

# 推进我国基础教育质量监测制度建设的基本战略与体系保障

◆ 李凌艳 陈慧娟

[摘要] 随着首份《中国义务教育质量监测报告》的发布,我国基础教育质量监测制度建设的推进走到了重要的历史关头。借鉴国内外的历史经验,在该制度建设的战略设计中应切实把握好监测的根本定位、内容导向、主要发展原则、实施布局及长效发展的侧重点。同时,需建立从国家到省、地市、县区的体系保障,区分各级的“为”与“不为”,建立内外相合的监测结果应用环境,切实推动我国基础教育质量监测制度的长效发展。

[关键词] 基础教育;质量监测制度;体系保障

[中图分类号] G521

[文献标识码] A

[文章编号] 1002-4808(2020)03-0068-06

2018年7月,首份《中国义务教育质量监测报告》发布,对我国义务教育阶段学生德智体美和学校教育教学等状况进行了客观呈现,并对如何进一步提升义务教育质量提出了建议。<sup>[1]</sup> 报告的发布意味着我国基础教育质量监测从国家的工作布局、内容研制到组织实施、结果发布等一系列环节的制度建设初步成型,国家基础教育质量监测制度建设将迎来新的机遇和挑战,如何进一步发挥监测的科学引领和质量保障监督的重要作用,需要进行更系统、全面和有效的反思和总结。

## 一、基础教育质量监测制度建设的相关历史经验

虽然我国基础教育质量监测制度的探索和确立仅始于最近十余年,但从世界范围来看,除了PISA(Program for International Student Assessment)和TIMSS(Trends in International Mathematics and Science Study)等国际比较项目以外,据联合国教科文组织2017年《全球教育监测报告》的统计,在209个国家中,已有108个国家发布了国家教育监测报告,在不同程度上建立了本国(地区)的基础教育监测制度,通过监测发挥其对教育政策咨询和教育发展的影响作用。<sup>[2]</sup>

随着现代测量与评估理论和技术的发展,监

测作为一种大尺度评估能够更好地实现传统评估难以达到的功能,如对教育事实的描述更加多维、全面、客观,对教育各要素之间关系的探讨挖掘更加深入、有效等。但与此同时,监测被过分功能扩大化、目的功利化的风险也会增大。“第四代评估”理论的提出者埃贡·G·古贝(Egon G. Guba)和伊冯娜·S·林肯(Yvonna S. Lincoln)认为,评估时代从“测量”“描述”到“判断”的过程与评估技术发展紧密相关,评估所能实现的目标会受到评估理论与技术水平的制约。<sup>[3]</sup> 世界各国教育评估发展与监测制度建设的相关历史经验不仅验证了这一点,反过来也提醒我们必须意识到:教育质量监测首先应以为教育发展服务为根本,如果滥用技术的力量、过度扩大评估技术本身的功用,则很可能适得其反地导致评估实际效能的衰减。

实际上,世界许多国家或地区已有相关历史经验和教训。例如,日本早在20世纪50年代就通过全国学力、学习状况调查项目开始了国家基础教育质量监测制度的探索,但由于过早、过多地突出了其评判功能而引发了学校和地方的过度竞争和全国范围的质疑,一度中断40多年。<sup>[4]</sup> 相对来说,美国国家基础教育质量监测以其国家教育进步评估项目(National Assessment of Education-

李凌艳/北京师范大学中国基础教育质量监测协同创新中心教授,博士生导师(北京 100875);陈慧娟/北京师范大学中国基础教育质量监测协同创新中心博士后,本文通讯作者(北京 100875)。

al Progress, NAEP) 为依托, 虽起步于 1969 年, 但在前 30 多年中多发挥教育评估描述与记录等基本功能, 一直到其制度设计较为完备、评估结果受到广泛认可后, 2004 年起才逐渐转向评估与问责功能的并存。<sup>[5]</sup>

我国基础教育质量监测制度的创建始于 2007 年, 由教育部基础教育质量监测中心在教育部领导下进行试点监测探索, 在八年的试点监测阶段, 一直坚持监测为教育发展服务的根本定位, 突出并坚持监测中的纸笔测验与传统统考、统测在目的、性质、功能、过程、结果运用等方面的不同。通过试点探索, 逐步实现了我国基础教育质量监测制度的基本设计与全国性实施, 在八年试点、三年第一轮正式监测后才发布首份国家义务教育质量监测报告。

