

# 教师远程培训效果评估指标体系构建

## ——基于德尔菲法的研究

武丽志<sup>1</sup> 吴甜甜<sup>2</sup>

(1. 华南师范大学 网络教育学院, 广州 510631; 2. 广州市比目网络科技有限公司, 广州 510006)

**【摘要】** 随着信息技术的迅猛发展以及“国培计划”等政府项目的推动, 远程培训作为大规模教师培训的有效方式在全国范围迅速推广, 业已成为我国当前中小学教师培训的重要途径之一。由于教师远程培训具有学员数量多、影响范围广、实践发展快等特点, 对其效果进行评估研究逾显迫切和意义重大。本研究在前人研究基础上, 以企业培训效果评估理论、成人学习理论、培训转化理论等为基础, 构建了教师远程培训效果的评估指标体系初稿, 并通过两轮德尔菲法对三个级别指标项进行意见征询, 对评估指标体系的科学合理性进行验证, 继而在反馈意见统计分析的基础上进行指标删除、内涵及表述修改、补增项目等, 最终构建了包含4个一级指标、12个二级指标和46个三级指标的教师远程培训效果评估指标体系。其中, 一级指标包含学员反应层、学习结果层、个人绩效层和组织绩效层。学员反应层包括学员对网络课程、支持服务、在线项目设计与实施、培训整体满意度; 学习结果层包括专业知识、专业技能、情感态度价值观; 个人绩效层包括学员对所学知识的直接迁移和间接迁移; 组织绩效层包括学员在自我提升基础上, 对学生、同事、学校的影响等。该评估指标体系具有较高的专家效度, 以及较好的适用性和可推广性, 在一定程度上丰富了教师远程培训效果评估的理论体系, 也为教师远程培训实践的健康、可持续发展提供了有益指导。

**【关键词】** 远程培训; 效果; 评估; 德尔菲法

**【中图分类号】** G644.4

**【文献标识码】** A

**【文章编号】** 1007-2179(2014)05-0091-11

百年大计, 教育为本; 教育大计, 教师为本。目前, 教师培训是加强中小学教师队伍建设的抓手, 是全面推进素质教育, 促进基础教育均衡发展, 提高我国基础教育整体质量, 实现教育公平的重要保证。近年来, 随着信息技术的日益发展以及“国培计划”等政府项目的推动, 远程培训作为大规模教师培训方式在全国范围迅速推广, 成为当前我国中小学教师培训的重要途径之一, 与传统集中培训、校本培训合称教师培训的“三驾马车”。

实践快速发展的同时, 教师远程培训实施效果究竟如何? 应当如何科学实施评估? 尤其是如何评估远程培训对学员个人实际教学及组织绩效的影响(或提升)? 目前, 我国教师远程培训实践尚停留在对学员学习参与、学习任务完成、过程性考核、终结性考核等表层数据的统计和分析, 缺乏对学员学习

过程及培训后在实践教学中行为持久变化的跟踪分析和比较, 无法获取学员全面、真实的培训效果。教师远程培训具有学员数量多、影响范围广、实践发展快等特点, 理论研究现已明显滞后于实践发展, 因此尽快研究、制订科学的教师远程培训效果评估指标体系对于其健康可持续发展意义重大。

### 一、基于层次模型的评估指标体系框架设计

“效果”是由某种动因或原因所产生的结果。培训效果主要指公司和受训者从培训中获得的收益(肖娟, 2007)。就教师远程培训而言, 其效果指教师经过远程培训后, 个人的知识、技能、情感、态度和价值观等方面的变化, 这些变化进一步影响了教师的个人工作绩效, 并带动对学生、同事、学校等组织层面的绩效变化。

**【收稿日期】** 2014-6-29

**【修回日期】** 2014-8-27

**【基金项目】** 国家社会科学基金“十二五”规划2011年度教育学青年课题“‘研训用’一体的农村教师远程培训模式及支持服务体系研究”(CKA110163)。

**【作者简介】** 武丽志, 博士, 华南师范大学网络教育学院发展规划部部长, 副研究员, 硕士生导师, 研究方向: 远程教育、教师教育(41245622@qq.com); 吴甜甜, 广州市比目网络科技有限公司, 研究方向为远程教育、移动学习。

### (一) 培训效果评估层次模型的产生与发展

培训效果评估产生于 20 世纪 50 年代,经历了从定性评估到定量评估的发展。其中影响最大、使用最为广泛的是唐纳德·L·柯克帕特里克(Donald L. Kirkpatrick)1959 年提出的柯氏层次评估模型,该模型将评估标准分为反应层、学习层、行为层、结果层(Kirkpatrick,1998)。其中,反应层关注学员对培训的直接感受;学习层关注学员受训后对知识、技能、态度等培训相关内容的理解和掌握程度;行为层关注学员受训后多大程度上改变了自己的行为,多大程度上应用了培训所学的知识、技能或技巧;效果层关注培训给组织带来了多大成效和变化。1974 年,威廉·H·汉姆布林(William H. Hamblin)在此基础上提出了“五层级评估模型”,增加了成本效益分析和培训对企业战略目标的影响(莱斯利,2003)。1991 年,杰克·菲利普斯(Jack Philips)改进了四层次模型,增加了投资回报率(Philips,1995)。之后,罗杰·考夫曼(Roger Kaufman)等也提出了“五层次评估模型”,将柯氏模型的反应层内涵扩展为“培训可能性”和“反应”,并增加了“社会效益”(Kaufman & Keller,1994)。纵观层次评估模型的发展与演变历程,虽划分层次存在差异,但都延续了柯氏四层次模型的主体思路和方法。

到 20 世纪 90 年代,托马斯·R·古斯基(Thomas R. Guskey)将柯氏四层次模型引入教师专业发展领域。古斯基认为教师专业发展指“增进教育者专业知识、技能和态度的过程和活动”,包括各种外部培训活动、基于现场的策略以及个体自主专业发展活动等,是一个有着多种形式,且有意识、持续和系统的过程(Guskey,2002)。为了更好地了解这个过程的效果,需要对培训项目和活动进行有效评估,他进而提出了五层次的教师专业发展评估模型(包含学员反应、参与者的学习、组织支持和变化、学员应用新知识和技能、学习结果)。该模型具备更多的教师教育色彩,具有界定清晰、操作简单、指导性强、更注重形成性评价等优点,为教师专业发展尤其是教师培训的评估提供了有益参考。

