Performance in Public Organizations: Clarifying the Conceptual Space

By Lotte Bøgh Andersen\textsuperscript{12}, Andreas Boesen\textsuperscript{1}, & Lene Holm Pedersen\textsuperscript{23},
\textsuperscript{1}Department of Political Science, University of Århus, \textsuperscript{2}KORA – Danish Institute of Governmental Studies, \textsuperscript{3}Department of Business and Politics, Copenhagen Business School, Denmark,

\begin{flushright}
Mailto: lhp.dbp@cbs.dk
\end{flushright}

Abstract

Performance in public organizations is a key concept in need of clarification. Based on a conceptual review of articles in ten public administration journals, this article proposes six distinctions to describe the systematic differences between performance criteria: From which stakeholder’s perspective is performance being assessed? Are the criteria formal or informal? Are the criteria subjective? What process focus and product focus do they have (if any)? What is the unit of analysis? Based on these distinctions, we classify the performance criteria of existing studies used in an empirical review of management and performance. Our results illustrate how a systematization of the conceptual space of performance in public organizations can help researchers select with greater accuracy what to study and what to leave out while also bringing greater clarity to public dialogues about performance.
Clarification and refinement of concepts is a fundamental scientific task (Adcock and Collier 2001, 529), and performance is an important — possibly the most important — concept in public administration. According to Rainey, “virtually all of management and organization theory concerns performance and effectiveness, at least implicitly” (1997, 125). Very broadly defined, performance is the actual achievements of a unit relative to its intended achievements, such as the attainment of goals and objectives (Jung 2011, 195). In the last decades, both quantity and quality of the empirical contributions to the performance literature have increased (Andrews, Boyne, and Walker 2006b, 14; Brewer 2006; Bommer 1995; O’Toole and Meier 2013; Meier et al. 2015; Meier and O’Toole 2013; Andrews, Boyne, and Walker 2006a, 14; Walker et al., 2015), but the performance concept itself is still not clear enough. In order to take the study of performance one step further and to bring greater clarity to public dialogues about what performance is, it is necessary to clarify the conceptual space of performance in public organizations.

25 years ago, Hirsh and Levin (1999: 208) prophesized that the performance concept would go through a life-cycle where it would fall by the wayside or narrow in scope, once better operationalizations were required. Coinciding with the increasing number of articles on performance, the concept of organizational effectiveness had disappeared as a construct in public administration, and Hirsch and Levin argued that this was due to questioning from ‘the validity police’, that is, critical reviewers. They predicted that this would also happen for the performance concept, and there are indications in the present performance literature that they were right, given that the strong demands for better research designs and operationalizations in future performance studies (e.g. Andrews et al, 2006a and Meier & O’Toole, 2013). However, as Hirsch and Levin (1999) stated, there is a tension between rigor and relevance. Especially in studies of performance in public organizations, it is problematic to focus on one performance criterion without discussing what is excluded, as most public organizations
have multiple, potentially conflicting goals. Public administration studies of performance that do not explicate their partial perspective will paint a biased picture of performance as a whole.

We argue that the solution is not to abandon the concept – as it happened to ‘organizational effectiveness’ - but rather to develop a valid conceptual space based on theoretical arguments and to exemplify how we can use this conceptual space to compare and develop studies of performance in order to systematize what is included and what is left out. Furthermore, a conceptual space can explicate the choice between different performance criteria. Our research questions are thus:

- What are the central distinctions in the conceptual space of performance in public organizations?
- How can these distinctions be defined and understood?
- How can the distinctions inform our understanding of existing studies of performance in public organizations?

A performance criterion is a theoretically defined standard based on achievements that can be evaluated. Effectiveness is an example of a specific performance criterion, and for a school it could be the extent to which formal learning standards are reached. In contrast, a conceptual space consists of one or more distinctions, representing various qualities or characteristics of criteria corresponding to how these criteria are judged to be similar or different. Which stakeholder has, for example, decided that effectiveness should be a performance criterion for the organization in question? And is the criterion formal or informal? A discussion of performance in public organizations that ultimately clarifies the conceptual space will allow us to understand and use specific performance criteria better, and a stronger conceptual framework can make performance research less data driven and sharpen the potential for theory building. Analyzing different performance criteria in different studies is not in itself
problematic, but it requires greater conceptual clarity as well as a framework that allows us to compare different performance criteria.

The literature has begun to develop the conceptual space for performance in public organizations, and important contributions have been made, even if no final consensus exists. The 3E model focuses on economy (production costs for a given quality), efficiency (cost per unit of output), and effectiveness (achievement of formal objectives) (Boyne, 2002; 2003a; 2003b). Similarly, the IOO model examines the sequence of inputs, outputs, and outcomes, where inputs are comparable with economy in the 3E model. Outputs concern the actions performed in the production process and include both quantity and quality. The ratio of outputs to inputs is defined as efficiency. Outcome concerns changes in external units (e.g., individuals or organizations), which are the object or target of the relevant policy or service intervention (Behn 2014, 137; Lynn and Robichau 2013, 208). It can include effectiveness as well as impact and equity in outcomes. The ratio of outcomes to inputs is defined as cost-effectiveness or ‘value for money’ (Walker et al., 2015).

