



# Comparability, Competition and Control: Performance Management in the Correctional Services of Germany and England and Wales

Nathalie Iloga Balep, Andrei Guter-Sandu, Andrea Mennicken und Christian Huber

## Contents

1	Introduction .....	357
2	Research Design and Method .....	360
3	Performance Measurement in the Penal Justice Sectors of Germany and England and Wales .....	361
4	Conclusion .....	377
	Appendix .....	379
	References .....	381

---

Die Originalversion dieses Kapitels wurde ursprünglich ohne Open Access veröffentlicht.  
Ein Erratum ist verfügbar unter  
[https://doi.org/10.1007/978-3-658-35620-0\\_19](https://doi.org/10.1007/978-3-658-35620-0_19)

---

N. Iloga Balep (✉)

Institut für Unternehmensrechnung und Controlling, Helmut-Schmidt-Universität  
Hamburg, Hamburg, Germany  
E-Mail: [nathalie.ilogabalep@hsu-hh.de](mailto:nathalie.ilogabalep@hsu-hh.de)

A. Guter-Sandu · A. Mennicken

Department of Accounting, The London School of Economics and Political Science,  
London, UK  
E-Mail: [A.Guter-Sandu@lse.ac.uk](mailto:A.Guter-Sandu@lse.ac.uk)

A. Mennicken

E-Mail: [A.M.Mennicken@lse.ac.uk](mailto:A.M.Mennicken@lse.ac.uk)

C. Huber

Department of Operations Management, Copenhagen Business School, Frederiksberg,  
Denmark  
E-Mail: [chu.om@cbs.dk](mailto:chu.om@cbs.dk)

© Der/die Autor(en) 2022, korrigierte Publikation 2023

W. Wirth (Hrsg.), *Steuerung und Erfolgskontrolle im Strafvollzug*, Edition Forschung und Entwicklung in der Strafrechtspflege,  
[https://doi.org/10.1007/978-3-658-35620-0\\_15](https://doi.org/10.1007/978-3-658-35620-0_15)

---

### Zusammenfassung

In den vergangenen drei Jahrzehnten hat Performance Measurement im öffentlichen Sektor über New Public Management stetig zugenommen, so auch im Strafvollzug. Dieses Kapitel vergleicht Leistungsmessungssysteme in den Strafvollzugssystemen von Deutschland sowie England und Wales und geht dabei auf drei Hauptthemen ein: Erstens beschreiben wir die grundlegenden Merkmale beider Strafvollzugssysteme. Zweitens gehen wir der Frage nach, welche formalen Gründe hinter der Einführung von Performance Measurement-Instrumenten stehen. Drittens untersuchen wir die Effekte von Leistungsmessungssystemen auf die Steuerung des Strafvollzugs. Abschließend diskutieren wir, unter welchen Bedingungen Leistungsmessungssysteme einen produktiven Austausch zwischen verschiedenen Akteuren, Interessen und teilweise konfliktären Werten und Zielen ermöglichen können. Dem in englischer Sprache verfassten Beitrag ist eine deutschsprachige Kurzfassung vorangestellt.

---

### Abstract

Over the last three decades, performance measurement has advanced to become a ubiquitous part of public management and has also spread to the penal sector. This chapter compares the performance measurement systems that have come to exist in the penal sectors of Germany and England and Wales by discussing three main themes: First, we describe and analyse the different measurement systems in use. Second, we compare different rationales and objectives underlying the introduction of the different performance measurement systems. Third, we examine uses and effects of the respective systems. We then discuss implementation challenges in both systems and potential negative unintended consequences. We suggest that by keeping the performance measurement systems open to negotiation and debate, such systems may help to mediate between different prison values and objectives. The article, written in English, is preceded by a German summary.

### **Deutschsprachige Kurzfassung: Vergleichbarkeit, Wettbewerb und Kontrolle: Performance Management in den Strafvollzügen von Deutschland und England/Wales**

Performance Measurement ist in den vergangenen drei Jahrzehnten zu einem festen Bestandteil der Steuerung öffentlicher Organisationen geworden (Hood, 1991; Pollitt et al., 2007). Der Trend zu Leistungsmessungen hat auch den Strafvollzug erreicht (Mennicken, 2013). Das Wie, Was und Warum von Leistungsmessungen

unterscheidet sich in den Strafvollzugssystemen verschiedener Länder allerdings beträchtlich (James and Hood, 2004). Solche Unterschiede lassen sich auch im Fall von Europas zwei größten Volkswirtschaften, Deutschland und Großbritannien (UK), beobachten. Deutschlands Strafvollzug verzeichnet seit 2003 eine sinkende Anzahl an Gefangenen (Drenkhahn, 2018). Aufgrund Deutschlands föderaler Struktur, gibt es zwischen den einzelnen Bundesländern mehr oder weniger große Unterschiede in der Organisation und Steuerung des Strafvollzugs. Charakteristisch für Deutschlands öffentlichen Sektor ist aber eine generelle Skepsis gegenüber der Forderung nach mehr Transparenz. Im Vordergrund stehen meist die Erfüllung gesetzlicher Vorgaben, die durch die jeweiligen Landesjustizministerien kontrolliert wird. Dabei spielt der Zugang der Öffentlichkeit zu administrativen Vorgängen traditionell eine geringe Rolle. In Großbritannien hingegen ist der Strafvollzug bereits seit zwei Jahrzehnten einem bedeutenden wirtschaftlichen Leistungsdruck ausgesetzt. Großbritannien hat die höchste Inhaftierungsrate in Westeuropa und viele Justizvollzugsanstalten sind überfüllt. Seit den britischen New Public Management-Reformen der späten 1980er Jahren stehen Wettbewerb, Vergleichbarkeit und Transparenz auch im Zentrum des Strafvollzugs. Vor diesem Hintergrund vergleichen wir in diesem Kapitel Performance Measurement-Systeme in den Strafvollzugssystemen von Deutschland und England und Wales, mit einem besonderen Fokus auf Leistungsmessungen und Leistungsmonitoring (s. Wirth in dem Einleitungsbeitrag „Steuerungsrelevante Erfolgskontrolle“ zu diesem Band) in den Strafvollzugsanstalten. Wir gehen den verschiedenen Zielen und Ursprüngen der Entstehung von solchen Messungen nach, wie auch ihren Effekten auf die alltägliche Steuerung der Anstalten. Im Fall von Großbritannien konzentrieren wir uns auf England und Wales und nicht auf Großbritannien als Ganzes, weil der Strafvollzug von England und Wales unter die Aufsicht derselben Verwaltung fällt, während Schottland und Nordirland in den Verantwortungsbereich dezentraler Ministerien gehören. Zudem ist ‚Her Majesty’s Prison and Probation Service of England and Wales‘ (HMPPS) bei Weitem der größte Strafvollzug Großbritanniens. Von 135 Strafvollzugsanstalten in Großbritannien, befinden sich 117 in England oder Wales, von denen wiederum 13 von drei privaten Vertragsnehmern geführt werden: G4S, Sodexo und Serco.<sup>1</sup>

Im Rahmen eines internationalen Forschungsprojektes haben wir das Aufkommen und die Verbreitung von Performance Measurement in drei unterschiedlichen öffentlichen Sektoren (Hochschulen, Gesundheit, Strafvollzug) in Europa

---

<sup>1</sup> Siehe <https://www.gov.uk/government/statistics/prison-population-figures-2021>, abgerufen am 26 April 2021.

(Deutschland, Großbritannien, Frankreich, Niederlande) untersucht.<sup>2</sup> Insgesamt wurden im Strafvollzug von drei Bundesländern in Deutschland 44 und im Strafvollzug von England/Wales 47 teilstrukturierte Interviews geführt. Nach einem kurzen Einblick in unsere methodische Vorgehensweise (2), konzentrieren sich unsere Erkenntnisse auf drei Hauptthemen: In einem ersten Schritt beschreiben wir die grundlegenden Merkmale der im Strafvollzug von England/Wales und Deutschland jeweils eingesetzten Leistungsmessungssysteme (Sect. 3.1). Bezeichnend für den deutschen Strafvollzug ist, dass jedes Bundesland sein eigenes Leistungsmessungssystem verwendet. Die eingesetzten Messinstrumente sind häufig klassische Controlling- und Budgetierungsinstrumente, die allerdings in mehreren Bundesländern durch moderne Steuerungsinstrumente wie die Balanced Scorecard (BSC) ergänzt werden, die neben den rein finanziellen und operativen Kennzahlen auch Informationen über qualitative Ziele wie Resozialisierung abbilden kann. In England/Wales sind die Messsysteme deutlich komplexer und elaborierter: So werden beispielsweise über die Ländergrenzen hinweg im ‚Prison Performance Hub‘ standardisierte Kennzahlen erhoben und fast tagesaktuell ausgewertet.

