

基于要素整合的教育评估方案设计

孟 洁,张令飞,冯修猛

(上海市教育评估院高等教育评估所,上海 200031)

摘要: 文章针对评估方案设计所涉及的核心问题,梳理评估方案设计五个关键环节,关注主体、客体、指标、条件和程序等要素的整合。提出主体要素需明确评估是外部评估还是学校自我评估,明确是多元主体还是单一主体;客体要素关注评估对象的范围和类型,需明确是专项评估还是综合评估,明确院校、学科专业、课程、人员、科研、教学等评估类型;指标要素需厘清评估指标的选取、指标体系设计和优化的原则和方法;条件要素解决收集指标测量、评判检验和结果反馈所借助的技术和工具;程序要素给出具体操作程序和完成时间节点等。文章尝试给出科学合理的评估方案的特性及方案框架。

关键词: 教育评估方案;要素整合;方案设计

中图分类号:G40-058.1 文献标识码:A 文章编号:2095-3380(2020)02-0028-06

Integration of the Key Elements in the Design of the Evaluation Program

Meng Jie, Zhang Lingfei, Feng Xiumeng

(Department of High Education Evaluation, Shanghai Education Evaluation Institute, Shanghai 200031)

Abstract: According to the core issues involved in the design of evaluation program, this paper sorts out five key parts of evaluation program design, and the integration of subject, object, index, conditions and procedures. We argue that the subject should be made clear whether it is an external assessment or university self-assessment, multi-subject or single-subject. The object should be narrowed down to the scope and type of the evaluation target, identify whether it is a special evaluation or comprehensive evaluation, and its evaluation type such as colleges, disciplines, courses, faculties, research, teaching, etc. Index elements need clarification on the index selection, index systems design and optimization. The content element means the technical tools used to collect index data, conduct evaluation tests and give feedback results. Procedures should clarify the specific operation procedures and completion time nodes, etc. This paper attempts to find the characteristics and framework of a scientific and reasonable evaluation program.

Keywords: Education evaluation program; Integration of the key elements; Program design

教育评估方案是根据一定的目的和教育活动一般规律及评估活动的范式,对评估内容和对象、方法和工具、程序和组织、监督和结果使用等加以规范,做出规定的基本文件,是评估活动的先行内

容。^[1]教育评估方案设计包括方案的总体要求和目标的确立、评估指标体系和评估标准的构建、评估方法的选择和应用、评估工具的研制和评估工作的计划安排等环节,围绕这些环节设计评估方案,需

要关注主体、客体、指标、条件和程序5要素的整合。其中,主体要素主要解决由谁来评的问题;客体要素关注评估对象的范围和类型;指标要素即目标的分解,解决评估的重点和具体内容;条件要素解决用什么方法和工具来评的问题;程序要素是解决如何评的问题,5要素整合过程需要教育评估理念的指导,在评估方案设计过程中要权衡和把握好5要素的逻辑关系。

