法治如何定量

——我国法治评估量化方法评析

周祖成　　杨惠琪*

内容提要：对法治状态的定量评价，可以为法治建设提供方向和技术指引。法治评估中所运用的量化方法科学与否，将直接影响评估结果的合理性、执行力及社会效果。通过对几种典型的法治量化评估体系在数据搜集处理、权重设置、指数计算中采用的计量方法进行比对分析，可以发现，当前实践与理论研究中量化方法存在着整体局面复杂混乱、具体运用多有瑕疵、实施过程不公开、缺乏结果校验机制等问题。必须正视当下法治评估背后所隐藏的计量隐患，以社会实效为导向探索和建构法治评估的方法论体系。

关键词：法治指数　法治量化评估　法治量化方法

一、问题的提出

国内多地法治评估活动相继开展，使得如何评估法治成为近年备受关注的理论热点问题。随着理论研究的不断深入与典型实践的依次开展，原本关于“要不要评估法治”的争论转变成了“如何评估法治”的争论。在此基础上，不少学者就法治评估的目标、意义、原则、体系构建方法、指标设计方案、评估对象选择等诸多方面展开了持续性探讨，“法治评估学”、“法治实践派”等理论也应运而生。不过，现有研究大多将“如何评估法治”等同于“如何建构评估体系”等基础定性问题，鲜有学者选择从“定量”这一极具评估特色的视角出发，针对法治评估量化方法作出全面的分析。现实法治评估中所涉及的量化方法

* 周祖成，西南政法大学行政法学院教授；杨惠琪，西南政法大学行政法学院博士生。
本文为中国法学会重点委托课题“全面推进依法治国背景下的地方法治建设问题研究”（CLS（2015）ZDWT10）、司法部国家法治与法学理论研究项目“法治指数及其中国应用研究”（14SFB2002）、2015 年重庆市研究生科研创新项目“地方立法的公共参与研究”（YKC201501085）成果。周祖成负责全文写作思路，杨惠琪负责资料收集及初稿撰写。

有哪些？这些量化方法在实践运用中存在哪些问题？怎样避免相关问题的发生？本文尝试就这些问题作出初步探讨。

法治评估活动在短短几年内积累了许多理论与实践成果，相应的评估体系也种类繁多，并且，以往许多成果已经对法治评估体系的概况、类型、特征等进行过深入研究，〔1〕由此，本文并非对所有体系作全面分析，而主要是从“量化”角度衡量、选取具有相应分析价值的典型体系，并对这些体系中的量化方法与量化过程做出梳理与评价，希望以此发现量化过程中可能存在的不足与值得改进之处。从量化方法的运用出发，我们选择了两类共计 8 个典型量化评估体系：

表 1　本文涉及的典型量化评估体系分类及概况〔2〕

分类	内容	主体	实施（公布）时间
实践体系	余杭体系	政府支持	2007 年至今
	昆明体系	政府支持	2010 年至今
	上海体系	学者自拟	2011 年、2014 年
	北京体系 Ⅰ	学者自拟	评价 1997—2001 年北京法治环境水平
理论体系	江苏体系	政府制定	2015 年试行及专家讨论稿
	北京体系 Ⅱ	学者拟定	2007 年出版专著说明
	社会主义法治综合评价	学者拟定	2009 年以论文形式说明
	法治国情指数	学者拟定	2014 年以论文形式说明

这些体系的典型与否并不以其自身影响力大小为评判标准，而是结合前述三个问题，以其是否具有相应的分析价值为选择依据。虽然理论体系的影响力和认知度远小于实践体系，并且仅北京体系 Ⅱ 较为整全完善，其他体系仍旧处于雏形阶段，但在纷繁的法治评估研究中，已有法学学者对量化方法给予了关注，并提出了一些理论设想供实践参考、检验，故有必要对这些理论体系做出回应。

（一）实践体系的中坚力量：政府支持类

1．余杭体系：法治余杭量化评估体系

2006 年，浙江余杭提出用一个量化的指数来衡量余杭法治建设水平、评估余杭社会管理能力，从而使余杭成为我国内地最早提出并运用法治指数评价体系的地区。余杭的法治指数评价体系结构可以概括为“149”。“1”就是指余杭法治指数；“4”是指 4 个评估层面：区本级总指标、区级机关指标、乡镇街道指标、农村社区指标；“9”是指人民群众关于党风廉政建设、政府行政工作、司法工作、权利救济、社会法治意识程度、市场秩序规范性、监督工作、民主政治参与、社会治安 9 个方面的满意度调查。余杭的法治指数主要运

〔1〕 代表性成果如周尚君、彭浩：《可量化的正义：地方法治指数评估体系研究报告》，《法学评论》2014 年第 2 期；孟涛：《论法治评估的三种类型——法治评估的一个比较视角》，《法学家》2015 年第 3 期；钱弘道等：《法治评估及其中国应用》，《中国社会科学》2012 年第 4 期。

〔2〕 备选体系仅局限于法治建设和法治环境类评估体系，即评估对象主要是一地区的法治现象，而不包括法治政府、立法状况、司法状况等具体的法治专项评估体系。

用了德尔菲法，是由内部评审组、外部评审组、专家评审组、群众满意度调查相结合计算得出的。整体上而言，余杭体系运作时间最长，指标建构较为完善，操作过程较为成熟，社会及学界影响也很大，是我国法治评估体系的典型代表。〔3〕

2．昆明体系：法治昆明综合评价指标体系

云南省昆明市在 2010 年发布了《法治昆明综合评价指标体系》。该体系独特之处在于，它在一个总指标体系之下囊括了三个分级指标体系（法治社会环境、法治制度环境和法治人文环境），三个分级指标体系又分别采用了不同的评估方法。其中，法治社会环境和制度环境借鉴余杭体系，由内部组、外部组、专家组独立评价；法治人文环境包括群众满意度调查和“治理能力评价指数”的相关数据。该体系是除余杭体系之外现有实践经验最为丰富的评估体系，也具有一定的社会影响力与认知度。

（二）实践体系的重要参照：学者自拟类

1．上海体系：上海法治建设指标体系

上海法治建设指标体系由上海社科院法治市情研究中心课题组拟定，主要由民主政治、法治政府、司法公正、社会治理等 4 个一级指标组成，内含 22 个二级指标和 52 个三级指标，其前身是 2011 年开展的“上海法治建设满意度调研”。从量化角度来看，该体系严谨性、科学化程度较高，各指标权重纯粹由层次分析方法确定，最终通过系统的问卷调研方式求得 2014 年度上海法治建设综合指数。但是，该体系运行年份很少，相关实践资料披露也较少，相较于余杭、昆明体系而言，缺乏政府的引导与有力支持，具体前景未知。〔4〕