In einem weiteren Schritt untersuchen wir, welche formalen und expliziten Gründe hinter der Einführung von Leistungsmessungssystemen stehen (Sect. 3.2). In Deutschland spielen Zielvereinbarungen zwischen den Justizministerien und den Anstaltsleitungen die größte Rolle, wobei sich systematische Vergleiche auf spezielle Interessensfelder (z. B. Jugendstrafvollzug) beschränken. In England/Wales sind die meistgenannten Gründe für den Einsatz von Messsystemen eine politische und administrative Kontrolle über bzw. die Schaffung von Rechenschaftspflichten für die Justizvollzugsanstalten.

Schließlich gehen wir der Frage nach, wie Leistungsmessungssysteme tatsächlich eingesetzt werden und welche Effekte sie verursachen (Sect. 3.3). In Deutschland tragen Messsysteme keineswegs zu einer Standardisierung bei, sondern verstärken in ihrer Diversität die vom Föderalismus verursachte Fragmentierung. Durch die Einführung unterschiedlicher Messinstrumente werden Vergleiche zwischen einzelnen Bundesländern nicht vereinfacht, sondern eher erschwert. In England und Wales hingegen zielen Steuerungsinstrumente insbesondere auf Vergleichbarkeit ab, stark gekoppelt an einen Markt- und Wettbewerbsgedanken. Tab. A.1 im Anhang fasst die Ergebnisse zusammen.

Abschließend heben wir in diesem Beitrag zwei grundlegende Herausforderungen von Leistungsmessungssystemen hervor, die wir im Kontext der Steuerung

---

<sup>2</sup> Projektbeschreibung unter: <https://www.lse.ac.uk/accounting/carr/research/quad>.

als besonders relevant erachten. Erstens ist es von Bedeutung, welche Organisationseinheit als Grundlage für die Leistungsmessung genutzt wird. Werden Justizvollzugsanstalten individuell betrachtet, droht die Leistungszuschreibung für das gesamte Strafvollzugssystem zu kurz zu greifen, weil nur abgebildet werden kann, was innerhalb einer einzelnen Einrichtung passiert, während Faktoren vor oder nach der Inhaftierung, wie beispielsweise die Rückfälligkeit und die grundsätzliche gesellschaftliche Bedeutung des Strafvollzugs, ausgeblendet werden. Zielt die Leistungsmessung hingegen auf das gesamte Strafvollzugssystem beispielsweise innerhalb eines Bundeslandes ab, gehen möglicherweise die Anreizwirkungen für einzelne Anstalten verloren. Zweitens ist es wichtig, eine kritische Distanz zu den Messsystemen zu wahren und sich ihrer teilweise verzerrenden und einschränkenden Effekte bewusst zu sein. Performance Measurement-Systeme sollten für Debatten und Verhandlungen offenbleiben. Mehr noch: Leistungsmessungssysteme sollten als „Brennglas“ für Introspektion und Reflektion verstanden werden und nicht als Instrument von Schuldzuweisungen. So können Performance Measurement-Systeme als Anstoß und Plattform für den Austausch zwischen verschiedenen Akteuren, Interessen und teilweise konfliktären Werten und Zielen, wie Ökonomie, Sicherheit und Resozialisierung, dienen und einen Raum für Diskussionen und Lerneffekte schaffen.

This work was supported by the Economic and Social Research Council (grant number ES/N018869/1) and the Deutsche Forschungsgemeinschaft (grant number: 627097) under the Open Research Area Scheme (Project Title: QUAD – Quantification, Administrative Capacity and Democracy). The QUAD project is an international project co-funded by the Agence National de la Recherche (ANR, France), Deutsche Forschungsgemeinschaft (DFG, Germany), Economic and Social Research Council (ESRC, UK), and the Nederlands Organisatie voor Wetenschappelijk Onderzoek (NOW, Netherlands).

---

## 1 Introduction

Performance measurement has advanced to become a ubiquitous part of public management over the last three decades (Hood, 1991; Pollitt et al., 2007) and the mega-trend towards measuring performance has also spread to the penal sector (Mennicken, 2013). Yet, the how, what and why of performance measurement differs significantly between different countries' penal systems (James and Hood, 2004). We can observe such differences also in the case of Europe's two largest economies, Germany and the United Kingdom (UK). Germany has a penal system that is characterized by a declining imprisonment rate. Germany's

penal system has a long tradition of law-based governance following largely the command and control model of regulation. Due to Germany's federal political structure, there is variation in how penal systems are organized and governed in different states (Länder). Traditionally, public agencies have had a rather sceptical stance towards transparency. In contrast, the UK has a penal system that has faced substantial economic and operational pressures over the past two decades. In western Europe, the UK has one of the highest imprisonment rates and many prison establishments are overcrowded. Since the late 1980s, the stimulation of competition and enhancement of comparability and transparency have been at the centre of British New Public Management reforms, also in the penal system.

Against this backdrop, this chapter compares and contrasts the performance measurement systems that have come to exist in the penal sectors of Germany and England and Wales, focusing in particular on performance measurement in prison establishments. It examines the different objectives and rationales underlying their introduction, as well as their effects on day-to-day prison management. In the case of the UK, we focus on England and Wales, rather than the UK as a whole, because of UK's devolved governance structure. The prison service of England and Wales is under the purview of the same administration, whereas Scotland and Northern Ireland fall under the jurisdiction of devolved ministries. Furthermore, Her Majesty's Prison and Probation Service of England and Wales (HMPPS) is by far the largest prison service that exists in the UK. In March 2021, out of a total of 135 prisons situated in the UK, 117 were located in England and Wales, out of which 13 were privately managed by three contractors: G4S, Sodexo and Serco.<sup>3</sup>

This chapter presents findings from an international comparative research project that examined the rise and spread of performance measurement in three different public services (higher education, healthcare, correctional services) in Europe (Germany, UK, France, Netherlands).<sup>4</sup> The findings presented in this

---

<sup>3</sup> See <https://www.gov.uk/government/statistics/prison-population-figures-2021>, accessed 26 April 2021.

<sup>4</sup> Amongst other things, this project was supported by the Economic and Social Research Council (grant number ES/N018869/1) and the Deutsche Forschungsgemeinschaft (grant number: 627097) under the Open Research Area Scheme (Project Title: QUAD – Quantification, Administrative Capacity and Democracy). The QUAD project is an international project co-funded by the Agence National de la Recherche (ANR, France), Deutsche Forschungsgemeinschaft (DFG, Germany), Economic and Social Research Council (ESRC, UK), and the Nederlands Organisatie voor Wetenschappelijk Onderzoek (NOW, Netherlands). More information about the project can be accessed on <https://www.lse.ac.uk/accounting/carr/research/quad>.

chapter are based on 44 interviews which we conducted in the German penal sector (in three different states) and 47 interviews which we conducted in the penal sector of England and Wales. Following a brief summary of the research design and methods we applied (Sect. 2), the presentation of our findings is structured around three main themes (for a summary of findings see also Tab. A.1): First, we describe and analyse the different measurement systems in use (Sect. 3.1). Whereas in the German case more reliance tends to be placed on traditional tools of budgetary control, which only more recently have come to be complemented with modern performance management instruments like the Balanced Scorecard, in the case of England and Wales, there exists an elaborate, highly centralized and standardized performance measurement system, which includes publicly available composite prison ratings (presented on a scale of 1 to 4).

Next, we compare different rationales and objectives underlying the introduction of the different performance measurement systems (Sect. 3.2). Here, we find that in Germany, budgetary planning has often been a main driver of quantification, whereas other concerns, such as learning, have been largely limited to special areas of offender management, like juvenile detention. In contrast, in England and Wales, political control, surveillance and accountability are amongst the most frequently stated ideas behind the introduction of performance measurement. Lastly, we examine uses and effects of the respective systems (Sect. 3.3). In the German case, the use of performance measurement has often lead to the reinforcing of existing fragmentations, also due to Germany's federal political structure, which makes comparisons across states or between different prison establishments difficult. In contrast, in England and Wales there exists a higher degree of comparability, which, amongst other things, has been facilitated by the standardized prison ratings that have contributed to the enhancement of competition between prison establishments, but also certain unintended negative consequences, such as gaming, disengagement and bureaucratization.

