

中小学教师薪酬激励项目对学生学业成绩的影响效应：基于实验和准实验研究的元分析

赵冉, 王艺博, 杨靖, 杜育红

[摘要]基于委托代理理论和团队生产理论的分析框架,利用33篇基于实验或准实验设计的教师薪酬激励项目研究中的312个效应量,运用元分析方法,剖析中小学教师薪酬激励项目对学生学业成绩的影响及调节效应。结果显示,中小学教师薪酬激励项目对学生学业成绩总体上存在负面影响,中学实施的教师薪酬激励项目效果要优于小学,但存在调节效应。在调节效应中,团体激励在小学教师中效果更好,个体激励在中学教师中效果更好;薪酬激励对数学学科有积极影响,对语言阅读学科有消极影响;发达国家或地区实施的教师薪酬激励项目整体起到消极影响,而在发展中国家或地区则表现出积极影响;激励金额多少对薪酬激励项目的实施效果无显著影响。我国基础教育阶段在进行教师绩效工资制度设计时,应充分关注教师工作所处的学科学段、地区经济发展水平等背景性特征,综合激励形式和强度科学设计教师绩效激励。

[关键词]教师薪酬激励; 学生学业成绩; 元分析

一、引言

世界各国普遍采用基于受教育水平和教龄的教师固定薪酬支付制度,但

[收稿日期] 2025-01-06

[基金项目] 北京师范大学教育学一流学科培优项目“新时代中国各级各类教育经费投入与体制机制创新研究”(YLXKPY-XSDW202208);北京师范大学教育学一流学科培优项目“高水平教育现代化目标下教育经费投入总量与结构的优化路径研究”(YLXKPY-XSDW202406)。

[作者简介] 赵冉,北京师范大学教育学部/首都教育经济研究基地,电子邮箱地址: zhaoran@bnu.edu.cn; 王艺博,北京师范大学教育学部,电子邮箱地址: 202321010159@mail.bnu.edu.cn; 杨靖(通讯作者),成都师范学院教育科学学院,电子邮箱地址: 201523010066@mail.bnu.edu.cn; 杜育红,北京师范大学教育学部/首都教育经济研究基地,电子邮箱地址: dyh@bnu.edu.cn。

众多学者认为这是一种错位的薪酬制度，不能奖励更好的绩效和激励教师努力，在私营部门取得巨大成效的薪酬激励项目(绩效工资项目)通常被认为是改善现行教师薪酬制度的有效工具(Murnane and Cohen, 1986; Hanushek and Jorgenson, 2000)。20世纪80年代开始，教师薪酬激励项目相继在世界各国的部分地区进行试点，在美国最具代表性的包括：路易斯安那州等地区实施的教师提升项目(Teacher Advancement Program, TAP)、田纳西州的教师职业阶梯项目(Career Ladder Program, CLP)和丹佛市的教师职业薪酬制度(ProComp)等。同时，也有众多非政府组织向肯尼亚、印度、坦桑尼亚等发展中国家部分地区的学校和教师提供薪酬激励项目。

理论上，当厂商和工人建立委托代理关系时，代理人关心的是自己付出努力的成本，而委托人关心的是代理人努力带来的收益。当委托人无法监控或因监控成本太高，难以获知代理人行为和关于成本的信息时，代理人具有了私人信息，出现信息不对称的情况(Bradley and Green, 2020)。此时，如果双方目标不一致，代理人会出现两种隐藏行为——道德风险和逆向选择，降低代理效率(Laffont and Martimort, 2009)，这就需要委托人和代理人之间建立某种机制(契约)，通过价格机制来提高代理效率，薪酬激励便是行之有效的方式。在私营部门的绩效工资项目设计中，通常会设计组织需要的绩效目标，并对超过预定绩效标准的代理人按相应计件工资率支付绩效工资，提高代理人出现道德风险和逆向选择的成本，从而通过价格机制将双方目标统一起来，实现薪酬的激励功能，提升组织绩效。

但由于与私营部门生产背景不同，在教育行业实施薪酬激励是否适宜一直存在争议。反对者认为以利润动机为导向的绩效工资与教育部门所需的公共服务精神不兼容(Kellough and Nigro, 2002; Lewis et al., 2009; Bregm, 2013)，提供经济激励会损害教师的工作动机和行为规范，提高绩效工资项目的隐藏成本。另外，教师绩效工资的锚定标准应是学生未来的工资收益，但学生成年后的工资收益需要数年以后才能观测，将教师绩效工资与学生收入挂钩几乎是不可能的，教师绩效工资与生产利润难以产生联系(Lavy, 2007)。最后，在学校现实环境中，教师工作具有多维特征，组织目标和员工任务是不容易定义的，也难以对个别教师或教师群体所提供服务的价值进行衡量(Podgursky and Springer, 2007)。而支持者则认为，教师年复一年重复着相同的工作，单纯依靠内生动机维持其对工作的热情和承诺是难以想象的，向教师提供薪酬激励是改善教师质量的必要措施(Frey and Jegen, 2001; Asch, 2005)。

在实践环节，教师薪酬激励项目对学生学业成绩改善效果的证据也并不

一致。部分基于回归分析的实证研究支持了教师薪酬激励项目对学生学业成绩的积极影响(Springer, 2009);但部分采用随机实验设计或准实验设计的研究发现教师薪酬激励项目的实施不仅没有提高学生的考试平均分,甚至起到了降低学生出勤率的反作用(Dee and Keys, 2004)。由此,部分学者认为教师薪酬激励的效果取决于其实施背景和设计,实施学段、激励方式、激励强度和激励对象都会对其实施效果造成影响,教师薪酬激励项目对学生学业成绩的影响可能具有结构效应(Neal, 2011; Pham et al., 2021)。已有随机试验和准随机试验研究多基于单一教师激励项目进行研究,缺乏对结构因素影响效应的探讨。本研究基于元分析方法,一方面,对教师薪酬激励项目和学生学业成绩相关文献的实证研究效应量进行整合分析,提供一个通用的方式来表征教师薪酬激励项目对学生学业成绩的实际影响力;另一方面,从教师薪酬激励项目特征的角度深层次检视,探析教师薪酬激励项目影响调节效应的来源,并验证教师薪酬激励项目特征对教师薪酬激励项目影响综合效应的调节效果;运用漏斗不对称检验法验证已有研究中存在发表偏误问题,进而在一定程度上检验教师薪酬激励项目对学生学业成绩的影响实证结果的可靠性与适用性。

