education sciences

MDPI

Article

When Is Faculty Pay-for-Performance Perceived as Effective? A Configurational Study of Organizational Discretion Across Chinese Public University Colleges

Zheng Yang [1,*] and Shiqin Xu [2]

[1] Department of Public Administration, California State University, Dominguez Hills, Carson, CA 90747, USA
[2] Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, CA 94305, USA
* Correspondence: zyang@csudh.edu

Abstract

Faculty pay-for-performance (PFP) is often evaluated primarily in terms of incentive design, yet its implementation also depends on the authority, resources, and organizational complexity of the academic units that enact it. This study examines how managerial discretion, financial discretion, implementation satisfaction, and college size combine in configurations associated with administrators' assessments of faculty PFP. Evidence was drawn from 22 semi-structured interviews with college-level administrators and university-level administrators across 19 colleges in three Chinese public universities, supplemented by institutional policy documents. Qualitative thematic analysis identified the relevant organizational conditions, followed by fuzzy-set qualitative comparative analysis of the 18 colleges with complete data. The results reveal three configurations associated with high perceived effectiveness and one associated with its absence, demonstrating that no single level of managerial or financial discretion is uniformly effective. Instead, favorable assessments arose through different alignments of discretion, implementation satisfaction, and organizational size. Conversely, high managerial discretion combined with limited financial discretion, low implementation satisfaction, and smaller college size was associated with the absence of high perceived effectiveness. By conceptualizing discretion as multidimensional and configurational, the study helps explain why formally similar faculty-performance policies may be perceived differently across academic units and provides an empirical baseline and organizational framework for evaluating subsequent reforms in Chinese higher education.

Keywords: higher education administration; faculty performance management; pay-for-performance; organizational discretion; fuzzy-set qualitative comparative analysis; China

Check for updates

Academic Editor: Jess Harris

Received: 5 September 2026
Revised: 23 October 2026
Accepted: 27 October 2026
Published: 30 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Performance-based evaluation and compensation have become prominent features of contemporary higher education, reflecting broader shifts toward managerialism, accountability, and quantified assessment that have reshaped university operations and organizational identities (Deem, 1998; Guthrie et al., 2024; Kallio et al., 2016; ter Bogt & Scapens, 2012). Yet faculty pay-for-performance (PFP) is not implemented in an organizational vacuum. Colleges and departments operate within university-level rules, budgets, and approval structures that can restrict their ability to adapt evaluation criteria and rewards to disciplinary and organizational needs. These constraints are especially consequential in public universities, where performance systems must accommodate multiple missions, professional norms, limited reward resources, and heterogeneous relational, professional,

and financial motivations among academic staff (Frey et al., 2013; Hitka et al., 2023; Kallio & Kallio, 2014; Sułkowski et al., 2020; Weibel et al., 2010). Understanding faculty PFP therefore requires attention not only to incentive design but also to the authority and resources available to the academic units responsible for implementation.

Research has established that faculty appraisal systems can affect job satisfaction, work motivation, academic identity, and the balance between measurable outputs and less readily quantified contributions (Dasanayaka et al., 2021; Kallio et al., 2016; Sułkowski et al., 2020; ter Bogt & Scapens, 2012). In China, publication incentives and tournament-like evaluation systems have likewise shaped research cultures and academic behavior (Gu & Levin, 2021; Xu, 2020; Xu et al., 2021). Most studies, however, emphasize the content and consequences of performance criteria or examine autonomy at the level of the university. Less is known about how a common university policy is enacted across constituent colleges that possess different combinations of managerial authority, financial discretion, organizational scale, and implementation experience. College size may shape implementation by increasing the diversity of faculty needs and the demands for communication, participation, and coordination in evaluation processes (Al-Omari, 2007; Smith et al., 2024). This academic-unit level is analytically important because formal institutional autonomy does not necessarily reveal the de facto discretion available to those who translate policy into evaluation and compensation decisions.

Organizational discretion provides a bridge between research on faculty performance systems and research on higher education governance. Autonomy is multidimensional rather than unitary: authority to design evaluation rules may coexist with limited control over the funds needed to make rewards consequential (Enders et al., 2013; Verhoest et al., 2004). Moreover, appraisal research emphasizes that reactions to an evaluation system depend on its social and implementation context, including perceived fairness, participation, feedback, and relations with evaluators (Levy & Williams, 2004). Implementation satisfaction concerns administrators' assessments of how PFP is enacted, whereas perceived effectiveness concerns their assessments of its outcomes. This distinction is consistent with performance-appraisal research that examines satisfaction with appraisal processes separately from perceived appraisal effectiveness (Abdullah & Malik, 2022; Maharvi et al., 2023). These insights suggest that neither high nor low discretion should be expected to produce uniformly favorable outcomes. Its relevance is likely to depend on how distinct forms of discretion combine with implementation quality and the coordination demands associated with academic-unit size.

This study therefore asks: Which combinations of managerial discretion, financial discretion, implementation satisfaction, and college size are associated with high perceived effectiveness of faculty PFP or its absence? The study combines qualitative thematic analysis with fuzzy-set qualitative comparative analysis (fsQCA) across colleges nested within three Chinese public universities. By comparing colleges operating under common university-level requirements, the study reveals differences in policy implementation that would be obscured by treating the university as the sole unit of analysis. It contributes to higher education administration in three ways. First, it shifts the focus from formal university-level policy design to college-level implementation. Second, it disaggregates managerial and financial discretion and examines their alignment with implementation quality and organizational complexity. Third, it uses a configurational approach to identify multiple pathways to the same perceived outcome rather than assuming a single average effect of autonomy. Because the evidence was collected in 2014, the study captures college-level PFP during the era of Projects 211 and 985, national initiatives supporting the development of selected universities, before the subsequent Double First-Class initiative to develop world-class universities and disciplines (Xu et al., 2021). It also predates the later "Breaking Five-

Only" evaluation reform, which sought to reduce reliance on narrow academic performance indicators (Gao & Wang, 2024). Rather than characterizing present-day Chinese policy, the study provides a historical baseline for future research examining continuity and change in how faculty PFP is implemented across academic units.

1.1. Pay-for-Performance in Higher Education: From Incentive Design to Organizational Context

Pay-for-performance has been adopted across educational settings to strengthen accountability and connect rewards to measurable contributions (Ballou, 2001; Hearn, 1999; Lavy, 2007; Proudfoot, 2024). Its use remains contested, however, because educational work is multidimensional, performance measures are imperfect, and performance incentives are interpreted within broader professional and accountability contexts that can redirect effort toward visible indicators or weaken professional motivation (Kallio & Kallio, 2014; Kallio et al., 2016; Mintrop et al., 2018; Proudfoot, 2024; Sułkowski et al., 2020). Research on appraisal also shows that acceptance depends on procedural fairness, transparency, participation, feedback, and the broader relationship between evaluators and those evaluated (Kimball, 2002; Levy & Williams, 2004). These issues are especially salient in universities, where teaching, research, and service are difficult to reduce to a single metric and where academic professionals place substantial value on autonomy.

Much PFP research has consequently focused on internal design: what motivates employees, how performance should be measured, which tasks should be rewarded, whether rewards are financially consequential, and how incentives should be linked to multiple goals (Chen, 2018; Dixit, 2002; Ethiraj & Levinthal, 2009; Holmstrom & Milgrom, 1991; Park, 2022). This work is indispensable, but it can implicitly presume that implementing units possess sufficient authority and resources to select and operate an appropriate system. In practice, performance appraisal is embedded in a social and institutional context, and the same formal design may be interpreted or enacted differently across institutions, organizational subunits, and professional groups (Levy & Williams, 2004; Sheikh et al., 2022; ter Bogt & Scapens, 2012). The central issue is therefore not only whether a PFP design is technically sound, but whether the implementing unit has a workable combination of authority, resources, and administrative capacity. For example, insufficient authority may limit a college's ability to adjust evaluation procedures, whereas restricted reward funds may constrain its ability to implement those decisions. Such constraints may influence both administrators' satisfaction with implementation and their separate assessments of PFP effectiveness.

