管机构本质上是孩子们的交际场所，很多因素都会影响小群孩子的聚散。有时“比尔·奈”节目成为孩子们聚在一起的理由，有时它又导致原本在一起的孩子们分开。举例来说，在一些托管机构，那些曾经（在家里或学校）经常看“比尔·奈”节目的女生离开了“比尔·奈”的活动，转而去参加其他的社交活动。此外，有人觉得，看“比尔·奈”节目这一活动非常需要有大人的监管和介入，但是这种寄希望于专门人员的想法所产生的问题可能比解决的问题还要多，尽管多数的托管机构认为组织和监管是一件容易的事。

我们调研的课外托管机构中，有几家机构支持孩子们参加学习及集体活动，故在这些托管机构，“比尔·奈”节目较容易畅通。托管机构的监护人员通常为孩子们安排进行学习活动的时间和地方，“比尔·奈”节目成为孩子们在家庭作业或课外辅导后的补充活

动。此外,该系列片不仅吸引了年纪较小的孩子,而且也吸引了年纪大的孩子,他们常常加入到正在看节目的孩子群中。我们可能认为,这是因为孩子们喜欢扎堆,而不是科学节目的吸引力,但是一些年纪较小的孩子确实成为了“比尔·奈”节目的忠实观看者。

我们对课外托管机构的孩子进行了与学校观看者组、家中观看者组和对照观看者组一样的测评。但是,托管机构的孩子数量有限,而且,这些孩子有时是观看者,有时不是观看者,其角色处在不断的动态变化中,因此妨碍了对有关问题的详细分析。

“比尔·奈”节目的外延活动和材料有助于扩大孩子们观看“比尔·奈”节目的机会,有助于鼓励和形成新的家中观看者、学校观看者和课外观看者。最好的外延材料可能不是教师指南,而是那些可在室内和室外进行的简单的实验和观察活动。对课外托管机构的相关情况进行认真分析,会确定出“比尔·奈”节目的外延材料和活动中,哪些是最好的。

费城地区调研概述

费城是一个典型的工业城市。城市有一中央核心区,毗邻港口并有一条铁路干线。有闹市区,然后是市内的联排式住宅,然后是老化的、已被遗弃的仓库和上个世纪90年代铁路全盛期时的工厂。客运铁路线从城市中心向四围辐射,犹如车轮的辐条。有各种各样的环城路,宾夕法尼亚州收费公路穿过北部环绕市郊的蓝色道路,一直通向西部地区。在环线的外部,是富裕的郊区,通往西方的干线,四处伸展的林荫道以及西北和北部的新郊区。许多国家药物企业巨头位于这个环线附近。在弧形以外,间杂着工业区、农场、小城镇、零星的城市和快餐店。

城市和邻近地区的混杂主要是由于旧式的人口迁移模式所造成的。威尔士人、德国人、意大利人、犹太人、美裔非洲人和祖先是英国或北欧新教徒的美国人。最近的文章称,该地区存在着严重

的黑人和白人隔离现象。

我们在费城地区选择了三家市内托管机构和两家郊区托管机构作为我们调研的对象,主要侧重于女孩和少数民族儿童托管机构。我们为这些机构提供了电视和录像机及一批“比尔·奈”节目录像带。我们请监护人员定期放映录像带,使看节目成为这些托管机构的备选活动。我们给监护人员一定的报酬,让他们每天观察和记录学生看节目的情况。市内的三家托管机构包括“加百利课外活动中心”、“日尔曼敦男女生俱乐部”和宾夕法尼亚大学“父母幼儿中心课外活动中心”。郊区的托管机构包括“梅因莱恩基督教青年会课外活动部”和“加勒特福特小学课外活动中心”。

我们在费城地区的总协调员负责处理看前看后问卷调查相关事宜,并与各个课外活动协调员进行了访谈。

以下是各个课外托管机构及其“比尔·奈”节目收看情况简介。我们选择的每一个托管机构都可以视为“比尔·奈”系列片用于不同的儿童群体和不同场所的小案例。从这些初步的工作中得出的概括性结论对以后的“比尔·奈”外延工作是有助益的,也值得该系列片扩大观看者时借鉴。

加百利课外活动中心

基本情况

加百利课外活动中心位于一圣公会教堂内,主要面向费城东北部的一家社区,该社区的居民主要是务工的贫民。虽然那里的孩子们说英文,但他们的家庭使用的语言总共有 37 种之多。

孩子们的运动场被铁丝网围绕,包括一个四轮汽车停车场和一个靠近教堂的草坪。汽车从教堂旁边的公路急行而过,这条宽阔的林荫大道是进入费城东北部的要道。不同状态的褐砂岩分布在林荫大道和小巷。

该中心现有 33 名一至八年級的托管儿童,还有 52 名儿童等

待加入。该中心管理员林达·阿洛西说：“我们每周只收 2 美元，这可能是我们的与众不同之处。其他的托管中心每名儿童收费约为 45~50 美元。一旦孩子们进到我们的中心，他们就会留在这儿。”

维持该中心运行的成本大约为每个儿童每周 80 美元。据林达介绍，该活动站已变成一家实体单位，虽然中心所在地仍位于教堂内，但它实质上是一个独立单位。

有 5 名拿薪水的工作人员和大约 20 名志愿者（年龄 15~18 岁），共有 33 个孩子，分布在 1~8 年级。每个志愿者负责照应 5~6 个儿童，这个数字远高于州政府所要求的比例（1:10~1:15）。

与本次调研的其他课外托管机构相比，加百利课外中心较重视孩子们的文化学习。林达解释说：“我们中心主要是一个艺术活动站，后扩展到文化学习辅导。”“我们有一个大班和一个小班，每班每天都有一个小时的 cultural 学习辅导。我们的孩子中，很少有人达到同级学生的水平，他们中很多人被确定患有注意力不集中多动症。”

“比尔·奈”节目观看情况

加百利的工作人员专门为“比尔·奈”节目提供了一个场所，使观看“比尔·奈”节目成为自容式的活动，且不与其他活动项目冲突。在加百利，每天下午被分成两部分：1 小时文化学习课，用于做家庭作业和观看“比尔·奈”节目；还有 1 小时的艺术活动课。大班和小班过 1 小时后会交换活动场地。

在做家庭作业的那 1 小时里，工作人员每星期为孩子放映两次“比尔·奈”节目。在看完一集节目后，孩子们将进行约 15 分钟的讨论，然后做家庭作业。在我们调研的三个多月的时间里，加百利的孩子共观看了 24 集节目，平均每月看 8 集。每周，林达都要询问孩子们，他们最喜爱的节目是什么，以及他们从一周来观看的节目中记住了什么。孩子们特别喜欢“消化”一集的节目，后来又

重新看过了一次。根据林达的介绍,孩子们记住了“比尔·奈”节目中的许多东西。她反映说:

不知道是怎么回事,孩子们说他们讨厌科学的,但奇怪的是事实并非如此。科学概念成为了他们交谈的内容。例如,我们有一些滑板(带有轮子的塑料板),孩子们会说,“由于重力的作用,我们下坡滑更容易”。还听孩子们说过,“我们吃的食物需要经过数小时后才能进入小肠”。

加百利的工作人员对“比尔·奈”节目和它在科学学习方面发挥的作用感到非常满意。林达称,“在学校,科学对于孩子非常难于理解,因为他们必须读懂它,现在他们可以通过电视学习科学了”。该中心决定今后继续使用这一节目。林达说:“我认为‘比尔·奈’节目是最好的教材之一。明年我还会使用这个节目。”

结论

加百利是一家虽普通却组织良好的课外托管中心的典型,其工作人员素质较高。该中心专门为“比尔·奈”节目提供了场地,使观看“比尔·奈”节目成为自容式的活动,且不与其他活动项目冲突。加百利服务于低收入少数民族社区的儿童,这些孩子在学校被认为是在科学方面不会取得成功的孩子。“比尔·奈”节目为这些在学校得不到扎实的科学课程学习的孩子提供了学习科学的机会,能弥补这些孩子在科学方面的落差。

日尔曼敦男女生俱乐部

基本情况

日尔曼敦是费城西北部的一个低收入区,那里50%的民众生活处于贫穷之中。它的看护率是该城市最高的。

日尔曼敦男女生俱乐部位于一个空置的停车场和一所教友派学校之间,是一座四层楼的文化娱乐中心,从下午3:30至晚上9:00开放。该中心开展了三个活动站,可为170名儿童提供服务:

娱乐活动站从每年的十月份至第二年的六月份开放,会员费为 10 美元;儿童下午托管站面向年龄为 6~12 岁的儿童,费用较高,每周收费 35~55 美元(可给予适当补贴)。还有一个是面向 7~9 年级学生的青少年活动站。

该中心同时拥有娱乐活动站和儿童托管站。娱乐活动站为孩子提供了多项活动,孩子们来去自由,这里给人留下的一个总体印象是监管不够。儿童托管站提供的是有执照的日托服务,看起来管理得更细致些。

在该中心,不同的楼层提供不同的活动。娱乐室提供有台球和乒乓球活动。还有多个家庭作业室、一个计算机房、一个电视房、一个健身房和数个工艺美术室。下午 3:00~4:00 是家庭作业时间,随后是活动时间,包括游戏、计算机实验室、工艺美术和体育运动。

“比尔·奈”节目观看情况

该中心安排在家庭作业时间播放“比尔·奈”节目。在三个多月的时间里,孩子们共观看了 20 集节目,平均每月看 7 集。播放活动常常与其他重要活动发生冲突,此外还有人手问题。二月份的“黑人历史月”占用了“比尔·奈”节目的播放时间,接下来的假期和数日的雪天也影响了节目的收看。孩子多以及数周的播放时间变动使得该中心没有一个固定的、可预测的“比尔·奈”节目时间表。

虽然有这些干扰,但“比尔·奈”节目的核心观看者不断发展,约一个半月之后,该节目已成为孩子们每周日常活动的一部分。约 10 名儿童每次都来观看这一节目。

然而,在观看了约一个月后,女孩们开始退出。她们似乎仍然喜欢这一节目,但同伴的压力使她们离开。此外,其他的活动也常常吸引她们,使她们不再看节目了。不过,有些男孩成为了热心观

看者。在我们的调研结束时,我们注意到,有两三个男孩带他们的大朋友或亲戚一起来观看节目。

虽然女孩开始退出,但有些大孩子加入了进来。一些六、七年级的学生对“比尔·奈”发生了兴趣,开始观看节目。

托管员克利夫顿说,她一定还会再用“比尔·奈”节目,因为它有积极的教育作用。谈到明年的打算,雷切尔说她计划把“比尔·奈”观看时间安排在冬季,从而不与棒球比赛撞车。雷切尔还说,她将在年初提醒孩子们观看“比尔·奈”节目,而不是到时候再把这个节目编入孩子们已经满满当当的活动时间表。俱乐部也考虑采取一些激励措施,以鼓励儿童观看。

结论

日尔曼敦男女生俱乐部无疑是最缺乏管制的课外活动中心,由于没有激励措施,加上其他的活动吸引孩子,使得维持节目观看者(特别是女孩观看者)存在困难。然而,该俱乐部的管理员表示,只要该节目及早纳入俱乐部的活动安排之中,他将来会使用该节目。

父母幼儿中心

基本情况

父母幼儿中心(PIC)是一家日托中心和课外托管中心联合体,主要服务于宾夕法尼亚大学教职员工。父母幼儿中心联合体位于一个很大的、很少使用的石头教堂的外围建筑中,课外托管中心与日托中心同在一处又彼此分隔。课外托管中心三分之二的儿童是宾夕法尼亚大学教职员工的子女,其余三分之一则来自附近务工的贫民家庭,他们从附近的学校搭乘公交车过来。

该中心有45个孩子,从幼儿园儿童一直到四年级学生。有些孩子每天都来,其他孩子偶尔来一次。该中心拥有运动设施和四

周围网的操场,供孩子们在天气晴好时在户外活动。遇上冷天或雨天——或进行室内活动的时候——所有的孩子呆在楼上的几间小教室内。

课外中心主任 Cogdell 说:“我们中心最重要的经营理念是让孩子有许多可选择的活动。一天中,我们都告诉孩子们做哪些事,让他们自己选择。我们设法提供很多可做的事情。”

该中心有一名艺术教师,给人的感觉是每天都组织艺术活动。课外活动的管理人员包括 Brad 和一名助手。

“比尔·奈”节目观看情况

在 PIC,“比尔·奈”节目与其他活动存在尖锐的时间竞争。Cogdell 说:“我们努力为孩子提供很多可做的事情,让他们自由选择。这就是为什么我们这里吸引孩子看节目较困难的原因。”

起初,该中心计划在每周的活动时间段中,安插三天观看“比尔·奈”节目:星期一、星期二和星期五。然而,课外中心主任 Cogdell 称,“大约一个月后,蜜月期结束了。‘比尔·奈’节目要与太多的其他活动竞争,而且孩子们的兴趣也不稳定。因此我决定改在孩子们吃加餐的时候放映,并做到让每个孩子都知道这件事”。

通过与孩子和与 Cogdell 的访谈,我们对加餐时间孩子们看节目的情景有了了解。在吃加餐的时候播放节目,孩子们边看边聊天,因此孩子们的注意力肯定不集中。吃完加餐,中心不会要求孩子呆在那里看完节目,尽管他们鼓励孩子们留下来再看上 10 来分钟。

甚至刚吃完加餐,就有很多的活动吸引孩子离开。女孩可能会到家庭作业室去学习或与其他孩子一起做什么。其他的孩子可能去外边玩。Cogdell 曾试验了不同的播放节目时间,结果发现还是加餐时间效果最好。在看节目的同时,孩子们会有交际活动。只有少数儿童会留下来看完节目,对此 Cogdell 解释说:“要记得,

我们给了孩子们许多选择。”

在父母幼儿中心的儿童总共观看了 27 集节目,平均每月看 9 集。我们的节目日程表显示,在三月份时观看者人数明显下降,而在四月份有所回升。在与孩子们的看后访谈中,我们的费城地区协调员听到孩子们说:“比尔·奈是傻瓜”,这是三、四年级的学生常说的。Cogdell 解释说:“孩子们会告诉朋友比尔·奈很傻。结果,孩子们不是蜂拥着去看节目,而是风行说它傻,就像电视‘金刚战士’和‘小恐龙巴尼’的情形一样。”

尽管有这样的压力,节目还是吸引了部分孩子。看节目的男孩明显多于女孩。Cogdell 说:“四年级的女孩觉得她们不喜欢这个节目。她们转而到其他房间去聊天或玩耍学习。”

Cogdell 认为,他会再次使用“比尔·奈”节目,但不是定期使用。他问道:“再次使用‘比尔·奈’?我愿意更多地在主题周期间使用它。”

结论

PIC 的情况使我们看到了教育计划和教育资源的一种不协调。在一个特别注重孩子的自由选择、而且有大量活动可供选择的托管中心,安排一个定期观看的节目会惹来很多麻烦。在像 PIC 这样的托管中心,如果只给它一两盘节目录像带,供其需要时使用,这样效果会更好。

梅因莱恩基督教青年会课外活动部

基本情况

阿德莫尔是在梅因莱恩外的一个中上产阶级居住的老式郊区(梅因莱恩是美国宾州东南部的一片郊区,它是以通往首府费城的主要铁路线命名的)。基督教青年会位于郊区广场路附近,周围是一个老商业区,有劳拉-阿什利等高档商店。基督教青年会托

儿部有五六十名孩子(日托,每天上午7点直到下午6点)。课外活动部有25~30名一至五年级的孩子,这些孩子中很少是少数民族儿童。

课外活动部在楼下的两三间教室里,以便于孩子到体育馆和户外运动场活动。该活动部对孩子的管理松散,安排有许多户内外活动。有些活动是有组织的,另一些则不是。课外活动协调员特伦斯说:“孩子们做的大都是自由选择的活动。他们在学校呆了一整天了,他们不需要再被强迫做什么了。”

“比尔·奈”节目观看情况

看“比尔·奈”节目的孩子每天都不一样,这也使人感到该课外活动部管理松散。一些孩子是常客,每周5天都来看,很多孩子按家长的选择每周来3天以上,其他时候要参加学校体育项目、上犹太学校等活动。

孩子们下午3:30开始过来,活动差不多从下午4:00开始。季节对孩子们的活动有一定的影响。特伦斯解释说:“在冬季月份,我们的活动更为组织化。春天,我们则去户外活动。”在明媚温暖的天气,孩子们下午4:15(或者更早)外出活动,一直到下午5:00家长接走。在冬季,孩子们下午4:00吃加餐,4:15~5:00选择上手工工艺课、到体育馆活动或去学电脑。“比尔·奈”节目也是孩子们可选择的活动之一。

在梅因莱恩基督教青年会或其姊妹会所,员工通常身兼多职。托儿部管理员迈克也是水上运动部的主任。特伦斯管理课外活动,白天还要在托儿部工作,并且管理着青年会的多个度假营和夏令营。工作人员的缺乏无疑对“比尔·奈”节目的收看产生影响,当工作人员同时兼顾几件事时,你无法期望他什么事都干得好。

儿童在单元室观看“比尔·奈”节目,该房间还放有课外活动部的计算机。特伦斯解释道:“我们不是强迫而是鼓励孩子们来这儿

看节目的。”其他孩子可以用计算机,但要音量关闭。当孩子们看节目时会时不时地有些议论。节目放完后,孩子们会转向其他活动或去玩耍。

在第一个月,该活动部每周放映2~3次“比尔·奈”节目。在三月份工作人员出现变动,然而节目放映仍稳定在一周两次。在四月份,度假营的开办让工作人员手忙脚乱,他暂停了“比尔·奈”节目的放映。度假营一结束,节目的放映又恢复正常。

特伦斯说:“一旦孩子们看节目,他们的反应大都很积极。但是刚开始看时,常常会有些阻抗。”他解释说,个别的孩子安定不下来,工作人员给他们一些小奖品或棒棒糖之类的东西,促使他们呆下来看。他说:“但孩子们随后就不再理会我们给他们的东西,呆在那看节目了。到最后,我只是在他们愿意看时才为他们放映节目。”

由于孩子们是在靠近课外活动站的房间里看节目的,那些低年级孩子也想观看,特别是一些二年级学生。在我们整个的调研过程中,几个二年级学生一直是“比尔·奈”节目的观看者。特伦斯说:“如果我再次使用‘比尔·奈’节目,我兴许会让二年级学生也加入进来。”总的说来,孩子们似乎十分喜欢这一节目。他们常常吟唱节目的主题歌。“恐龙”和“化学反应”是最受喜爱的两集节目,而“营养”是最不受欢迎的节目之一。

由于孩子们开始谈论“比尔·奈”节目,观看“比尔·奈”的消息传入家庭。那些曾填写过许可单因而知道该节目的家长,对他们的孩子正在做的事感到好奇。在我们进行看后访谈的那一天,较早来接孩子的家长表现出了积极的好奇心。

特伦斯说,他还会使用“比尔·奈”节目,“与那些电视枪战片比起来,‘比尔·奈’节目是一种积极的体验”。不过,他想采取更多的办法,通过为看节目的孩子发放一些奖品如T恤衫和奖旗的方式,争取更多的“比尔·奈”观看者。他解释说,“我想做更多的事,

使我们这儿成为孩子们的俱乐部”。

结论

基督教青年会课外活动部的经验显示出“比尔·奈”节目在一个管理松散的环境中具有的潜力。利用正面强化和奖品的俱乐部方式对保证节目的收看发挥了一定作用,尽管这样做需要额外的开销。该节目似乎适合在冬季放映,而在天气回暖的春季,放映任何电视节目看来都不合适。

加勒特·福特小学课外活动中心

基本情况

加勒特·福特小学坐落于德雷克塞尔希尔,有两层小楼,是一所不错的老式学校。它的课外活动计划是上达比学区几个日课延伸计划之一。

学校两旁是车身修理厂和便利商店。学校的前门外是绿色草坪和石垣,后面是柏油操场和球场。课外活动中心设在学校地下室,包括一个自助餐厅和几个小教室。两个侧面房间作为家庭作业室和“比尔·奈”电视室,孩子们主要是在自助餐厅和外面玩耍。