2．北京体系Ⅰ：城市法治环境评价体系

“城市法治环境评价体系”是中国人民公安大学课题组的研究成果，虽然研究年份较早，但其不仅注重对数据的量化处理，还开辟了综合运用权重确定方式的实践先例，具有较大的分析价值。但是，该体系构造过于简单，同当下主流体系风格并不一致，并且也仅实际运行了一次，还是对以往多年份的法治环境做量化考查，影响力、认知度较低。〔5〕

（三）理论体系的积极探索

1．政府制定类：江苏体系

江苏体系构成较为复杂。《法治江苏建设指标体系（试行）》2015 年上半年出台，在借鉴余杭体系基础之上，试行稿由七大类一级指标、29 项二级指标和 1 项综合评判指标构成，共涉及计分点 90 个，但并未开展相应的评估实践。2015 年下半年，江苏省又召开《江苏法治社会建设指标体系》专家讨论会，体系初拟由 5 个一级指标、19 个二级指标、61 个三级指标构成。其中，第五个一级指标“区域法治社会建设的社会感受和评价”，侧重从群众感受来考察法治社会建设的效果，使指标体系更为客观有效。关于量化方面，会议指出，需要在多方征询专家意见的基础上，确定最终指标体系和权重，建立定量评估模型，确定各项指标的权重百分比数值。虽然具体方案还未公布，但会议明确将在 2016 年初在宜兴、如皋、沭阳县（市）三地进行试测，并于 2016 年上半年全面展开实际运用。

〔3〕 相关可参考钱弘道：《中国法治指数报告（2007—2011）——余杭的实验》，中国社会科学出版社 2012 年版，第 5 页以下。

〔4〕 参见姚颉靖、彭辉：《上海法治评估的实证分析》，《行政法学研究》2015 年第 2 期，第 53 页以下。

〔5〕 参见王光：《城市法治环境评价体系与方法研究》，中国人民公安大学出版社 2004 年版，第 74 页以下。

2．学者拟定类：北京体系Ⅱ

该体系由北京市法治建设状况综合评价指标体系研究课题组于2007年创设。虽然它选取的测评样本同样是北京地区，但在指标建构、权重方式、计量程序方面均和北京体系Ⅰ存在较大区别。该体系的量化优势在于，对于相关的指标属性和数据处理方法做出了充分的论证与说明，权重设置也考虑了层次分析法，只是缺乏相关实践支撑。〔6〕

除此之外，学者自拟的理论体系还有“社会主义法治综合评价体系”和“法治国情指数”，它们均是以论文形式发表，并具备有价值的量化方法论述。

二、典型评估体系主要运用的量化方法

一般认为，法治评估主要涉及建立评估体系、收集处理数据信息、分配权重、计算结果、公布结果、后续分析研究等程序。在这一过程中，法治量化方法运用合理与否，将会直接影响评估本身及其结果的合理性、公信力与执行力。

（一）数据使用中的无量纲化处理

1．现有体系数据来源与使用策略各异

开展法治评估需要依托大量的数据信息，现有实践中，依照数据提供主体的不同，可将法治评估涉及的相关数据区分为政府数据和民间数据两类。

政府数据又可细分为统计数据和考评数据，其中，统计数据一般出自国家统计系统，考评数据则是指对政府具体工作情况的打分数据。统计数据本身的准确性与权威性是其他数据不能比拟的，由于不需要借助社会调研、问卷调查等手段去收集，统计数据也具有极大的便利性，其稳定性更能为法治评估的长效运行提供保障。考评数据则是对法治评估中具体指标的最直观反映，这类数据能够简洁地体现出相关法治指标的量化情况。政府数据虽然受到评估组织者的青睐，却并不一定能令法治建设的最终受益者——民众满意。有些民众倾向认为政府数据有自说自话之嫌，过多运用政府数据也容易使法治评估陷入“内部考核”的尴尬境地。

民间数据又可细分为公众调查数据和民间统计数据两类。公众调查数据多是评估组织者以问卷形式得到的具体调研数据，能够较大程度地反映民众对评估区域内法治状况总体评价。民间统计数据则是指一些民间管理咨询机构所提供的调研统计数据，在一定程度上填补了政府统计数据可能存在的盲区。民间数据虽然具有涉及面广、信息含量大的优势，却也存在数据过于集中的弊端，并且，以公众问卷调查这类主流民间数据为例，它在具体使用中还面临着如何合理地将主观性信息转换为客观统计数据等操作性难题。

政府数据和民间数据各有所长，各类法治评估体系对两类数据抱有的态度也大不相同。总体而言，目前政府数据在法治评估中仍旧占据了很大比重，〔7〕民间数据承担的多是辅助

〔6〕 参见王称心：《现代化法治城市评价——北京市法治建设状况综合评价指标体系研究》，知识产权出版社2008年版，第116页以下。

〔7〕 民间数据往往仅在评估体系的某一层次出现，而其他的层次往往都是对政府数据的不同运用。例如，北京体系Ⅱ就指明是“以政府统计数据为主建立的评价体系，但也少量设置了需从民间统计中获得数据的公共评价指标”。同上书，第145页。

者的角色，并且集中出现在评估体系中的公众评价层次。〔8〕也有一些地区另辟蹊径，在评估中主要考虑民间数据而非政府数据，如上海体系就仅仅涉及公众调查数据。〔9〕

不同评估系统在使用数据方面采取的策略也各不相同，大致可将数据的使用情况区分为直接使用和间接使用两大类别。直接使用是指将收集来的数据直接运用到指标体系中，如北京体系Ⅱ将法院一审结案率、政协委员提案采纳率、公众社会安全状况满意率等数据，直接体现在指标体系中。〔10〕间接使用主要表现为，收集到的数据对法治评估的结果起参照作用而非决定作用。以余杭体系中政府数据为例，该体系在评估的四大层次中均设置了详细的考评细则，并得出了明确的考评分数，但这些考评数据并不代表相应法治指标的最终得分，考评数据只是为评估主体提供的打分参照，最终决定余杭体系评估结果的，是评估主体参照多种数据信息而得出的指标具体评分与权重值。