Distinguishing between these different elements in the production process is useful and will also be part of this article’s proposed conceptual space, but many aspects of performance in public organizations are not included in the 3E and IOO models which draw on management of private firms. First, aspects related to the process are not included, i.e. accountability, probity, participation, due process etc., and these aspects can be very important in public organizations (Walker et al., 2010; Moynihan et al. 2011). Second, identifying the stakeholders is central in defining performance, because it is otherwise unclear who has the legitimate right to define what good performance is (Andrews et al., 2006: 29; Amirkhanyan, Kim, and Lambright 2014). Third, ambiguity about the relevant units of analysis can hinder comparisons of different performance studies. While performance analyses of individual surgeons and of hospital wards both contribute to our understanding of performance in
public organizations, it is necessary to be explicit about the level on which performance is conceptualized. Responding to the need for a coherent conceptualization of performance in the public sector, we identify six distinctions and show how they can differentiate between the current, disparate performance criteria in the public administration literature. This facilitates a more explicit discussion of the relation between the performance concept and the specific performance criteria used in public administration studies.

The article is structured in five sections. We next discuss the relationship between the general performance concept and the specific concepts used in individual articles and books and argue that a systematized conceptual space clarifies the relationship between these two conceptual levels. We then explain the methodology used to develop our proposed clarification of the conceptual space for performance and discuss the six identified distinctions. This is followed by a categorization of 110 performance criteria from 66 existing performance studies, which exemplifies how the distinctions can inform our understanding of performance in public organizations. In the final section, we discuss the benefits of a clear conceptual space for future research and practice.

**Conceptualizing performance**

The relationship between concepts and observations can be illustrated as consisting of at least three levels as shown in Figure 1. A given study’s performance concept is based on the general performance concept in the relevant literature (also called the background concept). A specific concept can contain one criterion or several criteria depending on the complexity included in the relevant research, and the same is the case for the indicators (also called measures).

[FIGURE 1 here]
Conceptualization is developed through downward and upward movement in figure 1. On the left-hand side, the specific concept used in a study should be based on the background concept. Here, a central choice relates to the number of criteria included. Background concepts often include a variety of meanings (this is especially true for performance), and this means that it is normally necessary to prioritize when the specific concepts are formulated.

Adcock and Collier (2001) warn researchers against three common traps when they establish their specific concept, and we argue that a clear conceptual space can help avoid all three. First, the flexibility inherent in the choice of concept does not suggest that everything goes. In most literatures, scholars commonly associate a matrix of potential meanings with the background concept, limiting the range of plausible options, and the researcher who strays outside it risks being dismissed or misunderstood. A clear conceptual space makes this explicit. Although performance of public organizations is inescapably contestable (Boyne et al. 2006: 6), a clarification of the concept gives a vocabulary to discuss what performance is and what it is not.

Second, scholars should refrain from claiming that their choice of a specific concept is the only possible choice. A clear conceptual space facilitates cumulative research if scholars recognize that it is fruitful to emphasize different aspects of a background concept in developing specific concepts. Rather than make sweeping claims about what performance in public organizations “really” means, scholars should present specific arguments, linked to the goals and context of their research, that justify their choices.

The third problem occurs when scholars stop short of accounting in detail for their specific concept. Adcock and Collier argue that scholars should specify the meaning and entailments of their specific concept in relation to the background concept. If the translation from background to specific concept in performance studies included a more systematic discussion of what is included and what is
left out, it would be easier to make a comprehensive review of the findings in the performance literature.

**Methodology for reaching the proposed conceptual clarification**

Performance in public organizations do, as mentioned, have highly variable and often implicit meanings, and this makes the concept indistinct and elusive, hindering cumulative research. Still, a proposed clarification of the conceptual space needs to be based on relevant literature, and this section discusses the basis for our conceptualization. Building on existing conceptual discussions in public administration books and journal articles, we have performed a focused, conceptual review of existing understandings of the performance concept. Our aim was to identify distinctions, which are relevant to conceptualize and include in the overall concept (Petticrew and Roberts 2006, 39), rather than to provide an overview of all the empirical results produced in this field.

