

Oct. 30  
2023

全球治理学科动态  
2023 年第 5 期 (总第 29 期)

# 全球治理中的 绩效评估

中国社会科学院 (CASS)  
世界经济与政治研究所 (IWEPI)

本期执笔

孟思宇 张尊月 杨嘉豪  
苏山岳

中国社会科学院世界经济与政治研究所  
全球治理研究团队成果发布  
Global Governance Perspectives



## 专题序言

如何评估国际组织的绩效直接关系到全球治理的权威和效能。本期学科动态以全球治理中的绩效评估为主题，探讨了评估国际组织绩效的文本分析方法，探讨了评估体系的制度设计如何影响绩效评估的客观性，展示了不同国家在推动改革过程中如何利用绩效评估结果。本期学科动态旨在通过多篇近期文献阐释绩效评估的意义，检视现有评估方法的优劣，为构建更科学的国际组织绩效评估体系提供借鉴。绩效评估体系的完善不仅有助于提高国际组织的运行效率，也将推动全球治理体系朝着更加公正合理有效的方向发展。

## 本期目录

1. Steffen Eckhard, et al., “The Performance of International Organizations: A New Measure and Dataset Based on Computational Text Analysis of Evaluation Reports,” *The Review of International Organizations*, forthcoming.
2. Steffen Eckhard, Vytautas Jankauskas, and Elena Leuschner, “Institutional Design and Biases in Evaluation Reports by International Organizations,” *Public Administration Review*, forthcoming.
3. Allison Carnegie and Richard Clark, “Reforming Global Governance: Power, Alliance, and Institutional Performance,” *World Politics*, Vol. 75, No. 3, 2023.
4. Dan Honig, Ranjit Lall, and Bradley C. Parks, “When Does Transparency Improve Institutional Performance? Evidence from 20,000 Projects in 183 Countries,” *American Journal of Political Science*, forthcoming.
5. Thomas Sommerer et al., “Decision-making in International Organizations: Institutional Design and Performance,” *The Review of International Organizations*, Vol. 17, No. 4, 2022.
6. Mirko Heinzl, “International Bureaucrats and Organizational Performance. Country-Specific Knowledge and Sectoral Knowledge in World Bank Projects,” *International Studies Quarterly*, Vol. 66, No. 2, 2022.

## 本期审校

陈兆源、杨嘉豪



**No. 1**

THE REVIEW OF  
INTERNATIONAL  
ORGANIZATIONS

Steffen Eckhard, et al

*The Review of International Organizations*

Forthcoming

**“The performance of international organizations: a new measure and dataset based on computational text analysis of evaluation reports”**

**《国际组织绩效：一个基于评估报告本文分析的新测量与新数据集》**

联合国体系下的国际组织每年会发布大约 750 篇评估报告以分析各个组织的项目、方案、机构和活动的表现。本文利用深度学习，提出了一种全新基于文本的绩效评价指标。在提出指标的基础上，作者建立了一个新的数据集，数据集包括系统下 9 个国际组织实施的 1082 项评估活动的绩效指标。

评价公共组织绩效最简单的方法就是考察其完成预定目标的能力。既有研究大多根据一系列关键指标来衡量组织绩效，但由于并非所有组织的目标实现都可以在指标层面得到反映。因此有学者已经开始对更广泛的数据来源进行分析，研究机构设计、人员配置和管理流程以及理事机构决议的相应指标。在开发性金融和发展援助领域，对绩效和有效性的评价指标已经有了大量进展，但在公共卫生，人道主义援助和社会政策等领域，仍然难以形成一致性的评价体系。

对国际组织的评估可以从组织外部展开，考察项目资助的审查机制、资金来源等问题，但考虑到既有国际组织的资助者大多是西方国家，外部资助本身存在“偏好性”，且资金相关报告往往难以获取，因此作者从联合国评估小组（UN Evaluation Group）的评价报告入手，对包括劳工组织、开发计划署、粮农组织在内的 9 个组织的数据进行了抓取和整理，建立了 IOEval 数据集，包括报告级别



的各个变量。

为实现对项目的绩效考核，作者主张对报告的本文进行分析，报告文本包含的语句越积极（消极），既评估整体越积极（消极）。在具体操作上，作者使用基于深度学习的 BERT 语言模型，将每个句子分别被标记为积极、负面和描述性评价，进行深度学习，取三个标签的最大概率值。随后作者对数据进行了验证。内容验证结果表明，语言模型以非常高的准确性对语句内容进行了分类；聚合分析基于外部绩效进行比较，作者模型结果和 IEG（世界银行独立评估小组）的评级数据相比较，将 2012-2021 年的世行报告编成数据集，输入语言模型并进行分布分析，统计结果显示二者可以相互匹配。