经过对比分析，可以将现有指标体系中的数据来源与使用情况归纳为下表：

表 2　典型体系数据来源及使用情况

分类	内容	政府数据	民间数据
实践体系	余杭体系	间接使用	直接使用
	昆明体系	直接使用	直接使用
	上海体系	无	直接使用
	北京体系Ⅰ	直接使用	直接使用
理论体系	江苏体系〔11〕	*	直接使用
	北京体系Ⅱ	直接使用	直接使用
	社会主义法治综合评价〔12〕	*	*
	法治国情指数〔13〕	*	*

* 号代表相关信息的缺失

上述类型区分仅是理论意义上的，实践中情况要复杂得多。由于各类数据来源不同、反映指标不同，数据单位也容易出现不统一的情况，这就为法治评估的运算过程带来了难度。为了规避数据统一化这一专业性强、运算复杂的处理步骤，有些地区选择对数据进行间接使用，但间接使用又势必会掺杂许多不确定的主观因素，故此，实践中仍有一些地区坚持直接使用数据，这就需要评估组织者对原始数据进行无量纲化处理。〔14〕现有实践中无

〔8〕 例如，昆明体系中专设“公众评价性指标”和“公众体验性指标”。前者数据出自“中国城市竞争力报告”，后者则是由问卷调查获得。

〔9〕 参见前引〔4〕，姚颉靖等文，第 56 页。

〔10〕 参见前引〔6〕，王称心书，第 128 页以下。

〔11〕 江苏体系的具体方案未明，但根据相关报道能够看出，该地区在民间数据上是提倡直接使用的，其一级指标“区域法治社会建设的社会感受和评价”就是使用民间的调研数据。

〔12〕 参见陈海燕、张庆旭：《社会主义法治评价指标量化研究》，《科学社会主义》2009 年第 4 期。

〔13〕 参见蒋立山：《中国法治指数设计的理论问题》，《法学家》2004 年第 1 期。

〔14〕 即数据的标准化处理，就是将数据转化，统一数据单位、格式，以便数据直接用于计算最终结果，从而最大程度地剥离主观因素的影响。无量纲化处理在统计学中是一项复杂的工作，主要有直线法、折线法、曲线法三种处理方法。在法治评估过程中，一般使用的都是最简单的直线法。

量纲化针对的主要是复杂的政府数据，尤其是政府统计数据，因其来自不同的部门，具有不同的性质，〔15〕所以在直接使用过程中大都需要进行标准化处理。

2．指标无量纲化处理过少

既然政府数据权威复杂、比重大又性质各异，在具体实践中理应都尽量做到无量纲化，但专门开展过相关研究的体系仅有北京体系Ⅰ和北京体系Ⅱ。这一现状不得不让人担忧评估数据的科学性、客观性与真实性。

北京体系Ⅱ选择的是“赋分法”。〔16〕这种方法从本质上看属于对直线法中“极值法”的运用，具体的操作规则为：“在起评年，对正向指标将参评单位中三级指标成绩最好的评价结果定为满分，其余单位按比例折算。……对逆向指标将参评单位中三级指标成绩最好的评价结果定为满分，成绩最差的评价结果定为6成分，其余单位按线性插值折算。……从第二年起，可继续按起评年的赋分标准赋分，从而实现时间上的纵向比较。”〔17〕

相较于北京体系Ⅱ而言，北京体系Ⅰ中涉及的具体指标项目较少，需要收集处理的数据也十分有限，故该体系采取的是较为简单的“相对数法”和“比值法”。其中，“相对数法”将2000年数据定为基期，以后各年数据看作报告期，将报告期的指标数据和相应的基期指标数据相比，得出的相对数即为无量纲化数据，这种方法侧重于对同一评估对象（城市）在不同年度的同类指标进行纵向比较。“比值法”则选择以城市人口数（万人）为基数，将各个指标的原始数据同人口数相比，确定一定单位下的积分，再将其用于最后的计算，该方法则侧重于对不同评估对象（城市）之间的比较，并被该体系确定为主要采用的方法。〔18〕

指标无量纲化处理的备选方法种类繁多且难易兼备，背后原理也并不难掌握，只是在操作时需要投入较多的时间与精力成本。理论界对于这方面的探讨多集中于统计学领域，法学界也已有少数学者涉及。例如，有学者曾在理论层面提出了“Z分数法”这一“指标标准化”的操作方案，〔19〕虽然实际上仍旧是对直线法中“标准差标准化法”的具体运用，但这一设想却为理论界后续深入思考预留了空间。

（二）赋权方法选择较为集中

权重代表了某项指标在评估体系内的重要性，反映了指标对评估结果的影响程度。由于法治评估体系内含多个层级和诸多指标，权重的设计分配是一项颇为复杂的技术性工作，可以说法治评估的量化色彩最主要就体现在赋权这一步骤中。面对相同的指标和数据，采用不同的权重分配方式将会得出差异较大的评估结果，因此权重体系〔20〕的设置是否科学合

〔15〕 不同数据对应的指标性质其实是有区分的。例如，“国家赔偿案件执行率”和“流动人口犯罪率”这两项指标性质就明显不同，前者数值越大越好，后者则越小越好，在统计中将前者称为正向指标，后者则称为逆向指标，在无量纲化过程中具体的处理方式也不同。

〔16〕 这种方法主要针对第三级指标，也即最末一级指标，只有这一最末层次需要面对各种数据的直接运用计算，而二级、一级指标结果的计算只要综合各自权重相乘即可。

〔17〕 前引〔6〕，王称心书，第141页。

〔18〕 参见前引〔5〕，王光书，第139页以下。

〔19〕 参见前引〔12〕，陈海燕等文，第72页。

〔20〕 在法治评估体系当中，权重分配呈现多层次的体系化特点。具体的评价指标和各评价维度（也可称为领域层）都需要相应的权重设置。对领域层而言，权重是各领域层对目标层重要程度的最直接量化反映。对具体评价指标而言，权重反映的是具体指标在该领域层内的重要性程度。

理，不仅能够决定评估结果的科学程度，更影响整个法治评估的公信力与参考力。一般而言，确定权重的备选方法就多达数十种，但法治评估中典型体系的选择却显得较为集中。经分析发现，当下评估主要青睐德尔菲法、层次分析法、均分法和直接赋值的方式，这些权重方案的应用程度也多有殊异。