二、理论框架

(一)委托代理理论

在委托代理理论的分析框架下,薪酬与个体能力、努力联系在一起,主要通过选择效应(selecting)和激励效应(incentive)两种机制来解决代理人的逆向选择和道德风险问题,已有教师薪酬激励研究主要关注其中的激励效应。在学校生产环境下,一方面,教师招聘依赖学历信号的筛选机制,对教师实际教学能力的预测效度存在偏差。并且,教育活动的多维性导致教师能力评估具有长期滞后性。在此情境下,固定薪酬制度易诱发高能力教师的逆向选择行为(Gibb, 1996; Kwon et al., 2006)。另一方面,教师绩效由努力强度、时间配置及教学策略三要素决定(Balapumi et al., 2016)。由于教师个体对工作质量的认知常与教育管理部门存在价值分歧,教学活动又具有高度情境依赖性,监督成本高昂,导致教师努力配置易偏离组织预期,带来道德风险问题(Neal, 2011)。由此,施行以结果为导向的、计件形式的绩效薪酬,理论上是激励教师揭示隐藏信息,有效分配努力水平,提高学校期望目标绩效的有效措施。

但是教育领域的绩效薪酬项目难以满足委托代理理论的一些基本前提。

首先,教师绩效薪酬缺少利润分享机制。教育作为准公共产品缺乏市场化利润分配基础,薪酬激励无法与教育价值创造直接关联,不存在利润分享的可能(Lavy, 2007)。其次,教师绩效难以衡量,高努力对应的产出概率分布难以满足一阶随机占优条件。教学本质假设认为教师的绩效较其他职业更难监控,受限于绩效评估的准确性、可靠性与有效性,难以为教师提供有效的教学改进信息,教师的高努力水平并不一定带来理想的绩效水平(Murnane and Cohen, 1986; Podgursky and Springer, 2007; Phillips et al., 2009)。最后,教师的风险规避程度较高,绩效薪酬的成本效益低。大众普遍认为教师之所以选择这一职业是因为其稳定性,具备较强的风险厌恶属性,绩效薪酬实际是在向员工转嫁风险,员工风险规避程度越高,就需要设计越高的计件工资率,教师的风险和激励权衡倾向会使其缺乏效用(Gelfand et al., 2011; Dang et al., 2017)。

同时,在多委托代理分析框架下,薪酬激励的效果取决于目标绩效向量与测量绩效向量的对齐性(Neal, 2011),并且,当教师同时承担存在时间配置替代性的量化任务和质量任务时(如提升学生学业成绩和专业发展),计件工资率设计需均衡两类任务的效用成本,数量与质量激励强度需遵循严格对应原则(Holmström, 2017)。在教育环境下,绩效测量维度多指向量化任务(学生学业成绩),教师易采用努力边际成本低的应试教学途径提升学生学业成绩,忽视努力边际成本高的专业发展途径,面临激励错位和操纵的风险(Baker, 2002)。但 Muralidharan 和 Sundararaman(2011)也指出,这种情况受学校系统教学技术条件与有效教学基准的双重调节,当教学标准较低,两类努力未形成替代时,虽存在激励操纵行为,教师绩效薪酬也能显著提升低投入教师群体的表现,形成帕累托改进效应。教师绩效薪酬的激励效果受到设计特征、受众群体风险偏好、教育发展程度以及激励强度方面的约束。

(二)团队生产理论

团队生产理论的核心假设在于协同效应——即团队整体产出超越个体产出的简单线性叠加,其效率源于专业化分工、知识共享与监督互惠三者的动态耦合。专业化分工通过成员聚焦比较优势任务提升局部效率,知识共享实现隐性经验与显性信息的跨主体流动,监督互惠则借助非正式约束抑制道德风险。但是,团队生产的不可分割性导致个体边际贡献难以精确测度(Alchian and Demsetz, 1976),由此引发双重困境:其一,搭便车效应随团队规模扩大呈指数级增长,成员合作承诺概率急剧衰减(Sen, 1977);其二,个体理性选择迫使系统陷入低效均衡——努力投入与收益补偿持续错配。Lazear(2000)由此提出激励阈值条件:只有当团队协同产生的净收益显著超

越个体激励弱化引致的效率损失时,团队生产模式方具备帕累托改进的可行性。

在众多学者看来,教师很大程度上是作为团队成员工作的,教师的教育生产函数并不一定具备可加分离性,并且在学生成绩的生产中,教师之间具有重要的互补性(Baker, 2002)。继而有学者认为,在教师个体层面引入绩效薪酬激励可能会减少教师合作的动力,从而降低学校整体的绩效表现;同时,教师个体层面的绩效薪酬激励也使教师承担额外的风险并迫使教师工资成本增加,从而使其失去效率(Podgursky and Springer, 2007)。而团队激励适合教师教育生产函数的非可加分离性,教师专注于自身学科的教学,并协同促进学生学业成绩的发展。考虑到教师工作通常是在较小的团队中进行,不同科目的教师间容易形成关于学生学业成绩发展信息的互动,并且易于相互监督,团队中个体教师的努力容易被同伴观测到,形成合作承诺,团队激励更适宜于教师行业。但这一理论假设也受到部分学者的怀疑,特别是在教师教育生产函数的非可加分离性上,教师绩效薪酬的激励效果受到不同学段和学科教学生产方式的约束。

基于理论框架的分析表明,教师绩效薪酬的激励效果受区域教育水平、学段学科特征等多维背景因素与制度设计参数的交互影响。传统单一情境研究范式难以揭示其非线性作用机制与情境异质性,而通过元分析方法整合跨研究的效应量异质性,可系统解析绩效薪酬对学生学业成绩的净因果效应,进而识别关键调节变量的差异化影响路径,对当下我国中小学教师的绩效工资制度改革具有重要的指导意义。