一、教育评估方案设计总体要求

1. 评估方案形成机理和核心关切

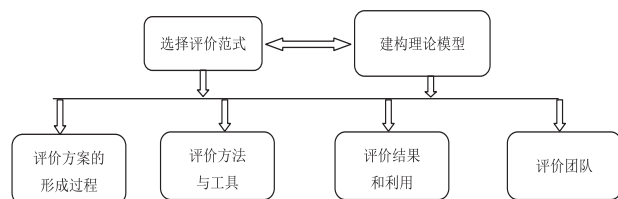


图1 评估方案设计概念图

王战军提出评估方案设计概念图,^[1]从中可以得出方案设计至少要考虑六个方面的问题:第一,评估方案达到什么样的目标?不同的评估范式中哪一种类型可说明评估方案的目的性?第二,要建构理论模型,主要指评估目标、对象性质与评估理念、模型契合性。第三,评估方案框架是什么?有哪些利益相关者参与运作?评估方案主要环节是什么?第四,评估方案采用何种方法和工具?量化评估主要涉及数据的有效采集和甄别,自行调查将会增加评估工作的各种成本,要衡量评估方案的成本和效益与效果的关系。第五,评估方案如何处理好评估结果的公布和使用问题,包括是否回应公众的问责,还是监督管理者的办学行为?第六,评估团队(组织方、专家、被评方等)的组织、专业评估和配合能力如何?团队对评估专项运作的把握程度如何?如何通过计划安排加强评估团队合作和协调?对上述这些问题的明确过程就是评估方案设计的过程,也是将主客体、指标、条件、程序等要素在评估方案设计中系统化的过程。

2. 循评估理论导向,选择适合主客体的评估理念和方法

设计评估方案需要评估理论的指导,包括选用评估范式,匹配评估模型与评估实践。要通过评估

活动获取评估对象的发展机理、特点和不足,深刻把握评估规律和经验。

纵观教育评估领域理论的发展历程,其变革基本走向是从测量时代(第一代)强调客观工具,到描述时代(第二代)的目标达成度,到判断时代(第三代)的定性定量标准鉴定,转向建构时代(第四代)重视协商与共识发展性评估。前三代的评估理论不足之处主要有三:一是管理主义倾向过重,把评估对象排除在评估活动之外;二是忽视价值的多元性,将评估者的评估意见作为评估的唯一标准和结论,忽略其他利益相关者的价值判断;三是过分依赖科学范式和数量测量,而忽视质性评估方法的运用。^[2]随着第四代评估理论、网络技术和多媒体技术的发展,以及相关利益人概念纳入评估领域,带来了平等协商、信息公开、服务学校、绩效问责、学校主体、学生本位、多元分类等理念,倡导评估工作中利益相关方享有平等权,将评估主体和客体对立关系转变为协商和共同构建的关系,改变评估客体消极配合评估主体的情况,进而转变为更加注重描述和澄清事实、注重运作和效果价值判断二方面的结合。

评估模型也在迭代,过去有目标导向模式、决策导向型(CIPP)模式、目标游离模式和应答模式等,各有适用范围,美国教育评估标准联合会也从22种模型中总结了被称为是21世纪适用的9大类最佳评估模型,^[3]有些模型在国内评估领域的应用,催生了新评估范式的成熟。如:按照被评对象状态和政策要求不同,产生了合格评估、水平评估、状态监测等范式;有按照评估强制程度和标准自定程度进行分类的质量认证、认可模式,督导、审核模式,培育、遴选模式等等,还有针对破“五唯”难题,克服定量评价和同行评价的一些缺陷而创新评价方法的尝试,如:融计量评价(定量、客观)和同行评价(专业、主观)为一体的融合评价方法也在一级学科评估和期刊论文评价方案中得到运用。^[4]

总体而言,评估活动更加关注事实的获取,而评估方案的设计也越来越注重相关利益群体的多元参与以及多学科方法融合和多样化技术投入。正如斯达福比姆(Daniel L. Stufflebeam)所说:“评估是指一项研究,其设计及实施是协助评估报告阅读

者评估任一对象的优点和价值。”^[3]

3. 凝练评估目标的价值,体现方案指导思想针对性

如果评估作为国家政策性要求或委托方的任务,政策目标、委托方要求自然成为评估的依据或目标,要凝练这些目标的价值导向并体现在指导思想中,而不同层次评估目标的实现依赖于选择的评估范式。评估方案指导思想应聚焦回答以下问题:哪一种评估范式可以解决评估活动所涉及的目标?衡量评估范式中涉及主体、方法元素是否最大限度发挥评估若干功能?指导思想所列原则对评估实践是否恰当?对上述问题认识和要素分析,直接决定了评估方案的适用性和指导思想的针对性。譬如,高等教育的三种评估范式就有显著区别(表1)。