在国内,赵德成(2009)、张慕华(2011)、闫冬(2012)等基于柯氏层次模型,结合具体教师远程培训项目进行了研究。王冬凌(2011)在分析培训评估内容的基础上构建了教师培训的六层次效果评估

模型,将学员反馈和对培训机构的评估纳入教师培训效果的评估模式,提出了“学员反应”“学习结果”“应用行为”“组织影响”“学生反馈”“培训机构”等六个层次,并据此提出在评估中要重视对学员的职场跟踪、教育思想的建构、发展状况的跟踪,以及强化培训机构的反思等。

### (二) 构建教师远程培训效果评估框架

为有效开展研究,笔者以柯氏四层次模型为基准,对几种典型的评估层次模型进行了对比分析(见表一)。在表一中,反应层、学习层的出现频率很高,且内涵较为一致。对行为层的描述虽各不相同,却也呈现一致趋势,均涉及参训学员个人在工作中应用培训知识的相关表现,因此笔者将其定义为“个人绩效层”。而将与组织相关的组织收益、组织支持、组织效益、社会效益等纳入“组织绩效层”。由此,在对概念进行界定后,本研究抽取多数模型的共有层次并结合教师专业发展尤其是教师远程培训特点,初步设定教师远程培训效果的评估框架为:学员反应层、学习结果层、个人绩效层、组织绩效层。

表一 典型层次评估模型的层次划分对比

|            | 反应层      | 学习层         | 行为层        | 结果层               | 其他     |
|------------|----------|-------------|------------|-------------------|--------|
| 柯氏模型       | √        | √           | √          | √                 |        |
| 汉姆布林的五层次模型 | √        | √           | 工作行为       | 成本效益分析及对企业战略目标的影响 |        |
| 菲利普斯的五层次模型 | √        | √           | 在工作中的应用    | 业务结果和投资回报率        | √      |
| 考夫曼的五层次模型  | 培训可能性和反应 | 掌握          | 应用         | 组织效益              | 社会效益   |
| 霍顿的柯氏修正模型  |          | √           | 个人绩效       | 组织结果              | 相关影响因素 |
| 古斯基的五层次模型  | √        | 参与者的学习和学生结果 | 学员应用新知识和技能 | 组织支持和变化           |        |
| 王冬凌的六层次模型  | √        | √           | 应用行为       | 组织影响和学生反馈         | 培训机构   |

1) 学员反应层:指学员对培训的反应,通常以满意度的形式出现。在表一所列的模型中,仅有希尔维亚·霍顿(Sylvia Horton)的模型未将学员反应层列入。他认为学习反应是衡量学习行为结果的标准之一。教师远程培训是依托信息技术开展的在线培训。霍顿认为,在线学习对学员来说是全新的体验,对反应层进行评估能帮助获知学员对在线学习

方式的情感接受度,并可以获取学员评价和其他有用数据,这些数据对接下来的学习能产生正面影响(Horton,2001)。大量研究也证明,在技术支持的学习中,学习者对技术支持的反应对学习具有重要作用。

2)学习结果层:指学员通过参与培训,对知识、技能、情感、态度和价值观等培训内容的理解和掌握程度。该层在各学者的模型中具有一致性,内涵是评估参训学员在知识、技能、情感、态度和价值观等方面的变化。学员的相关知识是否得到了丰富;学员是否习得了新技能、技巧及策略;学员的思想观念是否得到了转变;职业信念是否得到了强化等(王冬凌,2011)。本研究所界定的学习结果层是从知识获取而言的,侧重知识的内化,如学员掌握了教学设计的一般流程,能口头描述其操作过程。

3)个人绩效层:指学员将培训所学知识和参与培训过程中间接形成的能力、态度等应用于教育教学实践中所产生的行为和结果的综合。该层在各学者模型中都有涉及,并普遍认为是将培训所学知识迁移到工作中所引发的行为和变化。就培训效果评估而言,个人绩效层主要评估参训学员在培训之后,将培训所学的知识、技能、态度、情感和价值观运用到实践中的行为及相应结果。

4)组织绩效层:指学员个人绩效实现后对所在学校(或年级组、教研组、备课组等)产生的影响及结果。不同学者的模型都涉及培训效果之于组织的效益问题,有的限于组织层面,有的扩大到社会层面;有的强调经济效益(即投资回报),有的将经济效益与社会效益并举。组织的范围可大可小,考虑到评估指标体系构建的实践操作意义,我们将教师远程培训的组织绩效限于学校层面,重点考察受训教师在教育教学实践中运用培训所学知识给学生、同事、学校带来的影响。

## 二、指标设置

在评估指标体系框架确定后,本研究综合他人研究成果并充分结合教师远程培训实践的特点,尝试构建教师远程培训效果评估的具体指标。指标的设计具有一定的主观性,是基于研究者的认识对一级指标的逐步细化,目的是为施测提供具体依据和观察点。此外,本研究所构建的教师远程培训效果

评价指标体系采用问卷自测的方式进行施测,这也一定程度上影响了指标项的设定。

### (一)学员反应层

张慕华(2011)认为反应层主要包括学员对培训主题、网络培训课程、网络辅导老师、网络培训组织的整体反应四个方面。为充分体现教师远程培训特色,本研究将学员反应层所包含的内容界定为五方面(见表二)。

表二 学员反应层效果评估指标

| 一级指标    | 二级指标                | 三级指标  |
|---------|---------------------|---|
| A 学员反应层 | A1 参与在线学习的积极性       | A11 学员登陆次数<br>A12 在线学习时间<br>A13 任务完成情况<br>A14 互动交流等情况                   |
|         | A2 对网络课程的满意度        | A21 对网络课程专题的满意度<br>A22 对学习资源的满意度<br>A23 对主讲教师的满意度<br>A24 对培训平台用户体验等的满意度 |
|         | A3 对学习支持服务的满意度      | A31 对组织机构的满意度<br>A32 对辅导教师的满意度<br>A33 对协作交流工具等的满意度                      |
|         | A4 对在在线项目的设计和实施的满意度 | A41 对培训模式的满意度<br>A42 对培训任务量的满意度<br>A43 对进度安排的满意度<br>A44 对评价考核方式的满意度     |
|         | A5 对培训的整体反应         | A51 对培训的整体满意度<br>A52 对培训整体有用性的感知  |