三、研究设计

现有关于教师薪酬激励项目与学生学业成绩关系的综述和实证研究非常丰富,但鲜有运用元分析(Meta-analysis)方法对这类实证研究进行数量化的系统分析。元分析(荟萃分析)是基于实验与准实验研究设计的实证研究文献,依托某一可比特性,探究研究间异质性的来源及大小,可有效度量由于模型设定、统计方法及研究数据集等特征差异对研究结果不一致的影响程度(Weichselbaumer and Winter-Ebmer, 2005)。元分析方法克服了传统文献综述的缺陷,可以回答单个研究未提出的问题,并且有助于甄别在发表过程中可能存在的发表偏误问题,以提高研究的精度;通过运用系统全面的量化方法解释并综合相关文献中的某个参数,还在一定程度上回避了原始文献可能存在的选择性偏误等问题(赵冉等, 2022)。因此,针对现有研究中教师薪酬

激励项目对学生学业成绩作用方向及大小的争议，本研究比较分析了现有实证文献的结果差异，利用元分析方法，借助元分析软件 CMA 3.3，深入探讨教师薪酬激励项目与学生成绩之间关系的实证研究结果是否受到具体项目特征的影响。

(一)文献检索与筛选

文献检索分两轮进行。第一轮为数据库检索，通过 Web of Science 数据库、谷歌学术等外文数据库对 1990—2022 年篇名中含“performance pay”或“merit pay”且含“achievement”词语的文献进行检索。第二轮为手动检索，通过如下的具体的筛选标准对文献进行筛选：该研究样本为基础教育阶段(K—12)的教师和学生；教师所在的学校、地区、州或国家实施了教师薪酬激励项目；该研究报告了学生在标准化考试中的成绩；该研究采用随机对照实验或准实验设计。

此外，考虑到研究结果应具有可比性，第二轮检索中排除了以学生考试通过率指标作为学生变量的研究，仅保留报告学生考试成绩的研究。综合上述分析，最终本研究选取 33 篇教师薪酬激励项目实证研究作为元分析的主要样本。文献筛选的流程见图 1。

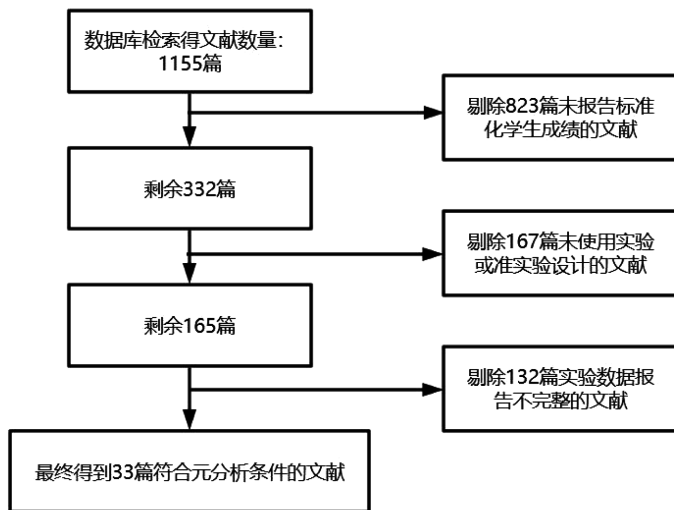


图1 文献筛选流程

(二)变量设定与编码

为便于分析，由研究者对所有样本研究的变量进行独立编码。编码内容包括：计算教师薪酬激励项目影响效应所需的变量、教师薪酬激励项目特征相关变量和教师薪酬激励项目研究特征相关变量三部分。

基于在文献回顾部分提到的已有评估教师薪酬激励项目研究主要争论点，对一系列先验变量进行编码，将教师薪酬激励项目特征作为调节变量。具体包括教师薪酬激励项目实施学段和实施学科的交互、教师薪酬激励项目实施学段和激励形式的交互、激励金额以及研究样本所在地区的交互。此外，为进一步从研究特征分析教师薪酬激励项目对学生学业成绩的调节效应，对发表类型和研究样本所在地区进行编码。表 1 显示了本研究对变量的编码结果。

表 1 变量编码

变量名	样本量	效应量均值	95%置信区间
因变量			
学生学业成绩			
学生数学成绩	167	0.665	[0.092, 0.268]
学生语言阅读成绩	145	-0.323	[-0.271, -0.159]
自变量			
实施了教师薪酬激励项目	312	/	/
分组变量			
实施学段和实施学科交互			
教师薪酬激励项目在小学阶段的数学学科实施	65	0.486	[0.211, 0.667]
教师薪酬激励项目在小学阶段的语言阅读学科实施	59	-0.200	[-0.024, 0.140]
教师薪酬激励项目在中学阶段的数学学科实施	37	1.697	[0.706, 1.638]
教师薪酬激励项目在中学阶段的语言阅读学科实施	28	0.053	[0.030, 0.077]
实施学段和激励形式交互			
教师薪酬激励项目在小学阶段为团队激励	31	0.269	[0.212, 0.325]
教师薪酬激励项目在小学阶段为个人激励	86	0.127	[-0.010, 0.319]
教师薪酬激励项目在中学阶段为团队激励	37	0.058	[-0.085, 0.201]
教师薪酬激励项目在中学阶段为个人激励	34	1.845	[0.805, 1.810]

续表			
变量名	样本量	效应量均值	95%置信区间
激励金额			
教师激励金额低于或等于中位数	140	-0.052	[-0.071, 0.013]
教师激励金额高于中位数	117	-0.115	[-0.127, 0.026]
样本地区			
研究样本位于中国	10	0.161	[0.095, 0.225]
研究样本位于发达国家或地区	221	0.165	[-0.298, -0.174]
研究样本位于其他发展中国家或地区	81	0.322	[0.282, 0.355]
发表变量			
期刊文献	189	0.126	[0.156, 0.208]
非期刊文献	123	0.328	[-0.223, 0.070]

注：效应量均值报告了教师薪酬激励项目干预效果的均值；分组变量的“效应量均值”报告了当样本限制在某一特征时教师薪酬激励项目干预效果的均值。95%置信区间为随机效应模型下加权合并效应量的置信区间。表中报告的效应量均值是未加权均值，置信区间是基于随机效应模型，根据研究样本量对效应量进行了加权。受效应量间高异质性的影响，加权均值可能偏向大样本研究，而未加权均值反映“原始分布中心”，两者可能存在偏离，导致效应量均值超出加权置信区间。