This contextual problem is pronounced in public organizations. Their missions, budgets, personnel systems, and performance expectations are shaped by legal mandates, political priorities, resource dependencies, and pressures for legitimacy (DiMaggio & Powell, 1983; Heugens & Lander, 2009; Meyer & Rowan, 1977; O'Toole & Meier, 2015; Perry et al., 2009; Pfeffer & Salancik, 2003). Public universities share these constraints while also depending on professional expertise and collegial participation. As a result, formal performance requirements imposed from above may interact with local discretion in nonlinear ways: central rules can restrict adaptation, but they may also provide consistency and legitimacy; local discretion can facilitate customization, but it can also shift responsibility to academic-unit leaders without supplying the resources required to meet expectations.

A principal-agent perspective helps explain why the locus of authority matters when a university establishes PFP requirements that colleges must implement. Universities and colleges may differ in priorities, information, and capacity, and delegated authority does not ensure that a college can use that authority effectively (Eisenhardt, 1989; Enders et al., 2013). More broadly, organizational autonomy has both formal and de facto dimensions: an academic unit may possess nominal decision rights yet remain constrained by budgets,

approval procedures, or stakeholder resistance (Bach, 2014; Verhoest et al., 2004). Examining the discretion college administrators can exercise in practice, rather than relying on formal university autonomy alone, therefore connects the design of PFP to the institutional arrangements governing its implementation.

1.2. Multidimensional Discretion and Configurational Expectations

Organizational autonomy generally refers to the scope for independent decision-making, but it varies by domain and by whether the emphasis is on formal competencies or the practical ability to use them (Bach, 2014; Enders et al., 2013; Verhoest et al., 2004). In this study, organizational autonomy refers broadly to decision-making independence, whereas discretion refers more specifically to the choices available to college administrators in implementing PFP. We focus on de facto discretion—the practical ability to exercise these choices, which may differ from formally granted authority—and distinguish two dimensions. Managerial discretion concerns a college's latitude to design and administer evaluation criteria, weights, and procedures. Financial discretion concerns its latitude to allocate or generate resources for performance rewards. The distinction matters because decision authority without commensurate resources may create responsibility without capacity, whereas financial flexibility without legitimate and workable procedures may not secure acceptance.

Discretion is often viewed as enabling managers to tailor performance systems to local tasks and to deploy resources more effectively (Birdsall, 2018; Ingraham, 1993). Nevertheless, autonomy is not inherently beneficial. Centralized rules can reduce flexibility, but they can also standardize expectations, constrain favoritism, and locate accountability at a level perceived as more legitimate. Conversely, local authority may improve fit while increasing coordination burdens and exposing college leaders to blame when resources or implementation processes are inadequate. These competing possibilities are consistent with evidence that autonomy does not operate independently of accountability and organizational context, but can condition how accountability arrangements relate to performance (Birdsall, 2018; Enders et al., 2013; Han & Hong, 2019; Nielsen, 2014).

A configurational perspective is therefore more appropriate than a universal hypothesis that greater discretion produces greater effectiveness. The relationship between managerial or financial discretion and perceived PFP effectiveness may depend on implementation satisfaction, which reflects administrators' assessments of how PFP operates, and college size, which captures differences in organizational scale and associated coordination demands. Rather than treating these conditions as independent predictors, we examine how they combine in different organizational settings. Equifinality is plausible: favorable assessments may arise through a high-discretion pathway when authority, resources, and implementation align, or through lower-discretion pathways when centralized rules reduce ambiguity and coordination demands. Causal asymmetry is also plausible: the configuration associated with the absence of high perceived effectiveness need not be the simple inverse of a favorable configuration.

China provides an informative setting in which to examine these relationships because public universities have combined strong state and university steering with differentiated implementation inside institutions (Mok, 2009; Mok & Han, 2017). During the period studied, Project 211 and Project 985 intensified pressures for measurable achievement and world-class status, while colleges retained varying practical latitude over faculty evaluation and rewards. Later initiatives, including Double First-Class and the Breaking Five-Only reform, sought to diversify evaluation criteria and reduce reliance on narrow publication indicators, but they did not eliminate the underlying governance question of how centrally initiated reforms are interpreted and implemented within universities (Gao & Wang, 2024;

Gu & Levin, 2021; Shu et al., 2022; Xu et al., 2021). The Chinese case thus offers theoretical leverage for understanding how multidimensional discretion operates under strong external steering.

2. Materials and Methods

2.1. Research Context, Case Selection, and Data Collection

This study used a comparative multi-case design involving 19 colleges in three Chinese public universities in which PFP formed part of faculty evaluation and compensation. In this study, a college refers to an academic unit within a university, comparable to a school or faculty in other higher education systems. The college was selected as the unit of analysis because university-wide requirements were translated into operational criteria, procedures, and reward decisions at that level. Evidence consisted of interviews with 19 deans or associate deans and three university-level administrators, supplemented by relevant university and college policy documents. Interview and documentary evidence were first analyzed thematically; the 19 colleges were then treated as cases, with 18 complete cases entering the configurational analysis that included college size.

The study captures the institutional setting of late 2014, when Project 211 and Project 985 structured the pursuit of world-class status and intensified competition around research performance. This timing is substantively important. The evidence predates the Double First-Class initiative and the later Breaking Five-Only reform, which sought to reduce excessive reliance on five types of academic evaluation criteria: publications, talent-program designations, professional titles, educational credentials, and awards (Gao & Wang, 2024; Xu et al., 2021). Accordingly, the cases provide a baseline from an earlier phase of Chinese performance governance rather than a cross-section of the current system. Subsequent reforms increase the value of making that temporal boundary explicit: they permit the findings to be used as a theoretically informed point of comparison for examining change and continuity in how central evaluation policies are enacted within universities.

The three universities were purposively selected to vary in administrative affiliation, geographic location, institutional orientation, and Project 211/985 status. Pseudonyms are used throughout. University Ocean was a highly ranked, research-oriented comprehensive university included in both Project 211 and Project 985. University Garden was a science-oriented Project 211 university in a developed eastern coastal province. University Mountain was a Project 211 university in western China with an emphasis on science and engineering. Recruitment began through each university's Office of Faculty Affairs. The final sample comprised 22 participants: 19 deans or associate deans (six from University Ocean, five from University Garden, and eight from University Mountain) and three university-level officials who provided overviews of institutional PFP policy. University- and college-level PFP documents were collected to contextualize and corroborate the interview accounts.

All interviews were conducted in person in Chinese during late 2014 and lasted approximately 40–60 min. A semi-structured protocol addressed PFP design and implementation, the discretion available to colleges, and administrators' assessments of implementation and outcomes. Interviews were digitally recorded and transcribed for analysis. The institutional descriptions and findings refer to conditions at the time of data collection. One college lacked information on college size; because fsQCA requires complete membership scores for all conditions in a truth table, that case was retained in the descriptive matrix but excluded from the configurational analyses, yielding an analytic set of 18 cases.

2.2. Qualitative Thematic Analysis

The analysis proceeded in two complementary stages. First, thematic analysis was used to identify how participants understood college discretion, implementation processes,

and PFP outcomes. Second, fsQCA examined how the conditions identified through the qualitative analysis and prior theory combined in cases with high perceived effectiveness and with its absence.

