

# 基于有效性验证经验的美国早期教育质量评价的元评价研究

孙 贺 群

(天津师范大学 教育科学学院, 天津 300387)

**摘要:**美国早期教育质量评级与促进系统(QRIS)是在世界范围内发展比较成熟且具有代表性的早教质量监督与保障政策机制之一。当前美国各州正逐步展开 QRIS 的有效性验证工作。其中,对评级系统的评价,即早期教育机构质量评价的元评价成为验证 QRIS 有效性工作的重中之重。美国各州 QRIS 评价系统元评价主要涉及以下四方面检验:质量要素划分的合理性检验,质量指标体系构建的科学性检验,质量评级结果的准确性检验,质量等级与儿童发展结果的相关性检验。美国 QRIS 元评价实施也面临着诸多挑战,需要去突破。借鉴美国的经验,我国应构建普适性早期教育质量评价体系,同时考虑元评价工作,并协调政府、市场、学术三种力量开展元评价以保证政策效能。

**关键词:**美国;早期教育;质量评价;元评价

**中图分类号:**G61 **文献标志码:**A **文章编号:**1000-5315(2018)03-0069-09

当前在国际上,如何才能让全部儿童接受高质量的早期教育仍是一个热点话题,备受关注。各国政府都在积极探索尝试建立儿童早期教育质量的监督与保障体系,针对儿童早期教育质量现状而开展系列研究、编订教育质量评估量表、构建教育质量评估模型并出台一系列政策措施等以帮助提升早教质量。其中,美国的早期教育质量评级与提升系统(Quality Rating and Improvement Systems,以下简称 QRIS)已经实施了十几年,自 1998 年俄克拉荷马州首个实施 QRIS,到 2015 年 2 月美国除了密苏里州没有推行 QRIS,以及包含阿拉斯加州、夏威夷州在内的 6 个州正在计划推行 QRIS 之外,其余各州均实施了 QRIS<sup>[1]</sup>。它是在世界范围内发展比较成熟且具有代表性的早期教育质量监督与保障政策

机制之一。尽管如此,还是有研究指出,QRIS 对早教机构进行评级的准确性已经遭到质疑,这将影响到整个 QRIS 系统的公众信服度,包括家长的信任程度以及早教机构的参与积极性<sup>[1]</sup>。因此,对 QRIS 进行有效性验证,已经成为一个亟须开展的工作。另外,在这十几年里,QRIS 在美国各州蓬勃发展的同时,也越来越受到联邦政府的重视。如今 QRIS 已经成为联邦政府为提升全美早期教育质量而资助的竞争性项目“力争上游——早期学习的挑战”(Race to the Top-Early Learning Challenge,以下简称 RTT-ELC)的核心考核部分,具体说来,就是各州政府是否能顺利申请到联邦政府的 RTT-ELC 项目基金,很大程度上取决于 QRIS 的实施情况。这样一来,各州政府在申请 RTT-ELC 项目基

**收稿日期:**2016-11-28

**基金项目:**教育部人文社科基金资助项目“早期教育机构课程质量监督与保障机制研究”(15YJC880066);天津市教育科学“十三五”规划课题“天津市幼儿园教育质量监督与保障机制研究”(BEYP5063)。

**作者简介:**孙贺群(1984—),女,河北平泉人,教育学博士,天津师范大学教育科学学院讲师,研究方向为学前教育。

金时,需要将 QRIS 有效性验证报告作为一份重要的申请材料递交联邦政府。这也成为最近各州政府积极关注并进行 QRIS 的有效性验证活动的原因。

美国各州对 QRIS 进行有效性验证的核心工作是对其等级评定系统进行评价。因为 QRIS 的核心工作就是对早教机构进行质量评级,其他活动都要围绕评级结果来进行。具体来说,QRIS 是一个比较复杂的政策机制。实施运行的基本逻辑是:通过建立等级评定系统对早教机构进行质量评级,与此同时,会根据 QRIS 的评级结果,为早教机构提供专业支持和财政激励,以帮助其继续提升质量等级。之后,将早教机构的质量等级信息公开发布,有计划地提供给家长,家庭会选择高质量等级的早教机构,从而保证儿童能够接受高质量的早期教育,在认知与情感发展以及入学准备等方面都收获更好的结果。而另一方面,QRIS 的质量等级也为早教机构提供了一个基准,早教机构为吸引更多生源,也会参照 QRIS 的“质量等级坐标系”来积极提升质量,各早教机构之间的质量等级差距缩小,早期教育质量的整体水平提高。因此,对评级系统的评价,即早期教育机构质量评价的元评价,成为验证 QRIS 有效性工作的重中之重。

### 一 美国早期教育质量评价元评价的意义

元评价(Meta-evaluation)是指按照一定的理论框架和价值标准对教育评价本身所做的评价与研究<sup>[2]</sup>。简而言之,元评价就是对评价本身的评价。当前,美国各州开始着手进行的 QRIS 有效性验证中,对评级系统的评价就是早期教育质量评价元评价。QRIS 评级系统评价是一个复杂的过程,需要长期持续、反复进行,面临着诸多困难与挑战,但这是美国近年来急需向前推进的一项工作<sup>[1]</sup>。从 QRIS 的评级系统评价的经验中,可获知早期教育质量评价元评价的意义。

