

## 论司法评鉴的必要性<sup>[\*]</sup>

周 贇

(厦门大学 法学院,福建 厦门 361005)

[摘要]人们往往基于司法的独立性、司法评鉴的逻辑倒置以及经验中司法评鉴效果不佳等方面反对设置专门而系统的司法评鉴机制。这些反对理由要么过分强调了司法或司法评鉴的某个侧面,要么陷入“倒洗澡水把小孩儿也倒掉”的经验主义泥淖中,因此,并不足取。虽然司法工作和法官职业具有特殊性,但这种特殊性,尚不足以构成对法官进行业绩考核评价否定的充足理由。在目前要塑造“无考核的法官”,可能会弊大于利。当然,各种反对理由的存在,也确实提醒我们,在设置司法评鉴机制时,应特别注意司法评鉴的复杂性、司法评鉴可能的负作用以及平衡司法评鉴与司法机关依法独立行使司法权等问题。

[关键词]司法评鉴;司法的独立性;行政化倾向

DOI:10.3969/j.issn.1002-1698.2023.08.011

### 一、反对司法评鉴的典型理由:逻辑的及经验的

欲系统探讨当下中国的司法评鉴,首先要面对的就是其必要性问题。从表面上看,司法评鉴存在的必要性本不应该成为问题,因为司法如其他有组织的公共活动一般,几乎天然地需要评鉴才可能对相关主体(包括个人和机构)进行更好、更准确的激励。但颇有趣味的是,在司法领域,反对绩效考核以及更为宏观层面的司法评鉴的声音,却并不少;甚至可以说,其理据还比较有力。大体而言,反对者的理由可以归纳为如下逻辑的和经验的两个方面:

首先,从逻辑上讲,第一,司法评鉴可能与司

法的独立性(在西方为“司法独立”,而在我国则是“司法机关依法独立行使司法权”)相冲突,因为评鉴就可能涉及相应司法结论,进而反过来影响司法的独立性,而司法机关独立行使司法权是讨论所有司法问题(当然包括司法评鉴)的前提,因为如果没有司法机关独立行使司法权,还存不存在真正的“司法”都将成为一个问题。显然,如果司法的独立性一旦被动摇、被伤害,那么,司法评鉴所带来的任何“收益”可能都无法抵消这一成本。<sup>[1]</sup>

此处或许有必要明确一个问题:很多活动都具有独立性,但为何独立性之于司法却是本质属性?进而为何唯独在司法领域会把对独立性的

作者简介:周贇,法学博士,厦门大学法学院教授、博士生导师,研究方向主要为法哲学、司法哲学。

[\*]本文系国家社会科学基金后期资助重点项目“司法评鉴的应然逻辑与制度构造”(22FFXA003)的阶段性成果。

伤害当作反对评鉴的理由?这是因为,司法的本质是判断,而“司法权既然是判断权,就必须赋予判断主体高度的独立性”,<sup>[2]</sup>“司法独立是司法公正和法治的前提,任何法官都应在个人以及职务活动中坚持和落实此一原则”。<sup>[3]</sup>更进一步讲,正是“司法独立赋予了法官以作出‘忤逆’民意之判断以可能性,从而可以避免法治社会中多数人暴政的出现”。<sup>[4]</sup>申言之,独立性是达致司法“判断”本质、司法之所以为司法的必要途径,因此相对其他活动而言,司法更加讲求独立性,也更加需要强调对独立性的坚守;也因此,如果司法评鉴确乎造成对司法独立性的伤害,那么,当然就得不偿失,进而司法评鉴也就没有了合法性和必要性。

第二,还有论者认为,至少今日中国并不需要专门而系统的司法评鉴、监督机制,因为中国已经存在“体系化、强密度的法官控制之制度系统”,而之所以仍有司法官<sup>[5]</sup>有“勇气”且有能力去违纪违法,根本上是因为这些制度在司法实践上功能甚弱、效力甚微,几乎都是一种无效的制度。既然如此,“那我们就应该断然制止周而复始的制度再生产步伐,因为我们稀缺的不是制度本身,而是制度要作为一种真正具备制约效能的实效制度存在”。并且,考虑到无论哪一种评鉴或监督,归根结底都是某种程度的不信任,因而“监督越多则信任度越低,司法监督亦不例外”,而“司法的基本逻辑是:因为信任司法所以将纠纷争议交给司法部门去裁判。来自社会民众的信任乃是司法之所以为司法的根本,一旦失去民众的信任,那司法也就失去了其立足之根基。而社会民众所信任的司法应该是自由的司法,即极少受各种内外机制监督制约的司法。因为被置于重重监督之下的司法必将是主体性深受重创乃至彻底沦丧的司法,而主体性不完整的司法当然是不自由的司法,这种自身主体性破碎的不自由的司法怎么能让社会民众信任呢?社会民众又怎么能信任这种自身主体的完整性都难保的司法呢?”<sup>[6]</sup>