(三)分析策略

1. 教师薪酬激励对学生学业成绩的影响效应

为了计算教师薪酬激励项目对学生学业成绩影响的综合效应，需要合并研究中报告的不同类型的效应量。标准化平均差 (Standardized Mean Difference, SMD) 是用以表征教师薪酬激励项目对学生学业成绩的影响程度的效应量，教育研究中常用的 SMD 有 Cohen's d、Glass's Δ、Hedges'g 等；经过比较，最终采用 Hedges'g 作为效应值指标，以校正 Cohen's d 产生的小样本效应值估计偏差 (Hedges, 1981)。

根据 Hedges 和 Olkin(1985)提出的标准化均差 SMD 计算方法，用模型 (1)来生成每个样本的效应量，并通过模型 (2)计算教师薪酬激励项目对学生学业成绩的综合影响效应。

$$d_i = \frac{\bar{x}_{1i} - \bar{x}_{2i}}{\sqrt{\frac{(n_{1i}-1)S_{1i}^2 + (n_{2i}-1)S_{2i}^2}{n_{1i} + n_{2i} - 2}}}, i=1, 2, 3 \cdots k \quad (1)$$

$$\bar{d} = \frac{\sum n_i d_i}{\sum n_i} \quad (2)$$

其中,第 i 项研究的实验组和对照组的样本量分别为 n_{1i} 和 n_{2i} , 均值分别为 $\bar{x}_{1i}$ 和 $\bar{x}_{2i}$, 标准差分别为 S_{1i} 和 S_{2i} , 自由度为 df 。权重系数 n_i 表示样本量之和($n_i = n_{1i} + n_{2i}$)。

一般地,固定效应模型假设所有研究的真实效应大小相同,而随机效应模型将效应大小视为以平均值为中心的分布(Borenstein et al., 2021)。基于在不同研究和项目特征的基础上对教师薪酬激励项目影响的综合效应进行推断的研究需要,固定效应模型的等方差假定不太符合现实情况,因此使用随机效应模型。

2. 教师薪酬激励对学生学业成绩的影响的调节效应

基于对文献的回顾结果,假设教师薪酬激励项目对学生学业成绩的影响效应存在异质性,并通过评估教师薪酬激励项目的不同特征的影响大小如何变化,尝试阐明围绕教师薪酬激励的一些政策辩论。本研究并未考虑使用元回归方法来探索教师薪酬激励的影响如何因项目特征而异,原因是 meta 回归的样本量是基于研究的数量,因此认为多变量 meta 回归是不合适的,33 篇研究的样本量意味着分析的把握度严重不足,可能会产生仅由少数研究驱动的回归系数。鉴于这些问题,正如 Brant 等(2010)所建议的那样,选择了亚组分析作为分析教师薪酬激励对学生学业成绩的影响的调节效应的方法,描述当将样本限制在具有某些特征的研究时,教师薪酬激励影响的综合效应如何变化。其中,在对教师薪酬激励项目的激励强度上,使用教师平均激励薪酬与激励项目实施时间内,区域教师平均薪酬的百分比进行处理,并根据金额是否高于中位数进行分组。

四、实证分析结果

(一)教师薪酬激励对学生学业成绩的影响效应

1. 发表偏误检验

一般来说,能够得出显著性的实证研究结果更容易被发表,这种现象对

元分析样本选择造成的影响叫做发表偏误(publication bias)，发表偏误较大
会影响元分析的可靠性。为评估和检验元分析的发表偏误，需要结合绘制效
应量样本的发表偏误漏斗图(funnel plots)、失安全系数(fail-safe N)和 Egger
线性回归检验(Egger et al，1997)等多种方法。

漏斗图是以纳入元分析中的教师薪酬激励项目对学生学业成绩影响效应
量为横坐标，以标准误为纵坐标，绘制而成的散点图，根据图中散点的对称
性可以判断纳入元分析的文献是否存在发表偏误；如果不存在发表偏误，则
漏斗图中的点会以真实值为轴对称分布，对称的程度越高，发表偏误就越小
(Egger et al.，1997)。失安全系数用于估计未发表文章中重要效应值对研究
结果的影响；当失安全系数大于 $5k+10$ (k 为样本数)时，表明即使存在发表
偏误，结果依旧较为稳健(Egger et al.，1997)。Egger 法通过构建线性回归
方程检验发表偏误，若结果显示的截距值为零，且检验结果不显著($p>0.05$)，
则说明元分析不存在发表偏误；反之，量化结果则揭示了发表偏误的存在
(Egger et al.，1997)。

表 2 报告了失安全系数和 Egger 法的检验结果，元分析的效应量样本数
为 312，失安全系数为 4924(大于 1570)，总体效应量的 t 值为 0.136($p>$
 0.05)。根据量化分析结果，元分析的样本不存在发表偏误。从漏斗图的检验
结果可以看出，元分析的样本基本均匀分布于中上部及总效应量两侧，但有
部分研究位于漏斗图的右下侧。在此基础上，采用剪补法(Trim and Fill)对
样本进行增补(Duval and Tweedie，2000)，纳入 73 个虚拟样本后重新对所
有样本进行合并分析(见图 2)，剪补前后元分析的结果保持了一致，揭示本
研究的元分析结果具有较好的稳定性。综上可以认为，教师薪酬激励项目元
分析的研究样本中，存在发表偏误的可能性较小，分析结果具有可靠性和稳
健性。

表 2 发表偏误检验

	样本量	失安全系数	Egger 回归	
			t 值	P 值
总体	312	4924	0.136	0.892
数学成绩	167	7157	0.435	0.664
语言阅读成绩	145	7949	0.805	0.422