(一)从评价系统外部看,元评价可以解决问责需要与赢得支持

教育政策的制定与实施需要监督评价,第三方评价是评判教育政策设计是否科学合理、实施是否高效有序的重要手段,也是对政策制定者和实施者进行问责的依据。质量评级系统是 QRIS 这一政策机制的核心组成部分,对其进行评价可以解决对 QRIS 进行问责的需要。美国各州政府已经在努力地推行 QRIS,并且给予很多的外部支持,但所有的

这些努力都是基于 QRIS 质量等级评定是非常准确的理论假设之上。如此一来,对 QRIS 评级系统的评价成为对整个政策进行问责的关键。对 QRIS 评级系统的评价可以回答这样的问题:其是否达成了对早教机构准确质量定级的目标?如果没能达成问题出在哪里?原因又是什么?导致理想政策或制度在现实中陷入困境的原因很复杂,不同问题背后的责任主体也不尽相同。而回答上述问题的过程,正是寻找原因和进行“追责”的过程。当然,“追责”并不是要对相关责任主体进行奖惩,而是希望不同的责任主体能够发挥各自的能动性来解决问题,保证 QRIS 评价系统能够准确对早教质量进行评价。

与此同时,对 QRIS 评级系统进行评价,有利于赢得政府、家长、早教机构及其他相关利益者对这一政策机制的支持。各州政府如果能够证明 QRIS 评级系统的准确性,并进一步说明其在提升教育过程中所发挥的效用,那么将更有可能获得联邦政府的财政资助。另一方面,QRIS 评级系统质量评级的准确性将直接影响到家长和早教机构对其的认可程度和参与的积极性。在 QRIS 理想的逻辑假设模型中,家长之所以会选择质量等级高的早教机构,是因为在这里儿童能享受更优质的保教,并且获得更好的发展;而早教机构会积极自愿申请加入 QRIS 的前提之一,也必须是其对等级评定非常信服,能够感受到评级过程和结果的科学与公正。总之,如果质量定级的结果与家长、早教机构负责人和教师们的亲身经验有着较高的匹配度,他们会更相信评级结果,进而提高 QRIS 的整体信誉,这一政策机制会得到更多的支持,走上良性运行的轨道,更好地发挥出应有的功效。

(二)从评价系统内部看,元评价有利于评价系统的改进与优化

元评价不是对评价进行简单的价值判断,它远远超越了“成功”或者“失败”的评判。元评价能找到一个评价系统在设计 and 实施等方面存在的问题,并为改进与优化评价系统提出对策和建议。一方面,目前“循证决策”尚未在教育领域广泛应用,QRIS 的设计决策很大程度上依靠专家的判断<sup>[3]10</sup>,QRIS 评级系统可以说更多是建立在理论推演的基础上。另一方面,理想的评价体系设计落实到实践之中时,不可避免地会出现各种问题,QRIS 评级系统在运行的过程中往往由于评价实施者和其他相关利益者

的原因而出现不同程度失误或偏差。而 QRIS 质量等级的评定是建立在这样的假设基础上:早期教育机构的质量可以被准确地测量,不同早教机构之间的质量差异可以依靠一整套质量指标来辨别。那么,在实践中是否真的能够准确地对早教机构进行质量评级,对 QRIS 来说至关重要。因此,美国各州对 QRIS 评级系统进行有效性验证活动是非常有意义的。

基于实证研究的元评价可以获得大量数据,支持我们去验证一系列问题:质量要素的划分是否合理?质量标准体系是否能够正确反映各质量要素的本质,进而准确地衡量质量?测量工具是否具有好的信效度?测量过程是否高效、精准?质量等级计算形式是否合理?建构的质量等级体系是否具有好的质量区分度?早教机构的质量等级与儿童的发展结果是否呈积极相关?通过实证研究去验证回答上述问题的同时,也是发现并矫正评价系统内部问题的过程。政策制定者将依据元评价提供的证据对评价系统进行改进优化,使其能够更准确地对质量进行测量,能够更有目标和针对性地帮助早教机构去提升质量。例如,在对 QRIS 评级系统进行有效性验证过程中,有研究者发现 QRIS 质量标准体系中所包含的“家庭参与”这一质量要素(quality component)对质量评级没有贡献。因为加入 QRIS 的所有早教机构在“家庭参与”这项质量要素测评中都获得了很高的分数,且“家庭参与”测评分数不存在统计学意义上的显著差异,所以其对于区分早教机构之间的质量差异就没有发挥作用<sup>[4][5]</sup>。QRIS 评级系统对各质量要素进行测评是需要资源投入的,如果某一质量要素对整体的质量评级没有贡献或者不能很准确地对其进行测量,可以删除这一质量要素或者修改测量标准、方式或工具。事实上,通过对评价系统进行有效性验证研究,可以提高测量的效率,降低测量的成本。QRIS 的有效性验证还能表明政策制定者对质量评级与促进系统改革是持开放的态度,愿意积极改进,并致力于提高其达到目标的可能性。所以说,元评价可以帮助评价系统变得更完善。