其中,A2、A4与张慕华的研究结果相同。考虑到教师远程培训以自主学习为主,在线学习的积极性直接决定培训效果,因此增加了A1指标。而学习支持服务是远程教育特有概念,在远程培训中亦有重要地位,关系学员学习发生及效果,因此增加了A3。另外,为了解学员对本次培训的整体反应尤其是满意度,增加了A5。

学员反应层三级指标的设计主要依据在线学习相关理论展开。其中,A1和当前教师远程培训中实践考核的过程性评价方式基本一致,包括学员登陆次数、在线学习时间、任务完成情况、互动交流情况等。A2包括学员对网络课程专题、学习资源、主讲教师、培训平台用户体验等的满意度。A3主要指对组织机构、辅导教师、协作交流工具等的满意度。A4主要涉及学员对培训模式、任务量、进度安排、评价考核方式等的满意度。A5主要从整体满意度和学员对培训的整体感知有用性两个方面进行评价。

### (二)学习结果层

教师远程培训的最终目的是促进教师个体及群体的专业化发展。培训课程大纲、主题选取、课程设

计与开发等均应服从和服务于这一目的。教师专业发展的内涵包括专业知识、专业技能、专业情意三个方面(叶澜,1998),而对学习结果的测量依据布鲁姆的三维教学目标分类亦是如此(安德森等,2009)。因此,本研究的学习结果层评估亦依此进行划分(见表三)。

表三 学习结果层效果评估指标

| 一级指标    | 二级指标    | 三级指标  |
|---------|---------|---|
| B 学习结果层 | B1 专业知识 | B11 学科知识是否得到了丰富<br>B12 学科教学知识是否得到了丰富  |
|         | B2 专业技能 | B21 学科方面的技能是否得到提升<br>B22 课堂教学技能方法是否得到内化<br>B23 班级管理技能方法是否得到内化<br>B24 教研方面的技能方法是否得到内化<br>B25 心理调适方面的技能方法是否得到内化 |
|         | B3 专业情意 | B31 专业精神<br>B32 专业性向<br>B33 专业自我<br>B34 专业情操  |

结合教师教育的特殊性,B1 应包括学科知识和学科教学知识两大类。其中,学科教学知识是在真实教学中使用,且有别于纯粹的学科知识和一般教学知识的知识,是教师被视为专业所必须具备的知识(Shulman,1986)。B2 主要考察学员是否习得了新技能、新技巧及新策略,在具体教师远程培训中需与培训主题及内容结合考虑,主要涉及学科技能、课堂教学、班级管理、教学研究、心理调适等方面。B3 借鉴经柏龙(2008)对教师专业发展中专业情意的结构划分,包含专业精神、专业性向、专业自我、专业情操四个方面。

(三)个人绩效层

根据蒂莫西·T·鲍尔文(Timothy T. Baldwin)等的分类,笔者将个人对培训所学知识的迁移分为直接迁移和间接迁移两个方面(Baldwin & Ford, 1988)。直接迁移指参训学员能将培训所学知识、技能等应用于实践教学;间接迁移指参训学员将培训过程中潜移默化学到的东西(与培训主题通常不直接相关)应用于实践。例如,团队协作、时间管理、自我效能感、交流能力、信息素养等都可通过远程培训获得,且能间接迁移到工作场所,对个人绩效产生促进作用(见表四)。

培训的迁移受多种因素影响。埃尔伍德·霍尔顿(Elwood F. Holton)在其培训评估模型中列举了三类因素:迁移动机(个人特征)、迁移设计(培训与

表四 个人绩效层效果评估指标

| 一级指标    | 二级指标    | 三级指标  |
|---------|---------|---|
| C 个人绩效层 | C1 直接迁移 | C11 是否具备使用培训所学知识的意识<br>C12 能否辨别实践中可用所学知识的场景<br>C13 能否灵活调整并运用培训所学知识<br>C14 课堂教学效率是否得到提高<br>C15 是否有更多的科研/教研作品输出<br>C16 个人心理状态和情绪调节更加容易<br>C17 知识应用和创新次数是否增加 |
|         | C2 间接迁移 | C21 自我效能感是否得到提高<br>C22 时间管理能力是否得到提高<br>C23 团队精神是否得到增强<br>C24 交流能力是否得到提高<br>C25 信息素养是否得到提高   |

组织因素交互作用)、迁移气氛(组织因素)(Holton, 1996)。对迁移影响因素进行探讨,可用于编制更为具体的三级指标。珍妮丝·L·鲁伊勒(Janice Z. Rouiller)和欧文·L·戈德斯坦(Irwin L. Goldstein)通过实证研究,确定了培训迁移气氛的结构包括情境线索(用于提醒受训者并为其提供在工作中应用培训所学的机会)和结果线索(受训者在实践中应用培训所学知识后得到的反馈)两个要素(Rouiller & Goldstein, 1993)。据此,研究者在二级指标“直接迁移”下设置了“能否辨别实践中可用所学知识的场景”条目。

(四)组织绩效层

教师存在于教师群体(即教师共同体)和教育组织机构之中,在通过受训获得自我提升基础上,其除了通过课堂教学影响学生外,还将通过日常工作与同事交往,促进所在单位工作成绩的提升,并对同事产生积极影响。因此,我们将组织绩效归纳为对学生的影响、对同事的影响、对学校的影响三个方面(见表五)。

表五 组织绩效层效果评估指标

| 一级指标    | 二级指标      | 三级指标  |
|---------|-----------|---|
| D 组织绩效层 | D1 对学生的影响 | D11 学生是否感知到教师上课行为的变化<br>D12 学生的学习积极性、主动性是否有所提高<br>D13 学生的学习成绩是否得到提高<br>D14 班级气氛是否得到改善 |
|         | D2 对同事的影响 | D21 同事之间的分享活动是否增多<br>D22 教研活动效率是否得到提高<br>D23 组织气氛(如团队斗志、教研组或年级组教学热情)是否得到提升            |
|         | D3 对学校的影响 | D31 培训效果是否得到学校相关领导的认可<br>D32 学校相关政策、实践工作是否改进<br>D33 家长对学校的态度是否有所改变                    |

