

RESEARCH ARTICLE

“What Gets Measured Gets Done”¹: Metric Fixation and China’s Experiment in Quantified Judging

Kwai Hang Ng¹ and Peter C. H. Chan^{2,*}

¹Department of Sociology, University of California, San Diego, La Jolla, CA, USA and ²School of Law, City University of Hong Kong, Kowloon, Hong Kong

*Corresponding author. E-mail: pchchan@cityu.edu.hk

Abstract

This article analyzes the ambitious Case Quality Assessment System (CQAS) that the Supreme People’s Court of China (SPC) promoted during the first half of the 2010s. It offers a case-study of Court J, a grassroots court located in an affluent urban metropolis of China that struggled to come out ahead in the CQAS competition. The article discusses how the SPC quantified judging and the problems created by the metricization process. The CQAS project is analyzed as a case of metric fixation. By identifying the problems that doomed the CQAS, the article points out the challenges facing the authoritarian regime in subjecting good judging to quantitative output standards. The CQAS is a metric that judges judging. It reveals how judging is viewed by the party-state. The article concludes by discussing the legacy of the CQAS. Though it nominally ended in 2014, key indicators that it introduced for supervising judges are still used by the Chinese courts today. The CQAS presaged the growing centralization that the Chinese judicial system is undergoing today. Though the SPC has terminated the tournament-style competition that defined the CQAS, the metric remains the template used to evaluate judging.

Keywords: judging; judgment; metric; professionalism; China

1. Introduction

Can judging be measured and quantified? The answer of the Supreme People’s Court (SPC), the highest court of China, is a resounding yes. In the early years of the 2010s, a comprehensive assessment programme of judicial performance was carried out in China. The project, known as the Case Quality Assessment System (*anjian zhiliang pinggutixi* 案件质量评估体系, CQAS), was unprecedented in breadth and depth, not just within China, but also compared to similar assessment tools in any jurisdiction in the world. The CQAS was an elaborate scheme of measurements that gauged the quality of judging. Its scope was at once sweeping and penetrating. It covered all types of cases: administrative, civil, and criminal cases, and any other types that appeared on the record books of courts. The system applied to all geographic regions, despite the vast differences in the degree of economic development between the coastal East and the inland West. It was unashamedly quantitative. The quality of judging was boiled down to a set of numbers. The CQAS represented the most large-scale project to date to measure and evaluate the performances of judges on the sheer basis of numbers alone.

¹ This is a famous phrase by business management writer Peters (1986).

The reasons that the SPC implemented the scheme were not obscure. The highest court of China believed the CQAS would strengthen the court system by objectively and reliably measuring the quality of judging.² Despite the vast power that the authoritarian party-state commands, it remains a challenge for it to evaluate its 120,000 judges in a manner that is both valid (evaluating what matters) and accurate (unbiased).³ The common criteria used in Western liberal democracies, such as professional reputation, peer evaluation, and academic scrutiny, are either unavailable or relatively undeveloped in China. With an embedded judicial system that is asked to contribute to the political goal of stability maintenance, the party-state does not evaluate the work of courts purely on the basis of law.⁴ Chinese leadership relies on lower-level officials to carry out the minutiae of day-to-day governance.⁵ The presence of this political duty compels front-line judges not only to rule (in the legal sense), but also to *rule* (in the political sense). Substantial discretion is given to them to operate flexibly and to decide when to use the law and when to resort to extra-legal measures. What happens on the ground is notably different from what is reported in the books. The SPC is thus forced to live with the “principal-agent problems” that accompany the discretion it must give to front-line judges for effective governance.⁶

Ever since the 1990s, the SPC has worked to promote a more professionalized judicial system. Effectiveness of results and efficiency in operation have become the twin pillars that replaced ideological purity as a measuring rod of the effectiveness of the work of the courts. In its many guiding directives and opinions, the SPC asks lower courts to follow its lead to use law responsibly to rule and govern. But the SPC continually faces the “principal-agent” problem. In many cases, grassroots courts pledge their loyalty to local governments. Proximity aside, until the recent judicial reform, local governments had been the primary source of funding for local courts.⁷ It is therefore problematic whether Chinese courts follow the directives of the SPC to carry out the law in ways in which the SPC wants it to be carried out.⁸ The central leaders are constantly looking for a reliable means to ascertain that local judges are complying.

The CQAS was the most elaborate and wholesale effort of the SPC to tackle this “principal-agent” problem. In 2005, the SPC published its “Second Five Year Reform Program for the People’s Courts (2004–2008).” One component of the reform was to establish a nationwide performance evaluation plan. Article 41 stated that it was the goal of the SPC to establish “scientific and unified systems for evaluating trial quality and effectiveness.”⁹ Beginning in the late 2000s, the SPC began to test various indices to improve the performances of grassroots judges. The system in its fullest form, the CQAS, was formally introduced in 2011. It was by far the most comprehensive measurement system of judicial performances in China, if not in the world. It lasted about four years. By the end of 2014, the SPC had aborted the project. There was no official eulogy; it just stopped. What happened? What led to the precipitous fall of the CQAS? What is most puzzling is that the CQAS did perfectly what it was supposed to do. From 2007 to 2014, the SPC steadily recorded a national year-to-year improvement in the overall index score.¹⁰ Why did the SPC decide to abandon a powerful tool that had seemingly addressed its perennial worry about front-line judges’ not following its lead?

² Supreme People’s Court (2011).

³ Before the latest judicial reform and the introduction of the quota system, there were even more judges in China. The number was about 212,000.

⁴ Ng & He (2017).

⁵ Minzner (2006).

⁶ O’Brien (2003).

⁷ Ng & He, *supra* note 4.

⁸ Ng (2019).

⁹ Supreme People’s Court (2005).

¹⁰ Yan & Yuan (2013); Yan & Yuan (2015).

This article aims to achieve two goals. First, it offers a post-mortem. We identify the problems that eventually led the SPC to terminate the CQAS. Through identifying the pernicious outcomes that the assessment system produced, we show the limits of quantification that the SPC eventually became painfully aware of. Second, we explore how judging is perceived, conducted, and evaluated in China. Through this study of the CQAS, we try to answer the question: “How is judging viewed by the Chinese state?” We further argue that the output model that informed the CQAS conflicted with the conceptual underpinnings of adjudication. Even for a regime that abhors unequivocal commitment to any valued-based notion of the rule of law, it became obvious to the SPC that the activity of judging could not be quantified as an aggregation of numbers. In the last part of the article, we address the legacy of the CQAS.

2. Background

The CQAS was a “trial management” system. Article 2 of the SPC’s “Guiding Opinion” on the subject states that the purpose of the system was to improve the quality of judges, and to promote their sense of responsibility.¹¹ The Chinese concept of “trial management” (also known as “judicature management”) (*shenpan guanli* 审判管理) is categorically different from the notion of “case management” that has become quite popular in the Anglo-American context. Trial management goes beyond simply improving judicial efficiency. It includes the administrative functions of managing the court institution, the individual judge’s conduct (including disciplinary matters), and the implementation of policy objectives in civil adjudication (e.g. the policy objective of preferring mediation to adjudication).¹²

The CQAS was made up of 31 indices. These individual indices were grouped under three secondary indices with the headings of “fairness” (*gongzheng* 公正), “efficiency” (*xiaoli* 效率), and “impact” (*xiaoguo* 效果). These labels were anything but neutral descriptors. They combined disparate measurements into seemingly coherent frameworks. They were a form of public transcript¹³—tailored descriptions that reflect the interests and priorities of the power holder (a point on which we will elaborate later). At the very least, the labels suggested what the SPC leaders wanted the public to believe the indices were measuring. Hence, indices such as the rates of judgments reversed and remanded for retrial and the rates of cases that were awarded judicial compensation were grouped under “fairness.” Under “efficiency,” the SPC includes indices such as average cases cleared per judge and the rate of cases resolved within statutory time limits, among others. And, under “impact,” there was an even more eclectic mix of indices that included the rates of mediation, case withdrawal, and cases in which neither retrial nor appeal is sought. Most importantly, the CQAS adopted a fully quantitative approach to assessing judicial quality (Articles 3, 21, 25, and 26 of the “Guiding Opinion”). The SPC states that assessment indices should be the exclusive criteria used to evaluate judicial performance (Articles 6–10).

The three secondary indices combined to produce a master index known as the “Integrated Case Quality Index” score. The 31 component indices were weighted differently. Two traditional indices that pre-dated the CQAS—the rate of first-instance cases in which, upon appeal, a judgment was corrected and the rate of cases in which, upon petition and supervision, a judgment that had already been entered was corrected—carried the heaviest weights of 19% and 21%, respectively, in the version of the indices

¹¹ Supreme People’s Court, *supra* note 2.

¹² Chan (2016).

¹³ Scott (1990).

published by the SPC.¹⁴ Local provincial and intermediate courts could, taking into account the unique work environment of their courts, adjust the weight given to their specific assessment indices (Articles 13 and 14). In practice, courts followed the percentage distribution suggested by the SPC, with only minor tweaks to individual indices such as the rate of mediation.

The SPC did not want lower courts to be one-trick ponies. The CQAS was designed to push courts to perform on all fronts (see Table 1). A court that disposed of many cases but received a lot of complaints and petitions was by the CQAS standard not a good court. Courts must be all-round in order to excel. To use a sports analogy, getting evaluated by the CQAS was like competing in the decathlon. As in the decathlon, it was impossible for a competitor not to have any weak spots but they could not be so serious as to depress the overall score. An athlete who is a soaring high jumper may not be an explosive shot-putter and a lightning-fast 100-metre sprinter may not be the most strenuous middle-distance runner. Each event in the decathlon requires different muscle strength, body type, and even mind frame for an athlete to excel, yet all events in the end must be aggregated into a single overall score. As stated, the CQAS scheme measured three aspects of performance that, in theory, were distinct. In fact, the three meta index types were of discordant qualities. For example, the efficiency index called for quick closing of as many cases as possible, yet the fairness index demanded that cases should not be “misjudged,” as measured by the reversal/retrial rates.