## 二 美国早期教育质量评价元评价的实施现状

元评价是一个非常复杂的过程,实质上元评价是对评价的准确性和可靠性进行考察。一般来说,原评价中的所有环节(包括评价目的、评价指标体

系、评价过程与方法、评价结果等)都是元评价要研究和评价的对象。美国各州对 QRIS 评价系统元评价工作也是各自为政、分头进行,联邦政府并没有做统一部署与规划。总体来看,各州元评价工作主要涉及对以下四方面进行检验。

### (一)质量要素划分的合理性检验

对教育质量进行测量,要有明确的可操作对象,这需要对其进行因素列举式理解,即将与提升教育质量密切相关的要素罗列出来,称之为质量要素(quality component)。美国大部分州(20个州以上)的 QRIS 都包含有六大质量要素:许可证的合规性、员工资质、课堂环境、家长伙伴关系、行政和管理、资格认证。另外,有一半左右的州还涉及其他三个质量要素:课程、师幼比例与班级规模、儿童评价<sup>[1][3]10</sup>。对 QRIS 的质量要素划分的合理性进行检验,是非常重要的验证活动,因为这是对早教机构进行质量水平定级最初的基础。

当前,美国各州对 QRIS 所含质量要素进行验证的主要方式是通过专家讨论和收集相关实证研究文献。一方面,一些州(包括特拉华州、罗得岛州、明尼苏达州和维吉尼亚州)已经启用了专家评审系统,通过征求专家们的意见来确定 QRIS 的质量评级系统中应该包括哪些质量要素。专家们也比较重视早教机构和家长对质量要素的看法,会让他们对与教育质量相关的因素进行排序,根据其选择的顺序判断他们优先关注的内容,并将此作为提出意见的依据之一。例如,明尼苏达州为了开发针对 QRIS 的评价工具,组织家长进行了焦点小组式的访谈<sup>[3]10</sup>。另一方面,各州通过收集关于现有的质量要素是如何为提高早教机构质量和促进儿童发展做出贡献方面的实证研究文献,来验证 QRIS 的各质量要素在何种程度上与早教机构以及儿童发展的结果联系起来。通过对文献进行深入分析,确定每个质量要素的实证支持水平,这被专家视为对质量要素进行评估的有效方法。文献分析法是当前 QRIS 对质量要素划分是否合理的科学有效的外部评价,其实施的具体过程是对现有质量要素进行“证据审查”。例如,在针对印第安纳州 QRIS 的研究中,若某一质量要素有 1 或 2 个精心设计的实证研究支持其对儿童发展有积极的促进所用,这一要素将被归类为“有一些证据支持”;若有超过 5 个类似的实证研究支持的质量要素则会被归类为“有实质性证据支持”。目前

绝大部分(大约 75%)的质量要素都被认为“有实质性证据支持”<sup>[6]5</sup>。文献分析法通过分析现有的证据为 QRIS 的质量要素划分提供合理性解释,即说明为什么 QRIS 包括某些质量要素而排除了另外的质量要素。文献分析工作也会反复进行,来判断 QRIS 的质量要素划分是否应该根据最新的研究结果进行修改。

## (二)质量指标体系构建的科学性检验

质量指标体系的构建,要建立在对各质量要素进一步分析的基础上。QRIS 中的早期教育质量评价指标体系是指早教机构为了促进早教教育质量提高和儿童获得更好的发展而应该达到的目标集合。通常来说,完整的指标体系至少应包含三层次或三级,QRIS 的质量指标系统也不例外,是由要素(component)、标准(standard)、指标(indicators)三层次组成。其中,作为一级指标的“要素”,是对质量进行广泛的分类;“标准”则是指是二级指标,是对各个要素要达成的目标方向做出精炼的描述;“指标”在此是指三级指标,对二级指标作出了量的规定,也是对其的进一步细化,比较方便被测量或校验。在 QRIS 中,一个质量标准可能有一个或多个质量指标。例如,在员工资质这一质量要素中,标准可能是“教学人员须经过专门培训和资格评定”,与此标准相关的一个指标可能是“师资队伍中至少有 50% 已完成全州指定的两门课程的培训并通过考核”<sup>[3]7</sup>。质量指标体系的制定是教育质量从理论走向实践的重要一环,这是将“教育质量”这一抽象概念转为可操作对象的关键步骤,科学的评价指标体系是保证对早教机构准确质量定级的前提。因此,对质量指标体系科学性进行检验,也是元评价的关键步骤。