表 3　典型体系权重确定方案总览

分类	内容	均分法	德尔菲法	层次分析法	直接赋值法
实践体系	余杭体系		√		√
	昆明体系				√
	上海体系			√	
	北京体系 Ⅰ		√	√	
理论体系	江苏体系				√
	北京体系 Ⅱ			√	√
	社会主义法治综合评价	√	√		
	法治国情指数	√			

1．均分法：停留于理论设想层面

根据基本统计学原理，权重设置方法虽多，但如果没有特殊的理由就需要选择“均分法”分配设置。〔21〕我国几类实践体系中均未采用该方法。不过，国内理论界对运用均分法有着一些初步设想，如蒋立山设计的“法治国情指数”方案就指明要采用“平均赋权法”，〔22〕陈海燕设计的“社会主义法治综合评价”模型部分采用了平均法。〔23〕

均分法简单便捷，但面对法治评估这类庞大、复杂的体系时，就难以凸显出体系中各层次和具体指标重要程度之间的差异性。虽然我国有学者对其做出过设想，但实践中却鲜有尝试，使得均分法目前还仅仅停留在理论探讨层面。

2．德尔菲法：影响最大、备受推崇却亟待完善

德尔菲法（Delphi Method）又称作专家咨询赋值法。该方法是在系统程序的基础上，召集相关领域多位专家，在保证各位专家之间不发生互相联系、讨论的基础上，通过多轮次调查专家对问卷（问题）的看法，经过反复、多次的征询、修改，最终取得专家之间较为一致的征询结果。它最初的创设，是为了避免集体讨论中出现屈从权威或盲从多数意见的状况，其实质是“利用专家集体的知识和经验，对那些带有很大模糊性、比较复杂且无法直接进行定量分析的问题，通过选择一批专家多次填写征询意见表的调查形式，取得测定结论的方法”。〔24〕在确定权重阶段运用德尔菲法，除了组织确定专家小组外，一般需要进

〔21〕 参见［美］艾尔·巴比：《社会研究方法》，邱泽奇译，华夏出版社 2009 年版，第 162 页。

〔22〕 参见前引〔13〕，蒋立山文，第 17 页。

〔23〕 该模型下辖两个层次，其中 5 项一级指标的权重确定采用的是平均法，即每个一级指标的权重各为 20%，但是每项二级指标的权重确定则又运用了专家咨询主观赋值法（即德尔菲法）。参见前引〔12〕，陈海燕等文，第 72 页。

〔24〕 王春枝、斯琴：《德尔菲法中的数据统计处理方法及其应用研究》，《内蒙古财经学院学报（综合版）》2011 年第 4 期，第 92 页。

行多轮次的征询，且每一轮对专家的征询都具有不同的价值。例如，第一轮征询主要是以开放的态度接受专家的匿名意见，并整理相关意见分布情况；第二轮是要求专家根据第一轮反馈结果给出进一步评价意见；第三轮是重审相关权重意见，并需要仍旧游离于上下四分位数〔25〕之外的专家写明具体理由，再次根据反馈统计专家意见趋势；第四轮则需要复核，要求组织者收回调查结果，统计上下四分位数和中位数，总结各种意见、观点与争论点。一般在经过四轮征询后，专家组成员的意见会逐渐收敛，从而取得较为一致的结果。德尔菲法的关键词就是匿名协商，通过组织者和各位专家之间的单独联系，使每位专家尽可能独立、匿名提出判断意见，并经过多次征询，在专家参考他人匿名评价意见的结果上，取得专家组关于问题的一致意见。〔26〕

在评估权重设置方面主要采取德尔菲法的是余杭体系，它分别召集了内部组、外部组、专家组三组评估专家团，通过征求专家组意见的形式，对评估体系内各指标权重值及得分值一同进行赋分，其中权重值赋分范围为1—10分，指标分值赋分满分为100分。另外，虽然该实践在“149体系”下创设了多达77项具体评估内容，但三组专家团所需要进行评价的对象仅仅是“九大项目”。

德尔菲法之所以备受推崇且影响广泛，很大程度上要归因于余杭体系的成功运作。德尔菲法具有匿名性、统计性、反馈性、收敛性的特点，在预测、评价、统计领域均得到了广泛的运用。但是，该方法也并不必然客观精准，在具体操作过程中，专家的专业程度、主观偏好、自信程度等都会对权重数值的精确性造成很大影响，并且，在多轮次求得意见收敛的过程中，不仅需要倾注大量的重复操作成本，还有可能使得专家有意朝中位数靠拢的情况发生。虽然余杭体系声称其在指标设计与权重分配环节运用的均是“德尔菲法”，〔27〕但其实主要是种简化的德尔菲法环节运用。

首先，德尔菲法的多轮征询在余杭实践中并未得到实施。余杭体系虽然成功组建了专家组，但从现有公开资料来看，该地区所进行的是单轮征询而非多轮征询。换言之，组织者在回收了初步反馈意见后便进入了计算得分与权重的阶段，并未针对回收的数据开展统计学意义分析，也并未向专家组成员进行信息反馈。简化模式虽然极大地节省了工作量，却也使得相关意见并未得到有效收敛。

其次，因为缺乏有效的意见收敛过程，无法对指标权重得出较为统一的意见，所以针对收集到的初步信息，余杭体系选择了去掉最高数值和最低数值后再求出平均值的做法，这样最终得出的平均值便是需要纳入最终计算程序的权重数值了。这种方式是单轮征询的必然选择，也是数据处理中较为常见的选择，但就复杂评估体系而言，简单的计算平均数未必能够得出令人信服的结果。从统计数据分析角度而言，每一个数据的存在都是有价值的，平均值很难反映出数据整体的波动情况，虽然在后续的数据比较中，余杭也根据反馈结果计算权重的分布情况与方差，但相关分析主要是针对评估人员对不同指标的看法差异，

〔25〕 四分位数（Quartile）是一种统计描述分析方法，用于描述任何类型的数据，尤其是偏态数据的离散程度，即将全部数据从小到大排列，正好排列在下1/4位置上的数就叫下四分位数，排在上1/4位置上的数就叫上四分位数，排列在中间位置的就是中位数。

〔26〕 参见陈敬全：《科研评价方法与实证研究》，武汉大学2004年博士论文，第32页以下；刘学毅：《德尔菲法在交叉学科研究评价中的运用》，《西南交通大学学报（社会科学版）》2007年第2期，第22页。