In 2008, the SPC began a pilot project to explore the feasibility of adopting a national system of judicial-performance assessment.¹⁵ The system was formally put in place in the spring of 2011. In the “Guiding Opinions” issued that year, the SPC outlined a framework to measure judicial quality and efficacy.¹⁶ There were four themes: (1) unifying standards in the CQAS with the view to promote efficiency and fairness in litigation; (2) constructing a sophisticated, scientific, and systemic evaluative framework encompassing adjudication and non-adjudicatory matters (e.g. enforcement work and management of judicial personnel); (3) institutionalizing the CQAS with clear delineation of the responsibilities of different divisions within the court, as well as clarifying the supervisory and evaluative powers of higher-level courts over lower courts in relation to case-quality evaluation; and (4) improving data collection, analysis, and management, and establishing effective reporting channels on matters concerning case-quality evaluation.¹⁷

Both front-line judges and senior judges with administrative roles were evaluated by the system. In the case of the former, the CQAS computed a personal score for each of

¹⁴ The two indices look similar, but they refer to the decisions challenged by two different mechanisms—appeal and petition for a retrial. Appeals against grassroots court judgments are heard by intermediate people’s courts. In fact, when a party appeals in China, the higher court is obligated to take on the appeal. For most cases, though, there is a very short time window for initiating an appeal, usually within 15 days from the date of service of the first-instance judgment. Parties can also challenge a decision by petitioning at the trial-court level for a retrial (*zaishen* 再审, sometimes translated as reopening a trial) rather than initiating an appeal. The retrial review usually happens after the trial court has entered its decision. Art. 199 of the Civil Procedure Law (2017) states: “Any party that considers a legally effective judgment or ruling to be wrong may apply to the immediate superior people’s court for retrial; as for the case where one party comprises of a large number of individuals or both parties thereto are citizens, the parties may apply for retrial of the case to the original people’s court. Nevertheless, the application for retrial does not mean that the enforcement of the judgment or ruling is suspended.” There is a similar provision in Art. 252 of the Criminal Procedure Law (2018) governing criminal retrials. Courts and people’s procuratorates may sometimes initiate retrial reviews on their own. But most retrials are initiated by parties who petition to the courts. The application is, however, subject to the approval of the court. A retrial may take place at the original court or the next higher court. Between the two mechanisms, appealing an adverse decision to a higher court is by far the more common choice.

¹⁵ Kinkel & Hurst (2015); Supreme People’s Court (2008).

¹⁶ Supreme People’s Court, *supra* note 2.

¹⁷ Chan, *supra* note 12.

Table 1. Indices of the Case Quality Assessment System, 2008–14

Index type	Name of index	Weight	Effect of higher scores on index
Fairness	Rate of cases in which key information is changed after filing	8%	Negative
	Rate of first-instance trials in which a lay assessor is present	7%	Positive
	Rate of first-instance cases in which, upon appeal, a judgment is corrected (changed or remanded for retrial)	19%	Negative
	Rate of second-instance cases in which the original judgment is corrected (judgment changed or remanded for a new trial)	5%	Positive
	Rate of second-instance cases in which a hearing is held	5%	Positive
	Rate of reopened trials (<i>tiqi zaishen</i>) among lower courts' judgments that were reviewed (for the court that conducts reviews)	6%	Positive
	Rate of cases in which, upon petition and supervision, a judgment that has already been entered is corrected (changed or remanded for a retrial)	21%	Negative
	Rate of retried (<i>zaishen</i>) cases in which the court corrects a lower court's judgment that has already been entered (judgment changed or remanded for a new trial)	5%	Positive
	Rate of hearing and inquiry held during retrial	4%	Positive
	Rate of cases awarded judicial compensation	10%	Negative
Efficiency	Quality of judgment	10%	Positive
	Rate of cases placed on file within statutory time limit	9%	Positive
	Rate of first-instance cases in which the simplified procedure is used	10%	Positive
	Rate of cases resolved within statutory (normal) time limits	15%	Positive
	Rate of the average time taken to resolve a case against the statutory time limit	9%	Positive
	Rate of the average time taken to enforce a judgment against the statutory time limit for enforcement	9%	Positive
	Rate of cases with their time limit extended against cases resolved within statutory time limit	9%	Negative
	Rate of cases in which judgment is given at the conclusion of the hearing	5%	Positive
	Index of resolving cases evenly throughout the year	12%	Positive
	Number of cases resolved per capita (court establishment) per year	11%	Positive
Impact	Number of cases resolved per judge per year	11%	Positive
	Rate of cases in which complaints are settled by the first judgment without further appeal or retrial	9%	Positive
	Rate of cases enforced	15%	Positive
	Rate of mediation	10%	Positive
	Rate of case withdrawal	6%	Positive
	Rate of cases in which enforcement objectives are realized	12%	Positive
Rate of cases in which judgment is voluntarily executed	11%	Positive	

(Continued)

Table 1. (Continued)

Index type	Name of index	Weight	Effect of higher scores on index
	Rate of mediated cases in which there is an application for an enforcement order	7%	Negative
	Rate of retrial reviews among judgments/mediation agreements that have already been entered	10%	Negative
	Rate of citizen petitioning against court	10%	Negative
	Index of level of public satisfaction	12%	Positive

them. Individual scores would then add up to the score of the division (*ting* 庭) to which judges belonged. Divisional scores would in turn be tallied to determine the collective score of a court. The full significance of the CQAS can only be appreciated in the context of annual performance evaluations (*nianzhong kaohe* 年终考核). As pointed out by Kinkel and Hurst (2015), ordinary judges could lose points on their records if they scored poorly on certain indices—for example, if cases were remanded for retrial by the higher court. By the same token, the promotional prospects of presidents and vice presidents would also be negatively affected if their courts ranked as low in the CQAS.

3. Data

Our primary data come from a grassroots court that we name Court J, which is situated in one of the most affluent metropolises of coastal China, with a gross domestic product (GDP) in 2011 (the year in which the CQAS was implemented) of about 93,000 yuan per capita. The court is among the busiest in this economically developed region. We closely studied the internal newsletters that circulated among the judges during the period and prepared by a unit of the court known as the trial management office (*shenpan guanli bangongshi* 审判管理办公室). The office is the research and publication arm of the powerful adjudication committee and is accountable to both the court president and the adjudication committee.¹⁸ It is given the mandate to manage all aspects of adjudication. It is also understood that the newsletters were prepared in consultation with the court-leadership team—that is, the president and the vice presidents of the court.¹⁹ As an update, we recently interviewed about ten judges from different parts of the country to see what has happened since the cancellation of the CQAS. We asked how their grassroots courts reacted to the CQAS during the peak years of its implementation. We also draw from relevant Chinese and English secondary literature, as well as official statistics.

The fact that the data are drawn mainly from a single court limits the representativeness of the behaviour we describe here. However, our study identifies broader structural tensions that affected grassroots courts whatever their location. Individual courts and judges throughout the country were pressured not to fall behind. While the intensity of pressure might have varied across courts, we believe the findings of this case-study have generalizability. Court J is a highly relevant case for the study of the metricization of judicial performance for the following reasons. It is a court that, despite its busy activities, struggled (and continues to struggle) to get on par in several core indicators. In 2011, it ranked last among courts in the same city in arguably the most important individual

¹⁸ *Ibid.*

¹⁹ The trial management office is responsible for managing court processes, evaluating case quality, assessing overall effectiveness of the court's adjudication work, and assisting in the evaluation of the work of individual judges. See *ibid.*, p. 19.

indicator of all—the rate of cases corrected. From analyzing its responses, we see how a grassroots court mobilized its resources and devised new strategies to cope. Its poor rankings were a constant theme in the newsletters. In this sense, Court J serves as an “extreme case” and it is the extremity of the situation that makes it an instructive case to study.²⁰ The senior judges of the court made it a top priority to improve the CQAS score. While most basic and intermediate courts produced monthly internal newsletters assessing the performance of judges and divisions, Court J circulated *weekly* newsletters in the year that we studied, and sometimes even published several issues in a single week.²¹ Senior judges openly and frankly discussed many of the problems they identified about the CQAS system in these internal documents. They also came up with a multitude of new requirements for the judges to boost their scores. Through personal connections with some of the judges who worked in the courts, we collected the newsletters of 2012, the year in which the CQAS was advancing in full force. The newsletters provided an interesting paper trail by which we were able to trace the pressure facing the grassroots courts of China and identify the coping strategies that they devised in response.

4. Theoretical perspective

The CQAS was at root a metric.²² It was a data-generating, indicator-based method of evaluation designed to quantify all facets of judicial performance. With the CQAS, any two courts in China could be in theory be compared and ranked. A big urban court in Shanghai could, for example, be directly compared to its rural counterpart in a modestly developed region of inland Shaanxi. This could be done despite their obvious differences in size, resources, and the quality of judicial personnel. In an important sense, the CQAS was a form of standardization. It simplified the messy reality that the courts of China deal with.²³ As a metric, it also created and quantified difference, with far-reaching consequences for deciding the status of a court among its peers.²⁴

We approach the rise and fall of the CQAS as a case of metric fixation. In his recent book, Jerry Muller identifies three key components of metric fixation that we use to structure our discussion of the CQAS:

1. the belief that it is possible and desirable to replace judgment, acquired by personal experience and talent, with numerical indicators of comparative performance based upon standardized data;
2. the belief that making such metrics public (transparent) assures that institutions are actually carrying out their purposes (accountability);
3. the belief that the best way to motivate people within these organizations is by attaching rewards and penalties to their measured performance, rewards that are either monetary (pay-for-performance) or reputational (rankings).²⁵

4.1 CQAS Is “impersonal” and therefore “objective”

As mentioned, the principal-agent problem has vexed the operation of Chinese state bureaucracies for years. The central government is well aware of the problem, yet each solution has created its own problems in turn. Full-scale centralization limits flexibility; and lack of flexibility in turn exacerbates the disconnect between policy and reality. The

²⁰ Seawright & Gerring (2008).

²¹ At least during the period of our study, i.e. 2011–12.

²² Mau (2019); Muller (2019).

²³ Star & Lampland (2009).

²⁴ Mau, *supra* note 22.

²⁵ Muller, *supra* note 22, p. 18.

state's policy is to tread a fine middle line. The central government demands its local bureaucracies to be agile and flexible yet at the same time faithful to the intent of the central government when carrying out policies.²⁶ The court system is no exception. The SPC is suspicious about the extent to which national law is carried out, or at least carried out in the way it expects. The courts' presidents and vice presidents, who represent the leadership strata, are often handpicked by the local people's congress and are naturally subject to the influence and control of local governments. The SPC also worries that the supervisory function that these senior judges are supposed to perform is undermined by the long-standing patronage system.²⁷ It is no secret that the SPC is looking for more transparent, objective, and comparable tools to evaluate performance.²⁸ The CQAS was hailed as an answer to this dilemma. It was a means to gather accurate and unfiltered information on the performance of the 3,000-odd courts in different parts of the country. Elaborate and data-based, the CQAS promised a more objective way to assess performance.