美国各州质量指标体系的科学性检验工作主要从以下方面进行。首先,为检验 QRIS 的质量评价标准体系中各项目的区分度,用统计学的方法考查参与评级的早教机构各质量要素分数的分布情况。例如,泽尔曼等人在 2008 年对科罗拉多州早教机构的 QRIS 评级分数进行了评价研究。最初的研究表明,早教机构在“家庭参与”这一项的得分的离散性很低,几乎所有的早教机构都能在这项测评中得到最高的分数,这项要素的测量结果对早教机构的质量等级几乎不会产生影响<sup>[7]</sup>。这或者意味着 QRIS 正在花费大量的精力与资源来测评一项在早教机构之间并没有什么差别的项目,或者说明了“家庭参

与”这一质量要素的测评指标和测评方式存在着问题。总之,尽管导致出现上述结果的原因有待于进一步探究,但当前 QRIS 的质量体系的确存在一定问题。

其次,为检验 QRIS 质量评价指标的相互独立性,采用相关系数分析法对其进行了相关分析。在这样的研究中,相关系数是非常重要的参考值。理想的假设是:由于 QRIS 质量等级评定属于综合评价,其指标体系包含不同性质、不同项目的评价,各项目测评结果之间呈较低相关,这将保证每一项测评都为早教机构最终等级的评定提供了有意义的信息。例如,当前有研究显示师生比与互动质量之间存在着较高的正相关,这表明同时测评这两项并不能为 QRIS 质量定级提供更多的信息,因此有人建议减掉对其中一项的测评<sup>[3]11</sup>;但也有专家指出,QRIS 质量评级系统包含的某些质量要素从测量学的角度来看并不是理想的测评对象,保留其主要目的是引起早教机构和家长对这方面的重视<sup>[3]11</sup>。

## (三)质量评级结果的准确性检验

质量评级结果准确性检验也是美国各州对 QRIS 质量等级评价系统的有效性验证工作中比较重要的一环。这一部分工作涉及到对单项质量要素所得分数以及早教机构质量等级评定是否准确的验证,检验过程是通过考察计分方法与等级计算形式是否科学,进而判断 QRIS 所评定的等级是否能够准确区分不同早教机构质量水平。具体说来,这部分研究主要回答如下问题:采用不同的计分方法与等级计算形式会导致早教机构的质量等级发生很大的变化吗?如果是这样,那么到底哪一种计分方法与等级计算形式更为合理呢?

目前专门针对 QRIS 的计分方法以及等级计算形式(即采用适宜的方法将各单项质量要素的测评结果合并在一起综合判定早教机构的等级)进行的深入探讨有限,但实际上科学合理的等级计算形式对保证质量等级准确评定非常关键,甚至可以说是对早教机构准确定级的最关键步骤。关于 QRIS 的多项研究发现,等级计算形式是影响 QRIS 系统中各个早教机构的等级分布的重要变量<sup>[8]</sup>。当前美国各州在为早期教育机构进行定级时会采用三种等级计算方式:分级制、打分制和两者综合制。分级制是指评级对象在达成某一级别的所有标准后才能晋

升下一级别;打分制则是依据质量标准对各单项质量要素进行打分,然后对各项要素得分加权之后相加算出总分,再根据总分确定评级对象的最终级别<sup>[1]</sup>。例如,托特等人研究发现,采用分级制的等级计算形式,大多数早教机构的质量评级都很低;而采用综合打分制更多的早教机构被评为中等或高质量<sup>[8]</sup>。由此可知,计分方法和等级计算形式的变化确实会影响早教机构在 QRIS 中的等级。鉴于此,各州在对 QRIS 进行有效性验证的过程中已经开始收集和统计相关原始数据,探索改变计分方式以及合并规则对早教机构的质量等级产生什么样的影响。例如,卡洛里和泽尔曼对加利福尼亚州的 QRIS 进行了研究,目的在于确定 QRIS 对早教机构各项质量要素(尤其是课堂环境质量)进行测评时所使用的具体分数分割点是否合理,为修改评分方法提供建议<sup>[9]</sup><sup>45</sup>。与此同时,还有研究关注了不同项目的质量要素与儿童发展之间的相关性,指出计算总分的时候要考虑到不同项目内容的重要性不同,重要项目的权重要增加。例如,越来越多的研究表明,高质量师幼互动与儿童良好发展以及学业准备密切相关<sup>[10]</sup>。因此,各州 QRIS 都非常重视“师幼互动”,将其作为单独的评价要素,并增加其在整个标准体系中的权重。

另外,为了检验评级结果的准确性,同时也是对质量评价方案本身进行评定,一些州对 QRIS 的质量评价结果进行了等值性验证。即出于相同的评价目的,采用不同的评价方案对同一群体进行评价,将得出的结果进行相关分析,研究和判断两种方案的异同。在两种方案具有高度相关的情况下,如果有资料说明其中的一个方案是高质量的,那么可据此推断另一个方案的质量高低<sup>[2]</sup>。例如,南缅因大学的研究者使用当前认可度较高、最为流行的环境等级量表(Environment Rating Scales)对缅因州早教机构的课堂环境进行了测评,缅因州的 QRIS 中并没有将这一量表作为测评工具,研究者主要目的是看不同质量等级的早教机构之间的 ERS 得分是否存在统计学意义上的差异,从而了解 QRIS 质量评价方案与 ERS 之间的相关性,在此基础上对 QRIS 质量评价方案的科学性作出判断<sup>[11]</sup><sup>6</sup>。