其中,D1 关注受训教师所在班级的学生体验、学生学习表现是否改善、班级学习氛围是否改善等,

这些是教师远程培训项目组织实施所期望的最重要结果。D2 关注对受训教师周边同事的影响,围绕教师实践共同体的积极作用展开,主要是交流、讨论、分享、建构、问题解决等(张兰,2010)。D3 关注对学校工作的影响,主要包括校方对培训项目的认可,以及是否促进了学校相关政策、实践工作的改进,获得家长的积极反馈等。

### 三、应用德尔斐法修正评估指标体系

#### (一)研究方法及专家遴选

本研究采用德尔菲法,以问卷形式对专家进行意见征询。为集中专家智慧并排除可能存在的权威干扰,在调查期间,专家只与研究者保持联系,不相互讨论(即背对背的通讯方式)。经过两轮意见征询和修订,专家意见趋于集中,最后由研究者汇总研究结果。

专家选择是影响德尔菲法研究结果质量的关键因素之一。本研究预选了 20 位专家,得到 16 位回应(来自高校的 8 位,中小学 8 位,见表六)。文森特·W·米切尔(Vincent W. Mitchell)认为德尔斐法中专家人数达到 13 人以上,误差降幅不明显(Mitchell,1991),因此本研究符合要求。16 位专家均参与过教师远程培训相关工作,并活跃于实践或研究第一线。其中,来自中小学的专家尽管研究相对薄弱,但对教师远程培训有较为全面、系统且独到的见解。

#### (二)第一轮专家问卷调查

##### 1. 问卷发放

第一轮专家问卷调查的目的是征求专家对研究者初拟的教师远程培训效果评估指标体系的修改意见,并做适切度问询。问卷包括简介、专家信息区、填写提示、一级指标划分咨询表(4 项)、二级指标划分咨询表(13 项)、三级指标划分咨询表(50 项)。为避免专家对指标项内涵产生误解,在一二级指标划分咨询表中均有简短的指标项描述。专家只需按项目勾选恰当的适切度值(1、2、3、4、5)。1-5 代表适切度,且依次增加。此外,问卷也声明评估指标将用于后续开发学员自我评估工具。

鉴于初拟框架在推导过程中采用了文献研究法、访谈法,且在框架和二三级指标项的修改过程中征询过部分专家的意见,指标项描述的修改也汇集

表六 专家背景信息

| 项目 |      | 人数 | 百分比(%) |
|----|------|----|--------|
| 性别 | 男    | 10 | 62.5   |
|    | 女    | 6  | 37.5   |
| 学历 | 本科   | 5  | 31.25  |
|    | 硕士   | 7  | 43.75  |
|    | 博士   | 4  | 25     |
| 职称 | 教授   | 6  | 37.5   |
|    | 副教授  | 2  | 12.5   |
|    | 中学高级 | 7  | 43.75  |
|    | 小学特级 | 1  | 6.25   |

了专家智慧和专业判断,因此问卷具有较高信度和效度。

第一轮问卷共发放 16 份,回收 16 份且全部有效,回收率和有效率均为 100%。

#### 2. 统计分析

专家对各级指标适切度的打分情况如表七、表八和表九所示。表格左侧是根据专家填写选项进行计数统计后的结果;右侧的描述项是研究者基于统计学原理计算得出的平均数、中位数、众数、标准差,目的是进一步分析专家对各指标项适切度评分的均值、集中趋势、一致性及离散分布情况。

如表七所示,4 个一级指标适切度调查的平均数均在 4.0(比较合适)以上,中位数、众数都为 4(比较合适)或 5(非常合适);4 个一级指标中专家认为“非常适合”与“比较合适”的比例合计达 90% 以上;标准差都在 1.0 以下,A 和 D 略高,表明专家对学员反应层、组织绩效层的适合性认识存在一定分歧,但尚在可接受范围。依据平均数,专家对 4 个一级指标的认同度由高到低依次为:学习结果层、个人绩效层、学员反应层、组织绩效层。

如表八所示,13 个二级指标中,平均数低于 4.0 的共 3 项(A1、A4、D3);除 A1 外,中位数、众数都为 4(比较合适)或 5(非常合适);二级指标中“非常适合”与“比较合适”的比例合计在 75% 以下的只有 A1 一项,A4、A5、C2、D3 为 75%;各项标准差都保持在 1.0 以下,其中 C1 的值最低,为 0,反映了专家对直接迁移的一致认可。相对于一级指标,二级指标各项标准差的数值较高,专家认识呈一定离散性。

三级指标调查统计结果如表九所示。与前两级指标征询不同,考虑到三级指标项目多,故问卷采用二分法,即调查专家对三级指标设置的态度是同意或不同意。专家勾选“同意”,计为“1”;勾选“不同意”,计为“0”;未标记视作不同意。由于专家赋值

只有两种可能,因此我们只计算平均数,不计算标准差。从调查结果看,共7项指标的同意比例达到了100%,10项指标的同意比例达到93.75%,13项指标的同意比例处于80%-90%之间,14项指标的同意比例在60%-80%之间,6项指标的同意比例在60%以下。从平均数来看,A11、A12、D11最低,且低于0.5,即半数以上专家不赞同。

3. 意见收集与指标修订

本研究的德尔菲问卷为半开放式,专家除填写评估指标的适切度外,还可通过填写“修改意见”的形式反馈。经筛选,最后采纳的意见或建议如下:

1)有关指标项名称及内涵描述

• B3(专业情意):名称内涵不清,三级指标也需相应调整;

- C2(间接迁移):内涵需进一步提炼;
- D31(培训效果是否得到了学校领导的认可):建议改为“培训是否促进了教师的专业发展”;
- B31、B32、B33、B34:表述过于专业化。

2)分类及指标领域归属

- A1:“学员参与在线学习的积极性”数据可从培训平台直接获取,无需作为自我评估指标,建议删除;
- D11:“学生是否感知到教师上课形式的变化”与D12、D13相关,建议删除。