第三,有论者则敏锐地观察到了司法评鉴中的“逻辑颠倒”现象,所谓:由于司法评鉴归根结底最终都只能通过影响具体的作为案件承办人的法官来实现,而法院审判的首要目的在于实现个案公正,当众多案件都达到这一目的时,案件质量就应当呈现较为理想的状态,这是实现司法正义的正常逻辑。但当前的案件质量管理等司法评鉴机制却将这一逻辑颠倒:也即,通过使司法活动符合某些指标来倒逼司法正义的实现;或者说,似乎只要经验中的司法活动符合某些指标,则司法正义就当然得到了实现。逻辑颠倒的司法评鉴似乎完全忽略了,由于在质量评估和绩效考核中,评估指数和考核指标的巨大压力形成于案件办理之前,这很可能不仅不利于正义的实现,反而导致个案办理的进程和结果被扭曲。事实上,经验中的司法官们已经“聪明”地采取如下策略来应对各项评鉴:首先是功利性地追求指标排名,并由此获取实质性的利益;如其不能,则尽可能上交矛盾,回避个人责任。<sup>[7]</sup>从这个角度看,司法评鉴的正面功能如何或许我们并不容易确定,但至少可以确定它具有如下消极影响:司法官的主体地位无法彰显。法官审判权被弱化,法官仍处于单纯的被管控者的角色,与管控者之间的张力增加,案件质量管理的认同度与参与度降低。<sup>[8]</sup>这几乎必然导致程序公正以及司法权威的受损,尤其体现在司法评鉴设置不合理的情形中,如过去(甚至现在仍有不少法院坚持)曾将“二审改判率”长期作为评鉴的重要指标,可想而知,这样的指标必定会导致法院系统内部的“沟通”,进而弱化上下级法院之间的内部独立性。

其次,从经验上看,一方面,由于司法官的工作性质不是体力劳动,也不是竞技比赛,而司法评鉴又总是无法绕开“以量论绩”的泥淖,最终必定某种程度地忽视对法官在具体个案中作出的创造性司法活动之考察,进而有损法官的职业尊严感。另外,基于“以量论绩”的逻辑,不可避免地要设置各种量化指标体系,而在没有对各项

指标设置的科学性和合理性作出充分研究之前,容易出现指标间彼此矛盾或人为利用漏洞操纵数据的情况,进而导致考核结果丧失参考性,甚至导致反效果。或许也正因如此,有学者就尖锐地指出:考虑到“一套量化的法官考核指标不可能有效测度法官的廉洁度、努力度和司法能力”,那么,“建立在这些指标考核结果之上的奖惩机制当然也就无法实现激励和引导法官廉洁、公正、高效地完成审判工作的预期目标”。<sup>[9]</sup>

另一方面,事实上,至少我国司法评鉴引入的各种指标,似乎并没有真正让司法变得更好、更良善,毋宁说,“逼着”法官们为了迎合指标而作出各种看上去更聪明但实则伤害司法的行为安排;相反,取消评鉴、考核之后,却使司法变得“更健康”了。经验中的某些数据似乎确实也佐证了这一点——有报道指出,司法评鉴中的各项指标让法官变成了埋头苦干的“计件工人”。法官们疲于奔命,苦不堪言,陷入“数字崇拜”,各种绩效考核直接影响干警待遇、评先评优、选拔提升;一些单位和部门单纯围绕考核项目安排工作,司法公正、司法效果和司法权威或多或少受到了影响。相反,在取消考核后,各级法院审判效率大幅提升。以南京市中级人民法院为例,2014年1月至9月,该院共受理各类诉讼、执行案件10616件,同比上升22%;审执结8022件,同比上升20.8%。各业务部门案件平均审理天数53.8天,同比下降7.4%。同一年,江苏全省法院受理案件1392440件,审执结1165234件,同比分别增长12.44%和9.52%。其中,江苏高院同比分别增长32.45%和29.86%。全省法院执结标的金额达626.26亿元,创历史新高,同比增长40.14%。一线法官人均结案157.48件,同比增加16.18件。同时,各项指标也均有不同程度的改善。以较早取消考核的徐州两级法院为例,2014年新收各类案件140126件,同比下降1.35%;而审执结140736件,同比上升10.45%。从收结案比看,徐州中院为90.39%,同比上升3.83%,基层法院为95.59%,同比上升10.11%。<sup>[10]</sup>

综合起来看,反对设置系统而专门的司法评鉴机制之理由大体上包括:第一,它可能与司法的独立性相龃龉;第二,它可能会导致司法公信、司法权威的损害;第三,它可能错置了目的与手段;第四,从实际效果来看,设置专门的司法评鉴机制也并不一定更好,毋宁说,很多时候主要表现为反效果。