#### (四)质量等级与儿童发展结果的相关性检验

对于 QRIS 来说,它的逻辑假设为高质量的保教会使儿童获得更好的发展。因此,非常重要的有

效性验证工作就是:高质量等级的早教机构中的儿童比低质量等级早教机构中的儿童获得了更好的发展。当前,美国各州围绕这方面的研究工作主要是验证早教机构的质量等级以及各质量要素的等级是否与儿童的发展水平成正相关。佛罗里达州的研究者做了系列研究,认为早期教育机构的质量等级与儿童入学准备之间是显著的正相关<sup>[12]</sup><sup>7</sup>。密苏里州的索恩伯格等人研究表明,参加了高质量等级早教机构的儿童在社会性和情感方面比参加低质量等级早教机构的儿童获得了更好的发展,这样的影响对低收入家庭的儿童更加明显<sup>[13]</sup><sup>7</sup>。弗吉尼亚州的研究则表明,比起在低质量等级早教机构就读的儿童来说,接受了高质量等级早期教育的儿童在幼儿园里的前阅读能力发展得更好<sup>[1]</sup>。然而,科罗拉多州的泽尔曼等人的研究发现,质量等级与儿童发展结果之间没有相关<sup>[7]</sup>;印第安纳州的埃利克(Elicker)和明尼苏达州的托特(Tout)等也都发现质量等级与儿童发展结果之间没有一致性关系<sup>[14]</sup><sup>33</sup>。总而言之,当前并没有足够多的证据支持 QRIS 所评定的质量等级与儿童发展的结果呈积极相关这一假设。尽管也有一些州得出了 QRIS 的质量等级与儿童发展呈正相关的研究结论,但这些州的验证性研究设计并不严谨,例如佛罗里达州对儿童发展水平进行评价的是教师而不是第三方评价者。而印第安纳州的研究则是对儿童发展水平数据的收集由独立的第三方来完成,并对儿童的家庭背景进行了严格的控制,其得出的结论是不同质量等级早教机构中的儿童发展水平没有统计学上的差异。针对这一现象,有人认为,这是出现了数理统计中的门槛效应,因为越来越多的低质量早教机构加入 QRIS,它们所得分数也许还没能到达必要的值域来验证质量等级与儿童发展之间的关系;也有人指出,这或许是因为研究设计中缺少家庭与早教机构之间的中介变量<sup>[3]</sup><sup>13</sup>。无论做出何种解释,QRIS 评级结果的有效性遭到质疑,必然会影响公众对 QRIS 的信服度,进而影响其参与 QRIS 的积极性。

### 三 美国早期教育质量评价元评价面临的挑战

尽管美国各州都已启动了 QRIS 有效性验证工作,特别是针对 QRIS 的评级系统开展了一系列研究,但在元评价实施的过程中面临着诸多困难与挑战。QRIS 作为美国早期教育质量监督与保障的政策机制,其有效运行离不开政府、市场、学术团体三

方力量的共同作用。而当对 QRIS 的有效性进行验证时,政府、市场和学术团体三方面均暴露出了一定的问题,给元评价工作的开展带来了一定的困难。

(一)来自政府力量方面的阻力:为维护政府权威讳疾忌医

政府方面作为政策制定者与执行者,其代表的是社会公共利益。政府参与到早期教育质量监督与保障工作中,其主要目的是希望通过一些政策措施使所有儿童都享受优质的早期教育。政策制定者为了赢得政治支持,必须对其所制定政策的价值做出有力论证,并要拿出确凿的证据来证明这一政策完全实现了它的目标。政策在赢得支持的过程中,容不得任何的怀疑和不确定。正如美国各州大力推行的 QRIS,这一政策机制运行需要大量的资源,如不对其进行绩效评估,继续的政策支持和财政支持就没有可靠的依据。但任何评价总是带着一定的风险,在评估进行前,没有人能保证一定会得到一个积极结果,甚至有时得到的结论是某项政策使局面变得更糟糕了。针对 QRIS 的有效性验证也是如此,有可能得到这样的结论:QRIS 的成效并不明显,早教机构的保教质量并未得到提升,儿童也未能获得更好的发展。如此一来,政府之前付出的全部努力都处于付诸东流的危险之中。在某种意义上说,每一项政策后面都隐含着政策制定者的威望与权力,出于对自己权威的维护,政策制定者并不想获得“此项政策低效或者无效”的负面信息。即使政策评价能够帮助他们了解哪些地方需要改进,进而对政策进行修正,最终促使目标更好地实现。以 QRIS 的有效性验证为例,假如对 QRIS 的评价表明这一政策机制包含的某些方法和措施没有很好的发挥作用,那么政府可以重新审视这个政策某些方面,努力尝试改进,使其更加有效地实施运转,而不至于发展到无法控制的局面,使政策被全面否定。但美国各州政府中负责 QRIS 的部分官员们并不想获得关于此项政策措施的任何负面反馈,在这种情况下,“即刻宣布胜利”的气场是强大的,在其位者为了维护自己的权威甚至不惜“掩耳盗铃”<sup>[15]6</sup>。当下一个比较明显的趋势是某些州政府在对 QRIS 的评级系统进行评估时仅限于收集既得利益者的看法与意见,没有设计出一整套科学严谨的评价方案。这样的元评价通常会产生赞美的结果,对问责或政策措施的改进没有任何作用。