We conclude by reflecting on two particular challenges of performance measurement in the context of prison management. First, we highlight how the choice of accounting entity underlying a performance measurement system (e.g. individual prison establishments, as in the case of England and Wales) can lead to undesirable consequences, such as the narrowing of accountability. Second, we discuss the risk that certain means (in our case individual performance measures) can shift attention away from broader ends (such as rehabilitation). We argue that investment in the building of analytical capacity and critical reflexivity, i.e. the development of a sensitivity to both uses and limits of performance measurement systems, are important ingredients of their successful implementation. Furthermore, performance measurement should be kept open to negotiation and debate.

Put differently, performance measurement systems should be thought of as a “burning lens” for the stimulation of introspection and reflection, rather than as a mechanism for the allocation of blame. Only then can they help to mediate between, and to further knowledge exchange and learning across, different, at times conflicting, prison values and objectives, such as those of security, rehabilitation, decency and economy.

---

## 2 Research Design and Method

This chapter is based on a multiple case study research design (Eisenhardt and Graebner, 2007; Yin, 2014). Amongst other things, we analysed a variety of different documents including publicly available prison data; governmental reports; inspection reports; reports from prison interest groups; as well as secondary literature. Further, we conducted 91 semi-structured interviews with representatives of prison management (top and middle management), prison officers, prison inspectors, regulators, criminologists, representatives of prison interest groups as well as employees of different Ministries of Justice. Between 2016 and 2020, we visited 10 prisons in three different states in Germany, and 6 prisons in England and Wales, conducting a total of 44 interviews in Germany and 47 interviews in England and Wales. Our interviews were based on a shared interview guide that, amongst other things, addressed questions about the measurement instruments in use, the relative importance of different performance indicators, effects of the performance measures on individual behaviours and organizational processes, and uses and effects in relation to broader regulatory objectives aimed at steering. In most cases, the interviews were conducted by at least two members of the research team, and recorded and transcribed. In addition, the German team participated in twelve day-long meetings held amongst prison governors and the Ministry of Justice in Land B in which they discussed the renewal of the current Balanced Scorecard in use.

The choice of cases and interviewees was driven by the goal of maximizing variation. For instance, in Germany, we chose to conduct our research in three different states (called Land A, B, and C for reasons of anonymity) in an attempt to represent Germany’s diversity: Land A is a large territorial state, with a dense population and many prisons; Land B is also a large territorial state, but situated in Eastern Germany (former GDR), with a small population and only a handful of prison facilities; Land C is a city-state and one of the most populated cities in Germany, and has a similar number of prisons as Land B. In England and Wales, we visited prisons of different sizes, security levels and management status



(publicly managed vs. contracted out). In the case of England and Wales, the state-mandated performance measures apply to both public and private prisons. In Germany, there are no fully private prisons. Out of the 179 prisons in Germany, only four are operated as private–public-partnerships, of which one was part of our sample.

In the following, we present our findings by, first, describing the different measurement systems in use (Sect. 3.1); second, examining the different rationales and objectives underlying them (Sect. 3.2); and, third, depicting different uses and effects (Sect. 3.3).

---

## **3 Performance Measurement in the Penal Justice Sectors of Germany and England and Wales**

### **3.1 Performance Measurement Systems in Use**

#### **Germany**

Owing to Germany's federal structure there is not one but sixteen different performance measurement systems in place in the penal justice sector (Iloga Balep and Huber, 2017). These systems differ in terms of the performance measurement instruments, individual indicators as well as the information technologies in use. Lower-Saxony, for example, uses an interconnected system (LoHN) that combines performance-oriented management control tools, such as quantified target and reporting systems, with budgeting as well as benchmarking information for the entire public sector (i.e. different public services). Other states use less integrated systems for each public service. Generally, a large part of the performance measurement instruments is made up of operational and financial indicators which are based on data concerning for instance: operational capacity measured by total number of prisoners within an establishment, number of prisoners in individual cells, number of prisoners in two-bed cells; staff capacity, measured, for example, by rates of absenteeism; financial measures, such as costs of prisoner per day, leasing costs, energy costs, as well as revenues made, for example through the selling of goods produced by prisoners. Such data are used by the Ministries of Justice, for example, to steer their state's prison estate and take long-term decisions (e.g. with regards to staffing or the building of new or closing of old prisons) and by prison governors for the day-to-day managing of their respective establishments.

While most performance measures relate to operational and financial indicators, there have been initiatives to introduce instruments that go beyond these parameters. For instance, many states have introduced a Balanced Scorecard (BSC) approach to prison performance measurement (see Iloga Balep in this volume) as part of a

'Neues Steuerungsmodell' (new governance model) built on the model of British New Public Management (NPM). Roughly half of the states' penitentiary systems use such instruments now, which allow them to oversee not only financial and operational information, but also data on security and resocialization. Indicators related to security, for instance, include the number of escapes, incidents of violence, and the amount of confiscated illegal drugs. Indicators related to resocialization refer often mainly to procedural information, such as the accomplishment of education and training programmes, overall behaviour, or work contracts and housing after release. Another common indicator in relation to resocialization is the percentage of sentence plans successfully updated every six months. The sentence plan is a written agreement between the prisoner and prison management (often middle managers with a psychological background) that sets out aims to be achieved during imprisonment, for example, with respect to completion of programmes, including drug rehabilitation, professional training and other educational programmes. Every six months, information about the progress made in relation to the set aims has to be noted within the sentence plan, together with potential new aims which are to be determined in cooperation with the prisoner concerned. Although the performance measurement systems are similar in their objectives, their operational design differs from state to state, for example with regards to the indicators that are used, set targets or underlying measurements. This applies also to the composition of the Balanced Scorecard (BSC).

At the level of individual prison establishments, prison governors, accountants and middle managers organize the collection of performance data. Prison governors are accountable to the state-based Ministries of Justice, particularly in relation to questions concerning budgetary planning and the allocation of resources, but also with regards to major incidents such as prisoner escapes. In day-to-day business, prison governors have the autonomy to decide on most aspects of operational work. Decisions at the administrative level of a prison establishment, concerning for example the management of personnel, are often based on performance indicators such as employees' sickness absence rates. As every activity and movement of prisoners within the prison establishments require accompanying staff, indicators on sickness absence are particularly critical for ensuring frictionless day-to-day operations. Street-level staff such as prison officers are involved in collecting and delivering data but otherwise are not in touch with the measurement instruments and rarely receive information about performance.

In Germany, criminological research institutes play a prominent role in developing performance measurement tools for the prison sector (see Wirth's introduction to this volume). Such institutes are, in contrast to the UK, not always attached to universities. They exist on the state or federal level and although they often work

on behalf of the Ministries of Justice, they also cooperate with individual prisons or universities. Criminological research institutes mainly undertake onetime evaluations and their studies tend to be focused on questions of quality and efficacy of the penal justice system. One major topic concerning prisons is how effective different treatment measures have been with regards to rehabilitation.

### **England and Wales**

In contrast to Germany, performance measurement in England and Wales is highly standardized and centralized, and it includes both public and private prison establishments. Since 2008, all national prison performance data is collected and reported on a rolling basis on p-NOMIS, the operational database used in the HM Prison Service of England and Wales for the management of offenders. In 2008, following the creation of a dedicated data science team in the Ministry of Justice and a drive towards digitization and modernization, a “Performance Hub” (originally NOMS Performance Hub) was created for the collection and reporting of Prisons and Probation Trusts data and management information.<sup>5</sup> Today, the prison Performance Hub (“the Hub”) provides prison governors, regulators and other stakeholders concerned with key performance data at individual prison as well as aggregate level in the form of dashboards and detailed reports. The Hub gives prison governors not only access to performance data concerning their own establishments, but also that of all other prisons in England and Wales; governors can benchmark and compare their prison’s performance with that of other prisons (e.g. prisons that belong to the same category of security) in relation to specific aspects of their choice with relative ease.

Primary responsibility for the monitoring and measuring of prison performance rests with Her Majesty’s Prison and Probation Service (HMPPS), an executive agency sponsored by the British Ministry of Justice.<sup>6</sup> In 2019/20, this agency measured and monitored prison level performance on the basis of what is called the Prison Performance Tool (PPT). As is stated on the relevant government website:

---

<sup>5</sup> See <https://www.justice.gov.uk/downloads/offenders/psipso/psi-2014/psi-28-2014-noms-performance-hub-data-quality-policy.pdf>, accessed 26 April 2021.