## 二、对司法评鉴必要性的证成:反向的及正向的

应该说,所有经验上的因素,或者说主要具有经验意味的因素,不成其为反对司法评鉴的充足理由,因为经验最多只能证明某一种具体的做法在某个具体语境中没有取得良好效果,它永远不能,也不可能证明逻辑上的必然。譬如,前述提及的考核量化倾向以及所导致的不良后果,就可以很容易地作出这样的反驳:其一,并不是所有的司法评鉴都必定以量化指标为主,因此,即便我们能够证成量化的司法评鉴可能并不总是能让司法变得更好,也并不意味着司法评鉴本身的不必要。其二,如果司法评鉴更多地重视、尊重个性,而非像当前的某些司法评鉴机制那样动辄讲求所谓“系统性”“全国一盘棋”,并尽力淡化其中的“横向竞争”“横向比较”之意味,而不是像当前那样动辄演变成所谓法官之间的晋升“锦标赛”,<sup>[11]</sup>则很可能可以避免,或至少大幅减少当前司法评鉴所导致的不良后果——事实上,法学界已经有人观察到这一逻辑在司法领域的具体表现。<sup>[12]</sup>

那么,前述逻辑上的反对理由又能否成立?首先,关于司法评鉴与司法的独立性冲突问题。司法评鉴是否会伤害到司法的独立性?回答应是肯定的,可以说,无论我们将司法评鉴机制设计得多么精巧,它的有效运作都必定会在某种程度上造成对司法独立性的影响,甚至伤害。但是,必须立即予以明确的是,司法的独立性之所以重要,并不是因为它本身重要,而是因为唯有保障独立性,才有更大的可能让每一个司法环节的每一个司法官都能作出自己独立的判断,从而

通过多环节的相互制约、叠加(当然也可能相互抵消)来更大概率地避免司法不公,甚至司法腐败。<sup>[13]</sup>换言之,司法公正才是比司法的独立性更具目的性的目的,也正是在这个意义上,我们确实可以说,“失去公正的法官就是一种‘法治’的摆设,公正不仅是法官必备的品质,也是法官的本质属性,还是法官的职业本色,并是法官威信之基础”。<sup>[14]</sup>这意味着,当司法的独立性与司法不公、司法腐败之间可能呈现出某种紧张关系时,即便我们不能说“司法公正”必然总是处于更为优先的地位,至少也并没有什么充足的理由让我们可以毫不犹豫且片面地肯认司法的独立性总是具有天然优先的地位。这当然也就意味着,为了更好地追求司法独立的目的(司法公正),有时候适度地牺牲或侵蚀司法的独立性也并非完全不可取。当然,由于司法独立毕竟是导致司法判断权本质的必要途径,从这个角度可以说,如果没有起码的独立,司法也就不存在了,所谓司法公正当然也就无从谈起,因此,即便是为了制约司法腐败、司法不公而设置必要的司法评鉴机制,该机制也应以尽可能地避免伤及,甚至仅仅是危及司法的独立性为第一要义。从世界各地的经验来看,应该说,各个国家和地区对此已经作出了很多卓有成效的尝试,典型者如我国台湾地区就特别强调“坚守法律见解不得作为个案评鉴事由之底线”。又如德国强调在法官的“职务核心领域”不允许有任何的外部评鉴——在这里,所谓“职务核心领域”,是德国联邦宪法法院提出的一种划分:德国联邦宪法法院将法官职务行为划分为核心领域和外部秩序领域两个部分,并认为法官职务行为的核心领域,即判决行为、直接或间接为了准备判决以及接续判决等行为,除非存在显著错误,否则不属于职务监督的范围。此种见解被称为“核心领域理论”<sup>[15]</sup>。这些具体经验我们当然不必一定要认同(最典型的问题是,“职务核心”如何与“非职务核心”相区分,进而完全分离?),但它们的存在以及实际上取得的不错效果,本身就已经表明,只要处理

得当,司法评鉴与司法的独立性可以大体保持一种相互平衡—紧张但并行不悖的格局。

其次,关于司法评鉴与司法公信力的冲突。确实,对于司法这样一种“既无军权、又无财权”“既无强制、又无意志,而只有判断”的“三权中最弱的一个”公权机关<sup>[16]</sup>而言,其生命力就在于也只在于公信力,也即社会大众的信任,所谓“司法以人民的信任为基础而存在”。<sup>[17]</sup>而显然,由公权机关组织的司法评鉴很容易给社会大众造成这样的观感:原来“官方”自己并不放心司法官们以及司法活动,原来还有一批人“高于”司法官们以及司法活动(以至于他们可以评鉴后者),(待司法评鉴结论中可能揭示出一些问题来之后)原来司法官们、司法活动中居然有这么多的问题,既然如此,为什么要信任司法官们以及司法本身?

