Spring 2024

The Elements of Governance and Their Association with Aviation Safety Oversight Effectiveness

João Souza Dias Garcia
Embry-Riddle Aeronautical University, souzadij@erau.edu

The Elements of Governance and Their Association with Aviation Safety Oversight Effectiveness

João Souza Dias Garcia

Dissertation Submitted to the College of Aviation in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Aviation

Embry-Riddle Aeronautical University

Daytona Beach, Florida

March 2024

The Elements of Governance and their Association with Aviation Safety Oversight Effectiveness

By

João Souza Dias Garcia

This dissertation was prepared under the direction of the candidate's Dissertation Committee Chair, Dr. Dothang Truong, and has been approved by the members of the dissertation committee. It was submitted to the College of Aviation and was accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Aviation.

Dothang Truong Digitally signed by Dothang Truong Date: 2024.03.21 13:58:54 -04'00'

Dothang Truong, Ph.D.
Committee Chair

Alan J. Stolzer Digitally signed by Alan J. Stolzer Date: 2024.03.25 08:41:01 -04'00'

Alan J. Stolzer, Ph.D.
Committee Member

Steven Hampton Digitally signed by Steven Hampton Date: 2024.03.22 17:51:09 -04'00'

Steven Hampton, Ed.D.
Associate Dean, School of Graduate Studies, College of Aviation

Steven Hampton Digitally signed by Steven Hampton Date: 2024.03.22 17:50:44 -04'00'

Steven Hampton, Ed.D.
Committee Member

Alan J. Stolzer Digitally signed by Alan J. Stolzer Date: 2024.03.25 08:41:22 -04'00'

Alan J. Stolzer, Ph.D.
Dean, College of Aviation

Committee Member

Rafael Digitally signed by Rafael Date: 2024.03.23 11:52:42 -05'00'

Rafael Echevarne, Ph.D.
Committee Member (External)

March 19, 2024

Signature Page Date

Abstract

Researcher: João Souza Dias Garcia

Title: The Elements of Governance and their Association with Aviation Safety Oversight Effectiveness

Institution: Embry-Riddle Aeronautical University

Degree: Doctor of Philosophy in Aviation

Year: 2024

States and their respective national civil aviation authorities promote the improvement of the global civil aviation system by developing and enforcing safety regulations for aircraft design, operations, personnel training, infrastructure, and air traffic management, among other topics. For many years, the International Civil Aviation Organization (ICAO) has driven the continuous improvement of states' safety oversight functions, supporting them in evaluating the effectiveness of these functions by conducting audits under the Organization's Universal Safety Oversight Audit Program (USOAP). While the benefits of these assessments have been broadly recognized, scholarly literature on the factors commonly associated with safety oversight effectiveness across states is minimal. This study aimed to bridge this gap by drawing knowledge and understanding from the scholarly literature on public administration and the evolving governance theory and testing the relationship between state governance measures and aviation safety oversight effectiveness. Two variables were selected from the Worldwide Governance Indicators to reflect state governance measures, namely Regulatory Quality (RQ) and Government Effectiveness (GE), and the result of the ICAO safety oversight audits, as expressed in the Effective Implementation (EI) metric, is selected as a measure of safety oversight effectiveness. The research methodology included exploratory and quantitative approaches. Various multivariate quantitative analytical models, including multiple linear regression, structural equation modeling, and data mining, were developed in the exploratory dimension. The non-linear data mining approaches included random forest, deep learning, and decision tree models. The analysis supported the validation of RQ and GE's factor structure and tested the relationships between these constructs and EI. These models were compared with respect to model fit and predictive performance. This approach was complemented by quantitative analyses of the association between states' RQ and GE dimensions of governance and EI and the evaluation of their relative importance. All three approaches presented relevant insights into the association under study. The findings indicate a statistically significant association between governance and aviation safety oversight effectiveness. Government effectiveness explained a notable portion of the variation in safety oversight effective implementation. Among the predictive data mining models, random forest showed better performance when compared with deep learning and decision tree models. Some of the theoretical contributions of the study include the added support for the factorial validity of the WGI structure for RQ and GE and the establishment of a framework linking broader dimensions of governance and aviation safety oversight, particularly between GE and EI. Practical contributions include recommendations for aviation safety oversight organizations to consider governance measures in safety oversight assessment mechanisms and to promote additional research on common factors associated with safety oversight effectiveness. The study also outlines recommendations for future research, such as exploring additional types and sources of governance metrics and expanding the tested theoretical framework to other safety-relevant industries beyond aviation.

Keywords: aviation safety oversight, civil aviation authorities, governance, public administration, safety, safety management

Dedication

To Julia and Lina.

Acknowledgments

Throughout this journey, I have been fortunate to receive encouragement and support from many individuals, more than I could possibly express in words. My deepest gratitude goes to Dr. Dothang Truong and Dr. Steven Hampton for their leadership and guidance. I am also forever indebted to Dr. Alan Stolzer and Dr. Rafael Echevarne for their generous and insightful contributions as members of my dissertation committee. This journey would not have been possible without the unwavering support of my family. I am deeply thankful for your kindness, patience, understanding, and support. Julia e Lina, vocês são os meus amores. Mãe, muito obrigado pelo incentivo e apoio ao longo da vida. I am also grateful for the support, inspiration, and guidance received from all colleagues at ANAC and ERAU. While you are too many to list individually, I am immensely thankful to each of you for your contributions. Alissa, também devo boa parte dessa conquista ao seu apoio obstinado. Obrigado. Finally, I would be remiss not to acknowledge the support and encouragement from my colleagues in cohort 10 of the Ph.D. in Aviation program.

Table of Contents

Page

List of Tables

Table Page

List of Figures

Figure Page

Chapter I: Introduction

In under a century, aviation has evolved from a dangerous enterprise into one of the safest means of long-distance travel. The continuous revision of regulatory requirements due to the identification of new contributing factors to incidents and accidents favored this evolution. Consequently, events with similar causes are prevented, and safety performance is maintained within acceptable levels (Hollinger, 2013; Stolzer & Goglia, 2015). Civil aviation has always relied on the regulatory process to improve safety performance. As early as the 1920s in the United States, the Morrow board recommended, among other things, the creation of safety standards for pilots and aircraft (Lawrence, 2008). Since then, several other strategic initiatives, such as the ones presented by Moak et al. (2020) and Stimpson et al. (2008), have ensured that safety oversight continues to support additional operational improvements. Nonetheless, even if some countries have strived to improve their safety oversight processes and organizations individually, aviation is an inherently global endeavor, requiring governments worldwide to push in the same direction.

The International Civil Aviation Organization (ICAO) recognizes civil aviation agencies or authorities (CAAs) as the institutions responsible for implementing regulatory measures for aviation safety within states[1]. When structuring their aviation safety functions, CAAs are typically assigned the responsibility to implement *safety oversight.* The term is defined by ICAO (2016) as "a function performed by a state to

[1] Unless otherwise expressed, whenever the term "state" is used in this dissertation, it conveys the notion of sovereign states eligible for ratification or adherence to the Convention on International Civil Aviation as per its Chapter XXI.

ensure that individuals and organizations performing an aviation activity comply with safety-related national laws and regulations" (p. 1-3). Such functions include: (a) the creation of primary aviation legislation and (b) specific operating regulations, (c) the establishment of an adequate and empowered civil aviation authority (d) with sufficient qualified technical personnel, (e) the publication of guidance material for regulated entities and CAA employees, and (f) the implementation of processes for certification and (g) surveillance that include (h) the possibility of taking the appropriate actions to resolve identified safety deficiencies (ICAO, 2016). This list of functions is referenced in ICAO documents as the eight "critical elements" of safety oversight. Starting in 1999, ICAO has audited states on the implementation of these functions through the Universal Safety Oversight Audit Program (USOAP). The program has been widely accepted as beneficial to supporting harmonization and "the safe and orderly growth of international civil aviation throughout the world" (ICAO, 1944, p. 23).

USOAP-CMA audit results indicate that, even if policy supporting the implementation of safety oversight functions is uniform and inscribed in the Annexes to the Convention on International Civil Aviation, implementation still presents challenges. The effective implementation of ICAO standards in 2021 ranged from 4.29% for Djibouti to 98.91% for the United Arab Emirates (ICAO, 2021a). These significant differences suggest that states' capability to implement global policies could be impacted by local factors such as resource availability, support from stakeholders, and organizational challenges (Andrews et al., 2017).

Despite the problem's relevance to aviation safety, a diligent literature review yielded few studies on how states structure their aviation safety oversight functions and institutions. Chapter II – Literature Review presents some of these references in more detail. A few studies, limited to a single civil aviation authority or those constrained to a specific geographic region, cover most of the published assessments (Barreto, 2002; Forsyth et al., 2014; Ghahremani, 2020; Hansen et al., 2006, 2008; O'Neil, 2008; O'Neil & Krane, 2012; Petras & Vaugeois, 2017; Swedavia AB & McGregor & Company, 1988). One noteworthy exception is an initiative by the Organization for Economic Co-operation and Development (OECD) to apply its methodology of the Indicators on the Governance of Sector Regulators to 29 Latin American CAAs (Durand & Pietikainen, 2022). The OECD methodology addresses CAA governance in three dimensions: independence, accountability, and scope of action. Other studies present broader analyses of global perspectives and the role of international organizations and standards (Broderick & Loos, 2002; Detra, 2006; Ratajczyk, 2014). Others still tackle authorities' processes and concepts more broadly but refrain from proposing ways to evaluate safety oversight effectiveness (Bahr, 2014; Klenka, 2017; Sønderby, 2016).

On the other hand, initiatives to study and measure public governance on a higher, more comprehensive level are prominent in scholarly literature. State-level policy-making and general government effectiveness have been widely studied, with added focus on the faster and more complex dynamics of public decision-making processes in a public-private network of actors (Ansell & Torfing, 2022; Bovaird, 2005; Bovaird & Loeffler, 2016; Ongaro & van Thiel, 2018; Osborne, 2010). The revised perspective has been referred to as governance, public governance, good governance, network governance, and new public governance (Bovaird & Loeffler, 2016; Vignieri, 2020). Several databases on regulatory governance quality exist and highlight initiatives by the World Bank and the OECD as particularly relevant (International Finance Corporation et al., 2010). The World Bank's Worldwide Governance Indicators (WGI) initiative periodically collects and publicizes indicators on over 200 countries' governance maturity, divided into six dimensions. The definition of governance used in the WGIs includes three critical aspects of governance: (a) the government selection process, (b) citizens' respect for the institutions and their interactions, and (c) the government's capacity to formulate and implement reasonable policies (Kaufmann et al., 2010b). The policy formulation and implementation aspect of governance comprises *Government Effectiveness* and *Regulatory Quality* dimensions. The *Government Effectiveness* (GE) dimension proposed by Kaufmann et al. (2010b) captures (a) the perceptions of public and civil service quality, (b) the level of political pressure, policy formulation and implementation, and (c) the level of commitment to said policies. The *Regulatory Quality* (RQ) dimension focuses on the ability of governments to formulate and implement relevant and adequate policies and regulations. The elements GE and RQ are particularly relevant to the discussion of CAA effectiveness and were selected as potential predictors for aviation safety oversight implementation in the present study.

Statement of the Problem

The global civil aviation industry has maintained extraordinary long-term growth despite the occasional reversals. Identifying ways to ensure that the CAAs' safety oversight processes become and remain effective is crucial to the industry's sustained development (Roberts & Safety Management International Collaboration

Group, 2013). The need for better safety oversight worldwide becomes even more evident when the global nature of air transportation is considered. Airlines from countries with poorly performing civil aviation authorities may eventually expose other countries' citizens to the risks associated with their operations. It is thus in the interest of all that the industry matures across boundaries. The drive to build more effective safety oversight systems has led ICAO and its member states to adopt performance-based regulatory structures and safety management principles and processes. The challenges in implementing these changes are significant, even though the industry generally recognizes the benefits of a move away from purely prescriptive safety oversight processes. The fact that such a change requires a new set of institutional capabilities presents additional difficulties. In a performance-based regulatory environment, the states' responsibility for identifying substantial problems and disposing of resources to mitigate associated risks becomes more relevant. Andrews et al. (2017) have advised that "premature load bearing" institutions with tasks exceeding their capacity can compromise current capability (p. 54). Still, the academic literature on the aviation safety oversight assessment and the factors contributing to its effectiveness is scant. On another hand, attempts to structure effective measures of governments' bureaucratic processes in an ampler manner are more prominent in the literature, with significant effort directed toward measures of governance. Still, the extant scholarly literature had not previously assessed the relationship between regulatory effectiveness metrics within the aviation sector and broader governance indicators. Identifying predictors to safety oversight effectiveness from the governance literature can help ensure that these organizations continue to contribute to a safer civil aviation system at a global level.

Purpose Statement

The main objective of the study was to identify predictors of effective aviation safety oversight implementation from public governance indicators and, ultimately, support states in improving safety oversight and contributing to a safer global civil aviation system. Using archival quantitative data on the assessment of states' governance and their level of effective implementation assessed through USOAP audits, the research set out to establish whether an association existed among such variables and the strength of that relationship. The use of governance metrics as predictors of aviation safety oversight would highlight structural difficulties states face when implementing or improving their civil aviation authorities' processes. The variables selected as independent predictors were those collected in the World Bank's Worldwide Governance Indicators (WGI) initiative. Greater focus is directed to regulatory practice and policy implementation, namely Government Effectiveness (GE) and Regulatory Quality (RQ), and their associated sources. The dependent or target variable was ICAO's Effective Implementation (EI) metric used to communicate USOAP-CMA safety oversight audit results.

Different quantitative and qualitative methods have been employed to measure effectiveness in public and private organizations. The present research expanded on these approaches and focused on their impact on aviation safety oversight processes and on a better understanding of their assessment mechanisms. Since all variables considered were scale/ratio measures, the application of distinct analytical approaches supported the identification of statistically significant relationships between these variables and the comparison of different models and their application to the identification of this relationship. The results contributed to the establishment of predictive models to support the understanding of government implication to oversight in states' CAAs.

An exploratory comparison of different types of analytical approaches provided more support for the identification of potential associations among the independent variables and aviation safety oversight implementation, while also providing insights into the models' explanatory power of such relationships. Testing models from different families, such as multiple linear regression, confirmatory factor analysis (CFA) and structural equation modelling (SEM), and non-linear data-mining models, such as decision trees, deep learning, and random forests, contributed to the study's exploratory dimension and provided a fresh perspective on the data structure and the interrelationships across factors and sources. The application of SEM and data mining on the sources used to derive the WGI and USOAP-CMA EI metrics provided additional support to their use as policy-informing instruments.

Significance of the Study

Theoretical Contributions

The primary focus of the study is on identifying and validating factors and elements of broader public administration and governance theories and their association with governments' aviation safety oversight. The study set out to contribute to the theory by developing a baseline relationship framework that includes aspects of governance and measures of safety oversight effectiveness, providing inter-disciplinary insights to the relevant problem of safety oversight effectiveness. The framework, in turn, provides a new perspective to the understanding of the contributions of public administration and governance to aviation safety oversight, adding value to the body of knowledge in aviation safety. It also establishes a foundation upon which further studies could be developed to test causal relationships, their directions, and the identification of specific confounders to these associations. Additionally, a more thorough understanding of the potential relationship between governance and aviation safety oversight may better support the development of improved instruments to assess aviation safety oversight effectiveness.

Practical Contributions

The ever-increasing complexity and globalization of the industry have called for even more effective safety oversight by states worldwide. An everyday staple of this evolution has been performance-based regulatory approaches, including the implementation of state safety programs (SSPs) (Roberts & Safety Management International Collaboration Group, 2013). Performance-based oversight allows states to address local safety issues more effectively while also promoting continuous improvement in safety performance as the industry evolves. In consequence, the evaluation of oversight processes becomes more challenging. ICAO has continuously promoted changes to the USOAP program to better gauge states' aviation safety oversight effectiveness. The current research undertook to inform those processes by providing novel perspectives into what makes states' civil aviation authorities more effective while supporting the identification of additional strategies to assess and enhance the effectiveness of safety oversight in civil aviation authorities. Comparing the performance of different types of predictive models, in turn, supported the identification of a suitable tool to study the association among the variables under study while also indicating potential opportunities for expansion for other aspects of safety oversight implementation. The current study supports the identification of opportunities for systemic progress and helps states improve safety performance, particularly those at lower safety oversight implementation levels. Finally, a more mature global regulatory oversight process would promote ICAO's strategic objectives of enhancing global civil aviation safety and fostering the development of a sound and economically viable civil aviation system (ICAO, 2022a).

Research Question and Hypotheses

Three research questions were selected to serve as directional guidance for the research. They also reflected the need to understand better the predictors of effective aviation safety oversight, and to help states continuously improve their civil aviation authority processes. This led to the definition of RQ_1 below.

RQ_1: What is the effect of public governance indicators of Regulatory Quality (RQ) and Government Effectiveness (GE) on the level of aviation safety oversight Effective Implementation (EI) amongst states?

RQ_2: What predictive models better predict Effective Implementation (EI) from public governance indicators of Regulatory Quality (RQ) and Government Effectiveness (GE)?

RQ_3: What elements of governance are more closely associated with increased USOAP effective implementation results?

Details related to variables RQ, GE, and EI are presented in Chapter II – Review of the Relevant Literature and Chapter III – Methodology. While different predictive models could be employed to assess the relationship among the variables under study, data characteristics could lead to the adoption of some types of analytical approaches over others. Identifying the most suited type would better establish the association's strength and relative importance, which supported RQ_2. Finally, RQ_3 addressed the relative importance of the independent variables in predicting the outcome variable. The results collected to answer these questions enhanced the understanding of the problem under study in terms of the contributions to safety oversight effective implementation.

Delimitations

Data availability imposed the most significant constraints on the scope of the current study, as only states with data available on all variables under investigation were considered in the analyses. Expanding assumptions or generalizing concepts beyond the states for which data were available for use in the study was only considered with caution. Conditioning the analysis on data availability may be considered a form of convenience sampling, posing threats to the study's external validity, and limiting its conclusions' generalizability to the states in the sample. However, with a representative sample including states from various levels of economic activity and geographic locations, it is possible to support generalizability with the necessary considerations.

Another important delimitation is that only the Regulatory Quality (RQ) and Government Effectiveness (GE) elements of governance were considered in the current analysis. Governance is a complex issue involving several factors, and other elements could also contribute to aviation safety oversight Effective Implementation (EI). Restraining analysis to RQ and GE as the elements seen as more directly related to EI could limit the identification of other factors or potential confounders to the relationship.

While CAAs commonly exert additional mandates in economic regulation and passenger protection, the focus of this study is restricted to social regulation aspects, more specifically on aviation safety oversight. Other subjects associated with regulatory aspects of civil aviation activities and additional factors to the relationship between governance and safety oversight could be the focus of future research.

Limitations and Assumptions

The main limitations of the current study were the ones related to the measures for the variables under investigation and the biases potentially extracted from their association. A discussion of such constraints was presented in Chapter II – Review of the Relevant Literature. Concerns raised regarding the validity and reliability of the WGI have been addressed by the authors (Kaufmann et al., 2007). One such concern is that minor changes to the survey data sources may be necessary after analysis, affecting longitudinal assessments of WGI data. Kaufmann et al. (2007) indicated, however, that historical data is commonly updated to reflect changes such as source-dimension loadings. The data sources that compose the WGI are also very diverse, including secondary survey references from other institutions, public sector data providers, non-governmental organizations, and commercial business information providers (Kaufmann et al., 2010a, 2010b). The relevance and validity of the WGI dimensions selected for the current analysis were supported by extant literature but additional unaddressed confounders can also contribute to variation in individual elements of the proposed framework and their association. Additional external variables can potentially contribute to the proposed association based on states' demographics and can help identify more complex interrelations relative to governance and safety oversight effectiveness. However, the literature on multivariate methods discourages the indiscriminate increase in model complexity, especially in problems constrained in terms of sample size. Still, no inference of causality was expected from the results due to the exploratory and correlational nature of the present study's design. Another important assumption is that no bias arise from the difference in data collection moments of different data sources for the same state. The absence of significant time differences in data collection for a single state between sources was ensured in analysis by matching metrics referring to the same calendar year. The use of secondary data also restricted the design choices commonly used to ensure the independence of observations (e.g., random sampling, adjusting data collection intervals and spatial distributions), which was assumed for the current study's linear regression analysis. While failing to meet the independence assumption could impact the statistics, ICAO audit measurements were considered sufficiently interspersed to minimize dependence, reflecting common practice in research involving cross-country comparisons.

Summary

Chapter I briefly presented the problem of aviation safety oversight and public governance. It introduced the most relevant concepts and arguments supporting the relevance of the research and its potential benefits. The following chapters II and III delve into more detail on the theory and current research supporting the analysis and the methodological aspects of governance and safety oversight. Essential concepts, relevant governance assessment initiatives, associated variables, and other relevant background information supporting the current study are also discussed. Chapters IV and V present the results and associated discussions and interpretations, respectively.