作者还对数据集进行了构建效用验证（Construct Validation）。构建效用主要是利用 IOEval 数据集中的评估条目进行理论假设的验证。作者在本部分对“项目目标的可实现性越强，项目绩效越高”进行了验证。项目良好的预期表现可以帮助项目获得更多的资金和支持，进而在下一阶段的目标中更容易成功，换言之每个更早阶段的成功，将帮助国际组织在下一阶段取得成功。通过固定效应分析，作者发现，国际组织项目的计划和结果存在差异，有更多可实现短期目标的项目，往往可以实现成功。

本文结合了三种验证策略检查了作者提出的国际组织活动绩效标准的有效性，开发了一种基于文本的绩效评估方法，同时开源的语言模型为数据集的扩展提供了可能性。新的数据集和模型增强了研究国际组织绩效的能力和路径，帮助将国际组织绩效从发展援助拓展到其他领域；新数据库还可以帮助研究国际组织绩效造成的影响与后果，同时帮助政策制定者评估各个组织之间的绩效差异。

编译：孟思宇（中国社会科学院大学）



**No. 2**

**Steffen Eckhard, Vytautas Jancauskas, and Elena Leuschner**

*Public Administration Review*

**Forthcoming**

**“Institutional Design and Biases in Evaluation Reports by International Organizations”**

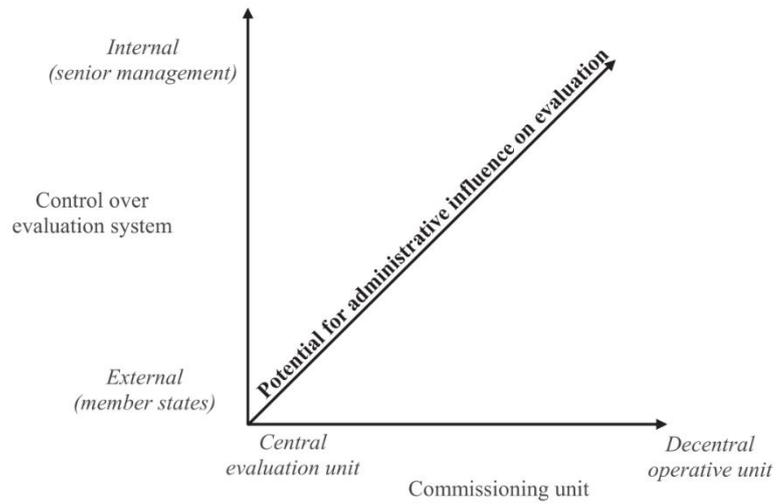
**《制度设计与国际组织评估报告中的偏见》**

近几十年，评估（evaluation）成为衡量国际组织业绩和影响的重要工具。各国政府有时花费数亿美元来评估组织绩效，但评估是否可能存在偏差？提交给决策者的冗长而又精心制作的评估报告，是否真的能为政治决策提供客观且有效的信息？本文对 2012 年至 2020 年间联合国系统中九个国际组织发表的 1082 份评估报告的分析显示，国际组织中评估体系的制度设计会导致评估结果存在差异，即存在偏见。总体而言，与中央评估单位（central evaluation units）相比，由业务单位（operative units）进行的评估更加正面。

评估有很多积极作用，比如将评估结果用于促进政策学习，确保问责制，追踪政策变化的路径等等。针对评估通常有两种观点，一是功能性观点（functional perspective），该观点认为评估是一种独立、无价值、基于研究的评价，有助于公共行政管理中的学习和循证政策；二是政治性观点（political perspective），该观点强调评估本身就是一项影响组织中资源和权力分配的政治活动。本文则聚焦现有文献忽视的评估过程，认为评估的进行依赖其发生时的组织环境，即组织设计