The identification of eligible studies was based on two principles. First, the review was delimited by articles published in top peer-reviewed journals in public administration. We included 10 journals in line with the selections in recent reviews of public administration concepts (e.g. Voorberg, Bekkers, and Tummers 2014; Kuipers et al. 2014; Tummers et al. 2015). Selecting the top journals should make it unlikely that we have omitted important trends in the existing discussions, and we did not find additional distinctions for the last many articles we conceptually reviewed. This indicates that a sufficient number of journals have been reviewed to identify the relevant distinctions. The second principle was specific usage of the terms “performance” and “public”, but not synonyms like effectiveness and results. As implied by Hirsh and Levin’s (1999) predictions about the rise and fall of the performance concept, there are several related concepts, but it would be another task to analyze the
relationship between the performance concept and other relevant concepts. Given that we aim to identify the state of the art in the conceptualization of performance, articles that focus empirically on the effect of specific explanatory variables were excluded unless they included a conceptualization of performance. As discussed below, studies were identified according to the steps depicted in Figure 2, and we reached a saturated conceptualization, where no more dimensions were identified.

[Insert figure 2 here]

In the first step, a search was conducted in Scopus for articles published before April 15, 2014 with “performance” and “public” in the title, abstract or keywords in 10 journals: Public Administration Review (PAR – 190 results), Journal of Public Administration Research and Theory (JPART – 130 results), Public Administration (PA – 97 results), American Review of Public Administration (ARPA – 50 results), International Public Management Journal (IPMJ – 26 results), Administration and Society (A&S – 35 results), International Review of Administrative Science (IRAS – 103 results), International Journal of Public Administration (IIPA – 64 results), Public Management Review (PMR—74 results), and Review of Public Personnel Administration (ROPPA – 25 results). In the second step, title, abstracts and keywords in step 1 articles (n = 794) were screened, and 635 studies (step 3) were excluded, because it was clear from title, abstract and/or keywords that they did not conceptualize performance in public organizations. The remaining 159 studies were screened on the specific eligibility principles via title, abstract and text (step 4), and 98 studies were then excluded in step 5, because they did not contribute with conceptual performance discussions. The remaining 61 articles were included in the conceptual review (n = 61), and 28 of these are referred to in the final article. Additionally, snowball sampling based on the reading of the 159 studies identified in step 4 led to identification of relevant studies, which were screened and included if relevant. Furthermore, references identified through peer discussions were included. This additional process (step 7) identified
27 references, nine of which are referred to in the final article. The process was similar for books (step 8).

This literature forms the basis of the six distinctions discussed below. Ideally, the distinctions should be comprehensive, mutually exclusive, and systematized in order to capture the most important aspects of how performance is conceptualized in public administration literature.

The Distinctions

Starting with the distinction consistently regarded as most important, namely who decides what performance is, this section discusses stakeholders, formality, subjectivity, process focus, product focus and finally unit of analysis for performance criteria in public administration research.

Stakeholders

The central question behind the first distinction is who decides what good performance is. It is broadly accepted that different stakeholders, inside and outside public organizations, hold diverse views on what it takes to perform well (Boyne 2003a, 368; Smith and Larimer 2004; Walker, Boyne, and Brewer 2010; Yang 2009). However, the distinction between different stakeholders’ performance criteria is not always explicitly defined, and it is seldom discussed who ultimately has the authority to determine what good performance is.

Stakeholders can be defined as “any group or individual who can affect or is affected by the achievement of the organization’s objectives” (Freeman 1984, 46). Thus, performance criteria and their relative weight vary across stakeholders, and performance criteria are not technical and universal, but rather politically constructed (Boyne 2003b, 224). The performance literature often points out that
giving one stakeholder priority may have distributive consequences (Andrews et al. 2012; Boyne 2003a; Moynihan et al. 2011), but the identity of prioritized stakeholder is seldom explicitly mentioned in comparisons of performance studies. Analyzing stakeholder identity as a distinction in the conceptual space of performance allows us to compare which stakeholders are taken into account and which stakeholders are left out.

The authority to define good performance differs in public and private organizations. While private organizations often have close ties to their owners and customers, public organizations have stronger relations with political and government authorities and are more exposed to political influence (Mintzberg 1973). Boschken (1992) distinguishes between performance indexes according to the source of authority. For organizations resting on political authority, politicians formulate the goals and objectives, and the relevant performance question is whether the organization meets its public mandate, in which case it is performing well. Boschken (1992, 272) labels this “social effectiveness”. Here, performance is about attaining democratically stated goals. For the market, the source of authority is economic, and the central question for the organization is: Do our customers like us? (Boschken 1992). Even when the introduction of market-based principles has created “customers” in public service industries, the public mandate is often still crucial to the legitimate claim on authority, and public organizations rarely operate according to ideal-typical market principles as the financial resources depend on political authority.

A third unit of authority, in addition to the market and the state, is the clan (Ouchi 1980). The clan is a group that may or may not be linked by kinship ties but is based on common, internalized goals and strong feelings of kinship (Scott 1995, 252). Clan systems are characterized by implicit, internalized control mechanisms, non-specialized roles and career paths, holistic rather than segmented concerns, and long-term employment (Ouchi 1980). The socialization process effectively eliminates
goal incongruence between individuals—that is, among the members of the clan—while incompatibility can remain between the clan and outsiders or their managers (Behn 2014).