表1 三种评估范式的比较

评估范式	合格评估	水平评估	监测评估
目的	保障教育的基本质量	测量教育的发展水平	监控教育的发展状况
功能	基于标准	基于比较的事实、价值判定	常态的事实、价值判定
方法	周期性的审核、认证	数据计量、定性分等	经济学方法、现代信息技术

合格评估是保证教育质量底线,水平评估是在评估客体之间作相对比较,以激励他们提高质量,都是对某一时点的教育发展状况通过一次性评估做出价值判断,而监测评估是掌握教育发展的状态,是连续性的评估。选择的评估范式不同,评估目的、功能和方法有所不同。

二、评估指标体系构建

评估标准和评估指标体系不仅需刻画评估对象的主要特征和属性,而且还与评估目标分解密切相关。构建指标体系应体现评估理念指导下的目标取向,评估目标不同,即使是同一评估对象,都可能构建出不同的指标体系和评估标准。

1. 厘清所选评估指标内涵、类型、权重和评价基准的方式,指标选取满足一定原则要求

指标内涵是指指标的实质含义,是评估对象的各构成要素,是对指标进行数据采集的详尽说明。而指标权重是指标对价值汇总贡献度的预判,反映对整个指标体系的重要程度。

指标类型按指标价值属性能否数量化可分为质性/量化指标;按重要程度不同可分为主要指标/次要指标;按指标属性值大小与对象总价值关系可分为效益型/成本型/适中型指标;按照质量特征可分为目标/过程/条件指标。当然,还需要关注是否存在特殊的开关型指标,比如师德师风,对最终目标贡献效果属于必备资格型,但运行时与其他指标无关,却是满足指标体系的完备性而无法剔除的指标。

指标选取要满足“SMART”原则。“S”——Specific,即指标具体而明确,指标切中评估模型,模型中每个模块都有对应的支撑指标,考察内容尽量全面。“M”——Measurable,指标最终是可衡量的、非主观笼统的描述,对变化或评估对象之间的差异敏感,可以明显区别分析对象的差异。“A”——Achievable,指标基准应该是可达到的、能实现的,切合实际不过高也不过低,兼顾发展的评估指标。“R”——Realistic,指标具有应用的现实性,符合本国或本地方的实际情况,方便分析应用。“T”——Time bound,指标内涵有一定的时限性。可根据学校教育的实际状况予以更新。^[5]

2. 构建评估指标体系要考虑指标间的层次关系和分支关系,整合成具有清晰指标内涵、权重和评估标准的有机整体

(1)坚持目标导向性要求。评估指标体系的整合要结合评估目标的层次性,突出选中的评估模型属性。比如对于选拔评估和合格评估,指标体系不能一样。选拔评估偏重于好中选优,指标体系需要对少数优秀的评估对象比较敏感,善于区分较优秀的评估对象;而在合格评估中,基于绝大部分评估对象都合格的客观事实,评估指标体系就需要对少数较差的评估对象比较敏感。指标要服务于目标。再如,评估对象比较复杂,有时需要按照分类评估模式,不能用同一套评估指标体系评估所有的评估对象。另外,指标体系中上下级指标导向要一致,体现评估目标的有效分解,分解时要考虑涵盖性和公正性。

(2)兼顾可比性和可靠性要求。检视指标体系中各项指标是否能够反映同类评估对象的共同属性,能够测定出不同对象价值以便比较,是否可以

相互配合,整体实现评估总目标,检视各项评估指标及观测点所需的资料和数据来源是否具有可获得性。数据最好来源于管理部门平时常态化积累的数据库。

(3)保持整合后独立性和简约性。指标体系层次结构尽量扁平,同一层级评估指标之间不能相互包含、重叠或者存在因果关系,评估指标不可互相矛盾。需分清主要和次要的、本质的和非本质的指标,尽量剔除一些次要指标,需要分析、归类、合并、筛选和剔除诸多存在交叉、重叠、包含、矛盾、因果关系的指标,避免过于繁琐。