该课外活动中心共有 57 名幼儿园至五年级的孩子,但事实上核心人数在 40~45 人。家长每天支付的托管费为 5.50 美元。大约 80% 的孩子一周来 5 天,另外 20% 的孩子偶尔来。3 名管理员轮流值班,两名辅工则全天上班。

该中心的主导思想是“有辅导的玩耍和日间托管”。辅工南希解释说,“我们这里实行的是松散管理,因为孩子们在学校已被管了一整天了”。

孩子们下午 3:30 来到中心,然后自己选择是去家庭作业室还是到外面玩。下午 4:00,所有人都回来吃加餐。之后孩子们可以有多种选择,玩游戏,去体育馆锻炼,看电影或做手工。一些家长

早在下午 4 :30 就来接孩子 ,但大多数的家长在下午 5 :00 至 5 :30 之间赶来接孩子。

“比尔·奈”节目观看情况

学生们在下午 4 :00 的活动时间观看“比尔·奈”节目。可供学生选择的活动中是观看“比尔·奈”节目或与朋友一起玩耍。

愿意看“比尔·奈”节目的孩子的名单经父母签名后收集上来。一些没有签名的学生出于好奇在节目放映期间会过来看看。节目放完 ,所有的孩子都去外面玩。在我们调研的三个多月的时间里 ,孩子们共看了 18 集节目 ,平均每月看 6 集。

如同在其他托管中心观察到的情况一样 ,签名同意观看节目的孩子被鼓励看下去。南希谈到 :“起初 ,我们说‘你们签名了 ,最好坚持看下去’。然后我们就任其发展。结果很快淘汰了一些孩子。那些坚持看下来的孩子大部分是三年级的男生。”

随着我们的调研的推进 ,女孩们开始退出 ,不再观看“比尔·奈”节目 ,转而去与她们的朋友聊天或玩耍。看节目的差不多是一帮三年级男生。“这些男生边看节目边讨论。他们唱节目的主题歌。”南希说 :“他们喜欢这一节目。他们最喜欢的两集是‘恐龙’和‘爬行动物’。”

此外 ,还有其他一些观看效果。比如 ,低年级的孩子渐渐对“比尔·奈”节目有了兴趣 ,一些二年级学生经常过来看节目。另一个附带效果是孩子们在家里也看节目了。一些学生成了“比尔·奈”的家中观看者 ,而其他一些学生成了该节目的新观看者。

南希认为他们一定还会使用“比尔·奈”节目 ,但只能安排在冬季的两三个月的时间。她觉得短期使用可以使节目有新鲜感 ,能发挥最大效用。长期使用时 ,如与某些其他活动(像天气晴好时的户外活动)竞争 ,会不利于节目的放映。

结论

像加勒特·福特小学办的这种日课延伸活动中心在这个地区很普遍。尽管这类活动中心具有娱乐功能,但它们还是在冬季为“比尔·奈”节目提供了观看场地。在加勒特·福特,社交群体活动看起来对女生的吸引力更大,使她们离开了观看现场。看来对于女孩,在课外场所观看“比尔·奈”节目是无法与社交群体活动竞争的,诀窍在于抓住时间空档。对女生特别是四、五年级的女生来说,与同伴交往是必要且高兴的事,对此必须有所考虑。

对教育电视频道的调研

我们调查了美国和英联邦国家公共广播公司的 200 多家电视台,以确定这些电视台是否使用了“比尔·奈”节目以及是如何使用的。我们将调查表寄给了各家电视台的教育电视部主任,回应率为 72%。调查表通常由教育电视部主任亲自填写,有些则是由电视台负责外延工作的人员完成的。

几乎所有的(93%)电视台都播放“比尔·奈”节目,从星期一至星期五定时播放。最多的两个播放时间档是在下午 5:30(31%)和下午 4:30(31%)。一些电视台还在全年不同的时间滚动播出该节目。

半数以上(58%)的电视台将“比尔·奈”作为他们的外延活动的一部分,而且几乎所有的这些外延活动(92%)是面向学校和教师的。教师指南是最大众化的外延宣传材料,如有人索取,电视台可寄发,或作为教师培训材料的一部分分发。教育电视部主任还报告了该节目的其他应用。例如,一家电视台将“比尔·奈”节目与托儿外延计划结合使用,另一家电视台则将它与家长教育研习班结合起来使用。

许多电视台说,它们通过 NTTI(国家教师培训研究所)使用

“比尔·奈”节目。例如,在奥斯汀的 KLRU 电视台,一些特级教师在教师培训研究所的教师演示课上使用该系列片,他们制定的课程计划中将“比尔·奈”节目的部分内容纳入到他们的课程中。

有几家电视台称,它们用于开展外延活动的经费已没有了。一些电视台特别提到,它们希望能有经费来实施“比尔·奈”教师培训。

在 1995—1996 年间,只有 14% 的电视台开展了系列片(包括“比尔·奈”)的使用和影响的调查。我们共收到了 5 家电视台的调查报告的副本,从中我们发现,在其中的 4 家电视台,“比尔·奈”系列片一直位居播出最多的三大教育片之列(另外两大教育片是“阅读彩虹”和“神奇的校车”),在第五家电视台,“比尔·奈”系列片位居第四位。

许多电视台反映,“比尔·奈”作为教学节目受到学校的青睐:

- 根据得克萨斯州圣安东尼奥 KLRN 电视台进行的一项调查,在接受调查的 65 所学校中,有 26 所学校报告说,它们在课堂上使用“比尔·奈”节目。

- 根据得克萨斯州埃尔帕索城 KCOS 电视台进行的一项调查,48% 的学校在课堂上使用“比尔·奈”节目。

- 俄亥俄州托莱多 WGTE 电视台报告说,课堂上使用的电视节目中,“比尔·奈”排在第三位。

很多电视台的负责人想利用更多的材料和更多的机会,把“比尔·奈”作为其外延工作的一部分推介出去。有关“比尔·奈”系列片的反馈大部分是非常肯定的。一些电视台的主任谈到:

观看者的反应非常积极,它真的激发了观看者对科学的兴趣。我们接到了许多观看者的电话要求回放。

在我们举办的研讨会上,一些教师过来告诉我们,他们非常喜欢使用“比尔·奈”节目。家长则告诉我们说,每天节目一放完,孩子们就想尝试着做节目中表演的实验。

“比尔·奈”节目大受欢迎。该节目太成功了。

然而,也有一些异议:

我们喜欢“比尔·奈”节目,并在我们开展的辅助学校教学活动及有关的研讨会中使用它,但是我们有一个很大的不满。我们不愿意该系列片同时在商业电视台上播放……覆盖全国的公共广播公司电视台一直在大量讨论保持公共广播公司身份独特的重要性……不愿意与商业电视台搅在一起。如果“比尔·奈”节目继续保持与商业电视台的关系,我们可能会把它从我们的节目表中剔出……我们希望它是一部公共广播公司独家播放的系列片。

我们很想播放“比尔·奈”节目,它很受欢迎。但公共广播公司将这一节目分入 NDS 类别,且对该节目的获取采取了限制措施,使得我们不能播放它。

我们使用的是 TTI 模式,我们支持合法的使用节目。由于权利有限,我们没能使用“比尔·奈”节目。教师们需要能够在媒体中心录制和转录节目。三年不播放权会有帮助,但不能解决合法使用的担忧。

结论与建议

我们在三种“自然的”观看场所(指学校、家庭、课外托管机构)进行了“比尔·奈”电视系列片观看情况的调研,我们对观看者的行为只采取了最小程度的控制和出于个人好意的限制。如同所有的实地调研一样,不管事先多么周密地设计了控制和条件,在执行过程中总还会有这样或那样的缺陷存在。本次调研亦是如此。如何让一些班级和孩子不看节目,而让应该看节目的孩子看足够的节目,这方面存在些问题。尽管如此,我们运用的设计为我们提供了足够多的样本,从而确保了我們得出合情合理的结论。

在一些地区的调查中,有很多重要和有利的结果显示,该系列片对儿童了解科学产生了影响。而在其他地区的调查中,我们却发现该系列片的影响很小,可能是因为我们问的问题有毛病,也可能是因为观看和随访的数量不足,无法产生有意义的变化,还可能

是由于构思或表现技巧的问题,使得该节目对这些观看者没有发生效果。调研的结果有些是十分合理的,有些则是反直觉的。

我们的调研结果中很多应被视作“初步性的”,因为不是所有的观看条件都会产生结果。也许是节目之外的一些情景事件对孩子们产生了重要的影响。有时候,学生在知识和理解方面的变化在统计上有显著意义,而按绝对值衡量差别却并不明显。有时候,我们看到了较大的变化,但由于样品数量不足,这些变化不具有统计学上的差别。但是有些调研结果来自于我们对类似问题的多次调查,并且是按照不同的评估任务以不同的方式得出的同样的结论,我们对所提供的这些方面的结果较有把握。

结论

根据我们从全国三个调研地点采集的数据,以及我们对孩子、班级和课外托管机构有关情况的观察,我们总结出以下结论。

- “比尔·奈”系列片在多种场合(如在家中、在学校以及在课外场所)影响了儿童对科学的了解。该节目是一个有效的教育工具,它提高了儿童解释科学信息的能力,加强了他们对科学已有的积极态度。

- “比尔·奈”节目最大的效果在于,它提高了儿童运用科学知识进行批判性思维和解决问题的技能。与看前相比,孩子们在看后能够对有关科学概念提出更全面更复杂的解释,他们还提高了识别和解释科学现象的能力。

- “比尔·奈”对女孩和少数民族学生产生了影响,但这种影响并没有使她(他)们迎头赶上。按科学思维技能的提高来衡量,女孩的进步常常与男孩一样大,甚至更大些,少数民族学生通常也取得了与多数民族学生一样的或更大的进步。然而,由于女孩和少数民族学生开始时的得分都低于男孩和多数民族学生,因此,尽管他们的得分差距缩小了,但她(他)们还没有完全地追上。

• 在我们的家庭调研中,看节目的女孩人数与男孩差不多,在学校中,两者也很相似;但在课外托管场所,当有其他社交性更强的活动可以选择时(如与别人一起做家庭作业或其他集体活动),看节目的女孩会变少。

态度

由于学生们起初就对科学抱有非常积极的态度,因此,不会因看节目再出现什么改善。看节目的孩子原有的对科学家的刻板认识似乎在消除,我们看到女孩更有信心认为她们也可以成为科学家。不过,大多数孩子不会轻易地将妇女视为科学家。

教师

虽然有许多教师把该节目用作教学工具,但大多数教师报告说,“比尔·奈”只是在有些时候配合了他们的课程。某些教师要求将该系列片的各集节目分成不同的主题单元,这样便于协调播放或者组合销售。教师们认为,目前的播放时间安排妨碍了节目的有效使用,因为他们常常直到课程单元讲完还没看到该系列片的相关节目。

教师播放节目时,常常没有一个明确的教学计划,也没有安排相关的课程,“比尔·奈”节目成为其科学课程的一部分常常出于偶然而非有意为之。许多教师报告说,他们没有组织与节目有关的活动或实验。然而,学生们对科学的了解的提高往往直接得益于他们在“比尔·奈”节目中看到的科学原理的演示。

家长

“比尔·奈”节目促进了家庭成员之间的互动。家长和孩子一起观看“比尔·奈”节目,并进行讨论,这样有助于儿童从节目中学习科学知识,而且有助于他们将来参与更多的科学活动。几乎所

有的家长打算继续与他们的孩子一起观看节目。

大多数的家长与他们的孩子一起进行科学活动,并报告说,他们的孩子在看过节目后,对科学更有兴趣了。几乎所有的家长都说,在我们过去数月的调研时间里,他们的孩子进行了科学活动。

课外场所

课外场所可以成为“比尔·奈”节目很好的观看和外延活动场地。总的说来,在课外中心,“比尔·奈”节目吸引了一些核心观看者,有时较小的孩子也加入了观看队伍。在冬季时看节目的孩子较多,因为那时看节目不会与其他备选活动(如体育活动)发生直接竞争。虽然多数情况下社交活动更能吸引孩子,但在某些课外中心,观看节目成为一种能使孩子们成群结伙的社交活动。在课外托管机构,孩子们很少有机会进行基于“比尔·奈”节目的实验或探究活动。

建议

本评估的目的是研究和评价“比尔·奈”系列片对儿童科学知识、科学态度和科学行为的影响。我们在对收集的信息进行分析解释的过程中,有必要就如何使该系列片对儿童、学校和家庭发挥更大作用提出我们的建议(或假设)。以下是我们向该系列片及外延计划的工作人员提出的一些建议想法,供他们将来工作时参考。

制作问题

根据我们的调研结果,有几个内容和制作问题值得“比尔·奈”制作组考虑:

- 节目中的科学演示是向观看者解释科学概念的最有效的方式。然而,节目中的演示并不总是适合儿童的年龄。制作者应考虑使演示更简单些,并对这些演示多做些解释。

- 孩子们常常难以将节目中讲的科学原理转移到其他情景

中。因此,演示应包含多角度的叙述,用以说明所演示的科学现象的各种应用。

- 正如节目制作人员已经知道的那样,不断反复就会奏效。对于这一年龄段的孩子来说,在不同情景中反复讲述相同的概念——并把它们相互联系起来——有助于加强孩子对概念的理解。这些问题对孩子理解节目的核心概念十分关键。

- 多集节目的内容交叉重复同样有助于孩子们掌握节目的核心概念。我们收集的数据显示,在多集节目中反复介绍的概念更易于儿童理解。

- 制作人员应考虑把同一科学概念放在实验室以及非实验室环境下来表现,以减少孩子们以为科学家总是呆在实验室的刻板印象。尽管“比尔·奈”节目中表现了很多的科学家和科学活动,但由于该节目把比尔弄成穿着实验室大褂呆在地下实验室的样子,结果强化了孩子原有的对科学家的刻板印象。

- 应特别重视在节目中选择一些女性角色榜样。该系列片中的“酷科学家”和“动物小姐”两集节目是采用正面女性角色榜样的极好例子。然而,节目中对女性的描绘也较刻板,如在“哺乳动物”这集节目中的“山谷少女”歌手以及50年代妈咪的重复角色。可以考虑给女生观看者(和少数民族学生观看者)更多的来宾镜头,以鼓励更多的女孩(和少数民族儿童)观看此系列片。

外延活动

对该节目的外延计划,我们有如下建议:

- 为了继续激发和培养女孩的兴趣,除要改变节目的制作使之更多地反映与女孩/妇女的良性互动外,应考虑采取专门针对女孩的外延活动,以鼓励和保持女孩观看者,比如创办一份面向女孩家中观看者的简报。

- 加强对教师和学校的外延工作对节目会有好处。使节目

与学校课程更为兼容,将节目进行包装,使之用于学校教学时更符合课程的需要,这些会有助于教师将该系列片用作他们的课程的一部分,而不是作为补充性的即兴课程。如果节目时间安排(乃至教师指南)能把内容相近的节目放在一起,教师可能更愿意使用“比尔·奈”系列片。

- 随着越来越多的教育工作者转向依靠因特网来获取信息,以及越来越多的网站连接到“比尔·奈实验室在线”,该节目网站上的课程材料和课程计划会成为吸引教师加盟的有益补充。该节目网站还可以为教师提供添加在线资源(诸如“比尔·奈”实验课程和备选材料)的机会。

- 其他的面向家庭的外延工作最终会增进孩子们的科学学习,使通过“比尔·奈”节目建立起来的家庭联系继续保持下去。可能的外延材料包括针对具体的“比尔·奈”节目的问答纸以及推荐阅读的科学书目,特别是女性科学家和少数民族科学家的传记。

- 课外托管机构可以成为孩子们观看“比尔·奈”系列片的重要辅助场所,能够促进更多的家庭观看该节目。那些重视文化学习、组织性强的课外托管机构最适宜开展外延活动。在天气不好的月份里,孩子们在课外场所观看节目的可能性更大。外延工作还应该包括给家长的信息,以便孩子可以将某些材料带回家,这样有助于促进家庭成员观看节目。

附件四 美国创新技术博物馆展厅 评估报告(2000)

评估机构 :兰迪科恩合伙公司

引言

在 1999 年 5 月 ,兰迪科恩合伙公司(RK & A)开始对加利福尼亚州圣约瑟市的创新技术博物馆的四个常设展厅和观众参观该博物馆的总体体验进行一项全面的总结性评估。评估的第一阶段共产生了四份报告——每个展厅一份评估报告。本报告是这 4 个展厅评估的总结概要。

这一阶段的评估数据是在 1999 年 7 月至 8 月间收集的。评估力求证明 4 个展厅产生的影响和效果 ,以及确定现有展览展项中那些要素需要调整以便提高效果。评估的具体目标是测定 :

- 观众在各个展厅驻留多长时间 ;
- 观众在个别展项上驻留多长时间 ;
- 哪些展项吸引观众逗留 ;
- 观众注意到的展馆环境特色 ;
- 观众等候使用交互性展项的频率 ;
- 观众遭遇损坏的展品的频率 ;
- 观众对各展厅的意见 ;
- 观众在各展厅的认知体验。

评估方法

为评估观众参观各个展厅的情况以及他们在各个展厅中的体验 ,我们采用了三种数据收集方法 :计时与跟踪观察 ,出口处的开

放式访谈以及焦点观察与访谈。

计时与跟踪观察

在进行总结性评估中,经常要对观众进行观察,因为观察能客观、定量地记录观众参观展览的行为和反应。观察获得的数据能使我们了解观众的行为范围,表明哪些展项吸引观众而且抓住了观众的注意力。

在各个展厅,所有九岁以上的观众都可以作为观察的对象。当他们参观展厅的展项时,调查者在一旁观察他们的行为。被观察的观众的选择采取连续随机抽样法。根据这个方法,一个经过培训的观察员站在各个展厅的入口处,跟随第一名观众进入展区,并在他(她)参观整个展览期间,记录他(她)在哪些展项处停留、在个别展项上驻留的时间以及参观整个展览所用的时间总量。在参观结束后,观察员回到入口处,等候下一名观众进入展区。

出口处开放式访谈

进行开放式访谈的目的是鼓励和激发受访者描述他们的体验,表达他们的意见和感受,与调查人员分享他们根据此次体验而构建的意义。开放式访谈得到的信息会很丰富,因为受访者能从非常个人的角度来谈论他们的体验。访谈所获得的信息意在使数字化的观察数据人性化。

参观完各展厅的16岁以上的成人均可选择(如上所述,采取连续随机抽样法)作为受访者,回答一些关于他们的参观体验的问题。我们的访谈提纲有意设计成开放式的,以便受访者能自由讨论他们觉得有意义的事情。所有的访谈在受访者知情的情况下录音并转录成文字,以便将来分析。

焦点观察和访谈

焦点观察和访谈旨在深入了解观众如何利用某些具体的展项,他们在认知和情感上对这些展项的内容和展示方法做出什么样的反应,以及他们从这些展项中学到了哪些知识。在每个展厅,我们都选择一两个展项进行观众调研。

采用先前所述的连续随机抽样法,调查人员请观众(年龄在16岁或16岁以上)使用某件交互式展品。这种征募调查对象的方法被称为提示式测试(cued testing)。当那些同意参与的观众使用交互性展品时,调查人员在旁进行观察。在观众参观完该展项后,调查人员与他(或她)进行开放式访谈。所有的访谈在受访者知情的情况下录音并转录成文字,以便将来分析。由于焦点观察和访谈比较琐碎,因此我们没有把它们纳入本评估概要中,有关内容请参照我们对各个展厅的评估所得出的相关调查结果。

评估数据分析

计时和跟踪调查所获得的是定量数据,这些数据被输入到电脑中进行统计分析。百分比和总计数字有中位数(指在一组应答者的数据中居于中间的数,即在这组数据中,有一半应答者的数据比它大,有一半应答者的数据比它小)、平均数和标准差,用于计算间距变数和比例变数。为研究两个不同观众组之间的关系(例如带孩子的观众组和没有带孩子的观众组),我们进行了T检验和U检验计算。