在多大程度上使被评估单位能够影响评估过程，决定了评估结果是否存在偏差。



**FIGURE 1** Two dimensions of evaluation systems' institutional design in IOs.

作者从两个维度来评估制度设计，一是政治主体（比如成员国理事会）对评估系统的控制程度；二是委托单位在国际组织的层次结构中的位置。据此本文提出三个假设：

H1：国际组织行政部门控制下产生正面评估的比例比成员国控制下产生正面评估的比例更高；

H2：由分散的业务单位委托的评估比由中央单位委托的评估产生正面评估的比例更高；

H3：由国际组织行政部门控制且由分散的业务单位委托的评估，产生正面评估的比例最高。

实证分析中，本文利用基于深度学习的语言模型，通过计算每份报告（主要是摘要）中积极评价的句子所占的比例来衡量评估报告的结果是否积极，并以其为因变量。为衡量评估系统制度设计的差异设置了两个自变量：一是成员国或国际组织的行政部门谁控制着评估系统的资源，比如任命评估单位负责人、决定评

估预算和评估议程；二是中央单位和业务单位谁委托评估。

本文的主要贡献在于，在既有研究常使用的访谈和调查数据之外，首次基于文本且跨多个国际组织对评估进行研究，充实了对评估及其潜在偏差的比较实证研究。同时，本文回应了功能性观点与政治性观点之间的的争论，强调评估者和被评估者之间的制度距离越小，评估结果会越正面，这解释了为什么有时评估在某个国际组织中具有功能性，而在另一个国际组织中则具有政治性。

编译：张尊月（中国社会科学院大学）



**No. 3**

**WORLD  
POLITICS**  
A Quarterly Journal of  
International Relations

Allison Carnegie and Richard Clark

*World Politics*

Vol. 75, No. 3, 2023.

## “Reforming Global Governance: Power, Alliance, and Institutional Performances”

### 《改革全球治理：权力、联盟与制度绩效》

长期以来，学术界一直试图研究全球治理变革的关键性因素。此类研究大多关注于系统性的政治和经济冲击带来的政治经济后果。在这些系统性冲击之外，国际组织同样会进行一些较为细微的改革以补偿部分成员国。尽管改革的措施微不足道，但其带来的政策后果可能是难以估计的，甚至导致根本性的结构与权力转变。本文研究了国际组织内部的主要结构变化，同时研究了这些组织在日常运作中发生的微小变动。作者试图解释：什么因素导致了国际组织成员在日常工作中进行改革？哪些成员在改革过程中收益最大？

作者认为，国际组织的表现以及成员和组织主要利益相关者的政治关系是改革发生的主要障碍。在日常运行期间，大国的支持将有助于其盟国推动国际组织改革，改革成功将有助于完成对盟国的奖惩。此外，国家也可以激励国际组织官僚以进行改革。原因有二，一是国际组织官僚希望实现自治以保留自主权，二是官僚往往是由大国推选出来的，其本身就具有意识形态偏见。因此可以预计国际组织官员倾向于支持大国盟友的改革目标，以寻求大国对机构的支持。此外，当大国对国际组织的表现不满时，组织官僚会试图安抚这些国家以防止其脱离国际组织，通过采取较小的让步说服其维持资助或是留在组织内。基于以上观点，作



者提出了理论假设：

H1:在组织常规运行期间，国际组织的改革会使最强大成员国的盟友获益；

H2:当最大成员国的非盟国因国际组织表现不佳而威胁脱离时，国际组织会进行改革以进行补偿。

此后作者通过检验世界银行的案例，对理论假设进行了检验。世界银行内的改革往往以微小和渐进为主要特点，同时世界银行是具有代表性的国际组织，可以有效对理论进行验证。作者主要考察的是国际复兴开发银行（IBRD）的变革过程，世行的主要工作是向中等收入国家提供发展援助资金，其依赖于每三年的成员国注资。美国是世行最大出资国，同时中国等借款国同样对其构成竞争。美国利用其主导权在世行内推动其特殊议程，例如借款国检测和评估标准等，同样也可以像其他成员国一样威胁拖欠贷款以造成银行违约。基于理论假说，可以预计美国的盟国会世行的渐进性改革中受益最大，同时当世行的表现不佳时，非美国盟友国将威胁脱离银行并从随后的改革中获益。

为系统性验证理论假设，作者根据 1944 年至 2018 年的新闻稿和复兴开发银行理事会决议内容进行了手工编码。1944 年以来世行进行了 108 次改革，包括贷款政策、扩张、新员工、规则、投票份额和其他六类内容。除份额改革外，世行章程也是改革的另一个主要内容，章程的变动往往会带来世行内权力的重新分配。为了明确受益者，作者将数据限制在分配性改革的过程中，总数据包括 83 项改革和 121 个受益国。在利用 IEG 数据、ATOP 数据集、世行改革数据集和其他控制变量进行测试后，作者的理论假设得到了验证。