相关历史经验表明, 基础教育质量监测应坚持为教育发展服务的根本定位, 而不是片面地对教育进行判定的“法官”, 否则, 评估功能就会被不恰当地扩大化, 走上被评估过程异化的可持续发展之路。

## 二、推进我国基础教育质量监测制度建设的基本战略

### (一) 根本定位: 具有技术先进性、为教育发展服务的评估

教育评估应服务于教育发展, 其根本定位是促进教育事业发展的—种手段和方法。然而长期以来, 我国基础教育领域的各类评估大多具有高评判性、标签性和行政性的特质, 始终扮演着“法官”的角色, 评估中的“造假”现象也成为众所周知的顽疾。造成这类问题的原因固然复杂而复杂, 但是, 评估技术手段的不足, 也是其中不可忽视的重要原因之一。

而质量监测作为—种大尺度评估, 借助现代测量理论与技术的新发展, 诸如矩阵抽样技术、项目反应理论、认知诊断模型等, 改变了传统的基于经典测验理论下“—张卷子考所有人”的考试技术, 在更大程度上实现了对包含学业成就在内的学生各方面综合发展信息更多维、多层的比较分析, 从而更加客观地揭示学生发展状况和规律, 并对差异原因进行针对性的挖掘和探索。因此, 就评估的专业性而言, 质量监测是—种更为客观、全面、深入的技术和手段, 能够在一定程度上突破传统评估的简单测量、描述和判断功能,

通过更好地响应式建构实现与教育过程的对话, 为教育政策制定、教育过程管理提供更有针对性的确切依据。因此, 要充分发挥监测的技术先进性优势, 始终坚持为教育发展服务, 而不是“以评代管”, 是推进教育质量监测制度建设的根本定位。

### (二) 内容设计: 促进学生全面发展的综合质量导向

监测内容与范围体现着监测的核心价值, 基于现阶段国情, 我国基础教育质量监测制度建设中, 监测“导向”比监测“多少”更为重要。目前, 评估在教育实践中发挥的导向作用相对比较简单, 用考试分数、升学率作为学校发展状况最主要甚至唯一—指标的现象仍比较突出。因此, 通过对包含学业成就在内、同时全面考查学生身心发展状况及其影响因素的科学监测, 导向学生全面发展的综合质量, 对于我国基础教育的可持续发展具有深远意义。

也正是基于综合质量导向, 2015 年发布的《国家义务教育质量监测方案》确定了我国义务教育质量监测的学科和领域为语文、数学、科学、体育、艺术和德育, 三年为—个监测周期, 第一年监测数学和体育、第二年语文和艺术、第三年科学和德育, 同时每年进行学生发展影响因素的监测, 这里的导向意义不言而喻。<sup>[6]</sup>可以说, 国家就是要通过这样的内容设计, 打造不同于高考的另一根教育“指挥棒”。不同于高考服务于学生个体选拔的性质, 教育质量监测服务于整个教育事业本身的健康发展, 因此, 监测的内容定要更为全面、综合, 而且针对基础教育实践中的薄弱环节有针对性地突破。

简而言之, 在监测的内容设计上, 做“对”比做“多少”更为重要, 促进学生全面发展的综合质量导向至为关键。

### (三) 发展原则: 在优先考虑实践适用性的同时确保专业科学性

监测内容在确保综合质量导向的同时, 也将随着基础教育质量监测制度建设的不断推进而日益丰富, 在这一过程中, 需立足实际, 按照适用性、专业性的优先顺序来发展。例如, 目前我国基本确定了义务教育阶段的监测内容, 未来随着时机成熟及技术进步, 监测将向学前和高中两端延伸, 逐渐覆盖基础教育全阶段。在这一过程中, 如何确定学前和高中的监测内容及方式、如何在

一次性横断监测项目的基础上丰富纵向追踪甚至专项监测项目，固然要坚持大尺度评估过程中的技术规范和流程专业性，但立足国情，这一过程需要首先考虑其适用性如何，即专业性原则要先符合适用性原则。