<sup>6</sup> The HMPPS works with the HM Prison Service and the National Probation Service. See <https://www.gov.uk/government/organisations/her-majestys-prison-and-probation-service>, accessed 26 April 2021. In England and Wales, executive agencies were established “following a recommendation made in 1988 to allow the delivery of executive functions of government to be carried out separately from—but within a policy and resources framework set by—a primarily policy-focused department” ([https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/690636/Executive\\_Agencies\\_Guidance.PDF](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/690636/Executive_Agencies_Guidance.PDF), p. 3, accessed 26 April 2021).

In the PPT, overall performance in each prison is rated on a 1 to 4 scale. The ratings are 4: Performance is exceptional; 3: Performance is acceptable; 2: Performance is of concern and 1: Performance is of serious concern.<sup>7</sup>

Such overall performance ratings are normally undertaken on an annual basis<sup>8</sup>, although prison governors have also access to more timely measurement information (e.g., with regards to incident reporting). In 2019/20, the PPT comprised 33 measures structured around six main categories which reflect HMPPS priorities: safety; security; rehabilitation and release planning; respect; purposeful activity; and organizational effectiveness.<sup>9</sup> Three new measures were introduced into the framework in 2019/20 in addition to those already used in the 2018/19 framework: accommodation on the first night of release; employment at six weeks following release; staff resignation rate.<sup>10</sup> Although the prison system is composed of establishments of various sizes and types, HMPPS is able to ensure commensurability through a mechanism of assigning a particular weight to particular measures in the overall scores. In the first place, each measure carries a weighting based on the policy priorities set out in the PPT framework. Then, specific types of prisons entail measures with different weightings, based on their function (e.g., male vs. female, short vs. long-term). The weighting is done in percentage points, with the 33 measures totalling 100 %. Passing a specific threshold results in a specific score (e.g., less than 82 % and greater than or equal to 61 % results in a score of 3).

The PPT is aimed at providing a balanced performance measurement framework across different dimensions of prison management (safety, economy, security, decency and rehabilitation). The data are drawn from various sources, including p-NOMIS, the Performance Hub, the HM Inspectorate of Prisons, and the Operational System and Assurance Group (which, amongst other things, runs audits, such as custodial operational assurance audits, and manages the Quality of Prisoners' Life (MQPL) survey results and reports. Thus, the actors involved in collecting and reporting data are varied. Normally, key performance indicators (KPIs) and targets

---

<sup>7</sup> See <https://data.justice.gov.uk/prisons/additional/performance-band>, accessed 26 April 2021.

<sup>8</sup> See <https://www.gov.uk/government/statistics/prison-performance-ratings-2019-to-2020>, accessed 26 April 2021.

<sup>9</sup> See [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/913004/annual-prison-performance-ratings-2019-20-bulletin.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/913004/annual-prison-performance-ratings-2019-20-bulletin.pdf), p. 2, accessed 26 April 2021.

<sup>10</sup> See [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/913004/annual-prison-performance-ratings-2019-20-bulletin.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/913004/annual-prison-performance-ratings-2019-20-bulletin.pdf), *ibid.*, accessed 26 April 2021.

are set by HMPPS. Neither prison governors nor prison officers participate in their development and elaboration. However, from time to time, HMPPS is assisted in the development of new measures by university-based criminological research centres, such as the Cambridge Institute for Criminology, which developed the globally unique Measurement of the Quality of Prison Life (MQPL) based on a survey with prisoners that seeks to capture the ‘moral performance’ of individual prison establishments (Liebling, 2004). This survey was launched in 2001 and has undergone several rounds of revisions and refinement since, consisting now of over 120 statements that offenders are asked to agree or disagree with on a scale from 1 to 5. These deal with questions of, for instance, humanity, relationships, fairness and respect in prisons, and are meant to score prisons according to their performance in these qualitative dimensions.<sup>11</sup> The MQPL is part of the PPT.

Responsible for delivering performance against the PPT are the individual prison governors, who are held accountable by the regulator (HMPPS) and the HM Inspectorate of Prisons, which conducts announced and unannounced inspections. Prison officers, on the other hand, are normally not directly accountable, except for specific circumstances where they are responsible for an individual item or group of measures, for which a prison governor might hold them accountable (e.g., time out of cell measures, or MQPL).

There are no immediate financial penalties or rewards related to performing well or badly against the PPT. Yet, achieving a good or bad overall rating has reputational consequences, and prison governors may be removed from a badly performing establishment, or their establishment can be put under special measures of regulatory scrutiny and intervention. Until a few years ago, the prison ratings were an important component in the context of “market testing”, as they permitted the identification of “failing prisons” whose management would then be put out to tender. Market testing permitted the private sector to compete directly with the public sector for the management of prisons that were considered to be “failing”, that were not meeting performance targets, for instance with respect to cost management or security standards, evidenced for example by prisoner escapes or riots (Black, 1993; Prison Reform Trust, 1994).

The current system is the result of more than three decades of continuous measurement reform and development. The first set of standardized performance metrics was introduced into the Prison Service in 1992/93 (Guter-Sandu and Mennicken, [forthcoming](#)). These metrics included, amongst other things: number of prisoner

---

<sup>11</sup> See <https://www.prc.crim.cam.ac.uk/directory/research-themes/mqpl>, accessed 26 April 2021.

escapes, number of assaults (on staff, prisoners and others), hours spend in purposeful activity, proportion of prisoners held in unit of accommodation intended for fewer numbers, and information about the average cost per prisoner place (Prison Reform Trust, 1996). In subsequent years, they were frequently revised and in 2003, for the first time, composite prison performance ratings were introduced (Guter-Sandu and Mennicken, [forthcoming](#)). Although composite prison ratings have been in place since then, the rating system itself (e.g. the weightings assigned) and the measures and targets underlying it were changed. Before the introduction of the PPT in 2018/19, for instance, annual prison ratings were based on the so-called Custodial Performance Tool (2017/18) and before that on the Prison Rating System (2009/10–2016/17). Although all these performance measurement frameworks are predicated on similar aims and objectives (e.g., to measure a prison’s performance on different dimensions: security, safety, rehabilitation, decency, economy and efficiency), at least some of the measures they include vary (e.g., in their calculation and composition) and, as mentioned already above, in 2019/20 three new measures were added (accommodation on the first night of release; employment at six weeks following release; staff resignation rate).

## 3.2 Rationales and Objectives

### Germany

We identified three main objectives underlying the design of performance measures in the German penal system: 1) to aid budgetary planning, 2) to enhance comparability, and 3) to create scientific knowledge.

*Facilitating budgetary planning.* In the wake of the German version of NPM reforms, the ‘Neues Steuerungsmodell’ (new governance model) was introduced, and state-level administrations moved from input control to output control, and a system of budgetary planning. In the context of these reforms, several states introduced a system of agreed targets (‘Zielvereinbarungen’) in the early 2000s. Despite referring to this system as a system of “management by objectives”, the set targets are not so much used for incentivizing or punishing prison management but more for informational purposes. A policy document from Lower-Saxony (Niedersächsisches Finanzministerium, 2005) gives a list of purposes ranging from operationalizing legal obligations, improving communication between ministries and prisons, to constant improvement. Although the autonomy of prisons has increased compared to the earlier systems of public administration, ministries still operate with detailed non-transferable budgets. The major advancement was that prison governors are now involved in negotiating these budgets once per year. For example, since 2006

in the state of Lower-Saxony, the Ministry of Justice and the prison governors agree upon targets on a yearly basis in accordance with the stated goals of a Balanced Scorecard, and they assign specific budgets to these targets (Niedersächsischer Landtag, 2010). Similar performance management systems are in place in other states as well, albeit not always with a Balanced Scorecard in use.

Apart from such changes based on mega-trends like NPM, the German penal sector has received relatively little attention and few issues have been problematized publicly. In the absence of public debates, scandals or financial problems, lawmakers have invested much less in introducing management systems into the penal sector compared to other public sectors such as healthcare or education.