注：*、** 和***分别表示 5%、1%和 0.1%的显著性水平，如无特别说明，以下各
表同。

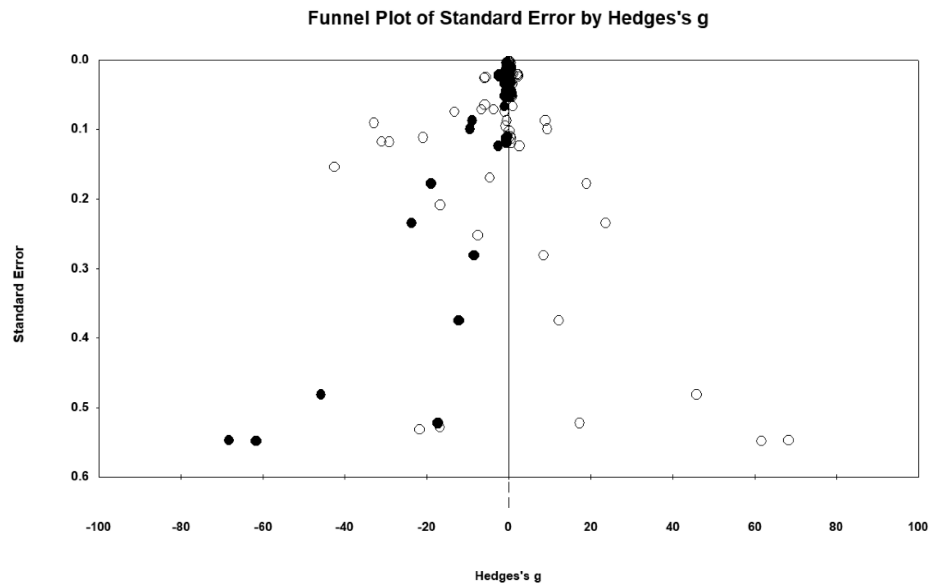


图 2 发表偏误漏斗图(剪补法)

2. 总体效应检验

表 3 报告了教师薪酬激励项目对学生学业成绩影响的随机效应估计。结果显示总体效应估计值为 $-0.084(p<0.01)$ ，说明教师薪酬激励项目对学生学业成绩存在负向影响。在分学科的估计中，教师薪酬激励项目对学生的数学成绩和语言阅读成绩的影响效应估计值分别为 $0.180(p<0.001)$ 和 $-0.215(p<0.001)$ 。教师薪酬激励项目对学生数学成绩存在显著的正向影响，对学生语言阅读成绩存在显著的负向影响，这表明，教师间的教学工作存在非可加分离性，教师薪酬激励易破坏教师教学原生的团队生产方式。结合激励措施与考试成绩挂钩更易鼓励教师只关注测试内容或应试技巧的研究发现(Pham et al. , 2021; Glewwe et al. , 2010)，以及众多教育随机干预实验中更易对数学学科产生影响的特点，这一结果可能是因为基于学生学业成绩的教师薪酬激励更易使学校或教师通过强化对数学学科测试内容或应试技巧的训练来提高绩效水平。然而，这种强化策略破坏了各学科教师之间原有的教学生产方式，挤压了其他学科的教学投入，从而总体上对学生学业成绩造成负面影响，并尤易在语言与阅读学科中表现出负面效应。

表 3 总体效应分析结果

	样本量	SMD	SE	异质性检验		
				I ²	τ	Q
总体	312	−0.084**	0.025	99.958	0.445	311***
数学成绩	167	0.180***	0.045	99.967	0.574	166***
语言阅读成绩	145	−0.215***	0.028	99.940	0.337	144***

3. 异质性检验

元分析样本研究间异质性的检验需要结合 I²、τ 和 Q 三个统计量综合判断。I² 代表由于异质性而非偶然性导致的研究总变异，表 3 中 I² 估计值均为 99 以上，可以被归类为“高度”异质性。τ 表示研究中真实效应大小参数分布的估计标准差，I² 和 τ 的值共同表明随机效应估计值在样本研究中并没有广泛分布，但大多数变化是由研究效应的异质性而非偶然性来解释的。统计量 Q 被经典地用于检验研究中同质性的零假设，Q 的估计值均达显著水平 (p<0.001)，说明样本研究的随机效应估计值不存在同质性。同时，这三个统计量提供了进一步研究教师薪酬激励对学生学业成绩影响的调节效应的证据。

(二)教师薪酬激励对学生学业成绩影响的调节效应

根据文献回顾结果，采用亚组分析的方法分别从教师薪酬激励项目的项目特征和研究特征来探究中小学教师薪酬激励对学生学业成绩影响的调节效应。表 4 报告了教师薪酬激励项目对学生学业成绩影响的调节效应的亚组分析结果。

表 4 调节效应分析结果

调节变量	类型	样本量	SMD	SE	组间异质性(Q)
实施学段和 实施学科 交互	小学数学	65	0.439***	0.116	78.288***
	小学语言阅读	59	0.058	0.042	
	中学数学	37	1.172***	0.238	
	中学语言阅读	28	0.053***	0.012	
实施学段和 激励形式 交互	小学一团队	31	0.268***	0.029	130.086***
	小学一个人	86	0.155	0.084	
	中学一团队	37	0.058	0.073	
	中学一个人	34	1.308***	0.256	

续表					
调节变量	类型	样本量	SMD	SE	组间异质性(Q)
激励金额	低于或等于中位数	140	-0.029	0.021	26.602***
	高于中位数	117	-0.051	0.039	
样本地区	中国	10	0.160***	0.033	229.153***
	发达国家或地区	221	-0.236***	0.031	
	其他发展中国家或地区	81	0.318***	0.019	
发表类型	期刊文献	189	0.182***	0.013	11.573***
	非期刊文献	123	-0.076	0.075	

1. 实施学段和实施学科交互

从表 4 中可知，小学阶段教师薪酬激励项目对学生数学成绩影响效应估计值为 0.439($p<0.001$)，而小学阶段教师薪酬激励项目对学生语言阅读成绩的影响效应在统计学意义上不显著。中学阶段教师薪酬激励项目对学生数学成绩和语言阅读成绩的影响效应估计值为 1.172 和 0.053($p<0.001$)，这说明教师薪酬激励项目对不同学科的影响效应在学段层面上存在异质性。

2. 实施学段和激励形式交互

从学段层面来看激励形式，在小学阶段，当激励对象为团队时，教师薪酬激励项目对学生学业成绩产生显著的积极影响，效应值为 0.268($p<0.001$)，而当激励对象为个人时，教师薪酬激励项目对学生学业成绩的影响效应不显著。在中学学段则恰好相反，当激励对象为个人时，教师薪酬激励项目对学生学业成绩产生显著的积极影响，效应值为 1.308($p<0.001$)，而当激励对象为团队时，教师薪酬激励项目对学生学业成绩的影响效应不显著。这可能与小学和中学阶段教师教学特点存在差异有关。小学学段学生学习科目较少，班级任课教师的团队规模通常较小，符合小团队生产易于形成知识互补、相互监督和合作承诺的团队生产特点，实施团队激励更易促进学生的学业成绩发展。而中学班级教学中涉及的学科较多，教师团队规模更大，教师间更倾向于以个体生产的方式进行教学，教师间不易形成知识互补和相互监督，并且以学校或年级为单位设计的团队激励项目还易产生“搭便车”问题，因此实施个体激励更能促进学生的学业成绩发展。