Numbers generated by a system like the CQAS could serve as a check and, more important, as a substitute for personal acquaintance and trust. Hence, the CQAS had multiple targets. It was not a system designed only to monitor and thereby improve the performances of individual judges. It also monitored the presidents and vice presidents of local courts who in turn were supposed to monitor their front-line judges. Quantitative data are particularly powerful for large-scale, remote control that is distinct from direct intervention yet potentially even more effective in checking supervision by local officials.²⁹

4.2 Did the CQAS promote accountability?

There was a strong ideological affinity between the instrumentalist view of law that the Chinese authoritarian state holds and the output-focused framework of the CQAS. Strong instrumentalism tends to debase professional values. Members of a profession jointly delimit the problems that fall in their demarcated domain and develop their own accepted ways of thinking about them.³⁰ In many countries with long-established legal systems, law is a classic *profession* in which specific standards of competency, knowledge, training, and ways of working are shared among its practitioners. In China, judges and lawyers have not yet achieved enough operational autonomy to develop a distinct sense of professionalism. The appeal of a metric such as the CQAS is that it provides a calibrated benchmark to hold judges accountable without subscribing to any fully developed notion of professionalism. The CQAS did not measure individual qualities such as independence and legal sophistication, nor did it touch on institutional qualities such as legitimacy, public acceptance, and decisional coherency.³¹ In many ways, it adopted a black-box approach to justice by measuring only input and output—that is, the number of cases filed compared to the number of cases closed. It further focused on the number of cases in which a litigant appealed or filed a petition, in which a court's decision could be reversed, or in which it might be remanded by the higher court for retrial. In short, the metric system was most concerned about the *processing* capacity of the court system. Many cases done is cases well done.

²⁶ Ng, *supra* note 8.

²⁷ Ng & He, *supra* note 4.

²⁸ Keith, Lin, & Hou (2013).

²⁹ Porter (1996).

³⁰ Dingwall & Lewis (1983), p. 5.

³¹ The Guiding Opinions in Supreme People's Court, *supra* note 2, note that public satisfaction can be measured by giving out survey questionnaires to people's congress representatives, members of political consultative committees, parties to litigation and their lawyers, and the public (Art. 18). In reality, it is seldom practised. See "Discussion" below.

Of course, from the perspective of the SPC, how well a court performed was reflected in the numbers of judgments corrected, when a higher court either reversed its decision or remanded the case for retrial. Too many judgments corrected was an indication of sloppiness. As we shall see, however, the reality turned out to be far more complicated.

Furthermore, the CQAS was not only designed to hold judges accountable. The system also promoted accountability. A number of social scientists have explored the effect of metrics on behaviour.³² They use the concept of reactivity to describe the interactions between measurement indicators and the actions they set out to measure.³³ A reactive measure is one that changes, often unwittingly, the very thing that a researcher is trying to measure.³⁴ In the case of the CQAS, it did not set out to just measure performance. On the contrary, one might say that inducing reactivity was one of its primary objectives, since the SPC intended to use the scoring system to improve the performance of front-line judges. Unlike ranking systems that set out merely to measure the performance of individuals and organizations (such as law schools³⁵), the CQAS was designed to improve what it measured. It pushed judges to improve their performance through enhanced scrutiny and self-criticism. As stated in Article 2 of the “Guiding Opinions on Conducting Case Quality Assessment,” the system was conducive to improving the quality of judges and to reinforcing their sense of responsibility.³⁶ The SPC in effect aimed to follow the famous adage of management guru Tom Peters: “What gets measured gets done.”³⁷

4.3 Did the CQAS determine reward and punishment fairly?

Chinese judges often complain about their meagre salary. Resentment is growing particularly among younger, college-educated judges working and living in big cities. For one thing, the living standards in cities such as Beijing and Shanghai have been rising rapidly for decades. For another, the judges who live there experience the familiar phenomenon of relative deprivation. They tend to compare themselves with lawyers who are earning much more money in the private sector. As the urban economy grows, the public–private pay gap widens. As a result, many urban courts suffer from growing rates of attrition. Judges in the big cities rely heavily on their year-end bonuses to make up for their lean base salaries. These bonuses are, however, performance-based. According to the Judges Law,

appraisal of judges shall be conducted by the People’s Courts the judges belong to The result of appraisal shall be taken as the basis for award, punishment, training, [or] dismissal of a judge, and for readjustment of his or her grade and salary.³⁸

During the years over which the scheme was in effect, performance as measured by the CQAS was an important determinant of bonuses or the lack thereof. As mentioned, the indices applied to multiple levels. At the individual level, court leaders used a judge’s performance scores to determine the size of his/her bonus pay. In some cases, a bonus could

³² Espeland & Sauder (2007); Mau, *supra* note 22; Muller, *supra* note 22.

³³ Campbell (1957); Espeland & Sauder, *supra* note 32.

³⁴ Campbell, *supra* note 33, p. 298.

³⁵ Espeland & Sauder, *supra* note 32.

³⁶ Supreme People’s Court, *supra* note 2.

³⁷ Peters, *supra* note 1.

³⁸ Judges Law of the People’s Republic of China (*zhonghua renmin gongheguo faguanfa*) 中华人民共和国法官法 (2019), Arts 38 and 42. Since the end of the CQAS, judges have been evaluated by a group of indices, some of them reminiscent of the CQAS indices. The inputs of senior judges also continue to play a prominent role in the exercise.

account for as much as the base salary of a grassroots judge.³⁹ Judges we interviewed in urban cities said that doing well at least in the core indices of the CQAS was a way to make sure one received a bonus payment at the end of the year. If a given judge's performance on a particularly important index was substantially weaker compared to those of other judges, the person might face a loss of bonus pay or even a fine; in more serious cases, the judge was transferred or demoted.⁴⁰

In sum, metric fixation fuelled the aspiration to replace evaluations based on experience with standardized measurement, since the former, as Muller points out, was viewed negatively as personal, subjective, and self-interested.⁴¹ One merit of the CQAS touted by the SPC was its objectivity and transparency. Indeed, as a data-based ranking system, the CQAS could be termed an "objectivity generator."⁴² What went wrong then? In the following section, we analyze how the CQAS in its implementation created unintended consequences that eventually led to its abrupt abolition.

5. Problems of implementing the CQAS

5.1 Selective objectivity

The SPC, in its "Guiding Opinions," outlined how case quality and judicial efficacy could be measured under the CQAS. But, because the SPC allowed provincial and intermediate courts to adjust the weight given to individual indices, there were small local variations in the application of the CQAS across provinces and even among prefectures in the same province. It was, however, in the interpretation of the meaning of the scores that courts differed sharply with one another.

In practice, the meaning of excellence became contested as provincial high courts engaged with one another in battles of one-upmanship. Provinces that performed well in the CQAS system liked to boast about their judicial excellence. In 2013, the Shanghai High People's Court claimed that it had been ranked "No. 1" in the "fairness" (*gongzheng*) score of the CQAS for the past five years. The Jiangsu High People's Court said it was "leading the country" in the overall CQAS score. The Guangdong High People's Court proclaimed it was "consistently in a leading position" in many of the indices. Even those that did not perform as well promoted themselves by highlighting the specific indices on which they excelled. The Henan High People's Court announced that it had improved its national rankings in many indices. The Shaanxi High People's Court said that it had more indices in an "advantageous position" than before. In short, courts picked and chose what to emphasize when publicizing how they performed in the CQAS.⁴³

Furthermore, beyond these interprovincial competitions, courts across localities put different emphases on the CQAS. The encompassing nature of the scheme meant that it was too broad for courts to focus on every item. There were over 30 indices altogether; and it was difficult for a court to excel in all of them. When a metric so comprehensive was adopted, it pushed people or organizations under evaluation to pick and choose.⁴⁴ Consequently, some provincial courts elevated specific indices to the status of "hard targets" (*ying zhibiao* 硬指标).⁴⁵ Hard targets were not only targets that the provincial courts

³⁹ Kinkel and Hurst suggest that most performance-based bonuses were paltry. They conclude that judges wanted good CQAS scores mainly not for financial reasons, but because of face. See Kinkel & Hurst, *supra* note 15, p. 948. Our research suggests that there were significant local variations.

⁴⁰ *Ibid.*, p. 947.

⁴¹ Muller, *supra* note 22.

⁴² Hornbostel et al. (2010); Mau, *supra* note 22, p. 41.

⁴³ Gao (2015).

⁴⁴ Whiting (2004), p. 116.

⁴⁵ Kinkel & Hurst, *supra* note 15, pp. 941–2; Minzner (2011).

pursued without compromise to achieve good results. They were also hard in the plainer sense that they were difficult to achieve. The SPC used hard targets to address the weak spots in the performances of some local courts.

In the case of Court J, the internal newsletters showed that not all indicators were given the same level of attention. By far, the two key indices that the court leaders most often emphasized were the rate of cases resolved within statutory time limits (*fading (zhengchang) shen xiannei jie'an lv* 法定(正常)审限内结案率) and the rate of first-instance cases in which judgment was corrected on appeal (mistake) (*yi shen panjue anjian bei gaipan fahui chongshen lv (cuowu)* 一审判决案件被改判发回重审率(错误)). Within the CQAS framework, the former measured efficiency and the latter fairness. Difficult cases take time to adjudicate, but the judges were nevertheless expected to resolve all cases within statutory time limits.⁴⁶

In the CQAS system, these two indices received the most attention because they were weighted heavily in the total score (see below). They were also emphasized because senior judges needed the total effort from front-line judges in order to score well on them. Of course, some indicators were within the control of the court's own management. For example, a resourceful court such as Court J could, to some extent, raise its mediation rate without the full co-operation of its judges. Li, Kocken, and van Rooij describe some of the strategies that court management could use to boost the mediation rate.⁴⁷ The same can be said about the rate of first-instance cases in which the simplified procedure was used (*yi shen jianyi chengxu shiyong lv* 一审简易程序适用率). Senior judges had control over what cases were tried summarily (i.e. the so-called simplified procedure) and what cases were tried by the use of a collegial panel, the Chinese equivalent of a judicial panel.⁴⁸ But the rate of cases resolved within statutory time limits and the rate of cases in which judgment is corrected on appeal were two indices that required strong efforts from front-line judges. As we will see, the first index (rate of cases resolved) was most strenuously resisted by front-line judges.

5.2 Sweeping objectivity

While the multi-index CQAS varied in its implementation, the composition of individual indices, as directed by the SPC, was so rigid and sweeping that they irked many grassroots judges. Some indices were just unrealistic. For example, “the rate of cases in which judgment is voluntarily executed” was an index so detached from reality that many grassroots courts, Court J included, simply ignored it.⁴⁹ Judgments were rarely carried out voluntarily without the involvement of the enforcement division of the court. Some other indices that were viewed as too elusive were also overlooked, including the index that is most closely related to the theme of this Special Issue—“quality of judgment.” We return to that index later in the discussion.

Perhaps no index better illustrated the rigid nature of the CQAS system than the *rate of cases resolved within statutory time limits*—one of the two indices to which Court J paid the most attention. The title of this index is clunkily named even by SPC standards. It was specifically developed by the architect of the CQAS. Before its adoption, courts just

⁴⁶ There are officially legitimate reasons for courts to “stop the clock,” i.e. to go beyond the statutory time limit of six months for civil cases. Yet, it is still considered a bad sign if too many cases legitimately go beyond the time limit. See “Discussion” below.