Definitions of Terms

Governance	"the traditions and institutions by which authority in a country is exercised. This includes (a) the process by which governments are selected, monitored, and replaced; (b) the capacity of the government to effectively formulate and implement sound policies; and (c) the respect of citizens and the state for the institutions that govern economic and social interactions among them." (Kaufmann et al., 2010b)
Safety oversight	"A function performed by a State to ensure that individuals and organizations performing an aviation activity comply with safety-related national laws and regulations." (International Civil Aviation Organization, 2016)

List of Acronyms

CAA	Civil Aviation Authority
EI	Effective Implementation
GE	Government Effectiveness

ICAO	International Civil Aviation Organization
RQ	Regulatory Quality
SSP	State Safety Program
USOAP	ICAO's Universal Safety Oversight Audit Program
WGI	Worldwide Governance Indicators

Chapter II: Review of the Relevant Literature

The current chapter presents the most relevant references related to the evolution of the study of public governance and the different initiatives that have been structured to measure it. It also presents the main concepts and assessment strategies associated with aviation safety oversight effectiveness. While studies about civil aviation authority effectiveness are scarce, the program implemented by ICAO to audit states on these processes is discussed to support the analysis.

Literature Review and Theoretical Foundation

This section presents the theoretical perspectives of the topics under consideration for the study. It starts with a summary of the different public administration traditions and the conceptual frameworks and theories. The discussion follows a review of various initiatives that measure "public governance" and a description of the World Bank's Worldwide Governance Indicators. The ICAO USOAP-CMA program is also presented with a brief overview of its results.

Advances in Public Administration and Governance

The field of public administration is a constantly evolving body of knowledge that draws from diverse areas of study, including social and political sciences, institutional economics, international relations, and development studies, among others. Several theories of public administration were also presented through the years, reflecting their contributions to the problem of government effectiveness. While it is impossible to separate individual perspectives and traditions in discrete steps, the study of public administration and public governance is commonly accepted as having evolved in three main phases: (a) "old" public administration, (b) new public management, and (c) public governance (Bovaird & Loeffler, 2016; Dunleavy & Hood, 1994; Osborne, 2010; Vignieri, 2020).

The principles of Public Administration were developed, at least from a western perspective, in the 1800s. They were substantively founded on the concepts and theories propounded by Max Weber, Woodrow Wilson, and Frederick Taylor (Hughes, 1992). Up to that point, most political scientists' discussions focused on the nature of states and the structure of governments. Studies focused mainly on "who shall make law, and what shall that law be" (Wilson, 1887, p. 198). Max Weber promoted the adoption of an ideal-type rational/legal bureaucracy, a hierarchical model of public administration bounded by the concepts of legal/rational authority, jurisdiction, hierarchy, and procedures (Pollitt & Bouckaert, 2011; Vignieri, 2020). Weber's foundation for such a system included: (a) the recognition of formal authority and its delimitation by established laws and rules, (b) the separation of positions and persons and their allocation based on competence and expertise, (c) the use of hierarchy as the primary tool to address coordination, (d) reliance on written records to ensure continuity, and (e) the need for disinterested administration (Dunleavy & Hood, 1994; Vignieri, 2020). The assumption was that bureaucracies played a crucial role in policymaking and implementation and should rely on these principles to provide equal services to all citizens (Osborne, 2006).

Woodrow Wilson (1887) proposed the study of public administration as a new science to support the efforts to improve the personnel, the organization, and the methods used in government. The new field would be removed from politics and focus on establishing "foundations laid deep in stable principles" (p. 210) for executive administration methods. Wilson's central thesis was that good practices of public administration work independently of the political structure or type of government in which the bureaucracy is placed. According to Wilson, to achieve good administration, it was necessary to have a hierarchically organized structure of ranks filled with well-trained and disciplined personnel (Ostrom & Ostrom, 1971). Taylor's (1919) work on scientific management completed the traditional model of administration used from the 1920s onward (Hughes, 1992), giving it "form and purpose, a self-confidence to both the practice and the study" (Dunsire, 1973, p. 94). Osborne (2006) proposed the critical foundations of the earlier days of public administration to be:

> The dominance of the 'rule of law'; a focus on administering set rules and guidelines; a central role for the bureaucracy in policymaking and implementation; the 'politics – administration' split within public organizations; a commitment to incremental budgeting; and the hegemony of the professional in the service delivery system. (p. 378)

The general principles driving public administration should be "hierarchy and rules; permanence and stability; an institutionalized civil service; internal regulation; equality" (Pollitt et al., 2007, p. 3). In this traditional public administration perspective, the role of politics was to select governors responsible for defining policy objectives which, in turn, were accountable to citizens through the election process (Bryson et al., 2014). Respect for such principles would naturally yield a bureaucracy with "the utmost possible efficiency and at the least possible cost either of money or of energy" (Wilson, 1887, p. 197). Andrews et al. (2017) contested this view of development "naturalness," highlighting the pervasive slow progress in building state capability across countries.

The late nineteenth and early twentieth centuries principles of public administration remained largely unquestioned until the 1970s, when, in the socio-political sphere, deference towards the authority of civil servants and politicians started to decline. Citizens started to voice their resentment toward the quality of public services (Pollitt et al., 2007, p. 4). Criticisms of the old public administration compounded by arguing that its principles were incompatible, and that no compelling argument indicated that hierarchical ordering was necessarily the most efficient organizational solution (Osborne, 2006; Simon, 2013).

Concerns about government failure and beliefs in the efficiency and effectiveness of markets and economic rationality led to a new public administration approach called New Public Management (NPM) (Bryson et al., 2014; Hood, 1991). Proponents of the NPM drew from varied theoretical stances in institutional economics, including public choice, transaction costs, and the principal-agent theory (Coase, 2004; den Hertog, 2010; Ostrom & Ostrom, 1971; Stoker, 2018; Vignieri, 2020). Definitions for what NPM stands for abound in literature, but Pollitt (2007) revised them proposing their interpretation as a two-level concept. On a higher stance, NPM reflected a theory or doctrine that argued for the benefits of applying business *concepts*, *techniques,* and *values* to improve public administration. This overarching concept was supported by a set of general principles, which included: (a) increased focus on *performance* and output measures; (b) preference for specialized, more straightforward organizations; (c) reduced reliance on hierarchy for coordination; (d) increased adoption of market-type mechanisms; (e) and an attempt to treat service users as *customers* with the support of quality management tools (Pollitt, 2007; Pollitt & Bouckaert, 2011). Hood (1991) is said to have "codified the nature of the New Public Management (NPM)" (Osborne, 2010, p. 1), proposing the following seven doctrinal components for NPM:

> "Hands-on professional management" in the public sector; Explicit standards and measures of performance; Greater emphasis on output controls; Shift to disaggregation of units in the public sector; Shift to greater competition in the public sector; Stress on private-sector styles of management practice; Stress on greater discipline and parsimony in resource use. (Hood, 1991, p. 5)

NPM did not escape its share of critics, both those that defended more hierarchical traditions and those that considered the use of business tools in government an overly simplistic solution. Pollitt and Bouckaert (2011) questioned the effectiveness of NPM: "Elements of the NPM have been widespread, but have they worked?" (p.15). Hood and Jackson (2013) criticized NPM, mentioning the lack of hard-data evidence supporting its claims and arguing that its principles may contribute to socially created disasters in Charles Perrow's definition (1984). The idea of public organizations dealing with citizens as "customers" has also seen its share of criticisms. Sparrow (2000) demonstrated the inadequacies of such an approach, particularly in regulatory and enforcement agencies responsible for delivering "obligations rather than services" (p. 2). Osborne (2006, pp. 377–379) enumerated the common criticisms of NPM but upheld its ability to unpackage policy implementation and management complexities. The author also posited that NPM was a brief transition from the "statist and bureaucratic" tradition of public administration to the new public governance tradition that addressed many such concerns (p. 377). Dunleavy and Hood (1994) structured the critics of NPM into four categories – fatalist, individualist, hierarchist, and egalitarian – and highlighted some contradictions.

To address the shortcomings of the early public administration and NPM and the “plural and pluralist complexities” of the twenty-first-century state, researchers increasingly embraced the concept of governance (Osborne, 2006, p. 381). The structural changes brought about by the increased perception of the network factors affecting governments drove scholars of public administration to the development of a new operating code (Stoker, 2018). Researchers referred to this new tradition as governance, good governance, public governance, network governance, and new public governance (Bovaird & Loeffler, 2016; Vignieri, 2020). While some debate on the best approach to define governance still occurs, these definitions generally referred to “governing styles in which boundaries between and within public and private sectors have become blurred” (Stoker, 2018, p. 15). A pervasive element of the different views of public governance was the departure from some of the central principles of earlier public administration traditions. Governance rejected vertically integrated policy making and implementation strategies and the need for hierarchical mechanisms to ensure accountability. It expanded the focus beyond the NPM’s on intra-organizational processes and management (Osborne, 2006).

The study of governance evolved from the NPM limitations in accommodating different values, policy-making processes, and types of organization/stakeholders presented by the more complex and dynamic contemporary public decision-making processes (Bovaird, 2005). Interest in the topic grew due to the increased recognition that the scale and complexity of relevant decision processes and functions were progressively organized around networks. Different governance concepts expanded the previous public administration traditions by recognizing “the role of non-state actors in decision-making on public issues” (Bovaird, 2005, p. 219). Osborne (2006, 2010) supported this view. He highlighted the plural – with multiple inter-dependent actors engaged in public service delivery – and pluralist – where several inputs and processes inform policy-making – characteristics of the contemporary state. These characteristics required the focus to be on “inter-organizational relationships and the governance of processes” while stressing “service effectiveness and outcomes” (Osborne, 2006, p. 384, 2010, p. 97).

Fukuyama (2013) took a somewhat different stance and, moving away from the discussion on the system such government is embedded in, argued that governance is related to its ability “to make and enforce rules, and to deliver services” (p. 350). Recognizing the challenges of adopting broader definitions, La Porta et al. (1999) attempted to restrict the concept of “good governance” to “good for economic development” (p. 223). This approach, in turn, was claimed to disregard significant non-economic consequences, such as interpersonal trust and subjective well-being, while leaving room for tautology (Rothstein & Teorell, 2005). In turn, Rothstein and Teorell (2005) argued that the more specific element of impartiality of government institutions exercising authority should be the central aspect of governance definitions. This idea has been contested as being unresponsive to the complexities of real-world problems.

However dissonant, the distinct views of governance broadly recognized “the role of non-state actors in decision-making on public issues” (Bovaird, 2005, p. 219).

They also suggested a move away from central principles of earlier public administration traditions such as hierarchy and vertical integration of policymaking and implementation (Osborne, 2006, 2010). Following the previous reform waves, the newer governance views of public administration have also amassed their share of critics. Some authors presented concerns about the increase in complexity these new traditions present, arguing that some states, especially those with lower capabilities, may have to prioritize the application of limited resources (de Vries, 2013; Grindle, 2004, 2007). However, reforms are seldom achieved in discrete steps. Detailed examinations commonly highlighted that elements of traditional approaches are gradually adapted, and features from different reform generations are usually preserved in more modern approaches (Christensen & Lægreid, 2008; Osborne, 2010). To avoid delving into an examination of a "complex 'archaeology' of reforms" (Christensen & Lægreid, 2008, p. 8), the current study adopted the approach to governance – or "good governance" – presented by Kaufmann et al. (2010b) and discussed in Andrews, (2008), Osborne, (2010), and Thomas (2010). Osborne (2010), a proponent of the term "new public governance," acknowledged that the expression "good governance" adopted by the World Bank carries the ordinary meaning of governance (p. 92). Kaufman et al. (2010b) defined governance as "the traditions and institutions by which authority in a country is exercised" (p. 4). The definition included the process used for selecting and replacing governments, the effectiveness of the policy-making process, and the respect of people towards institutions. This definition of governance supported the work done by the World Bank to develop the Worldwide Governance Indicators (WGI), an initiative to measure different elements of states' governance (Kaufmann et al., 2010b). The strategies used to measure governance and the details of the WGI initiative are presented in the following section.

Public Governance Indicators

Attempting to measure abstract concepts exposes problems and helps to break them down into tractable and prioritizable pieces, supporting political and administrative decision-making (Erkkilä et al., 2016). Rotberg (2004) argued that rating governance can create a "virtuous, competitive cycle" among states for effectiveness improvement (p. 74). However, the absence of a consensus definition of governance contributed to a proliferation of alternative approaches to measuring it. This effect was further motivated by the different uses of such metrics, such as informing the prioritization of public resource allocation, financial investments, civil society advocacy, and research. These initiatives also supported various types of stakeholders, including governments, NGOs, academics, development agencies, private entities, and the media (Sudders & Nahem, 2004).

Sudders and Nahem (2007) proposed that governance indicators should be considered in a three-tier framework with levels for (a) input/rights/commitment/de jure, (b) process/responsibility/de facto, and (c) output/outcome/enjoyment/performance/de facto. The authors also presented some general aspects that affected all forms of governance monitoring, including those related to (a) who gathered the data, how they gathered it, and with which tools and processes, (b) what were the primary data sources, (c) who was sampled in data collection, and (d) how the indicators were compiled. A few of the relevant initiatives to measure governance are briefly presented below.

The Bertelsmann Stiftung's Transformation Index (BTI) started in 2004 and "analyzes and evaluates whether and how developing countries and countries in transition are steering social change toward democracy and a market economy" (Hartmann et al., 2020, p. 116). It aggregated the results of surveys on transformation processes and political management in two indices related to the state's degree of political and economic transformation and governance. The BTI governance index evaluated the quality of political leadership supporting transformation initiatives (Hartmann et al., 2020). The concept of governance used in the BTI was composed of five criteria comprised of 20 indicators. The governance index criteria reflected difficulty level, steering capability, resource efficiency, consensus-building, and international cooperation metrics. Current results for the BTI presented data on 137 developing states and were published every two years (Hartmann et al., 2020).

The Ibrahim Index of African Governance (IIAG) has been published since 2007 and was reassessed between 2018 and 2020, addressing some of its criticisms (Farrington, 2009, 2010, 2011; Mo Ibrahim Foundation, 2020). It assessed governance performance in 54 African states in four categories: security and the rule of law; participation, rights, and inclusion; foundations for economic opportunity; and human development. Data was collected directly from citizens and official and expert assessments (Mo Ibrahim Foundation, 2020).

The World Bank Group is an international development institution that provides funding and knowledge for developing countries. The institution was responsible for the "Governance Matters" initiative to establish a structure of indicators and conduct periodic cross-country assessments on six governance dimensions (Kaufmann et al., 2009). The Bank’s Worldwide Governance Indicators (WGI) have been regarded as the most widely cited governance indicator source (Sudders & Nahem, 2007). It has been an ongoing initiative, with data collected periodically in over 200 countries since 1996. It was used primarily to support decision-making in development institutions related to loans and grants to developing countries' governments (Arndt & Oman, 2006; Kaufmann et al., 2010b). The WGI employed an unobserved components model to aggregate data collected from several different sources into six metrics reflecting governance dimensions: (a) Voice and Accountability, (b) Political Stability and Absence of Violence, (c) Government Effectiveness, (d) Regulatory Quality, (e) the Rule of Law, and (f) Corruption.

The sources for calculating the WGIs are subjective or perception-based data collected from various informed stakeholders and experts from public sector agencies, NGOs, and the private sector. These sources were categorized into three main groups: surveys of individual and domestic firms with knowledge of the countries' governance, reports from multilateral and public development agencies, and commercial business information providers (Kaufmann et al., 2010b). The authors argued for the importance of perception data, highlighting that (a) it usually is what agents use to inform decisions, (b) that more objective measures tend to capture a *de jure* notion of governance which may not be aligned with the results it delivers, and, more pragmatically, (c) that more objective data is simply not available (Kaufmann et al., 2010b). Erkkilä et al. (2016) supported this approach and highlighted that governance indicators were seldom purely objective but always involved some subjective assessment. The data for the six indicators were normalized into a -2.5 to 2.5 range, with higher numbers reflecting higher levels of governance, and were presented along with the margins of errors (Kaufmann et al., 2010b). Critics of the World Bank's approach have raised concerns about bias, lack of comparability among countries, and, potentially, problems with construct validity (Andrews, 2010; Arndt & Oman, 2006; Langbein & Knack, 2010; Thomas, 2010; van de Walle, 2007). The choice of the theoretical framework used by the World Bank to create the WGI was aligned with the organization's interest in developmental studies, compromising assumptions that it reflected a disinterested compilation of empirical data (Erkkilä et al., 2016). The authors have responded to the main arguments against the WGI (Kaufmann et al., 2007, 2010a). Arndt and Oman (2006), while admitting some concerns with using the indicators, regard the WGIs as “the most carefully constructed set of governance indicators” (p. 49). The authors also recognized the WGI’s extensive documentation on its construction and limitations. Erkkilä et al. (2016) acknowledged that new actors in global governance measurement often criticize existing products but adopt many of the existing practices in the field. Pollitt and Bouckaert (2011, p. 128) argued for the adoption of WGIs, stating that the international community widely knows them and that their extensive discussion provides relevant insights into how to measure effectiveness.

The Methodology of the WGIs

While *governance* has been defined and measured in many distinct ways, it was broadly regarded as crucial in sustained economic development (Kaufmann & Kraay, 2008). The WGI project collects data on countries’ quality of governance from approximately 30 sources and 25 organizations. It aggregates them into six dimensions: Voice and Accountability (VA), Political Stability and Absence of Violence/Terrorism

(PV), Government Effectiveness (GE), Regulatory Quality (RQ), Rule of Law (RL), and Control of Corruption (CC). The results have been reported every two years since its inception in 1996.

To build the WGI's six dimensions, data from several different sources are collected, including: (a) surveys of households and firms, (b) non-governmental organizations (NGOs), (c) commercial business information providers, and (d) public sector organizations. A complete list of the sources used for the 2022 update of the WGI is detailed in Table 1, along with information related to its type, the number of countries covered, and public availability. Types of expert assessment include CBIP - Commercial Business Information Provider, GOV - Public Sector Data Provider, and NGO - Nongovernmental Organization Data Provider.

Table 1

List of Sources for the Worldwide Governance Indicators

Code	Source	Type	Public	Coverage
ADB	African Development Bank Country Policy and Institutional Assessments	Expert (GOV)	Partial	54
AFR	Afrobarometer	Survey	Yes	22
ASD	Asian Development Bank Country Policy and Institutional Assessments	Expert (GOV)	Partial	28
BTI	Bertelsmann Transformation Index	Expert (NGO)	Yes	129
EIU	Economist Intelligence Unit Riskwire & Democracy Index	Expert (CBIP)	Yes	183
EQI	European Quality of Governance Index (Underlying Survey Data)	Survey	Yes	27
FRH	Freedom House	Expert (NGO)	Yes	198
GCB	Transparency International Global Corruption Barometer Survey	Survey	Yes	115
GCS	World Economic Forum Global Competitiveness Report	Survey	Yes	144
GII	Global Integrity Index	Expert (NGO)	Yes	62
GWP	Gallup World Poll	Survey	Yes	161
HER	Heritage Foundation Index of Economic Freedom	Expert (NGO)	Yes	183
HRM	Human Rights Measurement Initiative	Expert (NGO)	Yes	39
HUM	Cingranelli Richards Human Rights Database and Political Terror Scale	Expert (GOV)	Yes	194
IFD	IFAD Rural Sector Performance Assessments	Expert (GOV)	Yes	98
IJT	iJET Country Security Risk Ratings	Expert (CBIP)	Yes	197
IPD	Institutional Profiles Database	Expert (GOV)	Yes	143
IRP	African Electoral Index	Expert (NGO)	Yes	54
LBO	Latinobarometro	Survey	Yes	18
MSI	International Research and Exchanges Board Media Sustainability Index	Expert (NGO)	Yes	71
OBI	International Budget Project Open Budget Index	Expert (NGO)	Yes	100
PIA	World Bank Country Policy and Institutional Assessments	Expert (GOV)	Partial	136
PRC	Political Economic Risk Consultancy Corruption in Asia Survey	Survey	Yes	17
PRS	Political Risk Services International Country Risk Guide	Expert (CBIP)	Yes	140
RSF	Reporters Without Borders Press Freedom Index	Expert (NGO)	Yes	177
TPR	US State Department Trafficking in People report	Expert (GOV)	Yes	185
VAB	Vanderbilt University Americas Barometer	Survey	Yes	26
VDM	Varieties of Democracy Project	Expert (NGO)	Yes	171
WCY	Institute for Management and Development World Competitiveness Yearbook	Survey	Yes	59
WJP	World Justice Project Rule of Law Index	Expert (NGO)/Survey	Yes	97
WMO	Global Insight Business Conditions and Risk Indicators	Expert (CBIP)	Yes	203

As presented by Kaufmann et al. (2010b), the calculation of the WGI for a particular year followed three steps. First, the individual questions or variables in each reference data source were assigned to one of the six indicators reflecting the WGI dimensions of governance. The values for the individual questions were normalized to a zero to one range, with larger numbers reflecting a better contribution to the associated governance dimension. For instance, the Bertelsmann Transformation Index (BTI) was comprised of several individual metrics related to stateness, political participation, rule of law, stability of democratic institutions, political and social integration, among others (Hartmann et al., 2020). The variables related to distinct sources were normalized and associated with a WGI dimension. Finally, an Unobserved Components Model (UCM) made the rescaled data comparable across datasets by assessing the weighted average for each source-country (Goldberger, 1972; Kaufmann et al., 2010b). The UCM resulted in a weighted data average, with weights indicating the correlation level among different sources. The resulting metrics for each dimension fell within a normal distribution ($\mu = 0; \sigma = 1$).