Professions are often seen as the prime examples of clans due to the existence of inter-occupational norms and specialized theoretical knowledge (Ouchi 1980). Accordingly, a profession is a unit of authority that has standards for defining good performance, and this authority rests on expert knowledge and to some extent on a monopoly on the delivery of specialized services (Andersen 2005, 205; Collins 1990; Friedson 2001). In some types of service delivery, professions hold the authority to define good performance, but this may be challenged if other stakeholders find that professions fail to deliver high quality services. Professions can also have the power to influence the opinions of others based on recognized expertise and knowledge. In addition to public service professions such as teachers and nurses, researchers for example from the public administration scholarly community can also have this type of clan power and become the arbiters of performance in public organizations by virtue of their expertise and the choices they make in their research. In contrast, the logic of the hierarchical model as implemented in representative democracies is that elected politicians have the power to make decisions and enforce obedience based on recognized democratic legitimacy.

In the literature on performance in public organizations, the stakeholder model is pluralist, favoring the worldview that there are multiple legitimate interests surrounding an organization. The performance literature also tends to emphasize a more polycentric relationship with multiple principals (Boschken 1992; Boyne 2003b; Moynihan et al. 2011; Talbot 2008, 1579) rather than a simplistic vertical principal–agent relationship in the public sector. Stakeholders can, for example, be customers, clients, citizens, professional groups, organizations in the environment, governance networks, and international institutions (Stoney and Winstanley 2001). These various stakeholders possibly steer towards specific forms of performance in public organizations, including conformance to internal
process standards (Talbot 2008, 1580). In this complex institutional environment, it is relevant for analysts to ask who holds the authority to determine that a given criterion represents performance.

**Formality**

Research on organizational performance tends to see organizations as established to achieve specific objectives (Boyne 2003b, 214). The formality of these objectives may vary, however. Some public organizations do not have formal goals that are clearly expressed in legislation or other official documents, and formal goals (should they exist) can be broad mission statements rather than concrete objectives. Boyne (2003b) argues that a combination of the multiple stakeholder model as discussed above and the goal model is a positive basis for defining performance in public organizations. The key argument of the goal model is that the extent to which formal goals are attained should be used to assess performance. Although this steers us free of seeing performance as consisting only of the subjective impressions of powerful stakeholders (Boyne 2003b), it remains important to recognize that performance criteria can vary in formality.

The central question in the distinction concerning formality is the degree to which performance is formally or informally defined. Goals that are formalized in the sense that they are written, predefined, and/or made explicit are more formal than implicit expectations regarding the achievement of a unit. Compliance with professional norms, for example, can be central in public organizations, and these norms may be highly informal, even tacit. Politicians may also formally decide that satisfying a group’s preferences is the goal, making user satisfaction a formal performance criterion. It would still, however, be an inherently subjective performance criterion as discussed below.

**Inherent Subjectivity**
Subjectivity has been discussed at length with regard to operationalization of performance measures, but performance criteria may conceptually also have varying degrees of inherent subjectivity. The central question behind this distinction is the degree to which a performance criterion concerns interior experiences and perceptions versus exterior, observable phenomena. Some criteria such as trust in civil servants (Van Ryzin 2011) are inherently subjective. Boyne (2003b) mentions consumer satisfaction as either a criterion in itself or a proxy for another type of performance. As the former, it can be classified as inherently subjective, because satisfaction is an internal feeling. An example of a criterion related to exterior and observable phenomena is the physical health of hospital patients.

This is a conceptual distinction, but performance can also be measured more or less subjectively according to whether it concerns fixed criteria of performance, whether there is a process for verifying the accuracy of the measure, and whether there is external verification (Andrews, Boyne, and Walker 2006b). For instance, parents’ satisfaction with schools is an inherently subjective performance criterion and almost always measured subjectively compared to student stress, which is a school performance criterion with objective manifestations. Stress can be assessed in a relatively objective manner with a saliva test of cortisol levels, but it can also be measured subjectively by simply asking the students. Correspondingly, survey measures are not the only way to assess parent satisfaction; (objective) exit behavior from the school also serves as a measure of satisfaction.

Distinguishing between performance criteria with different subjectivity highlights the relevance of first discussing the distinction between degrees of subjectivity at the conceptual level and subsequently to make decisions regarding measurement. Research that postulates that objective measures of performance are always preferable to subjective measures ignores conceptual differences between performance criteria, whereas our discussion of the performance concept implies that the choice of indicator depends on the conceptual understanding of the relevant performance criterion.
Similarly, it is difficult to make an *a priori* decision in favor of either inherently subjective or inherently objective performance criteria, because this depends on what understanding of performance relevant stakeholders find important.