⁴⁷ One strategy is to have judicial clerks seek out cases that are already mediated and then encourage parties to file the cases officially. These cases go through the case-filing process so that they can be counted as court-mediation cases in the statistics. Li, Kocken, & van Rooij (2018).

⁴⁸ At the level of the grassroots courts, a collegial panel is a three-member panel formed to hear a case. Besides judges, many panels in grassroots courts also include laypeople's assessors.

⁴⁹ See also Kinkel & Hurst, *supra* note 15, p. 943.

submitted their clearance rate, measured by the number of outgoing cases as a percentage of the number of incoming cases. The clearance rate is arguably the most basic indicator used by courts everywhere in the world. But, by the 2000s, its validity had become more and more questionable in China. Many grassroots courts tried to raise their clearance rate not by increasing the number of cases disposed of, but by depressing the number of cases they took in. It was an open secret that, in many parts of China, it was difficult for litigants to file cases after October because courts were determined to protect their shiny clearance rates. The SPC was well aware of this kind of gamesmanship practised by the lower courts. The original intent of using the rate of cases resolved within statutory time limits was to assuage the concerns of grassroots judges and to eliminate their resistance to accepting new cases in the last months of the year—because only cases that were not disposed beyond the statutory limit of six months counted.⁵⁰

Yet, this newly coined index was almost universally loathed by grassroots judges. In dealing with difficult cases, judges sometimes asked for an extension, but senior judges were reluctant to grant them in normal circumstances. More often, grassroots judges attempted to buy time by “stopping the clock” on the six-month deadline (*kouchushen-xian* 扣除审限). Judges devised a set of well-rehearsed reasons to stop the clock, some of them valid, others less so. Sometimes litigants were nowhere to be found; sometimes new evidence had to be sought; sometimes a litigating party died in the middle of the trial and the next of kin stepped in; sometimes negotiations had to be conducted with other government bureaus; and sometimes judges believed a settlement was near and wanted more time. The days taken to complete the activity for stopping the clock did not count towards the total time taken to dispose of the case. For example, if it took two months for the plaintiff in a personal-injury case to obtain a medical report, those two months would not count toward the total days taken by the judge to dispose of the case. But the SPC became concerned that the clock was being stopped too casually. There were occasions on which judges stalled because they knew that, however they ruled, their decisions would be appealed by the parties concerned and they desperately wanted to avoid that. And there were also idle judges who took advantage of the practice and abused it to conceal their lack of effort.

Application of the rate of cases resolved within statutory time limits was akin to the legal concept of “strict liability.” It did not inquire into what could be causing the delay; a case was counted as overdue for whatever reason if it was not resolved beyond the normal statutory limit of six months. Many front-line judges were not happy about this new index, yet it was the most heavily weighted among the “efficiency” group of indices. As mentioned, despite occasional abuses, front-line judges do often need more time to resolve their cases. But the index made no distinction between legitimate and illegitimate excuses. In one internal newsletter (No. 23), the senior judges of Court J reported the findings of a random examination of 200 cases that stayed open beyond the statutory limits. The study concluded that about 47% of the 200 cases offered legitimate reasons for extension, including solicitation of forensic evidence and additional investigation, and followed the required procedures to suspend the statutory time limits. Another 47% provided reasons but did not observe all the required procedures. Only the remaining 6% (15 cases) used illegitimate reasons to suspend time limits. However, the way in which the index score was compiled meant that all these cases were lumped together as outstanding beyond the statutory or normal time limits. They were all deemed unresolved cases. They were also bothersome sore spots for the judges.

It is thus understandable that many front-line judges felt frustrated about the sweeping “objectivity” of the CQAS, which they viewed as crude and undiscerning. Front-line judges we interviewed lamented that the “objective” measures of CQAS overlooked the

⁵⁰ Tong & Yan (2011).

complexity of the cases they handled. It is noteworthy that, even in a judiciary lacking judicial independence, sweeping quantification provoked genuine resistance. The numbers generated by the CQAS offered a purportedly objective assessment without ever addressing the substantive question of procedural rightness or wrongness. In the eyes of the judges, these numbers simply failed to reflect reality. As simplifications and abstractions, they may have staked a claim for *the* reality, in the sense of what really mattered,⁵¹ but it was clearly a reality that reflected the specific interests of the SPC. Even senior judges in grassroots courts disliked the index. To them, it reflected a deep mistrust of their ability to supervise front-line judges. In an effort to achieve objectivity and to avoid possible abuses by such judges, the SPC resorted to a “strict” indicator that pushed them to act more quickly even when they were presiding over cases so complex that patience and tact were called for.

Frustration with the CQAS was particularly palpable for judges in Court J, who heard mixed messages. The county in which Court J is located is home to many foreign and overseas companies. There is a strong presence of Taiwanese investors. Reminding judges to be cautious and prudent when dealing with cases involving Taiwanese investment, the newsletter (No. 13) cautioned: “Adjudicating Taiwan-related cases is not purely a legal exercise. These cases are related to politics, economics, and society. They are politically sensitive not just because they are related to Taiwan; but also because these cases can affect our city.” The newsletter urged judges to rely even more heavily on mediation to resolve Taiwan-related cases. Yet they also had to consider the index of rate of cases resolved within statutory time limits. Thus, under the CQAS, judges could no longer afford to be too patient even when dealing with complex and politically sensitive cases.

5.3 Relative objectivity

Objectivity as understood in the CQAS was also a relative notion. This is where the analogy of the decathlon breaks down. Decathletes compete not directly with each other, but against a scoring table. There is, however, no preset, fixed-value scoring table in the CQAS. The system did not prescribe definite values of what qualifies as a good, or even a satisfactory, performance. The meaning of a CQAS score was determined relatively by rankings. Article 12 of the 2011 Guiding Opinions states that, when the SPC evaluates the performance of a court, it should refer to the national highest, lowest, and average values of indices to determine their “reasonable levels” (*helizhi* 合理值) and “warning levels” (*jingshizhi* 警示值). The document does not specify qualifying cutoffs for different indices. It adopted a purely relative (aka competition) approach—whether a court performs well or not was to be judged by how well it performed relative to other courts.

This notion of relative objectivity had far-reaching consequences for what the CQAS actually meant to courts. The scores did not in themselves provide yardsticks; they were simply the results of a competition. While the practice was not officially endorsed by the SPC, it was not unusual for provincial high courts to publish “league tables” that ranked the performances of their lower-level courts in individual “hard-target” indices. Court J’s province was among the few affluent provinces that enthusiastically promoted the use of such tables. A March newsletter (No. 17) stated:

Between January and March, our court’s rankings are in general unsatisfactory. Among the 17 fundamental indices and the 18 diagnostic indices that count the most, we are behind in many of them, compared with other grassroots courts in the city.

⁵¹ Mau, *supra* note 22, p. 18.

Ranking winners and losers became iterative at multiple levels. It started with provincial comparisons and moved down to the levels of prefectural and county courts. It then descended to the level of divisions within the same court and ultimately pitted individual judges against one another. This put extreme pressure on front-line judges. It also pressured court leaders to produce “shiny” results.

Unfortunately, not only did Court J perform below average in overall rankings; it also trailed in two “hard-target” indices. Its rate of cases closed within statutory time limits fell short of the standard set by the provincial high court (see “Culture of Discipline” below). Its rate of cases that were corrected on appeal was relatively too high. Grassroots courts hate having their decisions reversed or remanded for retrial. Before the implementation of the CQAS, the percentages of reversal and retrial were *the* indices that mattered. After implementation, this same perspective was reflected in the heavy weight that the CQAS assigned to the indices that used reversal and retrial rates as their components. The CQAS turned the rate of judgments corrected upon appeal into “hard-target” indices for many courts, especially courts in the most urbanized areas of the country, as these courts were expected to display a particularly high degree of professionalism (and hence low correction rates). In an issue of its internal newsletter (No. 4), the court leadership wrote:

Including the fourth quarter of 2010 and all of 2011, there were altogether 102 cases reversed or remanded on appeal. Our reversal/retrial rate is 0.65 per cent. Among all the grassroots courts in City X, we rank last. A total of 73 cases (among the 102 cases), or 71.57 per cent, were reversed. A total of 29 cases, or 28.43 per cent, were remanded for retrial. By types of cases, there were four criminal cases (3.92 per cent), 97 civil cases (95.10 per cent), one administrative case (0.98 per cent) . . . We hope divisions will carefully review each and every case that was reversed and each and every case that was remanded. We hope to find out the reasons for these decisions. We need to strengthen our quality control on judging so that we can improve the quality of judging.

The newsletter’s author asked all divisions to conduct a thorough post-mortem (*shenru pingchafenxi* 深入评查分析) of every case corrected by the intermediate court. Even without knowing the specifics of the cases that were either reversed or remanded for retrial, one thing is clear: civil cases almost single-handedly contributed to the last-place ranking. The numbers that came out of Court J were consistent in this respect with a broader national trend. Among the urban courts of China, civil cases have become more complex. In commercial cases (the fastest-growing category of civil cases), many courts are seeing litigation with much bigger financial stakes. Appeals are consequently more likely.

That said, by most standards, a 0.65% reversal rate is very low. Any municipal court in the US would gladly take a reversal rate of 0.65% as its own. One might say that Chinese courts abhor appeals with much greater intensity than courts in other countries. But, even within China, Court J faced unusually stiff competition. In the city of Shenzhen, Kinkel and Hurst reported that a basic-level court there was asked to keep the percentage of cases reversed or remanded on appeal to 12% or less.⁵² Court J would have done very well in Shenzhen rather than in its better-performing geographical area. But, with this system of relative ranking, a good number is worth little if others in the same area are performing noticeably better. Courts were rewarded or punished according to their place in the league table.

Relative objectivity raised the stakes of the CQAS competition. The system inevitably produced losers. There was no passing score, and the goalpost was always moving,

⁵² Kinkel & Hurst, *supra* note 15, p. 946.

depending on how competitive other neighbouring courts were. In the case of Court J, competition turned from intense to cut-throat. With all that was at stake, court leaders began to question the objectivity of several index scores; and, in particular, they questioned the rate of reversal and retrial in which their court ranked last. Advocates of metricization in general argue that publicized rankings can make a system more transparent. It is a way to hold institutions that are being evaluated accountable.⁵³ Yet the relativity of the CQAS scores created tension among all courts—not just courts that were competing with one another at the same level, but also courts related to each other in a vertical hierarchy, that is, the supervising and the supervised.

To manage its poor performance, Court J asked its divisions to conduct internal reviews of every case that was corrected. The review process was spearheaded by the court's adjudication committee. In the first quarter of 2012, a total of 16 cases were either reversed or remanded by the intermediate court. Court J reviewed all of them. In some cases, the court questioned whether the intermediate court's decision was either fair or correct. In other words, it challenged the rulings that had contributed to its poor showing. The findings were published in an April issue of the newsletter (No. 26): "After discussion by our adjudication committee, eight of the (reversed/remanded) cases were 'satisfactory.' Six cases were 'basically satisfactory.' Two were 'unsatisfactory.'"