在本报告中,我们大都采用中位数而不是平均数,这是因为观众使用展览展项的数量和观众驻留的时间在整个展区中的分布很不均匀。例如,尽管大多数的观众在一些展览展项上停留的时间较短,但少量观众却呆了特别长的时间。当数值分布极端不均匀时,平均数会极大地受到极值的影响,以致远离了分布的中心区。在这种情况下,中位数是较为可取的度量值,因为它对超过和低于

它的数值不敏感,只对数值的数目多少敏感。

对焦点观察所收集到的观察数据以及访谈调查获得的受访者对访谈提纲中访谈问题的回答的文字稿进行定性分析,意味着评估人员需要研究受访者的回答,归纳出若干有意义的回答模式,将类似的回答归为一组,并给每组起一个名字或者类别名称,以表达此类回答所包含的意义。

评估结果讨论

我们对创新技术博物馆的评估是全面的、复杂的。为能完整地描述观众的参观体验,我们采用了多种数据收集方法,包括:

- 在各展厅进行观众观察,以了解观众都参观了哪些展品以及他们在展品上驻留了多长时间;
- 观众参观完展厅后进行开放式访谈,以期将展厅观察所获得的观众行为数据人性化,定性剖析观众的参观体验;
- 在四个展厅的五个展项处进行观众观察和访谈,以了解观众在操作和理解展项上存在的问题;
- 进行参观后问卷调查,以了解观众参观整个博物馆的体验;
- 进行参观后开放式访谈,以定性地了解观众参观整个博物馆的体验;
- 在观众参观博物馆3个月(或更长时间)之后,对其进行电话访谈,以判定参观博物馆对观众的长期影响。

本评估概要是前三种调查数据的综合。

对创新技术博物馆进行的观众观察和访谈表明,与其他科学展览的观众研究相比,观众在参观创新技术博物馆展厅时的表现不太一样。兰迪科恩公司(RK&A)及其他评估机构曾发现:在参观科学展览时,观众往往从一个展项很快转到另一个展项,很多时候只是按按钮和拉操作杆,看展品会出现什么反应,并没有仔细观

察所发生的现象或探究展项所表达的内容。然而,在创新技术博物馆中,观众的行为却并非如此,观众往往用较多的时间来使用较少的展项,说明创新技术博物馆的展项对观众非常有吸引力。

观众总体的参观行为模式

计时和跟踪观察全面调查了观众在四个展厅中的行为经历。虽然观众的参观经历极具个性,而且四个展厅也各有特色,但在比较了四个展厅的观众调查结果之后,我们还是发现了某些带有普遍性的趋向。

观众经停的展项和总的驻留时间

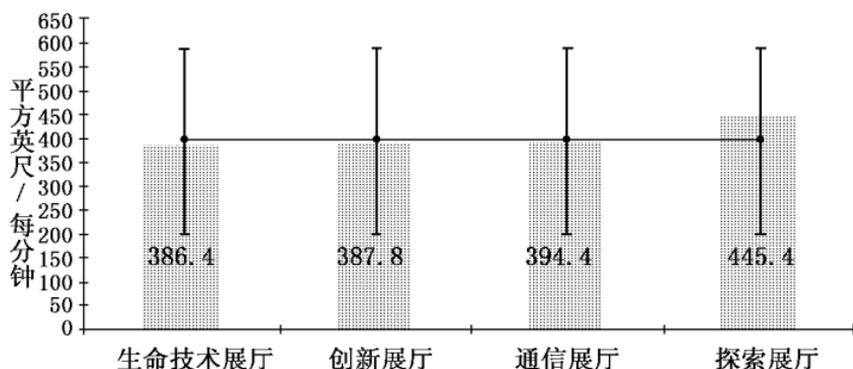
评价观众参观经历的一种方法是比较观众参观各展厅时经停的展项总数以及在各展厅的总的驻留时间。从观众在各展厅经停的展项数和在展厅的驻留时间来看,四个展厅的情况非常接近。观众在各展厅经停的展项为 7~11 个(中位数):观众在生命技术展厅经停的展项最多,在通信展厅经停的展项最少。观众在各展厅驻留的时间为 14~17 分钟(中位数),在创新展厅驻留的时间最长,在通信展厅驻留的时间最短。

为了更好地了解观众在创新技术博物馆各展厅的经停展项数和驻留时间,我们将观众参观创新技术博物馆各展厅的情况与参观规模相当的其他展览的情况进行比较。在进行此类比较,瑟雷尔(1997年)用观众花在展览上的时间来计算“扫描速率指数”(SRI),用观众经停的展项数来计算“用心的观众的百分比”(DV 百分比)。

SRI 的计算是用展览面积除以观众参观该展览的平均驻留时间。SRI 值越低,观众在每平方展览面积中驻留的时间就越长。图 1 给出了创新技术博物馆各展厅的 SRI 值。生命技术展厅的 SRI 值最低(每分钟 386 平方英尺),其次为创新展厅(每分钟 388

平方英尺)、通信展厅(每分钟 394 平方英尺)和探索展厅(每分钟 445 平方英尺)。从 SRI 值来看,观众在创新技术博物馆生命技术展厅、通信展厅和创新展厅的观看速度与参观同等规模的其他展览的观看速度差不多。观众在探索展厅的观看速度较高,但仍位于瑟雷尔 SRI 平均值的标准差之内。

图 1 观众在各展厅的扫描速率指数



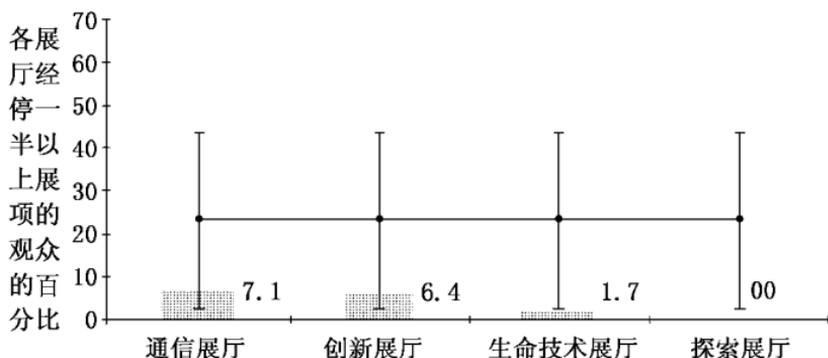
图中各条形棒中的竖线表示相对于瑟雷尔 SRI 平均值的大标准差。

“用心的观众的百分比”(DV 百分比)是用经停一半以上展项的观众的百分比计算的。DV 百分比越高,展览的利用率越高。图 2 显示了各个展厅的 DV 百分比。通信展厅的 DV 百分比最高(7%),其次为创新展厅(6%)、生命技术展厅(2%)和探索展厅(0%)。所有这四个展厅的 DV 百分比均低于瑟雷尔的 DV 百分比平均值,这意味着与同等规模的展览相比,观众在创新技术博物馆各展厅经停的展项较少。同样,图中各条形棒中的竖线表示瑟雷尔 DV 百分比平均值的大标准差。

一些策展人员反对使用瑟雷尔的 DV 百分比模型,认为该模型对展览成败的评判是基于观众经停的展品的数量多少而不是观众个人参观体验的质量,因为有些策展人员更关注的是驻留时间

而不是展览中展项的利用率。为此,当讨论 DV 百分比时,需要考虑策展人员的设计目标。如果策展人员设计他们的展览是为了观众有较长的驻留时间,并且他们对只有不到四分之一的观众使用不到一半的展品无所谓的话,那么创新技术博物馆的四个展厅应当是做的不错了。

图 2 各展厅“用心的观众”的百分比



然而,如果策展人员偏爱高 DV 百分比,在对此次调查结果变得过于惊慌之前,他们需要认识到,这些数据的某些限制因素会影响 DV 百分比。例如,观众可能觉得某些展项特别有意思,他们愿意把时间花在这些展项上,而不再去尝试其他很多不同的展项。事实上,我们在展厅出口处的访谈表明,观众往往先称赞展品总体的交互性,然后会具体说出有哪几个展项特别好。而且,观众或许打算什么时候还来展厅。由于观察调查的非介入性,数据采集者不知道对观众的跟踪调查发生在观众参观的哪一点(也就是说,不可能知道观众以前是否参观过这一展厅,或者是否打算日后再来参观)。而且,回头客再次参观时往往只光顾他们喜爱的展项,而把其他的展项排除在外。回头客还可能只着重参观他们所选择的展项,因为他们知道他们还能再来,不必将所有的东西一次就看完、做完。从本次观众调查的初步结果来看,大约五分之一的观众

(19%)是回头客,考虑到在我们进行本次调查的时候,创新技术博物馆还是一座开馆仅 8 个月的新馆,19%的回头客比例是相当大的。尽管如此,对策展人员来说,重要的是要考虑他们自己到底对观众行为有什么期望。他们是想为观众提供若干深入体验的展品?还是提供一系列快速的体验?策展人员或者满意观众目前使用展馆的方式,或者认为观众的行为应该有所不同。

尽管在理解观众行为的数据时还存在某些未知的变量,但是通过探究观众在展馆内是如何活动的以及他们选择在哪些展项前停下来,会使我们弄清楚展览的使用情况,有助于我们总结观众的参观体验以及展览的得失。从这个意义上来说,有必要弄清楚这个问题:即为什么和其他展厅相比,探索展厅的 DV 百分比最低。

在四个展厅中,探索展厅最大,展项数量也最多。瑟雷尔等人曾发现,观众参观大型展览所用的时间比小型展览短。对于观众,少往往就是多,这是对探索展厅 DV 百分比低的一个非常简单的解释。对这一问题,还可以从探索展厅为观众提供的展览类型加以解释。在探索展厅,有大量无器物展板(panels without artifacts),共有 21 个,是其他展厅的 2 倍多。本次各展厅调查以及以往的其他研究均表明,无器物展板对观众的吸引力和保持力不是很好。此外,探索展厅中计算机交互展品最少,而从其他三个展厅的情况看,无论是根据观众驻留的频次还是驻留的时间,计算机交互展品都是颇受观众欢迎的展项。由于探索展厅中不吸引观众注意的展项(如无器物展板)居多,而观众喜欢的展项(如计算机交互展品)却较少,这两个因素迭加起来造成探索展厅的低 DV 百分比表现。

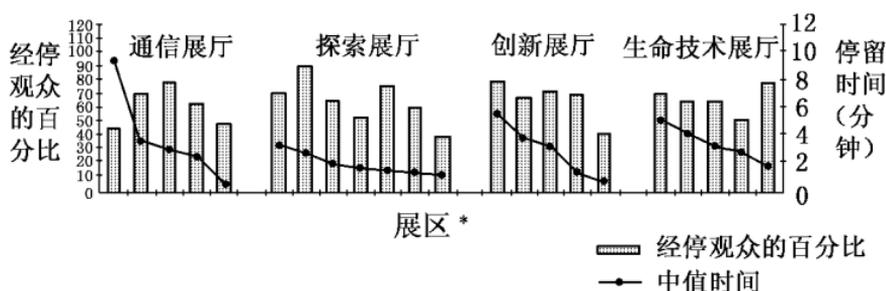
对生命技术展厅的低 DV 百分比(2%)也应作些解释。生命技术展厅是最小的展厅,但它有 65 个展项,只比探索展厅少 10 项,但比其他两个展厅都多。它有 15 个计算机交互展项和 14 个机械交互展项,有一半以上的观众在其中的三个或更多的展项前逗留,平均在每个展项上驻留的时间大约是 4 分钟(中值)。除计

算机和机械交互展项外,观众在生命技术展厅其他展项前经停的次数和时间则相当短。因此,观众选择把他们的时间花在少数几件展品上,而不是参观大量的展品。这对观众驻足参观的展品来说是个好消息,因为它意味着这些展品有吸引力,抓住了观众的注意力。如前所述,采用是否有超过四分之一的观众在超过一半的展品处驻足这一模型对判断展览成败并不总是合适的。

观众参观展厅各展区的情况

了解观众行为的第二种方式是分析观众参观各展厅展区的情况(参见图3)。我们的调查表明,至少一半的观众参观了大部分展区:至少一半的观众参观了生命技术展厅的所有展区、创新展厅和技术展厅中除一个展区以外的所有展区、以及通信展厅中除两个展区以外的所有展区。这表明,各展厅展区都有多种多样的展项,吸引了大部分观众。

图3 观众在各展厅经停的数量和时间



译注：* 图中每一竖形条的下方是各展厅不同展区的名称，由于英文原图中各展区名称字体非常小，无法辨认，故未将各展区的名称译出，这对我们理解评估者此处所要表达的评估思想应无大碍。

展区的吸引力有时是由其所处的位置决定的，因此，我们对生命技术展厅中的“透视”展区和通信展厅中的“全球网络”展区吸引了大多数观众并不感到奇怪(参见图3)，因为这两个展区都位于

所在展厅的入口处。然而,在探索展厅,“深度边界”展区吸引了大多数的观众,而在创新展厅,“虚拟设计”是最受欢迎的展区,这两个展区都不在入口处,说明展区位置似乎不是决定性因素。在探索展厅,较多的观众停留在“深度边界”展区而不是在入口的地方。“深度边界”展区之所以具有如此大的吸引力,是因为有一个引人关注的大型遥控潜水器(ROV),它成为这个展区的代表和象征。在创新展厅,很多观众也径直走过入口区,而直奔“虚拟设计”展区。这点特别值得注意,因为“虚拟设计”展区位于展厅入口的左侧,而在通常情况下,当观众参观展览时,他们习惯靠右走。“虚拟设计”展区有一些大型的、受欢迎的展品,这些展品可能是“虚拟设计”展区具有较高吸引力的原因。

展厅展区的观众保持力如何也能说明一个有意思的问题。与吸引力不同,保持力与展区的位置无关。虽然入口区有较高的吸引力,但它们对观众的保持力较低。事实上,在通信展厅和创新展厅,观众在入口区的停留时间还不到1分钟。同样,在通信展厅的所有展区中,“数码演播室”展区吸引的观众最少,但观众在该展区停留的时间最长,超过了9分钟(中值时间)。

探索展厅的“生活在不平静的星球”展区和生命技术展厅的“生命的新前沿”展区也都有较高的观众保持力。但是在所有四个展厅的展区中,只有一个展区同时具有高的吸引力和保持力,这就是创新展厅的“虚拟设计”展区。该展区不仅吸引的观众最多,而且对观众很有保持力,观众在该展区停留的时间达6分钟(中值时间)。

不同类型展项的吸引力和保持力

考查观众行为的另一种方式是分析他们对不同类型的展项都有什么样的行为反应。在创新技术博物馆,共有7大类展项:计算机交互类、机械交互类、仿真交互类、配备工作人员的展项、有器物

展板、无器物展板以及伦理问题展项。

在所有四个展厅中,一半以上的观众在一个或多个机械交互类展项和一个或多个仿真交互类展项处停留。在除探索展厅之外的其他展厅,一半以上的观众在一个或多个计算机交互类展项处停留。如表 1 所示,机械交互类展项和计算机交互类展项受到欢迎,部分原因是由于展厅中这类展项较多见。然而,在创新展厅、通信展厅和探索展厅,更多的是其他类型的展项(如无器物展板),但这些非交互类展项在吸引观众停留上仍比不上交互类展项,这表明交互性展品受欢迎的原因并非完全是由于其数量多。对于仿真交互类展项和配备工作人员的展项,情况的确如此,因为这类展项在各展厅的数量都不多(一般为 2~4 个),但大多数观众在一个或多个这类展项前停留。

对各类展项的保持力的考察进一步证实了机械交互类展项、仿真交互类展项、配备人员的展项和计算机交互类展项的受欢迎性。如表 2 所示,在各展厅中,观众在机械交互类展项和仿真交互类展项前停留的中值时间超过了 2 分钟。计算机交互类展项的保持力甚至还要高,在通信展厅、创新展厅和生命技术展厅,观众在计算机交互展项前停留都在 3 分钟以上。在通信展厅,观众在计算机交互类展项前停留的时间高达 9 分钟。在探索展厅,尽管那里的计算机交互展项吸引力较低,但保持力仍然较高,观众驻留的时间为 2 分钟。在除通信展厅之外的其他展厅,配了人员的展项也有较高的保持力,观众驻留的时间在 1~2 分钟。

除这些总体趋势之外,还有一个因展厅而异的调查结果值得注意。在创新展厅,有器物展板是最受欢迎的展项,观众平均在两个有器物展板前停留。此外,有器物展板有较高的保持力(中值时间为 1 分钟)。访谈数据使我们得以了解观众的行为。如本报告后面所讨论的那样,受访者对了解如何制造微芯片十分入迷,这说明他们花时间阅读了有器物展板。

表 1 在各展厅不同展项前经停的观众数量

展项类型	通信展厅			探索展厅			创新展厅			生命技术展厅		
	展项数量	经停观众的百分比	经停的展项中值数	展项数量	经停观众的百分比	经停的展项中值数	展项数量	经停观众的百分比	经停的展项中值数	展项数量	经停观众的百分比	经停的展项中值数
计算机交互类	13	83.5	3	5	46.2	0	7	76.2	1	15	85.2	3
机械交互类	5	65.2	1	18	90.4	4	10	61.9	1	14	73.9	3
仿真交互类	3	52.2	1	4	69.2	1	2	52.4	1	2	53	1
配备工作人员的展项	1	46.1	—	4	77.9	1	2	57.1	1	2	23.5	0
伦理问题展项	7	45.2	0	2	7.7	0	3	62.7	1	6	22.8	0
有器物展板	—	—	—	8	49	0	7	62.7	1.5	2	28.7	0
无器物展板	6	21.7	0	21	51.9	1	8	34.1	0	8	29.6	0

表 2 观众在各展厅不同展项前驻留的时间

展项类型	通信展厅		探索展厅		创新展厅		生命技术展厅	
	n	中值时间	n	中值时间	n	中值时间	n	中值时间
计算机交互类	96	8分36秒	48	1分36秒	96	4分47秒	98	3分54秒
机械交互类	75	2分20秒	94	2分17秒	79	2分20秒	85	3分38秒
仿真交互类	60	1分35秒	72	2分24秒	66	2分6秒	61	2分15秒
配备工作人员的展项	53	—	81	2分8秒	72	1分57秒	27	1分16秒
伦理问题展项	52	1分1秒	8	39秒	79	47秒	26	28秒
有器物展板	—	—	50	27秒	79	1分20秒	33	42秒
无器物展板	25	33秒	34	50秒	43	12秒		37秒

展览主信息的传递情况

观众对展览内容了解了多少? 衡量方法之一是在观众参观完展览之后与他们交谈。展厅访谈使我们能了解观众通过参观学到了什么,有什么体验。在通信展厅和探索展厅,观众在谈及他们的学习情况时显得有些含糊,也就是说,观众不一定能确切说出他们学到的具体概念,他们谈及的是其与新技术的接触和互动,有时候是对一种新现象(如地震)的体验。另外,对观众来说,看到肉眼无法看见的东西是其学习经历的一个有趣部分。

然而,在生命技术展厅和创新展厅,观众获得的是一种很不同的学习经历。在生命技术展厅中,观众回忆起他们参观某些展品后获知的一些具体事实,或许是因为那些事实与健康 and 人体有关,而健康与人体是一个与所有的观众都相关的主题。正如到探索展厅和通信展厅的观众喜欢看工艺机械的内部结构一样,到生命技术展厅的观众喜欢看人体内部结构。在创新展厅,超净室和手工台促进了观众的学习体验。例如,观众对了解芯片开发非常兴奋,

喜欢在超净室观看。其他观众则喜欢手工台所提供的有教育意义的体验。尽管观众不能确切说出他们通过手工台获知了什么,但他们喜欢参与这样的学习过程。

观众的学习体验非常宽泛,但正如各展厅的展览计划所阐述的那样,各展厅的目标是以内容为基础的。通常,当策展者着手开发一项展览时,他们会思考展览要传达的具体概念和信息。在最理想的情况下,展览是按照策展者起初想要传达的概念设计完成的。许多有经验的展览评估专家和策展人主张开发概念宽泛的展览,他们建议展览中的所有展品都应与主题概念相呼应。所以,最终展览的具体目标需具有一致性,而且各项展品必须反映出展览的目标。本调研并不是要评论创新技术博物馆各展厅的目标,我们的讨论是要考查另一个更大的话题,即要想使观众掌握展览的本质概念,展览应该包括些什么。展览除了要有传意的实物展品外,如果要让观众理解展览的概念及概念框架,还需要提供展览简介。