**Type of Process Focus**

Performance in public organizations involves more than services and decisions; it also includes the manner in which the service is delivered. In the context of the private sector, Grönroos (1984, 39) argues that both product and process are relevant for customers. How a good or service is made functionally (the process) is often important to consumers and their view of what they receive (the product—which can also be a service, e.g., a haircut). In the public sector, this argument is even more valid because the specific user is rarely the sole benefactor of the service, and Van Ryzin (2011) emphasizes that some specific governmental processes matter a great deal to citizens. Moynihan et al. (2011, 143) also criticize that performance measurement systems displace attention from democratic values linked to the working processes of the administration such as due process, equity, integrity and transparency, because these systems focus only on product criteria related to mission achievement and effectiveness and disregard process criteria related to traditional democratic values. Performance criteria can focus on process as well as product, but the distinction is not between means and ends (where means per definition is not performance), because democracy is a goal in itself in a democratic state; hence, administration and service delivery that respect democratic values is also a goal.

**Type of Product Focus**

Boyne (2002) derives an influential classification of performance criteria by combining the 3E (economy–efficiency–effectiveness) and IOO (inputs–outputs–outcomes) models with two additional
criteria (responsiveness and democratic outcomes). After omitting the redundant elements of the two models, Boyne (2002, 19) claims that he has a comprehensive checklist. He distinguishes outputs, efficiency, service outcomes, responsiveness, and democratic outcomes and specifies 15 performance criteria. The problem with combining two different models and including responsiveness and democratic outcomes in an ad hoc way is that it is unclear how the criteria differ.

We argue that the distinction in the original IOO model (i.e., production phase) can be useful to classify what type of product focus (if any) performance criteria have: Input, output, outcome, output per input or outcome per input. Outcome performance criteria such as effectiveness are normally closest to achievement of defined objectives in public organizations, while input-related criteria are rarely relevant for performance in this understanding of the concept. Although outcome criteria are often seen as the gold standard, Lynn and Robichau (2013, 208) argue that they can be problematic in public organizations for two reasons. First, they may be imposed on organizations by external stakeholders and might therefore be incompletely embraced by organizational actors. Second, since public organizations are often unable to control all factors that affect outcomes (Ashworth, Boyne, and Entwistle 2010; Behn 2014; Talbot 2010), outcome criteria are poor reflections of the actual organizational effort.

Concerning the second reason, Boyne (2003b, 218) discusses whether the difficulty in attributing outcomes exclusively to the work of service providers may imply that achievement of professional standards can be used as performance criteria. He argues that that evidence of following correct procedures or doing things right might be preferable to outcome information. This links back to our discussion of the relevant source of authority. If the profession is seen as a relevant source, then Boyne clearly has a point, and his argument in favor of the usefulness of the achievement of professional
standards can also be valid and useful if it is impossible to assess the achievement of outcome-related goals (given that evidence indicates that compliance with professional standards leads to these goals).

**Unit of Analysis**

Prior to assessing effectiveness, Cameron (1986, 542) argues that a number of questions must be answered. The same goes for performance. The first question concerns the perspective from which effectiveness is being assessed, which corresponds to the stakeholder distinction discussed above. The next two questions concern the unit of analysis. What domain of activity is being assessed, and what is the level of analysis? These questions are about the identity of the unit whose achievements we analyze. Who performs: individuals (Andersen, Heinesen, and Pedersen 2014), groups (Van Ryzin 2011), organizations (Angle 1981; Caillier 2011; Kim 2005, 245), or programs (Collins and Gerber 2008; Jung 2014)? Even when we analyze individuals, it might be relevant to consider performance at a higher level, because the extra-role performance of individuals captures their contribution to broader organizational goals, while the in-role performance is narrowly related to the individual’s own work task (Van Loon, Vandenabeele, and Leisink 2015; Williams and Anderson 1991). Although individuals are the agents behind extra-role performance, the organization might be the relevant level to analyze performance effects. For teachers, an individual in-role performance criterion could be their students’ academic skills in their subject controlled for performance in other subjects. Extra-role performance is typically more difficult to capture. For teachers, students’ academic skills at the school level could be relevant, but this would also be a consequence of efforts from other school employees and factors outside the school such as student background.

Analyses of performance carried out at different analytical levels may speak to each other, but they are not directly cumulative. For instance, the organizational performance investigated at the school
level (Andersen and Mortensen 2010) and the individual performance investigated for each teacher (Andersen, Heinesen, and Pedersen 2014) represent fundamentally different performance criteria, although knowledge of the factors that increase individual performance are crucial in order to improve school performance.

The distinction between different analytical levels calls for a conscious choice of the unit of analysis in research designs. The consequences of individual factors, such as public service motivation, job satisfaction, and intrinsic motivation, may be better investigated in terms of performance of individual employees, as much information is lost if individual scores are aggregated to the organizational level. In contrast, it may be less relevant to investigate individual-level performance in organizations such as hospitals where production is the result of teamwork involving different professions and organizational units. Linking the survival of cancer patients to individual surgeons makes less sense than using students’ math skills to assess the performance of their math teacher.