As many as half of the cases reversed or remanded by the intermediate court were, in the eyes of the senior judges of Court J, decided correctly ("satisfactory") at trial. "Basically satisfactory" was evidently an intermediate category whose meaning was never clearly explained in the newsletters. Reading the assessments in context, however, suggests that decisions labelled "basically satisfactory" involved cases in which Court J's leaders agreed that the intermediate court conducted a more thorough examination of facts, and their interpretation of the law was admittedly more sophisticated than that of the trial judge. Nonetheless, the trial judges' performance in these cases could be considered "basically satisfactory" if one took into account the heavy case-load and lack of resources available to them. To put it in plain terms, a judge who is given a "basically satisfactory" assessment could have done better, but he or she did not commit any egregious mistakes. The intermediate court may have re-examined the case more thoroughly, but that was to be expected, as it handled a much smaller number of cases. The trial judge in such circumstances did not deserve to be punished.

The authors of the newsletters did not hesitate to identify the "mistakes" of the intermediate court. In one issue, the newsletter singled out a case that was remanded for retrial after the intermediate court had decided to add a bank as a new plaintiff:

Our review concludes that the original trial was robust in its factual findings, and its conclusion was backed up by ample evidence. This case should not include the bank as the new plaintiff, and the decision by the appellate court to remand the case for retrial was unreasonable.

Even more revealing were the motives alleged by the leaders of Court J behind the intermediate court's decisions to correct cases. The same newsletter commented on a hit-and-run case that was remanded for retrial. It depicted the intermediate court in an unflattering light:

In that traffic accident case, the intermediate court said the facts of the case were not clearly identified in the first trial. It said that was why the case was remanded. But it did not offer any details about the facts that were not identified. The real reason this case was sent back to us is that the defendant's sister sent letters to everyone. The

⁵³ Muller, *supra*, note 22.

case is now a tangled web. The intermediate court sent it back for retrial. The reason this case was remanded is that the intermediate court didn't want to get itself into trouble.

The poor performance of Court J on this index that continued for more than a year (at the time at which the newsletter was published) had called into question the legitimacy of the court management. A common way of dealing with and compensating for this kind of dissonance within organizations is to resort to alternative narratives.⁵⁴ The post-hoc investigations reported in the newsletters are examples of generating alternative narratives. They challenged the infallibility of the appeal reviews and even questioned the integrity of the intermediate court.

5.4 Perverse consequences

Another symptom of metric fixation is the belief that attaching rewards and penalties to measured performance is the best way to motivate people to perform. But, often times, what is transformed is not just the outcome itself, but the way in which the measured outcome is achieved. To paraphrase Karl Marx, people change their behaviour in reaction to being evaluated, but they do not change it necessarily as the architects of the CQAS please.⁵⁵ Very often, the high stakes of the CQAS competition turned performance evaluation into a combat sport. Courts tried almost everything to game the system.

Again, the newsletters offer telltale clues about how the court made changes to its daily operations in order to climb in the CQAS rankings. In one issue (No. 19), a senior official offered an in-depth look at the composition of the CQAS indices. He wrote:

When dealing with related indices, divisions should rethink their approach to adjudication and adjust their working methods. Let's use the rate of mediation and the rate of case withdrawal as examples. The weights of the two indices are 12 per cent and 5 per cent respectively. So when litigants settle and apply to withdraw their cases, we must try our very best to ask the litigants to close their cases in the form of (court) mediation.

The newsletter author implies, in a not-too-subtle manner, that front-line judges should be more strategic if they want to come out ahead in the CQAS game. In the province in which Court J is located, the mediation rate (12%) was weighted more heavily than the case-withdrawal rate (5%) in the computation of the total CQAS score.⁵⁶ Usually plaintiffs withdrew their cases when they settled with defendants out of court. But these settlements were private, in the sense that the court did not formally endorse the settlement. As a result, the cases were marked as "withdrawn cases" (rather than "settled cases") in the official records. Instead, the newsletter author asked judges to formalize and endorse as many private settlements as possible, to transform these withdrawn cases into settlements in the books. This is simply playing the numbers game. Although it would produce more mediated cases in the books, there would be no more settled cases in fact, since the rise in the number of mediated cases would be generated by converting cases already settled privately by parties. There would be a corresponding drop in the number of

⁵⁴ Mau, *supra* note 22.

⁵⁵ The original quote from Marx is a line in *The Eighteenth Brumaire of Louis Bonaparte*: "Men make their own history, but they do not make it just as they please." Marx & Engels (1978), p. 595.

⁵⁶ Provincial high courts were allowed by the SPC to adjust the weight distribution of the indices. While the weights of hard-target indices were quite fixed, the weights of mediation rates and case-withdrawal rates varied across provinces.

withdrawn cases. But, because of their differential weights, the “rise” in the mediation rate would more than compensate for the “decline” in the withdrawal rate to produce a net gain in the CQAS score.

There was even more that a court could do to game the system. The same newsletter broke down the composition of the index of the mediation rate. The author suggests, not in so many words but perhaps to the same effect, that the most “effective” way to move up in the rankings is to mediate more standard-procedure cases, other things being equal. The simplified procedure (trial by one judge) dealt with straightforward cases; the standard procedure (trial by a collegial panel) dealt with more complicated ones. In a perverse way, some grassroots courts started most cases in the simplified-procedure docket. When a case was about to reach settlement, the court then moved it to the standard-procedure docket. Procedurally, it was doing things backwards. A case about to be settled did not need to be tried, let alone by a panel of judges. But this misuse of the two procedures had the effect of raising the CQAS score, since the mediation rate for standard-procedure cases was a bigger component in the overall mediation rate than the mediation rate for simplified-procedure cases. As mentioned, in some urban courts, performance bonuses that were almost equal to judges’ base salaries were calculated on the basis of the CQAS scores. This was enough of a perverse incentive for judges to achieve the right outcomes in the wrong way. Incidentally, this backward procedure also helped to soften the impact of difficult cases on the index score, since cases that proved impossible to mediate remained as cases handled by the simplified procedure.

5.5 Culture of discipline

Financial rewards were not the only factors that motivated judges’ performance. Rankings that fell short of expectations tended to provoke criticisms and ill feelings among the people who were being ranked.⁵⁷ The process is the punishment, to borrow the title of Malcolm Feeley’s well-known book.⁵⁸ And punishment comes in the form of public shaming, which is also a powerful motivator. As we mentioned earlier, one of the core indices was the rate of cases resolved without delay or, more precisely, the rate of cases resolved within statutory (normal) time limits (*fading (zhengchang) shenxian nei jie’an lǜ*法定(正常)审限内结案率). Very often, cases were delayed for reasons beyond the control of the responsible judges (see above).

The newsletters regularly included a full list of cases that remained unresolved beyond the normal statutory limits. The cases were grouped into three tables—cases that remained open for 12 to 16 months; cases that remained open for 16 to 18 months; and cases that remained open beyond 18 months. The tables were the literary totems of shame, one taller than the next, with the last table—cases that remained open beyond 18 months—the *tallest* of all (even though it is usually the shortest on the page). These tables were a visualization device that displayed the “faults” of individual judges and their divisions. They were “inscriptions” to remind everyone what was at stake.⁵⁹ For each of the unresolved cases, the tables listed the case number, the date of filing, the division of the court responsible, and, above all, the name of the responsible judge. The detailed naming practice was meant to put pressure on the judges and their divisions. In the first half of the year 2012 alone, these tables of outstanding cases appeared 12 times (Nos 5, 7, 9, 10, 14, 15, 21, 22, 27, 29, 30, and 31) for all judges to see. They covered page after page in the newsletters. In some issues, the entire newsletter consisted only of a few long tables of outstanding cases. Nothing more was said, because nothing needed to be said. But, at

⁵⁷ Elsbach & Kramer (1996).

⁵⁸ Feeley (1979).

⁵⁹ Latour & Woolgar (1986); Mau, *supra* note 22.

times, the newsletter authors had harsh words for judges who did not dispose of their cases in time. One issue (No. 23) states:

A small number of judges lack a sense of responsibility. Some cases are prolonged without reason. In the two hundred cases we analyzed, there are a total of 13 cases in which the responsible judge suspended the statutory time limit (i.e., stopped the clock) with lame excuses. In the judicial procedure management forms, the reasons that were put down to suspend the statutory time limit were arbitrary. The reasons given were random. Some judges abused mediation as a reason to extend their cases. In some forms, mediation was mentioned multiple times, but the case file never provided any details about mediation.

As we have seen, comparisons based on quantitative indicators fostered a competitive climate.⁶⁰ Judges in the same court may have been friends, but they were also rivals. As a result of the CQAS, courts established a clear pecking order of judges from the best-performing to the worst-performing. To push its judges to work harder to close cases, the leaders of Court J stepped up their pressure as the semi-annual deadline (the end of May) approached. In the newsletter circulated towards the end of March (No. 14), the court implemented the following measures:

- Judges have to give weekly progress updates to their division on unresolved cases
- Judges are required to tell the court a target date for resolving each new case they are assigned
- Judges will be held accountable to resolve each case by the target date they set for themselves

A month and a half before the mid-year deadline of 31 May, the newsletter reminded the judges, for the umpteenth time:

Based on our latest assessment, it remains a tall order for our court to clear all unresolved cases. We are just one and a half months from the target date, yet we are 51.24 per cent short of the target we set for ourselves. We hope all responsible judges will step up their efforts, overcome whatever obstacles, and get the job done.

Just ten days before the deadline, the persistent push by the court management finally yielded results. The court as a whole managed to raise the clearance rate to 88.37%; 45 cases still remained unresolved, but a total of 342 cases out of the 387 unresolved cases were cleared. The target was to achieve a clearance rate of 90%, and that meant there were only seven more cases to go. The newsletter (No. 31) announced:

Only ten days before the deadline. We must achieve the clearance rate of 90 per cent for cases that remain unresolved beyond statutory time limits. The difference is just 1.63 per cent. In other words, among the remaining 45 cases, there are seven more cases to go. The longer we wait, the more difficult it is going to be. We hope all responsible judges will work even harder and overcome obstacles to get the job done in these final ten days.

And then, on 1 June, the next newsletter (No. 34) was circulated. It proclaimed triumphantly:

⁶⁰ Mau, *supra* note 22.

Our job to clear outstanding unresolved cases is basically done. The outcome is significant. On the 31st of May, our court cleared a total of 367 cases; there is only a total of 20 cases that remain outstanding. The clearance rate is 94.83 per cent. We have already fulfilled the goal set by the High People's Court.