3. 分阶段采取合适方法及组合,构建科学、合理、可操作的教育评估标准体系

初拟指标阶段:通过文献分析法、头脑风暴法和专家咨询,总结可借鉴的指标体系优缺点,并根据评估目标和评估对象罗列一、二级指标;做到使每一个指标外延清晰,易于界定,内涵明确,易于理解。筛选指标阶段:通过德尔菲法、层次分析法等多种组合,进一步分解分析各级指标的内涵及指标相互关系,保证指标间的独立性与整体完备性,尽可能地采用较少指标满足评估实际需要。试验修订阶段:利用主成分分析法等统计方法,提高指标的可靠程度,保证权重的合理性,每一条指标应保证能搜集到有用的信息资料;另外要通过调整和试验,验证评估模型中参照基准,保证评估指标体系的可比性。

总之,科学、合理、可操作的评估指标体系有下列特点:一是能够关注评估对象和学校教育中重要方面;二是指标内涵和类型反映教育价值与发展趋势;三是指标收集有便利性和稳定性,考虑到评估主客体、资料 and 工具、时间和流程等因素的制约;四是指标测量与评估工具和方法的选择结合得比较好。

三、评估方法选择和应用

按照评估目的和基准不同,评估方法可分为相对评估法、绝对评估法和个体差异评估;基于指标类型和判断属性,评估方法大致可分为定性评估法、定量评估法和综合评估方法(融合两者及多学科方法)。评估者只有梳理当前教育评估面临的挑

战,客观分析不同方法的优势和缺陷,选择适用和便捷可行的评估方法,一方面搜集大量的第一手观测点证据,一方面处理好数据集成和评判,才能做出相对客观、公正的评估结果。否则即使评估理念正确,价值导向正确,但因方法不当导致评估结果不能反映事实,评估目标仍会落空。

1. 定性评估方法多应用于质性指标,选择时多注意工具的应用策略

质性指标是指难以用数字来表达、又难以捉摸的表现。定性方式是用丰富的文字来描述观察者对对象和教育质量的观感,如课堂气氛、学校文化、师生关系、学生需要、家校伙伴关系等。定性评估方法主要有观察法、访谈法、各类文字、音视频材料法等,被我们广泛应用于实地考察、专家通讯评议和会议评审形式中。

以访谈法应用为例,通过与被访谈者的谈话来搜集资料,通过评估人员与被访谈者面对面直接交流的方式实现。访谈有正式访谈和非正式访谈,深度访谈和集体访谈,还可分为一次性访谈和重复性访谈等。应用策略可根据考察需要和指标内容,应具有一定的灵活性和适应性。使用时应注意:一是对于谈话形式、人数及样本选取等问题要提前思考;二是集体访谈选取不同层面对象,要多角度挑选;三是访谈前应当制定一份较为周密、可行的访谈计划,目的明确,中心议题要集中,不可随意漫谈;四是对访谈对象介绍情况时,需要具体查找相应文档资料佐证,在没有相应证明资料时,可视为一般性参考;五是被访对象需要在较宽松环境下交流,力求访谈效果。

2. 定量评估方法的选择需要借助成熟的或专门开发的评估工具做支撑

量化指标多是指能够用数字来量度的指标变项,如公开考试成绩和合格率、学生出席率和逃学率、师生比、家长参与学校活动的频率、学生在学科上的增值表现等。量化指标中又可分为数量指标和行为指标。数量指标是指能用实际数据来测量的指标,行为指标是指用来表示组织或人员某种行为强度的高低指标,只能用相对数值比较显示行为强度的高低。定量评估方法常有:考试、测试法、问卷调查法、网络系统数据监测、心理或行为量表

法等。

3. 综合评估方法的选择要考虑方法的优缺点和适用范围

已有学者归纳了一些综合评估方法的性质、优缺点及适用范围,^[6]体现了评估工作者主动、积极面对新时代教育情境、解决新问题的态度。随着评估领域不断引入各类学科新方法及其组合应用,如融合评估方法吸收了定量评估和同行评议(定性)的优点,通过数量分析,为专家提供排序或对比供参考,帮助专家综合研判,并在实践中固化成“基于客观事实的同行评价”,解决评估可信度的新思路,但好方法也不可能适合于所有项目评估。