展厅简介

我们提供给调查人员的观察和访谈指南有一些旨在了解观众在展览简介区行为和体验的调查问题。观察调查结果显示,至少一半的观众在各展厅的简介区停留。这一调查结果并不意外,因为简介区位于展厅的入口处,这样无疑增加了它们对观众的吸引力。然而,保持力则不随场地而变化,观众在简介区驻留的时间微乎其微。在探索展厅和生命技术展厅,观众在简介区驻留的时间约为1分钟(中值),而在创新展厅和通信展厅,观众在简介区的驻留时间还不到1分钟。而且,值得注意的是,与各展厅的其他展区相比,观众在简介区停留的时间最短(参见图3)。这一点值得注意,因为其他的调研显示,观众在刚开始参观的时候,在展品处驻留的时间较长,在参观的末了,由于越来越疲劳,在后来参观的展

品前驻留的时间会较短。因此,如果观众在展厅简介区驻留的时间短,说明展厅的简介区不是很吸引人。

在访谈期间,观众被问及他们是否观看了作为展厅简介的展区。约一半的观众确切说出了探索展厅中的探索者的手以及创新展厅中的明信片。在20个受访观众中,只有一个观众答出了通信展厅的简介区,两个观众答出了生命技术展厅的简介区。与其他简介区相比,大多数观众能确认探索者的手是探索展厅的简介区,这是有道理的,因为探索展厅是惟一提供单独的场地安放展厅简介并用说明文本来介绍展厅主题的展厅。此外,一些观众还认为,探索者的手很适合作为展厅简介,因为它提供了一个镜头,观众可以通过镜头观看展览——而这正是展厅简介应该为观众做的事。

判断展厅简介区优劣得失的另一方面是考查观众是否能确切地说出各展厅的主题。有种假说认为:展厅简介,有时被称作“前导组体”,为展览提供了物质架构和概念架构,引导了展览的主题。有人认为,恰当的展览简介能使展览更宜人、更易于理解(格里格,1983年)。比如,在探索展厅,几乎所有的观众说出了人类三个主要的探索区域——陆地、海洋和太空,少数观众将展厅的主题理解为展示新的信息和技术。相比之下,在到过生命技术展厅的观众中,有一半的观众认识到该展厅是有关健康和人体的,较少的观众把技术和人体研究联系起来。同样,观众在参观过通信展厅后,仅仅感知到展厅展示内容表层的东西,说展厅是展示通信装置的。这就是说,观众到通信展厅参观后,只体会到展览表层的东西,而没有认识到展览背后的更大的主题思想,而在探索展厅,观众是在一个大主题中观看各个展项的,可能是因为该展厅向观众提示了其大主题的缘故。观众对创新展厅的基本主题的理解可以说是五花八门。例如,创新展厅中超净室给一些观众的印象非常深,这些观众认为该展厅就是展示芯片制作的。少数观众认为该展厅是展

示新技术进展的,还有一些观众不能说出该展厅的主题。

如果博物馆策展人员认为展览的主要信息对于观众十分重要,或者如果展览的展示内容存在一个概念架构,策展人员就应让观众了解这些信息和概念架构。展览简介区是提供这类信息概念的最合适的地方。观众在参观博物馆时常常会碰到难题。他们在展品前驻足,是因为想体验和了解它们。然而,如果观众对展览的主题缺乏深度的了解(很多观众是这种情况),而展览又没有采取任何手段引导观众以特定的方式思考和体验,那么观众参观的效果就不可能达到策展人员的期望。

如果有一个具体的概念架构能使展览主题结构紧密,那么展览中的各个单一展品应用来加强和支持这一概念架构。各个展品采取怎样的展示方式,即它们如何与整个展览的大主题相联系以及如何传达展览思想,也是值得进一步探讨的问题。

沉浸式展品

创新技术博物馆的四个展厅都包括沉浸式展品。策展人员想知道观众是否知道他们处于一种沉浸环境中,如果已知的话,需要提供哪些要素才能让观众体验沉浸式展品,以及营造的沉浸环境在传达展品内容方面是否有效。对于此类问题,我们的观众访谈调查获得了相关的具体信息。

我们对参观完创新展厅的观众的访谈表明,几乎一半的观众并没有进超净室参观,但进了超净室的观众基本上都认为,超净室是一种沉浸环境。对超净室有所了解的观众认为,超净室是真实的,从地板到明亮的光线直到测量人身体内的灰尘数量,都给人真实的感觉。毫无疑问,对于那些进入超净室的观众来说,超净室是一个令人信以为真的环境,不过有一位观众提出,如果能让超净室看起来确实真实,还应该让观众看到超净室的工作人员。观察数据显示,有54%的观众进入了超净室,但是这些观众中有90%的

人是从展厅的出口处进入的。虽然按照预定的入口进入超净室会更好,但观众逆向通过超净室似乎没有降低他们的参观体验质量。许多受访者称超净室提供了一种有教育意义的体验,说他们从中知道了芯片制作和硅是怎么回事。

手术室做得和超净室差不多一样好,因此观众也把它看作是一种沉浸环境,但是它没能为观众提供内容丰富的体验。四分之一的受访者没有进入手术室,其余 15 个受访者中,有 11 人认为手术室是一种真实的现实世界。针对如何提高参观体验,观众提了许多建议。聚乙烯地板和盖尼式床模型是令人信服的,但是将近一半的受访者提到,应该展示真正的器械,而不仅仅是一些图片;房间应该明亮,墙壁应搞成绿色或白色,门应是双旋转的以方便观众进出。只有一个受访者认为“行星基地”和“电子聊天室”是沉浸环境。

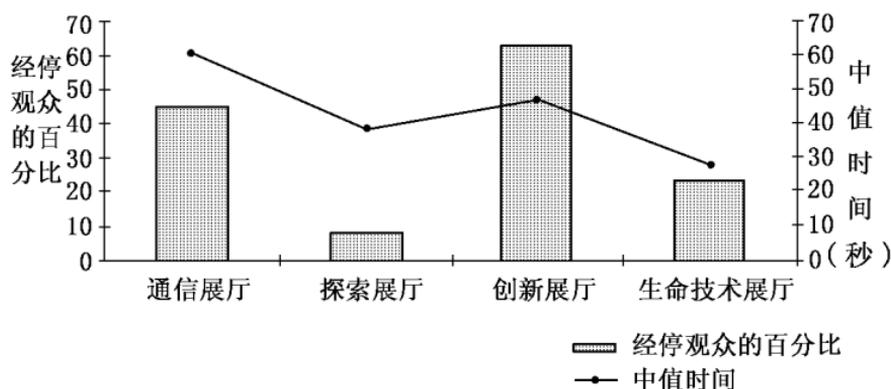
很显然,观众十分看重真实性,如果某个物件是真实的,他们似乎看得出来。超净室很容易被看出是真的,但其他展览环境则是些不足以让人相信的仿造品。在我们以前进行的调研中,观众都特别强调了真实的重要性:不仅表现特定事物的环境要真实,而且向观众传达的信息也应是真实的。“行星基地”和“聊天室”显得不够真实,观众说它们看起来太像是展览环境。如果是展览中的解释性材料妨碍真实性或许还可以接受(超净室就是如此);但如果是展件妨碍现实环境则属过度破坏了真实的环境,“行星基地”就是这样一种情形。在“行星基地”的演示现象应可以重复以便于观众观看,而且它应该是真实形成的。

伦理问题展项

创新技术博物馆的四个展厅都包括两个或两个以上的伦理问题展品。生命技术展厅和通信展厅分别有六七个伦理问题展品,探索展厅和创新展厅分别有两三个伦理问题展品。对观众的行为

观察数据表明,创新展厅中“创新论坛”伦理问题展品吸引了最大部分的观众,63%的观众在该展厅的伦理问题展品前停留,中值时间为47秒(参见图4)。通信展厅有七个伦理问题展品,45%的观众在其中的1~3个伦理问题展品处停留。与其他三个展厅中的伦理问题展品相比,通信展厅中的伦理问题展品保持力最高,观众在其文字展板前驻留的时间达1分钟。生命技术展厅有六项伦理问题展品,23%的观众在其中的1~3个伦理问题展品前停留。在四个展厅中,生命技术展厅的伦理问题展品保持力最低,中值时间为28秒。探索展厅有两项伦理问题展品,其吸引力在四个展厅的伦理问题展品中是最低的,只有8%的观众在其伦理问题展品前停留,驻留的中值时间是39秒。

图4 在各展厅伦理类展品处经停的观众数量和时间



表现伦理和社会问题的展品很自然地吸引了观众的注意力,吸引观众驻足思考,因为博物馆展品在展示有争议的问题时,所展现的往往不是特别主流的观念。一些受访者称,遇到这样一种展品时,所产生的惊奇感本身就对他们产生了潜在的吸引力。大多数受访者表示,伦理问题展品表现的问题很重要,他们支持博物馆展示这样的问题。如果伦理问题展品没有引起观众的注意,人们

可能推测那是展品的位置和摆放不够好,或者是展品的设计——不仅仅包括展品的外观,而且也包括展品的类型——不引人注目。例如,在创新展厅中,最吸引人的两个伦理问题展品都是计算机类的,它们吸引了29%的观众;而通信展厅中吸引观众最多的两个伦理问题展项分别吸引了18%和13%的观众,这两个展项都是有器物展板,其中一个还包含一项机械类展品。

就保持力而言,通信展厅中的伦理问题展品较为出色:“隐私:实施控制”展品让观众停留了44秒;“电话:实施控制”让观众停留了40秒。然而,有意思的是,当我们向参观通信展厅的观众询问有关伦理问题展品的问题时,他们说没有学到任何新东西。也许这些观众把时间花在了阅读文字展板上,这种阅读只证实了他们已知的事情。访谈结果显示,尽管观众觉得有关的信息不新鲜,但他们确实认为这些信息是与他们相关的。

通常,如果观众没有花时间思考有关的问题,人们可能推测是展品没有新内容,或者他们看不出展示内容与他们自身的生活有什么关系。我们采用焦点观测和访谈等数据调查方法,对生命技术展厅中的“生物技术始末”和通信展厅中的“谁在谈?谁在听?”两个伦理问题展品进行了深入调查。所得调查数据能使我们了解观众对这类展品到底有什么样的反应,以及如何改进观众参观这类展品的体验。

“生物技术始末”展品的吸引力和保持力均较低。对观众的焦点观察和访谈显示,那些停下来看展板的观众对展板中所呈现的生物技术伦理难题是认可的,他们赞同博物馆采取的解释策略,即以叙述人的不同体验的方式呈现伦理问题。受访者说,看其他人的观点促使他们思考自己的观点,有时甚至会激发他们与自己的孩子一起讨论。

如果一项展品促进了预期的社会互动(例如促使观众谈论展品的内容),就值得称道。人们认为博物馆是促进观众社会互动的

场所,但有时所观察到的观众社会互动却与展品或展示内容无关。显然,“生物技术始末”展品吸引观众停留,并且思考他们在那样的情景下会怎样做。“生物技术始末”的情景内容或许与大多数观众没有太大关系,但观众仍会设想自己是其中一员,思考他们会怎样做。对于那些将展板内容从头看到尾的观众,“生物技术始末”的内容有强烈的吸引力。“谁在谈?谁在听?”则没有这样的吸引力。当然互联网隐私问题对个人不是什么生死攸关的事,但几乎所有的受访者都认为,“谁在谈?谁在听”中有些东西与他们个人相关。尽管有相关性,除少数受访者之外,观众普遍认为通信展厅的这些展板上的内容并不新颖。个人相关性是展品开发者需要考虑的重要因素,但内容也同样重要,因为展品必须将观众引入一个新的领域。对某些观众来说,“生物技术始末”的内容不好理解,但其他观众则认为该展品的展示内容过于简单化,认为内容应更深些。

创新技术博物馆想了解观众是否认为它的展品均衡展现了各种伦理问题。在看有争议的伦理问题展览时,一些观众会看展览中是否有他们认同的那种观点。如果没有,他们会认为展品的内容展示不均衡。一些参观“生物技术始末”、“谁在谈?谁在听?”展览的观众就认为展示有失均衡。例如,有的观众在观看“生物技术始末”时想要看到保守的观点,而有些观众在观看“谁在谈?谁在听?”时想看到支持言论自由的观点。

伦理展板的设计也是我们调研的内容。在两个展厅,观众不全认为伦理展板有视觉吸引力。“谁在谈?谁在听”展板上的感叹号和观察孔被观众大大忽视和误解。“生物技术始末”展示的是严肃性的内容,照片和语言泡就显得不太相称。甚至放置“生物技术始末”展板的空地也让人感觉不舒服。展板的设计和摆放方式影响了观众的行为。另外,观众提到,在这两个展厅,伦理问题展品缺乏动手的选项,动手的展品可以起到吸引儿童注意的作用。“谁在谈?谁在听?”的观众更多地认为应让儿童了解这样的问题。

总之,博物馆人员需要考虑将展示内容与观众、展示内容与设计协调起来——展品如何展示,放置在什么地方,选择什么样的媒体。例如,如果展示的内容与儿童的生活有关,那么就应该提供动手展项,以吸引儿童。如果展示的内容主要是针对成人的,文本或许是合适的展示方法。如果展示的内容是针对成人的但能与儿童一起讨论,博物馆也许可以向家长们提供指导,使他们知道如何与他们的孩子更好地探讨这个话题。如果展示的内容属于隐私的问题,那么就应该将展品安排在私用空间,这样观众在观看展品或与他们的家人讨论时,会感到比较放松。

展厅的总体功能情况

关于展厅的功能,创新技术博物馆的工作人员提出了他们关切的一些问题:博物馆工作人员在观众参观体验中的作用、观众遭遇损坏的展品的数量、观众等候使用展项的频率以及计算机交互类展品和视频中使用西班牙语的问题。

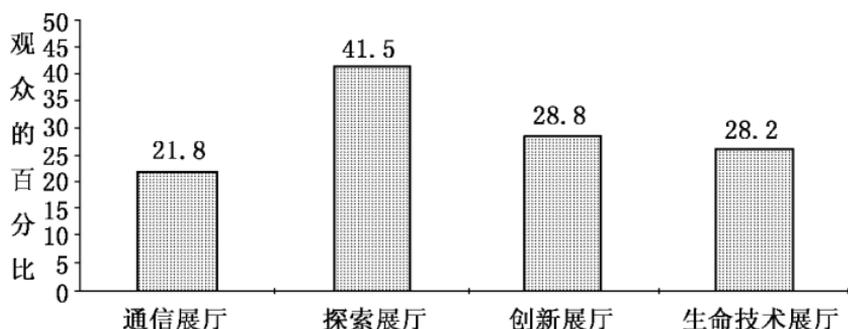
博物馆工作人员

观众利用展厅工作人员,并对工作人员的工作表示欣赏。对观众的计时和跟踪观察表明,在各展厅中,约三分之一的观众至少与一名工作人员进行了直接交谈(参见图5)。来自观众调查的初步数据显示,三分之二的观众在参观所有的展厅时至少与工作人员交谈过一次或接受过他们的帮助。这些调查结果证实了大部分的观众参观了一个以上的展厅。有趣的是,观众调查显示,观众喜欢展厅中有更多的工作人员,他们对工作人员可用性评分为4.89(按1~7分的量表,1表示差,7表示极好)。此外,当被要求描述他们在展厅中的体验时,一些观众选择了“我希望有更多的展厅工作人员来帮助我使用展项”。

除了利用展厅工作人员外,观众对他们与工作人员的交流感

到满意。调查的应答者对博物馆工作人员的礼貌评分为 6.04(按 1~7 分的量表,1 表示差,7 表示极好)。此外,如前面谈到的,配备工作人员的展品属高吸引力和高保持力的展项。在这一点上,创新技术博物馆的调查结果与我们以前进行的其他调查结果差不多,即到科学博物馆参观的观众格外看重配备了工作人员的演示项目。

图 5 与展厅工作人员进行交流的观众的百分比



损坏的展品

在调研期间,我们发现,创新技术博物馆的每个展厅都有 1~8 个损坏的展品。考虑到展厅中展项的总数,应当说损坏的展品所占百分比并不算高。然而,由于观众只在少量展品前停留(观众在各展厅经停的展品数为 7~11 个),所以即使遇到一个损坏的展品,也会影响他们的体验。事实上,重要的并不是损坏的展品的发生率,而是展厅中损坏的展品的数量给观众的不良感觉。

根据初步的观众调查结果,相当数量的观众受到展厅中损坏的展品的消极影响。当我们让观众选择他们对展品的体验的回答时,四分之一的观众选择的是“我感到失望,因为有的展品坏了”。另一方面,观众对展品总体维护状况的评分为 5.22(按 1~7 分的量表,1 表示差,7 表示极好例)。尽管这一评分很高,但我们仍能

看出观众对展品维护上的一些不满。四分之三的观众报告说,他们遇到过损坏的展品,其中约五分之二观众受到损坏展品的烦扰,认为它们实际上增加了门票(15%),减少了独特展品而增加了复制品数量(15%)以及减少了动手型展品的数量(11%)。

观众等候使用展品的情况

计时和跟踪观察的结果表明,观众总体上很少在展项前等候,很少花时间排队等候使用展项。如图6所示,在各展厅中,只有大约三分之一的观众排队等候使用展项。就时间而言,大多数排队等候的观众的等候时间占他们全部参观时间的10%或者更少。

图6 在各展厅等待使用交互性展品的观众的百分比

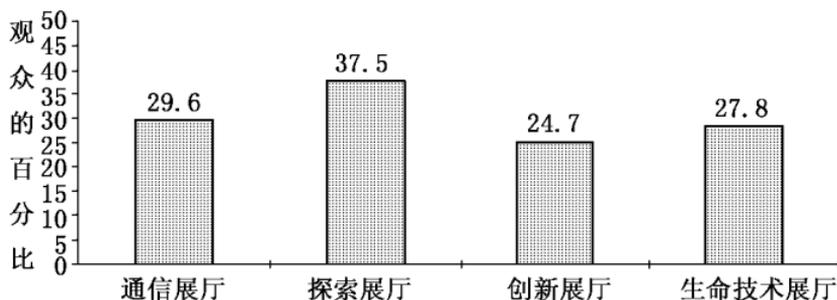
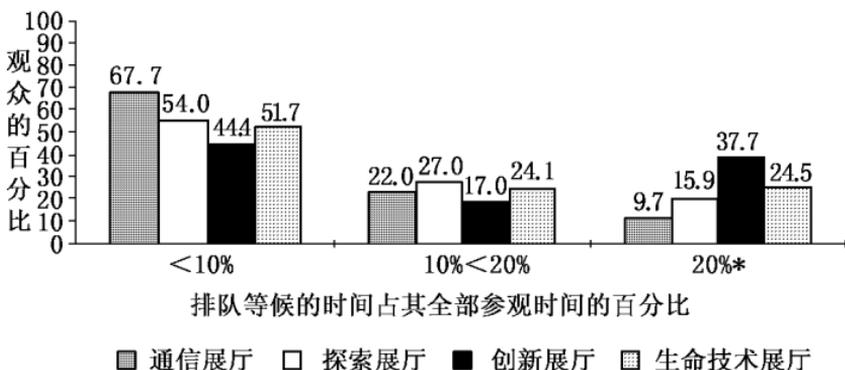


图7 观众在各展厅排队等候的时间占其全部参观时间的百分比



尽管跟踪观察显示观众在展项前等候的时候很少,但如前所说,等候给观众造成的感觉却很不好(换言之,观众可能觉得他们在很多展项前都要等候)。初步的观众调查结果显示,有相当多的观众指出,排队等候使用展项对他们的参观体验造成了负面的影响。几乎一半的受调查观众称,他们在某些展品处因排队太长,以至于他们最终不等了。尽管四分之一的观众称,展品很有意思,他们不介意排队等候,但其他观众则说,他们希望排的队短些,即使这意味着减少展厅中独特展项的数量。只有少数观众说他们感到失望,因为他们到博物馆是来看某一专门展品的,结果这一展品总是被人使用着。