3. 激励强度

表 4 的估计结果显示，激励金额是否高于中位数都没有对学生的学业成绩造成显著影响。教师薪酬激励的价格机制并没有在提升学生学业成绩方面得到体现，并且估计结果显示激励金额无论是高于中位数还是低于中位数，

都在一定程度上对学生学业成绩起到了负向影响，但并不具备显著性。由于教师薪酬激励项目的激励金额设计并不遵循市场机制，与生产利润并不产生联系，委托代理理论中，通过提高代理人出现逆向选择和道德风险成本来提高代理人努力的价格机制作用并不适用于教师行业。

4. 样本地区

由表4可知，在发展中国家或地区实施的教师薪酬激励项目对学生学业成绩的促进作用上要明显优于发达国家或地区，在发展中国家或地区实施的教师薪酬激励项目对学生学业成绩总体上起到了显著的促进作用，影响效应为0.318($p < 0.001$)，而在发达国家或地区实施的教师薪酬激励项目对学生学业成绩总体上起到的是消极作用，影响效应为-0.236($p < 0.001$)。这可能是由于发达国家或地区的教师薪酬激励项目的绩效指标设计通常包含学生学业成绩增值评价、教师专业发展和教学实践评估等教师教学绩效的高阶维度，而发展中国家或地区的教师薪酬激励项目的绩效指标设计通常包含的是学生的出勤率、学业成绩达标率和教师的出勤率等指向基本规范性的维度。由于发达国家或地区的教育发展水平较发展中国家或地区更高，教师工作要求、规范性和专业化程度都较其他地区处于更高的水平，致使发达国家或地区的教师在各项绩效指标下努力间存在替代性，调整努力方向的边际成本通常较高，薪酬激励对改变教师效用水平的作用有限。为了给我国教师薪酬激励项目提供本土化的实证证据，本研究将中国样本单独进行分组，进一步检验教师薪酬激励影响在不同样本地区的异质性，中国样本中教师薪酬激励项目对学生学业成绩影响的效应值是0.160($p < 0.001$)，可见，在我国实施的教师薪酬激励项目显著地促进了学生学业成绩的提升。但是从效应值大小来看，我国教师薪酬激励项目需进一步借鉴发达国家和其他发展中国家相关政策的经验和教训，因地制宜发挥我国教师薪酬激励项目的制度优势。

5. 发表类型

在期刊文献和非期刊文献的亚组分析中，期刊文献中教师薪酬激励项目对学生总体学业成绩的影响效应值为0.182($p < 0.001$)，而非期刊文献中教师薪酬激励项目对学生总体学业成绩的影响效应值为-0.076。这表明虽然估计结果通过了相关发表偏误的检验，但“结果显著的研究更容易被发表”的现象对元分析效应值的影响依然存在。

五、结论与讨论

研究基于33项教师薪酬激励实验或准实验研究文献的312个效应量，运

用元分析对教师薪酬激励与学生学业成绩的影响关系展开了探究。研究结果表明,教师薪酬激励总体上对学生学业成绩呈消极影响,但在学科间存在差异,在数学学科上对学生学业成绩起到了积极影响,而对语言阅读学科起到的是消极影响。对其影响的调节效应进行探析发现,在中学实施的教师薪酬激励项目和在小学阶段实施的教师薪酬激励项目在不同学科表现不同,总体来看在数学学科实施的教师薪酬激励项目相比于语言阅读学科实施的教师薪酬激励效果更好。在团队和个体激励设计方面,小学阶段实施团体激励相比于个体激励能表现出更大的积极作用,中学阶段则正好相反。激励强度方面,激励金额高于或低于中位数都没有对学生学业成绩造成显著影响。区域差异方面,在发达国家或地区实施的教师激励项目整体上呈现出消极作用,而发展中国家或地区则整体上呈现出积极作用。发表类型方面,得到显著积极影响效应结果的研究更易发表。基于实证结果和理论框架,研究得出如下启示。

(一)探索教师非货币激励形式,激发教师内生动机

从上述教师薪酬激励项目总体上对学生学业成绩的消极影响,以及激励金额对学生学业成绩无显著影响可以看出,由于教师薪酬激励项目的激励金额设计并不遵循市场机制,与生产利润并不产生联系,委托代理理论中,通过提高代理人出现逆向选择和道德风险成本来提高代理人努力的价格机制作用并不适用于教师行业。主导教师薪酬激励项目实施效果的机制可能更符合动机挤压理论(Motivation Crowding Theory)的解释:人们普遍认为教师选择教师职业是因为其对教师职业的高度认可,并且是具有风险厌恶、公平偏好的个体。而作为外部干预的薪酬激励会使其承担额外的风险,改变其动机因果关系点,严重情况下甚至对个体动机产生消极影响(Frey and Jegen, 2001),并且过度聚焦于价格机制可能对教师行业的规范产生挑战(Ryan and Deci, 2000)。由此,教师薪酬激励设计只有在与教师内生动机产生挤入效应,即激励薪酬使教师感受到内生动机驱动的努力得到充分肯定时,才能实现其激励功能。广义的教师薪酬分为货币形式和非货币形式,除了货币激励以外,还有专业发展、社会荣誉、工作自主权等激励内容。相比于完全货币形式的教师薪酬激励,将货币激励与非货币激励形式相结合不失为一种激发教师内生动机、提升教师薪酬激励成效的有效方式。在基础教育阶段进行教师绩效工资设计时,可充分借鉴美国部分教师薪酬激励项目中职业阶梯计划的设计,并与我国当下教师职称岗位晋升制度充分结合,将