The process further standardized results mitigating the impact of potential differences in underlying variable distribution, aggregated these variables into each dimension by applying a weighted average, and indicated the margins of error of each resulting metric reflecting the WGI dimension for a particular year and state (Kaufmann et al., 2010b). To do this, the researchers responsible for the WGIs proposed the observed score of a country j for indicator k, y_{jk}, to be a linear function of unobserved governance in country j, g_j, and an error term ε_{jk}, as

$$y_{jk} = \alpha_k + \beta_k(g_j + \varepsilon_{jk}), \tag{1}$$

with α_k and β_k being the mapping parameters that scale measures with different units into the standardized governance dimension being evaluated. Errors terms were assumed to be independent across sources j and m, $E[\varepsilon_{jk}\varepsilon_{jm}] = 0$ and to be normally distributed, and $V[\varepsilon_{jk}] = \sigma_k^2$. The mean of the normally distributed unobserved governance in country j conditioned on the observed data y_{jk} was calculated with the following equation.

$$E[g_j | y_{j1}, \ldots, y_{jK}] = \sum_{k=1}^{K} w_k \frac{y_{jk} - \alpha_k}{\beta_k}, \tag{2}$$

where the weight terms were given by

$$w_k = {\sigma_k^{-2}}\Big/{1 + \sum_{k=1}^{K} \sigma_k^{-2}}. \tag{3}$$

The framework presented above allowed researchers to also compare the standard deviation for each assessment and to define confidence intervals for these measures, which are dependent on the sources' number and precision (Kaufmann et al., 2010b).

Aviation Safety Oversight and ICAO USOAP

Among several expectations legislators place on aviation regulatory practices by civil aviation authorities, one is present in most, if not all, states: ensuring the traveling public's safety. Aviation is among the most regulated industries when safety requirements are considered (Gowrisankaran, 2002), which has contributed to the industry's commitment to ever-increasing levels of safety performance. However, a diligent literature review with the support of scholarly research platforms yielded few studies on how states structure their aviation safety oversight functions and institutions and what contributes to their effectiveness.

Regulating aviation safety is a risk control effort. Baldwin et al. (2012) posited that some decisions, including the one a passenger makes to fly with an airline, are not of absolute certainties like safe or unsafe but one of risk and uncertainty. The sometimes-abstract view people have of risks resulted in biased perceptions and responses, which have led, in turn, to widely different regulatory approaches to mitigate them across domains (Hood et al., 2001). In the US, OSHA-H's implementation of asbestos occupational exposure limits cost 74 million dollars in 1990 values per premature death averted, while aircraft cabin fire protection standards, for instance, cost around 0.1 million in comparable conditions (Sunstein, 2002). These differences indicated that the way society views and acts upon societal risks can be more associated with subjective matters than with more objective risk assessments. Thus, fears, anxieties, and moral panics may be more effective in supporting regulatory actions than structured data analysis (Baldwin et al., 2012).

According to ICAO, aviation safety oversight is composed of the enactment of (a) primary aviation legislation and (b) specific operating regulations, (c) the establishment of an adequate and empowered civil aviation authority (d) with sufficient qualified technical personnel, (d) the publication of guidance material for regulated entities and CAA employees, and (e) the implementation of processes for certification and (f) surveillance that include (g) the possibility of taking the appropriate actions to resolve identified safety deficiencies (ICAO, 2016; Ratajczyk, 2014). These processes compose what ICAO called the critical elements of safety oversight and encompass the broad set of regulatory tools agencies must use to achieve the public interest of having a safe aviation system. Starting in 1999, ICAO has audited states on implementing these processes through the Universal Safety Oversight Audit Program (USOAP) (Petras & Vaugeois, 2017). Some discussion has taken place on the legitimacy of such arrangements and a potential conflict with the principles ensuring states' sovereignty over the airspace above their territory (Detra, 2006). Regardless, the program has been widely accepted as beneficial to increasing harmonization and, consequently, supporting "the safe and orderly growth of international civil aviation throughout the world" (ICAO, 1944, p. 23). Still, USOAP audits presented mixed results even if policy supporting the implementation of safety oversight is uniform and inscribed in the Annexes to the Chicago Convention. The published reports indicated Effective Implementation (EI) results ranging from 4.3%-98.9% in 2021 (ICAO, 2021a). These results stressed the need for more information on what drives safety oversight performance across the globe to support the development of state capability in aviation safety assurance.

More recently, ICAO has coordinated international efforts to expand from the more prescriptive critical elements into performance-based regulatory structures based on safety management (Yadav & Nikraz, 2014). The implementation of safety management systems was ubiquitous in aviation service providers and showed positive results (Stolzer & Goglia, 2010). This approach presented significant benefits and fundamental challenges for regulators worldwide. ICAO developed analytical tools to support states in advancing their safety management capabilities using a data-driven approach (Jung et al., 2018). However, the results presented by Andrews et al. (2017)

advise that *premature load bearing* institutions with tasks that exceed organizational aptitude while intending to accelerate modernization can compromise any small capability present.

However difficult to implement, different initiatives proposed methods for the complex task of regulating safety risks. ICAO's critical elements of safety oversight suggested a master list of functions that states had to implement to ensure that the main aspects of safety oversight are present. Other researchers provided a problem-centered approach to solving the most critical problems those states face. Sparrow (2000) proposed that regulatory agencies "pick important problems and fix them" (p. 132), with specific steps suggested to support the process. The method addressed the main safety concerns within a civil aviation system and was compatible with the new state safety management approach proposed by ICAO. Regulators have also enforced the need to implement risk management practices in private organizations in many social regulation areas (Baldwin et al., 2012).

The OECD contributed to the discussion by publishing a study assessing states and CAAs from a governance perspective (Durand & Pietikainen, 2022). The OECD applied its Indicators on the Governance of Sector Regulators to evaluate 29 civil aviation authorities in the Latin American region. The methodology was supported by a survey applied to CAAs, which collected information as indicator scores anchored to the normative framework provided by the OECD Best Practice Principles on the Governance of Regulators. The study's indicators assessed governance from three dimensions: scope of action, independence, and accountability. Their results indicate CAAs' mandates vary in terms of scope of action and highlight the need for strengthened authority, financial and decision-making independence. The study also suggested opportunities for improvement in the arrangements affecting leadership (selection and appointment, termination of mandate, and post-employment restrictions), as well as in strengthening accountability and performance assessment.

Gaps in the Literature

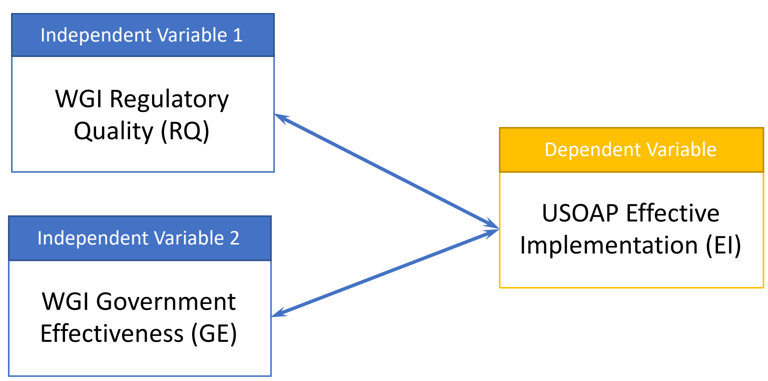
The review of the scholarly literature underscores the sharp contrast between the amount of sources available on the broader aspects of public administration and governance and the scarcity of studies on what makes the same governments better aviation safety regulators. The main initiative currently in place to assess state aviation safety oversight effectiveness is ICAO's USOAP. Even if the program is considered beneficial to aviation safety, the few scholarly articles discussing it and its impact have mainly addressed it from a legal perspective. Additionally, no evaluation of the factors commonly associated with states' effective implementation level in the program was carried out. The current study intends to bridge this gap by supporting the identification of predictors of effective aviation safety oversight from the public governance literature.

Theoretical Framework

To establish the theoretical framework for the research, the concept of theory in quantitative research proposed by Creswell and Creswell (2018, p. 52) is adopted and is defined as an interrelated set of variables or constructs and the hypothesized relationship among them. As no previous theoretical framework was available supporting analysis of governance factors and aviation safety oversight, one is proposed in Figure 1, supported by the recent developments in governance theory. While "more an acknowledgement of the empirical reality of changing times than it is a body of coherent theory" (Frederickson et al., 2012, p. 224), governance theory has been considered "one of the most significant developments" in the study of public administration (Kapucu et al., 2009, p. 41). The governance research framework supported the current study and the hypothesized cross-disciplinary theoretical model associating some of its measures with safety oversight effectiveness.

Figure 1

Theoretical Framework



The WGIs measure governance from six perspectives: voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and corruption. Regulatory Quality (RQ) is the variable capturing the "perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development" (Kaufmann et al., 2010, p. 4). Government Effectiveness (GE) was developed to reflect the perceived "quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (Kaufmann et al., 2010, p. 4). It seems self-evident that the characteristics measured by the two variables would contribute to superior performance in aviation safety oversight effectiveness.

While highlighting the challenges related to specific agencies' effectiveness, the literature provides strong support for the theoretical relationship of broader aspects of policy and standard setting and implementation with aviation safety oversight (Ahmad, 2010; Broderick & Loos, 2002; Groenleer et al., 2010; Jennison, 2013; Joosen et al., 2022; Pierre & Peters, 2009; Schout, 2008; Sønderby, 2016; Swedavia AB & McGregor & Company, 1988). Therefore, WGI's Regulatory Quality (RQ) and Government Effectiveness (GE) metrics were expected to reflect essential governance elements that were directly associated with aviation safety oversight effectiveness. As a result, the proposed structure presented in Figure 1 reflected the relevant variables for the current research and the expected relationship between states' RQ and GE as the independent variables or predictors and aviation safety oversight Effective Implementation (EI) assessed by ICAO as the dependent or target variable. In addition to testing the proposed framework and the expectation that states with more mature governance would present higher aviation safety oversight effectiveness, the study explores the depth of these associations by employing distinct analytical models. As more is learned on governance related to civil aviation safety oversight, new inferences on the relationships that may be present among broader aspects of governance and aviation regulatory process effectiveness can be pursued.

Summary

In Chapter II, an overview of the literature on public administration was presented, along with its waves of academic inquiry, the more recent aspects of governance, and the attempts to quantify good governance. The evolution of aviation safety oversight and the ICAO USOAP-CMA assessment program was also discussed. A theoretical framework was also created to support the structuring of the methodological aspects of the study, which is addressed in more detail in the next chapter.

Chapter III: Methodology

The present chapter addresses the methodological choices that support the identification of a potential association between elements of governance and the effectiveness of aviation safety oversight among states. The chapter also addresses the expectations placed on the exploratory comparison of approaches to studying such association. The advantages and drawbacks of such choices are also discussed in detail.

Research Method Selection

The research questions adopted for the study addressed the problem of civil aviation safety oversight and governance from two perspectives. First, and considering the preliminary stage of research on the topic, an exploratory approach of non-experimental research was taken to identify frameworks in which the relationship between broader aspects of governance and aviation safety oversight could be studied. This step supported the identification of different families of quantitative models that could be employed in evaluating said association. At least three types of models were investigated, including linear, e.g., multiple linear regression, structural equation modeling, and non-linear models, e.g., data-mining techniques.

A second perspective resulted in applying these models to quantitatively test the performance of the different models to the relationship between different governance dimensions and effective implementation of safety oversight. For this second perspective, a quantitative method for nonexperimental research was proposed, supported by archival numeric data generated in studies and assessments conducted by the World Bank and ICAO. Following the research design taxonomy proposed by Edmonds and Kennedy (2017), this perspective fell within an observational approach and an explanatory or predictive design. As all variables were scale/ratio measures, the predictive models could support the identification of a statistically significant relationship between them. If a significant relationship was confirmed, the amount of variation in the dependent variable (EI) explained by changes in the independent variables (RQ and GE) could also be established. Finally, comparing performance among different models would support the increased understanding of the relationship among relevant variables under study.

Population/Sample

Population and Sampling Frame

The population for the present study comprised the set of sovereign states eligible for ratification or adherence to the Convention on International Civil Aviation as enacted in Chapter XXI (ICAO, 1944). The sampling frame was defined by the list of 193 ICAO member states as of August 2022 (ICAO, 2022b). The list of states is presented in Table 2 for reference.

Table 2

List of the ICAO Member States as of August 2022

Afghanistan	Cook Islands	India	Myanmar	Singapore
Albania	Costa Rica	Indonesia	Namibia	Slovakia
Algeria	Côte d'Ivoire	Iran (Islamic Republic of)	Nauru	Slovenia
Andorra	Croatia	Iraq	Nepal	Solomon Islands
Angola	Cuba	Ireland	Netherlands	Somalia
Antigua and Barbuda	Cyprus	Israel	New Zealand	South Africa
Argentina	Czechia	Italy	Nicaragua	South Sudan
Armenia	Democratic People's Republic of Korea	Jamaica	Niger	Spain
Australia	Democratic Republic of the Congo	Japan	Nigeria	Sri Lanka
Austria	Denmark	Jordan	North Macedonia	Sudan
Azerbaijan	Djibouti	Kazakhstan	Norway	Suriname
Bahamas	Dominica	Kenya	Oman	Sweden
Bahrain	Dominican Republic	Kiribati	Pakistan	Switzerland
Bangladesh	Ecuador	Kuwait	Palau	Syrian Arab Republic
Barbados	Egypt	Kyrgyzstan	Panama	Tajikistan
Belarus	El Salvador	Lao People's Democratic Republic	Papua New Guinea	Thailand
Belgium	Equatorial Guinea	Latvia	Paraguay	Timor-Leste
Belize	Eritrea	Lebanon	Peru	Togo
Benin	Estonia	Lesotho	Philippines	Tonga
Bhutan	Eswatini	Liberia	Poland	Trinidad and Tobago
Bolivia (Plurinational State of)	Ethiopia	Libya	Portugal	Tunisia
Bosnia and Herzegovina	Fiji	Lithuania	Qatar	Türkiye
Botswana	Finland	Luxembourg	Republic of Korea	Turkmenistan
Brazil	France	Madagascar	Republic of Moldova	Tuvalu
Brunei Darussalam	Gabon	Malawi	Romania	Uganda
Bulgaria	Gambia	Malaysia	Russian Federation	Ukraine
Burkina Faso	Georgia	Maldives	Rwanda	United Arab Emirates
Burundi	Germany	Mali	Saint Kitts and Nevis	United Kingdom
Cabo Verde	Ghana	Malta	Saint Lucia	United Republic of Tanzania
Cambodia	Greece	Marshall Islands	Saint Vincent and the Grenadines	United States
Cameroon	Grenada	Mauritania	Samoa	Uruguay
Canada	Guatemala	Mauritius	San Marino	Uzbekistan
Central African Republic	Guinea	Mexico	Sao Tome and Principe	Vanuatu
Chad	Guinea-Bissau	Micronesia (Federated States of)	Saudi Arabia	Venezuela (Bolivarian Republic of)
Chile	Guyana	Monaco	Senegal	Viet Nam
China	Haiti	Mongolia	Serbia	Yemen
Colombia	Honduras	Montenegro	Seychelles	Zambia
Comoros	Hungary	Morocco	Sierra Leone	Zimbabwe
Congo	Iceland	Mozambique		

Note. Adapted from “Member States,” by the International Civil Aviation Organization, 2022 (https://www.icao.int/about-icao/Pages/member-states.aspx).

Sampling Strategy

Since the study relies on archival data on the states in the sampling frame, data availability determined the study’s sample. All states with data on all variables were added to the analysis. Chapter I indicated that conditioning the analysis on data availability may be considered a form of convenience sampling, which could challenge the study's validity and reliability. However, assessing that the final sample representativeness can potentially support the expanded generalizability of the study’s results (Campbell & Stanley, 1963). Ensuring sample representativeness was achieved by comparing the demographics of sample and all states, guaranteeing the sample included representation from states in all geographic areas and different levels of economic activity.

Sample Size

Inspection of the data sources used for the analysis confirms that they covered a significant fraction of the countries in the world, with the WGI collecting data on approximately 200 countries (Arndt & Oman, 2006; Kaufmann et al., 2010b) and ICAO USOAP currently carrying information on the effectiveness of over 180 states’ safety oversight (ICAO, 2021a). Even if both datasets do not cover some states, a significant portion of the world's countries was considered in the analysis, supporting the study's expanded generalizability. An adequate assessment of the sample's representativeness of the population also substantiated generalizability. The necessary sample size for the multiple linear regression analysis was calculated using G*Power 3.1 (Faul et al., 2009). For an exact two-tailed test of a linear multiple regression model with two predictors, the needed sample size was 64 data points, considering H_1 ρ^2 estimated to be 0.3, H_0 ρ^2 equals zero, α equals 0.01, and power ($1-\beta$) equals 0.95. These results indicated that the expected sample of over 150 states should satisfy the test's constraints. Having analysis constrained by sample size also impacts both CFA-SEM and data-mining models. Byrne (2016) highlights that sample size, number of estimated parameters, and confidence intervals are interconnected in SEM. Smaller sample sizes can impact the generalizability of multivariate models by contributing to over-fitting and reduced statistical power. To address these concerns, it is generally recommended that dimensionality is reduced by not adding variables to the analysis indiscriminately (Hair et al., 2014). Thus, model complexity must be restricted to ensure narrower confidence intervals in models with moderate samples. Commonly cited SEM references recommend minimum sample sizes larger than 100 cases, larger than 200-400 cases, and larger than 5 or 10 cases per estimated parameter (Anderson & Gerbing, 1984; Bentler & Chou, 1987; Jackson, 2001; Tanaka, 1987). For the present study, minimum sample size for the CFA-SEM analysis was estimated a-priori to be 100 observations using the criteria presented in (Soper, 2024; Westland, 2010) and an estimated effect size of .4, desired power level of .9, and alpha .05.

Data Collection Process

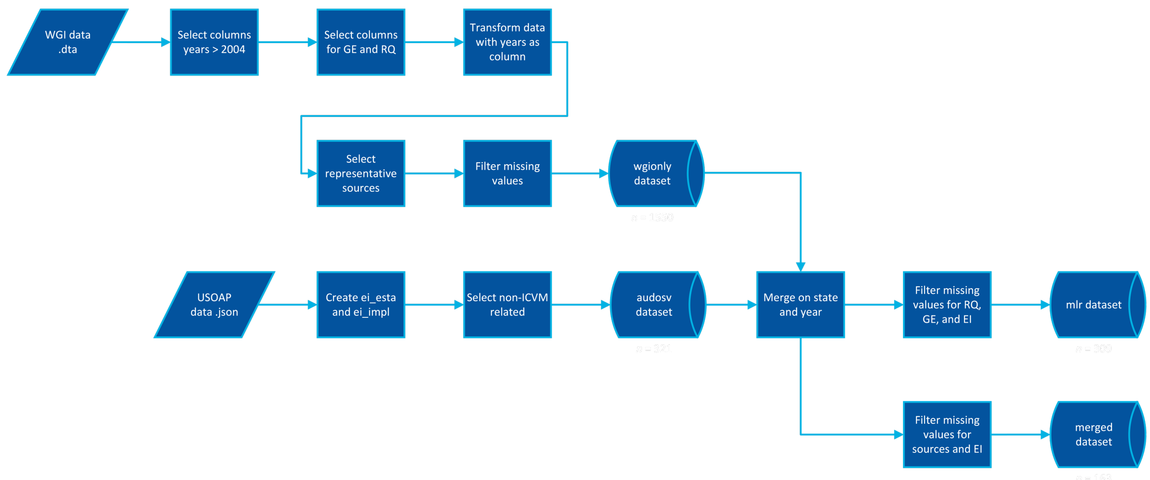
The data used for the study were openly available and obtained directly from the organizations responsible for them. The World Bank's WGI data are consolidated biannually online and are made available for download in several formats (Kaufmann & Kraay, 2023). The dataset that supported the current analysis was generated running the wgicode1.do Stata code made available by the authors in their WGI replication package. The consolidated data was structured in a single spreadsheet with rows for the states and relevant groups of states and columns reflecting different governance dimension metrics calculated for a particular state for each year that the WGIs were calculated. ICAO's USOAP data are available for public use on the Organization's website and through an Application Programming Interface (API) Data Service. With an individual key issued by the organization, these data can be accessed automatically in different forms of aggregation and file formats (ICAO, 2021a, 2021b, 2021c). Some preparation was necessary to adapt the collected World Bank and ICAO datasets from their original format to the format used by the analytical tools. However, no additional significant transformation, including coding, was necessary to support answering the research questions.

Design and Procedures

A multiple linear regression model was developed to support the identification of potential predictors of aviation safety oversight effectiveness among indicators of elements of countries' governance. Attesting compliance to the model's assumptions provided the necessary confidence in the model's results to support analysis. Ensuring assumptions of other linear and non-linear models employed for comparison also supported the study's validity and a better understanding of the models' strengths and weaknesses. An explanation of the process supporting data collection for ICAO audits in support of the construction of EI metrics is provided in Section Measurement Instrument below. A diagram reflecting the steps supporting the generation of the relevant datasets supporting the current research is presented in Figure 2.

Figure 2

Data Preparation Process



Sources of the Data

Both data sources used for the current research were made openly available by the organizations responsible, namely the World Bank and the International Civil Aviation Organization (ICAO, 2019; Kaufmann & Kraay, 2023). With both datasets coming from open sources, no additional approval was necessary for data collection in support of the current study. Detailed reference to the steps used to obtain access to the data used in the study and their respective formats was provided in the previous sections.

Ethical Consideration

The methodology presented in this chapter falls within the regulatory definition of research to assess Institutional Review Board (IRB) review requirements. However, no data on individuals, living or deceased, was used. Therefore, no IRB review was needed, and no IRB application was required per regulations or university policy, namely the 45 CFR 46 and Embry-Riddle Aeronautical University (ERAU) Administrative Policy and Procedure Manual (APPM), Section 12.2 - Human Subject Research Policy.