观众选择使用西班牙语的情况

创新技术博物馆的各个展厅中,有5~13个计算机交互类和视频展项提供了西班牙语的说明和内容,供观众选择使用。在创新展厅,5个观众在两个视频展项和一个计算机交互展项上选择了西班牙语,共计7次;在通信展厅,4个观众在三个计算机交互类展项上选择了西班牙语;在探索展厅,1个观众在一个视频展项上选择了西班牙语;在生命技术展厅,没有观众选择使用西班牙语。

博物馆人员看到西班牙语的使用率这样低,可能会感到沮丧。不过,我们必须在当前观众的构成的背景下看这些跟踪数据。观众调查的初步结果表明,到创新技术博物馆的观众中,12%的人说西班牙语,而在这12%的观众中不到一半的人在提供西班牙语选项的展项处选择了西班牙语。这些统计数字表明,博物馆需要更多地发展说西班牙语的观众。创新技术博物馆的工作人员已经采取了初步行动:在博物馆内设立有西班牙语选项的设施,以满足说西班牙语的观众的需求,对此他们应受到赞扬。接下来的工作是要培养一批说西班牙语的博物馆观众。

建议

• 博物馆应精心构思观众行为目标,以便据此量测被观察观众的行为。

• 在开发新展品时,博物馆应力求保持各展区中展品类型的多样性,以便均衡各个展区的吸引力。配备工作人员的展品、仿真交互式展品以及计算机交互类展品是最受欢迎的展项类型,需要仔细考虑这些展项的摆放。

• 通常,无器物展板不能吸引或留住观众,博物馆要考虑其他的解决办法,例如:尽量少用展板,将展板加以改进,加进一些器物或简单的交互性展品;或将展板信息嵌入到展馆的计算机交互类展品中。

• 观众喜欢伦理问题展品,要提高这种展品的吸引力,可以考虑加入一些交互性的内容以及考虑使用印刷文本之外的其他传播媒体。

• 对于伦理问题展品,策展人员进行概念化处理时,应以展示内容引导展品的最终设计。例如,对于很私人化的展览内容,应该安排私用空间进行展示;对于要一家人讨论的展览内容,应该安排在适合交流的场所进行展示。如果小孩也是展品目标观众的一部分,就应在展品中加入交互性的内容。

• 由于观众在参观展厅的过程中只在少数几项展品前停留,故有必要让展厅的主题在它的各项展品上得到有效、明显的体现。

• 对所有的计算机和机械交互类展品进行修正性评估,以确保展品的操作功能正常以及观众能够学习。

• 在通信展厅和创新展厅,应分别划出一独立的地方作展厅简介区。各展厅的简介区要介绍有关该展厅的总体情况,使观众了解展厅展区的物理布局和主题。由于观众是按展厅简介的字义理解内容,因此展厅的主题必须明显,不但要在展厅简介区阐明,

而且要在整个展厅反复提到。在简介区使用交互性展品会提高其对观众的吸引力和保持力,但这些交互性展品必须经过测试,以确保它们能够将预定的主题传达给观众。

- 生命技术展厅的视频简介区和“变换”照相室对观众很有吸引力。然而,观众并没有将它们当做展厅的简介区,他们也没有搞清该展厅的主题。可以考虑给视频加上字幕,以说明视频影像与展厅主题的关系。“变换”照相室是一个封闭区,可以考虑增加一个音响装置,以帮助观众了解他们在照相室中所看到的事情与展厅主题的联系。

- 观众可以感知展览的真实性,并且欣赏真实性。重新设计所有的展览环境细节,使观众犹如身临其境。创新展厅的超净室在这方面做得极好,从灯光到地板到实物,都给人真实的感觉。但医学技术展项需要改进,如果该展览展示的是真正的医学器械,场地的光线能明显区别于展厅的其他展区,“医学技术”展览给观众的感觉会更像是一个真正的手术室。

- 评估展厅展示内容的设计。例如,观众在“好奇心计数器”展项处向工作人员提问,这类活动为观众提供了学习机会。应培训工作人员简练、准确地表达展厅信息,以加强展厅的目标。

- 提高配备有工作人员的展项的数量。观众喜欢与工作人员互动,因此,配备工作人员的展项是一个表达内容的理想时机。例如,在“毁坏的建筑装置”展项处,操作展项的工作人员有时间解释将要模拟哪种地震,怎样将这样运动与真实的地震相比较,里氏等级计量意味着什么。

- 那些受欢迎的展项和接近进口区的展项的利用率最高,它们代表了观众的主要体验。要确保这些展项的正常操作。

- 鼓励更好地利用工作台及其他有着专业外观的机械交互式的展项,考虑在展厅中配备更多的工作人员,以吸引观众,并为他们提供帮助。

附件五 美国创新技术博物馆 评估报告(2000)

评估机构 :兰迪科恩公司

概要

本报告是我们对美国创新技术博物馆所做的第二阶段评估。评估结果得自于调查人员在博物馆出口处的开放式深度访谈、标准化的问卷调查和参观后的电话访谈。我们在 1999 年 7 月对观众进行了出口处访谈,在 1999 年 7 月、8 月和 11 月发放了标准化的调查问卷,并在观众参观后的 3~6 个月内对其进行了电话访谈。本次评估试图证明该博物馆取得了多大的影响和效果,评估的目标是要测定:

- 观众参观博物馆的总体体验,包括最积极的和最消极的体验;
- 观众参观博物馆所获得的认知体验;
- 观众对博物馆展品、展厅环境和服务的意见;
- 参观博物馆是否改变了观众对技术的态度;
- 观众对创新技术博物馆的使命的了解;
- 观众的人口统计学特征(如性别、年龄、受教育程度、居住地);
- 观众的参观特征(如参观频次,参观了哪些展厅);
- 在 7 月(夏季)和 11 月(秋季)参观人数的差别。

主要调查结果 :访谈调查的结果

我们总共进行了 50 次访谈 ,其中与男性观众访谈了 30 次 ,与女性观众访谈了 20 次。受访观众的平均年龄为 43 岁。

对展厅的描述

很多观众的参观体验是积极的 ,他们使用了诸如“有趣的”、“令人愉快的”、“有意思的”、“入迷的”和“有娱乐性的”等词语来描述他们参观展厅的体验。其他一些人称展厅很有教育性和知识性 ,还有个别人赞扬“展览的说明搞得不错”。在负面的评价中 ,一些观众觉得有些难以应对 ,一些观众发现有些展示说明对于外行来说过于复杂难懂 ,一些观众对看到有故障的展品感到失望。

约一半的受访者并没有将展厅视为独特的环境。其他观众则认为 ,展品的种类或展厅的主题使他们感到展厅是独特的环境。

使命陈述

约一半的观众说创新技术博物馆的使命是对公众尤其是对年轻人进行技术教育。五分之一的观众说该博物馆的使命是提高公众对技术与人们的日常生活的紧密联系的认识 ,使技术易于接近和理解。一些观众则认为 ,创新技术博物馆的使命是宣传和颂扬技术和硅谷 ,一些观众说创新技术博物馆旨在展示人类的成就 ,并为公众提供亲身体验技术的动手机会。

对技术的态度

一半以上的受访者称 ,参观并没有影响他们原有的对技术的想法和感觉。一些观众指出 ,尽管他们对技术的看法和态度没有因参观带来变化 ,但创新技术博物馆的展览确实为他们提供了新的信息 ,促使他们与同伴讨论了一些技术概念。一些观众说 ,展览增进了他们对技术的了解 ,展览展示技术是如何融入他们的日

常生活的,使他们懂得技术要比他们以前认为的易于使用。在回答有关展览是否改善了他们对技术的想法时,大多数受访者显得很难于回答。

参观中最有趣的部分

当我们询问观众参观中最有趣的是什么时,他们提到了很多不同的展项,这些展项涉及博物馆所有的展厅。约五分之一的受访者称,创新展厅或生命技术展厅中的某项展品最令他们感兴趣。多位受访者则提到其他的展品和观念,包括:看IMAX电影;了解光纤、芯片制作和计算机;认识到技术的变革非常快;看孩子们在工作台做实验;体验数字演播室和 underwater 勘探。

创新技术博物馆应增加的展品

针对创新技术博物馆应增加什么样的展品,受访观众提到最多的是,要为观众提供更多的“身体参与型”活动机会,诸如工作台活动,被观众最多地提到,其次是健康主题的展品。一些观众希望等待使用展品的排队更短一些,并提出了解决这一问题的建议。其他一些观众希望有更多的图形和动画展品。观众提到的(尽管提到的次数非常少)其他应增加的展品还有:内容更浅显易懂的展品、在成人展品旁设置儿童展品、仿真类展品以及可供一组观众而不是单名观众使用的展品。

创新技术博物馆应减少的展品

多达四分之三的受访观众认为,创新技术博物馆不应减少任何一种展品。有两个对芯片制作非常了解的受访者认为,展览中对晶片和结晶块的介绍有些过时,其他受访者对蜂窝电话展项也提出了同样的意见。一些观众希望减少那些他们无法触摸的展品,一些观众甚至提出了非常专断的意见,如:把那些没有人使用的展品撤掉,提高某些软件的运行速度,修复闪烁的显示屏。

参观中令观众不愉快的事

当被问到这一问题时,约一半的受访者认为,他们在参观创新技术博物馆过程中并没有觉得有什么不愉快的事。数名观众提到,博物馆营地活动的孩子让人烦心,还有几名观众提到了下面一些不愉快的事情:损坏的展品、一些人使用交互式展品的时间过长、展品说明差以及停车的地方离博物馆较远而停车费却很高。三分之二的观众说他们看到了告知展品损坏的牌子,但其中的很多人说他们并没有因为展品损坏而影响心情。

如何改善观众的参观体验

几乎一半的观众想不出创新技术博物馆还能做哪些改进,以改善他们下次的参观体验。其余的观众则提出几方面的问题,认为如果加以改进的话,将会很好地改善他们下次参观的质量。观众们提出的问题建议包括更换展品、调整门票价格、增加停车位的选择性以及加强展品维护等。

标准化问卷的主要调查结果

我们分别在1999年7月、8月(夏季)和11月(秋季)对总计1102名成人观众进行了问卷调查。

观众的基本情况

- 参观创新技术博物馆的男性观众略多于女性,分别为53%和47%。
- 总的说来,观众的年龄数据呈钟形曲线,即大多数观众年龄为35~44岁(29%)。25~34岁的观众和45~54岁的观众各占约五分之一。
- 参观创新技术博物馆的观众大多受过良好的教育。70%的观众是大学毕业生。将近三分之一的观众拥有研究生学位(32%)。

- 大多数观众是白人(71%)。不过,与其他博物馆相比,创新技术博物馆吸引了相对较高比例的亚裔(17%)和拉美裔(7%)观众。
- 大多数观众家中没有16岁以下的孩子(57%)。在那些家中有孩子的观众中,一半以上的人家的孩子的年龄在6~12岁之间(62%)。
- 不到五分之一的观众是创新技术博物馆的会员。

观众的参观行为

- 大多数观众是第一次参观创新技术博物馆(80%)。在回头客中,近三分之二的观众自创新技术博物馆1998年秋季开放以来已参观过一次或两次(63%)。三分之一的人参观过3次或更多次。
- 女性观众、25~44岁的观众以及家中有不足16岁孩子的观众是回头客的可能性较大。
- 大多数观众参观了所有展厅,而创新展厅吸引的观众最多(85%),其次是生命技术展厅(80%)。
- 探索展厅被观众选为最受欢迎的展厅,它获得了30%的观众支持率。观众对通信展厅的支持率是21%,在四个常设展厅当中排在最后。
- 将近一半的观众与博物馆工作人员之间有过一两次互动(45%),将近三分之一的观众没有与工作人员发生任何互动(31%)。

展厅环境评分

	平均值
无吸引力(1)/有吸引力(7)	5.87
感到吓人(1)/感到亲近宜人(7)	5.61
供观看和动手的东西太多(1)/供观看和动手的展品不够(7)	3.98

展品评分

	平均值
无教育性(1)/有教育性(7)	5.99
无信息(1)/富于信息(7)	5.79
无趣味性(1)/有趣味性(7)	5.71
乏味的(1)/娱乐的(7)	5.50
难用(1)/易用(7)	5.42
难懂(1)/易懂(7)	5.29

展厅评分——与展品有关的问题

	平均值
很差(1)/优秀(7)	
展品的安全性	6.29
展厅工作人员的礼仪	6.09
展厅的总体体验	5.61
展品的维护	5.23
(观众需要帮助时)展厅工作人员的可得性	5.05
观众使用展品的可得性	4.98
展厅休息座位的提供	4.75

- 我们通过计算相关系数,来确定哪些展厅展品问项(休息座位的提供、工作人员的可得性、工作人员的礼节、展品的维护、展品的安全性和观众使用展品的可得性)与观众的总体参观体验评分有关。有趣的是,所有的六个问项的计算结果都具有统计上的显著意义,表明它们都与观众的总体参观体验评分有关。在这六个问项中,四个问项有特别明显的关联,它们是工作人员的可得性、工作人员的礼仪、展品的维护以及观众使用展品的可得性。
- 大多数观众“喜欢观看其他人使用展品”(61%)。
- 三分之一的观众(35%)认为等候参观某些展品的排队过

长,决定不等着看了,而四分之一观众说他们不介意排队,因为要看的展品很特别(25%)。

- 四分之一的观众在选择描述他们的体验的选项中,还选择了“我失望,因为有损坏的展品”的选项。
- 观众被问到“你今天遇到过损坏的展品吗?”时,近四分之三(71%)的观众回答说他们确实遇到过损坏的展品。那些遇到过损坏的展品的观众随后被要求从所提供的四个备选答案中,选择一个答案。将近三分之二(62%)的观众选择了“我遇到的损坏的展品的数量并没有让我感到灰心丧气”。
- 超过四分之三(82%)的观众使用了计算机/多媒体展品。

观众对计算机/多媒体展品的评分

完全不同意(1)/完全同意(7)	平均值
展品内容有教育意义	5.61
我总是知道如何进入到下一屏	5.48
我总是知道如何获得帮助	5.25
我总是知道程序何时结束	5.16
操作说明总是清晰明了	5.05
我希望增加有关该主题的更深度的信息	4.65
我不总是知道如何重新回到上一屏	2.84
总的说来,我感觉程序不好用	2.69

观众对伦理类展品的评分

完全不同意(1)/完全同意(7)	平均值
我希望看到更多的探讨技术与社会问题的展品	5.35
我认为创新技术博物馆对伦理类问题提供了均衡的介绍	5.20
伦理类展品使我思考我以前从未考虑过的问题	3.80
在看了展品展示的各种观点后,我对有关问题的看法改变了	3.05

- 约一半的观众观看了伦理类展品(52%)。
- 90%的说西班牙语的观众($n = 101$)称,他们注意到在一些计算机/多媒体展品中有西班牙语选项。在这些观众中,约三分之一(35%)的人使用了该选项。
- 90%的观众称英语是他们的首选语言。在问及母语不是英语的观众($n = 104$)是否愿意租用使用其母语的语音向导时,四分之一的人说他们愿意租用(24%)。
- 按照分值为1(差)~7分(优秀)的量表,残障观众给创新技术博物馆的环境无障碍性打了5.8分。
- 当观众被要求写出参观中非常令其称奇的积极方面时,近五分之一的观众说他们没有发现什么特别值得称道的东西,而近五分之一的观众提到了某项具体的展品。
- 当观众被要求写出参观中非常令其不快的事情时,几乎一半(45%)的观众说他们没有感到有什么特别不好的事情。约十分之一的观众说损坏的展品是他们参观中遭遇的消极方面,约十分之一的人说拥挤令他们感觉不好。
- 当观众被要求写出通过参观学到了什么最特别的东西时,44%的观众说他们没有学到什么,其他41%的观众说他们学到了一些特别的东西。

展品如何影响了观众对技术的看法

	平均值
技术与我的生活不相干(1)/技术与我的生活有关(7)	5.99
技术是消极的(1)/技术是积极的(7)	5.97
令人难以应对(1)/令人惬意(7)	5.72
认为技术理所当然(1)/欣赏技术(7)	5.72
不安全的(1)/可靠的(7)	5.63
感到受胁迫(1)/感到受鼓舞(7)	5.29

电话访谈的主要调查结果

印象深刻的展项

大多数的受访者说出了某些令他们记忆深刻的展项。其他受访者则说不清楚,只记得展品的大概主题,或者对展品的交互性有较好的感觉。有几名受访者甚至记不得他们参观了什么东西。

为什么对这些展品记忆深刻,受访者说出了许多理由。大多数的观众说展品提供的独特的交互式体验令他们难忘。一些观众则强调说,那些引起他们特别关注的展品都是与他们的兴趣有关的展品。其他数名观众则喜欢提供了新信息的展品。有几名观众对某些展品有印象,是因为当时观众排着长队等待。

博物馆的使命

大多数受访者说,创新技术博物馆的使命是提高公众对技术的兴趣、意识和了解。有几名观众认为创新技术博物馆是展示和激励技术革新的。那些没有提到技术的受访者认为,创新技术博物馆的使命是为观众提供一种有趣的教育环境。

对创新技术博物馆参观经历的回想

在参观六个月后接受我们的电话采访时,四分之三的受访者想起了创新技术博物馆。许多人考虑再次参观博物馆,看新的IMAX电影,或再看某些展品。一些受访者曾和朋友或家人谈起过该博物馆,或回忆他们共同的参观经历,或把参观的情况告诉没有参观过创新技术博物馆的家人或朋友。几名观众回想起一些展品,因为这些展品与他们的日常生活有关。其中两个人在回忆他们在创新技术博物馆的经历时表示了失望。那些在参观后再没有想过博物馆的人认为,他们的参观对他们个人没有什么意义。

由参观创新技术博物馆而引发的行动

不到一半的受访者说,由于参观了博物馆的关系,他们后来有某些举动。其中的许多人说,上次对创新技术博物馆的参观是第一次,他们以后还会去,或者建议别人也去。有几名受访者的举动则更富于成果,和他们的学生一起利用博物馆的资源学习,或者让他们的孩子接触技术。

当被问到是否会向别人口头推荐参观创新技术博物馆时,几乎所有的受访者说,他们会建议朋友或家人参观创新技术博物馆。有五个人不想做这种口头宣传,其中只有一人强烈表示不愿这么做。

观众的其他意见

几名受访者还提了其他一些意见。他们抱怨博物馆会员部提供的服务太差、使用某些交互式展品需要排队、有损坏的展品以及博物馆环境不够宜人。两个受访者将创新技术博物馆与 Metreon 博物馆进行了比较。

讨论与建议

引言

2000年1月,兰迪科恩公司(RK&A)向创新技术博物馆提交了第一阶段的评估报告。在第一阶段的评估中,兰迪科恩公司共完成了五个报告——4个展厅各一份评估报告加上一个总报告,总报告讨论了各展厅的评估结果,并提出了相关建议(见附件四)。对创新技术博物馆第一阶段的评估试图证明该博物馆各展厅的展示影响和效果,以及确定现有展项中需要调整的要素以改进展览效果。

第一阶段的评估采用了下述数据采集方法:

- 在各展厅对观众进行观察,以确定观众参观了哪些展品,

他们在展品前驻留了多长时间；

- 在每个展厅,在观众参观后进行开放式访谈,以便将观众的行为数据人性化,并定性分析观众的体验；

- 在四个展厅中的五个展项处观察观众的参观行为并进行访谈,以便确定观众在操作和理解这些展项上是否存在问题。

本报告是我们对创新技术博物馆所做的第二阶段评估的结果。评估结果得自于我们所采用的以下 3 种不同的数据采集方法：

- 参观后的问卷调查,以评价观众参观整个博物馆的总体体验；

- 参观后的开放式访谈,目的是定性地了解观众参观整个博物馆的总体体验；

- 在观众参观博物馆 3 个月(或更长时间)之后,再对其进行电话访谈,以衡量创新技术博物馆对观众的长期影响。

第二阶段评估的具体目标是测定：

- 观众的人口统计学特征(如性别,年龄,受教育程度,居住地)；

- 观众的参观特征(如参观的频次,参观了哪些展厅)；

- 夏季和秋季参观人数的差别；

- 观众参观博物馆的总体体验,包括最积极的和最消极的体验；

- 观众参观博物馆的认知体验；

- 观众对展品、展厅环境和博物馆服务的意见；

- 博物馆对观众对技术的态度的影响；

- 观众对创新技术博物馆的使命的了解；

- 观众对他们参观创新技术博物馆的回忆程度；

- 参观创新技术博物馆对观众产生的较长期影响。

调查结果的讨论

观众参观博物馆能获得什么样的体验,取决于博物馆所提供的展教活动以及观众带到博物馆来的东西(如观众以前参观博物馆的体验、个人背景因素、观众对技术的了解和接触技术的经历、是否是带着孩子一起来的以及对创新技术博物馆的熟悉程度等)。尽管每个观众的博物馆参观体验各不相同,但观众的参观体验往往呈现某些倾向和类似之处,研究人员可从中总结出普遍性的东西。