*Enhancing comparability.* In 2006, a ruling of the Federal Constitutional Court concerning juvenile detentions demanded more benchmarking efforts in the prison system to ensure a high quality across different prison establishments and states (Iloga Balep and Huber, 2017). This ruling was interpreted by many practitioners in the field as a legal obligation to engage in benchmarking efforts, particularly in the context of juvenile detention (Bolay and Volz, 2013). As a consequence, several states included either mandatory or optional evaluations on the effectiveness of (juvenile) detentions into their penal state laws. These were supposed to allow for comparability with regard to success or failure of the penitentiary system, in particular with regards to recidivism (Bolay and Volz, 2013). In 2009, lawmakers added paragraph 91d to the German constitution which generally enabled and encouraged benchmarking efforts amongst the states' administrative bodies. This paragraph explicitly refers to benchmarking as an essential element of Anglo-Saxon administrative culture that had proven internationally to be an effective instrument of governance. Benchmarking in the constitution is referred to as a form of competition bringing about a continuous process of improvement within public administrations and strengthening parliamentary control (Deutscher Bundestag, 2006). This addition to the constitution affirmed that benchmarking within and between states was not only legally permitted but also politically desired at federal level. Amongst other things, lawmakers argued that benchmarking was a way for state-level administrations to learn from each another.

*Creating insight.* Another objective of performance measurement in prisons is the creation of scientific knowledge by criminological research institutes. Regularly, the states' local criminological research institutes (CRIs) produce scientific evaluations of prisons based on their own research agendas or ministries' ad-hoc concerns. The main purpose of these evaluations is to produce knowledge for the scientific communities in criminology and psychology. Yet, although the CRIs' research often concerns questions of performance, performance measurement in this context is not explicitly and primarily intended for everyday management, neither by the ministries

nor the individual prisons. Reasons for this lie in the relative autonomy of the CRIs and their focus on psychological and criminological expert discourses. However, ministries have in the past referred to selected results in their long-term policy-making.

### **England and Wales**

Given that performance measurement is so pervasive in the prison service of England and Wales, it comes as no surprise that it is one of the main tools for political control and accountability. Prisons that perform poorly can be easily identified through the prison ratings and be required to provide explanations and routes to improvement. Also ministers have an incentive to keep an eye on the prison ratings, as these can become the source of media attention, with readily available data being scrutinized and publicized by journalists, thus acquiring political salience. In this context, specific indicators, like overcrowding or self-harm, have become particularly resonant with the wider public.<sup>12</sup> At least three different sets of rationales and objectives can be attributed to the increasing use of performance measures in the prison service of England and Wales: 1) to enhance surveillance and control, 2) to facilitate comparison and competition amongst prison establishments, and 3) to facilitate the balancing of different, often conflicting, prison values and objectives.

*Enhancing surveillance and control.* It was, and still is, believed that performance measures augment the capacity for surveillance and control, which is a source of both rejoice and remorse, depending on the proclivity of the actors asked. Indeed, an often-quoted complaint of regulators is that before the introduction of quantified performance indicators, prisons were obscure entities inaccessible to ministerial oversight which allowed local prison governors to reign free and with impunity (Lewis, 1997). Some regarded the lack of indicators as one of the main reasons for which a number of high-profile scandals in the late 1980s and early 1990s were not foreseen, whereas this might have been forestalled had there been in place a system of performance indicators. In this context, it is also believed that the indicators are an important instrument to assure and enhance governmental control as an individual prison's performance is made visible and comparable through the ratings.

*Marketization.* Underlying the introduction of the prison ratings was also a belief in the power of market incentives (Carter, 2003; Guter-Sandu and Mennicken, forthcoming). With the help of the ratings poorly performing prisons could be easily and publicly identified and, at least in previous years, threatened to lose their operating

---

<sup>12</sup> See, for instance, frequent reporting on these measures in The Guardian, such as <https://www.theguardian.com/society/2020/jan/30/number-of-self-harm-incidents-in-prisons-reaches-record-high>, accessed 26 April 2021.



licence and be put up for market testing, with both the public and private sectors invited to place bids for the running of the “failing” establishment. The prison ratings are aimed at facilitating comparisons between public and private sector prison performance and, thereby, they help render ideas about competition and competitiveness operable. They redefine the prison as a separate “accounting entity” (Kurunmäki, 1999), a calculating, independent, performance-oriented unit, responsible for its own success and failure (Guter-Sandu and Mennicken, *forthcoming*; Mennicken, 2013). In so doing, the measures are also said to stimulate mutual learning. Even though the prison system is comprised of very different types of establishments, the Performance Hub allows prison governors to compare their own prison with other, similar ones, and prison governors are encouraged to use technological facilities like the Performance Hub to identify their standing in the panoply of different measures, and to reach out to better-rated prisons for the purpose of learning.

*Value balancing.* Finally, the measures can also be said to have had a “democratizing” ambition, namely to hold managers, public administrators and civil servants to account and to counteract nepotism and arbitrariness (Guter-Sandu and Mennicken, *forthcoming*). Furthermore, the performance measures are also perceived as an important mediating instrument (Miller and O’Leary, 2007), as a mechanism that can be utilized to link up and mediate between conflicting concerns and prison values, such as those of security, economy, and decency (Mennicken, 2014). Measuring the Quality of Prison Life (MQPL), for instance, brought prison values relating to questions of rehabilitation, care and decency back in, and to give prisoners “a voice” through the introduction of a standardized survey aimed at capturing their day-to-day experiences (Liebling and Arnold, 2004).

### 3.3 Uses and Effects

#### Germany

We found that the ways in which performance measurements were used produced three intertwined yet partly opposing effects: 1) rather than facilitating comparability, measurements tended to reinforce already existing fragmentations due to their diversity; 2) prison governors in Germany feel demotivated by the performance measures, especially when being punished for good performance; and 3) we observed that the incompleteness of measures and associated scepticism and critique lead to discussions and initiatives to make specific topics considered neglected by existing measures, for example topics related to resocialization, more visible.

*Reinforced fragmentation.* Even though the legal initiative on benchmarking in the penal sector mentioned in Sect. 3.2 indicates an interest in comparison at the federal level, such comparisons are not put into practice, at least not systematically. This is mainly due to the principle of subsidiarity typical for German public administration. An example for this is the reform of federalism in 2006 that allowed each of the sixteen German states to pass their own laws regarding penal administration. Practitioners and experts of the penal sector strongly disapproved of this change at the time, expecting major differences and thus injustices between the states to increase. Further, they anticipated in response to the benchmarking initiative at federal level a rise in comparisons and benchmarking amongst the states, which would lead to a race to the bottom regarding costs and cuts especially to programmes of resocialization (Iloga Balep and Huber, 2017). To the surprise of those experts, the feared “competition of shabbiness” (Düinkel and Schüler-Springorum, 2006) did not take place. Systematic comparisons between the states were not undertaken, apart from the annual summary statistics published by the German Federal Statistical Office (e.g. about the number and type of prisoners in different states) which have remained unchanged since the 1960s. The few comparisons that do exist are usually conducted by criminological research institutes either on behalf of Ministries of Justice or autonomously, and they are always research project-based (Bolay and Volz, 2009). Besides a general fear and scepticism of being compared in the states’ ministries, reasons for the lack of comparisons are also of a structural and organizational nature: Germany’s federal structure means that different laws, administrative processes, IT-Systems, and indicators are in use in the different states. Such a set-up and the historical use of different measures decreases possibilities for comparability and standardization across states. The lack of such comparability isolates the states’ penal systems and strengthens their autonomy, even if only unintentionally. Fragmentation of measures and an increasing number of different measures blur commonalities and increase the bureaucratic effort necessary to compare.

While there exist no systematic comparisons of states besides the federal statistics mentioned above, single indicators are quite frequently used and compared for political reasons. But identical labels can give a false impression of standardization, since each system has its own definition that underlies these indicators. Political actors use such incomparable indicators for their individual agendas. For example, ‘cost of prisoner per day’ is a politically highly relevant indicator that differs remarkably in its composition in the different states: while in one state detention facilities are rather old and the relatively high maintenance costs are included in the calculation of ‘cost of prisoner per day’, another state operates mostly new facilities and construction costs are not included in their cost indicator (Iloga Balep and

Huber, 2017; Meyer, 2003). Politically, low expenditure can be used to prove a well-functioning, cost-efficient prison system or, depending on the political agenda, high expenditure can stand for higher quality and better services. However, no matter the argument, the use of such indicators for comparative reasons can be misleading and we found no attempts aimed at harmonizing measurement here (Meyer, 2003).