Measurement Instrument

The World Bank created the six indices related to the dimensions of governance used to assess public governance based on the aggregation of 34 metrics from different sources. The variables related to the Regulatory Quality (RQ) and Government Effectiveness (GE) indices were selected as potential predictors for the proposed study. The purported intent for the GE metric was for it to capture the perceived "quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (Kaufmann et al., 2010b, p. 4). The RQ metric was expected to capture "perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development" (Kaufmann et al., 2010b, p. 4). These metrics were chosen as the potential predictors of the current study as they are more closely related to a state's regulatory capability. In the 2022 edition of the WGI study, the RQ and GE metrics comprised 14 and 17 aggregated metrics, respectively, with six of each set coming from representative sources. The Literature Review discusses some drawbacks of the WGIs and their methodological choices.

ICAO conducts USOAP-CMA audits on states using a set of standard protocol questions. The results of such audits, considering the percentage of compliant items divided by the number of protocol questions applicable to a particular state, are consolidated in the EI metric. The ICAO EI metric can also be organized in audit areas (e.g., aircraft operations, airworthiness, personnel licensing, aerodromes, aircraft accident, incident investigations, and air navigation services). Partial aggregation into critical elements of safety oversight can also provide a relevant perspective of a state's EI (e.g., primary aviation legislation; specific operating regulations; state system and functions; qualified technical personnel; technical guidance, tools, and provision of safety-critical information; licensing, certification, authorization, and approval obligations; surveillance obligations; and resolution of safety issues).

Variables and Scales

All variables were used as presented in their original sources, and no significant additional coding and transformation was deemed necessary after data collection. A consolidated vision of the selected variables is presented in Table 3 for clarification.

Data Analysis Approach

The development of a model for the proposed study compared the performance of three families of linear and non-linear analytical approaches, namely multiple linear regression, structural equation modeling, and data-mining non-linear models. Additional details on the models selected for the current work are provided in the following sections.

Table 3

Description of the Variables under Study

Code	Role	Name	Description	Data Type
RQ	Predictor	Regulatory Quality	Variable capturing the "perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development" (Kaufmann et al., 2010b, p. 4)	Continuous, with approximate range [-2.5; 2.5]
GE	Predictor	Government Effectiveness	Variable capturing the perceived "quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (Kaufmann et al., 2010b, p. 4)	Continuous, with approximate range [-2.5; 2.5]
EI	Target	USOAP Effective Implementation	Variable reflecting safety oversight effective implementation as the percentage of compliant items divided by the number of protocol questions applicable to a particular state as assessed by ICAO's USOAP CMA.	Continuous, with range [0; 100]

Multiple Linear Regression Analysis

Even if "grossly simplified descriptions of complex social reality" (Fox, 2016, p. 2), regression models are some of the most popular and powerful analytical tools used in a wide variety of research problems, including economics, business, social and sciences, among others (Hair et al., 2014). In regression analysis, researchers test the relationship between a single dependent variable (also commonly referred to as criterion or target variable) and several independent (or predictor) variables by considering a regression model. The general form of the linear regression model adopted for the current analysis is indicated by

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j, \tag{4}$$

where an input vector $X^T = (X_1, X_2, \cdots, X_p)$ is used to predict a real-valued output Y, $E(Y|X)$ is approximately linear, β_j are unknown parameters or coefficients, and the variables X_j are quantitative inputs (Hastie et al., 2009).

The general steps of multiple linear regression analysis were used for the first model and coded in a Jupyter Notebook using the Python 3.10 language for data preparation (van Rossum, 1995; Willing et al., 2016) and SPSS Statistics v27 for validation of model assumptions, bootstrapping, and parameter estimation. A multiple linear regression model is proposed to assess whether the RQ and GE dimensions of governance predict states' safety oversight EI. The data collected from the WGI and ICAO original repositories were uploaded to the Jupyter environment for pre-processing. Initial data clean-up ensured consistency of common references across data sources to support data fusion. Analyses were conducted by uploading the .csv dataset to SPSS Statistics, generating the visualizations and metrics needed to test the model assumptions and applying a bootstrapped multiple linear regression procedure. The demographics and general descriptive statistics were calculated for the analysis dataset, and the results reported. The assumptions of the linear regression model were tested with the available data, and the results of such tests registered for reference. In total, eight assumptions were tested for the multiple linear regression model, including (a) a continuous dependent variable, (b) two or more continuous or categorical independent variables, (c) independence of observations, (d) linearity between dependent and independent variables both individually and collectively, (e) homoscedasticity of residuals, (f) lack of multicollinearity, (g) absence of significant outliers, and (h) residuals approximately normally distributed (Berry, 1993; Laerd Statistics, 2015). Bootstrapping used simple sampling, 95% confidence interval, and a total of 1000 samples. The validated model supported the analysis of the degree of association between governance dimensions and aviation oversight implementation and the possibility of rejecting the null hypothesis. The results are presented in Chapter V.

Structural Equation Modeling Analysis

Structural equation modeling (SEM) is a type of multivariate statistical method used to simultaneously test structures of hypothesized relations among measured variables and their latent constructs. The method allows for the identification of support to the hypothesized relationship structure among variables in the study based on its consistency with data, supporting theory testing in non-experimental research (Byrne, 2016). While the application of the method is primarily directed to inferential and confirmatory approaches, prominent authors acknowledge its benefits to exploratory research (Byrne, 2016; Hair et al., 2014; Hair et al., 2022). Covariance-based SEM relies on regression equations to calculate model fit and estimate parameters. Still, the method has additional capabilities when compared to the multiple linear regression method. For instance, SEM can support the analysis of unobserved (i.e., latent) and observed variables and estimate error variance for individual variables in the model (Byrne, 2016). These characteristics increased the popularity of SEM among researchers from several fields (Byrne, 2016; Hair et al., 2022).

Version 27.0 of the IBM SPSS AMOS software was used for the confirmatory factor analyses and the structural equation modeling. Different from the process applied in the MLR analysis, the application of SEM to the analysis of governance and the implementation of aviation safety oversight considered the dimensions of regulatory quality and government effectiveness to be latent factors in the model and the sources used in the WGI structure to be the observed measures. When referring to the governance dimensions as latent factors in the CFA and SEM analysis, they are hereafter referred to as RQ' and GE' in contrast to RQ and GE, used when the values are calculated using the WGI methodology. In (Kaufmann et al., 2011), the authors discuss the allocation of different sources to the dimensions. As presented in Chapter II, some concerns have been raised regarding the WGI methodology and the proposed model's reliability and validity. Thus, confirmatory factor analysis (CFA) was used to test model fit, validity, and reliability measures for the WGI structure relevant to the present study. The assumptions of the WGI methodology provide support for the definition of a reflective measurement model, as the measures in different sources allocated to the dimensions are considered reflections of this construct (Hair et al., 2022; Kaufmann et al., 2011). The development of a confirmatory factor analysis (CFA) model on IBM SPSS Amos supported testing the proposed construct for the problem under study. It also supported the factorial validity of the hypothesized relationship among variables proposed for the study (Byrne, 2016).

Data Mining/Predictive Analysis

To support comparison among different models and identify potential differences in predictive performance, a non-linear data-mining model was also developed based on available WGI and ICAO data. A thorough discussion of the models' results and their implications for the theory and practice of aviation safety oversight followed, as indicated in Chapters IV and V.

The Modeling step in CRISP-DM encompasses four tasks, namely, modeling technique selection, generation of test design, model build, and model assessment. As an attempt to address potential nonlinearities in the data structure, the following modeling techniques were selected for predictive analysis: deep learning, decision tree, and random forest. Because the target variable EI in the model was numeric and continuous, root mean square error rates were selected as the predictive quality measure.

In this step of the analysis, different predictive models are tested and compared by using data mining and predictive analytics. According to Hastie et al. (2009), in predictive statistical learning processes such as the one discussed here, the goal is "to find a useful approximation to the function that underlies the predictive relationship between the inputs and outputs" (p. 27) even if that function is not deterministic. The process used for conducting data mining of the dataset on the relationship of the governance dimension and aviation safety oversight effective implementation was based on the phases defined in the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology and process model (Shearer, 2000; Wirth & Hipp, 2000). While other methodologies may be better suited for specific data analytics platforms (e.g., SEMMA and SAS Enterprise Miner), CRISP-DM is seen as a flexible approach for a wider range of applications and supporting platforms (Palacios et al., 2017; Schröer et al., 2021). Comparative studies indicate CRISP-DM as the "de-facto standard and an industry-independent process model for applying data mining projects" (Schröer et al., 2021, p. 526) and "still fit for purpose for data science projects" after over 20 years of existence (Martínez-Plumed et al., 2021, p. 3048). CRISP-DM proposes a six-phase cycle for data mining processes, including business understanding, data understanding, data preparation, modelling, evaluation, and deployment (Chapman et al., 2000; Shearer, 2000; Wirth & Hipp, 2000). For the study of governance dimensions and safety oversight achieved in this dissertation, the steps concerning business understanding, data understanding, and data preparation were largely those discussed and presented in detail in Chapters I, II, and III. Some complementary aspects of these steps are presented below (e.g., hardware and software resources). This section focuses on the modeling and evaluation steps for the analysis of the RQ and GE dimensions of governance and aviation safety oversight effective implementation. Some aspects of deployment are discussed in Chapter V.

Analysis of all data mining processes was conducted on RapidMiner Studio 10.2 based on either a MacBook Pro 15-inch 2018 with 2.9 GHz 6-Core Intel Core i9 processor and 32 GB 2400 MHz DDR4 memory or a MacBook Pro 16-inch 2021 with Apple M1 Pro processor and 32 GB memory.

In data mining, reliability relates to the quality of the quantitative input data used in the analysis (Odisho, 2020). The issues related to the quality of the WGI and USOAP-CMA datasets have been discussed extensively in Chapter II – Review of the

Relevant Literature. Details were provided on the processes for generation, compilation, and quality assurance for both datasets, and the credibility established for the institutions supporting them. This supported the consideration of data reliability.

Validity is associated with generalizability of the results. It addresses the accuracy of models' predictions in data not used for training. An initial concern to support validity should be on whether the dataset used for training reflects the characteristics seen in population data not available for training or testing. This issue is addressed in the sample representativeness discussions at the end of Chapter IV. A second aspect addresses the model's predictive accuracy in the test dataset (Truong et al., 2018). To improve performance and support generalizability of the results, the testing for all three models employed a combination of model parameter optimization and a bootstrap procedure. Bootstrapping is used to estimate the performance of a model when data availability does not allow for a validation set to be set aside (Hastie et al., 2009). The procedure repeatedly splits the dataset randomly into a training subset (which is used to train the model) and a test subset (to which the model is applied to estimate performance) (Hastie et al., 2009). The resulting performance is estimated by averaging the performance of the different models on the test sets (Xu & Goodacre, 2018).

For the present analysis, the base process was the same for all three model construction scenarios, with the bootstrap validation structure nested within a grid optimization parameter in RapidMiner Studio 10.2. A random seed was defined for the validation operator to ensure different models were tested using the same subsets of training data, following the process presented in (Garcia et al., 2023). The number of validations was fixed at 50, and the sample ratio was defined as .8, with examples not selected for the training set assigned to the test set.

Adopting cross-validation or bootstrapping techniques was expected to contribute to more robust model validation and, possibly, increased confidence in model generalizability (Fox, 2016; Freedman, 1991; Picard & Cook, 1984). While different methods for data-splitting into training and validation datasets for model validation have resulted in similar predictive performance, error estimation sensitivity is said to vary significantly, especially in small datasets (Xu & Goodacre, 2018). The application of bootstrapping was adopted in the current study supporting model generalizability and out-of-sample predictive performance.

Three non-linear predictive modelling techniques were tested in the current research, namely deep learning, decision tree, and random forest. Deep learning denotes a family of novel learning algorithms used to build complex predictive models. Deep learning has been used extensively in a wide range of applications due to its high predictive accuracy (Emmert-Streib et al., 2020). Deep-learning-based models have been shown to present superior performance to state-of-the-art methods in industry benchmarks such as the MNIST handwritten digit classification test (Emmert-Streib et al., 2020; Wan et al., 2013).

The model used for the current analysis is an open-source deep learning model based on multi-layer feedforward neural networks developed by H2O.ai (Candel & Ledell, 2023). The model was created to support businesses in developing AI and deep learning to complex problems and has been increasingly adopted to support research and publications in prominent journal across different fields of research (Anda et al., 2018; Candel & Ledell, 2023; Chen et al., 2023; Daugėla & Vaičiukynas, 2022; Ładyżyński et al., 2019; Senthilkumar et al., 2023). An additional feature of the h2o.ai deep learning model is that it allows for the estimation of input attributes' relative importance, offering an opportunity for enhanced model explainability by considering the weights of individual inputs to the models' initial hidden layers (Candel & Ledell, 2023).

In comparison to neural-network-based models like deep learning, decision trees and are seen as fast, easily applicable to a mixed data types without extensive preprocessing or model tuning and have the added benefit of producing interpretable results (Hastie et al., 2009). Hastie (2009) generalizes decision trees as an approach that produces discontinuous piecewise constant models by portioning the feature space into M regions $R_1, R_2, \ldots, R_M$ and modeling the response in each region with a constant c_m as

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m), \tag{5}$$

where the optimal c_m, for sum of squares $\sum(y_i - f(x_i))^2$, is the average of y_i in the region R_m.

The third predictive analytics approach selected for comparison was random forest, a type of ensemble learning method. Ensemble learning achieves superior predictive performance by combining the results of multiple weaker models (Breiman, 2001; Sirikulviriya & Sinthupinyo, 2011). Random forests are also robust to overfitting (Breiman, 2001; Han et al., 2012; Oshiro et al., 2012).

Reliability and Validity Assessment Method

Per Creswell and Creswell (2018), validity in quantitative research methods relates to whether meaningful and valuable inferences can be drawn from an instrument's results, while reliability refers to instruments' internal consistency and repeatability. The fact that both data sets are generated from third-party organizations imports some of the original studies' reliability and validity issues. A thorough discussion of the issues concerning the WGI is presented by Kaufmann et al. (2007), and Chapter II addressed these concerns following an extensive discussion in the literature on the WGI initiative. Similar considerations apply to the relevance of ICAO USOAP-CMA audit results. ICAO addresses such concerns by implementing a multi-level process to ensure audit results present internal reliability. The process includes a strict selection of auditors with extensive aviation auditing experience and providing classroom and on-the-job training followed by an assessment of auditors' performance. During on-site audits, auditors also communicate with the ICAO office responsible for the USOAP-CMA program to clarify questions that may arise during the process. Finally, before publication, results undergo a final review by ICAO before they are approved and published (ICAO, 2014). Additionally, for the current analysis, ICAO USOAP-CMA data was limited to audits and off-site validations, and results from ICAO coordinated validation missions (ICVM) were excluded to minimize selection bias (Bareinboim et al., 2022). Further assessing the biases of both the World Bank and ICAO (and, potentially, also the researcher) may help clarify threats to the reliability and validity.

Additionally, it must be recognized that observational studies are always constrained in terms of potential causal inferences drawn from the results compared to experimental research. However, the application of true experiments in state capability and governance research is limited. Thus, it is crucial to understand whether relevant external factors to the study may impact the association under analysis. In governance studies, economic, cultural, historical, and geographic factors have been mentioned as contributing to the maturity of state policies and organizations (Andrews et al., 2017; Maurseth, 2008) and could potentially affect safety oversight functions. However, as with any statistical model, potential factors impacting internal reliability and external validity must be addressed to ensure they do not introduce bias or hinder generalizability. Such an approach can ensure that the strength of the design's underlying logic is preserved to ensure confidence in its conclusions (Braverman & Arnold, 2008). This is further supported by the exploratory approach taken in assessing different approaches to data analysis, including the generation of linear and non-linear models. A review of datasets descriptive statistics including some relevant demographic variables supported further understanding of the theoretical framework under study and the aspects associated with selection bias.

Summary

Chapter III presented the most critical methodological choices for the analysis and data sources, their challenges, and potential reliability and validity issues. A detailed presentation of the databases and variables used to support the study of the research questions was also provided. Next, Chapter IV will address the results obtained by adopting the approach presented here.

Chapter IV: Results

This chapter presents the process and results obtained with the exploratory assessment of the relationship between the dimensions of governance selected for the current study, namely regulatory quality (RQ) and government effectiveness (GE), and aviation safety oversight effective implementation (EI). This was achieved by evaluating three different families of analytical approaches, including multiple linear regression, structural equation modelling, and predictive analytics/data mining. Three alternative approaches to data mining were also selected and compared. The details pertaining to the data sources selected to support the analyses are provided in Chapters II and III.

Demographics Results

As mentioned previously, data availability limited the sample size for the methods employed in this dissertation. To minimize data loss, an evaluation of the consolidated dataset was conducted before the application of each analytical approach to identify whether missing data would lead to a reduction in sample size. The dataset used in the multiple linear regression model created to support the investigation of the association of governance dimensions and aviation safety oversight effective implementation was a result of merging the WGI and ICAO USOAP-CMA data and selecting the values for RQ, GE, and EI for each state-year combination available in both datasets. After the removal of observations with missing variables, the process resulted in a dataset with 309 rows.

The initial dataset used to test the WGI structure was comprised of the values for each of the six representative sources allocated to the RQ and GE construct after

2004, resulting in a record with 1550 rows, each row reflecting the associated metrics for a state-year tuple. The time range was limited to that for which USOAP-CMA data was also available. For the structural model, a single dataset set comprised of merged USOAP-CMA results and WGI data for the states was used. The base USOAP-CMA dataset used was collected from ICAO through their API and included both audits and off-site validations. For each data point indicated in the USOAP-CMA dataset, the associated values of the representative sources assigned to RQ and GE for that state and year were collected and merged. Items with missing data for either each of the eight critical elements of safety oversight or the sources associated with the WGI dimensions were excluded from the dataset, leading to the final dataset employed in all analyses for the CFA and SEM analyses with a total of 162 observations. Two additional variables were calculated as an attempt to include an intermediate level of aggregation in USOAP-CMA data between the overall EI and the results for the eight critical elements of safety oversight. These additional variables followed the structure indicated by ICAO in (ICAO, 2017), in which the critical elements one through five are primarily associated with establishing the safety oversight system, and the critical elements six through eight are linked with its implementation. Figure 3 depicts the levels of aggregation in the USOAP-CMA data used in the SEM analyses.

The dataset employed in the data mining analysis is the WGI and USOAP-CMA merged dataset presented in Table 6. The dataset comprised 162 rows representing data points for state-year pairs for which EI and the 12 WGI sources were available. As such, no missing data or outliers were identified. Table 4 below shows the descriptive statistics in the datasets used for the multiple linear regression analyses (ref. mlr dataset). Table 5 shows the descriptive statistics in the dataset used for the CFA (ref. wgionly dataset) and Table 6 presents these set employed in SEM and data mining (ref. wgiaudosv dataset) analyses. Tables 7 and 8 present the distributions of states in the datasets by geographic and income level indicated in the World Development Indicators (WDI) a consolidated development-related dataset established by the World Bank (2023) based on officially recognized international sources.