3)增加指标项

- D:3位专家建议其下增加二级指标:对社会的影响;
- D3:增加三级指标,即学校相关方面的成绩

表七 一级指标适切度

| 指标项    | 专家填答题百分比(%) |      |     |       |     | 描述项  |     |    |      |
|--------|-------------|------|-----|-------|-----|------|-----|----|------|
|        | 非常适合        | 比较适合 | 一般  | 比较不适合 | 不适合 | 平均数  | 中位数 | 众数 | 标准差  |
|        | 5           | 4    | 3   | 2     | 1   |      |     |    |      |
| 学员反应 A | 33.3        | 60   | 6.7 | 0     | 0   | 4.31 | 4   | 4  | .602 |
| 学习结果 B | 93.3        | 6.7  | 0   | 0     | 0   | 4.94 | 5   | 5  | .250 |
| 个人绩效 C | 80          | 20   | 0   | 0     | 0   | 4.75 | 5   | 5  | .447 |
| 组织绩效 D | 26.7        | 66.7 | 6.7 | 0     | 0   | 4.19 | 4   | 4  | .544 |

表八 二级指标适切度

| 指标项                  | 专家填答题百分比(%) |      |      |       |     | 描述项  |     |    |      |
|----------------------|-------------|------|------|-------|-----|------|-----|----|------|
|                      | 非常适合        | 比较适合 | 一般   | 比较不适合 | 不适合 | 平均数  | 中位数 | 众数 | 标准差  |
|                      | 5           | 4    | 3    | 2     | 1   |      |     |    |      |
| A1 学员参与在线学习的积极性      | 0.0         | 12.5 | 37.5 | 50    | 0   | 2.63 | 2.5 | 2  | .719 |
| A2 学员对网络课程的满意度       | 75.0        | 25.0 | 0    | 0     | 0   | 4.75 | 5   | 5  | .447 |
| A3 学员对学习支持服务的满意度     | 68.8        | 31.3 | 0    | 0     | 0   | 4.69 | 5   | 5  | .479 |
| A4 学员对在线项目的设计和实施的满意度 | 25.0        | 56.3 | 25.0 | 0     | 0   | 3.94 | 4   | 4  | .680 |
| A5 学员对培训的整体反应        | 43.8        | 31.3 | 25.0 | 0     | 0   | 4.19 | 4   | 5  | .834 |
| B1 专业知识              | 75.0        | 25.0 | 0    | 0     | 0   | 4.75 | 5   | 5  | .447 |
| B2 专业技能              | 87.5        | 12.5 | 0    | 0     | 0   | 4.88 | 5   | 5  | .342 |
| B3 专业情意              | 31.3        | 56.3 | 12.5 | 0     | 0   | 4.19 | 4   | 4  | .655 |
| C1 直接迁移              | 100.0       | 0    | 0    | 0     | 0   | 5    | 5   | 5  | 0    |
| C2 间接迁移              | 37.5        | 37.5 | 25.0 | 0     | 0   | 4.13 | 4   | 4  | .806 |
| D1 对学生的影响            | 75.0        | 25.0 | 0    | 0     | 0   | 4.75 | 5   | 5  | .447 |
| D2 对同事的影响            | 50.0        | 43.8 | 6.3  | 0     | 0   | 4.44 | 4   | 5  | .629 |
| D3 对学校的影响            | 18.8        | 56.3 | 25.0 | 0     | 0   | 3.94 | 4   | 4  | .680 |

表九 三级指标适切度

| 指标项   | 同意 |        | 描述项  |
|---|----|--------|------|
|   | 人数 | 比例(%)  | 平均数  |
| A21 A22 A32<br>B12 B24<br>D12 D13                           | 16 | 100.00 | 1.00 |
| A23 A33 A43<br>B11 B25<br>C11 C13 C14<br>C24 C25            | 15 | 93.75  | 0.94 |
| A41 A42 A52<br>B21 B22 B31<br>B34<br>C15<br>D14             | 14 | 87.50  | 0.88 |
| A44<br>B23<br>C17 C22                                       | 13 | 81.25  | 0.81 |
| A24 A51<br>B32 B33<br>C12 C16 C23<br>D21 D22 D23<br>D32 D33 | 12 | 75.00  | 0.75 |
| C21   | 11 | 68.75  | 0.69 |
| A31   | 10 | 62.50  | 0.63 |
| A14   | 9  | 56.25  | 0.56 |
| A13<br>D31  | 8  | 50     | 0.50 |
| A11 A12<br>D11  | 7  | 43.75  | 0.44 |

表现是否得到提高。

根据量化的统计分析专家意见汇总,本研究对评估指标体系做了修改,具体如下:

一级指标方面,由于4项指标“非常适合”和“比较适合”之和达到93.5%(A、D)和100%(B、C),远大于最低标准75%,且标准差均在1.0以下,因此,本研究认为一级指标已达成一致,不做修改。下一轮问卷将不再对该层次的划分进行专家意见征询。

二级指标方面,除A1外,其余12项指标“非常适合”与“比较适合”比例均达75%以上。结合专家意见,并考虑到远程培训学员学习风格的个体差异,以及本指标体系拟采用学员自我评估方式进行施测,删除A1项。另将“B3专业情意”改为“B3情感、态度和价值观”;C2名称不变,凝练内涵描述;增加“D4对社会的影响”。

三级指标方面,重点结合同意率低于60%的指标(A11, A12, A13, A14, D11, D31)进行修改,删除A11、A12、A13、A14、D11;修改D31为“是否带动学校教师专业发展活动的开展”;添加“B13学员的教研知识是否得到了丰富”;将B31-B34修改为“教育

热情是否得到提升”“职业幸福感是否得到提升”“个人职业规划是否更加明确”“职业归属感和使命感是否得到提升”。

### (三)第二轮专家问卷调查

#### 1. 问卷发放

第二轮问卷的发放对象及发放方式与第一轮相同,共发放问卷16份,回收15份,有效问卷15份,回收率为87.5%。第二轮问卷的主要目的是咨询一级指标的权重,以及对二三级指标界定的再咨询。问卷所列指标项在第一轮问卷基础上进行了如前所述的修订。