教育行业的服务发展水平和最近发展区一定会跟社会文化和经济基础紧密关联，这也是我们经常会发现“中国教育的问题可以借用西方发达国家和地区的经验，但是必须且只能用中国自己的方法来解决”的深层原因。例如，由于我国以县为主的管理体制，加之基础教育的宏大规模，在首轮义务教育质量监测中，对监测对象的抽样方案就是与我国教育实践体系高度适配统一的、从省到县（市、区）再到学校、学生的多阶段不等概率抽样。<sup>[7]</sup> 我们知道，从抽样的代表性来说，固然是分阶段少的、基于所有学校的随机抽样最好，但由于我国基础教育的量大、面广，这种完全基于学校总体的从学校到学生的两阶段随机抽样在教育现实情景中几乎无法实现。再如，考察学生的艺术学习质量，除了通过学生对中外、民族艺术作品的风格特点及所表达主题和情感的感受与理解外，通过艺术创作来表达与交流自己情感与思想的能力等也是非常重要的艺术表现力。但根据我国义务教育质量监测的学生样本规模、国家对质量监测的现有投入水平以及目前艺术教育的现实基础，先从学生的歌曲演唱能力开始采集学生的音乐表现力数据，也是根据适用性优先、专业性为基础和保障的原则来实施的。

#### （四）实施布局：警惕过度监测，重视“留白”设计

尽管监测作为较为先进的大尺度评估手段，能在一定程度上改观传统教育评估所难以实现的更为全面、客观、真实地刻画教育发展状况的功能，但它并不是万能的，有些发展侧面或领域还需慎用“监测”，警惕过度监测。教育以人的成长为主体，而人的发展过程有很多无法具体、量化评估的方面，如果想要借助“监测”的量化评估“虎皮”而行“过度评估”之实，无疑会将监测带上不归之路。

所以，在监测制度的内容建设和实施布局上，不仅要立足实际坚持适用性、专业性的优先发展顺序，而且一定要有意识地设计“留白”，才能更好地发展和创新。例如，曾有报道说，有学校或技术公司试图通过对学生脑电波的跟踪去监测学

生的“学习”状态。这种监测不仅脑科学技术本身真假混呈，对于学生有效学习、主动学习的促进而言，脑电波是否应该被监测，也牵涉关于人的发展价值之论争等更多问题。也许在将来，随着技术对人类生活的改变，这个话题价值观层面的侧重会发生很大转变，但在现今时代背景下，强调有些学生发展的领域或层面，监测不要全覆盖、一定要留白，这才是真正的实事求是。一味地唯技术有可能导致伪技术的态度和做法。

无论基础教育的哪个阶段，一定要选择该阶段学生发展的关键领域和指标，在明确的教育导向下，有侧重地设计相关的监测内容，不宜平铺式、各阶段全科全领域地监测。例如，学前阶段或高中阶段的监测重点和内容应不同于义务教育阶段，而这种不同可能率先体现在内容上，继而会在形式、手段等方面体现。即便是义务教育阶段本身，国家监测目前也只是涉及六个领域，包含数学、语文和科学三个学科，当然这里的设计首先是基于现阶段国情的适用性原则，但也是因为监测的内容要留白，不是学生发展及其影响因素的各个侧面都能监测，或者都要监测的。

#### （五）长效发展：推进“以监促改”，重视结果应用

从宏观制度建设的意义而言，旨在促进学生全面发展、具有新的“教育体检仪”和“指挥棒”作用的教育质量监测，应发挥全面诊断（诊）、决策咨询（咨）、客观监督（督）、促进提升（促）和舆论引导（导）的五大职能。从监测内容确立、监测指标和工具研制、监测组织实施，到监测数据分析、报告撰写和发布，可以说，基本完成了以“诊”和“咨”为主的监测制度建设部分，而“督”“促”“导”功能则必须通过以监促改、加强监测结果应用才能得以实现，这也正是监测制度长效、可持续发展的生命力所在。因而，长远的监测制度推进中一定要完成各级监测后以监促改的相应配套制度建设。

首先，确保监测结果得到有效应用的前提是监测结果的定期公示或发布制度。通过对基础教育质量的持续监测，并定期将结果公之于众，让基础教育规模发展不断取得进步的同时，更让规模初具后的质量内涵发展状况接受相关部门、社会公众的知情和监督，这是监测结果能否被实际应用、发挥作用的基石。当然，如何正确、有效地引导社会，并重视结果的恰当发布，是结果发