Figure 3

Aggregation Levels for USOAP-CMA Data

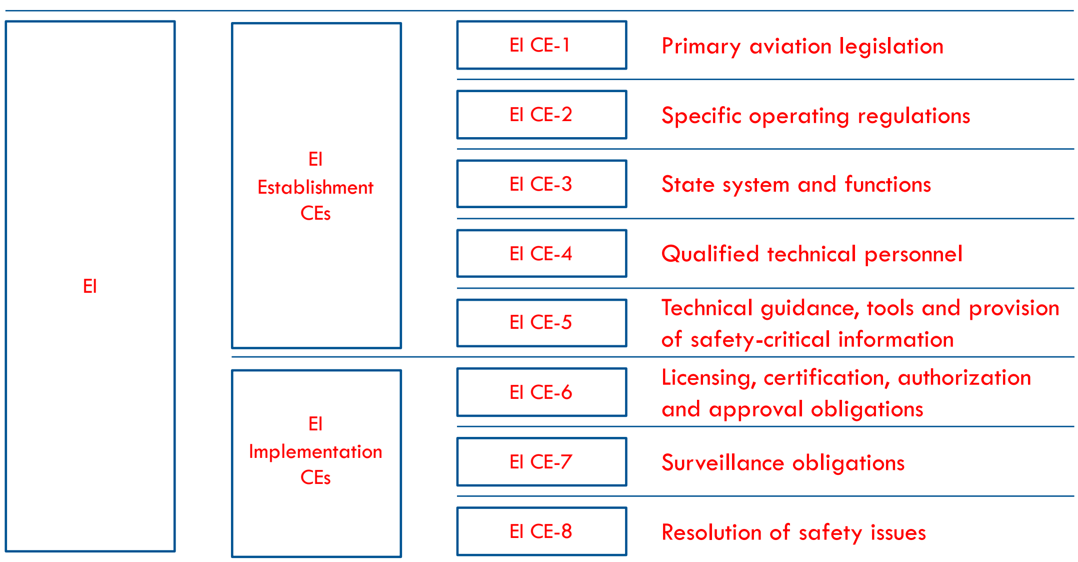


Table 4

Descriptive Statistics for the mlr Dataset

variable	count	mean	std	min	25%	50%	75%	max
ei	309	62.4885	21.8199	4.2918	48.9160	65.4219	81.1978	98.6041
rqerep	309	-0.0702	0.8622	-2.2447	-0.6022	-0.1832	0.5429	1.7863
geerep	309	-0.1217	0.8849	-2.1218	-0.7811	-0.2803	0.4053	2.1087

Table 5

Descriptive Statistics for the wgi Dataset

variable	count	mean	std	min	25%	50%	75%	max
eiurq1	1550	0.5958	0.2080	0.0500	0.4500	0.6000	0.8000	1.0000
gcsrq1	1550	0.5164	0.1010	0.2315	0.4455	0.5054	0.5841	0.8026
herrq1	1550	0.5787	0.1833	0.0500	0.4750	0.6000	0.7250	0.9000
ipdrq1	1550	0.5908	0.1937	0.0833	0.4444	0.5625	0.7500	1.0000
prsrq1	1550	0.7064	0.1811	0.0000	0.5909	0.6818	0.8636	1.0000
wmorq1	1550	0.6331	0.2125	0.0000	0.4375	0.6667	0.8125	1.0000
eiuge1	1550	0.4537	0.2678	0.0000	0.2500	0.3750	0.6250	1.0000
gcsge1	1550	0.5233	0.1702	0.1278	0.3922	0.5071	0.6618	0.9496
gwpge1	1550	0.5881	0.1274	0.1350	0.5043	0.6000	0.6800	0.9398
ipdge1	1550	0.6204	0.2884	0.0000	0.3750	0.6250	0.9167	1.0000
prsge1	1550	0.6035	0.2558	0.0000	0.5000	0.5000	0.7500	1.0000
wmoge1	1550	0.6486	0.1972	0.1111	0.5000	0.6667	0.8125	1.0000

Table 6

Descriptive Statistics for the merged Dataset

variable	count	mean	std	min	25%	50%	75%	max
eiurq1	162	0.5543	0.1996	0.0500	0.4000	0.5500	0.7000	0.9500
gcsrq1	162	0.4982	0.0948	0.3051	0.4371	0.4802	0.5618	0.7696
herrq1	162	0.5503	0.1739	0.0500	0.4500	0.5500	0.6750	0.8750
ipdrq1	162	0.5662	0.1878	0.0833	0.4375	0.5341	0.7083	0.9583
prsrq1	162	0.6883	0.1861	0.0000	0.5909	0.6818	0.8182	1.0000
wmorq1	162	0.5966	0.2166	0.1667	0.4167	0.5833	0.8125	1.0000
eiuge1	162	0.4028	0.2566	0.0000	0.2500	0.3750	0.5000	1.0000
gcsge1	162	0.4773	0.1676	0.1601	0.3497	0.4774	0.5809	0.9062
gwpge1	162	0.5690	0.1307	0.1350	0.4705	0.5889	0.6696	0.9000
ipdge1	162	0.5401	0.2813	0.0000	0.3333	0.5000	0.7917	1.0000
prsge1	162	0.5478	0.2592	0.0000	0.3750	0.5000	0.7500	1.0000
wmoge1	162	0.6090	0.1931	0.2500	0.4444	0.6111	0.7500	1.0000
ce1	162	78.0342	17.3821	6.4516	68.7500	81.2500	90.3226	100.0000
ce2	162	76.8809	15.4369	5.3191	70.6596	79.6604	87.6264	99.0741
ce3	162	70.6836	20.4426	8.0000	60.2740	74.0372	86.5854	100.0000
ce4	162	55.0141	27.0074	0.0000	33.5821	56.7253	79.2255	100.0000
ce5	162	72.9723	19.2945	0.8197	62.6567	75.9982	88.3188	98.5714
ce6	162	73.3774	20.8373	0.9901	61.7829	79.2710	88.9709	99.5671
ce7	162	61.5764	22.7396	0.0000	44.6970	62.2664	79.6768	100.0000
ce8	162	57.7678	25.6327	2.3256	39.5470	60.4651	79.4265	100.0000
ei_esta	162	70.7170	17.6173	9.6813	61.7313	73.1435	84.3718	99.0096
ei_impl	162	64.2405	22.2271	1.1052	49.8237	68.4137	82.8186	99.1342
ei	162	69.9010	18.3874	5.9155	59.7016	71.8066	85.0670	98.6041

Table 7

Dataset Distribution of States by Geographic Region.

Region	wdi	merged	mlr	audosv
East Asia & Pacific	17.4%	11.7%	14.2%	15.0%
Europe & Central Asia	26.6%	30.2%	25.6%	26.3%
Latin America & Caribbean	19.3%	18.5%	18.4%	17.8%
Middle East & North Africa	9.6%	11.1%	9.4%	9.1%
North America	1.4%	0.6%	0.6%	0.6%
South Asia	3.7%	4.3%	4.2%	4.1%
Sub-Saharan Africa	22.0%	23.5%	27.5%	27.2%

Table 8

Dataset Distribution of States by Income Level.

Income level	wdi	wgiaudosv	mlr	audosv
High income	38.1%	30.9%	26.5%	27.5%
Low income	11.9%	6.8%	11.0%	11.3%
Lower middle income	24.8%	32.7%	34.0%	33.1%
Upper middle income	24.8%	29.0%	28.2%	27.8%
Unavailable	0.5%	0.6%	0.3%	0.3%

Note. States for which the WDI did not indicate an income level are presented as "Unvailable".

Validity and sample representativeness

To test whether the process of selective exclusion of missing data in the analysis datasets led to important differences in terms of sample representativeness, six demographics data associated with each country were selected from the WDI (World Bank, 2023) and means and distributions compared across the datasets. These variables were GDP (current US$), GDP per capita (current US$), land area (sq. km), total population, air passengers carried, air freight carried (million-ton km). The corresponding values for each variable were merged to each data point of the dataset used in the current analysis. Visual inspection of the kernel density estimates plots for the six variables did not indicate important differences among datasets (see Figure 16 in Appendix A). Further support was provided by the results of a one-way ANOVA with the three datasets, one reflecting the USOAP-CMA dataset (audosv: n = 320), one reflecting the data used for the MLR model (mlr: n = 309), and one reflecting the merged dataset used in the SEM and data mining models (wgiaudosv: n = 162). States with missing WDI data were evenly excluded from each of the datasets. For all six variables, the data did not support an alternative hypothesis indicating differences in means in the three datasets (GDP: p = .170, GDP per capita: p = .425, land area: p = .123, population: p = .088, air passengers carried: p = .252, air freight carried: p = .582). Individual analyses were followed by pairwise group comparisons using Tukey's Honestly Significant Difference (HSD) test and all resulting p-values were also above .05. The use of non-parametric models (e.g., Kruskal-Wallis) was dismissed as the incidence of tied ranks would have created additional complexity.

Multiple Linear Regression

The initial exploration of the association of dimensions of governance and aviation safety oversight was conducted using a multiple linear regression (MLR) model that employed the WGI RQ and GE dimensions as predictors and USOAP-CMA EI as target. The results of the application of the model are presented in the coming sections.

Data Description

The application of multiple linear regression employed the mlr dataset presented with values for country code, RQ, GE, and EI for each state-year combination available in both datasets (N = 309). Summary statistics for the dataset are presented in Table 4.

Reliability and Validity Testing Results

The assumptions for the MLR model were validated and the results presented below. There was no independence of residuals, as assessed by a Durbin-Watson statistic of 1.599 (Field, 2009). Durbin-Watson also contributed to the assumption of independence of observations. The assumption of linearity was tested, first, by ensuring that independent variables showed collective linear relationship the dependent variable by inspection of the scatterplot of the studentized residuals against the unstandardized predicted values. Additionally, no significant non-linearities were found when inspecting the scatter plots of independent variables (RQ and GE) against the dependent variable (EI). Visual inspection of a plot of studentized residuals versus unstandardized predicted values supported the validation of homoscedasticity. Multicollinearity was considered acceptable, with a VIF 4.796, based on the references proposed by (Hair et al., 2014). No outliers were identified, with the highest studentized residuals being -2.663 and 2.051, highest leverage point .0404 (Huber, 1981), and highest Cook's distance less than .0466 (Cook & Weisberg, 1982). The highest calculated Mahalanobis distance was 12.454 ($p > .001$), further indicating the absence of points improbably far from the centroid, under normality assumptions (Byrne, 2016; Hairet al., 2014). Finally, inspection of the normal P-P plot of the standardized regression residual and the normal Q-Q plot of the studentized residual supported the assumption that the residuals are approximately normally distributed.

The coefficient of determination R^2 for the overall model was 18.5% with an adjusted R^2 of 17.9%. As an estimate of effect size, Cohen (1988) places this value of adjusted R^2 as a medium effect size. This indicates the proportion of variance on the dependent variable (EI) that can be explained by the variance in the independent variables in the model (RQ and GE). Even if the complex interactions of factors that can play a role on increasing EI levels in a state are accounted for, these results indicate that governance, more specifically government effectiveness, is significantly positively associated with aviation safety oversight effective implementation.

Hypothesis Testing Results

A multiple regression was run in SPSS to predict aviation safety oversight effective implementation from the regulatory quality and government effectiveness dimensions of governance. Assumptions of a multiple linear regression model were assessed as indicated in the previous section. For the multiple linear regression model being tested, the null hypothesis H_0 tested was the following.

H_0: There is no linear relationship between RQ or GE and EI.

The multiple regression model statistically significantly predicted EI, $F(2, 306) = 34.660$, $p < .001$, adj. $R^2 = .179$. Only GE added statistically significantly to the prediction, $p < .05$, supporting the rejection of the null hypothesis H_0. Model coefficients indicate a rate of 8.505-point increase in EI for a unit increase in GE, 95% CI [3.020, 13.991]. Regression coefficients and standard errors can be found in Table 9.

Table 9

Multiple Regression Results for EI

EI	*B*	95% CI for *B*		*SE B*	β	R^2	ΔR^2
		LL	*UL*				
Model						.185	.179***
Constant	63.689	61.446	65.446	1.140			
RQ	2.354*	-3.276	7.983	2.861	.093*		
GE	8.505**	3.020	13.991	2.788	.345**		

Note. Model = “Enter” method in SPSS Statistics; B = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; β = standardized coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2.

*$p = .411$, **$p < .005$ ***$p < .001$.

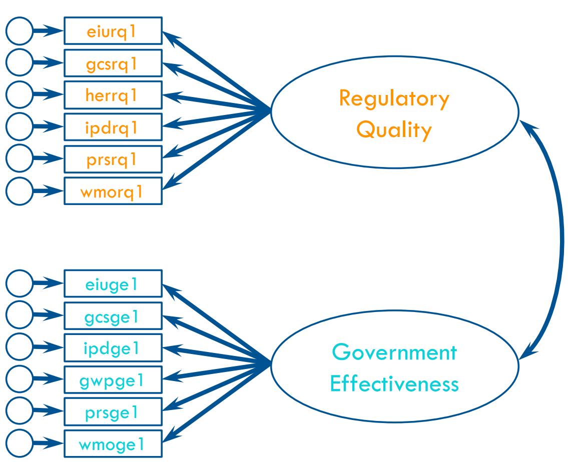
A preliminary analysis of the relationship of governance and aviation safety oversight using multiple linear regression provided initial support for exploration of the relationship between governance dimensions and safety oversight implementation by indicating a positive association between aggregated measure of government effectivess and safety oversight effective implementation. With the application of the MLR model, no significant support was found in the data to the hypothesis that regulatory quality would also be positively associated with EI. These results provided added support to the need for a more detailed investigation of the factors involved in said association, which is conducted employing structural equation modeling in the next section.

Structural Equation Model

An initial confirmatory factor analysis (CFA) was used to test model fit, validity, and reliability measures for the WGI structure relevant to the present study. A visual representation of the WGI model (model one) tested is presented in Figure 4.

Figure 4

Initial CFA Model (Model One)



Data Description

The evaluation of the proposed theoretical framework was conducted sequentially, with initial testing and re-specification of a CFA model using WGI sources, followed by the testing of the complete model including USOAP-CMA data in the full structural model using SEM. The WGI dataset indicated in Table 5 (*N* = 1550) was employed for the CFA analysis and the merged dataset (*N* = 162) used for the SEM analyses.

Reliability and Validity Testing Results

A confirmatory factor analysis (CFA) of the WGI measurement model for the RQ and GE dimensions presented in Figure 4 was conducted by using the wgi dataset, and the elements of reliability, validity, and model fit were assessed following the references in (Byrne, 2016; Hair et al., 2014; Hu & Bentler, 1999; Pan & Truong, 2018; Truong et al., 2020). The following model fit indices and thresholds were selected as reference for the analyses: comparative fit index (CFI) > .93, goodness-of-fit index (GFI) > .90, adjusted goodness of fit (AGFI) > .90, normed fit index (NFI) > .90, and normed chi-square (chi-square/df) <= 3. Some research has indicated concerns with the dependence of chi-square/df and Root mean square error of approximation (RMSEA) on sample size and degrees of freedom while others have indicated RMSEA to support "good" fit in the .5-.8 range (Byrne, 2016; Hair et al., 2014; Kenny et al., 2015; Maccallum et al., 1996).

The measurement models' convergent validity, internal consistency reliability and discriminant validity were also assessed. Convergent validity was tested by assessing the observed variables' loadings (> .7) and indicator reliability (> .5) and the factors' average variance extracted (AVE > .5) (Hair et al., 2022). Hair et al. (2022) acknowledge, however, that in the early stages of scale development, factor loadings below .7 are acceptable, and their retention could support content validity. However, the authors highlight that loadings below .4 should be removed from the model. Internal consistency reliability of the models' factors was tested using Cronbach's alpha (.60-.90) and composite reliability (.60-.90). Finally, discriminant validity was tested using the heterotrait-monotrait ratio (HTMT < .9), deemed superior to the traditionally adopted Fornell-Larcker method (Fornell & Larcker, 1981; Hair et al., 2022; Henseler et al., 2015). The results for the initial theoretical structure presented in Figure 1, as defined by Kauffman et al. (2011) and referred to here as the model one, are presented in Table 10.

The overall model fit and discriminant validity were deemed insufficient, with the highest modification indices (MI) for the covariances associated with the error terms for gcsrq1-gcsge1 and wmorq1-wmoge1. In the model proposed by (Kaufmann et al., 2011), distinct source variable subsets are assumed to load into different unobserved components, which reduces the chance of introducing artificial cross-loadings across factors. On the other hand, common influences associated with individual source methodologies could contribute to the elevated MI values, which, in turn, could provide practical-theoretical justification for the consideration of said covariances in the model. To test this effect, a second full structural model referred to as model two and presented in Figure 5 was evaluated, including covariances across said error terms. The results are shown in Table 11.

Table 10

Confirmatory Model Fit and Reliability and Validity Assessment (Model One)

Factors	Items	Standardized Factor Loading	Indicator Reliability	AVE	Cronbach's alpha	Composite Reliability
RQ'	eiurq1	.899	.808	.685	.922	.9976
	gcsrq1	.744	.554			
	herrq1	.821	.674			
	ipdrq1	.774	.599			
	prsrq1	.803	.645			
	wmorq1	.913	.834			
GE'	eiuge1	.921	.848	.680	.916	.9964
	gcsge1	.856	.733			
	ipdge1	.791	.625			
	gwpge1	.548	.300			
	prsge1	.845	.714			
	wmoge1	.926	.857			

Note. Chi-square = 2031.275, df = 53, $p<.001$. HTMT=.9365. Fit indices: GFI=.815; AGFI=.728; NFI=.893; CFI=.896; chi-square/df=38.326; RMSEA=.155.

Figure 5

CFA Model with Error Covariances (Model Two)

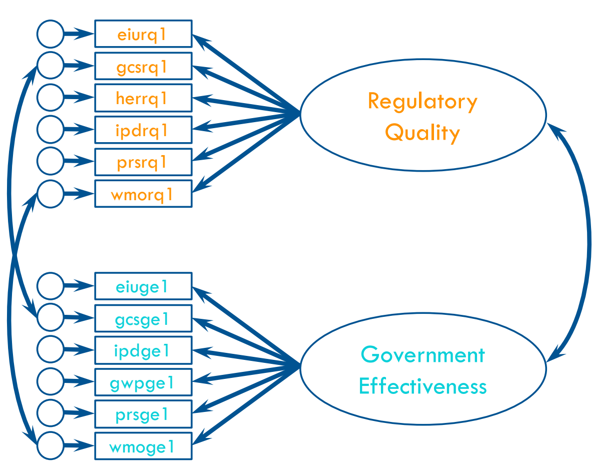


Table 11

Confirmatory Model Fit and Reliability and Validity Assessment (Model Two)

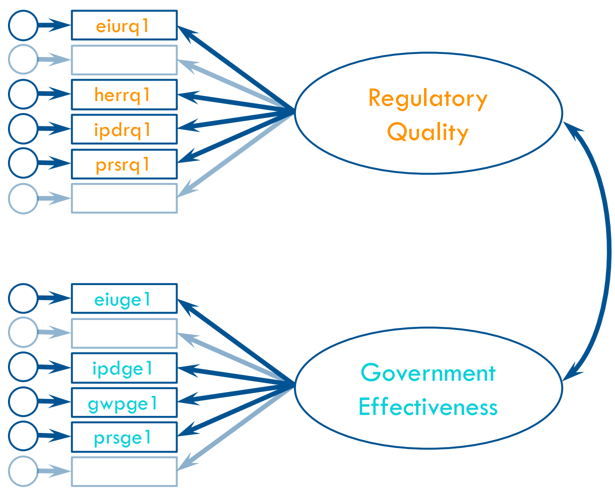
Factors	Items	Standardized Factor Loadings	Indicator Reliability	AVE	Cronbach's alpha	Composite Reliability
RQ'	eiurq1	.907	.822	.684	.922	.998
	gcsrq1	.731	.534			
	herrq1	.830	.689			
	ipdrq1	.784	.615			
	prsrq1	.801	.642			
	wmorq1	.897	.805			
GE'	eiuge1	.928	.861	.677	.916	.996
	gcsge1	.846	.716			
	ipdge1	.794	.630			
	gwpge1	.545	.297			
	prsge1	.852	.726			
	wmoge1	.912	.831			

Note. Chi-square = 1362.311, df = 51, $p < .001$. HTMT = .9365. Fit indices: GFI = .861; AGFI = .788; NFI = .929; CFI = .931; chi-square/df = 26.712; RMSEA=.129.

The model fit parameters showed overall improvement when compared to model one, but could still be considered unacceptable. Some discussion on the appropriateness of adding covariances across error terms persists in the literature, particularly when done across distinct factors (Hermida, 2015). Even if novel approaches have been proposed to relax error independence assumptions in the WGIs (Magnusson & Tarverdi, 2020), an alternative parsimonious modification to the model was then tested, where the variables associated with gcs and wmo sources were removed. This is indicated in the proposed model three presented in Figure 6.

Figure 6

Modified CFA Model (Model Three)



The results for model three are presented in Table 12. The results showed overall superior results, with GFI, NFI, and CFI within aceptable ranges, and AGFI marginally acceptable. The calculated values for chi-square/df and RMSEA were outside the expected range but, as mentioned previously, literature indicates concerns with their use in problems with small df and overly small or high sample sizes (Byrne, 2016; Hair et al., 2014; Kenny et al., 2015). Kenny et al. (2015) have indicated that, in "models with small df, the RMSEA can exceed cutoffs very often, even when the model is correctly specifed" (p. 501). Model three presented a marginally acceptable model fit and was selected to support further exploration of the factors associated with the WGI dimensions of governance and aviation safety oversight effective implementation. The chi-square value of the final specified model was 416.662 (df = 19, $p < .001$).

Table 12

Confirmatory Model Fit and Reliability and Validity Assessment (Model Three)

Factors	Items	Standardized Loading	Indicator Reliability	AVE	Cronbach's alpha	Composite Reliability
RQ'	eiurq1	.925	.8556	.712	.906	.9964
	herrq1	.857	.7344			
	ipdrq1	.804	.6464			
	prsrq1	.781	.6100			
GE'	eiuge1	.949	.9006	.631	.857	.9928
	ipdge1	.768	.5898			
	gwpge1	.537	.2884			
	prsge1	.863	.7448			

Note. HTMT = .8804. Fit indices: GFI = .938; AGFI = .882; NFI = .958; CFI = .960; chi-square/df = 21.930; RMSEA = .116.

Model three was used for data imputation in AMOS to obtain RQ' and GE' values and compare them with the original WGI calculated RQ and GE. Normalized RQ' and RQ showed an R^2 of .9631 and a two-sample Kolmogorov-Smirnov test did not support the rejection of the null hypothesis indicating differences in variables'distributions (p = .842). A similar test of normalized GE' and GE resulted in an R^2 of .9417 and also non-significant KS statistic (p = .580), supporting the assumptions of a lack of significant differences in estimated values distribution when compared to the WGI calculated values. In other words, even after model modification, the imputed values for unobserved RQ'and GE' indicated no significant differences with the original WGI values for RQ and GE. Figures 7 and 8 below show the scatter plots and empirical cumulative distribution functions for normalized model three imputed and WGI generated values.