货币形式的薪酬激励和非货币形式的激励相结合，探索更有效的非货币激励形式。

（二）依据任教学段特点确定激励类型，关注学科特点对激励效果的影响

研究结果显示，总体上在小学阶段实施教师团体激励优于教师个体激励，而在中学阶段实施教师个体激励要优于团体激励。这与小学和中学阶段教师教学特点存在差异有关。小学教学中，学生学习科目较少，班级任课教师的团队规模通常较小，符合小团队生产易于形成知识互补、相互监督和合作承诺的团队生产特点，实施团队激励更易促进学生学业成绩发展。而中学班级教学中涉及的学科较多，教师团队规模更大，教师间更倾向于以个体生产的方式进行教学，教师间不易形成知识互补和相互监督，以学校或年级为单位设计的团队激励项目还易产生“搭便车”问题，实施个体激励更易促进学生的学业成绩发展。因此，我国基础教育阶段教师绩效工资设计应避免在所有学段实施“一刀切”的激励类型，而是根据教师任教学段特点选择效果更好的设计形式。此外，虽然总体上教师薪酬激励项目对学生学业成绩具有负向影响，但在不同学科中影响方向并不相同。其中教师薪酬激励项目对学生数学学科成绩具有显著的正向影响，而对学生语言阅读学科成绩具有显著的负向影响。这也启示在基础教育阶段进行教师绩效工资设计时需关注学科特点对激励效果的影响，根据学科特点设计激励类型。

（三）根据地区教育发展水平制定政策，构建科学的绩效指标设计

区域背景性特征对教师薪酬激励实施效果的影响是政策评估的重要关注点之一。对比发达国家或地区和发展中国家或地区教师薪酬激励实施效果和绩效指标设计的差异可以看出，一方面，发达国家教师薪酬激励项目的绩效指标设计通常包含学生学业成绩增值评价、教师专业发展和教学实践评估等教师教学绩效的高阶维度，而非洲和南亚地区发展中国家教师薪酬激励项目的绩效指标设计通常包含的是学生的出勤率、学业成绩达标率和教师的出勤率等指向基本规范性的维度。由于发达国家或地区教育发展水平较发展中国家或地区更高，教师工作要求、规范性和专业化程度都较其他地区处于更高的水平，致使发达国家或地区教师在各项绩效指标下努力间存在替代性，调整努力方向的边际成本通常较高，薪酬激励对改变教师效用水平的作用有限。另一方面，对比发达国家或地区和发展中国家或地区教师薪酬激励项目的绩效指标设计还可以发现，发达国家教师薪酬激励项目的绩效指标设计不仅存

在难以观测的问题,还存在绩效目标和期望目标间的不对齐性,学者在众多基于美国教师薪酬激励项目的研究中也指出这些激励项目的绩效指标设计通常过于复杂,并且不一定与学生学业成绩间存在必然的联系,不能提供给教师改进绩效的相关信息(孟卫青,2016)。而发展中国家或地区实施的教师薪酬激励项目绩效指标设计相对简单,易于测量,并且当地对教师工作中课堂教学技术和有效教学时间标准相对较低,教师的各项努力通常不存在替代性。同时,也由于课堂教学技术和有效教学时间标准足够低,众多研究中即使发现教师中存在操纵行为(孔令帅和王超,2022),但依然对学生学业成绩起到了促进作用。我国现阶段区域间的教育发展水平差异巨大,经济发达地区、区域中心城市教师的工作通常具有较强的规范性以及较高的专业化程度和工作要求,工作内容多,负荷大;而部分偏远地区,教师的工作重点仍停留在维持正常教学秩序上。这意味着应从美国和其他国家或地区的教师薪酬激励项目中吸取经验和教训。在制定和实施基础教育教师绩效工资制度时应该充分结合区域教育发展的特征,因时因地制宜地设计教师绩效工资制度,规避教师绩效工资可能对教师工作和学生学业发展产生的消极影响。

综上,教师薪酬激励由于成本约束和教师职业规范的约束,并不是通过提高教师出现逆向选择和道德风险问题成本的价格机制影响学生学业成绩的,在对教师薪酬设计时更应注重教师对薪酬激励的主观感受。同时,在对教师薪酬激励进行类型和绩效指标设计时,需审慎考虑实施对象所处的背景特征,不同学段教师教学中的合作方式。激励指标的设计应充分考虑到实施对象工作任务中的替代性,在工作强度已经很大的教师群体中并不适宜实施薪酬激励,多维绩效指标下,重要且不易测量绩效指标应主导易测量指标的计件工资率设计。总之,要使教师薪酬激励真正实现促进学生学业成绩发展的激励功能,需严格审视教师薪酬激励的实施背景和设计。

[参考文献]

- 孔令帅、王超, 2022:《教师表现性评价的困境与超越》,《教育研究》第7期。
- 孟卫青, 2016:《义务教育学校奖励性绩效工资制度设计的研究》,《教育研究》第2期。
- 赵冉、杜育红、朱倩文, 2022:《教育对中国经济增长的影响效应——基于元回归分析方法的分析》,《教育研究》第1期。
- Alchian, A. A. and H. Demsetz, 1976, “Production, Information Costs, and Economic Organization”, *American Economic Review*, 62(5): 777—795.
- Asch, B. J., 2005, *High-Performance Government: Structure, Leadership, Incentives*, Santa Monica, Calif.: RAND Corporation.
- Baker, G., 2002, “Distortion and Risk in Optimal Incentive Contracts”, *Journal of Human Resources*, 37(4): 728—751.
- Balapumi, R., et al., 2016, “Factors Influencing University Students’ Self-Regulation of Learning: An Exploratory Study”, *Proceedings of the Australasian Computer Science Week Multiconference*.
- Borenstein, M., L. V. Hedges and J. P. Higgins, et al., 2021, *Introduction to Meta-Analysis*, Hoboken: John Wiley & Sons.
- Bradley, S. and C. Green, 2020, *The Economics of Education: A Comprehensive Overview*, North-Holland: Elsevier.
- Brant, X. M., N. Neagos and O. Oluwadara, et al., 2010, *Evidence-Based Practice: Toward Optimizing Clinical Outcomes*, F. Chiappelli (Ed.), Berlin Heidelberg: Springer.
- Bregm, K., 2013, “Detrimental Effects of Performance-Related Pay in the Public Sector? On the Need for a Broader Theoretical Perspective”, *Public Organization Review*, 13(1): 21—35.
- Dang, V., G. Gorton, B. Holmström, et al., 2017, “Banks as Secret Keepers”, *American Economic Review*, 107(4): 1005—1029.
- Dee, S. and J. Keys, 2004, “Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment”, *Journal of Policy Analysis and Management*, 23(3): 471—488.
- Duval, S. and R. Tweedie, 2000, “Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis”, *Biometrics*, 56(2): 455—463.
- Egger, M., D. Smith and N. Phillips, 1997, “Meta-analysis: Principles and Procedures”,