布过程中必须研究的技术性问题,但无论该问题如何重要、复杂且专业性强,也不应影响有无结果发布制度的本身。

其次,内部约谈、限期整改等一定的问责制度是促进监测结果应用的重要推手。<sup>[8]</sup>基础教育质量是民族和国家发展的原动力,各级政府与教育行政部门对其发展应负有主要责任,在这一前提下,才能吸纳更多社会力量积极参与办学。同时,基础教育质量监测后相应问责制度的力度和效率,在很大程度上与整个教育管理体制机制的水平和自身健康状况密切相关,问责制度建设也应循序渐进,过于简单的问责不仅不利于监测事业的长效发展,还可能阻碍教育的健康发展。从内部约谈、限期整改,逐渐到问题公示、接受社会监督、质量改进效果评估与改进问责,需要结合国情、区域实情及监测发展状况,逐步形成监测—改进—再监测—再改进的良性循环。在这一过程中,建立各级问题的责任制、问题整改的部门联动制和首责制等,都是保证有效问责的重要工作制度。

必须强调的是,在监测结果定期发布并促进问责制建设的过程中,还要高度重视监测组织和实施的责任主体制确立,建立预防作假、打击作假的防作弊制度以保障监测结果真实性。为帮助各个国家和地区收集、分析有关教育政策的信息并确证其有效性,世界银行组织于2011年起开展了“为了更好的教育系统”(System Approach for Better Education Result, SABER)项目,已有100多个国家和地区参与其中,在其用于分析教育系统质量的整体指标框架中包含了一系列丰富的教育管理信息系统(Education Management Information System, EMIS)指标,凸显出教育信息库的质量将在很大程度上反映一个国家教育系统的水平。<sup>[9]</sup>从这个意义上说,真实、客观的监测数据不仅是监测事业的生命线,而且也是国家教育发展整体水平的重要显见度指标之一。

### 三、基础教育质量监测制度建设的体系保障

确立基本战略是制度建设的重要基础,建立保障体系则是制度建设的内在实质。这里的保障是指搭构怎样的工作体系以及确立体系内各部分之间关系的“规则”——这应该是某项工作从一般的、带有短期特征的项目探索,上升至长远的制度建设后最核心的内容。

#### (一) 建立与教育管理体系相适配的监测评估工作体系

正由于监测是一种大尺度评估,就专业属性而言属于大的教育评估领域的一种形式、技术和手段;而从教育实践层面而言,评估的工作体系势必与管理的工作体系相适应,因此,保持两个体系的逻辑一致性,才能使为教育发展服务的根本定位落地。

我国的教育管理体系与政府和行政管理体系一致,均以县(区)为主,教育人事权和财权也都落在县(区)一级,从中央到省(自治区、直辖市)、地市级,则更多侧重于对县(区)一级的业务(即“事权”)领导、指导和支持,以及宏观政策的调控。那么,与之相适应,应建立从国家到省(自治区、直辖市)、地市,最后落脚在县(区)一级的监测评估工作保障体系。

截至2019年底,全国共有28个省(自治区、直辖市)成立了省级监测机构,据不完全统计,28个地市级、90余个县区级监测机构也已正式建立。至此,我国基础教育质量监测工作已经逐步形成了“以机构建立带动工作发展”的体系建设思路。虽然这些监测机构尚存在职能定位模糊、工作任务边界不清、专业力量及系统支持缺乏等困难,但也因地制宜开展了区域基础教育质量监测的实践探索,甚至部分地区结合监测结果采取了相应改进措施并取得了一定成效<sup>[10]</sup>,但同时,由于各地对监测的认识不清晰、不一致,对自身所应做工作的定位和内容更是模糊等原因,监测在我国“方兴未艾”的今天,已然出现了较多的以“监测”之名行“统考”之实现象。这不仅给地方教育部门、学校增加了负担,不利于教育的良性发展,更严重损害了基础教育质量监测事业的健康成长。

随着我国基础教育质量监测从国家、省级到地市、县区工作体系保障性网络的初步形成,未来制度建设中还需厘清该网络中各级的职责、分工和协作关系,区分各级的“为”与“不为”,通过体系形成合力,而不是“多头监测”导致滥用,以体系建设的不断完善推进监测事业的有序、健康、可持续发展。