那么,这是否就可以构成反对司法评鉴的充足理由?回答仍然是否定的。这是因为,司法评鉴、监督机制的存在,只是承认司法一如其他所有的公权,都有可能被滥用、误用进而导致腐败,它的目的并不是要向大家“证明”或“表明”当前的司法有多少问题,毋宁说,是为了制约各种可能的腐败进而预防问题的产生。或者也可以作相反的思考,在一个所有其他条件都相当的两个不同社会中,一个社会中设有司法评鉴、司法监督机制,而另一个没有相应机制,大众将对哪一种司法抱以更大的信任?答案似乎更可能是前者。因此,当司法评鉴制度刚刚启动之时,或许会发现较多的问题,一时之间也确实可能会给社会大众造成前述观感,但一方面,这些问题只是通过司法评鉴被发现出来,它们本身并不是司法评鉴的“错”,而大家真正关注、担心的不应该是这“发现”或“发现机制”,而是相应问题本身;另一方面,如果没有司法评鉴的发现,这些本就存在的问题并不会自动消亡,甚至可能会变本加厉,进而从长远来看,一定会给司法公信、司法权威造成更大的伤害。或者再退一步讲,即便承认司法评鉴会造成观感的下降,我们也可以通过更

少地公开评鉴结论、更合理地运用评鉴结论等方式改良司法评鉴的运作机制,从而降低它可能给司法带来的负面冲击。换言之,司法评鉴导致司法公信、司法权威的损害,可能仅仅是经验中我们操作不当,而非司法评鉴本身的必然之恶(果)。

至于司法评鉴中的“逻辑颠倒”现象,首先必须承认,这确实无可否认,因为所有的评鉴都注定只能如此:通过规制事后的结果来倒逼事先、事中行为或活动的改良。但也正因为所有评鉴都注定只能如此,所以,除非我们可以证成司法活动具有某些或某种特别的个性,以至于“逻辑颠倒”的评鉴机制在其他领域都行得通却偏偏在司法的领域没有存在的余地,否则,我们就没有理由仅仅因之就否定司法评鉴。换个角度看,虽然在任何领域,评鉴都几乎不可避免地会导致某种扭曲,但如果我们可以找到相应的办法避免或减小这些扭曲,或者仅仅是认定这种扭曲带来的负效果比相应评鉴制度带来的正效果更小,甚至小得多,那么,就没有理由在司法领域否弃评鉴机制。

总之,尽管反对司法评鉴的各种理由都各自有一定的道理,但根本上讲,都多少具有“倒洗澡水把小孩儿也倒掉”的意味:它们虽然各自都指出了司法评鉴所可能带来的弊端,或弊端的某个层面,但应该说都不构成我们拒斥司法评鉴的充足理由。在这个意义上,确实可以这样来论断司法评鉴的必要性问题:至少就当下中国而言,虽然司法工作和法官职业具有特殊性,但这种特殊性,尚不足以构成对法官进行业绩考核评价否定的充足理由。在目前要塑造“无考核的法官”,可能会弊大于利。一是从目前“队伍”状况看,一部分进入员额的法官,其品德、业务素质距离理想法官的要求,仍存在一定差距,加之职业保障机制、督促激励机制缺失的因素,将难以保障案件质量和效率;二是目前法院面对着“喷发”般增长的诉讼案件,法院应对诉讼案件的数量多,难度大,不强化管理手段,包括加强绩效考

核,法院将难以有效运行,司法公信力也无从提升;<sup>[18]</sup>三是司法评鉴能够“正确把握审判形势、总结审判经验、增强审判能力、改进审判工作、推动国家法治化进程”,“评估结果是人民法院审判工作情况的综合反映,是进行审判工作管理、决策和评价考核各级人民法院的重要依据之一”。<sup>[19]</sup>因此,总体而言,司法评鉴确实具有很大的必要性。

### 三、构建司法评鉴机制的前置问题:语境上的及关键事项上的

换一个角度看,如果一个国家司法队伍基本没有贪腐、渎职现象,或至少没有贪腐、渎职文化,如果一个国家的司法水准足够高,一个国家司法的公信力足够高,一个国家整个民众的法律意识也足够高,以至于在这样的国家中,任何一个法官都有足够的意愿自觉抵御各种腐败的侵蚀,任何一个案件的司法结论都能得到普遍的认可和尊重,即便有些人不认可也仍然惯于从司法系统内部寻求救济(也就是说整个司法系统的公信力较强),任何一个民众都能自觉自愿地按照司法的逻辑和规律来看待司法活动,并积极地帮助、监督司法机关依法独立行使司法权,那么,在这样的国家,大概也没有必要构建司法评鉴机制,因为它根本就不需要司法评鉴,至少不需要成建制的、系统的司法评鉴。必须承认,虽然我国已经为达致这种状态付出了卓绝而有效的努力,但实事求是地说,至少我国当前尚未完全具备这些条件。

概言之,在当下中国,有很大的必要性设置司法评鉴机制——当然,前提是该司法评鉴机制本身具有较大的合理性。这或许也正解释了何以司法评鉴否定论始终没有成为理论界的主流,并且实践中也始终在进行着某种程度的司法评鉴,所谓“对司法质量进行评估,中国自古有之,但从未作为一个独立的体系存在,……20世纪90年代以来,对案件质量进行监督和评价的做法已经十分普遍,裁判文书评查、庭审质量评查