(二)来自市场既得利益者的阻力:为维护既得利益不予合作

市场力量的主体是早期教育的提供者,即各类早教机构,市场力量作用的发挥遵循资本逻辑,以逐利性为基本特性,它对早期教育质量的诉求就是追求收益的最大化<sup>[16]</sup>。对 QRIS 的评级系统进行评价可能遭到既得利益者的反对,因为评价的结果可能会动摇或破坏现状。早教机构的负责人已经熟悉了现有的资源分配标准和方式,他们并不太关心当前政策是否有效地解决了问题。各州对 QRIS 评级系统进行元评价,如果评价结果表明当前的物质激励和专业支持不足以使早教机构和儿童产生预期的变化,那么这就会给 QRIS 继续按原有程式推行带来压力。这可能意味着政府会对资源进行重新配置,设计新方案,对早教机构进行重新定级的过程也将被启动。因此,在 QRIS 运行比较成熟的州,既得利益者们不愿意就 QRIS 评级系统有效性验证这一问题进行探讨。在实践中,他们的具体表现为对有效性验证活动并不支持。例如,当想要了解资金奖励对早教机构质量提升的影响,需要早教机构提供财务数据,这恐怕是非常困难的<sup>[15]7</sup>。另外,由于 QRIS 的有效性验证工作是一项大工程,要有一定资金投入。精心设计评价方案并完美实施评价活动的花费是相当大的,早教机构在利益驱使下倾向于将预算高估,政策制定者往往会被早教机构提供的评估预算金额所震惊<sup>[15]7</sup>。一些早教机构如此的立场和举动,对于元评价工作的推行是一种阻力。

(三)来自学术研究过程的诸多限制:对象不稳定、条件不可控、工具受质疑、数据难搜集

对 QRIS 有效性验证离不开学术力量的支持,学术研究结果是对 QRIS 进行评价的客观依据。美国诸多学者对 QRIS 评价系统进行了相关研究,例如在质量标准体系的科学性、质量测评工具的有效性、质量评级结果的准确性等方面都进行了探索,但在实际研究的过程中仍面临着许多困境。首先,针对儿童的纵向追踪研究难以为继。例如,为了研究早教机构的质量等级对儿童阅读准备的影响,必须对儿童进行长时间的追踪。一旦儿童在研究的过程中离开了这家早教机构,那么对其未来发展结果的追踪就是无效的。事实上,由于搬家或父母换工作,儿童在早教机构中中途离开是十分常见的,这就让儿童纵向追踪研究难以为继。其次,针对质量要素

与儿童发展结果的相关性研究挑战性很大。问题一是不能控制所有的无关变量。儿童自身发展具有较强的个体差异性,不同家庭又会为儿童选择不同类型与质量等级的早教机构,选取的样本具有较强异质性,无关变量的控制较为困难。问题二就是测量儿童发展的量表是否具有普适性。美国早教机构中的儿童来自不同的种族,他们成长的文化背景和语言背景具有较大的差异,用一套指标去测量不同种族儿童在认知、情感、动作、社会性等多方面的发展结果并不科学。再次,基础数据的收集存在着困难。在 QRIS 已经比较成熟的州,一些早教机构出于对自己既得利益的保护,并不情愿如实提供相关数据;而在新建立 QRIS 的州,尚需要进行一定时间的积累,才能获得能够用于评价的数据。总之,基础数据的收集工作并不是很好开展。

#### 四 启示

(一)在构建我国的早期教育质量评价体系的同时,即须考虑元评价工作

当前我国并没有类似美国 QRIS 这样的系统性早教教育质量评级与提升政策机制,也没有国家层面制定的早期教育质量评价系统的评价标准,早期教育质量监督和保障工作处于探索期。目前情况是,各省和直辖市教育行政部门基本都制定了本土化的幼儿园质量评价标准,但各省市的评价标准差异较大,而且标准大都是只针对幼儿园,不涉及招收 0—3 岁儿童的早教机构。各省市教育行政部门负责组织对幼儿园进行等级评定,但均未对幼儿园教育质量评价进行元评价。从美国 QRIS 建设以及对其有效性验证的经验中,可以知道从国家层面推行系统性的早期教育质量监督与保障政策措施意义非常重大,这是保证所有儿童接受高质量早教的必要条件。国家性政策推行的首要步骤就是构建统一的早期教育质量评价体系,其中质量标准是质量保障机制运行的起点。对早教机构的教育质量进行测量与评级,必须首先具备一套标准,其次就是制定适宜质量测评工具。因此,构建国家性早期教育质量评价标准与开发普适性的测评工具,应该成为政府工作亟须考虑的问题。在建构国家性早期教育质量评价标准时,可借鉴美国的一些经验。在质量要素划分和指标体系建构时,可参照美国现有的标准体系,对国外相关文献进行分析,同时结合我国的实际情况,注重收集学前教育领域专家学者、早期教育工作