〔27〕 参见钱弘道：《法治评估的实验——余杭案例》，法律出版社2013年版，第303页。

并未将数据的分析落脚到对权重赋值本身是否科学合理的考量。

最后，在计算最终结果时，面对内部组、外部组、专家组、群众问卷调查四类不同评估主体给出的不同得分，余杭体系却又选择了直接赋值方法去分配权重。其中，内部组、外部组得分分别占 17.5%，专家组得分占 30%，群众满意度调查得分占 35%。这一做法不仅与评估体系整体的方法论要求不符，也显得尤为突兀，毕竟对缘何做出这种权重分配，评估组织者也并未给出科学合理的解释。

3．层次分析法：量化程度、技术要求、应用难度均属最高

层次分析法（Analytic Hierarchy Process，简称 AHP）〔28〕是美国运筹学家 T. L. Saaty 于 20 世纪 70 年代初提出的一种定性和定量相结合的多目标决策分析方法。〔29〕这种方法通过将评估对象分解为不同的影响因素，使之形成层次结构体系，再引入“两两比较”的方法确定各层次中具体因素的重要程度，从而得出每一因素、每一层次的具体权重。换言之，AHP 是将复杂的问题条理化、系统化、清晰化，尤其适合一些层次较多、体系庞大复杂、需要引入较多主观判断的评估对象。

以北京体系 I 为例。评估组织者首先对目标进行层次结构转换，然后召集 5 人专家小组，根据所需评价集合及元素情况，制作并发放专家评分表，再依据专家反馈信息构造出判断矩阵。当然，得到的判断矩阵需要经过复杂的运算与检验过程，如果有数值没有通过单层一致性检验或总排序一致性检验，则需要重新调整矩阵，直至权重数值通过检验为止。与德尔菲法不同的是，专家给出的评分并非每项指标的直接权重数值，而是各个指标之间两两比较所得出的标度值，这种汇总意见的方法也是层次分析法的特色之一。〔30〕在运用 AHP 计算出具体权重结果后，评估组织者又采用专家直观综合法〔31〕得出另一组权重数据，并对两者进行综合计算后得出该体系的最终权重结果。总之，北京体系 I 虽然主要采用了 AHP，但其权重的分配则是 AHP 和德尔菲法的综合运用结果。

北京体系Ⅱ也声称采用了 AHP，具体实施过程和北京体系 I 大致相同。不过，北京体系Ⅱ的指标设计更为全面、复杂，面对被细分为三大层次的百余项具体指标，组织者共邀请了三十多位相关专家参与了此次评估。出于简化计算过程的考虑，北京体系 II 仅有一级指标的最终权重是按照 AHP 计算得出的，而三级指标中多数权数拟定为 1，少数比较重要的拟定为 2，二级指标则由三级指标的权数相加得出。因此，北京体系Ⅱ可以看作是 AHP 和直接赋值法的综合使用。

上海体系是国内少有的纯粹运用 AHP 的实践体系，其在构建之初就依循了目标层、准则层、指标层的层次分析的思路，整个体系的权重确定严格遵循了 AHP 操作流程，在咨询了 11 位专家的基础上，借助专业的统计分析软件，对体系内 52 项指标进行了科学赋权。

因为涉及数学建模与构造矩阵，AHP 已成为评价精度、技术要求、应用难度均属最高

〔28〕 层次分析法所需要进行的运算量很大，并且较为复杂专业，本文仅提供一种对 AHP 的通俗性解释。AHP 在实际运用中往往需要借助专业软件进行相关计算。

〔29〕 参见王莲芬、许树柏：《层次分析法引论》，中国人民大学出版社 1990 年版，第 2 页。

〔30〕 参见前引〔5〕，王光书，第 132 页以下。

〔31〕 专家直观综合法其实就是专家打分法，只不过在收回专家意见、利用算术平均数计算出每项指标权重之后，又利用了比例法对权重数值进行了修正，使得每层次权重相加总值均为 1，以便于后期和 AHP 得出的结果相统一处理。该方法和余杭体系采取的德尔菲法原理基本一致。

的备选方案，但是，这种定位并不必然意味着它就是最优选择。首先，在涉及的指标较多时，AHP 的标度工作量十分巨大，容易造成法学评估专家的判断失误与混乱。其次，该方法对于标度专家的数量和质量重视不够，检验矩阵的标准也多关注数值的一致性而非合理性。〔32〕最后，对于刚刚接触量化工具的法学实践界而言，应用技术要求较高，难以高质量地转化为实践。

4．直接赋值：简便易行却解释力差

除了前述常用的几种统计学方法之外，现有地方实践中还经常采用直接赋值方式。简言之，就是评估组织者在设计指标体系时，已经自行将每部分的权重情况进行了规定。除了前文提及的余杭体系、北京体系Ⅱ之外，江苏体系（试行）、昆明体系采用了此种方法。

前文已指出，在未有明确特别说明情况下，分配权重通常采取均分法，德尔菲法与 AHP 则是因其已被证明的科学性而广泛运用于权重确定过程中。但对于直接赋值法而言，现有实践并未对其有效性、可信度等选用理由给出科学说明，我们既不了解权重如此设置的理由，也不了解权重具体得出的过程，总体而言，该方法说服力较差，不建议使用。

（三）运算规则选择一致又多有变通

现有评估体系在计算最终结果时都基本采取了加权求和法，但在具体运用方面又根据评估体系状况各做变通，归纳后大致有以下几类：

表 4　典型量化实践体系计算方案总览

主体区分	内容	典型加权求和	借助均值法	计算软件
政府支持	余杭体系		√	
	昆明体系	√		
学者自拟	北京体系Ⅰ	√		√
	上海体系	√		√

1．典型加权求和法最为直观

一般而言，在分配权重时采用 AHP 的评估体系，计算结果时多会选择直接运用加权求和法，上海体系就是典型代表。由于前期已对各指标数据进行了无量纲化处理，组织者又藉由 AHP 求得了确定权重值，故上海体系结果运算仅需要借助加权求和公式即可。可以看出，AHP 帮助上海体系完成了大量的前期数据处理工作，又因其涉及的评估主体单一，故最终结果计算过程也显得相对简单，可谓是加权求和方法的最典型应用。除此之外，一些采用直接赋值法和均分法的评估体系，由于权重值已被明确给定，故仅需要对相关指标数据进行处理后，便可采用加权求和公式计算出最终结果，也可看作加权求和法的典型应用。