等工作成为法院保证案件质量的重要举措。2000年以来,经过各级人民法院的不懈探索和积极实践,案件质量监督管理工作有了显著的发展。2002年,最高人民法院提出建立审判质量与效率综合评估体系,在经过若干轮的调研、征求意见和专家论证、测试后,最高人民法院于2008年发布了《关于开展案件质量评估工作的指导意见(试行)》(简称《意见》),初步建立起人民法院案件质量评估体系,……”<sup>[20]</sup>——事实上,自《人民法院第三个五年改革纲要(2009—2013)》发布以来,人民法院系统内部从未放弃过开展这项工作,甚至可以说,越来越重视、加强这项工作。

当然,否定论者的观点、意见仍然值得尊重,这不仅仅因为任何严肃且真诚的异议本就值得予以尊重,更因为这些反对意见至少可用来作为我们构建司法评鉴机制的“清醒剂”:即司法评鉴机制的设计,确实需要足够的慎重,否则就极有可能“好心办坏事”。基于此,在对司法评鉴的必要性作了肯定之后,有必要对司法评鉴机制的设置作出至少如下预警式提示:

第一,从姿态上,首先,应充分意识到司法评鉴的复杂性。“司法工作并不是一项完全以结果产出为最终指向的活动,而是一项包含广泛政策目标与更多价值实现的工作,对这些政策与价值的实现程度,其实很难进行准确的测量与评价”。<sup>[21]</sup>承认这一点,首先意味着当前五花八门、叠床架屋式地不断颁布各种司法评鉴制度的做法肯定不可取,因为既然它如此复杂,就显然不可能、也不应该如此频繁地以政出多门的形式颁布如此多的制度来进行规制;进而也就意味着,今后如果要再颁行相应制度,必须慎之又慎,否则就不如不颁行。其次,应充分意识到司法是一种特殊的公权活动,这特殊性尤其体现在它的专业性和独立性上,因此可以说,唯有法官(包括退休的法官)才最有资格评鉴司法,或至少应该是让职业法律人或者具备相关能力和思维方式、理解法官司法活动的“行家里手”来主导司法评鉴。司法评鉴的目的应该是促进司法职业

共同体的自律、自治。最后,还要充分意识到司法评鉴,哪怕是所谓“科学、系统”的司法评鉴也并不能包治百病、手到病除,换言之,不宜对司法评鉴有过高的期望——这可称之为司法评鉴中的“谦抑原则”或“谦抑姿态”。

之所以应当在司法评鉴中坚持谦抑姿态或原则,一是因为有太多的实际上影响,甚至决定司法质量的因素,未必能够进入到司法评鉴的视野中,典型者如司法官为消解某些特定案件所带来的压力而作出的种种努力,又如司法官为协调各方立场所做的协调工作,等等。这使得司法评鉴的结果注定只是司法人员行为表现“不完全意义上的描述,并非终极评价。……不能把它当作评估司法人员行为表现或个人业绩的最终依据”;<sup>[22]</sup>二是因为司法评鉴只不过是司法系统人事管理与服务中的一个环节而已,它必须、也只能与其他既有人事管理或服务机制相配合才可能发挥出一定的作用,而并非像有些研究说的那样,是“既决定法官入额遴选的思维导向,又决定法官奖惩的评价标准和等级待遇的提升条件,是沟通法官单独职务序列、司法责任制以及法官职业保障制度的中枢机制”;<sup>[23]</sup>三是因为说到底,司法的良善运作最终取决于整个社会,尤其是整个国民的法律素养。

坚持司法评鉴的谦抑原则,就意味着欲更好地发挥其效能,必须有意识地将司法评鉴回嵌、并契合于司法系统内既有的人事、监察等制度,进而有机结合于一国人事、监察等制度,<sup>[24]</sup>而不应该动辄另起炉灶进而陷入叠床架屋的困境中。当然,最根本的还是要匹配一国基本的法律文化、法律观念状况,正如习近平总书记所言,“司法体制改革事关全局,要加强顶层设计,自上而下有序推进。要坚持从实际出发,结合不同地区、不同层级司法机关实际情况积极实践,推动制度创新”。<sup>[25]</sup>

第二,从评鉴目的来看,要避免陷入行政化,也即片面人事管控面向的泥淖中,并且不宜过分突出司法评鉴的横向竞赛、比较意味。无论司法

评鉴的目的是什么,首先可以肯定的是,它不应该是行政化的,或至少不应该是单纯行政化的。因为行政化的司法评鉴,几乎注定将导致司法机关内部的“管控—被管控”“领导—被领导”关系的强化,而这种强化本身几乎不可避免地会伤害,甚至根本动摇司法机关独立行使司法权(这也正是“逻辑颠倒”论者反对司法评鉴时所强调的);<sup>[26]</sup>与此相关联的另一个面向的伤害则是,行政化的司法评鉴几乎一定会导致对个案以及对司法官个性的忽视甚至压制,进而间接造成对司法独立性的伤害。当然,它也不应该主要具有的是横向竞赛、比较意味,因为过分地突出司法评鉴的这个面向,就将不可避免地导致评鉴对象对各种指标的片面迎合,进而扭曲具体的司法本身,并导致类似当前某些司法评鉴所带来的种种弊端。可以说,如果所谓司法评鉴最终主要具有的是行政化或者横向竞赛意味,则这样的司法评鉴不要也罢。