Figure 7

Scatterplots for Normalized model three Imputed and WGI Generated RQ and GE Values

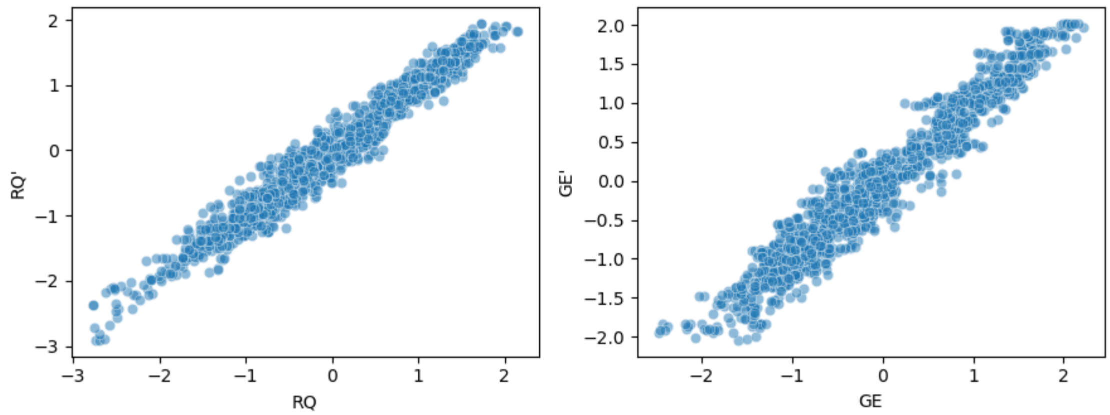
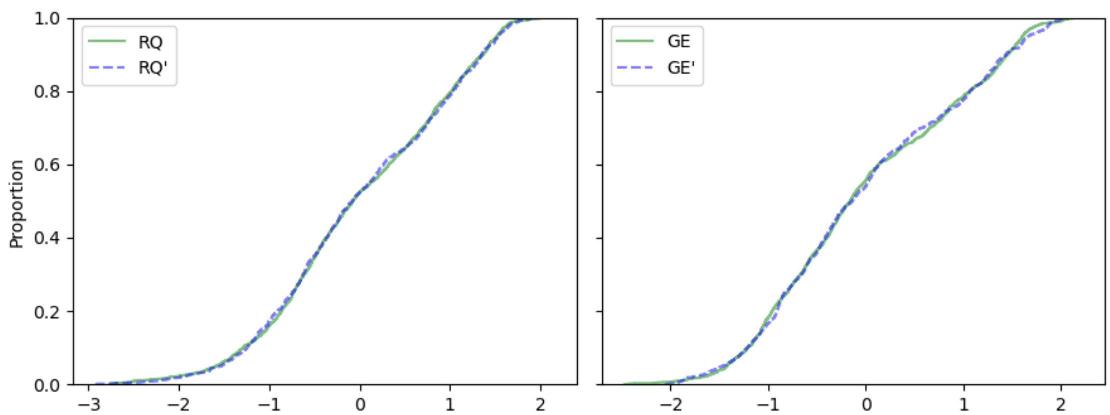


Figure 8

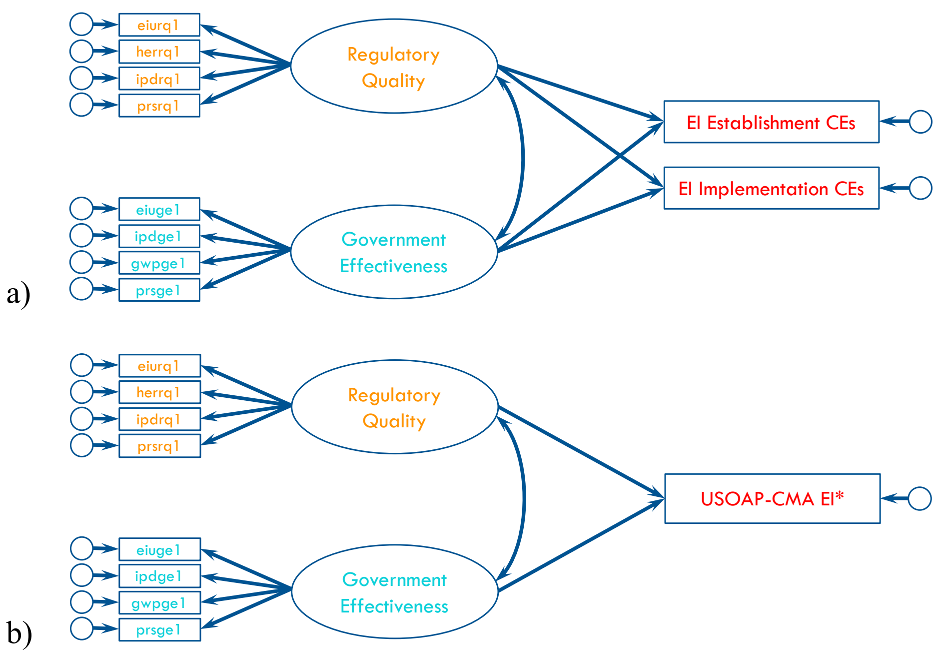
Empirical Cumulative Distribution Functions for Normalized model three Imputed and WGI Generated RQ and GE Values



The final specified CFA model was used as the foundation in the assessment of the structural models created to test the relationship of the RQ' and GE' dimensions of governance and safety oversight effective implementation. In the SEM models tested, one-way arrows represented the hypotheses for testing. A covariance arrow was added connecting exogenous factors RQ'and GE'. Two full structural models were tested, as presented in Figure 9. Their distinction was related to whether one or two variables were included in the right hand side of the model to represent the USOAP-CMA oversight effective implementation.

Figure 9

Full Models Used in the Structural Model Assessment



The consolidated USOAP-CMA metric used in audit reports is the effective implementation (EI) indicator, but the variable reflects a higher level of aggregation of more granular audit results. ICAO (2017) also publishes a breakdown of audit results into the eight critical elements (CEs) of safety oversight. According to the Organization (2017), the CEs represent "the essential components of a [s]tate safety oversight system" (p. v) and are needed for "the effective and sustainable implementation of a safety-related policy and associated procedures" (p. 2-5). ICAO (2017) defines the structure for the eight CEs as: primary aviation legislation (CE-1); specific operating regulations (CE-2); state system and functions (CE-3); qualified technical personnel (CE-4); technical guidance, tools and provision of safety-critical information (CE-5); licensing, certification, authorization, and approval obligations (CE-6); surveillance obligations (CE-7); and resolution of safety issues (CE-8). While remarking on some interconnectedness, ICAO (2017) considers CEs one through five as "establishment CEs" and six through eight as "implementation CEs" (p. 2-5). This concept is used to create an intermediate aggregation level with additional variables by averaging the results of the establishment (EI_e) and implementation (EI_i) CEs for each country. Figure 3 depicts this relationship.

For testing the structural models, an initial full structural model with two variables representing the EI_e and EI_i was assessed (ref. Figure 9-a) with the merged WGI USOAP-CMA dataset employed (n = 162). All of the model's fit indices now approached the selected thresholds, but the overall model fit for the proposed model was still deemed insufficient (GFI = .861; AGFI = .761; NFI = .895; CFI = .917; chi-square/df = 4.097; RMSEA = .139).

As such, testing proceeded with individual consideration of USOAP-CMA EI variables using the base model presented in Figure 9-b. Initially, tests for EI, EI_e, and EI_i were conducted. The results are presented in Table 13, and the following section presents a more detailed discussion of these results and the associated hypotheses. For all three, model fit was considered acceptable.

Table 13

Structural Model Results for Individual USOA-CMA Variables EI, EI_i, and EI_e

		EI*		
Hypothesis	Indicator	EI	EI_e	EI_i
RQ'→EI*	SRW	-.257	-.231	-.183
	t-value (CR)	-1.291	-1.127	-.970
	p-value	.197	.260	.332
	Result	Not sup.	Not sup.	Not sup.
GE'→EI*	SRW	.758	.623	.778
	t-value (CR)	3.747	3.013	4.043
	p-value	< .001	.003	< .001
	Result	Supported	Supported	Supported

Note. Fit indices: EI (chi-square = 62.791, df = 25, $p < .001$; GFI = .927, AGFI = .868, NFI = .938, CFI = .962, chi-square/*df* = 2.512., RMSEA = .097), EI_e (chi-square = 60.153, df = 25, $p < .001$; GFI = .931, AGFI = .877, NFI = .940, CFI = .963, chi-square/*df* = 2.406., RMSEA = .093). and EI_i (chi-square = 68.309, df = 25, $p < .001$; GFI = .919, AGFI = .854, NFI = .935, CFI = .957, chi-square/*df* = 2.732, RMSEA = .102).

Hypothesis Testing Results

The results presented in Table 13 provided additional insights with regards to the relationship of the governance dimensions RQ' and GE' and three different metrics reflecting USOAP-CMA assessments of EI. Two main hypotheses were tested for the three analyses, where EI* indicates the choice of EI, EI_e, or EI_i.

H_1: RQ' positively influences EI*.

H_2: GE' positively influences EI*.

For the analysis involving the aggregated overall USOAP-CMA EI score, Hypothesis 1 (H_1) was not supported, with a *p*-value larger than .05 (p=.204) and a *t*-value greater than 1.96. Hypothesis 2 (H_2) was supported, indicating that GE' positively influences EI. The *p*-value for this association was less than 0.5 ($p < .001$), and the *t*-value was greater than 1.96. These results indicate consistency with those obtained with the MLR tests, in which support was found for the hypotheses indicating an association between GE and EI, but not between RQ and EI. However, the added complexity of the model allowed it to explain 29.5% of the variance in EI, a larger portion when compared to the 17.9% optained with the MLR model. Employing the standardized regression weight for the GE'→EI as a measure of the estimated effect size would rank it as a large effect (Cohen, 1988).

A new segmented test was then conducted to evaluate the association with specific elements of the EI metric, namely the EI associated with the Establishment CEs or (EI_e) and the EI associated with the Implementation CEs (EI_i). When testing for EI_e, Hypothesis 1 (H_1) was not supported with a *p*-value above .05 (p = .260) and *t*-value greater than 1.96. A *p*-value below .05 (p = .003) and a t-value above 1.96 supported Hypothesis 2 (H_2). Similar results were obtained with EI_i selected as the right-hand side variable. Hypothesis 1 (H_1) was not supported (p = .332; $t < 1.96$), and Hypothesis 2 (H_2) was supported ($p < .001$; $t > 1.96$).

A final structural test explored a narrower aspect of the theoretical framework under investigation. The RQ dimension of governance is purported to address the state's regulatory process more specifically, namely “the ability of the government to formulate and implement sound policies and regulations” (Kaufmann et al., 2011, p. 223). The ICAO USOAP-CMA framework for aviation safety oversight implementation includes metrics more directly associated with the outcomes of the states’ legislative and regulatory process in CEs one and two. In order to test whether RQ’ could be associated with narrower regulatory-focused elements of aviation safety oversight, two additional models were tested with the states’ results for EI in primary aviation legislation (CE-1; chi-square = 58.788, df = 25, $p < .001$), and specific operating regulations (CE-2; chi-square = 58.592, df = 25, $p < .001$). In both tests, Hypothesis 1 was not supported ($EI_{CE\text{-}1}$: $p = .337$; $t = .959$; $EI_{CE\text{-}2}$: $p = .156$; $t = -1.419$). Hypothesis 2 was not supported for $EI_{CE\text{-}1}$ ($p = .472$; $t = .719$) and supported for $EI_{CE\text{-}2}$ ($p = .007$; $t = 2.703$).

Predictive Analytics/Data Mining

This section presents the results obtained with the exploratory analysis of the relationship governance-safety oversight using nonlinear data mining methods. Three families of models are employed and compared, including deep learning, decision tree, and random forest.

Data Description

The dataset employed for the data mining analyses was the merged dataset utilized in the SEM analyses with descriptive statistics shown in Table 6 and comprised of 162 observations.

Reliability and Validity Testing Results

An extensive discussion on the issues of data mining process reliability and validity was presented in Chapter III. Issues related to data quality that could potentially impact the study's reliability were addressed both for the WGI and USOAP-CMA in detail in Chapters III and IV, supporting the current analysis' reliability. The comparison in demographics for the three datasets was also addressed in the previous sections and the adoption of bootstrapping to improve the model's out-of-sample performance and overall generalizability also supports the current analysis' validity.

Modeling and Evaluation

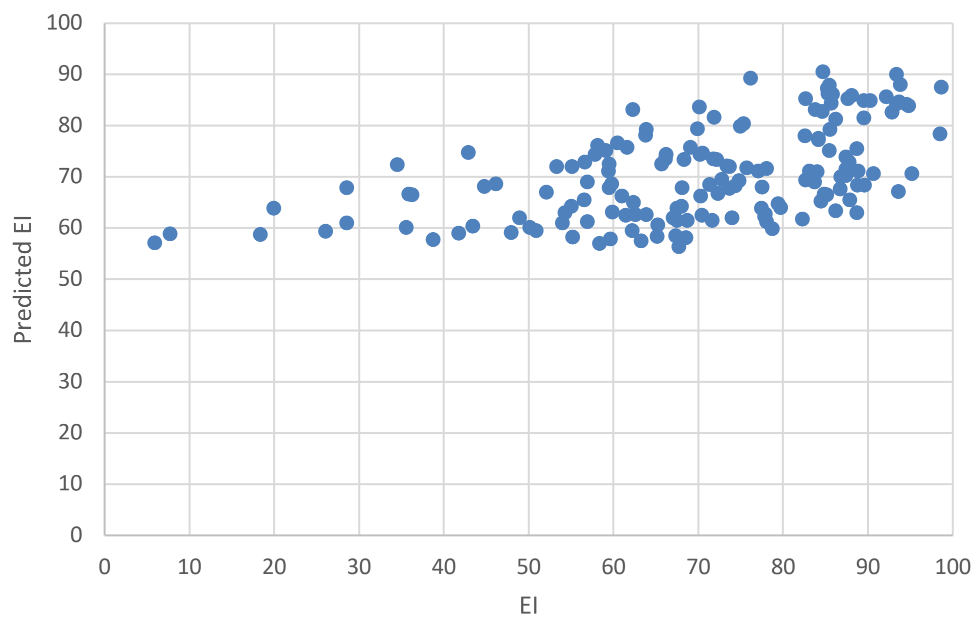
Deep Learning. The deep learning algorithm used in the current analysis was based on the H2O package ver. 3.30.0.1. The proposed base deep learning model was defined as having three hidden layers, each having varying numbers of the hidden neurons (two, four, eight, 16, 32, and 64) and types of hidden neuron activation functions (Tanh, Rectifier, Maxout, and ExpRectifier), and adaptive rate enabled. Parameter combinations leading to numerical instability were discarded. The remaining parameters were maintained at the default values.

The winning deep learning model, with hidden layers presenting eight, four, and 16 neurons and the ExpRectifier activation function, showed a root mean square error (RMSE) of 16.649 and standard deviation of 1.230. It is noted however, that RMSE did not change considerably across different combinations of parameter configuration. The deep learning model's explanatory capability based on the method proposed by Gedeon (1997) resulted in the following order of parameters and relative importance between parentheses: wmorq1 (1.0000), gwpge1 (0.8970), ipdrq1 (0.8688), wmoge1 (0.8534), herrq1 (0.82328), ipdge1 (0.7917), eiuge1 (0.789407), gcsge1 (0.7569), eiurq1 (0.7454), prsrq1 (0.7067), prsge1 (0.6377), and gcsrq1 (0.5900). It is interesting to note that variables wmorq1 and wmoge1 that were excluded in the final SEM model showed comparatively high relative importance in the deep learning model.

The results for the winning deep learning model are presented in Figure 9. Visual inspection of the relationship between actual EI and the values predicted by the model indicates poor predictive performance in low EI levels. As such, other modeling techniques were tested for comparison.

Figure 10

True and Predicted Values for EI with the Winning Deep Learning Model



Decision Tree. The application of the decision tree modeling technique to the merged dataset employed a least square approach to minimize the squared distance between the average values in the node and the true value. Maximal depth varied from two to 20 in two-unit steps and the remaining parameters remained with the default values. The comparative results for RMSE are shown in Figure 10. The winning model had a maximal depth of four and presented an RMSE of 17.869 with standard deviation of 1.391. The structure of the winning decision tree model is presented in Figure 11. The resulting decision tree model is considerably simpler than the deep learning model and, while indicating higher RMSE values compared to the deep learning model, it could better capture lower EI points using ipdge1 and wmoge1 as splitting variables. The scatter plot for predicted and actual values for the winning decision tree model are presented in Figure 12.

Figure 11

RMSE for Decision Tree Models with Distinct Maximal Depth

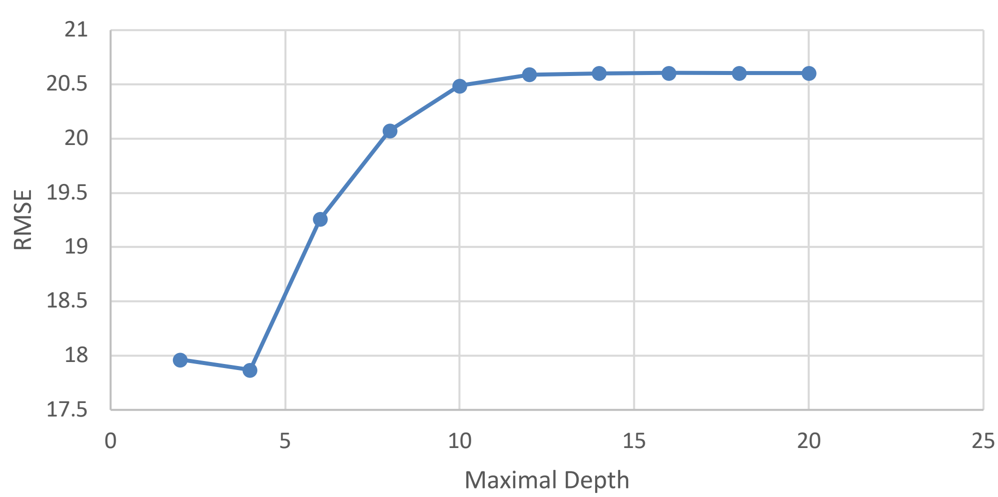


Figure 12

Winning Decision Tree Model

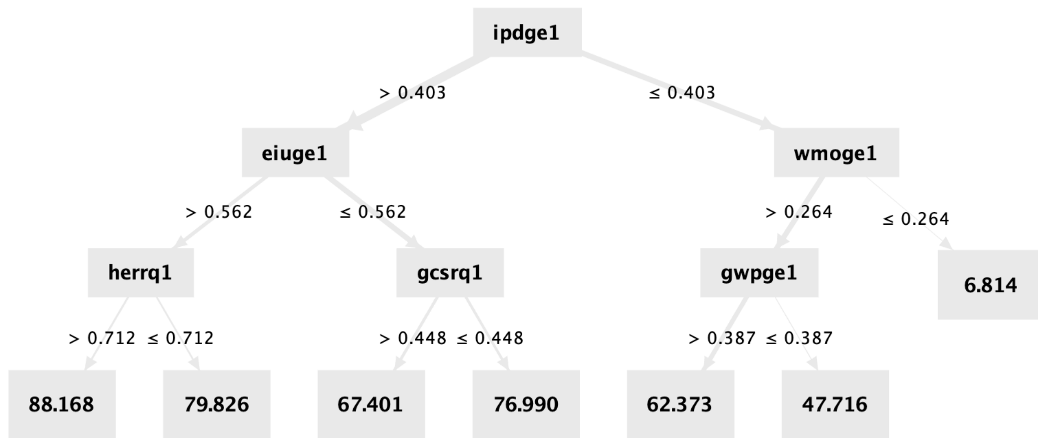
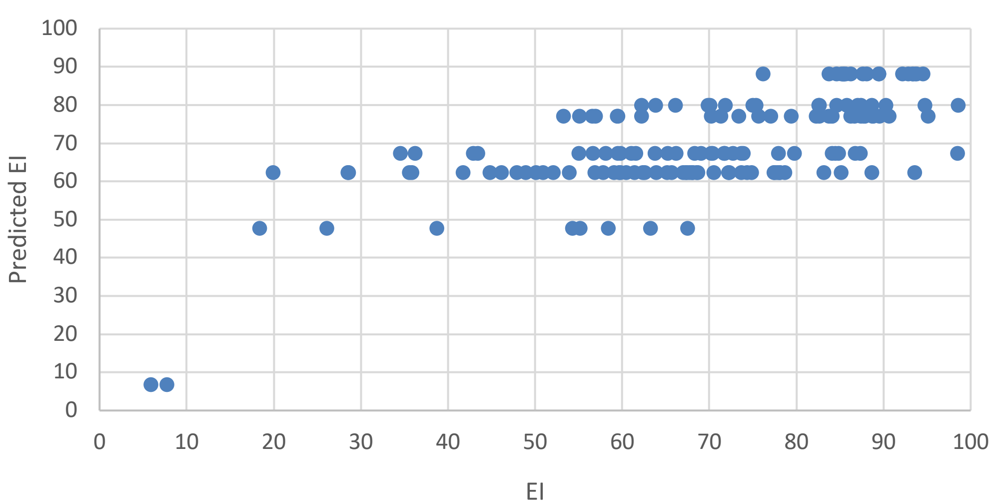


Figure 13

True and Predicted Values for EI with the Winning Decision Tree Model



Random Forest. The application of the random forest model to the merged dataset was tested with the number of trees defined as 100, following the discussion presented in (Oshiro et al., 2012), and the maximum depth parameter ranging from two to 20 in increment steps of two. The comparison is shown in Figure 13 with the maximum depth of 18 showing the lowest RMSE of 15.596 and standard deviation of 1.100. The scatterplot for actual and predicted values for the winning random forest model are presented in Figure 14. The plot showed improved performance both in overall RMSE and in residual distribution for lower EI values.