*Demotivation.* On the state level, Ministries of Justice have an interest in comparing individual prisons' performances, for example, with regards to costs or incidences of violence or drug abuse amongst prisoners. These comparisons, however, are not systematically attached to a reward and punishment scheme like a bonus-malus system. Such comparisons allow the ministries to stay informed at one glance and to stay close to operational business, at least through the numbers. The general lack of incentives and the wider absence of marketization and privatization in the German prison sector goes along with the absence of competition amongst prisons. Often, the Ministries of Justice are not interested in the performance of single establishments but rather treat all prisons within a state as one entity. In one of the states, for example, the Ministry of Justice did not find it important to differentiate between individual prison establishments, for example with regard to how much effort individual prisons put into communication with journalists and media (one indicator of their BSC). Information was aggregated into one number for all establishments. In another case, the Ministry of Justice did compare the financial performances of individual prisons, however, chose not to reward good performance. Instead, the prisons which did well lost their excess budget to the establishments that had financial shortcomings. As a consequence, some prison governors reported losing motivation to do particularly well since either no differentiation amongst the establishments was made by the Ministries of Justice, or good performance led to a reduced budget in the next year. Other prison governors, though, felt pressure to perform well in particular out of fear of humiliation in direct comparisons with other prison establishments. Nevertheless, due to the lack of a systematic bonus-malus system, adverse effects and gaming are very limited. Instead, dissatisfaction with existing measures within the prison establishments, from administrative to street level, has led to local initiatives aimed at creating more customized and individualized forms of measurement.

*Enhanced discourse.* Another effect of measurement in the German penal sector is that it leads to additional discourses and initiatives amongst different actors, since the Ministries of Justice aim to enhance not only their own steering capacity through the measures, but also learning and cooperation amongst different prison establishments or other parties concerned. On the one hand, we observed general scepticism towards performance measurement, benchmarking and comparisons within prisons. Many of our interviewees criticized the time-consuming nature of

performance data collection, the lack of communication by the ministries regarding measurement outcomes, as well as the choice of indicators, which they often felt did not measure what really mattered in their work. On the other hand, we noted that especially the latter criticism led many middle managers, prison governors and accountants to use and develop their own measurement instruments to capture what was of importance according to them. These instruments were usually additional excel-sheets or handwritten notes for personal use only. In some cases, however, such instruments travelled also upstream and started being used in the entire department, whole establishment, and in a few cases even in all prisons of one state and beyond.

Further, scepticism towards the measurements, or adverse behavioural effects, can be buffered through the close cooperation that exists between the Ministries of Justice and prison governors not only in the development of the indicators, but also their use and evaluation. During regular meetings (up to four per year) among the Ministries of Justice and prison governors, performance results of individual prison establishments are openly communicated and discussed. While some of our interviewees showed some fear of having their prisons highlighted as underperformers in comparison to other establishments, other prison governors perceived these comparisons as an opportunity to learn more about how to most effectively achieve certain targets. Also the choice of indicators was discussed in-depth at least on the regulatory level within the Ministries of Justice. Often prison governors were already participating in the process of their development (Iloga Balep and Huber, [forthcoming](#)). Also, the evaluation of performance in accordance with certain indicators, even when linked to a traffic light system, did not automatically trigger preconfigured responses. Instead, such assessments of performance opened the floor for explanations, refinements and further discussions amongst prison administrations and Ministries of Justice. While the Ministries base decisions about resources on numbers and have the final say, they also leave space for finding common ground in exchange with prison governors, thereby triggering processes of reflexivity, mutual exchange and learning.

Generally, measures related to resocialization are still more difficult to integrate next to security measures or financial indicators. Performance measurement tools developed by criminological research institutes aim at making visible specific issues that according to criminologists tend to receive scant attention. These topics often concern measures associated with resocialization, such as a tool aiming at gauging the effectiveness of detention by comparing the prisoners' attitudes, degrees, training, drug abuse, etc., at the beginning and at the end of detention, measuring them on a scale of 1–5 (Suhling et al., [2015](#)). Mainly due to bureaucratic and budgetary hurdles not all the performance instruments developed by criminological research

institutes are in the end included in the prisons' performance measurement systems. Still, criminological research institutes play a major role in instilling scientific discussions into the field, be it through direct involvement with the Ministries of Justice or the prison governors or by enhancing practitioner discourses in conferences and through publications.

### **England and Wales**

In England and Wales, responding to the demands for standardized performance information coming from HMPPS and government is an accepted and established practice at individual prison level. However, the extent to which such data travel within establishments is varied. For instance, in some establishments, even though data are collated, assured, and fed upwards, they do not seem to expand the capacity of local management of prisons or to change established practices. In other establishments, data and performance management tools may be enthusiastically embraced by particular prison administrators and included in, or built upon, towards a wider strategy for prison management. In the following, we draw particular attention to three types of effects we observed: 1) bureaucratization; 2) value hierarchization; 3) entity biases.

*Bureaucratization.* Although the rise of prison performance measurement in England and Wales was largely animated by market-oriented reform ideals, the introduction of standardized performance measures gave also rise to the creation of an unwieldy bureaucratic apparatus and new information systems that need to be managed and fed (Guter-Sandu and Mennicken, [forthcoming](#)). Prison governors have to meet increasingly detailed reporting demands, facing “constant oversight from internal auditors and external inspection bodies” (Coyle, [2005](#), p. 97; but see also (Bennett, [2016](#)). Such reporting demands can take them away from “walking the floor” and face-to-face interactions with prisoners and prison officers. As ex-prison governor Coyle ([2005](#), pp. 49) puts it, the introduction of formalized performance measurement led to “a concentration on process, on how things are done, rather than on outcome, that is, what is being achieved” (see also Bennett, [2016](#), [2019](#)). One of our interviewees, a prison governor of a high-security prison at the time of interview, highlighted the danger of producing a “virtual organization” that is de-coupled from day-to-day prison life. Such a de-coupled, virtual organization is sustained by fantasies of controllability which are nurtured by the fleet of performance measures, dashboards, ratings, etc. Yet, such fantasies of controllability may not match with what can actually be controlled. Many determinants of underperformance, such as overcrowding, budget cuts, a prison's age and building structure, are out of the control of management. This in turn can lead to demotivation and disengagement, or

other undesirable behavioural consequences, such as gaming or misreporting.<sup>13</sup> Further, it is difficult to attribute re-offending rates to individual prison performance, as prisoners are regularly transferred between prisons, transcending prison boundaries and the accounting for them (Bastow, 2013; Mennicken, 2013).

It is also important to remember that many of the performance measures underlying the prison ratings do not provide a “real-time” window into a prison establishment. Many of the measures are lagged and the reporting frequency hugely varies across measures. Some measures are updated on a monthly basis, others, such as the MQPL only every three years. If measures are not frequently updated, they can quickly become regarded as outdated and no longer relevant. Also comparability of performance across years is quite difficult, due to the frequent revisions that have been undertaken to the performance measure framework and the measures themselves. This undermines the consistency of the measures across years. Even though the PPT was introduced in the 2018–19 rating exercise, it is not possible to compare these composite ratings with the ratings from 2019–20 due to the changes that have been made to the PPT in 2019/20.

All this puts ideals of comparability and inter-organizational competitiveness into question. At the same time, the majority of the prison governors we interviewed highlighted that they feel still spurred by the ratings, that they want to ensure that their rating does not fall into the 1 or 2 category (although often that is unavoidable, e.g., when major incidents such as prisoner escapes or riots happen). Yet, how much time and energy is invested in analysing and making sense of the measures varies. Prison staff receive only limited training from the centre (i.e. HMPPS or the Ministry of Justice) in the use of these and other management information tools (such as the Performance Hub). It is down to each individual prison to decide how much time and energy it wants to invest in the building of analytical capacity. And often, external constraints, such as staffing shortages, budget cuts, and day-to-day operational demands make it nearly impossible to build such spaces for learning and reflexivity (e.g., about what worked well and why and what worked not so well and why).

*Value hierarchization.* As already highlighted above, the prison service’s performance measurement system seeks to balance, and mediate between, often conflicting, values and objectives of prison management. The plethora of performance measures is entered into a weighted scorecard, drawn up in the fashion of

---

<sup>13</sup> However, it should also be noted that internal and external audit functions have been built and extended over the years to assure the quality of the data underlying the performance measurements, and the Ministry of Justice’s data science team rated that quality of the data as generally high.