本报告讨论了观众参观创新技术博物馆的总体体验倾向。与我们第一阶段的评估报告不同的是,本报告基本上不涉及各个展厅的信息。本次评估(第二阶段评估)的焦点是考查观众参观整个博物馆的总体体验,而不是逐个探讨观众对各个展厅或展品部件的参观体验。与第一阶段的评估相同的是,此次评估的调查数据同样包括数字性的和描述性的,而且我们尽可能呈现各数据集之间以及同一数据集内不同的数据间的关系。

观众的背景构成

创新技术博物馆的观众情况有点类似于其他博物馆。一个长期以来众所周知的事实是,一个人的教育水平决定了他参观博物馆的情况。美国的博物馆吸引了受过良好教育的人们,创新技术博物馆也是如此。70%的创新技术博物馆观众是大学毕业生,几乎三分之一的观众拥有研究生学位。兰迪科恩公司收集的其他数据表明,在所有博物馆的观众当中,科学和艺术类博物馆的观众受教育水平是最高的。来科学博物馆的男性通常略多于女性,但是RK&A对创新技术博物馆的调查表明,在夏季来创新技术博物馆参观的女性高出男性2个百分点,在秋季女性则超出男性20个百分点。兰迪科恩公司对创新技术博物馆的秋季调查数据大都是在周末收集的,这可能是造成女性观众数量偏多的原因。将创新技

术博物馆自己的营销调查与兰迪科恩公司的调查对比后发现,创新技术博物馆七月份营销调查获得的观众性别构成与兰迪科恩公司夏季所调查的观众性别数据相同,创新技术博物馆十二月的营销调查获得的观众性别构成与兰迪科恩公司秋季所调查的观众性别数据类似。

如果用图表示创新技术博物馆观众的年龄结构,它将形成一条钟形曲线,有一半的观众的年龄是在35~54岁。与其他科学博物馆的观众相比,创新技术博物馆的观众年龄更大些:其他科学博物馆的观众中超过三分之一的人年龄在35~44岁,而在创新技术博物馆的观众中,不到三分之一的人处于这个年龄段。在结伴来的参观小组的组成上,创新技术博物馆与其他的科学博物馆也不相同。在其他的科学博物馆,一半以上的观众是与家人一起来的,而在创新技术博物馆,不到一半的观众家中有不到16岁的孩子。

与美国的其他博物馆的情况相类似,创新技术博物馆的观众大多数为白种人。不过,与其他博物馆相比,创新技术博物馆吸引了相对较高比例的亚洲裔(17%)和拉美裔(7%)观众。这一结果与创新技术博物馆自己搞的多次营销调查获得的观众种族构成非常相似(惟十月份的那次调查不太一样)。创新技术博物馆十月份的营销调查表明,只有56%的观众是白种人,而创新技术博物馆其他各次的营销调查和兰迪科恩公司的调查都一致显示,约70%的观众是白种人。

创新技术博物馆想吸引多样化的观众、尤其是吸引圣约瑟的拉美裔居民的愿望是值得称道的。在加州,其他博物馆也面临类似的挑战。比如,盖地博物馆在开发专门面向拉美裔人口的项目上,投入了相当多的努力和资金,它专门为当地拉美裔居民开发的项目吸引了很多拉美裔观众。而盖地博物馆的年度观众调查表明,它的观众构成与创新技术博物馆差不多。在盖地博物馆,拉美裔观众占观众总数的5%,亚洲裔观众占9%。像创新技术博物馆

一样,盖地博物馆也力图使它的观众能代表博物馆附近的居民,即24%拉美裔人和22%的亚洲裔人/太平洋岛民。与许多博物馆相比,创新技术博物馆开头并不太好。虽然吸引多样化观众的愿望是好的和合理的,但管理人员必须明白,受过良好教育的欧洲人有参观博物馆的传统,而对迄今还没有把去博物馆作为休闲活动的其他种族的人来说,养成去博物馆的习惯可能尚需一代人或多代人的时间。一种促进观众变化的方法是,博物馆加强与某些社区组织的合作伙伴关系,这些社区组织是为博物馆想吸引的观众群体提供服务的组织。

一些博物馆开展的旨在吸引新观众的活动在博物馆界很有影响。波士顿儿童博物馆参与的起步教育(Head Start)计划即为其一(译者注:起步教育计划是美国政府针对低收入家庭及儿童提供综合性服务的计划)。波士顿儿童博物馆了解到,在参观博物馆的大人中,有一半的人还参观儿童博物馆,因此博物馆打算在低收入家长儿童中培养参观博物馆的风气。为此,波士顿儿童博物馆加入了波士顿起步教育计划,获得该计划的资助。由此以来,受起步教育计划服务的所有的孩子及其老师每学年可以参观博物馆两次,受起步教育计划服务的儿童的家长也可以免费参观博物馆。初步的调查显示,在受起步教育计划服务的低收入家庭中,高达90%的家庭参观了博物馆和博物馆组织的活动。

参观创新技术博物馆的观众中,大部分是第一次参观这一新馆。考虑到在我们夏季调查时博物馆仅开放了八个月,在我们秋季调查时也只开放了一年,因此这样的统计数字并不令人感到惊奇。与其他年龄段的观众相比,24岁以下的观众第一次来创新技术博物馆的居多。来创新技术博物馆的人中,20%是回头客,这说明他们在这一新馆开馆后,已参观过3~4次了。回头客往往是25~44岁的妇女,家中有不足16岁的孩子。

创新技术博物馆自己做的营销调查与兰迪科恩公司的调查结

果相比,第一次参观创新技术博物馆的观众的比率低些,回头客的比率要高些。这个差异可能是因为数据收集方式不同造成的。对于营销调查来说,观众通常站着回答调查员的提问,调查过程较快。而兰迪科恩公司做调查时,由于所问的一系列问题需要观众花时间来思考,因此观众通常是坐着回答问题的,那些回头客可能不想花费较长的时间接受调查。此外,回头客还可能由于以前参观该博物馆时已经接受过营销调查,因此他们可能不愿意再接受兰迪科恩公司的调查。

展品环境

设计人员花费大量精力概念化博物馆各个展厅的背景或者环境。观众注意到博物馆的美学细节了吗?各个展厅的环境特征是否有助于观众对各展厅的体验?观众在各个展厅的体验是否是独特且互不相同的?对设计者和那些管理设计者预算的人来说,这些是很重要的问题。

可以回想一下,我们在观众参观完一个展厅之后所做的出口处访谈(参见第一阶段的评估报告)曾显示,一些观众记住了该展厅的独特的环境特征,例如光线,超净室和地面装饰。然而,在他们参观完博物馆,即在参观了二至四个展厅之后,他们似乎记不清各个展厅的独特的环境特征了。与展厅出口处的受访者一样,博物馆出口处的受访者同样认为,是展厅的展示内容构成了各个展厅彼此不同的品质。对于观众来说,内容使得各个展厅各具特色,因此,当观众走过各个展厅时,他们注意到展示的主题发生了变化,这让他们知道他们到了一个新的展区。如果观众不能辨别展厅中的独特的设计环境,这究竟意味着什么?无论如何应该探讨这个问题。

美术馆的装设与博物馆有着很强的类比性。美术馆的馆长和设计人员共同营造用来展示艺术品的展厅环境。他们仔细地选择

合适的绘画颜色、灯光照明、长椅的摆放等所有可能影响作品观赏的事情。当然,他们的目标是平衡各个方面,以使艺术品光彩夺目。如果到头来观众更多地感受到的是展厅环境而不是艺术品的话,那么馆长和设计人员的工作就是失败的。如果展厅环境过于抢眼,艺术品的重要性就会相对降低。尽管在创新技术博物馆,设计者花费了相当多的时间和资金来创造独特的展品展示环境,但是,事实上展品比环境更重要——至少就观众的体验而言应是如此。如果观众想起的是奇怪的自行车展示而不是计算机辅助设计活动,又或者观众想起的是工作台的精美而不是激光器活动,那意味着什么?展品的开发者甚至是设计者可能会以为他们在提供体验和传递概念上的努力失败了。

在对观众参观后进行的访谈中,虽然观众谈话中很少称赞展厅环境设计的好,但这并不意味着观众没有看到和不欣赏每个展厅的独特的环境。观众没有对展厅环境表达看法,只表明观众在陈述他们的博物馆体验时,环境设计要素没有进入他们意识的最中心。

间接地与观众讨论展厅环境的设计特征比较困难。为了不引导观众的思考,我们的访谈问题设计成开放性的和中立的。我们只想弄清楚,展厅的什么方面最为观众注意。令人欣慰的是,展品及其展示内容形成了观众的主要体验。我们在博物馆出口处访问的观众都说,博物馆各展厅的展示内容构成各个展厅的显著特征。正如我们在展厅出口处访问的观众所说的,是各展厅中展品的变化告诉他们正在从一个展区来到另一个展区,比如从探索展厅的地震展区来到探索展厅的水下展区。我们在博物馆出口处访问的观众也是根据展厅展示内容的变化知道他们在从一个展厅来到了另一个展厅。

在我们的问卷调查表上,有三个问项是针对展厅环境的,观众被给予三个7分量表,分别是“感到挺吓人的”(1)、“感到亲近宜

人”(7)，“无吸引力”(1) / “有吸引力”(7)，“可供参观或动手的东西太多”(1) / “可供参观和动手的东西不够”(7)。总的来看，展厅在吸引力方面得分较高(5.87分)。女性观众、年长的观众以及家中有不足16岁孩子的观众分别比男性观众、年轻的观众以及家中没有孩子的观众更能体验到展厅的吸引力。女性和年长观众的打分要高于男性和年轻的观众的打分，这种情况并不罕见。这被认为是礼节上的偏差，即为了不得罪有关的组织，回答者往往打高分。一旦女性受访者打了低分(本调查没有出现这种情况)，问题就很值得关注了。

对展厅“可供参观或动手的东西太多”(1) / “可供参观或动手的东西不够”这一问项，女性观众的打分较低，表明女性比男性易于感到压力。我们不知道女性的这种反应是针对展厅内容的，还是对展示主题表现出的不安，抑或只是对展厅的物理属性(诸如相对于展厅的面积而言展示的信息过多了)做出的反应。女性对“感到挺吓人的”和“感到亲近宜人”的打分与男性的打分并没有什么不同(平均5.61分)，这说明对展览环境女性并不比男性感觉有什么吓人，但展厅提供的活动多的话，她们会有所反应。

对展品的体验

我们在博物馆出口处调查的观众都对参观给出了积极的评价，他们使用了诸如“有趣的”、“愉快的”、“有意思的”、“令人着迷的”等字眼来描述他们的参观体验。一些人提到了展品的教育性，称“展示解说很棒”。当我们让观众说出参观中感觉最有意思的部分时，受访者提到了很多的展项，哪个展厅的都有。比如，正像我们在展厅出口处访谈所显示的那样，观众喜欢了解超净室中的芯片制作，家长愿意看他们的孩子在工作台上操作，观众喜欢看“数字头”上显示他们的3维头像，他们感觉在生命技术展厅的参观很有意思。观众也喜欢通信展厅的光纤展区和数字演播室，他们喜

欢操作水下展区中的海盗船。看 IMAX 电影也使观众感到高兴。在观众参观六个月后我们又进行了电话回访,结果观众又提起了他们在博物馆出口处访谈时提到的某些展品。

调查显示,一半以上的观众参观了所有四个常设展厅,约五分之一的观众参观了两至三个展厅,各个展厅都有四分之三以上的观众光顾。不过,女性观众比男性参观的展厅要少。

本次调查是要了解观众的参观体验。我们询问观众关于展品和展厅的一些非常具体的问题,采用的是语义差异法(反义词在 7 分量表的两端),让观众按照“很差”(1) / “优秀”(7) 打分。我们的调查获得了一些重要的调查结果,我们将之归纳成以下若干个对创新技术博物馆具有重要价值的题目,我们先从博物馆学习谈起。

博物馆学习(Learning)

当我们让受调查的观众按照一系列语义差异量表对博物馆展品进行打分时,打分最高的两个调查项分别是“无教育性”(1) / “有教育性”(7),以及“无知识性”(1) / “有知识性”(7)。观众认为展品有很好的教育性(按 7 分量表为 5.99 分)。有趣但并不令人奇怪的是,女性观众、受教育程度低的观众以及回头客对展品的教育性的打分要高于男性观众、受过良好教育的观众以及第一次来博物馆的观众。女性以及受教育低的观众对展品的知识性评分也高于男性观众和受过良好教育的观众。与家中没有孩子的观众相比,家中有不足 16 岁孩子的观众对展品的知识性的评分也更高些。

同样,与男性观众、家中没有孩子的观众以及第一次来博物馆的观众相比,女性观众、家中有不足 16 岁孩子的观众以及回头客更认为展品有趣味性和娱乐性。普遍认为,兴趣和好玩能刺激人们的学习,这样的话,那些觉得活动或创意有意思的观众很可能比那些觉得活动没有什么意思的观众更能获得有教育意义的参观体验。

在博物馆出口处的访谈以及后来的电话访谈中,我们就一些

开放式问题向观众进行了调查,以判定观众是否领会了展品的教育意义。一个问题是让观众谈一谈创新技术博物馆的使命是什么。在两种访谈中,大多数受访者说出了创新技术博物馆的使命是向公众进行技术教育,或者提高公众对技术是如何融入人们的日常生活的了解。毫无疑问,这是一个好结果。

博物馆出口处访谈的另一个问题是请观众说出他们感觉参观中最有趣的部分。一些家长认为,看他们的孩子学习是参观中最有趣的部分。这些家长显然把“学习”认定为是阅读说明、按说明操作展品和进行步进式活动。因此,看到学习行为发生很重要,而观众在访谈中并没有提及展品传递的学习内容。也许这些家长意识到“学习”是终生的事情,他们的孩子最终会整合其参观创新技术博物馆的体验,并且在将来能具体地说出所学到的知识。当然,另一个可能性是,观众没有真正理解工作台活动的展示内容。创新技术博物馆认识到,需要对某些展品尤其是工作台做一些改进,以使观众能理解这些展品要表达的概念。

本次调查还包括一个开放式问题:“你今天和创新技术博物馆有什么最意想不到的学习收获?”不到一半的观众写道,他们想不起来任何具体的事情,或者干脆简单地回答说“没有什么”。另有41%的观众表示,他们已经知道了某些事情(如超声波、生物工程食品、外科手术、地震、光纤)。没有一个回答是针对“学习”本身的,但观众的答复表明他们对某个展示主题的好奇心和兴趣增强了。事实上,观众的回答过于支离破碎,以至于无法按照展厅的目标去考察。创新技术博物馆的观众对“学习”问题的这种回答结果并不为奇。更多时候,观众不能描述他们所学到的东西,因为以一种具体的方式描述一个人所学到的东西是一项困难的事,它可能是也可能不是一次博物馆参观的结果。不过,某些电话受访者记得一些展品的具体内容,特别是与他们的职业相关的展品(例如计算机硬件/软件专业人员可能记住了微芯片的历史),或者是与

个人生活有关的展品(例如遗传疾病患者可能回想起 DNA 展品)。

测量学习是一项困难的事,特别是测量非正规场所发生的学习更为困难,在非正规教育环境中,学习是在社会体验的情境中发生的。甚至在正规教育中,对如何才能可靠地测量学生在课堂中的学习效果,研究人员仍面临挑战,因为学习研究还是一门发展中的科学。观众不能明白地说出他们学到的新概念,这不足为奇。博物馆大多数的展品没有简练的故事情节。对于观众来说,某些展品的展示概念有时太过宽泛,以至于观众无法理解最相关的信息。也许最主要的原因是,创新技术博物馆的工作人员还没有真正明白什么是“博物馆学习”。如果要测量观众在博物馆的学习,就必须很好地了解观众在看到展品时想知道什么。来自展厅出口处的访谈数据表明,生命技术展厅是观众发生学习最多的展厅,因为我们把观众的谈论与各展厅的目标相对照,发现四个展厅中生命技术展厅的目标与观众的谈论最为接近。

一些博物馆评估人员对他们评估的每件展品都打出了学习的旗帜,但在考查过相关的调查数据之后,人们不禁纳闷:他们是如何得出“有学习结果”的结论的。例如,福尔克和迪克给出的结论说:“观众的这些回忆证明,他们获得了广泛的多样化的学习”。他们提供的支持该结论的调查数据包括观众的如下表述:“我记得我们看了昆虫,我们看了毛虫,在回家的路上天下起冰雹,天降大雨,我们所有人被淋湿了。”还有“当我从动物园回到家中,我把我拿回的树籽种在我家的后院,它长成了树,我每天都可以看见它。”显然,人们有理由怀疑,观众的这种表述是否能证明他们进行了“博物馆学习”,或者仅仅是些难忘的体验。“博物馆学习”是一个崇高的目标,是一个几乎所有的博物馆都力图实现的目标。创造一种支持学习的环境不是一件容易的事,很多博物馆至今还没能做到。由于大多数博物馆目前仍在考虑如何创造最佳的学习环境的问

题,研究人员也在研究如何测量非正规场所的学习的同时,力图帮助博物馆定义什么是“博物馆学习”,所以,创新技术博物馆的观众学习效果评价应被视为一项进行中的工作。

在我们设计本次评估计划时,我们知道,要了解观众的“博物馆学习”情况,标准化的问卷调查可能不是一个好方法。但考虑到我们的展厅出口处访谈提纲中已相当多地涉及观众学习的问题,我们觉得在问卷调查表中设立观众学习的问题没有什么不妥。博物馆人员也认为,展厅的教育效果是一个重要的问题,他们在审定展厅宗旨和设计展品时都关注这个问题,这点值得称道。

早先提到过,回头客给展品教育性、知识性和趣味性的打分往往较高。按照“差”/“优秀”量表中,回头客对其参观体验的打分也高于首次来访的观众。如果要测量学习,也许应该以回头客为测试对象。回头客再次光顾博物馆,是因为博物馆有好的展品吸引他们,他们是有动机的,而动机与学习有很强的联系。随着参观次数的增加,学习将成为观众更实在的目标。博物馆人员应不断改进各项展品,力求使观众能按策展人员预想的方式操作展品,并能较为准确地感知展品内含的信息。

伦理类展品

我们编制的两个访谈大纲(博物馆出口处和展厅出口处)都没有设立伦理类展品的问题,因为在我们对创新技术博物馆的第一阶段评估中已对这类问题进行了相当多的调查。然而,在本次评估的书面问卷调查中有两个与伦理类展品有关的问题。在调查问卷中,我们给出了博物馆各项伦理类展品的名称,以便回答者明白调查中所指的“伦理类展品”是什么。约一半的观众称,他们参观过其中的一项或多项伦理类展品。没有孩子的观众比有孩子的观众更多地参观伦理类展品,这点并不奇怪,因为伦理类展品是面板展示的,而面板是一种不吸引孩子的展示方式。