第三,从评鉴标准的设置来看,要避免“一刀切”的妄想,要充分兼顾顶层设计的一致性与具体情况的特殊性。就当下中国的司法评鉴领域而言,尤其值得注意的“具体性”至少包括以下两个层面(以法院为例):一是不同审级和不同层级的法院应当设置不同的评鉴标准。大体上可以说,法院的审级、层级越高,收案数量越来越少,反之,则数量越来越多,这一格局本身就已经意味着任何“一刀切”的量化评鉴标准都不合适。另外,从功能架构角度讲,基层法院的职能定位于分流案件、初步定调纠纷解决方案,而中级人民法院的职能则侧重于依法监督、纠错基层法院的司法活动,高级人民法院以及最高人民法院更多地可能具有的是司法政策性、方向性功能<sup>[27]</sup>——这两个层级的法院当然也会从事具体案件的审理工作,但即便是针对这些具体案件的审理,其也具有导向性、方向性、政策宣示性的功能。如果我们承认各层级法院的功能定位本就不同,则毫无疑问地,任何“一刀切”的评鉴标准也都不合适。<sup>[28]</sup>二是不同地域的法院应当设置

不同的评鉴标准。虽然法治、司法本质上讲求普适性、同一性,但对于“大”中国来讲,有些地域差别确实有必要予以充分的考虑,尤其是东西部以及城乡差别:以笔者现在所在的闽南地区为例,城区尤其是城乡接合部的基层人民法院工作量,与内地农村地区、更不必说西部农村地区基层人民法院的工作量,根本不是一个数量级;而在沿海发达地区,很多案件都明显比中西部地区的更为复杂,譬如说其中往往关联着的涉台、涉港(澳)、涉外因素(尤其是现在各自贸区的管辖法院)就往往是中西部地区法院很少需要面对的。因此,如果盲目地、“一刀切”地设置一个评鉴标准来“统一”评鉴不同地域的法院,很可能造成各具体法院的极大困扰。总之,很大程度上确实可以说,“对各法院而言,制定的考核评价办法只要能促进执法办案任务顺利完成,审判质效有所提升,就是好的考核办法,不存在‘放之四海而皆准’的考核办法”;<sup>[29]</sup>我们甚至可以说,任何片面追求“统一”“一致”的司法评鉴机制,不管它是否打着“科学、合理、有效”的旗号,都属于典型的不合理因而也不可能真正有效的机制,都应该坚决予以摒弃。

第四,从内容上看,要谨慎处理司法的独立性、终极权威性(所谓司法是法治系统内最后的救济措施)与司法评鉴之间可能存在的紧张关系。当司法评鉴不可避免地涉及司法官职务核心时,一定要慎之又慎,其中,尤其不能让社会大众产生这样的观感:司法评鉴者是司法官的“顶头上司”,可以对司法结论品头论足,甚至径直否定。因为一旦如此,则一方面,可能与既有的上诉审、再审、重审机制相纠缠、重叠;另一方面,还可能造成一种类似于当前涉讼信访之于司法的伤害,<sup>[30]</sup>进而严重伤害司法权威、司法公信。

#### 注释:

[1] 这可以说是反对以及担心司法评鉴的最常见、最根本的理由。详可参见赵昌平等:《“法官、检察官评鉴制度之探讨”专案调查研究》(我国台湾地区“监察院”97年度调查研究报告)等。

[2]《全面深化人民法院改革的理论基点和实现路径——解读〈最高人民法院关于全面深化人民法院改革的意见——人民法院第四个五年改革纲要(2014—2018)〉》(最高人民法院司法改革领导小组办公室,2015年)。

[3] The Bangalore Principles of Judicial Conduct, Value 1: Independence, 2002.

[4] Sandra Day O'Connor, "Judicial Accountability Must Safeguard, not Threaten, Judicial Independence: an Introduction", in 86 *Denver University Law Review*, 1 (2008—2009)。当然,这并不是奥康纳大法官的发现,有关独立司法的这一价值,早在托克维尔那里就已经得到了系统的揭示和特别的重视,详可参见[法]托克维尔:《论美国的民主》,董果良译,北京:商务印书馆,1989年,第318—320页。

[5]在本文中,“法官”与“司法官”两词有时可能会混用。之所以作这样的安排,是因为一方面,在我国,“司法机关”以及“司法官”实际上包括“法院”“检察院”以及“法官”“检察官”;但另一方面,没有人否认,“法院”(或“法官”)才是更典型的司法机关(司法官)(有关这一点的理论辨析,详可参见周赞:《论我国检察院的角色定位、具体权能及其尴尬》,《江汉论坛》2016年第4期)。基于这两个因素,本文实际上面向的讨论对象就是法院以及法官,但当相应结论同样可以适用于检察院和检察官时,也会根据上下文语境将其表述为“司法官”。