But the celebration was swift and short-lived. The newsletter then read:

Even though we have finished our campaign to clear unresolved cases and achieved the numbers the High People's Court asked of us, there is no room for complacency. We have a persistently high number of unresolved cases. This drags down other CQAS indices that are related to it. Clearing outstanding cases is therefore a long-term mission.

We have no data to tell what the judges in Court J did to close cases that they had hitherto failed to resolve. But it is obvious that top-down pressure from senior judges of Court J worked magic, judging by the numbers alone. In our fieldwork, we have seen front-line judges resorting to extraordinary measures to resolve difficult cases, including paying multiple personal visits to recalcitrant litigants, pulling *guanxi* to get help from officials of other government bureaus to address litigant's concerns, and, in some cases, using the court's own funds to offer extra compensation to litigants.⁶¹ These tactics, we suspect, did not disturb the SPC and provincial high courts too much. The CQAS was not designed merely to measure the quality of judicial work, but also to measure the effectiveness of judges performing their roles as grassroots bureaucrats.

Finally, we must briefly mention the cases that judges failed to resolve. Table 2 lists the remaining cases in the three civil divisions of Court J. These were the cases that, despite the court's all-out effort, remained unresolved past the biannual deadline. Several of these cases involved Taiwanese and Korean defendants; some involved the time-consuming procedure of issuing public notices (when defendants could not be tracked down); a few involved investigation or auditing; one involved a jurisdictional dispute; and another required adding new parties. In short, these cases reflected the growing complexity of civil cases that big city courts like Court J were asked to deal with. In particular, the steady presence of cases involving defendants from Taiwan meant that Court J was at a regular disadvantage in its competition with neighbouring courts.

6. Discussion: judging about judging

What does the CQAS tell us about judging and judgment in China? It is important to point out that the CQAS is a project of judging about judging. The point is made evident in actor-network theory's treatment of judging (*putare*) and calculating (*computare*) as two activities that are intimately linked.⁶² By computing a list of indices to measure judicial performance, the CQAS defines what good judging is. Hence, one way to probe deeper into the meaning of judging is to take a close look at what the CQAS is said to measure. As mentioned, three secondary-level labels are used to describe what the CQAS substantively measures—fairness, efficiency, and impact. The ways in which the indices are bundled together inform us about how the SPC leaders understand and define these concepts.

Let us start with the notion of impact. The Chinese courts are known to be pragmatic and even populist.⁶³ Judicial rulings must consider their potential social impact. This pragmatic orientation was reflected in the construction of the CQAS. Much of what the

⁶¹ Ng & He, *supra* note 4.

⁶² Callon & Law (2005); Callon & Muniesa (2005).

⁶³ Liebman (2011).

Table 2. List of unresolved civil cases of Court J on 31 May 2012

Case no.	Date filed	Case name	Responsible judge	Action plan to close case
Civil Court No. 1				
(2011) J Civil 1518	9 May 2011	Rental-contract dispute: <i>XXX Trade Co. v. XXX Employment Agency Co.</i>	Judge XI	New notices sent to defendant; public notices; scheduled to close in June
(2011) J Civil 1577	12 May 2011	Property-sale dispute: <i>XXX & XXX v. XXX</i>	Judge Y	Case with complicated details; second hearing scheduled; plan to close in June
(2011) J Civil 1087	31 March 2011	Private-loan dispute: <i>XXX, XXX v. XXX</i>	Judge W	Defendant is Taiwanese; judicial assistance in service of judicial documents failed; public notices issued; case unlikely to be resolved by end of May
(2011) J Civil 1458	27 April 2011	Rental-contract dispute: <i>XXX Consultant Co. (Shanghai) v. XXX</i>	Judge W	Defendant is Taiwanese; judicial assistance in service of judicial documents failed; public notices issued; case unlikely to be resolved by end of May
(2011) J Civil 1467	28 April 2011	Claimed repayment by guarantor: <i>XXX v. XXX (Taiwan-related)</i>	Judge W	Defendant is Taiwanese; judicial assistance in service of judicial documents failed; public notices issued; case unlikely to be resolved by end of May
(2011) J Civil 1360	18 April 2011	Personal-injury dispute: <i>XXX v. XXX, XXX Insurance Co.</i>	Judge Z	Public notices issued. Plan to close by July
(2011) J Civil 1418	22 April 2011	Probate dispute: <i>XXX v. XXX</i>	Judge S	Complex unresolved disputes; investigation underway; plan to close case by June
(2011) J Civil 1642	18 May 2011	Private-loan dispute: <i>XXX Interior Design Ltd. v. XXX</i>	Judge C	Public notices issued after defendant was not found; defendants reappeared and contested our court's jurisdiction; that dispute is in process; can't close case by the end of June
Civil Court No. 2				
(2011) J Commerce 0572	25 May 2011	Partnership-agreement dispute: <i>XXX, XXX & XXX v. XXX</i>	Judge H	This case involves auditing; plan is to close the case in June
Civil Court No. 3				
(2011) J Commerce Foreign 0015	1 April 2011	Claimed repayment by guarantor dispute: <i>XXX v. XXX</i>	Judge X2	Defendant died while the case was in progress; his beneficiaries in Taiwan and illegitimate son in China are now the defendants; five beneficiaries in Taiwan; additional identification documents needed; judicial assistance from Taiwan sought; hearing scheduled for 20 June 2012; judge in liaison with Taiwan court; plan to close case by end of June
(2011) J Commerce Foreign 0023	31 May 2011	Dispute of contract for work: <i>XXX Dyeing Co. v. Shanghai XXX Fashion Co.</i>	Judge X2	Since defendant is Korean, judicial documents delivered through diplomatic channel; hearing scheduled for 13 June 2012; plan to close case by end of June

“impact” indices measured was about enforcement. While it was not exclusively concerned with enforcement work, no other aspects of the courts were as closely identified with this group of indices, which included the rate of cases enforced, the rate of cases in which enforcement objectives were realized, and the rate of cases in which judgment was voluntarily executed.⁶⁴ “Impact” is a powerful proxy for gauging the efficiency of enforcement. It determines whether a judgment actually has its presumptive effect. This is of course not a trivial matter, as enforcement has long been a challenge for Chinese courts.⁶⁵ Important as impact was, however, the CQAS construed it quite narrowly. Impact was measured by the courts’ ability to follow through on what they decided. It was conceived, in Weberian terms, as the coercive power of law.⁶⁶ What was missing from this conceptualization was *judicial* impact in its entirety. Did a decision offer insight to help solve a novel legal problem? Did it clarify the law on some particular issue? Did it shape the development of the law in some area? In the US, there is a sizable literature dedicated to identifying the criteria for measuring case impact or importance, including the frequency of citation in subsequent decisions, the extent of news coverage, the number of law review notes, and the inclusion of the case in popular casebooks and textbooks, among others.⁶⁷ These criteria are meant to gauge the substantive significance of a decision. In the case of the CQAS, measurements of substantive significance were conspicuous by their absence. It is not an exaggeration to say that the impact of a decision was measured without considering its content. It is, however, difficult to measure the substantive impact of a Chinese judgment. The Chinese courts do not adhere to a system of precedents. The SPC has promoted a Chinese-style system of impact cases,⁶⁸ but the purpose is more for top-down control to ensure that lower courts follow the lead of the SPC.

The other CQAS indices that measured “impact” seem to negate the very concept of the direct impact of law. In the CQAS, the rate of case withdrawal and the rate of mediation both contributed positively rather than negatively to the impact index! When cases were withdrawn, however, the plaintiff had actually decided not to pursue a legal resolution at all; and, when cases were mediated, the parties had pursued a non-legal resolution. It is clear, therefore, that, when the CQAS designers envisioned “impact,” their focus was more on the administrative impact of the court bureaucracy and less on the impact of legal norms or principles. From the perspective of the SPC, grassroots courts augmented their impact if they used the law judiciously. The wisdom of judging in China was not just about knowing how to apply the law rightly; it was also about knowing when it was the right moment to apply the law and when it was not.⁶⁹

It is hardly surprising that the group of indices that made up “fairness” were mostly measurements of an internal, bureaucratic type. The “fairness” index included, as we have seen, reversal and retrial rates. It also included the rate of appeal. It was, in other words, a thinly veiled proxy for measuring how well grassroots courts delivered the law to the satisfaction of the higher courts and, above all, the SPC. Other items included in the measurement of fairness tended to be formalistic, such as the rate of trials in which a lay assessor was present and the rate of cases in which the filing was changed. What was noticeably absent again is any subjective fairness assessments by litigants about the decisions of the

⁶⁴ Other than enforcement-related items, petitioning by litigants is also counted as a negative impact in the impact score. Petitioning a decision means launching a complaint against the judge’s handling of the case. Petitioning is outside of the formal appeal channel. If a litigant petitions against a judge’s decision, the judge’s CQAS score suffers as a consequence.

⁶⁵ He (2012).

⁶⁶ Weber (1978).

⁶⁷ Cross & Spriggs II (2010).

⁶⁸ Wang (2019).

⁶⁹ Ng & He, *supra* note 4.

courts.⁷⁰ Judges in our interviews contended that the litigants' assessments had little value, since the parties themselves would consider a judgment as fair only if it was favourable to them, and an unfavourable judgment would, according to these same judges, invariably be seen as unfair. It testifies to the lack of trust of the Chinese judicial system that judges do not believe there is room for non-outcome-based assessments about fairness. "Fairness" in the CQAS might even be viewed cynically as a means for the higher courts to tighten their grip on lower courts, since decisions of lower courts were deemed fair if they were not changed by the higher courts. In any court system, the appeals mechanism, by giving the appellate court the last word, confers power to the higher court over the lower court. Yet, by labelling this a matter of fairness, the CQAS implied that not only were the higher courts more technically *correct*; they were also *fairer*.

Among the 11 indices in the "fairness" group, there was one index that stood out for its substantive nature: the quality of judgment. However, in many grassroots courts, this was an index purely for show. Court leaders did not devote much effort to raising the score of this particular index. In the absence of an established system of precedents, it is difficult to talk about the quality of a judgment *qua* judgment. What most court leaders asked of its grassroots judges was that they follow standardized templates issued by the SPC when producing judicial documents. Other courts that bothered to pay attention to this index similarly chose to focus on the most formalistic attributes of the quality of judgment. In the province of Shaanxi, for example, its High People's Court issued in 2010 a draft rubric to all of its grassroots courts. The rubric proposed a four-category grading system of judgments: "excellent," "good," "pass," and "fail."⁷¹ This rubric rewarded summaries of facts and legal arguments that were both clear and comprehensive. It commended applications of the law that it considered correct and logical. But these are rather abstract criteria to operationalize. The most elaborate and concrete part of the rubric actually addressed the mechanics of a judgment rather than its substance. Points were deducted, for example, if "a judgment made more than six misspellings or typos" or if it "deviated from the standard template of judgment." More points were deducted if a judgment "failed to mention case type" or if it did not "state the full name of the court." This focus on formal elements continued beyond 2014. In 2016, the SPC issued a new "Specifications for Preparing Civil Judgments by the People's Courts,"⁷² which still serves as an updated template for civil judgments in all of China and preserves many of the stylistic requirements set forth in the CQAS. Surely there is more to a good judgment than being impeccable in craft and style?