者以及家长等相关利益者的意见。而且在评价体系建立之初,应该选择小部分省市作为试点开展评价工作,根据评价结果对评价指标与工具进行调整修改,保证合理性与科学性。

另外,通过美国 QRIS 的推行及其有效性验证活动的开展考察,可以更加明确地意识到对教育质量评价工作进行元评价的重要性。尤其是当教育质量评价作为一个质量保障政策机制的重要一环时,探索其是否发挥了效用是非常重要的。因为政策的实施往往会涉及政府财政支持,政府与公众都需要确定这些财政支持是否有所收益。持续一个无效政策的成本是巨大的,只要不对政策进行评估,那就没有机会去改进政策或者用一个更有效的措施来代替它。但开展元评价并不是轻而易举的事情,需要资金支持,同时数据的收集也面临着诸多挑战。这说明元评价的进行需要更加完善的组织与监管机制,需要顶层设计。如果在建构早期教育质量评价体系的同时就能考虑到元评价工作的开展,将为后继工作带来极大的方便。就像 QRIS 的设计者如能在实施之前就为后面有效性验证做好安排,为质量评级收集的基础数据也可以为元评价所用,避免额外的数据收集,那将会大大降低元评价成本。例如,在 QRIS 对早教机构提供支持 with 专业援助的过程中,可以收集关于早教机构和儿童发展状况的相关数据。总之,建议我国政府部门应将元评价工作的开展与质量评价工作同时进行考虑。

(二)协调政府、市场、学术三方力量开展元评价,以保证政策效能

国家制定并实施早期教育质量监督与保障政策,根本目的是为了所有儿童均能够享受到高质量的早教,促进儿童更好的发展。为了保证政策的有效性,应该对早期教育质量评价开展元评价。但从美国 QRIS 有效性验证的过程中,可以看到元评价开展在诸多方面都面临着困境。

面对各种阻力,必须强调把儿童的需要与利益放在第一位,而不是优先考虑政策制定者和早教机构的管理者等其他利益相关者自利性诉求。如能将政府行政部门、早教机构以及学术研究团队三方面力量合理协调配合起来,可以使元评价顺利开展,使政策效能得到保证,实现制定与推行政策的根本目的。

首先,元评价工作应该由专业人员组成的第三

方学术团体来负责。元评价工作专业性非常强,其所涉及的利益相关者关系又较为复杂,政府行政人员与早期教育机构管理者都具有自利性,而专家学者远离利益纠纷而又具有高水平的专业知识与能力,因此是承担这一工作的最佳人选。但政府作为公共利益的代表,其依然要发挥主导力量,负责协助第三方评价机构的组建,并赋予其专业自主权,重视元评价的结果。

其次,专家学者在元评价的过程中应考虑到政府、家长和早教机构等相关利益者的主观感受和利益诉求。教育政策不同于经济政策、科技政策等其他公共政策,正如之前所述,其制定和实施的根本目的是为了促进儿童的更好发展。评价指向“人”的政策并不能全部依靠数字标准,元评价实施者要带着人文关怀去倾听相关利益的主观福祉感受与利益诉求。原因之一是儿童的发展结果具有隐效性和长期性,不是全部都能即时量化,还有一些诸如态度、品质等方面的发展并不是很容易测量。这样,当我们

去检验早教机构的质量等级与儿童发展的相关性时,结合定性描述,会使元评价的结果更加客观。原因之二是政府政策制定者与早教机构管理者作为早期教育事业的主要建设者与推动者,他们的主观感受和合理利益诉求是在元评价的过程中应该给予考虑的。元评价目的是为了提提高政策效能,而绝不是为了打消有关工作者的工作积极性。专家学者通过访谈、实地调研等方式,倾听政府行政人员、早教机构管理者们的心声,能够对之前的评价工作有更深入的了解,也能与他们建立起良好的情感联结。而良好人际情感关系的建立,则有助于相关利益者理解并接受元评价的结果,顺利接受政策的重新调整,从而保证不至于因为政策调整而影响到早教机构的精神环境并最终影响儿童发展。

最后,学术团体为开展元评价而进行的数据收集以及后继的研究工作都离不开政府部门的大力支持,政府部门应该投入更多的人力、财力,确保专业评价团队力量的壮大和元评价工作的资金支持。