本文证明了国际组织员工和管理层常常通过推动改革以使组织中最大参与国的盟国收益，但这一过程中非盟友国往往也同样可以获得收益。作者进一步证



明,当组织绩效下降时,国际组织官员将试图利用改革以确保非盟国留在组织内。除此之外,作者还为研究国际组织改革提供了一个新的原始数据集,并且其理论框架有潜力推广至其他国际组织。

编译: 孟思宇 (中国社会科学院大学)



**No. 4**

**Dan Honig, Ranjit Lall, and Bradley C. Parks**

*American Journal of Political Science*

**Forthcoming**

**“When Does Transparency Improve Institutional Performance?  
Evidence from 20,000 Projects in 183 Countries”**

**《透明度何时能提升制度绩效？——来自 183 个国家 20000 个项目的证据》**

近年来，信息公开（Access to information, ATI）政策已经成为政府和国际组织制度透明度最突出的表现形式。这一政策保障了向公共机构索取信息的权利，为利益攸关方监测官僚活动和获取公共部门信息创造了新的机会，进而改变了他们与这些机构的关系。本文在国际发展援助的背景下，分析了采用 ATI 政策的双边和多边援助机构会如何影响它们在低收入和中等收入国家援助项目的结果。该数据集涵盖了 1956 年至 2016 年期间 183 个国家的 12 个援助机构资助的 2 万多个项目的绩效，这是迄今为止编制的最大的外援项目成果数据集。

ATI 政策建立了一个正式的程序，任何国家的公共或私人行为体都可以通过该程序请求获得援助机构掌握的信息，包括它们资助的外援项目信息。对于双边援助机构，ATI 政策采取政府信息公开（governmentwide freedom of information, FOI）的法律形式，这些法律大多以 1966 年的《美国信息自由法》（U.S. Freedom of Information Act）为基础设置；对于多边援助机构，ATI 政策采取捐助国理事会批准的约束性规则，这些规则也通常以《信息自由法》为蓝本，体现了公众有权了解政府间机构活动的原则。独立申诉程序（Independent Appeals Processes）是执行 ATI 政策的重要环节，使信息寻求者能够在请求被拒



绝时向一个独立的机构提出上诉。为了探究 ATI 政策对援助项目绩效的影响，本文提出以下假设：

H1：ATI 政策本身并不会改善援助项目的执行情况；

H2：采取包含独立申诉机制在内的 ATI 政策有助于提升项目的绩效；

H3：采取包含独立申诉机制在内的 ATI 政策将与某些受援国的项目绩效成更强的正相关，这些受援国自下而上问责制水平较高，治理质量水平较低，国内 ATI 制度薄弱或没有。

本文使用双重差分方法（Difference-in-Difference），以援助机构-受援国-年份为分析单位，为了验证假设 1，构建如下方程：

$$Project\ Success_{r,d,t} = \alpha + \beta_1 ATI_{d,t-1} + \beta_2 Controls_{r,d,t-1} + \delta_d + \gamma_r + \psi_t + \varepsilon_{r,d,t} \quad (1)$$

其中， $Project\ Success_{r,d,t}$  为被解释变量，衡量项目成功度； $ATI_{d,t-1}$  为处理变量，表示是否存在 ATI 政策； $Controls_{r,d,t-1}$  为控制变量，包括受援国 GDP 增长率、对数化人均 GDP 和官方援助净额占国民收入比重。纳入捐助机构、受援国和年份的固定效应。所有协变量滞后一期。

为了验证假设 2，构建如下方程：

$$Project\ Success_{r,d,t} = \alpha + \beta_1 IAP_{d,t-1} + \beta_2 Controls_{r,d,t-1} + \delta_d + \gamma_r + \psi_t + \varepsilon_{r,d,t} \quad (2)$$

$$Project\ Success_{r,d,t} = \alpha + \beta_1 ATI_{d,t-1} + \beta_2 IAP_{d,t-1} + \beta_3 Controls_{r,d,t-1} + \delta_d + \gamma_r + \psi_t + \varepsilon_{r,d,t} \quad (3)$$

为了验证假设 3，在方程（1）、（2）中分别引入以下交互项：公民社会组织的普遍参与度；媒体自由综合指数；法治和腐败指数；是否存在国内信息自由法；是否存在具有申诉机制的国内信息自由法。