#### 2. 统计分析

##### 1) 一级指标权重分析

对4项一级指标的经验权重赋值进行平均权重、标准差统计,并按平均权重进行排序(见表十)。依据平均权重,4项一级指标的重要性从高到低依次为:个人绩效层、学习结果层、学员反应层和组织绩效层。这表明专家认为培训效果最重要和最关键的体现是提高教师的个人工作绩效,其次是学习结果。这与“研训用”一体的教师远程培训理念相吻合,即培训以教师专业化发展为指向,旨在提高教师素养和实际施教水平(武丽志等,2014)。学员反应层和组织绩效层权重相当且在4个项目中较低。据访谈了解,专家尽管认同学员反应层、组织绩效层是评估教师远程培训效果的重要维度,但对其反映培训效果的直接性、测量的可操作性及效度等存在一定顾虑。

表十 一级指标权重

| 一级指标    | 平均权重  | 标准差  | 排序 |
|---------|-------|------|----|
| A 学员反应层 | 0.203 | .048 | 3  |
| B 学习结果层 | 0.289 | .060 | 2  |
| C 个人绩效层 | 0.304 | .055 | 1  |
| D 组织绩效层 | 0.203 | .055 | 3  |

##### 2) 二级指标分析及修订

二级指标方面,第二轮专家问卷调查仍然从百分比、平均数、中位数、众数、标准差5个方面进行统计分析。通常,当80%以上的专家选项相同,称为通过一致性(Murry & Hammons, 1995)。本研究认为“非常适合”与“比较适合”均代表同意,两项所占比例之和达到80%以上,且标准差不大于1.0,则认为该指标项通过了一致性检验。

如表十一所示,13 个二级指标中,有 12 项的平均数达到了 4.0 以上。新添加项 D4 赞同率较低,平均数仅为 2.6,“非常适合”和“比较适合”的百分比之和为 26.7%,且标准差也超过了 1.0,结合专家的意见,培训效果在社会影响上的体现甚微,且在效果评估时无法清楚区分,故可忽略不计,建议删去。除此之外,其它二级指标“非常适合”与“比较适合”的比例合计均达到 80% 以上且达到了一致性。因此,删除 D4(对社会的影响)项。

### 3) 三级指标分析及修订

三级指标方面,第二轮专家问卷仍询问专家是否同意项目设置,进而计算同意比例、平均数、标准差(见表十二)。

从表十二可知,48 项三级指标中,除对社会影响的两个三级指标 D41、D42 同意率低至 26.67% 外,其余 46 项均不低于 80%,达到一致性标准。在第一轮调查中同意率稍低的项(B31、B32、B33、B34)也因修改而获得提高。因此,删除与 D4 对应的 D41、D42 项。

### 3. 共识率判断

通常认为,专家对所有题项的一致性达到 80% 以上才算达成共识,可以考虑停止继续进行新一轮德尔菲专家调查。本研究中,二级指标项的共识率

表十二 三级指标适切度(第二轮)

| 指标项  | 同意 |        | 描述项  |      |
|--|----|--------|------|------|
|  | 人数 | 百分比    | 平均数  | 标准差  |
| A11 A12 A13 A22 A23 A31<br>A32 A33 A34 A42<br>B11 B12 B13 B22 B24 B25<br>C11 C13 C14 C25 D11 D12 | 15 | 100.00 | 1.00 | .000 |
| A14 A41<br>B21 B23 B31 B32 B34<br>C12 C15 C16 C17 C22 C24<br>D13                                 | 14 | 93.33  | .93  | .258 |
| B33<br>D21 D22 D31 D33   | 13 | 86.67  | .87  | .352 |
| A21<br>C21 C23<br>D23 D32  | 12 | 80.00  | .80  | .414 |
| D41 D42  | 4  | 26.67  | .27  | .458 |

为  $12/13=92.3\%$ ; 三级指标项的共识率为  $46/48=95.8\%$ ,均达到 80% 以上。因此,我们认为拟定的中小学教师远程培训效果评估指标体系经两轮修改已达成共识,专家意见一致。加之问卷的整体稳定性较好,因此可结束德尔菲问卷调查。

## 四、结果与讨论

### (一) 研究结果

经过两轮调查,我们探索性地建立了包含 4 个一级指标、12 个二级指标、46 个三级指标的远程培训效果评估指标体系(见表十三)。由于受研

表十一 二级指标适切度(第二轮)

| 指标项                | 专家答题百分比(%) |      |      |       |     | 描述项  |     |    |       |
|--------------------|------------|------|------|-------|-----|------|-----|----|-------|
|                    | 非常适合       | 比较适合 | 一般   | 比较不适合 | 不适合 | 平均数  | 中位数 | 众数 | 标准差   |
|                    | 5          | 4    | 3    | 2     | 1   |      |     |    |       |
| A1 对网络课程的满意度       | 100        | 0    | 0    | 0     | 0   | 5.00 | 5   | 5  | .000  |
| A2 对学习支持服务的满意度     | 73.3       | 20   | 6.7  | 0     | 0   | 4.67 | 5   | 5  | .617  |
| A3 对在线项目的设计和实施的满意度 | 66.7       | 33.3 | 0    | 0     | 0   | 4.67 | 5   | 5  | .488  |
| A4 对培训的整体反应        | 73.3       | 26.7 | 0    | 0     | 0   | 4.73 | 5   | 5  | .458  |
| B1 专业知识            | 86.7       | 13.3 | 0    | 0     | 0   | 4.87 | 5   | 5  | .352  |
| B2 专业技能            | 66.7       | 33   | 0    | 0     | 0   | 4.67 | 5   | 5  | .488  |
| B3 情感、态度和价值观       | 66.7       | 33.3 | 0    | 0     | 0   | 4.67 | 5   | 5  | .488  |
| C1 直接迁移            | 100        | 0    | 0    | 0     | 0   | 5.00 | 5   | 5  | .000  |
| C2 间接迁移            | 13.3       | 86.7 | 0    | 0     | 0   | 4.13 | 4   | 4  | .352  |
| D1 对学生的影响          | 100        | 0    | 0    | 0     | 0   | 5.00 | 5   | 5  | .000  |
| D2 对同事的影响          | 60         | 33.3 | 6.7  | 0     | 0   | 4.53 | 5   | 5  | .640  |
| D3 对学校的影响          | 20         | 66.7 | 13.3 | 0     | 0   | 4.07 | 4   | 4  | .594  |
| D4 对社会的影响          | 6.7        | 20   | 6.7  | 60    | 6.7 | 2.6  | 2   | 2  | 1.121 |