比如智能化评估方法。通过人工神经网络的学习和训练获取知识,能够揣摩、提炼相同属性评估对象的实际价值,能充分考虑评估专家的经验,又能降低人为不确定性。虽然有应用前景,但需要有大量学习样本的评估项目才可应用。

以学位论文抽检项目为例,评估采取的是专家评估法和多元评估标准进行类别内的比较,先通过专家主观判断进行绝对打分,再随机挑选更多的专家取平均值来判定合格与否,优点是能够汇集更多专家的不同意见,能够克服指标导向的同质化倾向,相对较为公平。缺点是工作量大而任务密集,分类角度、类别定位难以达成共识,对评审系统中论文抽检(项目领域)与专家匹配度要求高。

四、评估工具研制与应用

单一的评估工具已经不能适应新时代教育评估工作的要求,随着人工智能和大数据的普及,综合运用各类评估工具、合理运用评估工具,将成为实现评估现代化的一项重要路径。应根据不同项目选取最优工具,从而满足评估要求、提升评估质量。以下为几种主要评估工具。

1. 系统平台

系统平台是基于和项目实施细则紧密契合的算法,衍生出的一种评估工具,如“上海市研究生信息化平台”“本科教学工作审核评估信息化平台”等等,这些平台的创制和应用,不仅降低了文本使用、提升了学校申报效率,而且体现了教育评估和测量的整合、集合的概念。评估主要在测量所获得的客

观信息的基础上进行,同时,教育测量的结果要通过评估才能获得实际意义。^[7]系统对采集到的信息进行“深加工”,如论文抽检成绩出来之后,在系统中是否对某一评估指标(如某个学科、某一专家导师、某所院校、某个专业)近几年的抽检成绩形成曲线图?能否横向比较发现一些规律?这些重要信息的整合和分析,是评估系统平台发展的方向和趋势。

2. APP填报和评分系统

APP是手机软件的简称,广泛应用于各类手机系统,相对于传统的专家会评方式,APP的应用会是今后的趋势,大大缩减时间成本和空间成本。专家可以随时随地打开学校报送的电子材料,进行初审、填写意见。APP工具的应用对客户端的安全性要求高,维护和更新的成本较高。在青年教师激励计划中采纳了平板端数据采集系统,但仅限于专家自行记录和上传,还远未达到数据分析、数据共享、实时更新等功能。

还可研制专家意见和评分系统的小程序,生成一套专家可以实际使用的评分表,具备填写、更改和提交功能,而评估人员只需要在后台及时跟踪专家提交情况,这种工具有效地避免了人为二次录入可能产生的错误,也提升了整个项目的保密性和安全性。

3. 问卷星系统

调查研究中用来搜集问卷的一种工具。它的形式是一份问题表格,用途在于测量人们的态度、行为等特征。问卷星简便好用,通过微信即可参与,应用时,对于制作者专业能力的考量要求较高,选取哪种问卷题目和类型?如何设计问题和取样?关键要保证制作者对评估主客体有较为深刻的了解,并能保证问卷结果有效。

五、基于流程的项目计划

项目计划管理是项目执行、监控和收尾方式的说明,它整合了所有子管理系统中有关程序、结果及达到标准的有关方面规定,以及管理项目所需的其他信息。教育评估项目计划管理主要应明确以下三个步骤中的各项要求。

1. 前期工作安排

确定项目负责人和各级项目责任主体,签订项目主管责任书以及相关责任人责任承诺书。与院校协商项目初步流程与材料提交安排,该环节应留下会议纪要和会议文件存档备案。签订委托的预算文本,该环节应注意遵循财务制度,留存协议原件备案。选择部分评估对象开展试点或对评估对象先期评估,以检验评估方案的有效性,检查发现评估方案中可能潜在的错误,对偏差修正完善。^[7]