All 22 interview transcripts were analyzed through an iterative thematic coding process supported by ATLAS.ti 8.5 (Braun & Clarke, 2006; Patton, 2002). Initial open codes were assigned to passages concerning college discretion, implementation, and perceived consequences of PFP. Codes were subsequently compared, merged, divided, or refined; related codes were grouped into candidate themes and reviewed against the interview material before themes were defined. Coding was conducted by two researchers. Both researchers first independently coded a subset of transcripts and then compared their coding decisions, discussed interpretive differences, and refined the coding framework. They subsequently coded the remaining transcripts, cross-checked one another's coding, and resolved discrepancies through discussion and consensus. Policy documents were used to contextualize institutional rules and compare formal requirements with participants' accounts of college-level practice.

2.3. Configurational Analysis

In the second stage, fuzzy-set qualitative comparative analysis (fsQCA), implemented in fsQCA 3.0 (Ragin & Davey, 2016), was used to examine configurations associated with high perceived effectiveness of college PFP. FsQCA is suitable for small- and intermediate-N studies concerned with conjunctural causation, equifinality, and causal asymmetry (Ragin, 2008b; Thomann & Maggetti, 2020). It therefore allows managerial discretion, financial discretion, implementation satisfaction, and college size to be analyzed as combinations rather than as independent net effects. The outcome is administrators' perceived effectiveness of PFP, not an independently observed measure of faculty productivity. The complementary analysis modeled the negation of high perceived effectiveness rather than a separately measured construct of ineffectiveness.

The analysis was condition-oriented and substantively interpretable: it examined a theoretically specified set of conditions while using qualitative case knowledge to interpret the resulting configurations (Thomann & Maggetti, 2020). In set-theoretic terms, a condition may contribute to an outcome only in combination with other conditions, and different combinations may be sufficient for the same outcome. Accordingly, the reported pathways are configurational set relations within the cases studied; they should not be read as causal effects identified through experimental or longitudinal designs.

The study followed a primarily deductive logic in selecting the broad concepts of organizational discretion and implementation context, while using the thematic analysis to refine their meaning and operationalization in the cases. This alignment between theory, qualitative evidence, and set calibration is central to the study's mixed qualitative-configurational design.

Perceived PFP effectiveness was measured by asking the participating college-level administrator representing each college to rate, on a scale from 1 (extremely ineffective) to 10 (extremely effective), the overall effects of PFP in the college, including perceived consequences for faculty motivation, performance, satisfaction, and college performance. This measure captures the assessment of the administrator responsible for college-level implementation. It is substantively relevant because implementers' judgments can shape continued use, adaptation, legitimacy, and organizational support for an appraisal system. Such perceived or symbolic performance is organizationally consequential even though it remains analytically distinct from independently measured substantive performance (Yang & Nowell, 2021). Accordingly, the measure should not be interpreted as an objective indicator of productivity or as evidence of faculty consensus.

Four conditions were included based on the qualitative themes and relevant theory: managerial discretion in PFP, financial discretion in PFP, satisfaction with PFP implementation, and college size. Managerial discretion was defined as a college's practical latitude to design and administer its PFP system. Financial discretion was defined as its latitude to allocate funds or generate revenues for performance rewards (Andrews et al., 2011; Bozeman & Moulton, 2011; Meier, 1980). Interview passages were coded and compared across cases, after which explicit criteria were developed for assigning each college an initial qualitative score of 0, 0.2, 0.4, 0.6, 0.8, or 1 on each discretion dimension. These initial scores served as raw values in the subsequent direct-calibration procedure. Appendix A reports the criteria used to assign them.

Implementation satisfaction was measured using the participating college-level administrator's rating, on a scale from 1 to 10, of the extent to which PFP was implemented consistently, systematically, and equitably in the college. This rating focused on the implementation process, rather than the perceived consequences for faculty motivation, performance, satisfaction, and college performance captured by the effectiveness measure described above. College size was measured as the number of full-time faculty members reported on college websites. Size was included as a contextual condition because larger units generally face greater demands for coordination, communication, representation, and accommodation of heterogeneous disciplinary or faculty needs.

For perceived PFP effectiveness, implementation satisfaction, college size, and the two discretion conditions, final fuzzy-set membership scores were generated through direct calibration (Ragin, 2008a). Perceived effectiveness and implementation satisfaction were initially measured using 1–10 ratings, and college size was measured as the number of full-time faculty members. Managerial and financial discretion were first assigned qualitative scores of 0, 0.2, 0.4, 0.6, 0.8, or 1 using the substantively defined coding criteria reported in Appendix A. These initial discretion scores were then treated as raw values in the direct-calibration procedure. For all five measures, three calibration anchors specified the raw values associated with full membership (0.95), the crossover point (0.50), and full non-membership (0.05), respectively. These case-referenced anchors were informed by the observed measurement ranges rather than calculated as empirical percentiles; Table 1 reports the corresponding calibration anchors, and Table 2 reports the resulting fuzzy-set membership scores. In the absence of established external benchmarks, we selected case-referenced calibration anchors to distinguish relatively low, intermediate, and high values within the observed ranges, with crossover points near their midpoints and outer anchors toward either end. Accordingly, membership in the high sets is interpreted relative to the observed cases rather than as an externally validated absolute standard. Given the context-specific nature of these choices, we assessed the sensitivity of the findings to alternative calibration anchors (see Appendix B for specifications and results).

Table 1. Direct-Calibration Anchors.

	Full Membership Anchor (0.95)	**Crossover Anchor (0.50)**	**Full non-Membership Anchor (0.05)**
Perceived effectiveness of PFP	9.325	7.75	6.175
PFP Implementation satisfaction	9.75	8.5	7.15
College Size	295.45	173.5	48.55
Managerial discretion in PFP	0.58	0.39	0.18
Financial discretion in PFP	0.95	0.5	0.05

Table 2. Final Fuzzy-Set Membership Scores after Direct Calibration.

	Conditions				Outcome
Case	**I**	**S**	**M**	**F**	**O**
1	0.77	0.12	0.96	0.03	0.62
2	0.25	0.53	0.06	0.03	0.92
3	0.25	0.06	0.54	0.03	0.03
4	0.77	0.08	0.54	0.12	0.62
5	0.25	0.04	0.54	0.03	0.92
6	0.97	0.44	0.54	0.03	0.92
7	0.77	0.07	0.54	0.12	0.62
8	0.77	0.22	0.96	0.12	0.62
9	0.97	0.29	0.54	0.12	0.92
10	0.97	0.04	0.54	0.34	0.62
11	0.97	0.84	0.96	0.66	0.97
12	0.77	0.58	0.06	0.03	0.92
13	0.92	0.65	0.06	0.03	0.19
14	0.97	0.35	0.96	0.97	0.92
15	0.77	0.26	0.06	0.03	0.38
16	0.03	0.41	0.06	0.03	0.92
17	0.97	—	0.06	0.34	0.62
18	0.10	0.27	0.54	0.12	0.03
19	0.97	0.97	0.96	0.03	0.19

Note. I = implementation satisfaction; S = college size; M = managerial discretion in PFP; F = financial discretion in PFP; O = perceived PFP effectiveness. The em dash indicates missing data. The managerial- and financial-discretion scores shown are final calibrated membership scores, not the initial qualitative scores reported in Appendix A.