研究者能力及可获取支持的专家资源的限制,研究成果可能存在疏漏和偏颇,还不成熟,但该指标体系尝试将企业培训效果评价模型引入教师远程培训领域,并与实践紧密结合,力图反映教师远程培训的在线教育和教师教育特色,具有较强的针对性,并经专家背对背两轮修订,因此具有一定的科学性和先进性,对于当前我国教师远程培训效果评估工作具有积极指导意义。

(二) 讨论

教师远程培训的效果评估是一项复杂的系统工程,基于我们构建的评估指标体系,现做如下讨论,亦是反思与展望:

1) 教师远程培训效果的核心指向问题。概括而言,教师有组织的专业发展主要包括接受培训、参与研究和实践教学三大类。其中,培训是主动解决实践教学矛盾与冲突、提升教师专业知识与教学技能的关键,是打破教师个体专业发展原始平衡,引导建立更高层次发展水平的能动因素。教师远程培训作为教师培训的重要途径,其目的是为了提升个体及组织的工作绩效,增强实践教学能力,最终实现教育教学质量的不断提升。

2) 评估指标体系的适用性和推广性问题。通过分析构建教师远程培训评估指标体系的过程及结果,我们发现,尽管研究者试图尽可能地在指标设置时充分反映远程培训特点,但最终设置的指标项目具有较强的通用性和推广性,绝大多数指标亦适用于传统教师培训,只有学员反映层的部分指标和远程培训紧密相关。这也印证了网络教育仅是教师培训的途径和方式之一,并没有改变教师培训的本质和基本规律。

3) 教师远程培训效果评估的操作问题。评估指标体系的构建是教师远程培训效果评估的基础,但如何科学实施评估更富有挑战。柯氏四层次模型提供了不同层次评估的资料收集方法和途径,如反应层的衡量方法可采用问卷调查、评价调查表填写、评价访谈;学习层可采用评价调查表填写、绩效考试、案例研究;行为层可采用绩效考核、观察和绩效记录等;结果层可考察组织的故事率、生产率和士气等。这些方法为我们提供了较好的借鉴,但实践工作受到诸多客观条件的限制,精确评估不一定可操作,必须综合考虑可以利用的评估资源(如人力、物

表十三 教师远程培训效果评估指标体系

| 一级指标          | 二级指标              | 三级指标                          |
|---------------|-------------------|-------------------------------|
| 学员反应层 A       | A1 对网络课程的满意度      | A11 对网络课程专题的满意度               |
|               |                   | A12 对学习资源的满意度                 |
|               |                   | A13 对主讲教师的满意度                 |
|               |                   | A14 对培训平台用户体验等的满意度            |
|               | A2 对学习支持服务的满意度    | A21 对组织机构的满意度                 |
|               |                   | A22 对辅导教师的满意度                 |
|               |                   | A23 对协作交流工具等的满意度              |
|               | A3 对在线项目设计和实施的满意度 | A31 对培训模式的满意度                 |
|               |                   | A32 培训任务量是否合理                 |
|               |                   | A33 进度安排是否合理                  |
|               |                   | A34 对评价考核方式等的满意度              |
|               | A4 对培训的整体反应       | A41 对培训的整体满意度                 |
| A42 对培训有用性的感知 |                   |                               |
| 学习结果层 B       | B1 专业知识           | B11 学科知识是否得到了丰富               |
|               |                   | B12 学科教学知识是否得到了丰富             |
|               |                   | B13 教研知识是否得到了丰富               |
|               | B2 专业技能           | B21 学科方面的技能是否得到提升             |
|               |                   | B22 课堂教学技能方法是否得到内化            |
|               |                   | B23 班级管理技能方法是否得到内化            |
|               |                   | B24 教研方面的技能方法是否得到内化           |
|               |                   | B25 心理调适方面的技能方法是否得到内化         |
|               | B3 情感、态度和价值观      | B31 教育热情是否得到提升                |
|               |                   | B32 职业幸福感是否得到提升               |
|               |                   | B33 个人职业规划是否更加明确              |
|               |                   | B34 职业归属感和使命感是否得到提升           |
| 个人绩效层 C       | C1 直接迁移           | C11 是否具备使用培训所学知识意识            |
|               |                   | C12 能否辨别实践中可用所学知识的场景          |
|               |                   | C13 能否灵活调整并运用培训所学知识           |
|               |                   | C14 课堂教学效率是否得到提高              |
|               |                   | C15 是否有更多的科研/教研作品输出           |
|               |                   | C16 个人心理状态和情绪调节更加容易           |
|               |                   | C17 知识应用和创新次数是否增加             |
|               | C2 间接迁移           | C21 自我效能感是否得到提高               |
|               |                   | C22 时间管理能力是否得到提高              |
|               |                   | C23 团队精神是否得到增强                |
|               |                   | C24 交流能力是否得到提高                |
|               |                   | C25 信息素养是否得到提高                |
| 组织绩效层 D       | D1 对学生的影响         | D11 学生的学习积极性、主动性是否得到提高        |
|               |                   | D12 学生的学习成绩是否得到提高             |
|               |                   | D13 班级气氛是否得到改善                |
|               | D2 对同事的影响         | D21 同事之间的分享活动是否增多             |
|               |                   | D22 教研活动效率是否得到提高              |
|               |                   | D23 组织气氛(如团队精神、教研组教学热情)是否得到改善 |
|               | D3 对学校的影响         | D31 是否带动了学校教师专业发展活动的开展        |
|               |                   | D32 学校相关政策、实践工作是否改进           |
|               |                   | D33 家长对学校的态度是否有所改变            |

力、财力、时间等),以及评估方法和评估工具的可操作性,而这无疑是复杂的。所幸的是,随着信息技术的发展,基于大数据的学习分析技术将为教师远程培训效果评估提供重要支持。此外,自我评估尽管具有一定的主观性,但作为一种将培训效果评估