问卷中也提供了一系列陈述,让观众按照“完全不同意”(1)/

“完全同意”(7)的量表打分。对于“我想看到更多的探讨技术与社会问题的展品”这一陈述,观众平均打分 5.35,对于“我认为创新技术博物馆对伦理类问题的各种观点做了均衡的介绍”这一陈述,观众平均打分 5.20。展厅出口处访谈的结果显示,观众对伦理类展品感兴趣,而且许多观众(但不是所有的观众)认为展品对各种观点的介绍是均衡的。如同在第一阶段的评估报告中所述,如果观众看到他们的个人观点没有被介绍,他们就会认为展品的介绍是不均衡的。展厅出口处访谈所显示的并在本次调查中得到证实的事实是,伦理类展品不一定会让观众思考新问题,它们也不一定会让观众改变他们对问题的看法。该问卷调查中旨在测量观众这方面想法的两个陈述是:“伦理类展品使我思考我以前从未考虑过的问题”和“在看了不同的观点之后,伦理类展品使我改变了对某一问题的想法”。按照“完全不同意”/“完全同意”的 7 分量表,这两项测题分别得了 3.80 分和 3.04 分。与该问卷调查其他问题的评分相比,这两项测题的评分是相当低的。不过,伦理类展品对受正规教育程度低的观众的影响还是较大的。观众的受教育水平越低,他们越同意这些陈述:“伦理类展品使我思考我以前从未考虑过的问题”和“在看了不同的观点之后,伦理类展品使我改变了对某一问题的想法”,以及“我认为创新技术博物馆对伦理类问题的各种观点提供了均衡的介绍”。而受过四年或更多年高等教育的观众则难以认同这样的陈述。

有趣的是,回头客对伦理类展品也获得了较强的体验。与第一次来访的观众相比,回头客更加同意“伦理类展品使我思考我以前从未考虑过的问题”,“我认为创新技术博物馆对伦理类问题的各种观点提供了均衡的介绍”。第一次来访的观众可能热衷于体验那些“非看不可”的展品(例如技术飓风和单人喷气飞行器),不想花太多的时间看面板展示和思考面板上所提出的问题。伦理类展品需要时间来阅读并且思考——这两种行为在科学博物馆并不

常见。不过,如果观众以后还会参观创新技术博物馆的话,他们就很可能将注意力集中在观看那些首次参观时没注意看的展品。如同在此处及更早的调查资料中所反映的情况那样,与第一次来访的观众相比,回头客认为他们的参观更有教育意义。

改变观众对某个伦理问题的看法是一个冒进的目标。虽然观众喜欢接受挑战,但期望他们改变对十分复杂的问题的思维方式未必完全合适。观众可能觉得伦理类展品增进了他们对某个问题的认识,或者为他们提供了思考问题的另一种思路,然而并非所有的观众能够完成从增进知识到改变思想或者态度的跨跃。如果展品是体验型的或者涉及到角色扮演,观众可能会设想他们在某种特定情景下将会如何行动,因为在这种情况下他们必须做出决定。然而,展品是相当被动性的。如在第一阶段的报告所表明的那样,在为观众提供意想不到的、有教育意义的体验方面,伦理类展品还有很大的潜力。不过,展品展出前需先进行一些试验以确保发挥最佳展示效能。

观众对技术的态度

在博物馆出口处进行访谈时,我们让观众谈一谈,与他们参观创新技术博物馆之前相比,展览是否促使他们改变了对技术的看法或感受。对于观众来说,这是一个很难回答的问题,一些观众甚至没有回答。要回答这个问题,观众必须知道他们对技术抱有自己的看法。如果他们对技术有看法,那么那些看法必须在访谈的那个时刻及时地呈现于他们的意识之中。一半以上的受访者说,展览没有促使他们改变对技术的看法或感受,但他们确实从展览中获得了有关技术的新信息。另外几名受访者说,展览大大增进了他们对技术的欣赏。还有几个受访者说,展览使他们认识到技术是如何融入了人们的日常生活。调查使用的措辞被仔细地琢磨过,以使应答者能按照六个7分语义差异量表给出展览促使他们形成的对技术的看法。在“技术与我的生活毫不相干”(1)“技术

与我的生活有关”(7)量表中,观众的打分最高,平均值为5.99。其次是“技术是消极的”(1) / “技术是积极的”(7)量表,平均得分5.97。随后的两个量表“技术使我感到难以应对”(1) / “技术使我感觉惬意”(7)和“认为技术理所当然”(1) / “欣赏技术”(7)的得分均为5.72。“技术是不安全的”(1) / “技术是可靠的”(7)量表的得分为5.63,而“感到受胁迫”(1) / “感到受激励”(7)的得分排在最后,为5.29。

与男性相比,展览令较多的女性认为技术难以应对和不可靠,但是展览也让女性较男性更加欣赏技术。女性对技术感到难以应对和不可靠,这点可以部分地解释以下现象:当我们要求观众描述其在展厅参观时的体验时,女性比男性观众更多地选择“我想要有一些工作人员在旁边,帮助我使用展品”的陈述。虽然女性在参观中可能确实需要帮助,但造成她们与男性打分差异的另一个原因可能是:与男性相比,女性更倾向于认为她们对任何特定主题的知识有限。

这些数据本身实际上并没有显示出观众对技术的态度变化,它们所表明的是,通过参观,观众对技术有了积极的感受,并认识到技术与他们的生活有关。通过看展览,观众觉得技术令人惬意,并且欣赏技术。本调研的目的不是要测定观众在参观创新技术博物馆之前对技术的看法和态度,因此,如果工作人员想要做一个准确的“进口处”量测,他们可以在观众参观开始时对观众进行调查,让他们回答前述的问题量表。通过比较观众参观前后评分的差别,判定观众对技术的“态度变化”。

工作人员、损坏的展品以及场馆拥挤情况

我们的调查中有一个简单的问题,即请受访者说出在参观时,他们与工作人员有多少互动。几乎一半的观众说,他们与工作人员之间有一两次互动,四分之一的观众说他们与工作人员有三次以上的互动。如前所述,大约一半的观众参观了所有的四个展厅,

有五分之一的观众参观了两三个展厅。因此,我们不应将观众所说的与工作人员的互动次数与我们跟踪观察所获得数据进行比较,因为我们对观众的跟踪观察是在各单独的展厅进行的,所得到的观众与工作人员的互动次数只说明观众在某一展厅参观时的情况。有关每个被跟踪的观众的经历,还有太多的东西是我们不知道的,包括观众当天还参观了其他的哪些展厅。另外,别忘了,在一个较大的观众群体中,某个观众也许已经看到群体中别的观众与工作人员发生了互动,但是我们的调查只是询问受调查者本人是否与工作人员发生过互动。我们知道,场馆工作人员感到他们整天都在与观众进行互动,而调查显示的观众与工作人员的互动却较少。设想一下,在观众参观某一展厅的15分钟的时间内(15分钟是观众参观4个展厅的中值时间的平均值),工作人员要与这一展厅中的一半的观众进行交谈,这在实践中显然是不可能的。

我们的调查还让受访者按“差”(1)、“优秀”(7)的评分量表对涉及工作人员的两个问题项打分,这两个问题是“展厅中的工作人员的礼仪”以及“观众需要帮助时展厅工作人员的可得性”。工作人员的礼仪得到了很高的总体评价,得分为6.09。“观众需要帮助时展厅工作人员的可得性”的得分较低,为5.05。受教育程度低的观众对该问题项的评分要高于受教育高的观众,表明受教育高的观众对于展厅工作人员有着更高的期望。调查中,我们还请观众从我们提供的八个选择句中选择描述其展厅体验的语句。关于工作人员配备的语句是:“我想身边有展厅工作人员帮助我使用展品”。少量的观众(15%)选择了后面的描述他们对展厅看法的语句。不过,女性观众较多地选择了关于工作人员配备的语句,这一点值得关注,因为其他调查数据显示,女性对问题项的评分通常高于男性。尽管我们不很清楚为什么女性想要更多的工作人员来帮助她们,但我们猜测,她们可能认为工作人员的存在有助于她们建立对展示内容理解上的信心。女性观众可能还希望她们的孩子能得

到有教育意义的体验,如果她们自己搞不懂展示内容,她们会认为工作人员能帮助她们的孩子获得丰富的体验。

调查中有些问项涉及展品的维护问题,具体地说,就是损坏的展品。在博物馆出口的访谈中,观众被问到“参观中有什么令你不愉快的事?”。如果观众没有提及场馆拥挤或者损坏的展品,我们会使用一些特殊的方式鼓励观众无顾虑地表达他们的负面感受。总的说来,约一半的受访者感到参观中没有什么令人不愉快的事。几名观众提到他们碰到了损坏的展品,而其他的观众在被要求具体说一下损坏的展品的情况时,三分之二的观众说他们看见了标示展品损坏的牌子。有趣的是,这些受访者中很多人说他们没有受到展品损坏的影响。

调查结果显示了损坏展品情况的类似反馈。这一调查同样允许观众从五个可能的答案中选择一个解决展品损坏问题的方法,其中一句陈述是:“我碰到的那些损坏的展品没有让我不快”。碰到损坏的展品的观众中,三分之二的人选择了此语句。另外,在描述观众展厅参观体验的八个语句中,有一个语句是关于损坏的展品的,即“我感到失望,因为有损坏的展品”。四分之一的观众选择了此语句来描述他们的参观感受。受教育程度高的观众和第一次来访的观众比那些受教育水平低的观众和回头客更多地选择了此陈述。

我们还让观众对展品的维护情况按(“差”/“优秀”)的7分量表打分,他们给出的平均分为5.23。受教育程度高的观众和首次来访的观众比那些受教育水平低的观众和回头客打分要低,这一结果与前面的调查结果吻合。它说明了一种趋势,即受过四年以上高等教育的观众和首次来访的观众更难以满足。

我们认为,创新技术博物馆损坏的展品的数目没有让较多的观众不愉快,这一点实属幸运。很多观众在工作 and 在家中遇到过计算机故障,他们明白技术有时会出问题,所以他们对展品损坏表

示理解。其他观众则简单地认为,一个收门票的博物馆不应该有出故障的展品。在科学中心进行调查时,我们时常听到观众抱怨附近城市的另一家科学中心,其抱怨多集中在他们在其他科学中心参观时遇到了很多损坏的展品。如果人们在参观一家科学中心时,谈论另一家科学中心的展品损坏问题,这说明礼节的考虑是影响观众自由表达意见和看法的一个重要因素。更值得关注的是,抱怨的观众中较多地是首次来访的观众和受过四年以上高等教育的观众,而到创新技术博物馆参观的观众以这种人居多。

拥挤是另一个问题。博物馆出口处的一些受访观众称,他们已被事先提醒博物馆的观众比较多,因此拥挤在预料之中;其他人则没有想到在博物馆会有这么多的观众。我们在书面调查中测度了观众对拥挤的感受。在描述观众展厅参观体验的八个选择语句中,六个语句与拥挤有关,但其中只有三个语句被较多的观众选择。近三分之二的观众选择了“我喜欢观看其他人使用展馆中的某些展品”。显然,由于观众喜欢观看其他人使用一些展品,这在某种程度上缓解了观众排队等待的不快。不过,约三分之一的观众选择了“某些展品前的排队太长,我不想在一些展品前排队等待”。观众不愿排队等待,说明在创新技术博物馆还有其他一些展品是这些观众感兴趣的。四分之一的观众表示他们“不介意排队等待,因为展品很特别”,不过,选择此语句的观众以中学文化程度的观众为多,受过四年以上高等教育的观众选此语句的较少。总的说来,只有3%的应答者选择了“我感到失望,因为我来创新技术博物馆,是想参观某一展品,但它总是被别人使用着”。一些在电话中接受采访的观众也回想起排队等待使用某些展品的经历,而且令人遗憾的是,他们对排队等待的记忆超过了对在创新技术博物馆其他体验的记忆。

最后是计算相关系数,以确定哪些观众参观体验的总评分有关。根据计算结果,与观众的总评分强烈相关的四个问题

是:工作人员的可得性、工作人员的礼仪、展品的维护以及展品的可得性。有趣的是,对所有这四个因素,创新技术博物馆都可以加以控制。

雇用更多的场馆工作人员虽然费钱,但会提高观众的满意度。创新技术博物馆在工作人员礼仪上得分尤其高,应通过为员工举办有关客户服务和展示内容解说的专业培训班,继续保持这方面的卓越水准。展品维护仍将是创新技术博物馆需要面临的挑战之一。对这一问题,可以考虑仿照其他博物馆针对所保藏的展品进行“幕后”展的方式开发展品或观众体验。例如,美国国家历史博物馆撤换了美国国旗的展示,博物馆人员没有将国旗拿到场馆外的保藏室,而是在一公共区建造了一个现代化的保藏室。现在,观众可以站在玻璃后面,观看保管人员清洗和修缮国旗。一些观众如果发现美国国旗没有在原有的场地展出时,可能会感到失望。但是当看到国旗放在一个巨大的工作台上,观众就会被国旗的巨大尺寸以及工作人员为保存好国旗所做的工作而惊奇。

很显然,创新技术博物馆应该采取一切可能的措施,减少排队等待使用展品的情况。一个合理的解决办法是,如果展品前有人排队等待,可以要求观众使用展品的时间不能超过多少分钟。举止糟糕的孩子是一个严重的问题,因为他们妨碍其他观众的参观。应该让所有的观众感觉到,博物馆工作人员很注重保护他们的最大利益。在展厅中安排足够的工作人员负责监视此类情况将有益于增进观众的总体参观体验。

挑战

创新技术博物馆有一项困难的工作——它的名称决定了它必须展示尖端技术。该博物馆位于硅谷,它同技术公司合作,向公众展示先进技术,并期望“激励每个人的创新意识”。本评估报告认为,创新技术博物馆的挑战来自于四类观众:受教育程度高的观

众、女性观众、回头客以及首次来访的观众,他们对创新技术博物馆的未来发展至关重要。

创新技术博物馆的观众中,近四分之三受过四年以上的大学教育,他们是博物馆的典型观众。然而,受过四年以上大学教育的观众对该博物馆的教育性和知识性的评分低于受过四年以下教育的观众。在“展厅工作人员的可得性”上,受教育程度高的观众打分也较低,尽管他们并没有明确表示希望工作人员帮助他们使用展品。至于展品维护,与受过较少正规教育的观众相比,受教育程度高的观众更多地选择了“我失望,因为有损坏的展品”来描述他们的体验,而且他们对展品维护的打分也较低。还有就是,伦理类展品对受过四年以上高等教育的观众的影响也比受过中等教育的观众小。

造成上述现象的原因可能是,受教育程度高的观众对参观博物馆有很高的预期。他们可能与高技术产业有某种关系,也可能经常参观其他的博物馆。总之,需要对他们多加注意和关照,因为他们对博物馆的评分表明,他们是自由发表意见的观众群。如果他们不喜欢某些东西,他们会直言不讳。他们评分的一致性说明他们是诚实的。他们比其他的观众或许有更多的知识,或许没有,但是他们自己确信比其他人知道的多。因此,展品必须是他们以前从未看到的——他们需要被震撼,需要被带到一个新高度,而且他们可能想接触到一些新思维,他们希望展厅配备工作人员,但不一定需要工作人员帮助他们使用展品。这类观众不断来参观将有益于博物馆的发展。

在某些方面,女性观众与受教育程度高的观众获得的参观体验是相反的。她们对创新技术博物馆的评价比男性观众要高,更倾向于认为创新技术博物馆有较高的教育性、知识性、趣味性等等。然而女性比男性更希望在展厅有工作人员可以帮助她们。同样,与男性相比,女性对技术缺乏信心,接受技术也较困难。女性

也比男性参观的展厅少,说明她们可能有些吃力。但是有较多的女性再次回访了创新技术博物馆。总的来说,女性不像男性那样爱抱怨。因此,她们是一个重要的观众群,创新技术博物馆应该与她们保持牢固的联系,并且重视她们的特殊需要(例如,在展厅配备工作人员提供帮助,提供展示解说等)。

与第一次来访的观众相比,回头客更多地认为创新技术博物馆有教育性、知识性和趣味性。从教育的立场来看,伦理类展览对回头客的效果也更好。第一次来访的观众很可能对损坏的展品感到失望。创新技术博物馆面临的一个挑战是让第一次来访的观众成为回头客。调查结果表明,观众初次参观时可能会感到不知所措,而且侧重于体验有很强吸引力的展品,但观众下次再来参观时则可能将注意力集中于展品的内容上。如果回头客获得的参观体验更丰富,那么创新技术博物馆就应为第一次来访的观众提供卓越的参观体验,以激发他们成为回头客。

以下是我们围绕展品、观众开发、客户服务和学习研究四个方面提出的一些建议。

展品

- 考虑在每个展厅中划出一块地方作为演示和试验区,通过观众测试新创的展品。受过四年以上大学教育的观众约占所有观众的四分之三,可以请他们就改进现有展品和开发新展品提出意见和建议。有很多事可以让观众参与。例如,可以制作一个问题导向的展览,让观众解答;可以开发一个没有解说的展品,让观众对正在发生的事情写一个解说;或者针对如何修复损坏的展品请观众出谋划策。这样一来,观众会觉得他们似乎可以为创新技术博物馆提供一些东西。开辟原型或试验区会得到观众的积极参与。

- 技术之艰难已为技术使用者所熟知。创新技术博物馆应

该在其展品中以及在对展品损坏的解释中确切地反映技术的这一现实。技术是一个过程,它目前还处于发展的初期,还不尽善尽美,单单把技术的这些问题讲出来,就是一个吸引人的故事。

观众开发

- 同那些决定不再参观创新技术博物馆的观众举行专题小组座谈会。据调查,这些观众中包括家中没有16岁以下孩子的成人。努力弄清楚他们不愿回来的原因。如果他们觉得参观没有得到预期的满足,那么创新技术博物馆就要确定能作些什么,以便为他们提供更好的服务,使他们再来博物馆。

- 召集女性观众参加专题小组座谈会,了解创新技术博物馆可以做些什么来提高她们在博物馆的舒适水平。兰迪科恩公司曾为某一科学项目举行过一次全部由女性参加的座谈会。参会女性首先谈到的就是当她们发现没有男性在场时,她们感觉非常放松。

- 由于创新技术博物馆要服务于周围的拉美人社区,故应考虑设立一项计划,为观众创造可以多次参观的机会,因为本次调研表明回头客要比首次来访的观众获得的教育体验丰富。

观众服务

- 在调查观众对博物馆的总体评价时有四大问题,其中两个问题与工作人员有关。对于首次来访的观众、女性观众甚至受教育程度高的观众,展馆现场工作人员是绝对必需的。观众对工作人员的礼仪给予了高度评价,博物馆应继续对工作人员开展观众服务方面的培训。如果工作人员在接受展示内容解说方面的培训,也应继续下去。如果工作人员没有接受解说培训,那么应开始此类培训。

- 展品维护是第三个问题——特别是对于首次来访的观众和受教育程度高的观众,这个问题更为重要。应考虑制作能直接说明具体维护问题的展品。

- 展品的可得性是让观众的总体评价紧密相关的第四个问题。当博物馆观众特别多时,要请使用展品的观众体谅其他等待使用该展品的观众。在观众使用多的展品旁放一个五分钟的计时器,请观众开始使用展品时进行计时。

- 需要特别重视首次来访的观众。应使他们在创新技术博物馆中得到一种超乎寻常的体验,这样他们才会再来。建议给首次来访的观众佩戴一种特殊的不干胶标签,以便现场的工作人员可以识别他们并给予他们特别关照,就像一些饭店向新顾客介绍它们饭店的特色菜和招牌菜一样。

学习研究

- 作为一家致力于公共教育的机构,为了证明“博物馆学习”,博物馆应对“博物馆学习”给出自己的定义,并且要研究回头客在创新技术博物馆的体验。

附件六 英国“地铁中的科学”招贴画项目 评估报告(1998)

概要

校外教育基金会(Education Extra)是英国一家支持课外活动的基金会,该基金会应邀对1998年3月在伦敦地铁展示的科学概念卡通系列招贴画进行评估。主要评估结果如下:

- 地铁乘客对这些招贴画的知晓度、兴趣和回忆非常高。值得注意的是,这一结果具有一致性,与乘客的年龄、性别或科学背景无关。

- 媒体对这些招贴画的报道总体上不错。电视和广播都进行了广泛的宣传,许多大报也进行了报道。然而仍然有些遗憾,通俗小报以及儿童出版物对这些招贴画的报道不多,一些主要赞助机构的支持没有得到广泛的承认。