We first assessed whether any single condition was necessary for either high perceived effectiveness or its absence. We then constructed truth tables using implementation satisfaction, college size, managerial discretion, and financial discretion. Because one case lacked college-size information, the truth-table analyses used 18 complete cases. No individual condition or its negation met the 0.90 necessity consistency threshold for high perceived effectiveness. For its absence, low financial discretion was the only condition exceeding this threshold (consistency = 0.979; coverage = 0.432). However, because low financial discretion was prevalent across the sampled colleges, this finding should be interpreted cautiously.

Truth-table analysis was conducted separately for high perceived effectiveness and its negation. A frequency threshold of one case was appropriate for the small-N design. Consistency cutoffs of 0.818 for high perceived effectiveness and 0.8476 for its negation were selected based on observed gaps in the respective ordered truth-table consistency distributions. The lowest included truth-table row consistencies were 0.818182 and 0.847682, respectively. To assess the sensitivity of the findings, we repeated the analyses using an alternative calibration specification and, separately, a lower consistency threshold of 0.80 with the original calibration (see Appendix B for specifications and results). Complex solutions are reported because they rely only on empirically observed configurations and do not incorporate simplifying assumptions based on logical remainders. In the present analyses, the complex and intermediate solutions were identical for both high perceived effectiveness and its negation.

3. Results

3.1. Qualitative Findings: Formal Rules and De Facto College Discretion

All 19 colleges used PFP as part of faculty performance management, although the share of compensation linked to performance varied: approximately 20% in University Ocean colleges, 30–40% in University Mountain colleges, and 10–40% in University Garden colleges. Each university issued an overarching policy that served as a template for its colleges. Although colleges were required to implement PFP and follow core university requirements, they possessed varying latitude to specify evaluation criteria, assign weights to teaching, research, and service, determine which activities qualified, and decide how performance information affected rewards. Colleges generally submitted their plans for periodic university approval, but implementation and outcomes were not routinely monitored. Thus, a common formal framework coexisted with meaningful variation in college-level enactment.

Table 3 summarizes the locus of control over major elements of PFP. University administrations generally determined whether PFP was required, the evaluation cycle, broad implementation steps, and the approximate share of compensation tied to performance. Colleges had varying discretion over criteria and weights, process customization, staffing of evaluation activities, and reward forms. This division of authority created the possibility that managerial and financial discretion would not move together.

Participants identified five recurring features they associated with effective PFP: sufficient reward resources, adaptation to disciplinary and college needs, faculty participation and support, attention to the diversity of faculty work, and consistent and equitable implementation. Although each university established the basic structure and requirements, colleges differed in the practical latitude with which they implemented those requirements. This de facto discretion—the latitude colleges could exercise in practice, rather than simply the authority granted in policy—varied with revenue sources, unit size and mission, collegial culture, prior performance, and relations with university administrators (Bach, 2014; Verhoest et al., 2004). The variation across colleges provided the empirical basis for examining discretion configurationally.

Two distinct dimensions emerged from the interviews. Managerial discretion concerned a college's ability to design and administer PFP, including assigning personnel, specifying procedures and criteria, weighting performance dimensions, and linking assessments to rewards. Financial discretion concerned its ability to allocate or generate funds for performance rewards. Participants associated insufficient managerial discretion with poorly fitting standards, administrative burden, short-termism, and interpersonal conflict; they associated insufficient financial discretion with rewards too limited to differentiate or motivate performance. The dimensions were also interdependent. As one dean from University Garden stated, "The decentralization of power is only possible with the decentralization of financial resources; without the shift of resources from the university to the colleges, we [at the college level] are just going through the motions and meeting the imposed requirements."

Formal discretion did not necessarily translate into substantial local adaptation. Although colleges were nominally permitted to modify criteria, weights, incentives, and procedures, few departed markedly from university templates. Deans described the work and political risk involved in reconciling university requirements with faculty expectations. One explained that an ambitious dean could redesign the system, but a dean wishing to avoid blame, complaints, and additional workload could simply reproduce the university guidelines. This pattern illustrates the distinction between formal and de facto autonomy: decision rights may exist on paper but remain difficult to exercise without resources, support, and perceived legitimacy (Bach, 2014; Verhoest et al., 2004).

Table 3. Summary of Locus of PFP Control between University and College Administration.

	University Ocean	**University Mountain**	**University Garden**
Number of colleges interviewed	6	8	5
University Orientation	Research-oriented	Research-oriented	Teaching-oriented
Percentage of compensation tied to PFP	Approximately 20%	30–40%	10–40%
PFP required?	Yes	Yes	Yes
University–college division of PFP control	The university administration provides general guidelines and requires all colleges to design and implement their own annual PFP plans. Colleges must submit their PFP plans to the university for approval; these plans specify the implementation processes and performance criteria. Colleges have moderate discretion to develop specific performance criteria and determine the weight assigned to each criterion. Colleges also have discretion to determine how performance information is used to reward faculty, subject to the resources available to each college.	The university administration uses a contract-based approach to implement PFP at the college level. Under this approach, the university administration establishes annual goals and objectives for each college, signs a contract with each college dean, and provides bonus funding that colleges may use to reward faculty performance. Colleges have moderate discretion to develop specific criteria, assign weights to those criteria, and determine how performance is rewarded within the general guidelines established by the university.	The university administration provides general guidelines and requires all colleges to design and implement their own annual PFP plans. Colleges must submit their PFP plans to the university for approval; these plans specify the implementation processes and performance criteria. Colleges have moderate-to-high discretion to develop specific performance criteria, assign weights to those criteria, and determine how performance is rewarded within the general guidelines established by the university.

Participants also identified college size as an important contextual condition. Larger colleges faced greater complexity in representing diverse faculty interests, organizing participation, communicating standards, and coordinating decisions. One participant explained: "We have more faculty than other colleges in our university; some colleges can include all their faculty on the College Faculty Committee [for PFP], [but] we simply cannot.... A small family is easy to manage.... It is much harder to have 30 people meet at the same time [to discuss PFP standards].... An effective PFP system won't be possible without sufficient discussion with faculty." The account illustrates why the same degree of discretion may impose different implementation demands across units.

Implementation quality was the fourth recurring theme. Participants emphasized how standards were developed, whether representative committees were used, how extensively faculty were involved, how clearly decisions were communicated, and whether procedures were perceived as equitable and transparent. These accounts support treating implementation satisfaction as a condition that may enable or constrain the productive use of discretion rather than merely as a downstream reaction to PFP.

3.2. Configurational Results

The fsQCA examined which combinations of managerial discretion, financial discretion, implementation satisfaction, and college size were associated with high perceived PFP effectiveness and its negation. The analysis used the 18 colleges with complete information on all four conditions. Table 4 reports three pathways to high perceived effectiveness, and Table 5 reports one pathway to its absence.

Table 4. Configurations Associated with High Perceived PFP Effectiveness.

	Path 1	Path 2	Path 3
Satisfactory PFP implementation	●		●
Large college size	⊗	●	
High managerial discretion in PFP		⊗	●
High financial discretion in PFP	⊗	⊗	●
Consistency	0.81	0.81	0.96
Raw coverage	0.54	0.28	0.24
Unique coverage	0.29	0.07	0.12
Cases	7, 1, 8, 15, 9, 10, 6	13, 12, 2	14, 11
Overall solution consistency		0.81	
Overall solution coverage		0.74	

Note. Frequency cutoff = 1; truth-table consistency threshold = 0.818. The lowest included truth-table row consistency was 0.818182. Consistency and coverage statistics are rounded to two decimal places. Listed cases exhibit each configuration; their inclusion does not necessarily indicate high membership in the outcome. Blank cell = condition not specified in the configuration. ● = Presence of condition. ⊗ = Absence of condition.