主动权交给受训教师的方法,更有利于教师反思自己的收获和专业发展轨迹,推进自我改进和提高。这样的效果评估过程有利于引发教师更深层次、更为主动的学习,不失为一种值得研究和探索的方法。

### [参考文献]

- [1]安德森等(2009). 布鲁姆教育目标分类学:分类学视野下的学与教及其测评[M]. 蒋小平等译. 北京:外语教学与研究出版社:1-266.
- [2]Baldwin, T. T. & Ford, J. K. (1988). Transfer of training: A review and directions for future research [J]. *Personal psychology*, 41(4):63-105.
- [3]Guskey, T. R. (2002). Professional development and teacher change [J]. *Teachers and Teaching: Theory and Practice*, 8(3):381-391.
- [4]Holton, E. F. (1996). The flawed four-level evaluation model [J]. *Human Resource Development Quarterly*, 7(1):5-25.
- [5]Horton, S. (2001). Competency management in the British civil service[J]. *International Journal of Public Sector Management*, 13(4):354-368.
- [6]Horton, W. (2001). Evaluating e-learning (The ASTD E-Learning Series) [M]. Alexandria: American Society for Training & Development:1-150.
- [7]经柏龙(2008). 教师专业素质的形成与发展研究[D]. 长春:东北师范大学.
- [8]Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: Beyond Kirkpatrick[J]. *Human Resources Development Quarterly*, Winter, 5(4):371-380.
- [9]Kirkpatrick, D. L. (1998). Evaluating training programs: The four levels(2nd Edition)[M]. San Francisco: Berrett-Koehler Publishers, Inc:1-288.
- [10]莱斯利(2003). 培训效果评估[M]. 牛雅娜等译. 北京:中国劳动社会保障出版社:4-5.
- [11]Mitchell, V. W. (1991). The Delphi technique: An exposition and application [J]. *Technology Analysis & Strategic Management*, 3(4):333-358.
- [12]Murry, J. W., & Hammons, J. O. (1995). Delphi: A versatile methodology for conducting qualitative research[J]. *The Review of Higher Education*, 18(4):423-436.
- [13]Phillips, J. (1995). Return on investment—Beyond the four Level[C]. In *Academy of HRD conference proceeding*. E. Nolton (ED), 1995: 42.
- [14]Rouiller, J. Z., & Goldstein, I. L. (1993). The relationship between organizational transfer climate and positive transfer of training [J]. *Human Resources Development Quarterly*, 4(4):377-390.
- [15]Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching[J]. *Educational Researcher*, 15(2):4-14.
- [16]王冬凌(2011). 建构教师培训效果评估模式:内涵与策略——兼论教师培训机构走向专业化的路径[J]. *大连教育学院学报*, 27(4):4-8.
- [17]武丽志,许晓艺(2014). “研训用”一体的教师远程培训研究与实践[J]. *继续教育*, (6):16-17.
- [18]肖娟(2007). 培训效果评估方法及其应用研究[D]. 南京:南京理工大学.
- [19]闫冬(2012). 教师教育技术能力培训的有效转化策略研究[D]. 天津:天津师范大学.
- [20]叶澜(1998). 新世纪教师专业素养初探[J]. *教育研究与实验*, (1):43-48.
- [21]张兰(2010). 教师实践共同体建构研究[D]. 重庆:西南大学.
- [22]张慕华(2011). 中小学教师教育技术能力网络培训效果评估研究[D]. 西安:陕西师范大学.
- [23]赵德成,杜屏,杜育红(2009). 参与式教学培训效果:基于四层次评估模式的分析[J]. *教师教育研究*, 21(6):47-52.

(编辑:魏志慧)

## A Delphi Study on The Evaluation Index System of Training Effect for Teachers' Distance Training

WU Lizhi & WU Tiantian

- (1. School of E-learning, South China Normal University, Guangzhou 510631, China;  
2. Bmob Network Technology Limited, Guangzhou 510631, China)

**Abstract:** Distance training, which is popularized through nationwide massive teachers training and government projects like "National Training Plan," has been one of the most important approaches to national teachers training. Such training approaches have characteristics like massive trainees, great influences, and rapid changes. Evaluation research is critical and valuable for such a training system. Through the literature review, and guided by adult learning, training, and assessment theories, this research is intended to establish an evaluation index system of training effect for teachers distance training.

*The Delphi method is used twice to collect suggestions for three sub-indexes. Before the first round of Delphi application, researchers draw up Evaluation Index System through the literature review. Researchers then collect experts' feedback on indexes. They pick out indices which are suitable for measuring effects of training according to their experience and knowledge. Based on expert suggestions, original indices and expressions are revised. In the second round of Delphi application, researchers investigate the experts' opinion on the weight of index, and revise three levels of indices in system. Consensual validity is testified effective as values are above 80%. This new evaluation index system is proved to be up to the requirement of Delphi method. The soundness and reasonableness of index system is validated. Finally, two authors of this article establish the Evaluation Index System for online teacher training, which consist of 4 first level indices, 12 second level indices and 46 third level indices. There are four first level indices in this evaluation system, including learners' reaction, learning results, personal performance and organizational performance. Learners' reaction can be measured by four second indices, which are the learners' satisfaction degree of online courses, learning support service, online projects design and implementation, overall training. These indices focus on learners' direct feeling for training. It is a way for trainers to find out learners' reaction to online course content, training platform, resources, teaching, communication tools, learning support service and learners' general ideal about this training etc. Learning results contain three second indices as professional knowledge, professional technologies, learners' faith, attitude and values. These three indices help to learn people's development in subject knowledge, teaching knowledge, teaching technologies and their attitude etc. in training. These direct and indirect transfer indices measure personal performance. These indices mainly evaluate learners' levels in transferring training results into personal performance in their practice of real world. Organizational performance contains three sub-indices of training's influence on learners' students, their colleagues and workplace (school). Authors use these indices to evaluate training's influence on learners' workplace and their members. The Evaluation Index System has high validity of expert, better applicability and extension, and it improves the theory systems of distance teacher training. In addition, it provides instructions for the sustainable and healthy development of practices in distance teacher training.*

**Key words:** distance training; effect; evaluation; Delphi