While the way in which "fairness" and "impact" were constructed by the CQAS appears strained and often arbitrary, "efficiency" was by comparison clear and straightforward for the most part. This group of indices were meant to push grassroots courts and their judges to be faster and more productive. As seen in the foregoing discussion, the key indices were the rate of cases resolved within statutory time limits and the number of cases resolved per judge per year, among other measurements. It was this aspect of the CQAS that made judging resemble a tournament. Judging became no different from other administrative activities of the party-state. The central government in general favours competition to incentivize other branches of local governments to perform all of their functions better, as exemplified in the winner-takes-all tournament among local governments to achieve higher GDP growth. Like other governmental units, Chinese courts at different levels (provinces, municipalities, counties, and townships) were made to compete against each other in the CQAS system in a zero-sum performance-rankings contest in which passing grades were always relative but rankings with reference to one's peers were all-important.⁷³

⁷⁰ Tyler (1990).

⁷¹ Shaanxi High People's Court (2010).

⁷² Supreme People's Court (2016).

⁷³ Xu (2011).

7. Conclusion: the legacy of the CQAS

The CQAS was ended abruptly in the last days of 2014. By the numbers alone, the CQAS was an unqualified success. In analyzing the scores of 2012, the Deputy Director of SPC's Research Bureau Yan Ge wrote that courts reached new heights in achieving fairness, efficiency, and impact. The numbers regressed slightly in 2014, the final year in which the system was in place, but the decline was so marginal that it could hardly be considered the reason why the system was scrapped.⁷⁴

Why, then, did the SPC terminate what appeared to be a very successful policy? The highest court of China offered scant explanation, but the answer is already evident in the foregoing analysis. As we have suggested, the rise and fall of the CQAS should be considered a case of metric fixation. Underneath the surface of the “objective” numbers lay the real reasons for its demise. For judges and courts, much was at stake in the zero-sum game it created. The CQAS was set up in such a way that it bred perverse incentives. High scores on individual indices were achieved through means that the SPC did not anticipate. Just before the policy was scrapped, the SPC issued some cryptic comments about how the performance metrics caused the “reification of data” (*weishujulun* 唯数据论) and a “meta-physical” (*xingershang* 形而上) attitude toward the numbers being reported.⁷⁵ This Marxism-inflected jargon alluded to the doctoring of numbers that diverged from reality. For many of the component indices, both the front-line judges who performed routine judging and the senior judges who oversaw their work wanted to score well. Trial judges were of course motivated to come up with shiny scores for their year-end bonuses and a better career future. Court leaders who supposedly oversaw the work of front-line judges also benefited from exaggerated positive reports. They put pressure on their subordinates to resolve the maximum number of cases within the shortest possible amount of time. Their prospects of promotion to bigger and more important courts turned on how well their courts ranked against others. Much was at stake, perhaps too much for all the parties involved, from the top of the judicial system to the bottom. We have discussed the perverse ways through which some of the results were achieved. The underlying incentivizing scheme of the CQAS was asymmetrically one-sided, all “push” and no “pull”—neither front-line judges nor their superiors would want to see their CQAS scores go down.⁷⁶

In many other ways, the cut-throat nature of the CQAS was a breeding ground for sabotage. A payoff system determined by relative performance breeds subversion perhaps out of fear of sabotage.⁷⁷ We observed that judges tended to find their best friends in other divisions or courts where direct competition was absent. We also noticed that intense evaluations at times sowed distrust between front-line judges and their immediate superiors. Front-line judges constantly complained about the cases that their division head assigned them, since a problem case with difficult or elusive litigants might remain unresolved for months beyond the statutory deadline and could potentially be a career killer. Their complaints persisted even after many grassroots courts adopted computerized systems to randomly assign cases to judges. Perceived favoritism further undermined trust and authority.⁷⁸ The grass always seemed greener on the other side, and the cases assigned to other judges just seemed easier to close. In interviews conducted during the CQAS years, some judges even suggested to us that the computer program used by their court was rigged!⁷⁹

⁷⁴ Yan & Yuan, *supra* note 10.

⁷⁵ People's Court Daily (2013).

⁷⁶ This is an example of the promotion-tournament model of China.

⁷⁷ Except for the judge responsible for the case, other members of a collegial panel can be fellow judges from the same division or lay assessors. See Section 5.1 for further details.

⁷⁸ Prendergast & Taper (1996).

⁷⁹ Ng & He, *supra* note 4.

But the most natural sites of sabotage were found in the two workhorse indices of the CQAS—the rate of appealed cases in which judgment was subsequently corrected and the rate of petitioned cases in which a final judgment was corrected.⁸⁰ While grassroots courts dedicated much of their energy to minimizing the value of the two indices, their desire to do so was in conflict with intermediate courts' desire to promote their own corresponding indices—the rate of reopened trials among lower courts' judgments that were reviewed and the rate of retried cases in which the court corrected a lower court's judgment that had already been entered (*dui xiaji fayuan shengxiao anjian tiqi zaishen lv* 对下级法院生效案件提起再审率; *dui xiaji fayuan shengxiao anjian zaishen gaipan fahui chongshen lv* 对下级法院生效案件再审改判发回重审率). A higher court could only promote its own rate of correcting judgments at the expense of the courts below it. This, as we have seen, strained the relationship between Court J and its intermediate court. The leadership team of Court J who wrote the newsletters devised counter narratives to suggest that some cases were remanded by the intermediate court out of self-preservation. It is worth noting that, when the SPC decided to scrap the CQAS system, they did it, among other things, upon the request of lower courts.⁸¹

The comprehensive scope of the CQAS was meant to make all the grassroots courts in China comparable along a unified scale. It was, in the sense that Star and Lampland have proposed, an act of standard-making.⁸² Yet this grand experiment in the end proved unsuccessful. Even in a highly bureaucratic system that adhered to no single substantive notion of legal validity, the goal of imposing a unified scale proved to be elusive. As the Chinese legal system has grown more complex, different branches of law have become more specialized. It is simply unrealistic to expect a complex intellectual property case, for example, to be resolved in the same timeframe as a personal loan dispute in a small-claims court.⁸³ Legal specialization aside, there are other aspects of judging and judgment that cannot be easily quantified. In China, these include political acumen, communications with the media, public perception, and social stability maintenance, among others. In the case of Court J, as we have discussed, the court had to tread carefully when dealing with cases involving Taiwanese investments. In many ways, there proved to be a limit to which the CQAS metric could simplify social and political complexity.

The SPC is of course well aware of the reality of judging in the context of China. The CQAS revealed not so much the SPC's ignorance, but its mistrust of the grassroots courts. The SPC questioned the commitment of lower-court judges to their work. As mentioned in our analysis of the rate of unresolved cases, the index highlighted the struggles of the SPC in its cat-and-mouse game with the lower courts. But the frustration was reciprocal. There was strong pushback against the CQAS by grassroots courts. More than anything, this struggle shows the limits of administration supervision and oversight as a substitute for professionalism. The resentment towards the CQAS experiment challenged the belief that only numbers mattered. The CQAS's approach was determinedly externalist—that is, it refrained from evaluating the legal impact of judging. Judging was treated as an administrative activity. Judging cannot be *judged* by efficiency and output alone, but this was precisely what the CQAS attempted to achieve.

The CQAS was by no means a side note for understanding the judicial system of China. Recently, we asked some Chinese judges what they thought of the legacy of the CQAS. They

⁸⁰ Scott, *supra* note 13.

⁸¹ Hu (2014).

⁸² Star & Lampland, *supra* 23.

⁸³ In some areas of civil law, e.g. intellectual property, the Chinese courts are piloting more complex procedural mechanisms such as expanded discovery, evidence preservation, and burden-of-proof reversals, and clearer rules regarding the obligations of parties to produce evidence. See Chinaipr.com (2020).

commented that the CQAS may have disappeared in name but not in substance. Front-line judges we talked to said that, even though the SPC stopped the use of some CQAS indices, including the much-resented rate of outstanding cases beyond the statutory time limits, things have not changed fundamentally. What led the judges to this conclusion, given that the termination of the CQAS was widely regarded as a major victory for the lower courts?⁸⁴ This is especially puzzling in light of the fact that the courts, under the instructions of the SPC, have reverted to using the clearance rate as the primary statistic to measure judicial efficiency. A middle-aged judge in Beijing explained:

Yes, they now use the “clearance rate” again. But to me it’s just another set of numbers trying to do the same thing. Whether it is the CQAS indices or the clearance rate now, it is all the same. There is no fundamental difference.

The judge added: “You are still pushed by your superiors to close outstanding cases.” What the judge suggested is that the meaning of the clearance rate is now understood through the lens of the CQAS “strict-liability” index. In practice, grassroots judges are no longer at liberty to “stop the clock” when it is needed. They are subject to the ever-tightening scrutiny of their superiors, who in turn have to answer to their superiors in the provincial court and ultimately the SPC. The CQAS has been abolished, yet it is enjoying a robust afterlife. In many grassroots courts, judges now get automatic reminders of the deadlines of their unresolved cases when they log into their desktop computers. If a case remains open beyond the statutory time limit, the judge responsible for the case receives a flashing reminder on his or her computer screen to attend to it.

The themes that were emphasized by the CQAS continue to dominate the agenda of the SPC.⁸⁵ Some metrics, though no longer subsumed within the unified framework of the CQAS, continue to serve as important tools for the SPC and other provincial courts to assess the performance of judges. For example, as we have seen, the “impact” indices of the CQAS placed a heavy emphasis on enforcement. In 2018, SPC President Zhou Qiang delivered a report at the Sixth Session of the Standing Committee of the 13th National People’s Congress, highlighting various measures to strengthen enforcement in many areas of government. It is quite revealing that, four years after termination of the CQAS, Zhou in his report continued to boast of a countrywide judicial enforcement rate of over 90% in civil cases.⁸⁶

Thus, the CQAS, or one should say the metric fixation that fuelled the creation of the CQAS, continues to shape the work environment of front-line judges in China. In 2019, the World Bank ranked China as one of the world’s top ten most improved economies for ease of doing business, touting its judicial efficiency in enforcing contracts and resolving insolvencies.⁸⁷ The legacy of the CQAS is shown in the continued use of metrics to drive the courts to be more efficient, particularly in the business and commercial sector.⁸⁸ Though it ended in 2014, the twin pillars that defined the CQAS project—the drive to quantify and the focus on output—have continued to shape the bureaucratic orientation of the Chinese courts even today.