研究表明：

1) 单独采用 ATI 政策并不会改善援助项目的绩效，但是，当采用包含独立申诉机制在内的 ATI 政策时，援助项目的绩效会得到改善。独立申诉机制允许寻求信息披露的行为体通过独立管理的流程对被拒绝的请求提出上诉，这一机制实际上可以防止官僚们逃避信息披露。

2) 当受援国拥有更高水平的公民参与度和新闻自由指数时，具有独立申诉机制的 ATI 政策与援助项目成功度的关系更强。自下而上的问责制形式使公民更容易利用 ATI 政策和申诉程序，向援助机构和受援国政府施压，要求其对表现不佳的证据做出回应。

3) 当获取项目信息和行使政治影响力的现有渠道较少时，该相关性也会得到加强。研究表明，强有力的援助机构 ATI 政策是国内 ATI 制度的替代品，而不是补充品。ATI 政策所能达到的效果存在一个上限，在一个运作良好的问责制度中加入此类政策，对绩效结果产生的影响微乎其微。

4) 收集、评估和处理利益相关者投诉的程序可以成为阻止违规行为的有力工具，这种工具可以利用自下而上的监督和自上而下的执行。

编译：杨嘉豪（中国社会科学院大学）



**No. 5**

THE REVIEW OF  
INTERNATIONAL  
ORGANIZATIONS

Thomas Sommerer et al.

*The Review of International Organizations*

Vol. 17, No. 4, 2022.

## “Decision-making in International Organizations: Institutional Design and Performance”

### 《国际组织中的决策：制度设计与绩效》

不同的国际组织在决策表现、政策产出随时间的波动方面存在明显差异，一些国际组织运行的颇有成效，另一些则深陷困局、举步维艰。为什么会产生这种差异？本文的研究表明，国际组织的制度设计在其中发挥着关键作用。

国际组织的绩效通常包含两个方面，一是过程，主要关注组织能力；二是结果，主要关注目标实现。决策绩效作为国际组织发挥作用的关键部分，关注的是过程和结果之间，国际组织采纳决策的能力。决策绩效不仅与制度效率密切相关，即制度本身在减少或解决社会问题方面是否成功，也与能否解决制度复合体等制度自身的问题紧密相关。

作者设置了三个制度设计的特征，包括汇集（pooling）——通过多数投票汇集权力等方式取代国家的决策否决权；授权（delegation）——将权力授予独立的超国家机构；跨国行为体的准入（access for transnational actors, TNAs）——给予跨国行为体特权。基于三个制度设计特征，文章提出了三个假设：

H1：增强汇集将对决策绩效产生积极影响；

H2：增加授权将对决策绩效产生积极影响；



H3: 增加跨国行为体准入将对决策绩效产生积极影响。

制度设计特征之间的结合与互动,比如通过相互加强优势收益来实现集体决策从而降低交易成本等,最终影响了国际组织的决策绩效。对此,作者提出了有关特征之间互动的三个假设:

H4: 增加跨国行为体准入将在更高层次的汇集下对决策绩效产生更强的积极影响;

H5: 在较低的授权水平下,增强汇集将对决策绩效产生更强的积极影响;

H6: 在更高的跨国行为体准入的水平下,增加授权将对决策绩效产生更强的积极影响。

实证方面,作者引入了一种基于国际组织内部决策产出年增长率的新衡量方法,更加关注增长曲线而不是绝对数字,试图系统地分析决策绩效随时间的变化。文章统计模型控制了特定国际组织的固定效应,对 30 个国际组织自 1980 年至 2011 年内的上万个决策进行分析。结果表明,国际组织之间及国际组织内部的决策表现都存在差异。当汇集特征更明显时,跨国行为体准入对决策绩效有积极影响,授权的影响随跨国行为体准入程度的不同而有所不同,汇集对决策绩效有独立的积极影响。

文章的主要贡献在于,除了将决策绩效作为整个国际组织表现的一部分,并在理论方面对制度设计对决策绩效的影响机制与路径进行明晰之外,还首次对国际组织的决策绩效进行了比较评估,补充了国际组织决策绩效一般模式和解释的研究。

编译:张尊月(中国社会科学院大学)



**No. 6**

INTERNATIONAL STUDIES  
 QUARTERLY  
 academic.oup.com/isq

**Mirko Heinzl***International Studies Quarterly***Vol. 66, No. 2, 2022.**

**“International Bureaucrats and Organizational Performance.  
 Country-Specific Knowledge and Sectoral Knowledge in World Bank  
 Projects”**