Table 5. Configuration Associated with the Absence of High Perceived PFP Effectiveness.

	Path 1
Satisfactory PFP implementation	⊗
Large college size	⊗
High managerial discretion in PFP	●
High financial discretion in PFP	⊗
Consistency	0.85
Raw coverage	0.38
Unique coverage	0.38
Cases	18, 5, 3
Overall solution consistency	0.85
Overall solution coverage	0.38

Note. Frequency cutoff = 1; truth-table consistency threshold = 0.8476. The lowest included truth-table row consistency was 0.847682. Consistency and coverage statistics are rounded to two decimal places. Listed cases exhibit the configuration; their inclusion does not necessarily indicate high membership in the negated outcome. Blank cell = condition not specified in the configuration. ● = Presence of condition; ⊗ = Absence of condition.

The solution for high perceived effectiveness had an overall consistency of 0.81 and overall coverage of 0.74. In set-theoretic terms, consistency indicates the extent to which cases displaying a configuration also display the outcome, whereas coverage indicates the proportion of the outcome represented by a configuration (Ragin, 2008b). The truth-table consistency threshold (0.818182 for the lowest included row) and overall solution consistency (0.810628) represent different calculations; therefore, the slightly lower solution consistency does not indicate an inconsistency in the analysis. Pathway 1 had the greatest empirical coverage, but the three pathways represent alternative sufficient configurations rather than competing estimates of a single net effect.

Pathway 1 combined satisfactory implementation, small college size, and low financial discretion; managerial discretion was not specified in the solution. Seven cases exhibited this configuration (Cases 7, 1, 8, 15, 9, 10, and 6). The interviews suggest that smaller colleges

faced fewer coordination demands and could implement relatively straightforward systems even when reward resources were constrained. Under these conditions, satisfactory and accepted procedures may have mattered more to favorable assessments than local control over funding. Participants in several small colleges associated PFP with increased attention to visible tasks, publications, or teaching loads. These are the participating college-level administrators' perceptions and should not be treated as independently verified changes in faculty behavior.

Pathway 2 combined large college size with low managerial and low financial discretion; implementation satisfaction was not specified. Three cases exhibited this configuration (Cases 13, 12, and 2). This configuration suggests that limited college-level discretion may coexist with favorable administrator assessments in some larger colleges, although the pathway warrants caution given its limited coverage and sensitivity to calibration choices. Qualitative accounts suggest one possible mechanism: in large and heterogeneous colleges, university-wide rules may reduce the burden of negotiating numerous interests, provide common standards, and locate responsibility for contested design choices at the university level. This interpretation does not imply that centralization is generally superior; rather, it shows that limited discretion may be compatible with favorable assessments under a particular organizational configuration.

Pathway 3 combined high managerial discretion, high financial discretion, and satisfactory implementation; college size was not specified. Two cases exhibited this configuration (Cases 14 and 11). Here, authority to customize evaluation was accompanied by control over reward resources and a satisfactory implementation process. The qualitative evidence suggests that this alignment enabled colleges to adapt PFP to disciplinary and organizational circumstances. At the same time, the pathway appeared in only two cases and likely demanded substantial leadership, consultation, and administrative capacity; it should therefore be interpreted as a theoretically meaningful configuration rather than a universally preferred model.

The analysis of the negated outcome identified one pathway to the absence of high perceived effectiveness, with a consistency of 0.85 and raw and unique coverage of 0.38: small college size, high managerial discretion, low financial discretion, and unsatisfactory implementation. Three cases exhibited this configuration (Cases 18, 5, and 3). Case 3 illustrates the pattern. College leaders had responsibility for adapting PFP but insufficient financial latitude to differentiate rewards, while the required process added administrative work and was perceived as unsatisfactory. The combination suggests a responsibility–capacity mismatch: local leaders possessed managerial latitude and exposure to blame without corresponding resources or an accepted implementation process.

Sensitivity analyses examined the stability of these findings under alternative calibration anchors and a lower truth-table consistency threshold (see Appendix B for detailed specifications and results). With alternative calibration, two of the three configurations associated with high perceived effectiveness retained their original structure, while the third additionally required low implementation satisfaction. Overall solution consistency increased from 0.811 to 0.834, and coverage decreased from 0.741 to 0.721. However, no configuration met the original consistency threshold for the absence of high perceived effectiveness under the alternative calibration. Lowering the truth-table consistency threshold to 0.80 while retaining the original calibration produced no changes in either solution. These findings indicate partial configurational stability, with the results for the absence of high perceived effectiveness particularly sensitive to calibration choices.

4. Discussion

This study examined how organizational discretion and implementation conditions combined in administrators' perceptions of faculty PFP across public university colleges. Its central finding is not that more or less autonomy is uniformly preferable. Rather, managerial discretion, financial discretion, implementation satisfaction, and college size formed different configurations associated with high perceived effectiveness, while a distinct configuration characterized its absence. The results therefore recast discretion from a single organizational advantage into a relational question of fit among authority, resources, process, and complexity.

The first contribution is to higher education governance and faculty performance management. Existing research has shown that quantified performance systems can reshape academic motivation, identity, and work priorities (Kallio & Kallio, 2014; Kallio et al., 2016; Sułkowski et al., 2020; ter Bogt & Scapens, 2012). This study adds that administrators' assessments of PFP vary across colleges with different combinations of discretion and implementation conditions. Colleges operating under the same institutional policy may not possess the same practical capacity to translate evaluation criteria into legitimate procedures and meaningful rewards. The academic unit is therefore not merely an administrative conduit; it is a consequential site at which performance policy is interpreted and enacted.

The second contribution is conceptual. Distinguishing managerial from financial discretion reveals why broad measures of university autonomy may conceal consequential mismatches. Managerial authority can support local adaptation when it is accompanied by financial capacity and satisfactory implementation, as in Pathway 3. Yet managerial discretion combined with limited financial discretion and unsatisfactory implementation was associated with the absence of high perceived effectiveness in the original negated-outcome analysis. This contrast suggests a possible responsibility–capacity interpretation: delegation may disappoint when it transfers responsibility for contested decisions without the resources or process legitimacy needed to carry them out.

Third, the results demonstrate the value of configurational reasoning for heterogeneous academic units. Pathways 1 and 2 show that limited discretion can coexist with favorable administrator assessments under particular conditions. In smaller colleges, satisfactory implementation may compensate for limited financial latitude; in larger colleges, centralized rules may provide coordination and a defensible common standard. These findings do not establish that constraints improve performance. Instead, they challenge the assumption that autonomy has a constant positive effect and identify combinations that warrant testing in larger and longitudinal studies. The original solutions also suggest causal asymmetry: the absence of high perceived effectiveness is not necessarily explained by simply reversing a favorable configuration. The sensitivity analyses further indicate that two favorable pathways retained their structure under alternative calibration, whereas the remaining configurations were more sensitive to calibration choices (Appendix B).

The findings also connect performance appraisal to its social context. Fairness, participation, communication, and evaluator–employee relations influence how appraisal systems are received (Levy & Williams, 2004). In this study, implementation satisfaction captured part of that context and repeatedly conditioned the meaning of discretion. Authority over evaluation design is unlikely to generate favorable assessments when procedures are experienced as inconsistent or inequitable; conversely, accepted procedures can make a constrained system more workable. Performance-system reform should therefore align decision rights with resources and implementation capacity rather than treating formal decentralization as an outcome in itself.