- British Medical Journal*, 315(7121): 1533—1537.
- Frey, S. and R. Jegen, 2001, “Motivation Crowding Theory”, *Journal of Economic Surveys*, 15(5): 589—611.
- Gelfand, J. , et al. , 2011, “Differences between Tight and Loose Cultures: A 33-Nation Study”, *Science*, 332(6033): 1100—1104.
- Gibb, A. , 1996, “Entrepreneurship and Small Business Management: Can We Afford to Neglect Them in the Twenty-First Century Business School?”, *British Journal of Management*, 7(4): 309—321.
- Glewwe, P. , N. Ilias, and M. Kremer, 2010, “Teacher Incentives”, *American Economic Journal: Applied Economics*, 2(3): 205—227.
- Hanushek, E. A. and D. W. Jorgenson, 2000, *Improving America's Schools: The Role of Incentives*, Washington, D. C. : The National Academies Press.
- Hedges, L. V. , 1981, “Distribution Theory for Glass's Estimator of Effect Size and Related Estimators”, *Journal of Educational Statistics*, 6(2): 107—128.
- Hedges, L. V. and I. Olkin, 1985, *Statistical Methods for Meta-Analysis*, Cambridge: Academic Press, Inc.
- Holmström, B. , 2017, “Pay for Performance and Beyond”, *American Economic Review*, 107(7): 1753—1777.
- Kellough, J. E. and G. Nigro, 2002, “Pay for Performance in Georgia State Government: Employee Perspectives on GeorgiaGain after 5 Years”, *Review of Public Personnel Administration*, 22(2): 146—166.
- Kwon, O. N. , J. H. Park and J. S. Park, 2006, “Cultivating Divergent Thinking in Mathematics Through an Open-Ended Approach”, *Asia Pacific Education Review*, 7(1): 51—61.
- Laffont, J. J. and D. Martimort, 2009, *The Theory of Incentives: The Principal-Agent Model*, Princeton: Princeton University Press.
- Lavy, V. , 2007, “Using Performance-Based Pay to Improve the Quality of Teachers”, *The Future of Children*, 17(1): 87—109.
- Lazear, E. P. , 2000, “Performance Pay and Productivity”, *American Economic Review*, 90(5): 1346—1361.
- Lewis, C. , R. Perry and J. Hurd, 2009, “Improving Mathematics Instruction Through Lesson Study: A Theoretical Model and North American Case”, *Journal of Mathematics Teacher Education*, 12(4): 285—304.
- Muralidharan, K. and V. Sundararaman, 2011, “Teacher Performance Pay: Experimental Evidence from India”, *Journal of Political Economy*, 119(1): 39—77.

- Murnane, R. and D. Cohen, 1986, "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive", *Harvard Educational Review*, 56(1): 1—18.
- Neal, D., 2011, "The Design of Performance Pay in Education", In E. A. Hanushek, et al. (Ed.), *Handbook of the Economics of Education*, Vol 4, North-Holland: Elsevier.
- Pham, L. D., T. D. Nguyen and M. G. Springer, 2021, "Teacher Merit Pay: A Meta-Analysis", *American Educational Research Journal*, 58(3): 527—566.
- Phillips, D. C. K., M. E. Bardsley and T. Bach, et al., 2009, "But I Teach Math! The Journey of Middle School Mathematics Teachers and Literacy Coaches Learning to Integrate Literacy Strategies into the Math Instruction", *Education* 3—13, 129(3): 467—472.
- Podgursky, M. J. and M. G. Springer, 2007, "Teacher Performance Pay: A Review", *Journal of Policy Analysis and Management*, 26(4): 909—950.
- Ryan, R. M. and E. L. Deci, 2000, "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being", *American Psychologist*, 55(1): 68—78.
- Springer, M. G., 2009, *Performance Incentives: Their Growing Impact on American K—12 Education*, Maryland: Rowman & Littlefield.
- Sen, A. K., 1977, "Rationality and Morality: A Reply", *Erkenntnis*, 11(2): 225—232.
- Weichselbaumer, D. and R. Winter-Ebmer, 2005, "A Meta-Analysis of the International Gender Wage Gap", *Journal of Economic Surveys*, 19(3): 479—511.

The Impact of Compensation Incentive Programs for Teachers of Primary and Middle Schools on Students' Academic Performance: A Meta-Analysis of Experimental and Quasi-Experimental Studies

ZHAO Ran¹, WANG Yi-bo¹, YANG Jing², DU Yu-hong¹

(1. Faculty of Education/Capital Research Base for Education Economics, Beijing Normal University;

2. School of Education Science, Chengdu Normal University)

Abstract: Based on the analytical framework of principal-agent theory and team production theory, this study employs meta-analysis to examine the impact of teacher compensation incentive programs on student academic performance, utilizing 312 effect sizes from 33 experimental or quasi-experimental studies. The results reveal that teacher compensation incentive programs in primary and secondary education generally exert negative

effects on student academic performance overall. However, secondary school implementations demonstrate superior effectiveness compared to primary schools, with significant moderating effects observed. Key moderating effects include: (1) Incentive Structure: Team-based incentives show better outcomes in primary schools, whereas individual incentives prove more effective in secondary schools. (2) Subject Variation: Compensation incentives positively influence mathematics performance but negatively impact language and reading subjects. (3) Regional Disparity: Programs in developed countries/regions exhibit adverse effects, while those in developing areas yield positive results. (4) Financial Intensity: The monetary amount of incentives shows no statistically significant correlation with program effectiveness. These findings suggest that the design of teacher performance-based pay systems in China's basic education should holistically consider contextual factors such as academic disciplines, educational stages (primary vs. secondary), and regional economic development levels. A scientifically calibrated approach integrating incentive structures and intensity is recommended to optimize pedagogical outcomes.

Key words: teacher compensation incentive; student academic performance; meta-analysis

(责任编辑: 刘泽云 责任校对: 刘泽云 胡咏梅)