#### (二) 确立各级监测工作体系的“为”与“不为”

就监测的专业性而言,监测作为评估的一种手段,由于各级监测机构所要开展的教育评估工作目的不尽相同,因而其工作内容和性质也应有

所不同(见图1)。

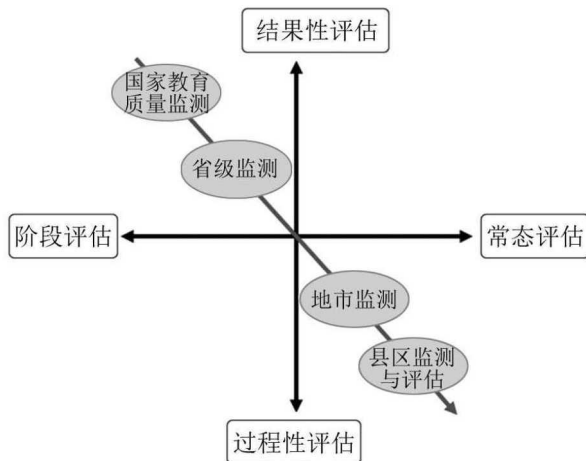


图1 各级监测工作保障体系应侧重的评估类型

相对而言,国家监测和省级监测更侧重通过大尺度评估了解整体质量状况,因而其监测更具有阶段性、结果性评估的性质;而地市监测和县区监测与评估,由于范围和规模要小得多,其目的更在于过程性监控和指导。尤其是县区一级,形式应不同于国家或省级监测常常采用的一次性测验,而重在过程性及常态的表现性评估。

基于此,各级监测的基本内容、功能和作用可以进行如表1所示的区分。不难发现,到了县区一级,虽然工作仍然是围绕“监测”而展开,但其主要内容和作用已经不在于“测”本身,而在于通过结果的实际应用发挥教育评估对教育发展的引领和规范作用。这不仅是因为评估工作内容要与评估目的保持高度一致才能落地的专业应然,也是因为目前我国县区一级的力量不可能进行需要较为复杂的现代测量与统计专业技术支撑的大尺度监测的实然。

表1 各层面基础教育质量监测工作重点一览表

各级层面	教育质量监测工作重点
国家层面	制定监测标准,研发监测指标和工具,建立监测工作规范,建立国家常模和国家监测数据库; 把握全国基础教育质量总体状况,了解不同区域、省份,不同类型学校、不同学生群体的状况; 了解国家各项教育政策的落实情况与落实效果; 发布国家监测报告,引导全社会树立科学、全面的教育质量观

续表

各级层面	教育质量监测工作重点
省级层面	配合做好国家质量监测; 有条件情况下,确定省域监测领域和主要内容,研发监测工具,建立省域常模和监测数据库; 了解不同地市、区县,不同类型学校,不同学生群体的发展状况; 了解省域各项教育政策的落实与效果情况; 面向社会,引导大众树立科学、全面的教育质量观
地市级层面	在国测、省测工作基础上,量力而行地开展一些专题监测,研发监测指标和工具; 了解不同区县,不同类型学校,不同学生群体的发展状况; 基于国测、省测及市级专项监测结果,指导、督促区县进行整改和提升
区县层面	研读国测、省测、市测结果报告,列出问题,深入分析原因,形成整改方案; 细化整改责任分工,落实到责任部门,采取有力措施,促进监测发现问题得到有效解决

只有明晰了从国家、省级到地市级、区县级监测中各自应侧重什么又不应该做什么,才可能形成协作、互补的监测工作保障体系,开展不同层次的监测,为各级教育决策、教育管理咨询和监督提供精准而高效的支持。简言之,应高度重视监测工作保障体系的顶层设计和整体规划,实现不同层级监测机构的错位发展、特色发展,从而避免重复监测和无效监测。

### (三) 创设内外相合的监测结果应用环境

以机构建立带动工作发展是监测体系保障的开始,使长效的监测制度发展具有了基本的条件保障,但制度的内涵发展离不开“监测—改进—再监测”的螺旋式提升过程,“以监促改”是制度建设的根本推动力,创设内外相合的监测结果应用环境则是保障相关工作顺利开展的重要前提。

监测结果应用的外部保障体现为国家—省—地市—区县的多级联动。随着国家监测的制度性开展,国家每年都会为各省和样本县区出具监测

结果报告,同时,越来越多的省份开始尝试省级自主监测,并为地市和区县提供监测结果报告。以国家和省级监测结果报告的应用活动为推手,加强其在地市和区县教育管理实践中的落实,是建设监测工作保障体系的重要依托。尤其,省和地市一级督导和监测部门在这一过程中能很好地弥补县级督导部门权威性不足的缺陷,也会为内部整改机制不畅的区县提供有效监督。