Figure 14

RMSE for Random Forest Models with Distinct Maximal Depth

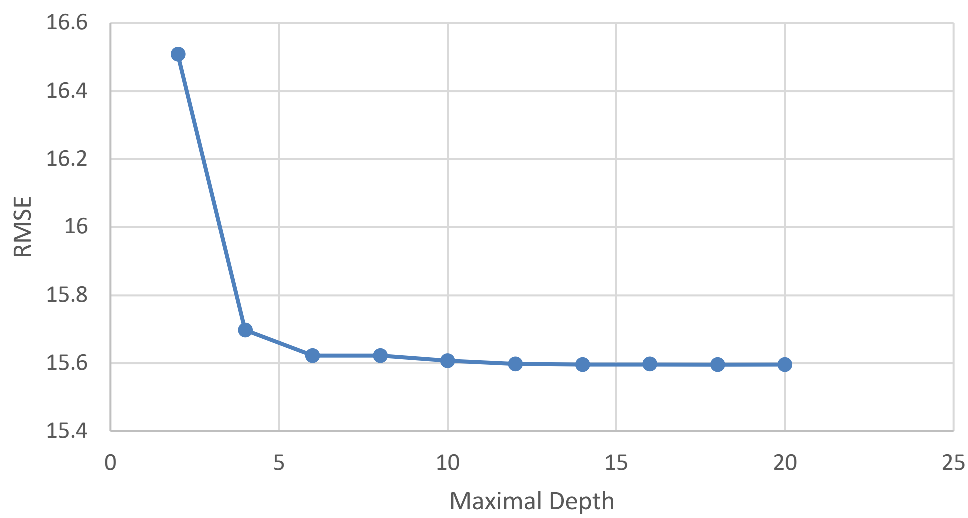
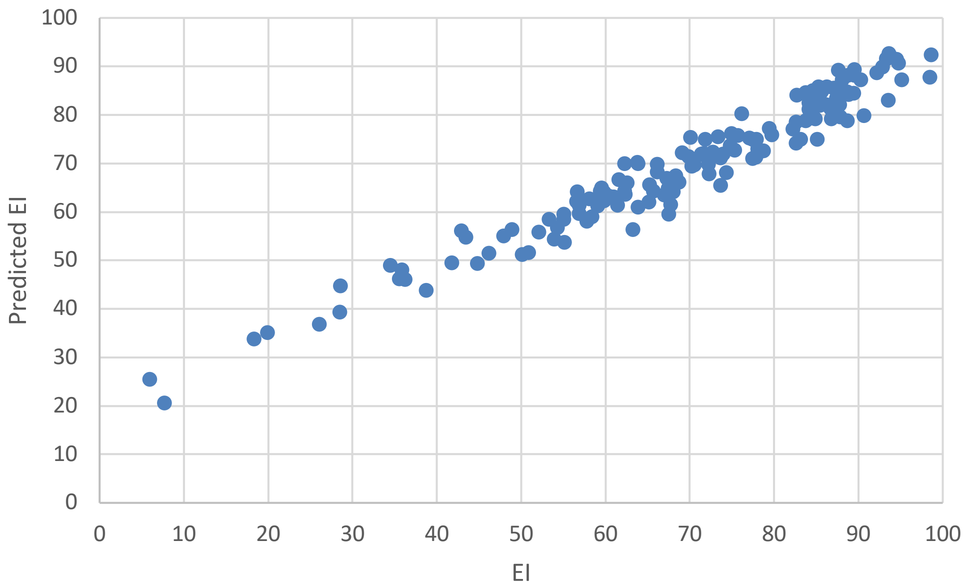


Figure 15

True and Predicted Values for EI with the Winning Random Forest Model



Comparison

Inspection of scatterplots for the true and predicted values for EI and RMSE values indicate that the random forest (RF) model presented superior predictive results (RMSE = 15.596) when compared to the multiple linear regression, deep learning, and decision tree models, with better distribution of residuals throughout the range for EI. Due to the nature of RapidMiner's ensemble models such as random forest, interpretability of results and the understanding of the relative importance of predictors on performance is relatively opaque. Some authors have suggested the use of explanatory models in parallel with such "black box" predictive models, but the approach is still under study (Carvalho et al., 2019) and was not used in the current research. Model comparison results are presented in Table 14 with a comparison to a baseline constant naïve model with predicted value as the mean of EI values in the mlr dataset (*M* = 62.488).

Table 14

Comparison of Performance Across Predictive Models

Metric	Naïve predictor	MLR	DL	DT	RF
RMSE	21.785	19.670	16.649	17.869	15.596
Rel. to Naïve predictor	-	-9.7%	-23.6%	-18.0%	-28.4%

Summary

Chapter IV presented the results of the application of three families of analytical methods to the study of the association among the dimension of governance measured by the WGI and the effectiveness of aviation safety oversight measured by USOAP-CMA EI metrics. Chapter V presented some discussions and conclusions related to these results and their implication for the research questions, as well as some recommendations for practitioners and future research.

Chapter V: Discussion, Conclusions, and Recommendations

This chapter summarizes the findings presented in the previous chapters. It discusses their application and relevance to understanding the relationship between public governance and aviation safety oversight implementation by states. It also addresses the comparative assessment of distinct methodologies used to explore this relationship, discussing the implications of their results and areas for potential future research on governance and aviation safety oversight.

Discussion

This dissertation's earlier chapters demonstrated the importance of civil aviation authorities and regulatory processes in supporting safe and efficient air travel. ICAO implemented a comprehensive framework to assess the effectiveness of member states' safety oversight functions. However, as research is still scant on what drives effective implementation of aviation safety oversight functions among states, the present study set out to identify whether individual elements of public governance would be associated with improved safety oversight. An exploratory approach was also taken to evaluate distinct analytical approaches supporting the identification of such a relationship. A summary of the findings is presented in the following sections in support of the three Research Questions selected for this study. They are replicated below for simplicity.

RQ_1: What are the effects of public governance indicators of Regulatory Quality (RQ) and Government Effectiveness (GE) on the level of aviation safety oversight Effective Implementation (EI) amongst states?

RQ_2: What types of models better predict Effective Implementation (EI) from public governance indicators of Regulatory Quality (RQ) and Government Effectiveness (GE)?

RQ_3: What elements of governance are more closely associated with increased USOAP effective implementation results?

Research Question One – RQ_1

Under RQ_1, the focus was assessing a potential association between states' public governance and aviation safety oversight. This objective was carried out by employing different types of analytical approaches to evaluate the association between Regulatory Quality (RQ) and Government Effectiveness (GE) dimensions of governance and their underlying sources, as proposed by (Kaufmann et al., 2011) and the results of ICAO USOAP-CMA evaluations of states' level of Effective Implementation (EI) for aviation safety oversight functions. All three approaches, namely multiple linear regression, structural equation modeling, and data mining, presented relevant insights into the association. A multiple linear regression model was applied to identify whether a linear association existed among predictors RQ and GE and the target variable EI. The data supported rejecting the null hypothesis and corroborated a linear relationship between GE and EI but not RQ. These results suggested a relevant association between public governance and aviation safety oversight effective implementation. Similar results were obtained by developing a structural equation model and considering RQ and GE as unobserved reflective factors. The WGI dataset supported the construct validity and reliability of a modified model. These results supported the use of the modified model in the analysis of governance and safety oversight. They also confront some of the concerns indicated in the literature regarding the WGIs like those presented in (Langbein & Knack, 2010). The hypothesis indicating that the unobserved factor GE' positively influences EI was also supported in the SEM model. An additional test of the influence of RQ' on safety oversight was conducted by evaluating EI related to primary aviation legislation (CE-1) and specific operating regulations (CE-2), and the only alternative hypothesis supported was the one indicating a positive influence of GE' on $EI_{CE\text{-}2}$.

As a result, the application of these analytical approaches to the identification of a relationship between WGI governance dimensions and ICAO assessments of aviation safety oversight functions suggests that a significant association exists, specifically between the government effectiveness (GE and GE') and USOAP-CMA effective implementation (EI). Linear assessments suggest that the variance in the governance measures can explain 29.5% of the variance in aviation safety oversight effectiveness. The results of applying each analytical approach are presented in detail in Chapter IV.

Research Question Two – RQ_2

RQ_2 addressed the exploratory comparative performance of the analytical approaches selected to address the relationship between governance and safety oversight. Three analytical approaches were employed: multiple linear regression, structural equation modeling, and data mining. Additionally, three predictive machine learning models were selected for data mining: deep learning, decision tree, and random forest.

Results presented in Chapter IV and discussed in the previous section provided general support for the association under study. However, the exploratory comparison of the performance of these methods is still warranted and dealt with under RQ_2. All three analytical models employed have been extensively used in different research fields, and their use typically serves complementary purposes. Multiple linear regression (MLR) models are commonly used in economics and public management. While MLR models' simplicity allows for a cursory initial evaluation of relationships among variables, potentially supporting the development of predictive linear models, they can also present restricted applications to more complex relationships. For the current research, the more complex examination of the relationship between governance and safety oversight was conducted using structural equation modeling (SEM). The method has been extensively used in management, psychology, and social sciences research. In the context of the current study, SEM allowed for considering the WGI governance dimensions as latent factors and the respective measurement errors for the measurement model. SEM also provides for testing the suggested model structure, supporting the proposal, and validating the theoretical framework. Finally, data mining encompasses a broader set of principles and tools for discovering patterns in datasets. Data mining includes the application of non-linear models, which are constructed directly from the data and rely less on the underlying assumptions of the methods.

All three methods brought complementary perspectives to the problem, each contributing to understanding the association under study in various ways. MLR initially indicated a significant association between the government effectiveness dimension of governance (GE) and aviation safety oversight measured by the USOAP-CMA EI metric. This association was further explored by applying the structural equation modeling method, with the WGI sources employed as observed measures and the WGI governance dimensions considered latent factors. This step provided additional support to the discussion of the construct reliability and validity of the WGI by testing the structure associated with the regulatory quality and government effectiveness dimensions of governance proposed by (Kaufmann et al., 2011) and supporting these characteristics for a modified model.

Finally, the three data mining approaches selected also provided complementary views on the relationship between the sources of the WGI governance dimensions and aviation safety oversight effective implementation. The non-linear models chosen for the data mining analyses were deep learning, decision trees, and random forests. Data mining methods, especially the random forest method, showed superior predictive performance compared to MLR and SEM. They also provided new insights into the relative importance of different sources of RQ and GE in predicting EI. It was also interesting to note that the WGI sources contributed differently to other models. While removing the GCS and WMO sources contributed to an improvement in model fit and discriminant validity in the CFA analysis, WMO presented the highest predictive importance in the deep learning model. It was also employed with IPD to discriminate low EI levels in the winning decision tree model. Despite the model's challenges in explainability, the random forest model presented the best predictive performance among all models, with a mean RMSE of 15.596 (*SD* = 1.100).

Research Question Three – RQ_3

Question RQ_3 sought to compare the contributions of the elements of governance and their association with aviation safety oversight effective implementation. The MLR and SEM models consistently supported the alternative hypothesis, indicating (a) a relevant relationship between government effectiveness (GE and GE') and EI and (b) no support for rejecting the null hypothesis on the lack of association between regulatory quality (RQ and RQ') and EI. The expectation that the regulatory quality dimension of governance would better reflect rule-specific elements of safety oversight, more precisely those associated with primary aviation legislation (CE-1) and specific operating regulations (CE-2), could not be confirmed in models where the specific EI values for these CE were used as the right-hand-side variables of the structural model. As such, the research supported the concept that government effectiveness is more directly associated with EI.

Conclusions

Theoretical Contributions

The increased recognition of the evolving complexity of decision-making in the context of public organizations and its relevance for regulatory effectiveness is prominent in research in economics, sociology, and political sciences. While abounding literature has been published on the broader perspectives of public administration and governance, studies on aviation safety oversight effectiveness are scant and generally confined to more legal domains. The aviation industry faces many complex issues, including the global impact of safety concerns in specific states, the increased pace of aerospace technological advances, and the importance of air connectivity in states of all income levels. This context further highlights the importance of understanding factors associated with CAAs' oversight effectiveness.

The application of public governance indices as analytical support for specific policy contexts has been extensively addressed in the literature (Erkkilä, 2015; Homer, 2022; Magnusson & Tarverdi, 2020; Sarwar & Alsaggaf, 2021; Sudders & Nahem, 2007; Zhang & Moffat, 2015). Using such metrics in aviation safety oversight supported identifying an association of broader aspects of public governance measured by the WGI initiative and safety oversight effective implementation measured by ICAO USOAP. While the current structure has been deemed positive to promote aviation safety worldwide, it reflects earlier normative public administration perspectives. Additional inquiry into its theoretical underpinnings could support further improvement. The present study initiates the development of a theoretical framework supporting the study of aviation safety oversight effectiveness by drawing on more modern governance literature. Among the WGI dimensions of governance, government effectiveness (GE) was constructed to indicate the perceived "quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (Kaufmann et al., 2010, p. 4). A statistically significant relationship was confirmed, providing significant support to the GE-EI association in the theoretical framework presented in Figure 1. An additional contribution of the present work relates to the structure of the WGIs and associated discussions regarding reliability and validity. Using the CFA+SEM approach, the WGI structure, including RQ and GE, was validated after modification, achieving acceptable levels of reliability and validity, supporting their use as predictors, and contributing to the extension of the theoretical underpinnings of aviation safety oversight effectiveness. These results further contribute to the knowledge of governance research, as while some concerns on the WGI's reliability had been the topic of discussion, the tests conducted here provide additional support to the general structure of the indicators' methodology.

Practical Contributions

The impact of aviation on the economic development of states with all income levels cannot be overstated, but said development is tied to the industry's continuous advance in safety performance. Safety oversight effectiveness is inherently connected to the system's performance. All organizations involved in improving safety oversight, e.g., ICAO, regional safety oversight organizations, civil aviation authorities, and industry organizations, can benefit from an improved understanding of the factors associated with its effectiveness. The industry's aspiration to continuously contribute to superior safety performance levels can help promote these stakeholders' engagement in future research to provide further insights into factors associated with safety oversight effectiveness. Establishing and confirming a significant relationship between WGI's GE and USOAP EI is an initial step in supporting the identification of more specific aspects of public governance associated with, or potentially leading to, better aviation safety oversight. It also enables additional research efforts to be directed towards the testing of these factor and the identification of additional potential connections. These advances can, in turn, support ICAO and oversight organizations in prioritization and performance evaluation using the WGIs or other governance measures to inform data-centric assessment, monitoring, and audit planning. Finally, as modern research on governance and the understanding of the collaborative and interconnected nature of public decision-making evolved, these new perspectives can be employed in the development of performance-based evaluation strategies to support the added effectiveness of ICAO state safety oversight audits and assessments.

Limitations of the Findings

While the current study provided some light on aspects associated with improved aviation safety oversight effectiveness, it is important to acknowledge and reinforce some of the limitations mentioned in the previous chapters as they can potentially impact the study's results' reliability and validity. The limited availability of the secondary data used for public governance and aviation safety oversight effectiveness restricted the sample size used in support of the quantitative analytical models employed. Measures were taken to ensure datasets used in the analysis showed consistent characteristics in terms of relevant demographics, and that predictive models were tested in out-of-sample data. Still, intepretation of the results must take said contraints into account when aspects of the results' generalizability are concerned. As the World Bank and ICAO continue to gather data for the metrics employed in this study, new reassessments of the proposed framework, with potentially larger samples and improved statistical power, could be achieved. Another notable constraint of the study lies in the exploratory, non-experimental nature of the study. While the results contribute to an initial understanding of the potential governance aspects associated with improved aviation safety oversight effectiveness, further studies would be necessary to test additional aspects of the proposed framework, including the identification of additional confounders and the introduction of additional control measures to support the analysis and potential confirmation of causal relationships among the factors under study. Despite these limitations, the current study contributes to establishing an initial groundwork for the scholarly assessment of factors associated with aviation safety oversight effectiveness and invites researchers in other fields to contribute to developing its associated theoretical framework.

Recommendations

The following sections provide opportunities for practical applications of the study's findings and conclusions and their extension, both for aviation safety oversight practitioners and researchers interested in further expanding the understanding of the factors associated with improved safety oversight effectiveness.

Recommendations for Aviation Safety Oversight Organizations

ICAO, regional safety oversight organizations, industry standard-setting associations, and state regulatory bodies, all considered aviation safety oversight organizations in the context of the present research, play crucial roles in improving aviation safety oversight effectiveness. They are recommended to actively promote and engage in research to identify factors associated with aviation safety oversight effectiveness, which could contribute to improved safety oversight by CAAs across the globe and superior safety performance by industry. A mature high-trust safety oversight landscape can promote, while ensuring safe operations, more efficient global regulatory mechanisms, including the increased adoption of common standards and acceptance criteria.

These organizations are also invited to consider the evolving nature of the public decision-making process in the current oversight assessment mechanisms. The adoption of performance-based mechanisms has the potential to increase the effectiveness of regulatory strategies while reducing compliance costs and promoting tailored interventions for audited organizations, considering differences in complexity and maturity. Finally, safety oversight organizations are invited to include governance measures as an additional source of information to support prioritization and decision-making in safety oversight measurement and assessment mechanisms.

Recommendations for Future Research

All three families of analytical methods employed in the current research contributed to understanding the association of elements of public governance and aviation safety oversight effectiveness, but each could have been further explored. Contemporary developments have proposed some benefits of partial least-squares structural equation modeling (PLS-SEM) compared to traditional covariance-based structural equation modeling, especially in its focus on explaining the variance in the models' dependent variables (Hair et al., 2022). While not necessary for the models considered in the application of SEM in the current research, further exploration of the aviation safety oversight effectiveness measures as unobserved factors could also be explored with PLS-SEM as the data structure would potentially lead to the need for formative measurement models. This approach presents an opportunity for applying PLS-SEM as a potential future research methodology development to support the study of aviation safety oversight effectiveness. The "causal-predictive" approach to SEM supported by PLS-SEM, along with the method's robustness to small sample size and variable distribution assumptions (Hair et al., 2022), can be further used to advance the test of causal relationships among governance and safety oversight measures.

The lack of support for a significant relationship between the regulatory quality dimension of governance (RQ) and the overall EI (as well as the EI associated with critical elements one and two) also creates opportunities for additional research. Further exploration could help identify whether lack of sufficient power, issues in scale development, or unaccounted sources of error could be potentially increasing the chances of a type-II error (failing to reject the null hypothesis when it is false) or provide added support for the lack of an actual association.

Another opportunity to expand the research methodology applied to the analysis of safety oversight concerns could focus on comparisons of more direct measures of governance aspects and their associated assessments of citizens' perceptions. Initiatives like the IIAG (Mo Ibrahim Foundation, 2023) already provide these complementary views of different elements in efforts to assess governance among African states. This approach could potentially contribute to the establishment of more complex theoretical relationships between de jure, de facto, and perceived governance metrics, as discussed in (Erkkilä et al., 2016; Sudders & Nahem, 2007). Such examination can form the groundwork for further studies concerning perception, trust, and confidence in civil aviation authorities and their implications.

Expansion to additional governance dimensions or measurement frameworks beyond the WGIs can also provide new insights with regard to additional elements of governance that could be eventually associated with improved aviation safety oversight effectiveness, potentially expanding the proposed theoretical framework. Similar added contributions could come from the analysis of factors beyond the governance literature. Finally, exploration of the association of governance dimensions and oversight effectiveness in other industries with relevant historical contributions of social regulation, such as nuclear energy, food, drug, and workplace safety, could contribute to the identification of potential similarities and differences in the application of the proposed theoretical frameworks, advancing the conclusions presented in the present research.

References

Ahmad, Md. T. (2010). *Adapting the existing regime for the contemporary world to achieve global civil aviation safety: A developing country perspective* [Master's thesis, McGill University]. https://escholarship.mcgill.ca/concern/theses/g158bh64j

Anda, A., Simon, B., Soós, G., Menyhárt, L., da Silva, J. A. T., & Kucserka, T. (2018). Extending Class A pan evaporation for a shallow lake to simulate the impact of littoral sediment and submerged macrophytes: A case study for Keszthely Bay (Lake Balaton, Hungary). *Agricultural and Forest Meteorology*, *250–251*, 277–289. https://doi.org/10.1016/J.AGRFORMET.2018.01.001

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling rrror on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*(2), 55–173. https://doi.org/10.1007/BF02294170

Andrews, M. (2008). The good governance agenda: Beyond indicators without theory. *Oxford Development Studies*, *36*(4), 379–407. https://doi.org/10.1080/13600810802455120

Andrews, M. (2010). Good government means different things in different countries. *Governance*, *23*(1), 7–35. https://doi.org/10.1111/J.1468-0491.2009.01465.X

Andrews, M., Pritchett, L., & Woolcock, M. (2017). *Building state capability: Evidence, analysis, action*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198747482.001.0001

Ansell, C., & Torfing, J. (Eds.). (2022). *Handbook on theories of governance* (2nd ed.). Edward Elgar. https://doi.org/10.4337/9781800371972

Arndt, C., & Oman, C. (2006). *Uses and abuses of governance indicators*. Development Centres, OECD Publishing. https://doi.org/10.1787/9789264026865-en

Bahr, N. J. (2014). Government regulations and safety oversight. In *System safety engineering and risk assessment: A practical approach* (2nd ed., pp. 301–338). CRC.