2．引入均值法配合简化德尔菲法

引入均值法计算评估结果的方式由余杭体系首创。由于仅进行了单轮征询且评估主体多元化，单纯运用典型加权求和法无法得出最终结果，故此，组织者先去掉专家打分的最

〔32〕 参见吴殿廷、李东方：《层次分析法的不足及其改进的途径》，《北京师范大学学报（自然科学版）》2004 年第 2 期，第 264 页。

高分和最低分，计算出每项法治条件的平均得分、平均权重及其在总体系中的相对权重，随后再通过加权求和法计算出最终结果。这种方式较好地配合了专家意见并不趋同的现实，却有些为计算而计算的倾向，期待组织者针对为何引入均值法而不进行多轮征询取得一致意见，做出科学的解释。

3．利用评估软件简化计算过程

法治评估体系的运行需要投入大量的人力物力资源，如需实现其纵向横向对比功用，就需要在多个年份、多个地区开展多次评估，这不仅会产生极高的运行成本，还必然会衍生出极大的运算工作量。并且，出于稳定性考虑，指标体系在具体运用中不会出现较大变动，长效、多地评估机制的建立就意味着一系列运算方面重复工作的产生，这无疑又是不必要的成本投入。考虑到这些因素，北京体系 I 开发了相应的评估软件，并将其运用于具体数据处理及计算，希望以此实现评估体系的长效、多地运行。启动“城市法治环境评价软件”后，仅需依照提示输入搜集到的数据，便能直接计算出结果。这种做法十分科学便捷高效，不仅提高了结果的准确性，更为评估系统的长期运用打下了良好的基础，但是，由于开发时间较早，这一软件也存在很多需要完善之处。例如，该软件仅能提供最终结果计算和数据查询功能，具体数据比较、多年结果比较等重要后续功能还有待进一步开发。

三、量化方法运用中凸显的问题

总体而言，作为法治评估最具特色也最为核心的部分，当下量化评估工作中还存在不少问题，有待未来作进一步完善。

（一）整体局面复杂混乱

法治评估量化方法内部蕴涵了多种可能与组合。这一方面体现出法治量化方法运用中的复杂、混乱局面，另一方面也带来了一系列追问：这些量化方法科学与否？选择一种量化方法的理由为何？采用不同方法得出的结果之间是否具有可比性？这些追问背后隐含的，其实是当前学界对于法治量化方法认识的匮乏。在大规模援引统计学知识、政府绩效评估实践、国际评估方法的同时，法学界并未就“法治”需要何种量化方法进行深入的思考，由此造成的最直接结果便是评估数据比较功能的虚置。时间、地域意义上的评估结果与数据可比性的缺失，无疑不利于对国家整体层面法治建设状况做出客观判断。对量化方法认识不足所造成的运用混乱，使得国内法治评估陷入了一种尴尬境地：各地的指标体系与评测方法看似能够自成一派，但其解释力与认可度却很有限，甚至于不同评估体系之间都难以获得互相认可。故此，如何选择科学适当的方法，力争在保有“个性”的情况下挖掘量化方法背后的“共性”，并最终形成一套科学的法治量化方法论体系，可能是今后评估法治工作中需要引起重视的一大重点与难点问题。

（二）具体应用多存在瑕疵

现有量化方法在实际运用过程中也存在瑕疵。这里以余杭体系为例。首先，虽然组织者声明主要采取的是德尔菲法，但对于缺乏有效意见收敛的单轮征询结果而言，其科学程

度本身就值得商榷。毕竟单轮征询得出的结果只是专家们对指标情况最原始的主观看法，它必定极大受制于评估人员自身知识水平、身份环境等诸多不确定因素，看似剥离了权威意见对他人可能造成的影响，却也丧失了通过匿名信息反馈过程取得对评估事项全面了解、客观对待的机会。面对宏大的法治建设状况，希望专家组成员在短时间内仅靠参考性资料，通过一次性打分程序就得出科学合理的结果，恐怕并不现实。参与过评估的专家就曾直言："专家打得这么准确，好像真是专家厉害，作为打分者之一的我有点担心，我是否真的这么了解余杭的法治建设现状，把握得这么准吗？答案是不敢肯定。"〔33〕其次，德尔菲法在运用中需要进行较多的数据处理，专家积极程度、专家权威系数、变异系数、协调系数等等，都是需要组织者做出的基本反馈，这些工作不仅能够帮助我们做出科学判断，更有助于反向检验评估体系内指标的合理程度，但现有资料中却并未对此做出过任何说明。有学者曾根据2010年余杭体系专家组的打分情况，利用克朗巴哈系数法对指标本身进行了信度检测，发现该体系的9大指标中有3个指标的克朗巴哈系数值未达到0.6，并且指标之间的关联程度也并不能令人满意。〔34〕再次，对数据的处理并不慎重。对于法治评估工作来说，误差的存在是不可避免的。但是，在余杭实践中，对于未能收到的回复信件和存在部分数据信息遗漏的评分信件，评估组织者都统一采取了删除有误样本的处理方式，这种直接减少样本数量的方式看似简单，却可能造成更大测量误差。〔35〕

一些指标体系的直接赋值方式难以寻得科学解释支撑。例如，北京体系Ⅰ采用多种方法综合计算权重，本来是综合德尔菲法与AHP的有益尝试，却并未给出这种综合处理的理论依据，显得过于草率。上海体系虽然贯彻AHP，数据来源却过于单一，法治评估成为了地区范围内的民意调查，体系内的组合权重计算也存在令人费解的地方。

（三）实际过程缺乏公开

当前，仅有余杭、北京、上海地区针对法治的量化过程做出过不同程度的信息公开：余杭地区在每年发布评估报告时都会对量化量过程做出简要说明，在相关专题书籍中也曾对量化问题做出过专题探讨；北京、上海地区也曾以书籍或论文形式对量化问题做出较为专业的解释。除此之外，其他地区对于量化方法、量化过程都仅有一笔带过的说明，更遑论对相关数据进行完善解释与分析了。

进行法治评估的最终目的并不仅在于得出一个具体分数。作为一种以纠错为主导功能的评价工具，计量过程中所涉及的方法、数据等都需要尽量做到透明公开，这不仅是提升法治评估公信力的最佳选择，也是保证评估科学性的必然要求。只有对相关量化过程进行公开，才能受到专业化的质询与审查，毕竟，只有经得起理论与实践检验的评估体系，才能真正称得上客观、科学、合理的法治评估工具。另一方面，对于那些带有瑕疵的方法而言，信息公开更有助于组织者对评估方法、乃至评估体系本身做出改良，这就又助推了法