The fact that performance exists at different levels does not mean that all levels should necessarily be analyzed, but rather that we must prioritize between different units of analysis and take the explanatory variables into account when deciding the relevant analytical level. If explanatory variables with a causal impact on performance are conceptualized at a given analytical level, it speaks in favor of also conceptualizing performance at this level.

The unit of analysis is the last distinction. Table 1 summarizes our discussion; it lists the key question for each distinction and provides examples of how performance criteria can vary based on the distinctions.

[TABLE 1 HERE]
Compared to Boyne’s conceptualizations of performance (Boyne 2002), our conceptual space explicates the distinctions between the criteria. As mentioned, Boyne’s classification contains five main dimensions (outputs, efficiency, service outcomes, responsiveness, and democratic outcomes) and 15 sub dimensions. In our terminology, Boyne’s dimensions can be seen as criteria, which be classified on one or more of our distinctions. For instance, output and service outcomes are different product focuses, while different types of democratic outcomes would be classified differently on our distinctions. While Boyne focuses on developing the criteria (called dimensions in his terminology), we focus on the differences between the criteria. It is possible to map on which distinctions the criteria are placed, and the two conceptualizations thus supplement each other in building the overall conceptual space.

**Using the Distinctions to Classify Specific Performance Criteria**

Deriving and discussing the six distinctions is merely the first step. It is also central to test the usefulness of the distinctions on empirical studies of performance in public organizations. As shown in detail in the online Appendix, we have classified 110 performance criteria used in Walker and Andrews’ (2015) review of local government management and performance. The advantage of this is that the six distinctions are not tested on the studies which were used to derive them. The focus on local governments in Walker and Andrews’ review (2015) means that there is limited variation in units of analysis, but our classification of the 110 performance criteria reveals that all six distinctions are relevant. It is for example important to know whether performance is analyzed for local authorities or police departments, and even for seemingly similar studies it can be relevant to highlight the difference between looking at performance for schools and school districts. The articles’ descriptions of their
performance criteria made it possible to classify almost all of them on all six distinctions. Some articles did not explicitly mention the information needed to make the classification, but it is often possible to infer from the context of the articles how the performance criteria should be classified.

The classification of the empirical studies (see Table A2 in the online Appendix) draws attention to two groups of performance criteria. First, research connected to Cardiff University (e.g. Andrews, Boyne, Law and Walker 2005) often combine many different criteria in indexes, thus covering performance of the investigated organizations (typically British local authorities) broadly and with central government represented by the audit commission as a key stakeholder. Second, research connected to Meier and O’Toole’s research program on public management typically focus more narrowly on specific performance criteria, often related to student academic performance and often analyzed for school districts in Texas (e.g. Meier and O’Toole 2003).

The substantial number of strong contributions from these two traditions means that many of the criteria used are very similar. 29 of the 110 performance criteria are proportions of students with acceptable academic competences (typically operationalized using TAAS test scores), while additional 14 criteria concern equity in proportions of students with different ethnic backgrounds with acceptable academic competences, and 13 additional criteria concern proportions of students with high quality academic competences (e.g. above 1110 on the SAT college entrance examination). In terms of comparability of different results, it is definitely not a problem that performance criteria in existing studies are similar, but given that we also work towards a comprehensive picture of performance in all types of public organizations, it is advisable to also apply performance criteria which differ more on the selected distinctions. Specifically, our classification highlights that quite few existing performance criteria (20 out of 110) have a process focus. Existing examples include equal access to services, equity in the use of disciplinary actions and user satisfaction with the process.
This highlights the usefulness of including distinctions concerning both process and product focus. While the distinction concerning product focus ensures comparability to existing classifications – illustrated by the fact that our classification on this distinction is almost similar to Walker and Andrew’s (2015) classification – it does not capture all relevant performance criteria. Examples of alternative criteria which can be systematized using our distinctions are participation, probity and responsiveness in service delivery processes. Although Walker and Andrew (2015:104) argue that “these additional dimensions of performance are critical for scholars to understand and assess the performance”, they do not present a systematic way to conceptualize them. We hope that our inclusion of the process-focus distinction will highlight that performance can relate to process as well as to product aspects. In line with this, our classification of existing performance criteria shows that it is relevant to differentiate between different types of equity and user satisfaction. Equity can for example concern either the product, as illustrated by Meier and O’Toole’s (2003) analysis of whether outcomes are equally good for students with different backgrounds, or the process, as exemplified by Andrew et al.’s (2010) analysis of how fairly services are distributed amongst citizens. Similarly, it can make a difference whether performance is conceptualized as user satisfaction with the process or as user satisfaction with the product delivered by public organizations.