Kaplan and Norton's Balanced Scorecard (Kaplan and Norton, 1992).<sup>14</sup> Yet, despite these attempts aimed at "value balancing", the different performance measures, and the values they represent, do not exist on an equal plane. In the day-to-day management of a prison, they can easily come to be hierarchized (Mennicken, 2013). We observed, for instance, that issues of security and cost tend to be prioritized over measures of decency and the quality of prison life (Bennett, 2019; Lieblich and Crewe, 2013). The infrequent conduct of the Measuring the Quality of Prison Life (MQPL) survey (every three years) supports such hierarchization, as these measures get easily "out of sight" given that they can become quickly seen as outdated, due to the fast moving prison population that feeds into them. Although all our interviewees highlighted the importance and usefulness of the MQPL surveys, we also heard concerns regarding their validity given that the surveys are capturing the views and experiences of "a moving target" – those of individual prisoners with different histories at one specific point in time; at another point in time, such views and experiences could very well be very different, given another set of prisoners with different individual histories and needs. Put differently, questions were articulated about the extent to which the surveys can capture a prison's broader organizational culture versus specific, individual experiences which can hugely vary and are not necessarily attributable to the organizational culture of a particular establishment.

Furthermore, the power of the MQPL initiative was undermined by the government's austerity politics, where the Prison and Probation Service in 2012/13 alone had to make savings of £246 million on top of the £228 million savings delivered in 2011/12 (HM Chief Inspector of Prisons for England and Wales, 2013, p. 7). As a result, concerns with cost and "economics of scale" came to the fore, and definitions of failure were narrowed to definitions of failure in economic (i.e. cost management) terms (Guter-Sandu and Mennicken, forthcoming). In response to these budget cuts, the Prison Service reduced its costs through a changed estate management strategy, which included not only land sales, but also the closing of smaller, older prisons, which are costlier to run than large establishments with more than 1500 places, but not necessarily of worse quality. For instance, of the 18 prisons closed or identified for closure by December 2013, eight were considered "high performers" according to the MQPL survey (ibid.).

*Entity biases.* The prison service's performance measurement system in England and Wales places emphasis on making individual entities – individual prison establishments – comparable and accountable. Performance is adjudicated at the level of individual prison organizations, rather than the system as a whole. Since

---

<sup>14</sup> See also [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/218344/prison-rating-system-technical-note.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/218344/prison-rating-system-technical-note.pdf), accessed 28 April 2021.

the late 1980s, the Prison Service of England and Wales has undertaken a series of steps to transform prisons into competitive, performance-oriented “accounting entities” (Mennicken, 2013). Yet, not all dimensions of prison performance can be measured at the level of the individual prison organization. For example, because of overcrowding, prisoners are often transferred between prisons, which makes it very hard to assess the effect a particular establishment has (e.g. with regards to the programmes it offers) on an individual prisoner’s rehabilitation prospects. This applies also to the case of re-offending rates more generally, which are difficult to attribute to an individual prison’s performance.

Furthermore, a preoccupation with organizational entity-based performance measures can undermine inter-organizational cooperation and coordination between prisons. The current performance measurement system does not recognize mutual aid and support that prison entities provide to each other, for example, when helping each other out in cases of staffing shortages, or in the form of collaborations aimed at mutual learning. Although we found ample evidence of such mutual support, it is not officially accounted for and hence difficult to explicitly encourage via the existing performance measurement system. Finally, organizationally based performance measures can undermine policies aimed at the system-level, for example, policies seeking to better integrate the prison and probation services in England and Wales. At least until 2019/20, the performance measurement system did not account for what happened to offenders once they left prison, i.e. the interface between prison and probation management was left unaccounted for. In 2019/20 an attempt was made to change this with the introduction of two new performance measures: accommodation on the first night of release; and employment at six weeks following release. These measures lie outside the normal remit of prison management, yet, individual prison establishments have been made responsible for them in order to encourage more collaboration with the probation service. It remains to be seen how successful these measures will be in encouraging more intra-system collaboration and integration. Some of our interviewees articulated already concerns over the lack of their controllability. In sum, individual prison entity-based performance measurement can detract attention away from system-wide issues of offender management, including the management of relations between prison and probation, and the reintegration of prisoners into society.



## 4 Conclusion

In both settings that we investigated, in England and Wales as well as in Germany, performance measures have increasingly come to matter in the management of prison establishments. Yet, whereas in England and Wales performance measurement has been an established part of prison management since the early 1990s, in Germany, the introduction of more elaborate prison performance measures, following the model of the Balanced Scorecard, is relatively recent. In England and Wales, the performance measurement system is highly centralized and standardized, whereas in Germany, partly because of its federal political structure, performance measurement is decentralized and varied. Such variation makes not only cross-state comparisons but also comparisons between different prison establishments difficult, which in turn undermines the very ideas of benchmarking and “best practice”. Furthermore, in England and Wales information about an individual prison’s performance is readily available and publicly accessible. In Germany, it is more difficult to publicly access such information, as reporting formats differ and publication channels vary. Yet, although prison performance measurement in England and Wales is highly standardized and comparability amongst individual prison established, we also observed elements of fragmentation.

Although the prison performance measurement system that exists in England and Wales is much more elaborate than its German counterpart, it is important not to lose sight of implementation challenges and potentially negative unintended consequences. Attention to such challenges is important to develop a better understanding of uses and limits of existing performance measurement systems, which in turn will be useful for policy makers, regulators and public administrators when undertaking reforms.

First, when dealing with questions of prison performance measurement design it is important to reflect on what the underlying “accounting entity” is (e.g. an individual prison organization or the entire system of offender management). Questions of what entity to account for, and where one accounting entity ends and another begins, are not only consequential for definitions of risk and responsibility but also ethical commitments and organization values. A performance measurement system that focuses solely on the performance and accountability of individual prison establishments runs the risk of narrowing views to what is going on inside a prison, rather than what happens before and after prison. It can make it difficult to evaluate the prison service as a whole, for example with regards to its role in the broader context of society, or in relation to assessments of alternatives to prison or broader criminal justice issues. Furthermore,

one should not forget how difficult it is to actually draw boundaries around the performance of individual prison entities. As we highlighted above, the reduction of re-offending, for example, can often not be measured at the level of individual prison establishments, because of the frequent movement of prisoners in between prisons. Also inter-organizational activities, for example, in the form of information exchanges and mutual aid, often remain unaccounted for. All this can contribute to a further undermining of systemic accountability. On the other hand, defining the accounting entity of a performance measurement system too broadly can make it difficult to provide meaningful incentives for individual prison establishments (see here our findings for the German case). Taking a birds-eye-view on all prison establishments in a region can lead to the transfer of funds from well-performing establishments to poorly-performing ones. While this may be appealing to regional administrators, it eliminates incentives for local prison management to perform well with regard to the measures.

Second, performance measurement often suffers from a bias towards the administratively actionable. Also in the cases studied here, means (e.g. process measures) became at least in some instances more central than ends (e.g. rehabilitation). We also found that there often exists a lack of investment in analytical capacity. Although resources are spent on data and information management (especially in the case of England and Wales); investment in collective understanding and sensemaking, reflexivity and learning is often lacking. It is important that sufficient room is made for training, exchange and discussion. It is also important that such collective sensemaking happens in a “safe space” where performance measurement is not so much used to hold to account and to punish and reward, but for learning and improvement. One should not be too quick in dismissing the potential that benchmarking and performance measurement can have for animating and focusing debate (see here also our findings for the German case). The obvious limitations of performance metrics and ratings—their proneness to failure, misrepresentation and narrowing—can paradoxically function as an important platform for debate about prison values and reform. When different actors engage with performance measures and their effects in a critical and reflexive manner, this can lead to debates which go beyond explaining why certain targets or objectives were not met. In such a case, the performance measures can come to play the role of an important catalyst that draws together different parties and views, including disparate prison values and objectives, such as those of security, rehabilitation, decency and economy opening them up for scrutiny, negotiation and recalibration. The importance of such debates should not be underestimated, as it is here where important foundations for learning and reflexivity are laid and change can be initiated. More generally, this chapter has

shown that it is important to attend to the different modalities and operations of prison performance measures, and their ability or inability to reform practices and redefine possibilities for action.