#### 参考文献:

- [1]刘昆,郭力平,钟晨焰.美国学前教育质量评级与提升系统:实施现状及面临的挑战[J].外国教育研究,2016,(5):110-128.
- [2]侯光文.试论教育评价元评价[J].教育理论与实践,1998,(4):23-26.
- [3]ZELLMAN G L, FIENE R. Validation of quality rating and improvement systems for early care and education and school-age care[EB/OL]. [2016-10-10]. [https://www.acf.hhs.gov/sites/default/files/opre/val\\_qual\\_early.pdf](https://www.acf.hhs.gov/sites/default/files/opre/val_qual_early.pdf).
- [4]ZELLMAN G L, PERLMAN M. Parent involvement in child care settings: Conceptual and measurement issues[J]. *Early Child Development and Care*, 2006,(5):521-538.
- [5]MCGRATH W H. Ambivalent partners: Power, trust, and partnerships in relationships between mothers and teachers in a full-time child care center[J]. *Teachers College Record*, 2007,(6):1401-1422.
- [6]ELICKER J, LANGILL C, RUPRECHT K, et al. Paths to quality, A child care quality rating system for Indiana: What is its scientific base[EB/OL]. (2007-10-23)[2016-10-10]. <https://www.researchconnections.org/childcare/resources/13236/pdf>.
- [7]ZELLMAN G L, PERLMAN M, LE V, et al. Assessing the validity of the qualistar early learning quality rating and improvement system as a tool for improving child-care quality[EB/OL]. [2016-10-20]. [http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND\\_MG650.sum.pdf](http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG650.sum.pdf).
- [8]TOUT K, CHIEN N, ROTHENBERG L, et al. Implications of QRIS design for the distribution of program ratings and linkages between ratings and observed quality[EB/OL]. [2016-10-05]. <https://www.researchconnections.org/childcare/resources/27349/pdf>.
- [9]KAROLY L A, ZELLMAN G L. How would programs rate under California's proposed quality rating and improvement system? Evidence from statewide and county data on early care and education program quality[EB/OL]. [2016-10-05]. [https://www.rand.org/pubs/monographs/2008/RAND\\_MG650.sum.pdf](https://www.rand.org/pubs/monographs/2008/RAND_MG650.sum.pdf).
- [10]KAROLY L A. Validation studies for early learning and care quality rating and improvement systems: a Review of the Literature[EB/OL]. [2016-10-18]. [https://www.rand.org/content/dam/rand/pubs/working\\_papers/WR1000/WR1051/RAND\\_WR1051.pdf](https://www.rand.org/content/dam/rand/pubs/working_papers/WR1000/WR1051/RAND_WR1051.pdf).
- [11]LAHTI M C, SABOL T, STARR R, et al. Validation of quality rating and improvement systems(qris): examples from

- four states [EB/OL]. [2016-10-18]. <http://www.childtrends.org/wp-content/uploads/2015/08/2013-61Validation-of-Quality-Rating-and-Improvement-Systems-QRIS-Examples-from-Four-States.pdf>.
- [12] SHEN J P, TACKETT W, MA X. Second evaluation report for Palm Beach County quality improvement system [EB/OL]. [2016-10-18]. <http://cache.trustedpartner.com/docs/library/000238/ECEQISEvaluation2008.pdf>.
- [13] THORNBURG K R, MAYFIELD W A, HAWKS J S. The Missouri quality rating system school readiness study [EB/OL]. [2016-10-18]. <http://www.elcmdm.org/Knowledge%20Center/reports/MOQRSreport.pdf>.
- [14] ELICKER J G, LANGILL, C C, RUPRECHT K M, et al. Evaluation of “Paths to QUALITY”, Indiana’s child care quality rating and improvement system: final report [EB/OL]. [2016-10-18]. <http://www.state.in.us/fssa/files/PTQFinalReportRev11012.pdf>.
- [15] ZELLMAN G L, BRANDON R N, BOLLER K, et al. Effective evaluation of quality rating and improvement systems for early care and education and school-age care [EB/OL]. [2016-10-18]. <https://www.researchconnections.org/childcare/resources/21555/pdf>.
- [16] 刘昊, 夏婧. 美国学前教育质量保障体系的演变及启示——基于国家、学术、市场“三角协调模型”的分析 [J]. 上海教育科研, 2015, (3): 31-35.

## Meta-Evaluation Research of American Early Education Quality Evaluation Based On Validation Experience

SUN He-qun

(Educational Science School, Tianjin Normal University, Tianjin 300387, China)

**Abstract:** The American Quality Rating and Improvement System is one of the representative strategies to enhance quality of early childhood education. At present, the United States is carrying out the validation of QRIS. The evaluation of the rating system, i.e., the meta evaluation of early education quality evaluation becomes the most important work. The validation includes four complementary and interrelated parts, namely, examination of the validity of key underlying concepts, examination of measurement strategy and the psychometric properties of the measures used to assess quality; assessment of outputs of the rating process; and examination of how ratings are associated with children’s development. QRIS meta-evaluation is also faced with many challenges. Drawing on the experience of the United States, China should construct a universal quality evaluation system of early education, take the meta-evaluation design into account, and coordinate government, market and academic strength to carry out meta-evaluation to ensure policy effectiveness.

**Key words:** the United States; early childhood education; quality evaluation; meta-evaluation

[责任编辑:罗银科]