**《国际官僚与组织绩效——世界银行项目中的国别知识和部门知识》**

国际组织需要将国别和部门知识纳入决策，以便在成员国成功实施发展项目。整合不同类型知识的核心手段之一是通过员工的专业和经验。当下的研究缺乏对不同类型的知识如何塑造工作人员为国际组织的绩效做出积极贡献的能力的评估。本文通过选取世界银行员工的案例，运用一个关于世界银行工作组组长的任期、国籍和教育背景的数据库进行检验，探究其国别知识与部门知识的储备对于提升世行绩效的作用。本文认为，国别知识、部门知识以及它们之间的结合对于世界银行的项目具有积极的促进作用。

首先，作者提出了本文的论点：员工的知识差异影响了世界银行的项目绩效。本文经过充分的文献回顾，指出国别和部门知识能对国际组织的绩效产生贡献。作者因而提出了本文的三个假设：

- H1: 员工拥有的国家特定知识越多，绩效提升越多；
- H2: 员工拥有的部门知识越多，绩效提升越多；
- H3: 员工拥有的国家和部门特定知识越多，绩效提升就越多。



其次，作者介绍了本文的研究设计。本文利用了一个数据库，该数据库包含了世界银行对 3800 多个项目的项目评估以及 10000 多个项目中人员安排的新数据。作者通过使用将世行工作人员的新数据与世行独立评估小组的绩效评估相结合的实证方法，对世行相应的项目绩效进行了实证分析。其中，因变量是基于世界银行独立评估小组（IEG）的项目评估。自变量是基于世界银行任务组负责人（TTL）的相关数据，包括负责人负责项目的所属地区占比以及负责项目所属部门的占比。为了尽量减少项目或国家之间的差异导致有问题的推论的可能性，作者还选取了负责人过去监督过的项目数量，并对其取对数计数作为控制变量，以确保成果不仅仅由更全面的经验驱动，而是由国别或部门背景的专业知识驱动。

接着，作者对国别和部门知识在世界银行绩效中的作用进行了实证检验。文章先是对负责人数据进行了描述性统计，结果发现，在大多数地区，约有 20-30% 的项目是由具有国别经验的负责人监督。并且在世行中，部门知识似乎比国别知识更为普遍。接着，文章采用简单最小二乘（OLS）模型对变量进行回归检验，结果表明：国别知识对于世界银行项目绩效具有较强的的重要性，然而所有衡量部门知识的变量都未能达到惯例水平的统计显著性。作者对世界银行相关工作人员的直接采访也证实了上述判断。因此，该分析证明了特定国别知识（H1）的重要性，但没有证据证明部门知识（H2）对促进世界银行绩效的直接作用。部门知识似乎间接重要，因为国别经验似乎依赖部门经验来产生实质性的积极影响（H3）。为最大限度地减少这些结果被数据或模型的特殊性所影响的可能性，作者又进行了一系列的稳健性检查。

最后，作者讨论了研究结果对国际官僚机构和世界银行项目研究的影响。本文的结论是，通过将国别知识与部门知识相结合，员工可以对国际组织的绩效产生积极影响。其中，员工的国别经验对于提升世界银行的项目绩效具有重要影响，



而员工的部门知识则有可能为项目绩效带来额外增益——尽管证据并不充分。但上述结论仍需接受进一步检验。本文有三项引人注目的发现：

第一，总体而言，国别知识似乎很重要，而部门知识则不重要；

第二，有证据表明，兼具这两种知识的员工有能力对绩效做出更积极的贡献；

第三，经验的多样性和相关性，而不是任期的长短，与更多的成功相关。

本文的贡献在于：研究结果表明了员工的个人知识差异是如何影响其决策和绩效的，从而为有关国际官僚机构的讨论做出了贡献。作者指出，国际组织可以更好地利用其官僚机构中的现有资源，通过减少工作人员在不同工作地点之间的轮换频率来提高其绩效。

编辑：苏山岳（中国社会科学院大学）

## 全球治理研究团队

任 琳 熊爱宗 韩 冰 吴国鼎  
陈兆源 黄宇韬 李 冲 韩永辉  
宋 锦 田 旭 沈 陈 彭 博

## 研究助理团队

兰馨彤 苏山岳 孟思宇 杨嘉豪

声明：对观点的摘录和引用不代表编者本人及其所属单位对观点的认同。