Discussions about the stakeholder identity are especially exemplary in the Cardiff-tradition, but their combination of different criteria (such as user satisfaction and effectiveness) in index measures sometimes means that there is little information about each criterion. Oppositely, the parsimonious criteria in the Meier/O’Toole tradition makes it easy to classify for example TAAS pass rates as effectiveness. Increased awareness of this trade-off between parsimony versus broad performance conceptualizations has at least two advantages: Researchers can make more explicit choices between stringency and completeness based on systematic knowledge about other performance
studies, and it can bring greater clarity to public dialogues about what performance is and what performance criteria are reasonable and legitimate.

Talking about legitimacy, our classification reveals that a large majority of the performance criteria are determined by external stakeholders. Democratically, this is reassuring, especially because many of these external stakeholders are democratically elected politicians. While only five of the 110 classified performance criteria are purely internal (and thus determined by the units whose performance are evaluated), 90 criteria have external stakeholders. Three performance criteria combine external and internal stakeholders, and seven are determined by the researchers alone. Only for five criteria, it is totally implicit who the stakeholders are. Still, for 23 of the criteria classified as having external stakeholders, this could not be seen directly from the text and was therefore inferred from context information in the article. This means that research could still benefit from more clarity. This would also make the role of the researcher clearer; while only seven criteria are explicitly researcher determined, more explicitness would facilitate a systematic evaluation of the extent to which researchers themselves are stakeholders or important arbiters, deciding how performance should be (measured) in public organizations.

Our last observation is directly relevant to the ongoing debate about subjective versus objective data. Although only a limited proportion of the classified performance criteria (approximately 20 percent as shown in Table A2) are inherently subjective on the conceptual level, we argue that being aware that subjectivity can be present at both conceptual and operational levels will qualify the discussion about the desirability of different types of performance data. While subjectivity in measures of performance criteria which concern observable outcomes is hardly desirable, it can be necessary to use subjective measures of interior experiences and perceptions, and criteria such as user’s subjective experiences with public organizations can be very important pieces of the total performance puzzle.
Conclusion

It is essential for public administration research to be aware of how we evaluate performance in public organizations, and this article aims to identify, define and discuss the central distinctions in order to build a conceptual space. Clarifying how performance in public organizations is assessed facilities better evaluation of existing performance studies, improves our ability to design new performance studies which fill important gaps in the literature, and clarify public dialogues about what performance is and what performance criteria are reasonable and legitimate. These three contributions are discussed in more detail after a short overview of the six distinctions which can be used to differentiate between different performance criteria.

A key insight in the performance literature is that public organizations often have multiple stakeholders and multiple goals and that the trade-offs between them are political choices (Radin 2006), and the most important distinction therefore concerns the identity of stakeholder behind a given performance criterion. Such criteria can be more or less formally stated, and although vague goals are seen as an inherent feature of public organizations (Heffron 1989), a classification of existing performance studies shows that there is substantial variation on this second distinction.

Turning to the contents of the performance criterion, the third distinction differentiated between different degrees of subjectivity. Clarifying whether a performance criterion conceptually concerns interior experiences and perceptions contributes greatly to the discussion on subjective versus objective performance measures, because it highlights that subjectivity is both a question about conceptual definition and data sources, and the critique of subjective measures sometimes fails to differentiate between these two levels. Existing studies (O’Toole & Meier 2013; Meier & O’Toole 2013; Meier et
al. 2015) have convincingly shown that it can be problematic for measurement validity to use perceptual measures to measure objective performance criteria, but our third distinction implies that it is worthwhile to discuss how we measure performance criteria which on the conceptual level concern interior experiences and perceptions. If our effort to increase measurement validity means that we do not include any inherently subjective performance criteria, it could be a problem for the conceptual validity of performance research in public administration, because important aspect of the theoretical performance concept would be ignored. It is for example difficult to measure user satisfaction without using perceptual measures.

The fourth and fifth distinctions concern performance criteria’s process and product focuses. Distinguishing between different types of product focuses corresponds to the IOO and 3E models in the existing literature, while the explicit inclusion of process answers several calls for attention to this type of performance in public organizations ((Van Ryzin 2011; Moynihan et al. 2011, 143)).

Finally, the last distinction relates to the level of analysis, which varies considerably. This choice depends on the explanatory variables and on the actor who is expected to produce a specific type of performance. For example, if several health workers are required for a surgical procedure, it is less meaningful to conceptualize performance as individual.

Successful application of the distinctions requires that we are able to describe the central differences among relevant studies that use them. The fact that all of the distinctions can be used to describe key differences between existing performance criteria in the literature (illustrated by our classification of 110 performance criteria) indicates that we do not have too many distinctions. There is, however, a balance between accuracy and detail on the one hand and parsimony and applicability on the other. We find that the distinctions are comprehensive and provide oversight of the central
differences and similarities between the performance criteria, and as such the distinctions can inform our understanding and comparison of performance in at least three ways.