与此同时,区县一级的部门联动制是监测结果应用有效落实、深化与创新的内部保障。在监测结果应用过程中,从监测结果分析与反馈到制定措施、落实行动,区县一级各部门要注重主体的转换和移交——监测的开展和组织要以县级监测部门为主,实现监测发现问题的目标。而到解决问题环节,则一定要争取行政的介入与支持,多部门共同参与形成合力。县级监测部门作为监测的归口部门,在结果应用过程中的主要职责是积极推动相关部门的联合作战,为结果应用提供支撑,并督促改进取得实效。如依据国家、省级等监测结果发现问题后,与课程设置、教学计划和办学行为有关的问题应交由基教和督导部门落实整改,而与教育投入、资源配置有关的问题应归口计财部门,与师资配备、人员培训和师德师风建设相关的问题应归口人事部门,与课堂定位、教学方法和教研交流等相关问题应归口教研部门,通过科室备案、部门联动的方式,从问题出发,通过问题解决的责任制来推进。

综上,只有建立与教育管理体系相适配的监测评估工作体系,并明确各级工作体系的职责定位,持续建设内外相合的监测结果应用环境,我国各级监测工作才能促进区域教育管理从基于经验向基于实证的转变,向教育治理现代化迈进,

在此过程中也才能真正推动我国基础教育质量监测制度的长效发展。

### [参考文献]

- [1]我国首份《中国义务教育质量监测报告》发布[EB/OL].(2018-07-24)[2019-11-18].[http://www.moe.gov.cn/jyb\\_xwfb/gzdt\\_gzdt/s5987/201807/t20180724\\_343663.html](http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/201807/t20180724_343663.html).
- [2]UNESCO. Global education monitoring report 2017/8: accountability in education; meeting our commitments [R]. Paris: UNESCO publishing, 2017.
- [3]古贝埃贡,林肯伊冯娜.第四代评估[M].秦霖,蒋燕玲,译.北京:中国人民大学出版社,2008:1-20.
- [4]浦岸英雄.全国学力テストはなぜ実施されたのか(全国学力测验为什么会实施)[J].园田学园女子大学论文集,2010(44):27-39.
- [5]Council of Chief State School Officers. The Development Process of the Common Core State Standards[EB/OL].(2015-08-01)[2019-11-18].<http://www.corestandards.org/about-the-standards/development-process/>.
- [6]国家义务教育质量监测制度建立[EB/OL].(2015-04-16)[2019-11-18].[http://www.moe.gov.cn/jyb\\_xwfb/xw\\_fbh/moe\\_2069/xwfbh\\_2015n/xwfb\\_150415/150415\\_mtbd/201504/t20150416\\_187220.html](http://www.moe.gov.cn/jyb_xwfb/xw_fbh/moe_2069/xwfbh_2015n/xwfb_150415/150415_mtbd/201504/t20150416_187220.html).
- [7]国务院教育督导委员会办公室关于印发《国家义务教育质量监测方案》的通知[EB/OL].(2015-04-15)[2019-11-18].[http://www.kmsx.com.cn/news\\_show.aspx?id=547](http://www.kmsx.com.cn/news_show.aspx?id=547).
- [8]董奇.质量监测如何为教育把好脉[N].光明日报,2018-08-18(06).
- [9]World Bank. The what, why, and how of the Systems Approach for Better Education Results (SABER). SABER Overview [R]. Washington, DC: The World Bank.
- [10]李凌艳,任晓琼,江照富.我国区域基础教育质量监测的探索与思考[J].中国教育学刊,2017(12):37-41.

(责任编辑 蒲丽芳)

## 《中国教育学刊》投稿指南

1. 文章须针对基础教育层面展开研究(包含宏观性教育理论研究),高等教育内容不在本刊编发范围。
2. 文章写作需突出学术引领和实践导向,要求论点明确、逻辑严谨、论述清晰、独具新意。
3. 本刊对于省以下级别及与文章内容无关的基金项目不予标注。
4. 文章中的图、表要精选且精心设计与绘制。
5. 参考文献著录采用顺序编码制,具体格式参见官网说明。
6. 稿件审稿周期通常为3个月,请勿一稿多投。