Acknowledgements. The authors thank David Engel for his detailed and extensive comments on early versions of this article. We also thank the anonymous reviewer for helpful feedback.

⁸⁴ Hu, *supra* note 81.

⁸⁵ And, in fact, the social credit system recently promoted by the Chinese government could be said to be motivated by the same metric-based approach to governance.

⁸⁶ Court.gov.cn (2018).

⁸⁷ Worldbank.org (2019).

⁸⁸ Supreme Peoples Court Monitor (2020).

References

- Callon, Michel, & John Law (2005) "On Qualculation, Agency, and Otherness." 23 *Environment and Planning D-Society & Space* 717–33.
- Callon, Michel, & Fabian Muniesa (2005) "Economic Markets as Calculative Collective Devices." 26 *Organization Studies* 1229–50.
- Campbell, Donald T. (1957) "Factors Relevant to the Validity of Experiments in Social Settings." 54 *Psychological Bulletin* 297–312.
- Chan, Peter C. H. (2016) "An Uphill Battle: How China's Obsession with Social Stability Is Blocking Judicial Reform." 100 *Judicature* 14–23.
- Chinaipr.com (2020) "SpC's 2020 Ip-Related Judicial Interpretation Agenda," <https://chinaipr.com/2020/03/31/spcs-2020-ip-related-judicial-interpretation-agenda/> (accessed 29 July 2020).
- Court.gov.cn (2018) "Report of the Supreme People's Court on the Work by the People's Court to Solve the 'Enforcement Difficulty'," <http://www.court.gov.cn/zixun-xiangqing-124841.html> (accessed 29 July 2020).
- Cross, Frank B., & James F. Spriggs II (2010) "The Most Important (and Best) Supreme Court Opinions and Justices." 60 *Emory Law Journal* 407–502.
- Dingwall, Robert, & Philip Simon Coleman Lewis (1983) *The Sociology of the Professions: Lawyers, Doctors, and Others*, New York: St. Martin's Press.
- Elsbach, Kimberly D., & Roderick M. Kramer (1996) "Members' Responses to Organizational Identity Threats: Encountering and Countering the Business Week Rankings." 41 *Administrative Science Quarterly* 442–76.
- Espeland, Wendy Nelson, & Michael Sauder (2007) "Rankings and Reactivity: How Public Measures Recreate Social Worlds." 113 *American Journal of Sociology* 1–40.
- Feeley, Malcolm (1979) *The Process Is the Punishment: Handling Cases in a Lower Criminal Court*, New York: Russell Sage Foundation.
- Gao, Xiang (2015) "Zhongguo Difang Fayuan Jingzheng De Shijian Yu Luoji [The Practices and Logic of the Competition among Chinese Local Courts]." 21 *Law and Social Development* 80–94.
- He, Xin (2012) "A Tale of Two Chinese Courts: Economic Development and Contract Enforcement." 39 *Journal of Law and Society* 384–409.
- Hornbostel, Stefan, Jürgen Kaube, Alfred Kieser et al. (2010) "Unser Täglich Ranking Gib Uns Heute. Über Das Vertrauen in Ratings, Rankings, Evaluationen Und Andere Objektivitätsgeneratoren Im Wissenschaftsbetrieb," in A. Kehnel, ed., *Kredit Und Vertrauen*, Mannheim: Frankfurter Allgemeine Buch, 51–76.
- Hu, Weixin (2014) "Zuigao Renmin Fayuan Jueding Quxiao Dui Quanguo Ge Gaojirenminfayuan Kaohe Paiming [The Supreme People's Court Decided to Revoke the Assessment and Ranking of the High Courts]," *People's Court Daily*, 27 December.
- Keith, Ronald C., Zhiqiu Lin, & Shumei Hou (2013) *China's Supreme Court*, Hoboken: Taylor and Francis.
- Kinkel, Jonathan J., & William J. Hurst (2015) "The Judicial Cadre Evaluation System in China: From Quantification to Intra-State Legibility." 224 *China Quarterly* 933–54.
- Latour, Bruno, & Steve Woolgar (1986) *Laboratory Life: The Construction of Scientific Facts*, Princeton, NJ: Princeton University Press.
- Li, Yedan, Joris Kocken, & Benjamin van Rooij (2018) "Understanding China's Court Mediation Surge: Insights from a Local Court." 43 *Law & Social Inquiry* 58–81.
- Liebman, Benjamin L. (2011) "A Populist Threat to China's Courts?" in M. Y. K. Woo & M. E. Gallagher, eds., *Chinese Justice: Civil Dispute Resolution in Contemporary China*, New York: Cambridge University Press, 269–313.
- Marx, Karl, & Friedrich Engels (1978) *The Marx-Engels Reader*, New York: Norton.
- Mau, Steffen (2019) *The Metric Society: On the Quantification of the Social*, London: Polity Press.
- Minzner, Carl (2006) "Xinfang: An Alternative to Formal Chinese Legal Institutions." 42 *Stanford Journal of International Law* 103–79.
- Minzner, Carl (2011) "China's Turn against Law." 59 *American Journal of Comparative Law* 935–84.
- Muller, Jerry Z. (2019) *The Tyranny of Metrics*, Princeton: Princeton University Press.
- Ng, Kwai Hang (2019) "Is China a 'Rule-by-Law' Regime?" 67 *Buffalo Law Review* 793–821.
- Ng, Kwai Hang, & Xin He (2017) *Embedded Courts: Judicial Decision-Making in China*, New York: Cambridge University Press.
- O'Brien, Kevin (2003) "Neither Transgressive nor Contained: Boundary-Spanning Contention in China." 8 *Mobilization* 51–64.
- People's Court Daily (2013) "Yonghao Anjian Zhiliang Pinggu Tixi Zhezhang Tjianbiao [Editorial: Make Good Use of the CQAS as a 'Health Check Form']," <https://www.chinacourt.org/article/detail/2013/06/id/1016368.shtml> (accessed 29 July 2020).
- Peters, Tom (1986) "What Gets Measured Gets Done," <https://tompeters.com/columns/what-gets-measured-gets-done/> (accessed 29 July 2020).

- Porter, Theodore M. (1996) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*, Princeton, NJ: Princeton University Press.
- Prendergast, Canice, & Robert H. Taper (1996) "Favoritism in Organizations." 104 *Journal of Political Economy* 958–78.
- Scott, James C. (1990) *Domination and the Arts of Resistance: Hidden Transcripts*, New Haven: Yale University Press.
- Seawright, Jason, & John Gerring (2008) "Case Selection Techniques in Case Study Research—a Menu of Qualitative and Quantitative Options." 61 *Political Research Quarterly* 294–308.
- Shaanxi High People's Court (2010) "Caipan Wenshu Zhiliang Pingding Biaozhun (Zhengqiu Yijian Gao) [A Rubric for Evaluating the Quality of Judicial Documents (Consultative Draft)]," on file with authors.
- Star, Susan, & Martha Lampland (2009) "Reckoning with Standards," in M. Lampland & S. L. S. Ithaca, eds., *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Ithaca: Cornell University Press, 3–25.
- Supreme People's Court (2005) "Renmin Fayuan Di Er Ge Wunian Gaige Gangyao [Second Five Year Reform Program for the People's Courts (2004–2008)]," <https://www.cecc.gov/resources/legal-provisions/second-five-year-reform-program-for-the-peoples-courts-2004-2008-cecc#body-chinese/> (accessed 1 January 2021).
- Supreme People's Court (2008) "Guanyu Kaizhan Anjian Zhiliang Pinggu Gongzuo De Zhidao Yijian (Shixing) [Guiding Opinions on Conducting Case-Quality Assessment (Trial Implementation)]," http://pkulaw.cn/fulltext_form.aspx?Gid=104209 (accessed 1 January 2021).
- Supreme People's Court (2011) "Guanyu Kaizhan Anjian Zhiliang Pinggu Gongzuo De Zhidao Yijian [Guiding Opinions on Conducting Case-Quality Assessment]," <http://www.court.gov.cn/zixun-xiangqing-2298.html> (accessed 1 January 2021).
- Supreme People's Court (2016) "Renmin Fayuan Minshi Caipan Wenshu Zhizuo Guifan [Specifications for Preparing Civil Judgments by the People's Courts]," http://www.pkulaw.cn/fulltext_form.aspx?Db=chl&Gid=274653/ (accessed 1 January 2021).
- Supreme Peoples Court Monitor (2020) "Supreme People's Court's New Vision for the Chinese Courts," <https://supremepoplescourtmonitor.com/2020/05/04/supreme-peoples-courts-new-vision-for-the-chinese-courts/> (accessed 30 July 2020).
- Tong, Ji, & Pingchao Yan (2011) "Fading (Zhengchang) Shen Xiannei Jie'an Lü Zhibiao De Shiye [Questions about the Rate of Cases Resolved within Statutory Time Limits]," *People's Court Daily*, 29 October.
- Tyler, Tom R. (1990) *Why People Obey the Law*, New Haven: Yale University Press.
- Wang, Shucheng (2019) "Guiding Cases and Bureaucratization of Judicial Precedents in China." 14 *University of Pennsylvania Asian Law Review* 96–135.
- Weber, Max (1978) *Economy and Society: An Outline of Interpretive Sociology*, Berkeley: University of California Press.
- Whiting, Susan H. (2004) "The Cadre Evaluation System at the Grass Roots: The Paradox of Party Rule," in B. J. Naughton & D. L. Yang, eds., *Holding China Together: Diversity and National Integration in the Post-Deng Era*, New York: Cambridge University Press, 101–19.
- Worldbank.org (2019) "Doing Business 2020: China's Strong Reform Agenda Places It in the Top 10 Improver List for the Second Consecutive Year," <https://www.worldbank.org/en/news/press-release/2019/10/24/doing-business-2020-chinas-strong-reform-agenda-places-it-in-the-top-10-improver-list-for-the-second-consecutive-year> (accessed 29 July 2020).
- Xu, Chenggang (2011) "The Fundamental Institutions of China's Reforms and Development." 49 *Journal of Economic Literature* 1076–151.
- Yan, Ge, & Chunxiang Yuan (2013) "2012 Nian Quanguo Fayuan Anjian Zhiliang Pinggu Fenxi Baogao [Report of the National Case Quality Assessment in 2012]." 13 *Renmin Sifa [The People's Jurisprudence]* 50–52.
- Yan, Ge, & Chunxiang Yuan (2015) "2014 Nian Quanguo Fayuan Anjian Zhiliang Pinggu Fenxi Baogao [Report of the National Case Quality Assessment in 2014]." 9 *Renmin Sifa [The People's Jurisprudence]* 82–84.

Cite this article: Ng, Kwai Hang and Chan, Peter C. H. (2021). "What Gets Measured Gets Done": Metric Fixation and China's Experiment in Quantified Judging. *Asian Journal of Law and Society* 8, 255–281. <https://doi.org/10.1017/als.2020.28>