Our classification of existing performance criteria illustrated how a clear conceptualization can facilitate better evaluation of existing performance studies. One key insight from our analysis of existing criteria were that the literature is dominated by criteria focusing only on product, but that it is also possible to analyze process aspects of performance. Another insight is that many criteria are implicit and that articles should be more explicit about their performance criteria. A last example is that comparisons of different performance studies should be very aware of the units of analysis, because performance can be conceptualized at many different levels in the public sector, and these different conceptualizations are not necessarily comparable.

This implies that the conceptual distinctions can help guide central choices related to research questions, data and methods in future performance studies. There is no logical method of having new ideas, but most researchers try to formulate research questions that contribute to an identifiable scholarly literature, and the continued discussion in the discipline of what is conceptualized as performance will therefore affect our future research questions. For example, the distinction between process and product criteria can highlight that traditional democratic values can also be seen as performance (Moynihan et al. 2011, 143). The distinctions developed in the article can also draw attention to new types of research designs (e.g. comparisons of organizations with different types of stakeholders), and they might contribute to more informed choices of data sources (e.g. basing the data choice on the degree of subjectivity of the relevant performance criterion) and higher comparability between studies (ensuring that we do not compare apples and oranges). It is necessary to have a sense of the larger conceptual space to discuss the delimitation of performance studies systematically, and the distinctions can also be used when we evaluate the contributions of existing studies. Finally, the
conceptual discussion of performance is an example of how a background concept can be systematized. This might inspire future research to clarify the conceptual space of other key public administration concepts such as public service motivation and red tape.

A systematic understanding of performance can also clarify public dialogues about what performance is and what performance criteria are reasonable and legitimate. The conceptual space developed in this article can potentially help public managers systematize their thoughts and decisions about which criteria to prioritize in their organizations. According to Behn (2003), public managers often measure performance because it helps them to achieve different managerial purposes (evaluate, control, budget, motivate, promote, celebrate, learn, and improve), and different purposes require performance criteria with different characteristics (Behn 2003, 593). We think that it is very important for managers in public organizations and for other participants in public dialogues about performance to understand the entire conceptual space of performance before they select the most important performance criteria, which are then measured. Hopefully, this article is helpful in that regard. Although completeness in measurement is not possible (or even desirable), it is very useful to know explicitly what types of performance you measure—and which you do not.

References


doi:10.1017/S0003055401003100.


Figure 1: Illustration of the conceptualization process

<table>
<thead>
<tr>
<th>Background concept</th>
<th>Specific concept</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>The broad constellation of meanings and understanding associated with a given concept in a given literature. Can be systematized in a conceptual space.</td>
<td>The formulation of a concept used by given scholar(s). Can contain one criterion or several criteria, depending on the complexity included in the relevant research.</td>
<td>Also referred to as measures and operationalizations. Should reflect the criteria included in the specific concept.</td>
</tr>
</tbody>
</table>

Note: The figure is inspired by Adcock and Collier (2001, 531), who also include a fourth level (Scores for Cases) because they are also interested in specific measurement whereas we focus on conceptualization.
Figure 2: Flowchart on the process for identifying and retaining studies

Studies identified through search on performance and public in the title, abstract and keywords of 10 international journals of public administration (step 1)

Studies selected on general eligibility criteria by screening title and abstracts (n=794) (step 2)

Studies excluded based on screening of title, abstract and keywords (n=635) (step 3)

Studies screened on eligibility criteria by screening title, abstract and full-text (n=159) (step 4)

Full-text studies excluded based on reading of text (n=98) (step 5)

Studies included in systematic review (articles n=61) (step 6)

Snowballing (step 7) identifying relevant references based on the articles read in step 4 (n=27)

Studies referred to in the conceptual review (articles n=37 and books = 7) (step 9)

Snowballing. Books identified through the reading of the studies included in the review (n=7) (step 8)
Table 1: A conceptual space of performance

<table>
<thead>
<tr>
<th>Distinction</th>
<th>Question</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stakeholders</td>
<td>Who decides what good performance is?</td>
<td>E.g., politicians, users, or professionals</td>
</tr>
<tr>
<td>2. Formality</td>
<td>To what degree is performance formally or informally defined?</td>
<td>E.g., written goals vs. implicit expectations</td>
</tr>
<tr>
<td>3. Subjectivity</td>
<td>To what degree does performance concern interior perceptions versus exterior phenomena?</td>
<td>E.g., consumer satisfaction vs. actual physical health as the performance criterion</td>
</tr>
<tr>
<td>4. Type of process focus</td>
<td>What type of process focus does the criterion have (if any)?</td>
<td>E.g., fair process or user participation in the decision-making process</td>
</tr>
<tr>
<td>5. Type of product focus</td>
<td>What type of process focus does the criterion have (if any)?</td>
<td>E.g., effectiveness or output quantity</td>
</tr>
<tr>
<td>6. Units of analysis</td>
<td>Who performs and on what level?</td>
<td>E.g., individual, organization, or program</td>
</tr>
</tbody>
</table>