## Appendix

**Tab. A.1** Performance measurement in the penal justice sectors in Germany and England and Wales

GERMANY	ENGLAND and WALES
<b>Measurement systems</b>	
<ul style="list-style-type: none"> <li>• Sixteen different states amount to sixteen different measurement systems in technology and content which result in different performance-oriented administrative quantification tools (BSC, budgeting, target and reporting systems)</li> <li>• Strong focus on measuring financial and operational performance, but instruments such as the BSC increasingly include indicators on security and resocialization</li> </ul>	<ul style="list-style-type: none"> <li>• Share a prison performance database (the Prison Performance Hub); measurement data are standardized; all have access to the same management information systems</li> <li>• Next to the national performance measures, some prisons might construct their own informal statistics for internal management use</li> <li>• Recently, outcome-based measures have been introduced that seek to account for life after prison (focusing on employment, accommodation)</li> </ul>
<b>Rationales and objectives</b>	
<ol style="list-style-type: none"> <li>1) Facilitating budgetary planning: Ministries of Justice and prison governors agree upon targets and specific budgets are assigned to these</li> <li>2) Enhancing comparability: benchmarking amongst states is supported by addition of paragraph 91d to the German constitution in 2009 encouraging states to compete over quality and to use learning effects. This applies, in principle, also to prison performance measurement</li> </ol>	<ol style="list-style-type: none"> <li>1) Enhancing surveillance and control: quantified performance indicators were introduced to allow for ministerial oversight and control</li> <li>2) Marketization: measurements and ratings are aimed at facilitating comparisons between public and private sector prison performance and, thereby, help render ideas about competition and competitiveness operable</li> </ol>

(Fortsetzung)

**Tab. A.1** (Fortsetzung)

GERMANY	ENGLAND and WALES
<p>3) Creating insight: evaluations are often produced by criminological research institutes, partly working on behalf of the ministries and mainly focusing on criminological and psychological objectives. Ministries of Justice partly ground long-term policy-making on such evaluations</p>	<p>3) Value balancing: performance measures are perceived as democratizing (all are held to account in the same way) as well as mediating instrument that can be used to (re)assess and mediate between conflicting concerns and prison values, such as those of security, economy and decency and rehabilitation</p>
<b>Uses and effects</b>	
<p>1) Reinforced fragmentation Diversity of measurements fortifies already existing fragmentation amongst the states making comparisons even more difficult</p> <p>2) Demotivation: Prison governors partly feel demotivated by measurement systems because of a lack of incentives</p> <p>3) Enhanced discourse: Skepticism and critique of existing measurement systems can lead to discussions and initiatives amongst the different actors to make specific topics considered neglected by existing measures more visible</p>	<p>1) Bureaucratization: standardized performance measures lead to increased bureaucratization and a decoupling between day-to-day prison life and management by target, creating a somewhat “virtual organization” at the level of management next to the lived reality of prison life “on the ground”</p> <p>2) Value hierarchization: one of the greatest challenges of performance measurement continues to exist in the balancing of different measures and values. Despite attempts aimed at value balancing, attention is often skewed towards measures concerning issues of security or economy. Such skewed attention is supported by the frequency with which different dimensions of performance are accounted for (real-time, quarterly, annually, every three years) and the broader context (e.g. governmental austerity measures)</p>
	<p>3) Entity biases: emphasis on individual prison performance makes it difficult to account for system-wide performance or dimensions that are difficult to attribute to one establishment alone, such as reduction of reoffending</p>

## References

- Bastow, S. (2013). *Governance, performance, and capacity stress*. Palgrave Macmillan.
- Bennett, J. (2016). *The working lives of prison managers*. Palgrave Macmillan.
- Bennett, J. (2019). Reform, resistance and managerial clawback: the evolution of 'Reform Prisons' in England. *The Howard Journal of Crime and Justice*, 58(1), 45–64. <https://doi.org/10.1111/hojo.12290>.
- Bolay, F. W., Volz, J. (2009). Wirkungsorientierte Verwaltungssteuerung als Modernisierungskonzept des Justizvollzugs? *VM*, 15(2), 57–65.
- Bolay, F. W., Volz, J. (2013). Länderübergreifendes Benchmarking im Justizvollzug. *VM*, 19(2), 90–97.
- Carter, P. (2003). *Managing offenders. Reducing crime: a new approach*. <https://doi.org/10.1002/9780470713532.ch1>.
- Coyle, A. (2005). *Understanding prisons: key issues in policy and practice. Crime and justice*. Open Univ. Press.
- Deutscher Bundestag (Ed.). *Gesetzentwurf der Fraktionen der CDU/CSU und SPD: Drucksache 16/12410 (2006)*.
- Drenkhahn, K. (2018). Entwicklung und Prognose der Gefangenenpopulation und ihrer Merkmale. In B. Maelicke & S. Suhling (Hrsg.), *Edition Forschung und Entwicklung in der Strafrechtspflege. Das Gefängnis auf dem Prüfstand: Zustand und Zukunft des Strafvollzugs* (S. 51–72). Springer.
- Dünkel, F., & Schüler-Springorum, H. (2006). Strafvollzug als Ländersache? Der „Wettbewerb der Schädigkeit“ ist schon im Gange! *Zeitschrift Für Strafvollzug Und Straffälligenhilfe (ZfStrVo)*, 55(3), 145–149.
- Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50(1), 25–32.
- Guter-Sandu, A., & Mennicken, A. (forthcoming). Quantification = Economization? Numbers, ratings and rankings in the prison service of England and Wales. In A. Mennicken & R. Salais (Eds.), *The new politics of numbers: Utopia, evidence and democracy*. Palgrave Macmillan.
- HM Chief Inspector of Prisons for England and Wales. (2013). *Annual Report 2012–13*. TSO.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(Spring), 3–19.
- Iloga Balep, N., & Huber, C. (forthcoming). Werte vollziehen: Verhandeln über Ziele und Kennzahlen im Controlling des Strafvollzugs. In F. Meier & T. Peetz (Eds.), *Organisation und Bewertung* (S. 321–343). Springer VS.
- Iloga Balep, N., & Huber, C. (2017). Quantification times 16: When decentralization stands in the way of markets. *Risk & Regulation*, 34(Winter), 16–19.
- James, O., & Hood, C. (2004). Prisons: Varying oversight and mutuality, much tinkering, limited control: Overview. In C. Hood (Hrsg.), *Controlling modern government: Variety, commonality and change* (S. 25–37). Elgar.
- Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard: Measures that drive performance. *Harvard Business Review*, (January–February), 71–79.
- Kurunmäki, L. (1999). Making an accounting entity: The case of the hospital in Finnish health care reforms. *European Accounting Review*, 8(2), 219–237.

- Lewis, D. (1997). *Hidden agendas: Politics, law, and disorder*. Hamish Hamilton; Penguin.
- Liebling, A., & Arnold, H. (2004). *Prisons and their moral performance: A study of values, quality, and prison life* (Bd. 45). Oxford University Press.
- Liebling, A., & Crewe, B. (2013). Prisons beyond the new penology: The shifting moral foundations of prison management. In J. Simon & R. Sparks (Hrsg.), *The SAGE handbook of punishment and society* (S. 283–307). Sage.
- Mennicken, A. (2013). ‘Too Big to Fail and Too Big to Succeed’: Accounting and privatisation in the prison service of England and Wales. *Financial Accountability & Management*, 29(2), 206–226. <https://doi.org/10.1111/faam.12012>
- Mennicken, A. (2014). Accounting for values in prison privatization. In S. Alexius & K. Tamm Hallström (Eds.), *Configuring Value Conflicts in Markets* (S. 22–42). Elgar.
- Meyer, S. (2003). Die Tageshaftkosten der deutschen Strafvollzugsanstalten: Ein Überblick. *Darmstadt Discussion Papers in Economics*, (121). Darmstadt: Technische Universität Darmstadt.
- Miller, P., & O’Leary, T. (2007). Mediating instruments and making markets: Capital budgeting, science and the economy. *Accounting, Organizations and Society*, 32(7–8), 701–734.
- Niedersächsischer Landtag. (2010). *Kleine Anfrage mit Antwort: Wohin steuert die Landesregierung? - Zielvereinbarungen und Kennzahlen für den niedersächsischen Justizvollzug. Drucksache 16/2374*. Niedersächsischer Landtag - 16. Wahlperiode.
- Leistungsorientierte Haushaltswirtschaft Niedersachsen (2005, June 2). Leitfaden Zielvereinbarungen. LoHN Dokument: No. 14. Niedersächsisches Finanzministerium.
- Pollitt, C., van Thiel, S., & Homburg, V. (2007). *New public management in Europe: Adaptation and alternatives*. Palgrave Macmillan.
- Suhling, S., Budde, S., & Häßler, U. (2015). MeWiS: Messinstrument der Wirksamkeit des Strafvollzuges: Konzept einer länderübergreifenden Arbeitsgruppe. *Forum Strafvollzug. Zeitschrift Für Strafvollzug und Straffälligenhilfe*, 64(2), 109–115.
- Yin, R. K. (2014). *Case study research: Design and methods*. Sage Publications; Sage.

**Open Access** Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