The historical setting sharpens rather than eliminates this theoretical contribution. The cases document how faculty PFP operated during the Project 211/985 era, before Double

First-Class and Breaking Five-Only reforms altered the national evaluation environment. Later research shows both continuing reliance on competitive and output-oriented evaluation and efforts to diversify criteria and reduce narrow metric dependence (Gao & Wang, 2024; Gu & Levin, 2021; Xu et al., 2021). The precise rules observed in 2014 should not be generalized to the present. However, the implementation problem revealed here remains relevant whenever higher-level reforms delegate interpretation to colleges while controlling key resources or accountability expectations. The findings offer a baseline for asking whether later reforms changed not only formal criteria but also the alignment of authority, resources, and implementation processes within universities.

For university leaders, the results suggest that decisions about decentralization should be made by function rather than through a blanket grant or restriction of autonomy. If colleges are expected to customize faculty evaluation, they also require sufficient control over rewards, clear procedural safeguards, and capacity for consultation and communication. Where financial discretion cannot be delegated, universities should avoid creating the appearance that college leaders control outcomes they cannot materially influence. Large colleges may need additional representative structures and administrative support, whereas smaller units may benefit most from transparent and proportionate processes. These implications concern organizational design and perceived implementability; the study does not establish which arrangement maximizes objective faculty or educational performance.

5. Limitations

Several limitations bound the claims. First, the evidence was collected in 2014 and cannot establish whether the same configurations characterize Chinese public universities today. The data are therefore presented as a historically situated baseline from the Project 211/985 period, not as a current-policy snapshot. Its continuing value lies in documenting the organizational mechanisms through which central rules were translated into college practice before major subsequent reforms. Direct comparison with post-Double First-Class and post-Breaking Five-Only cases is needed to determine which relationships persisted, disappeared, or changed. Second, the study includes 19 colleges from three purposively selected universities, with 18 complete cases in the fsQCA. The findings support context-bound configurational interpretation and theory development, not statistical generalization to Chinese or international higher education.

Third, the outcome rating and implementation-satisfaction rating were provided by the participating college-level administrators, who were themselves involved in administering PFP. Their assessments are consequential for policy implementation but may reflect role-based interests, self-presentation, or incomplete knowledge of faculty experience. Because both ratings came from the same administrators, common-method bias may also contribute to their association. The study therefore concerns perceived effectiveness and should not be read as evidence of objective productivity gains or faculty endorsement. Future studies should triangulate administrator, faculty, and staff perspectives with longitudinal indicators of teaching, research, service, turnover, and well-being. Fourth, the discretion sets were coded from qualitative evidence, and fuzzy-set membership scores were calibrated using case-referenced anchors rather than independently established external benchmarks. The sensitivity analyses showed that two favorable configurations retained their structure under alternative calibration, whereas the third favorable configuration changed and the original negated-outcome configuration no longer met the specified consistency threshold (Appendix B). These results warrant caution in treating individual configurations as stable empirical patterns beyond the cases studied. Finally, the cross-sectional design identifies set-theoretic associations, not temporal or causal effects. Longitudinal and comparative

designs could test the proposed responsibility–capacity mechanism and assess whether it travels across institutional and national contexts.

Future research can build on these findings in several complementary ways. Most immediately, the evidence provides a pre-reform baseline for examining organizational change in Chinese higher education. Future studies could replicate the analysis in the same or comparable colleges to assess whether Double First-Class and Breaking Five-Only reforms changed the alignment among managerial authority, financial resources, and implementation capacity. Such comparisons could reveal which configurations persisted, disappeared, or emerged after the reforms. Multilevel and longitudinal designs could combine the perspectives of university administrators, college leaders, faculty, and professional staff with indicators of teaching, research, service, turnover, and well-being. Such evidence would help determine whether configurations associated with favorable administrative assessments also generate outcomes valued by faculty members and universities. Research involving larger and more diverse case populations could further assess the stability of these configurations using independently justified calibration criteria and examine whether similar configurations arise across institutional types and national governance systems. Finally, process-oriented comparative case studies could investigate the proposed responsibility–capacity mismatch more directly, tracing how delegating managerial authority without corresponding financial resources or implementation legitimacy influences policy acceptance and performance over time.

6. Conclusions

This study shows that understanding the perceived effectiveness of faculty pay-for-performance requires attention not only to incentive design but also to the organizational conditions of implementation. Across the colleges studied, favorable assessments arose through multiple configurations of managerial discretion, financial discretion, implementation satisfaction, and college size. The original analysis also identified a configuration associated with the absence of high perceived effectiveness: high managerial discretion, limited financial discretion, low implementation satisfaction, and smaller college size. Organizational discretion is therefore neither inherently beneficial nor inherently constraining. Its value depends on whether decision-making authority is matched by adequate resources, workable implementation processes, and the organizational capacity to exercise it effectively. The central implication is not that universities should uniformly centralize or decentralize faculty PFP, but that they should align responsibility with capacity at the academic-unit level. As a historically situated baseline from an earlier period of Chinese higher education reform, the study also provides a configurational framework for assessing subsequent institutional change. Ultimately, formally similar faculty-evaluation policies may be perceived differently across academic units, depending not simply on how much discretion colleges possess but on how authority, resources, processes, and organizational complexity fit together.

Author Contributions: Conceptualization, Z.Y. and S.X.; methodology, Z.Y. and S.X.; software, Z.Y.; validation, Z.Y.; formal analysis, Z.Y.; investigation, Z.Y. and S.X.; resources, Z.Y. and S.X.; data curation, Z.Y. and S.X.; writing—original draft preparation, Z.Y. and S.X.; writing—review and editing, Z.Y. and S.X.; visualization, Z.Y.; supervision, Z.Y.; project administration, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was reviewed and determined exempt by the Institutional Review Board of North Carolina State University (IRB No. 3002; 26 November 2014; exemption under 45 CFR 46.101(b)(2)).

Informed Consent Statement: Informed consent for participation was obtained from all participants involved in the study.

Data Availability Statement: The underlying interview data are not publicly available because of confidentiality commitments made to participants and the risk of deductive disclosure in this small, institutionally bounded sample. The de-identified analytical information supporting the reported configurational results is presented in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Criteria for Assigning Initial Qualitative Scores to Managerial and Financial Discretion in PFP

The appendix presents the criteria used to assign initial qualitative scores to managerial and financial discretion based on interview evidence, prior to their direct calibration into fuzzy-set membership scores.

Condition	Initial Qualitative Score	Coding Criteria
Financial discretion in PFP	0	The college reported no independently generated funding sources and no discretion over allocating funds or generating revenue for PFP purposes.
	0.2	The college had very limited discretion over allocating funds or generating revenue for PFP. The university controlled or regulated most funding decisions.
	0.4	The college had some discretion over allocating funds or generating revenue for PFP, but the university retained primary control over funding decisions.
	0.6	The college had moderate discretion over allocating funds or generating revenue for PFP and could independently determine the allocation of more than half of its PFP funding.
	0.8	The college had substantial discretion over allocating funds or generating revenue for PFP. It had relatively diverse funding sources and generated a substantial share of its revenue independently.
	1	The college had complete discretion over allocating funds or generating revenue for PFP and could independently determine how its own revenue was used.
Managerial discretion in PFP	0	The college had no discretion over the design or implementation of its PFP system. It was required to follow university policies and guidelines without modification.
	0.2	The college had very limited discretion over the design and implementation of its PFP system. It primarily executed university policies and could make only minimal adjustments to accommodate disciplinary needs.
	0.4	The college had some discretion over the design and implementation of its PFP system but was still required to follow most university policies and guidelines. It could adjust selected elements, such as the weights assigned to different performance indicators.
	0.6	The college had moderate discretion over the design and implementation of its PFP system. It could introduce meaningful modifications or innovations based on its organizational and disciplinary needs and played an active role in managing the system.
	0.8	The college had substantial discretion over the design and implementation of its PFP system, including decisions about evaluation criteria, the weights assigned to different indicators, and how evaluation results were linked to faculty compensation.
	1	The college had complete discretion over the design, implementation, and use of its PFP system and evaluation results.