- 运用卡通画和简单易懂的语言来描述科学原理,成功地吸引了方方面面的人,而又无“恩赐”之嫌。

- 不过,卡通的使用使得一些人认为这些招贴画是给小孩看的。校外教育基金会建议,将来有必要进一步阐明这些招贴画是面向广大公众的。

- 招贴画非常成功地引导人们去思考科学,这是因为:

- i)它们易于看到和理解。

- ii)没有现成的答案,每一个人都必须自己得出结论。

- iii)它们减轻了人们长时间一个人坐地铁的单调无聊。

大多数具有中等科学背景的人们认为,这些招贴画促使他们去思考,单从这一点上说,这些招贴画已经物有所值了。具有较高

科学知识水平的人为他们所看到的感到高兴,认为这是揭开科学神秘面纱并引导人们思考科学的一种尝试。

- 这些招贴画在激励人们思考科学原理方面是成功的,得到了所有群体的人的承认,尤其得到教师和家长的利用。这些招贴画的创作者显然很成功地提高了概念卡通的形象,并使人们认识到它们的价值。

- 这些招贴画成功地激发了人们此后对“地铁中的科学”网站的兴趣。该网站的点击次数在 2300 次以上。不过,给科学热线打电话询问答案的人数相对较少。这表明很多人是想把有关的科学问题和答案从网上下载下来供日后使用,而不仅仅是获知正确的答案。

- 具有较高教育背景的人显然更多地通过上网或者亲自进行科学探究来进一步弄明白招贴画中所提的问题。

“地铁中的科学”项目简述

该项目设计了四组招贴画,张贴在伦敦地铁的 4000 节车厢中,从 1998 年 3 月 2 日开始,到 4 月 6 日结束。这些招贴画设计成彩色卡通画的形式,提出了一些简单的科学问题。招贴画没有给出正确的答案,但公众可以通过打电话给科学热线(一家大众科学信息服务机构)或登录“地铁中的科学”网站或通过参观伦敦交通博物馆得到答案。

这项活动得到了许多组织的资助:英国公众理解科学委员会(COPUS),葛兰素威康公司和英国物理学会。另外,曼彻斯特城市大学、科学热线、伦敦交通博物馆、英国科促会甚至包括一名个人也给予了支持。

该项目的设计者是曼彻斯特大学的 Brenda Keogh 和 Stuart Naylor,他们设计该项目的目的是要:

- 把科学放到一个不会被忽视的场所;

- 让乘客每天在乘坐地铁的时候可以思考科学；
- 以栩栩如生的形式表现科学,使科学易于被公众所接受；
- 利用图书馆、互联网或科学热线等信息资源,促进后续的科学活动；
- 创造运用概念卡通的新机会,以达到促进公众了解科学的目的,比如可以将概念卡通用于其他交通工具上以及用于彩色科普读物上。

评估方法

“地铁中的科学”项目的评估采用以下方式进行：

- 1998年4月2日(星期四)和4日(星期六)两天在伦敦交通博物馆对30人进行了深入的面对面问卷调查。招贴画与问题答案同时张贴出来。大多数的问卷调查是在星期六进行的,主要是因为这一天能访问到平时工作、周末来博物馆参观的伦敦市民；
- 观察地铁车厢乘客对招贴画的反应。观察调查选在1998年3月25日(星期三)和3月28日(星期六)在中央线 Bethnal Green 站与 North Acton 站间的路段进行。乘坐中央线的人来自社会各行各业。星期三的调查安排在放学时间,目的是为了不仅测度成年人,而且测度孩子们的反应。观察的目的是测度在整个行程中有多少人阅读并讨论这些招贴画以及他们占车厢乘客总数的比例；
- 分析100多份网上评估表答卷；
- 分析科学热线提供的数字及打电话者所说的轶事性证据；
- 分析媒体对该活动的报道情况。

要收集公众对招贴画的看法存在一些困难。尽管选择在地铁里进行问卷调查会更好,但由于法律及实践上的原因,这样做会非常困难。因此,问卷调查被安排在贴出招贴画的伦敦交通博物馆进行。不过这样做会引起对问卷所做的答复有多大代表性的问

题,因为来伦敦交通博物馆参观的人中受过高等教育者较多,而且他们的好奇心较强。

同样,对网站问卷的答复也只是代表了那些有上网愿望并且能够上网的人们的观点。

不过,根据受调查者对三个调查问题的非常清楚而一致的回答,我们可以得出有效的结论。我们调查的三个问题是:

- 对这项招贴画活动公众有多大兴趣,知道多少?其结果是否因受调查者年龄、性别、对科学的理解水平或所属社会群体的不同而有差异?

- 招贴画采用的图案和语言能在多大程度上吸引广大观众关注科学?招贴画对具有“较高”科学背景的人有吸引力吗?

- 招贴画在多大程度上激发了人们了解、查询以及未来进一步探究的兴趣?它们产生的影响是否因受众年龄、性别、对科学的理解水平或所属社会团体的不同而有差异?

结论

公众对此项活动的了解

新闻报道

公众对这项活动了解的增加来自于媒体对该活动的广泛报道。1997年11月至1998年3月间,英国共有11家报纸刊发了新闻稿,其中包括卫报、泰晤士报、每日邮报、独立报、星期日独立报等大报。日本和德国的报纸甚至也作了报道。

在新闻报道方面,有三点值得关注:

- i) 尽管有关该活动的新闻稿也发给了小报,但大多数的报道出现在大报上。因此,受教育程度高的人以及专业人员对这些招贴画的了解可能从一开始就比较高。不过,没有一家报纸是将招贴画的报道安排在报纸的科学版上,这表明报纸编辑认为招贴画

报道会吸引广大读者的关注。

ii) 卫报和星期日独立报对招贴画活动的报道不如其他报纸的报道效果好, 因为这两家报纸是在 1997 年 11 月发布消息的, 比“地铁中的科学”招贴画项目启动早了 4 个月。

iii) 据我们所知, 没有一家儿童出版物报道过招贴画活动。

在该活动从 1998 年 3 月 2 日到 3 月 14 日进行的两周时间里, 英国电台对招贴画活动共报道了 9 次, 包括广播 1 台、广播 4 台、谈话电台以及伦敦与东南部的两个地方电台。

从电视来看, BBC1、英国独立电视台(ITV)以及 BBC24 都作了报道。东南和西北地方电视台的新闻节目中也作了报道。这些电视报道大都集中在 1998 年 3 月间, 但 BBC1 的实用秀(Really Useful Show)节目是在 1997 年 11 月对该活动进行报道的。另外, 如果儿童电视节目能报道这一活动会更好。

广播与电视报道有助于各方面的人知晓这一活动。

网站也进行了相关报道, 并被一家专业化学互联网站点——ChemWeb 在 1998 年 3 月底评为“本周精选”。这必然会提高专业人士对招贴画活动的关注。

乘地铁的人对招贴画的了解与回忆

我们在地铁中央线 Bethnal Green 站到 East Acton 站段进行了两个往返行程的调查, 在伦敦交通博物馆进行了访谈, 以分析人们对招贴画的注意程度。我们想了解招贴画的吸引力是否因受调查者年龄和性别的差异而有所不同。结果发现:

- 15% 到 20% 的乘客看过这些招贴画, 不过, 在行程中非常拥挤的一些路段, 由于环境不舒服和过度拥挤的原因, 致使该数字下降到不足 10%。

- 值得注意的是, 儿童与成人的答复和反应很相似。

- 在看过招贴画的乘客中, 约 25% 的人表现出对这些招贴画“真正感兴趣”, 他们要么在跟前看, 要么走近些以便能仔细地看这

些招贴画。

- 再一次得到印证的是,儿童与成人对这些招贴画的兴趣一样大。

在伦敦交通博物馆所做的访谈,旨在调查乘坐地铁的人对招贴画的回忆程度,结果发现:

- 每周乘坐地铁不足两次的人中,有40%的人能回想起看过这些招贴画,表明这些招贴画给人们留下了很深的印象。

- 每周乘坐地铁三次或三次以上的人中,50%的人能够回想起看过这些招贴画。

- 男性和女性记住这项活动的比率是相同的。这从打到科学热线的电话与对网站的访问可以得到证实,打热线电话和访问网站的人当中男女各占50%。

- 在想起看过“地铁中的科学”招贴画的人中,略低于50%的人的科学层次为“A”级(高中水平)或更高,略低于50%的人为“O”即GCSE级(初中水平),不足10%的人没有资格证书。值得注意的是,那些没有想起看过招贴画的人中,影响似乎差不多:50%的人的科学层次为“A”级或更高,而另50%的人为GCSE即“O”级。因此,对招贴画的回忆程度是一定的,与被访者的科学背景无关。不过,对这些数据应当慎重看待,因为它们是从具有自我选择性而且光顾博物馆的回答者中收集来的。

招贴画的设计

对招贴画的设计进行评估,目的是评价招贴画中的图案和语言是否受欢迎,以及它们多大程度上吸引了不同年龄和科学背景的人。

图案

招贴画的图案设计受到普遍广泛的好评。在不同年龄段和不同科学背景的受访者中,几乎每一个都认为:

- 明亮的色彩以及卡通画的运用使这项活动引人注目,有吸引力。

- 通过采取上述形式,这些招贴画的创作者成功地将科学以毫不乏味而且易于理解的形式表现出来。这对于那些具有较强的科学背景的人以及非科学人员来说同样重要,前者因看到科学“以令人兴奋且易于接受的形式表现出来”而感到非常高兴。

不过,受访者还提到了以下意见:

- i)关于招贴画的文本字号。许多人发现字号太小,难于阅读,特别是当乘客坐在地铁的另一侧时更难于看清。一些乘客只能看到最大的字。

- ii)少部分成人不喜欢卡通形式,觉得这些招贴画在表现形式上应该更现实一些。

语言

语言也普遍受到好评。在伦敦交通博物馆接受问卷调查的人都理解招贴画中提出的问题(只有一人例外)。年仅7岁的儿童对招贴画文字的理解也没有问题。

然而,有一种抱怨意见,认为简单语言的运用会影响招贴画中科学的正确性。例如,一位被调查者注意到,“重力”这个词被用在了“密度”上,这“正是地铁中的科学应努力去错的错误”。

总体影响

虽然招贴画的总体设计在各方面的人群中产生了积极的影响,但是出现了如下两个主要问题:

- i)招贴画目标受众定位有些混乱。卡通的运用使一些人认为目标观众是儿童。即使那些认识到招贴画对成人也有价值的人,也非常多地认为这些招贴画对孩子们最有用。不过,也有一些人认为,由曼彻斯特城市大学设计这个项目,意味着这些招贴画是针对成人的。将来,通过某种方式明确目标观众或许是一种可取的做法。

ii)少数人认为这些招贴画太孩子气,所提问题过于简单。一位受访者建议科学招贴画应分级。通过提供一两张真正有难度的招贴画,创作者或许能成功地抓住一些成人的注意力。

招贴画的效果

我们对招贴画的效果进行了调查,调查内容包括招贴画在多大程度上促使人们思考科学,招贴画是否教给他们一些新的东西,人们是否因此把科学与日常生活联系起来以及人们是否由此想弄清楚更多的东西或者按招贴画的内容去做一做。

思考科学

完成网上调查问卷的人、在伦敦交通博物馆接受访谈的人以及只具有中等科学背景(即“O”级或初中级或无资格证书)的人中,绝大多数人认为,这些招贴画之所以能促使他们思考科学是因为以下一些原因:

- 多数情况下,使用简单的语言和简单的问题会很受欢迎。成人与儿童的看法差不多,他们认为这项活动促使他们思考那些此前自己无法明确提出的一些问题,因为这些问题技术性不是太强,而且也没有强加于人的感觉。然而,许多人发现这些问题并没有像它们乍看上去那样简单,因为要想出答案会有很多潜在的变量需要考虑。这受到大多数人的欢迎。

- 许多成人认为招贴画中的问题能够提神,而且琢磨这些问题能够防止“经常乘地铁者的衰老与智力萎缩”。

- 同样值得注意的是,由于这是一项招贴画活动,没有正确与错误的压力,因此人们可能比在其他环境下更愿意去思考科学。

对招贴画上不提供答案这一做法,人们的反应不一。多数人认为不提供答案促使他们更多地思考问题,而那些带孩子的人认为,不能即刻获知答案是对孩子们耐性的考验。一些人建议,在地铁的其他地方张贴答案招贴画或者在现有招贴画的下端印制倒置

的答案会很好。不过,后一种方法应避免公布详细的答案。

具有较高科学知识(即“A”级或更高)的人对网站调查和问卷调查问题的答复几乎都集中在招贴画促使他人思考科学这一效果上(见下文)。这些人显然认为自己已经学过科学,只是需要鼓励其他人去思考科学。

学习科学

大多数具有较高科学知识水平的人认为,他们从招贴画没有学到任何新东西,他们只是把招贴画看作是教育他人的一种有用的资源。普遍的看法如:

- “它们没有给我什么帮助,但我想它们有助于我思考自己的科学问题。”

- “这些招贴画加强了基于动机的有效教学思想。”

看过答案和具有中等科学背景的人认为,这些招贴画教给他们大量有关科学的东西。这些人理解并喜欢解释中运用的简单语言,同时也认为这些问题由于要考虑大量的因素而显得有些难。

然而,没有看到招贴画答案的人对学到了多少新东西有不同的回答。一些人认为,这些招贴画是一种受欢迎的消遣,但由于没有答案,他们没有学到任何有根据的东西。不过,如上面所提到的,许多人承认这些招贴画使他们进行了思考。一些人把这种认知过程本身看作是一种正确的学习方式,其他人则不然。

科学与日常生活的联系

虽然具有不同科学知识水平的人大都认为,这些招贴画在帮助人们联系科学与日常生活方面是有用的,但认为用处非常大的人极少。这个问题的出现并非源自人们难于把招贴画中的问题与科学联系起来,而是由于许多人认为这些问题本身与他们的生活没有什么关系。

在具有较高科学知识水平的受访者中,一些人建议应当提及有关的科学原理(例如:传导、反射等),以加强科学与日常生活的

联系。不过,这或许不会太成功,因为许多年轻人以及没有科学背景的人恰恰是由于简单的语言才对这些招贴画感兴趣的。另外,如前所述,科学与日常生活的联系可以通过改变招贴画的内容而不是试图加强科学与这些招贴画的联系来更有效地实现,这一点大多数人已经认识到。

后续行动

上网看答案的需求非常高。到1998年4月16日,网站的访问者已超过了2300人。与之相比打给科学热线的电话只有84个。打电话给科学热线的人与访问网站的人的个人背景明显不同。30%的打科学热线电话的人具有高水平的科学理解能力,而50%以上的访问网站的人具有高水平的科学理解能力。不过,由于以下两个原因这些数字必须慎重对待:

i) 这些数字得自于相对较少的样本。我们只获得了18位打电话者和27位网站访问者的科学背景资料。

ii) 具有较高水平科学理解力的人比一般科学背景的人们更可能上网。

对打科学热线电话的人数(大部分能打电话)与访问网站的人数(少数人能上网)的差异,有许多可能的解释:

- 人们或许不愿意承担电话的费用,特别是在没有明确说明科学热线是本地电话号码的情况下。

- 具有较高科学背景并且掌握高级IT技术的人能较容易地找到关于这一活动的进一步的信息,因而更愿意这样做(遗憾的是,我们不能通过我们从调查问卷收集的数据中确定这一点,因为没有一个人被调查者登录过这个网站)。

- 与听答案相比,人们更愿意读答案,而且可能希望下载这些招贴画和答案以备将来之需。

最后这一点由以下事实得到证实:许多具有较高科学理解力的成人提到,他们想利用这些招贴画来提高“其他人”的知识面。

受过较高教育的有小孩者或者其工作与小孩有关系的人尤其这样认为。例如,许多教育工作者说:

“我为索思沃克的所有小学都买了一套。”

“我在我的工作中运用这些知识来培训没有科学背景的小学老师,以改进他们的科学教学。”

“作为一个科学带头人,我希望运用这些思想来激发我的学生(和同事)去更多地思考科学。”

同样,许多成人说,他们想利用招贴画中的问题教孩子,无论是简单地讨论这些问题,还是在家中进行一些实际可行的实验。至少一位网站访问者的目的不是为自己寻找答案,而是想知道怎样以一种简单的、易于理解的方式向孩子们解释问题。

这再一次肯定了许多人认为这些招贴画尤其对孩子有用的观点。不过,这次招贴画活动在促进概念卡通作为学习资源的利用方面取得了显著成功。

相比之下,具有一般科学背景的人在家里做实验或者去图书馆的可能性显然小得多。只有两位具有一般科学背景的网上回答者说他们愿意这样做或者已经这样做了,而没有一位调查问卷的回答者说他们愿意这样做。

可能的变化

有很多人建议,将来可以做其他科学主题的招贴画。具有较高科学知识水平的人最经常提到的科学主题是:

- 环境与生态问题;
- 伦理与时事的问题,如疯牛病、克隆等。

所有的人,不论科学背景和性别,都希望在将来看到涉及如下学科的招贴画:

- 天文学;
- 解剖学与生物学。

赞助机构

约一半看过这些招贴画的人记得葛兰素威康公司赞助了这一招贴画活动。然而,很少有人能想起该活动的其他赞助者的名字。这或许反映了这样一个事实:葛兰素威康在某种程度上是家喻户晓的名字,而其他组织则不然。

报纸对一些赞助机构有所表现。许多相关的报道文章登载了招贴画的示例,招贴画旁有英国公众理解科学委员会、英国物理学会和英国科学促进会的名字。然而,遗憾的是,作为样例的招贴画很多不是最终的版本,因而没有包括葛兰素威康或者伦敦交通博物馆的名字。独立报只提到葛兰素威康,对伦敦交通博物馆则根本没有提及。

建议

地铁中的科学显然非常成功地实现了许多目标。不过,为了进一步扩大这种成功,校外教育基金会提出以下十点建议:

招贴画

- 招贴画的图案深受人们喜爱。创作者应当确保将来的招贴画具有同样清晰的语言,并且一样色彩鲜艳。不过,字号应加大以便于人们阅读。

- 希望能够更加清楚地标明目标观众。这可以通过一系列分级的招贴画明确地或巧妙地实现。

- 尽管一些人认为,如果能够提供答案,效果会更好,然而对许多人来说,正是因为没有提供现成的答案,这些招贴画才特别有效。我们认为这种形式应当保留,不过应该通过以下方式使人们更容易获知答案:

- i)明确说明科学热线是本地电话号码。

- ii)确保将来的网址更简单。

- 在每张招贴画上也应加上关于本系列招贴画中其他招贴画的提示,或者加上“为什么不在家里试试这个?”之类的话,以鼓

励更多的人继续探究招贴画中的内容。

主题领域

• 创作者应当认真研究下一组招贴画的主题选择。天文学、解剖学和生物学内容的招贴画在社会各阶层都将会受欢迎,不过创作者可能也希望选择一些只受特定群体欢迎的主题。大众科学领域的选择,不仅要吸引人们的兴趣,而且要有助于加强科学与日常生活的联系。

后续学习材料

• 有明显的证据表明,教育工作者与父母对概念卡通和其他后续材料非常感兴趣。“地铁中的科学”的创作者也许希望通过进一步提升概念卡通的思想或者创作出更多的学习材料来对此作出回应。

宣传

• 大报、电视和电台等媒体的报道很出色。不过,我们认为,如果小报及儿童出版物也参与报道下一轮招贴画活动,对这一活动会非常有助益。如果使下一轮招贴画的报道时间尽可能靠近招贴画推出的时间,也会大有裨益。

• 同样,如果互联网特别是非专业网站对这一活动进行更多报道的话,将会更好。

• 必须努力确保使这一活动的所有赞助机构得到最大程度的报道。必须鼓励新闻界报道这类商业赞助,而且必须让新闻机构收到带有所有赞助机构名字的招贴画样本。

其他展示地点

• 这些招贴画成功地唤起了不同群体的人的科学兴趣和探究精神。我们认为,如果可能的话,应考虑在其他场所也张贴这些招贴画。没有理由表明,如果把它们张贴于公共汽车上、出租车里或其他类似的地方,效果会有什么削弱。