〔33〕钱弘道：《中国法治增长点——学者和官员畅谈录》，中国社会科学出版社2012年版，第47页。

〔34〕克朗巴哈系数是信度测量的基本方法。一般认为，研究中该数值至少应达到0.7，实际操作中该数值最少也需要达到0.6。具体检验数据，参见孟涛：《法治指数的建构逻辑：世界法治指数分析及其借鉴》，《江苏行政学院学报》2015年第1期，第126页。

〔35〕相关余杭实践，参见前引〔3〕，钱弘道书，第229页以下。关于样本误差的其他处理办法及理由，参见上引，孟涛文，第125页。

治评估整体质量的提升。综上，法治的量化过程、量化方式都需要公开，这种公开不仅需要面向决策者，更需要直面研究者和公众。

（四）缺少结果论证机制

法治评估的结果是否科学可信，是各大评估体系必须予以回答的首要问题，但现有典型实践体系均未对该问题给出令人信服的解答。虽然各类体系的评估活动声势浩大，但都没有采用科学的方法对其评估结果的可靠程度做出检验。这种结果论证机制的缺失成为当下各类评估体系自说自话的重要原因。

一般认为，对评估体系进行科学审查需要关注信度和效度两大技术性指标。从学理层面解释，信度是指使用相同研究技术重复测量同一个对象时得到相同研究结果的可能性。〔36〕例如，如果让余杭体系中的专家组针对同样的指标进行多次打分，是否会得到一致的结果？如果运用昆明体系针对昆明地区开展多次评估，得到的结果会不会有差异？这些都是信度检测需要回答的问题。信度检测可采用的方法有很多种，前文提及的克朗巴哈系数法就是典型之一。虽然几类典型评估体系并未就该问题给出过相应的说明，但有学者曾利用余杭体系公布的数据小范围内检测了信度状况，得出了该体系相关信度值不高的结论，〔37〕值得我们反思。效度是指实际评估在多大程度上反映了概念的真实含义。〔38〕评估体系的效度越高，则评估结果越能展现评估对象的真实状况。对于法治评估而言，如果评估体系的指标设计不能较好反映地方法治的基本特征与主要内容，那么该体系的效度也不会很高，即便相应评估结果信度再高，也无法证明评估结果具有实际价值。目前，学界对于法治评估指标体系的效度检测还未有涉及。

世界正义工程的法治指数评估项目早在 2010 年就采用统计审查制度，来检验指数的信度、协调性和稳健性，涉及信度、内部协调性、外部协调性和稳健性四种目的不同的操作方法，运行较为成熟。〔39〕新加坡南洋理工大学也曾针对我国 11 项政府绩效第三方评估项目做出过评估，具体囊括了“独立性、相关性、效度、信度、易懂性、功能性”6 个维度，并得出我国相关第三方评估在功能性和信度方面表现最差的结论。〔40〕

四、完善法治量化的可行性路径

通过对典型量化体系的梳理和评述，我们认为，当下法治评估量化方面存在的问题，从源头上看都是由对量化方法的重视不足、了解不深所引发的。具体而言，相关认识不足使得一些评估体系在选用量化手段时天马行空、随意创新，形成了混乱复杂的运用局面；这种局面又反过来影响了组织者对具体量化方式的把握，从而出现了方法应用上的瑕疵；对于方法把握不足又催生出不自信感，使得组织者选择淡化定量过程，当然也就无法做到

〔36〕 参见［美］艾尔·巴比：《社会研究方法》，邱泽奇译，华夏出版社 2005 年版，第 137 页。

〔37〕 参见前引〔34〕，孟涛文，第 126 页。

〔38〕 参见前引〔36〕，巴比书，第 140 页。

〔39〕 参见孟涛：《法治的测量：世界正义工程法治指数研究》，《政治与法律》2015 年第 5 期，第 24 页。

〔40〕 See W. Yu & L. Ma, *External Government Performance Evaluation in China: A Case Study of the "Lien Service-oriented Government Project"*, 35(6) Public Money & Management 431 – 437 (2015).

相关信息的透明公开；而量化过程的封闭和相应结果审查程序的缺失，则直接导致了评估体系无法接受质询、获得改进，这又更加不利于量化方法科学认识的形成。四大问题以相互勾连的形式结成了恶性循环怪圈，将法治评估牢牢圈定在内，而摆脱这一桎梏的最佳路径无疑就是从根源入手——重视并开展法治定量方法相关理论研究，以形成体系化的法治量化方法论，指导未来实践。诚然，建构法治量化方法论体系需要具备多学科的知识储备、积累丰富的实践经验，也必然将是一项长期且缓慢的任务。本文无法对该体系做出宏大完善的构想，仅就当下的法治评估实践，提出几点建议。

第一，理论研究中对定量问题给予更多关注，凸显量化方法的重要地位。一方面，法治作为社会因素的存在状态是可以量化的，以法治指数为代表的法治评估，本质上就是一套量化数据的综合展现，能够对法治建设起到评价和指引作用，对实践具有重要意义。另一方面，对于量化方法的强调又源自法治评估领域的研究现状：有关法治评估的定性研究层出不穷却难有突破，而较为技术性的定量研究却备受冷待。故此，在完善法治量化的过程中，确立定量方法的主导地位，是对法治评估自身特性与研究态势的双重回应。当然，强调定量分析并不意味着抛弃定性分析，而是要求站稳科学化定量的基本立场，视评估的具体阶段与需要不同，综合运用两种手段。具体而言，在信息收集阶段，虽然可以选用无量纲化方法对数据进行量化处理，但依据政策分析学相关理论，还有必要针对取得的信息开展系列化的预测定性分析。〔41〕在评估具体开展过程中，权重方法、计算规则的运用，的确是定量分析发挥效用的主要阵地，但当评估进入结果论证与数据分析阶段时，单纯的定量方法就不足以支撑起整个评估结果的理论架构，这时需要综合运用定性与定量两种分析方法，既用实证数据说话又有理论价值分析，这样才能比较全面反映法治的状态与效果，为法治评估整套体系的运行画上圆满的句号。〔42〕