Appendix B. fsQCA Sensitivity Analyses

The sensitivity analyses used the 18 colleges with complete data and the same four conditions as the main analysis: implementation satisfaction (I), college size (S), managerial discretion (M), and financial discretion (F). The outcomes were high perceived PFP effectiveness (O) and its set-theoretic negation (~O). All analyses used a truth-table frequency threshold of one case and report complex solutions.

Appendix B.1. Calibration Specifications

The original calibration anchors are reproduced from Table 1 of the manuscript. In the alternative specification, the upper and lower anchors were placed at 90% and 10% of each variable's observed raw-data range, while the original crossover anchors were retained. Both specifications used direct calibration, with the upper, crossover, and lower anchors corresponding to fuzzy-set membership scores of 0.95, 0.50, and 0.05, respectively. The resulting membership values used in the truth-table analyses were recorded to two decimal places.

Table A1. Original and Alternative Direct-Calibration Anchors.

Measure	Original Upper	Original Crossover	Original Lower	Alternative Upper	Alternative Crossover	Alternative Lower
Perceived effectiveness	9.325	7.75	6.175	9.15	7.75	6.35
Implementation satisfaction	9.75	8.50	7.15	9.70	8.50	7.30
College size (faculty)	295.45	173.50	48.55	281.90	173.50	65.10
Managerial discretion	0.58	0.39	0.18	0.56	0.39	0.24
Financial discretion	0.95	0.50	0.05	0.90	0.50	0.10

Note. Upper, crossover, and lower refer to raw-data anchor values mapped to fuzzy-set memberships of 0.95, 0.50, and 0.05. The 10% and 90% positions used for the alternative specification are proportions of the observed ranges, not empirical percentile ranks. The original anchors were not calculated as empirical 5th/50th/95th percentiles.

Appendix B.2. Sensitivity of Configurational Solutions

The original truth-table consistency cutoffs were 0.818 for O and 0.8476 for ~O. Two checks were performed: (a) alternative calibration while keeping the original cutoffs, and (b) lowering the consistency cutoff to 0.80 while retaining the original calibration.

Table A2. fsQCA Results across Six Analytic Specifications.

Run	Calibration	Outcome	Applied Cutoff	Configuration(s)	Solution Consistency	Solution Coverage
A1	Original	O	0.818	I*~S*~F; S*~M*~F; I*M*F	0.810628	0.740512
A2	Original	~O	0.8476	~I*~S*M*~F	0.847682	0.383808
B1	Alternative	O	0.818	I*~S*~F; I*M*F; ~I*S*~M*~F	0.833669	0.721011
B2	Alternative	~O	0.8476	No qualifying configuration	—	—
C1	Original	O	0.80	Same as A1	0.810628	0.740512
C2	Original	~O	0.80	Same as A2	0.847682	0.383808

Note. The asterisk (*) denotes logical conjunction ("AND"), and the tilde (~) denotes set negation (absence of a condition or outcome). The 0.80 cutoff was evaluated only with the original calibration. In B2, the highest observed truth-table row consistency was 0.830389, below the applied 0.8476 cutoff; hence there was no qualifying positive row for ~O. Cutoffs apply to truth-table row selection, not to the consistency of every minimized pathway or of the overall solution. The lowest qualifying row consistencies were 0.818182 (A1/C1), 0.847682 (A2/C2), and 0.829114 (B1).

Table A3. Pathway-Level Consistency and Coverage under Alternative Calibration (Run B1).

Configuration	Consistency	Raw Coverage	Unique Coverage
I*~S*~F	0.807388	0.533566	0.374019
I*M*F	0.970909	0.232781	0.127289
~I*S*~M*~F	0.895238	0.163906	0.060157

Note. The asterisk (*) denotes logical conjunction ("AND"), and the tilde (~) denotes set negation (absence of a condition or outcome).

Appendix B.3. Interpretation of Sensitivity Findings

Under alternative calibration, two of the three configurations associated with high perceived effectiveness retained their original condition structure (I*~S*~F and I*M*F). The original S*~M*~F pathway became ~I*S*~M*~F, indicating that the alternative solution additionally specified the absence of high implementation satisfaction. The overall consistency of the favorable solution increased from 0.810628 to 0.833669, while coverage decreased from 0.740512 to 0.721011. The original pathway for ~O did not meet the prespecified consistency cutoff under alternative calibration. Lowering the threshold to 0.80 with the original calibration yielded unchanged solutions for both outcomes because no additional observed truth-table rows met that threshold. These results indicate partial stability of the favorable configurations and greater calibration sensitivity for the configuration associated with the absence of high perceived effectiveness.

References

Abdullah, Z., & Malik, S. K. (2022). Perceived appraisal fairness effect on performance appraisal effectiveness and appraisal satisfaction in higher education. *Journal of Business and Social Review in Emerging Economies*, *8*(2), 287–300. [CrossRef]

Al-Omari, A. A. (2007). The organization of academic departments and participation in decision making as perceived by faculty members in Jordanian universities. *International Studies in Educational Administration*, *35*(2), 82–100.

Andrews, R., Boyne, G. A., & Walker, R. M. (2011). Dimensions of publicness and organizational performance: A review of the evidence. *Journal of Public Administration Research and Theory*, *21*(Suppl. 3), i301–i319. [CrossRef]

Bach, T. (2014). The autonomy of government agencies in Germany and Norway: Explaining variation in management autonomy across countries and agencies. *International Review of Administrative Sciences*, *80*(2), 341–361. [CrossRef]

Ballou, D. (2001). Pay for performance in public and private schools. *Economics of Education Review*, *20*(1), 51–61. [CrossRef]

Birdsall, C. (2018). Performance management in public higher education: Unintended consequences and the implications of organizational diversity. *Public Performance & Management Review*, *41*(4), 669–695. [CrossRef]

Bozeman, B., & Moulton, S. (2011). Integrative publicness: A framework for public management strategy and performance. *Journal of Public Administration Research and Theory*, *21*(Suppl. 3), i363–i380. [CrossRef]

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. [CrossRef]

Chen, C.-A. (2018). "A little is better than zero" or "pay enough or don't pay at all"? Evidence on the size of pay-for-performance across the sectors. *Public Personnel Management*, *47*(2), 119–143. [CrossRef]

Dasanayaka, C. H., Abeykoon, C., Ranaweera, R. A. A. S., & Koswatte, I. (2021). The impact of the performance appraisal process on job satisfaction of the academic staff in higher educational institutions. *Education Sciences*, *11*(10), 623. [CrossRef]

Deem, R. (1998). 'New managerialism' and higher education: The management of performances and cultures in universities in the United Kingdom. *International Studies in Sociology of Education*, *8*(1), 47–70. [CrossRef]

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160. [CrossRef]

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, *37*(4), 696–727. [CrossRef]

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, *14*(1), 57–74. [CrossRef]

Enders, J., de Boer, H., & Weyer, E. (2013). Regulatory autonomy and performance: The reform of higher education re-visited. *Higher Education*, *65*(1), 5–23. [CrossRef]

Ethiraj, S. K., & Levinthal, D. (2009). Hoping for A to Z while rewarding only A: Complex organizations and multiple goals. *Organization Science*, *20*(1), 4–21. [CrossRef]