2. 项目实施管理

工作方案沟通和定稿,该环节要充分与委托方或院校沟通协商,共同确认完成,并严格遵照方案执行。时间进度管理,各流程风险点管理及完成进度表,主要包括:学校方任务完成时间节点;组织方文件、报告提交和财务结算时间节点、评审具体环节时间计划等等。专家遴选管理,包括预选专家、综合分析专家性格、资历、专业背景和利益关系等特征,遴选符合条件的专家团队。评估现场管理,包括工作团队组成和分工(工作手册、专家手册、记录和录音等)、结果统计和监督等要求(计分方式、统计口径,计票人、监票人等)。

3. 项目结果管理

明确评估报告撰写和评估结果发布等方面要求:评估报告应当遵循的范式和结果公布的规则,专家评估报告和总体分析报告的内涵要求,包括需要达到的分析效果预期。

教育评估方案的设计过程就是对上述五大环节中所涉及的理念和原则、指标和标准、方法和工具、主体和程序进行系统化的思考,整合要素确定文本内容以期评估活动有依有据有序进行,从而保障评估工作能够公平、高效和高质量地完成。由此可见,一个科学合理的评估方案应具有:①合目的性,评估是一种有目的的社会活动,有明确政策或评估目标,在此基础上确定评估方案的原则或指导思想以便共同遵守;②合规范性,评估指标构建、方法和工具选择、程序和安排贯穿一定的标准、要求和规定,文本框架及格式一定程度上有统一性要求;③可操作性,作为对评估活动有具体指导作用的文件,必然要求方案可以实施。

总结大多数的评估方案文本框架形成,应该包

括以下方面的内容:指导思想及评估重点、评估性质和对象的说明;关于评估内容、准则和评估指标的说明,包括关于评估标准和观测点的解读和说明;关于评估信息及其采集和处理方法的说明;关于评估程序、评估方法及评估工具的说明;关于评估结果、报告及其运用方式的说明;各流程时间进度计划和事务组织管理的说明。

参考文献

- [1] 王战军.学位与研究生教育评估理论与方法[M].北京:高等教育出版社,2012:201-203.
Wang Zhanjun. Degree and Graduate Education Evaluation Theory and Methodology [M]. Beijing: Higher Education Press, 2012: 201-203.
- [2] Cuba E C, Lincoln Y S. Fourth Generation Evaluation[M]. Newbury Park: Sage Publications. 1989: 31-32.
- [3] [美]Stufflebeam D.L.评估模型[M].苏锦丽,译.北京:北京大学出版社,2007:43-96.
[US]Stufflebeam D L. Evaluation Models: Viewpoints on Educational and Human Services Evaluation [M]. Translated by Su Jinli. Beijing: Peking University Press, 2007: 43-96.
- [4] 林梦泉,任超,陈燕,等.破解教育评价难题,探索“融合评价”新方法[J].学位与研究生教育,2019(12):1-6.
Lin Mengquan, Ren Chao, Chen Yan, et al. Solve Education Evaluation Problem Using New Methodology “Integration Evaluation” [J]. Academic Degrees & Graduate Education, 2019(12): 1-6.
- [5] 顾奎华,张茂昌,洪雨露,等.小学学校教育评估指标研究[M].北京:高等教育出版社,2014:191.
Gu Kuihua, Zhang Maochang, Hong Yulu, et al. Research on Elementary School Education Evaluation Index [M]. Beijing: Higher Education Press, 2014: 191.
- [6] 冯晖.教育评估计算学[M].北京:高等教育出版社,2012:169-174.
Feng Hui. Education Evaluation Computing [M]. Beijing: Higher Education Press, 2012: 169-174.
- [7] 金梯,王刚.教育评价与测量[M].北京:教育科学出版社,2002:83.
Jin Di, Wang Gang. Educational Evaluation and Measurement [M]. Beijing: Educational Science Publishing House, 2002: 83.