[6]刘练军:《如何控制法官》,《东方法学》2012年第4期。

[7]南京市中级人民法院课题组等:《法官业绩考核评价制度研究》,《中国应用法学》2018年第1期。持类似观点的文献还有杨铜钢:《法官绩效考核制度的非司法化困境及其调试》,《法制与社会发展》2022年第3期;祖鹏、王艳华:《对法官办案业绩量化考核制度的检视与反思》,《法律适用》2023年第2期;等等。

[8]许建兵:《中国特色审判管理机制构建之构想》,《法律适用》2009年第9期。

[9]艾佳慧:《中国法院绩效考评制度研究——“同构性”和“双轨制”的逻辑及其问题》,《法制与社会发展》2008年第5期。类似的担忧,亦可参见张建:《论法官绩效考评制度改革及其实践效果》,《法学》2019年第11期;张曦:《审判绩效考核的困境、缘由与脱困路径》,《上海交通大学学报(哲学社会科学版)》2019年第6期;等等。

[10]《江苏法院取消指标考核后审判质效情况之调查》,人民法院报网, [http://rmfyb.chinacourt.org/paper/html/2015-03/21/content\\_95681.htm?div=-1](http://rmfyb.chinacourt.org/paper/html/2015-03/21/content_95681.htm?div=-1)。当然,必须明确的是,并非所有研究都显示加强考核后反而取得更差的效果,譬如《武汉市中级人民法院关于2020年度办案质量检查情况的报告》(湖北人大网, <http://www.whrd.gov.cn/html/rdzl/cwhgh/14373839/2021/0930/20254.html>)、《上海一中院法官考核新模式详解》(该文以2019年上半年为样本,人民法院报网, [http://rmfyb.chinacourt.org/paper/html/2020-02/18/content\\_165082.htm](http://rmfyb.chinacourt.org/paper/html/2020-02/18/content_165082.htm))等报告就显示出相反的样态。但可以确定的是,近年来,各界已经开始反思这种做法,并且已经局部取消了相应“排行榜”。这种反思、取消

本身或许就意味着,反对者持有的这种担心、质疑,恰恰有其合理性。

[11]关于这一倾向在司法领域的体现,详可参见高翔:《中国地方法院竞争的实践与逻辑》,《法制与社会发展》2015年第1期;孙启福、吴美来:《案件质量精细化管理的局限及其克服——以最高人民法院〈关于开展案件质量评估工作的指导意见〉为中心》,《法律适用》2012年第6期;等等。

[12]王禄生:《我国司法改革的内卷化风险及其治理》,《法商研究》2022年第3期。

[13]司法决策是一个典型的可错性活动:任何一个环节、任何一个司法官都无法保证自己的结论客观、准确(譬如所谓“铁案”)。这意味着,在司法决策过程中,法官必定具有较大的裁量空间以及相对应的寻租空间,这进一步意味着有特别的必要设置多环节的诉讼体制:因为通过不同环节、不同具体主体的独立判断,既环环相扣、相互制约又相互叠加,才可更好地地致或接近更符合诉讼正义标准的结论。有关这一点,参见周赞:《司法决策的过程——现实主义进路的解说》,北京:清华大学出版社,2015年。

[14]王中:《法官的道德理性论》,北京:法律出版社,2017年,第3页。类似的观点及其论证,还可参见邢鸿飞、侯玲:《论我国司法制度的价值取向》,《学术界》2017年第7期;董翠翠、王淑荣:《新时代司法公正的社会主义核心价值观导向》,《社会主义核心价值观研究》2023年第1期。

[15]关于德国法中的“职务核心领域”理论,详可参见最高人民法院政治部编:《域外法院组织和法官管理法律译编》上册,北京:人民法院出版社,2017年,第436页。对德国“职务核心领域”理论的详细讨论,详可参见许政贤:《德意志联邦共和国法官职务法——以法官职务法庭的任务为中心》,许政贤:《民事法学与司法制度》,台北:新学林出版股份有限公司,2009年。

[16]参见[美]汉密尔顿等:《联邦党人文集》,程逢如等译,北京:商务印书馆,1980年,第391页。

[17][德]拉德布鲁赫:《法学导论》,黎冷、潘嵩译,台北:台湾实业文化有限公司,2004年,第118页。

[18]龙宗智:《法官业绩考核怎样才能“相对合理”》,《人民法院报》2017年7月9日。持类似观点的还有华小鹏:《法官绩效考核的终极目标及实现路径研究》,《法学杂志》2020年第10期;杨铜钢:《法官绩效考核制度的非司法化困境及其调试》,《法制与社会发展》2022年第3期;等等。

[19]《最高人民法院关于开展案件质量评估工作的指导意见(试行)》(2008年1月11日)。或应明确的是,这一主旨也体现在后来颁布的相应新文件之中,如《人民法院案件质量评估指标体系》(最高人民法院,2011年3月),又如《人民法院案件质量评估指数编制办法(试行)》(最高人民法院,2013年)以及《关于加强和完善法官考核工作的指导意见》(最高人民法院,2021年10月),等等。