Frey, B. S., Homberg, F., & Osterloh, M. (2013). Organizational control systems and pay-for-performance in the public service. *Organization Studies*, *34*(7), 949–972. [CrossRef]

Gao, Y., & Wang, X. (2024). When policy meets practice: Evaluating breaking five-only policy through academic production in China. *Publications*, *12*(4), 44. [CrossRef]

Gu, J., & Levin, J. S. (2021). Tournament in academia: A comparative analysis of faculty evaluation systems in research universities in China and the USA. *Higher Education*, *81*(5), 897–915. [CrossRef]

Guthrie, J., Manes-Rossi, F., Orelli, R. L., & Sforza, V. (2024). Performance management and measurement impacts on universities: (Re)viewing the past, present and future. *Journal of Public Budgeting, Accounting & Financial Management*, *36*(6), 1–25. [CrossRef]

Han, Y., & Hong, S. (2019). The impact of accountability on organizational performance in the U.S. federal government: The moderating role of autonomy. *Review of Public Personnel Administration*, *39*(1), 3–23. [CrossRef]

Hearn, J. C. (1999). Pay and performance in the university: An examination of faculty salaries. *The Review of Higher Education*, *22*(4), 391–410. [CrossRef]

Heugens, P. P. M. A. R., & Lander, M. W. (2009). Structure! Agency! (and other quarrels): A meta-analysis of institutional theories of organization. *Academy of Management Journal*, *52*(1), 61–85. [CrossRef]

Hitka, M., Ližbetinová, L., Javorčíková, J., Lorincová, S., & Vanderková, K. (2023). Managing employee motivation in Slovak universities from the perspectives of time and age. *Education Sciences*, *13*(6), 556. [CrossRef]

Holmstrom, B., & Milgrom, P. (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*, 24–52. [CrossRef]

Ingraham, P. W. (1993). Of pigs in pokes and policy diffusion: Another look at pay-for-performance. *Public Administration Review*, *53*(4), 348–356. [CrossRef]

Kallio, K.-M., & Kallio, T. J. (2014). Management-by-results and performance measurement in universities—Implications for work motivation. *Studies in Higher Education*, *39*(4), 574–589. [CrossRef]

Kallio, K.-M., Kallio, T. J., Tienari, J., & Hyvönen, T. (2016). Ethos at stake: Performance management and academic work in universities. *Human Relations*, *69*(3), 685–709. [CrossRef]

Kimball, S. M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, *16*(4), 241–268. [CrossRef]

Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *The Future of Children*, *17*(1), 87–109. [CrossRef]

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, *30*(6), 881–905. [CrossRef]

Maharvi, M. W., Kumar, A., Channa, K. A., & Mahmood, A. (2023). How leader member exchange affects effectiveness of performance appraisal system: A chain of reactions model. *Cogent Business & Management*, *10*(2), 2238392. [CrossRef]

Meier, K. J. (1980). Measuring organizational power: Resources and autonomy of government agencies. *Administration & Society*, *12*(3), 357–375. [CrossRef]

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*(2), 340–363. [CrossRef]

Mintrop, R., Órdenes, M., Coghlan, E., Pryor, L., & Madero, C. (2018). Teacher evaluation, pay for performance, and learning around instruction: Between dissonant incentives and resonant procedures. *Educational Administration Quarterly*, *54*(1), 3–46. [CrossRef]

Mok, K. H. (2009). The growing importance of the privateness in education: Challenges for higher education governance in China. *Compare: A Journal of Comparative and International Education*, *39*(1), 35–49. [CrossRef]

Mok, K. H., & Han, X. (2017). Higher education governance and policy in China: Managing decentralization and transnationalism. *Policy and Society*, *36*(1), 34–48. [CrossRef]

Nielsen, P. A. (2014). Performance management, managerial authority, and public service performance. *Journal of Public Administration Research and Theory*, *24*(2), 431–458. [CrossRef]

O'Toole, L. J., Jr., & Meier, K. J. (2015). Public management, context, and performance: In quest of a more general theory. *Journal of Public Administration Research and Theory*, *25*(1), 237–256. [CrossRef]

Park, J. (2022). What makes performance-related pay effective in the public sector? Target, pay design, and context. *Review of Public Personnel Administration*, *42*(3), 416–443. [CrossRef]

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). SAGE Publications.

Perry, J. L., Engbers, T. A., & Jun, S. Y. (2009). Back to the future? Performance-related pay, empirical research, and the perils of persistence. *Public Administration Review*, *69*(1), 39–51. [CrossRef]

Pfeffer, J., & Salancik, G. R. (2003). *The external control of organizations: A resource dependence perspective*. Stanford University Press.

Proudfoot, K. (2024). Teachers' perspectives on pay incentives in England: Performance evaluation in a context of high-stakes accountability. *Educational Assessment, Evaluation and Accountability*, *36*(4), 529–557. [CrossRef]

Ragin, C. C. (2008a). Measurement versus calibration: A set-theoretic approach. In J. M. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *The Oxford handbook of political methodology* (pp. 174–198). Oxford University Press. [CrossRef]

Ragin, C. C. (2008b). *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago Press.
Ragin, C. C., & Davey, S. (2016). *Fuzzy-set/qualitative comparative analysis* (Version 3.0) [Computer software]. Department of Sociology, University of California.
Sheikh, A. Z., Chandler, J., Hussain, B., & Timmons, S. (2022). Performance measurement and management in the British higher education sector. *Quality & Quantity*, *56*(6), 4809–4824. [CrossRef]
Shu, F., Liu, S., & Larivière, V. (2022). China's research evaluation reform: What are the consequences for global science? *Minerva*, *60*(3), 329–347. [CrossRef]
Smith, J. L., Mendez, S., Poe, J., Johnson, C., Willson, D. K., Daniels, E. A., Song, H., & Skop, E. (2024). From mandate to co-create: Leading the development of inclusive performance evaluation criteria. *Equality, Diversity and Inclusion: An International Journal*, *43*(9), 88–102. [CrossRef]
Sułkowski, Ł., Przytuła, S., Borg, C., & Kulikowski, K. (2020). Performance appraisal in universities—Assessing the tension in public service motivation (PSM). *Education Sciences*, *10*(7), 174. [CrossRef]
ter Bogt, H. J., & Scapens, R. W. (2012). Performance management in universities: Effects of the transition to more quantitative measurement systems. *European Accounting Review*, *21*(3), 451–497. [CrossRef]
Thomann, E., & Maggetti, M. (2020). Designing research with qualitative comparative analysis (QCA): Approaches, challenges, and tools. *Sociological Methods & Research*, *49*(2), 356–386. [CrossRef]
Verhoest, K., Peters, B. G., Bouckaert, G., & Verschuere, B. (2004). The study of organisational autonomy: A conceptual review. *Public Administration and Development*, *24*(2), 101–118. [CrossRef]
Weibel, A., Rost, K., & Osterloh, M. (2010). Pay for performance in the public sector—Benefits and (hidden) costs. *Journal of Public Administration Research and Theory*, *20*(2), 387–412. [CrossRef]
Xu, X. (2020). Performing under 'the baton of administrative power'? Chinese academics' responses to incentives for international publications. *Research Evaluation*, *29*(1), 87–99. [CrossRef]
Xu, X., Oancea, A., & Rose, H. (2021). The impacts of incentives for international publications on research cultures in Chinese humanities and social sciences. *Minerva*, *59*(4), 469–492. [CrossRef]
Yang, Z., & Nowell, B. (2021). Network isomorphism?: A network perspective on the symbolic performance of purpose-oriented networks. *International Public Management Journal*, *24*(3), 350–377. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.