第二，培育综合评估机制，激发量化法治的社会活力。当前主要存在三种不同的法治评估推进模式：政府实际操作的内部型评估，第三方独立操作的外部型评估，政府发起委托第三方开展的综合型评估。内部型评估更侧重于考核，虽然推行阻力较小且容易制定规范化标准，却在公信力与说服力方面留有缺憾。外部型评估则因其专业性与大众化，在民众中具有较大的信服力与亲和力，较为科学且容易被公众接纳，但这种模式的权威性又会大打折扣，并且评估结果难以进入决策层面，缺乏执行力。综合型评估则能够扬长避短，兼采二者之长。其一，该模式能够得到政府部门的鼎力支持，在数据收集处理阶段具备天然优势，且数据的具体运用者又是独立第三方机构，极大避免了主观选用数据的情况发生。其二，指标体系建构、权重分配、结果计算等环节均对专业水平要求较高，而第三方机构不仅立场中立，还往往比政府具备更为充分的智识资源；由政府实际操作评估不仅容易受利益因素干扰，繁琐的评估流程也会极大占用政府机构的有限资源。从量化方面而言，如果学界能够通过讨论尽早形成相对统一的认识，将法治评估模式确定化，也有助于终结量化方法运用混乱的局面。特别是，一个成熟的现代社会，应该有相对发达和权威的社会评价体系，这也是促进法治良性发展的重要社会机制与条件，需要政府和社会共同营造，从

〔41〕 参见刘作翔、冉井富：《立法后评估的理论与实践》，社会科学文献出版社 2013 年版，第 133 页。

〔42〕 类似主张在立法后评估理论研究中也有提及，参见上引刘作翔等书，第 131 页以下。

政策、信息公开等方面促进和规范。

第三，推动法治评估的整体理性化公开，以打破现有的评估量化封闭格局。法治评估无疑需要极大提升其公开程度，但从量化角度考虑，需要的并不是一刀切式的全盘信息公布，而是一种理性整全公开。这种公开包含了评估过程与评估结果两大板块，并需要根据不同的角色立场做出区别要求。为了便于理解，可以将法治评估涉及的角色大致区分为两类：推动者与接纳者。推动者主要包括评估组织者与信息提供者。结合主流实践状况来看，前者主要是组织评估的科研院所、高校等，后者则主要是掌握大部分权威数据的政府机构或一些民间统计调查组织。接纳者是指法治评估活动的受众群体，主要有决策者、学界研究人员、普通民众三类。评估过程的公开需要从推动者角度考虑：相关数据信息的主要提供者需要对评估数据的收集处理提供支持，对于不涉及保密的相关信息进行公开，便于评估组织者选用；评估组织者则需要公开具体量化过程，包括评估主体概况与选择理由、评估内容选择依据、数据收集处理状况、量化（权重与计算）方法的选取理由等等。评估结果的公开主要是从接纳者角度考虑，在全面公开的基础上，合理考虑不同受众的接纳需求，做到详略得当、有的放矢，最大程度地提升评估结果的实际效益。例如，针对普通民众而言，需要将公开侧重点放在对评估整体结果说明与热点关注项目分析上，对具体的量化方法与过程则需要尽量做到通俗化、大众化。针对学界研究人员而言，需要评估组织者对评估中的量化技术性与理论性问题进行详实、科学、严谨的论证说明，并以开放态度接受业内同行评议与质询，以此为契机进一步完善相应量化模式与方法，最大限度保证评估结果与实际状态相符合。对于决策者而言，则需要注重揭示量化数据背后隐含的现实问题，将公开侧重点放在量化评估的导向与预测功能方面，充分发挥社会评议机构的优势。

第四，破解现有评估的单线性局面，通过多维互动建构反馈与修正机制。评估过程公开的不足与量化结果论证机制的缺失，使得当前法治评估活动在操作完善性方面大打折扣。并且，反馈、论证这些重要环节的丧失，又造成了法治量化的封闭性与单线性，这不仅有碍法治评估体系理论实践的良性发展，更不利于社会评价机制的培育完善。现今法治评估活动的单线性主要表现为，法治评估同被评估者互动较少，难以接收有价值的评估结果反馈，缺乏合理的检验与修正机制。依据系统论的基本观点，一个系统的发展能力主要来自于其内部的反馈机制，特别是对负反馈的回应式修正，更是维持系统向特定目标运行的重要方式。〔43〕 因此，破解法治量化评估的单线局面，需要引入评估结果的反馈机制。为了最大限度保留社会评价机制的中立优势，被评估者仅能在评估结果公布后进行反馈。在法治评估方案还未成熟之时，所有实践都是一种实验主义下的“试错”活动，被评估方作为评估活动的切实相关者，当然有立场也有理由就“实验结果”给出自身的合理看法。另外，作为主要的评估信息提供者，被评估者对各类数据的性质与量化处理方式把握也最为直接深刻，在评估后充分吸纳其关于数据运用的反馈意见，无疑能够助推法治评估量化模式成型，并促进法治建设的各项具体工作推进。

以上四点建议并非孤立存在，它们是一个完整的逻辑发展过程，单独做到其中一点并

〔43〕 See A. Rosenblueth, N. Wiener & J. Bigelow, *Behavior, Purpose and Teleology*, 10 (1) Philosophy of Science 18 – 24 (1943).

不能对法治量化起到太大的改进作用。总之，加强量化理论研究，科学对待定量问题是我们需要做到的前提性应对；在进行了充分深入研究，能够将跨学科的量化方法同法治评估良好结合之后，再选用综合评估模式，充分调动社会力量践行评估活动，并将得出评估结果依循理性化方式公开，以便于接受同行评议、政府反馈与民众质疑；再从评议、反馈与疑问中发现法治评估的不足之处，从而能够有针对性地改进法治评估体系，丰富法治量化理论与实践研究成果。由此，完善法治量化就成为一条线索，串联起法治评估的整个流程与所有相关主体，成为提升法治评估实践生命力、公信力与执行力的重要突破点。

Abstract: Quantitative evaluation of the rule of law can provide direction and technical guidance for the construction of the rule of law. The scientificity of quantitative methods used in the assessment of the rule of law will directly affect the rationality, execution, and social effect of the results of evaluation. A comparative analysis of some typical quantitative assessment systems in terms of the methods for data collection and processing, weight setting, and index calculation shows that there are many problems with the quantitative methods currently used in practice and theoretical studies, including complicated and disorganized assessment methods, unsound applications, non-transparent assessment process, and the absence of result verifying mechanism. China needs to face up to the problems in the current system of assessment of the rule of law, focus on practical social effect, and explore and construct a sound methodological system for the assessment of the rule of law.

Key Words: index of the rule of law, quantitative evaluation of the rule of law, quantitative methodology of the rule of law