[20]严文:《人民法院案件质量评估工作的历程》,《人民法院报》2011年10月23日。

[21]张榕:《中国法院能动司法机制研究》,北京:中国政法大学出版社,2015年,第189页。

[22]郭松:《绩效考评与司法管理》,《江苏行政学院学报》2013年第4期。作出类似观察和判断的文献,还可参见张建:《论法官绩效考评制度改革及其实践效果》,《法学》2019年第11期;张曦:《审判绩效考核的困境、缘由与脱困路径》,《上海交通大学学报(哲学社会科学版)》2019年第6期;等等。

[23]上海市第一中级人民法院课题组:《审判绩效考核与管理问题研究》,《中国应用法学》2019年第3期。

[24]司法评鉴与一国的一般公务员评鉴(在我国,由《中华人民共和国公务员法》及相应配套制度规范)的关系,不同的人有不同的认识、判断。本文认为,它们构成特殊与一般的关系,也就是说,司法评鉴是公务员评鉴的特殊分支。这意味着,第一,公务员评鉴制度当然地可以适用于司法领域;第二,如果司法评鉴制度作了特别规定的,应从司法评鉴制度——这一判断的反面是,司法评鉴只应着力于司法领域特有的问题、事务,对于已经或可以为一般公务员评鉴制度所涵括的,应尽量避免叠床架屋式的关注。

[25]习近平:《论坚持全面依法治国》,北京:中央文献出版社,2020年,第148页。

[26]有关这一点,可参见方乐:《法官责任制度的司法化改造》,《法学》2019年第2期;杨铜铜:《法官绩效考核制度的非司法化困境及其调试》,《法制与社会发展》2022年第3期;等等。

[27]必须明确的是,对不同层级法院的功能定位,是一个颇具争议的大课题,不同论者站在不同角度可能得出不同的结论。我国最高人民法院[《全面深化人民法院改革的理论基点和实现路径——解读〈最高人民法院关于全面深化人民法院改革的意见——人民法院第四个五年改革纲要(2014—2018)〉》,最高人民法院司法改革领导小组办公室,2015年]认为,“一审距离案件争议的事实更近,便于及时查明事实;二审重在解决诉辩双方对一审认定事实和适用法律的争议,实现二审终审;再审重在审查终审裁判的正当性,维护裁判的权威性、稳定性,最终实现法院裁判的终局性”。或有必要明确的是,类似的认识同样体现在《最高人民法院关于深化人民法院司法体制综合配套改革的意见——人民法院第五个五年改革纲要(2019—2023)》(最高人

民法院,2019年)以及《关于完善四级法院审级职能定位改革试点的实施办法》(最高人民法院,2021年)等文件中。但无论如何,大家都承认这一点:各不同层级、审级的法院,职能有所不同(学界的有关讨论及其梳理,详可参见龙宗智:《审级职能定位改革的主要矛盾及试点建议》,《中国法律评论》2022年第2期)。而只要承认这一点,对于此处的论述而言,就已经足够了。

[28]有意思的是,当前在面对是否“一刀切”的问题上,呈现出明显的矛盾性。其具体表现是:一方面,在观念上,几乎没有人否定各层级司法机关的不同,进而也没有人否认应当针对不同层级司法机关设计相应的司法评鉴机制,如《关于进一步全面落实司法责任制的实施意见》(最高人民法院,2018年)就明确,“根据各级人民法院承担职能的不同,科学设置司法人员业绩考核内容”(第22条);但另一方面,在具体制度设计、更不必说制度实践过程中,能够真正坚持个性、地方性的制度安排和做法,却并不多见——只要考察一下各地方司法机关颁行的关于司法评鉴之规定就会发现:它们基本就是对国家或中央层面相关文件的细化,有时候甚至只是“移植”而已。当然,这不单是司法领域独有的现象,应该说,尽管当前有所改善,但所有领域的立法仍明显都存在这一倾向(对我国立法领域此种倾向的考察、判断,详可参见杨景宇:《加强地方立法工作,提高地方立法质量》,《求是》2005年第14期;江材讯:《地方立法数量及项目研析》,《人大研究》2005年第11期;等等)。

[29]南京市中级人民法院课题组等:《法官业绩考核评价制度研究》,《中国应用法学》2018年第1期。

[30]笔者曾经这样评论过涉讼信访为何不可取:“它把一种与司法工作逻辑相左的工作机制(信访)引入到了法院系统内部,进而一方面造成法院将可能无法继续遵循基本的司法逻辑运行,更恶劣的是,它甚至会造成如下观感:原来法院系统自己也不相信自己按照司法逻辑得出的结论,否则,它何以在司法逻辑之外又设定信访来对根据司法逻辑得出的结论进行制约?……(它)很难不让人认定,原来法院自己也不承认依法判决的终审效力。而这无疑意味着法院系统公信力的全面崩塌。”参见周贇:《对法院系统内涉讼信访的思考》,《法制日报》2014年10月9日。

[责任编辑:邹秋淑]