Baldwin, R., Cave, M., & Lodge, M. (2012). *Understanding regulation: Theory, strategy, and practice* (2nd ed.). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199576081.001.0001

Bareinboim, E., Tian, J., & Pearl, J. (2022). Recovering from selection bias in causal and statistical inference. In H. Geffner, R. Dechter, & J. Y. Halpern (Eds.), *Probabilistic and Causal Inference* (pp. 433–450). ACM. https://doi.org/10.1145/3501714.3501740

Barreto, O. (2002). Safety oversight: Federal Aviation Administration, International Civil Aviation Organization, and Central American Aviation Safety Agency. *Journal of Air Law and Commerce*, *67*, 651–683. https://scholar.smu.edu/jalc/vol67/iss3/3/

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Berry, W. D. (1993). *Understanding regression assumptions*. SAGE. https://doi.org/10.4135/9781412986427

Bovaird, T. (2005). Public governance: Balancing stakeholder power in a network society. *International Review of Administrative Sciences*, *71*(2), 217–228. https://doi.org/10.1177/0020852305053881

Bovaird, T., & Loeffler, E. (Eds.). (2016). *Public management and governance* (3rd ed.). Routledge. https://doi.org/10.4324/9781315693279

Braverman, M. T., & Arnold, M. E. (2008). An evaluator's balancing act: Making decisions about methodological rigor. *New Directions for Evaluation*, *2008*(120), 71–86. https://doi.org/10.1002/ev.277

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Broderick, A. J., & Loos, J. (2002). Government aviation safety oversight - Trust, but verify. *Journal of Air Law and Commerce*, *67*, 1035–1056. https://scholar.smu.edu/jalc/vol67/iss4/2/

Bryson, J. M., Crosby, B. C., & Bloomberg, L. (2014). Public value governance: Moving beyond traditional public administration and the new public management. *Public Administration Review*, *74*(4), 445–456. https://doi.org/10.1111/PUAR.12238

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge. https://doi.org/10.4324/9781315757421

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

Candel, A., & Ledell, E. (2023). *Deep learning with H2O* (A. Bartz, Ed.; 6th ed.). H2O.ai. https://docs.h2o.ai/h2o/latest-stable/h2o-docs/booklets/DeepLearningBooklet.pdf

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(832), 34. https://doi.org/10.3390/electronics8080832

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc. https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf

Chen, Y., Liu, X., Gao, L., Zhu, M., Shia, B. C., Chen, M., Ye, L., & Qin, L. (2023). Using the H2O automatic machine learning algorithms to identify predictors of web-based medical record nonuse among patients in a data-rich environment: Mixed methods study. *JMIR Medical Informatics*, *11*(1). https://doi.org/10.2196/41576

Christensen, T., & Lægreid, P. (2008). NPM and beyond - Structure, culture and demography. *International Review of Administrative Sciences*, *74*(1), 7–23. https://doi.org/10.1177/0020852307085730

Coase, R. (2004). The new institutional economics. In E. Brosseau & J.-M. Glachant (Eds.), *The economics of contracts: Theory and applications* (pp. 45–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511613807.002

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. Chapman and Hall. https://hdl.handle.net/11299/37076

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Daugėla, K., & Vaičiukynas, E. (2022). Real-Time anomaly detection for distributed systems logs using Apache Kafka and H2O.ai. *Communications in Computer and Information Science*, *1665 CCIS*, 33–42. https://doi.org/10.1007/978-3-031-16302-9_3

de Vries, M. (2013). The challenge of good governance. *The Innovation Journal: The Public Sector Innovation Journal*, *18*(1), 1–9.

den Hertog, J. (2010). Review of economic theories of regulation. *Tjalling C. Koopmans Institute Discussion Paper Series*, *10*(18), 1–59. https://EconPapers.repec.org/RePEc:use:tkiwps:1018

Detra, Z. D. (2006). *The legitimacy of the International Civil Aviation Organization's Universal Safety Oversight Audit Programme* [Master's thesis, McGill University]. https://escholarship.mcgill.ca/concern/theses/zg64tm19p

Dunleavy, P., & Hood, C. (1994). From old public administration to new public management. *Public Money and Management*, *14*(3), 9–16. https://doi.org/10.1080/09540969409387823

Dunsire, A. (1973). *Administration: The word and the science*. Martin Robertson.

Durand, A., & Pietikainen, A. (2022). *The governance of civil aviation authorities in Latin American countries: Evidence from ICAO's North American, Central American and Caribbean and South American regions*. https://doi.org/10.1787/e8bdf362-en

Edmonds, W. A., & Kennedy, T. D. (2017). *An applied guide to research design: Quantitative, qualitative, and mixed methods* (2nd ed.). SAGE. https://doi.org/10.4135/9781071802779

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An introductory review of deep learning for prediction models with big data. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.00004

Erkkilä, T. (2015). Global governance indices as policy instruments: Actionability, transparency and comparative policy analysis. *Journal of Comparative Policy Analysis: Research and Practice*, *18*(4), 382–402. https://doi.org/10.1080/13876988.2015.1023052

Erkkilä, T., Peters, B. G., & Piironen, O. (2016). Politics of comparative quantification: The case of governance metrics. *Journal of Comparative Policy Analysis: Research and Practice*, *18*(4), 319–328. https://doi.org/10.1080/13876988.2016.1145871

Farrington, C. (2009). Putting good governance into practice I: The Ibrahim Index of African governance. *Progress in Development Studies*, *9*(3), 249–255. https://doi.org/10.1177/146499340800900305

Farrington, C. (2010). Putting good governance into practice II: Critiquing and extending the Ibrahim index of African governance. *Progress in Development Studies*, *10*(1), 81–86. https://doi.org/10.1177/146499340901000106

Farrington, C. (2011). Putting good governance into practice III: Measuring intrinsic and instrumental empowerment in local government contexts. *Progress in Development Studies*, *11*(2), 151–161. https://doi.org/10.1177/146499341001100205

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). SAGE.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.2307/3151312

Forsyth, D., Spruston, D., & Whitefield, R. (2014). *Aviation safety regulation review* (Issue May). Australian Government. https://www.infrastructure.gov.au/sites/default/files/documents/ASRR_Report_May_2014.pdf

Fox, J. (2016). *Applied regression analysis and generalized linear models* (3rd ed.). SAGE.

Frederickson, H. G., Smith, K. B., Larimer, C. W., & Licari, M. J. (2012). *The public administration theory primer* (2nd ed.). Westview.

Freedman, J. (1991). Bootstrapping regression models. *Annals of Statistics*, *9*(6), 1218–1228. https://doi.org/10.1214/aos/1176345638

Fukuyama, F. (2013). What is governance? *Governance: An International Journal of Policy, Administration, and Institutions*, *26*(3), 347–368. https://doi.org/10.1111/GOVE.12035

Garcia, J. S. D., Jaedicke, C., Leng Lim, G., & Truong, D. (2023). Predicting the severity of runway excursions from aviation safety reports. *Journal of Aerospace Information Systems*, *20*(9), 555–564. https://doi.org/10.2514/1.I011145

Gedeon, T. D. (1997). Data mining of inputs: Analysing magnitude and functional measures. *International Journal of Neural Systems*, *08*(02), 209–218. https://doi.org/10.1142/S0129065797000227

Ghahremani, M. (2020). Analysis of the ICAO USOAP results in the Persian Gulf region and its relation to the economic indicators. *International Journal of Aviation, Aeronautics, and Aerospace*, *7*(3). https://doi.org/10.15394/ijaaa.2020.1505

Goldberger, A. S. (1972). Maximum-Likelihood estimation of regressions containing unobservable independent variables. *International Economic Review*, *13*(1), 1. https://doi.org/10.2307/2525901

Gowrisankaran, G. (2002). Competition and regulation in the airline industry: Puppets in chaos. *FRBSF Economic Letter*, *2002–01*, 1–4.

Grindle, M. S. (2004). Good enough governance: Poverty reduction and reform in developing countries. *Governance*, *17*(4), 525–548. https://doi.org/10.1111/J.0952-1895.2004.00256.X

Grindle, M. S. (2007). Good enough governance revisited. *Development Policy Review*, *25*(5), 533–574. https://doi.org/10.1111/J.1467-7679.2007.00385.X

Groenleer, M., Kaeding, M., & Versluis, E. (2010). Regulatory governance through agencies of the European Union? The role of the European agencies for maritime and aviation safety in the implementation of European transport legislation. *Journal of European Public Policy*, *17*(8), 1212–1230. https://doi.org/10.1080/13501763.2010.513577

Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson.

Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE. https://doi.org/10.1007/978-3-030-80519-7

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Hansen, M., McAndrews, C., & Berkeley, E. (2008, July). *History of aviation safety oversight in the United States* [DOT/FAA/AR-08/39]. U.S. Department of Transportation. https://www.tc.faa.gov/its/worldpac/techrpt/ar0839.pdf

Hansen, M., McAndrews, C., Berkeley, E., Gribko, J., Berkey, D., & Hasan, S. (2006, July). *Understanding and evaluating the Federal Aviation Administration Safety Oversight System*. National Center of Excellence for Aviation Operations Research.

Hartmann, H., Donner, S., & Schwarz, R. (2020). *Transformation Index BTI 2020: Governance in International Comparison*. Bertelsmann-Stiftung.

https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/imported/leseprobe/1900_Leseprobe.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: Concerns and considerations. *Computational Methods in Social Sciences*, *3*(1), 5–17. https://www.ceeol.com/search/article-detail?id=417198

Hollinger, K. (2013). *Safety management systems for aviation practitioners: Real-world lessons*. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/4.102110

Homer, J. (2022). Can good government save us? Extending a climate-population model to include governance and its effects. *Systems, 10*(2). https://doi.org/10.3390/systems10020037

Hood, C. (1991). A public management for all seasons? *Public Administration*, *69*(1), 3–19. https://doi.org/10.1111/j.1467-9299.1991.tb00779.x

Hood, C., & Jackson, M. (2013). The new public management: A recipe for disaster? In D. Parker & J. Handmer (Eds.), *Hazard management and emergency planning: Perspectives in Britain* (pp. 109–123). Routledge.

Hood, C., Rothstein, H., & Baldwin, R. (2001). *The government of risk: Understanding risk regulation regimes*. Oxford University Press. https://doi.org/10.1093/0199243638.001.0001

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huber, P. J. (1981). *Robust statistics*. John Wiley & Sons. https://doi.org/10.1002/0471725250

Hughes, O. (1992). Public management or public administration? *Australian Journal of Public Administration*, *51*(3), 286–296. https://doi.org/10.1111/j.1467-8500.1992.tb02614.x

International Civil Aviation Organization. (1944, December 7). *Convention on international civil aviation* [Doc 7300]. https://www.icao.int/publications/Pages/doc7300.aspx

International Civil Aviation Organization. (2014). *Universal safety oversight audit programme: Continuous monitoring manual* (4th ed.) [Doc 9735, AN/960]. https://standart.aero/en/icao/book/doc-9735-universal-safety-oversight-audit-programme-continuous-monitoring-manual-en-cons

International Civil Aviation Organization. (2016). *Annex 19: Safety management* (2nd ed.) https://www.icao.int/safety/SafetyManagement/WebsiteDesignJuly2016/Flyer_US-Letter_ANB-ANNEX19-SM_2016-10-03.AP.pdf

International Civil Aviation Organization. (2017). *Safety Oversight Manual: Part A - The establishment and management of a state safety oversight system* [Doc 9734] (3rd ed.) https://www.icao.int/APAC/Meetings/2019%20COSCAPSEA%20iSTARS/9734_parta_cons_en.pdf

International Civil Aviation Organization. (2019). *Universal safety oversight audit programme: Continuous monitoring approach results, 1 January 2016 to 31 December 2018*. https://www.icao.int/safety/CMAForum/Documents/USOAP_REPORT_2016-2018.pdf

International Civil Aviation Organization. (2021a). *ICAO API data service: USOAP EI by state*. https://www.icao.int/Aviation-API-Data-Service/Pages/default.aspx

International Civil Aviation Organization. (2021b). *ICAO API data service: USOAP Progress by State*. https://www.icao.int/Aviation-API-Data-Service/Pages/default.aspx

International Civil Aviation Organization. (2021c). *Terms and conditions for accessing and using the ICAO Application Programming Interfaces (API) data service*.

International Civil Aviation Organization. (2022a). *Business plan 2023-2025*. https://www.icao.int/Meetings/a41/Documents/Business_Plan_23_25.pdf

International Civil Aviation Organization. (2022b, June 1). *Member states*. https://www.icao.int/about-icao/Pages/member-states.aspx

International Finance Corporation, Multilateral Investment Guarantee Agency, & World Bank. (2010). *Regulatory quality indicators*. World Bank Group. https://doi.org/10.1596/27877

Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling*, *8*(2), 205–223. https://doi.org/10.1207/S15328007SEM0802_3

Jennison, M. (2013). The future of aviation safety regulation: New US-EU agreement harmonizes and consolidates the Transatlantic regime, but what is the potential for genuine regulatory reform? *Air and Space Law*, *38*(4 & 5), 333–350. https://doi.org/10.54648/aila2013022

Joosen, R., Haverland, M., & de Bruijn, E. (2022). Shaping EU agencies' rulemaking: Interest groups, national regulatory agencies and the European Union Aviation Safety Agency. *Comparative European Politics*, *20*, 411–442. https://doi.org/10.1057/s41295-021-00268-z

Jung, H., Merens, M., Valipour, M., Liang, X., Abboud, D., Wen, H. A., Yuan, Y., & Zimmerman, R. (2018). Data-driven decision-making processes, data services and applications for global aviation safety. *ITU Journal: ICT Discoveries, Special Issue*, *2*. https://www.itu.int/dms_pub/itu-s/opb/journal/S-JOURNAL-ICTF.VOL1-2018-2-P07-PDF-E.pdf

Kapucu, N., Yuldashev, F., & Bakiev, E. (2009). Collaborative public management and collaborative governance: Conceptual similarities and differences. *European Journal of Economic and Political Studies*, *2*(1), 39–60.

https://www.researchgate.net/profile/Erlan-Bakiev-2/publication/342851405_European_Journal_of_Economic_and_Political_Studies_Collaborative_Public_Management_and_Collaborative_Governance_Conceptual_Similarities_and_Differences/links/5f08cc55299bf1881610eddc/European-Journal-of-Economic-and-Political-Studies-Collaborative-Public-Management-and-Collaborative-Governance-Conceptual-Similarities-and-Differences.pdf

Kaufmann, D., & Kraay, A. (2008). Governance indicators: Where are we, where should we be going? *The World Bank Research Observer*, *23*(1), 1–30. https://doi.org/10.1093/WBRO/LKM012

Kaufmann, D., & Kraay, A. (2023). *Worldwide governance indicators*. World Bank Group. https://info.worldbank.org/governance/wgi/

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2007). *The worldwide governance indicators project: Answering the critics* (Issue WPS4149). World Bank Group. https://doi.org/10.1596/1813-9450-4149

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2009, June). *Governance matters VIII: Aggregate and individual governance indicators 1996-2008*. World Bank Group. https://doi.org/10.1596/1813-9450-4978

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010a). Response to ‘What do the Worldwide Governance Indicators measure?’ *The European Journal of Development Research*, *22*(1), 55–58. https://doi.org/10.1057/EJDR.2009.49

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010b). The Worldwide Governance Indicators: Methodology and analytical issues. In *World Bank Policy Research Working Paper No. 5430*. https://doi.org/10.1017/S1876404511200046

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and analytical issues. *Hague Journal on the Rule of Law*, *3*, 220–246. https://doi.org/10.1017/S1876404511200046

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods and Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Klenka, M. (2017). Aviation safety: Legal obligations of states and practice. *Journal of Transportation Security*, *10*(3–4), 127–143. https://doi.org/10.1007/s12198-017-0183-8

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (1999). The quality of government. *Journal of Law, Economics, and Organization*, *15*(1), 222–279. https://doi.org/10.1093/jleo/15.1.222

Ładyżyński, P., Żbikowski, K., & Gawrysiak, P. (2019). Direct marketing campaigns in retail banking with the use of deep learning and random forests. *Expert Systems with Applications*, *134*, 28–35. https://doi.org/10.1016/J.ESWA.2019.05.020

Laerd Statistics. (2015). *Multiple regression using SPSS Statistics*. Statistical Tutorials and Software Guides. https://statistics.laerd.com/

Langbein, L., & Knack, S. (2010). The Worldwide Governance Indicators: Six, one, or none? *The Journal of Development Studies*, *46*(2), 350–370. https://doi.org/10.1080/00220380902952399

Lawrence, H. (2008). *Aviation and the role of government* (2nd ed.). Kendall/Hunt.

Maccallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. In *Psychological Methods* (Vol. 2). https://doi.org/10.1037/1082-989X.1.2.130

Magnusson, L. M., & Tarverdi, Y. (2020). Measuring governance: Why do errors matter? *World Development*, *136*. https://doi.org/10.1016/J.WORLDDEV.2020.105061

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., & Flach, P. (2021). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, *33*(8), 3048–30361. https://doi.org/10.1109/TKDE.2019.2962680

Maurseth, P. B. (2008). *Governance indicators: A guided tour*.

Mo Ibrahim Foundation. (2020). *2020 Ibrahim Index of African governance - Index report*. https://mo.ibrahim.foundation/sites/default/files/2020-11/2020-index-report.pdf

Mo Ibrahim Foundation. (2023). *2022 Ibrahim Index of African governance - Index report*. https://mo.ibrahim.foundation/sites/default/files/2023-01/2022-index-report.pdf

Moak, L., McDew, D. W., Haskins, G., Hylander, K., & Grizzle, D. (2020). *Official report of the special committee to review the Federal Aviation Administration's aircraft certification process*. U.S. Department of Transportation. https://www.transportation.gov/sites/dot.gov/files/2020-01/scc-final-report.pdf

Odisho, E. V. (2020). *Predicting pilot misperception of runway excursion risk through machine learning algorithms of recorded flight data* [Doctoral dissertation, Embry-Riddle Aeronautical University]. https://commons.erau.edu/edt/503/

O'Neil, P. D. (2008). *Building a reliable organization: The evolution of error-intolerance in the FAA* [Doctoral dissertation, University of Nebraska at Omaha].

O'Neil, P. D., & Krane, D. (2012). Policy and organizational change in the Federal Aviation Administration: The ontogenesis of a high-reliability organization. *Public Administration Review*, *72*(1), 98–111. https://doi.org/10.1111/j.1540-6210.2011.02479.x

Ongaro, E., & van Thiel, S. (Eds.). (2018). *The Palgrave handbook of public administration and management in Europe*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-55269-3

Osborne, S. P. (2006). The new public governance? *Public Management Review*, *8*(3), 377–387. https://doi.org/10.1080/14719030600853022

Osborne, S. P. (Ed.). (2010). *The new public governance? Emerging perspectives on the theory and practice of public governance*. Routledge. https://doi.org/10.4324/9780203861684

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? In P. Perner (Ed.), *Machine learning and data mining in pattern recognition. MLDM 2012. Lecture Notes in Computer Science: Vol. 7376 LNAI* (pp. 154–168). Springer. https://doi.org/10.1007/978-3-642-31537-4_13

Ostrom, V., & Ostrom, E. (1971). Public choice: A different approach to the study of public administration. *Public Administration Review*, *31*(2), 203. https://doi.org/10.2307/974676

Palacios, H. J. G., Toledo, R. A. J., Pantoja, G. A. H., & Navarro, Á. A. M. (2017). A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change. *Advances in Science, Technology and Engineering Systems*, *2*(3), 598–604. https://doi.org/10.25046/AJ020376

Pan, J. Y., & Truong, D. (2018). Passengers' intentions to use low-cost carriers: An extended theory of planned behavior model. *Journal of Air Transport Management*, *69*, 38–48. https://doi.org/10.1016/j.jairtraman.2018.01.006

Perrow, C. (1984). *Normal accidents: Living with high-risk technologies*. Basic Books.

Petras, C. M., & Vaugeois, M. (2017). Safety regulation and oversight. In P. S. Dempsey & R. S. Jakhu (Eds.), *Routledge Handbook of Public Aviation Law* (pp. 64–87). Taylor & Francis Group.

Picard, R. R., & Cook, R. D. (1984). Cross-Validation of regression models. *Journal of the American Statistical Association*, *79*(387), 575. https://doi.org/10.2307/2288403

Pierre, J., & Peters, B. G. (2009). From a club to a bureaucracy: JAA, EASA, and European aviation regulation. *Journal of European Public Policy*, *16*(3), 337–355. https://doi.org/10.1080/13501760802662706

Pollitt, C. (2007). The new public management: An overview of its current status. *Administrație Și Management Public*, *8*, 110.

Pollitt, C., & Bouckaert, G. (2011). *Public management reform: A comparative analysis—New public management, governance, and the Neo-Weberian state* (3rd ed.). Oxford University Press.

Pollitt, C., Van Thiel, S., & Homburg, V. (2007). *New public management in Europe: Adaptation and alternatives*. Palgrave Macmillan. https://doi.org/10.1057/9780230625365

Ratajczyk, M. A. (2014). *Regional aviation safety organisations: Enhancing air transport safety through regional cooperation* [Doctoral dissertation, Leiden University]. https://scholarlypublications.universiteitleiden.nl/handle/1887/29759

Roberts, S., & Safety Management International Collaboration Group. (2013). New role for the regulator. *AeroSafety World August 2013.*

Rotberg, R. I. (2004). Strengthening governance: Ranking countries would help. *Washington Quarterly*, *28*(1), 71–81. https://doi.org/10.1162/0163660042518215

Rothstein, B., & Teorell, J. (2005). What is quality of government: A theory of impartial government institutions. *Governance: An International Journal of Policy, Administration, and Institutions, 21*(2), 165-190. https://doi.org/10.1111/j.1468-0491.2008.00391.x

Sarwar, S., & Alsaggaf, M. I. (2021). The role of governance indicators to minimize the carbon emission: A study of Saudi Arabia. *Management of Environmental Quality*, *32*(5), 970–988. https://doi.org/10.1108/MEQ-11-2020-0275

Schout, A. (2008). Inspecting aviation safety in the EU: EASA as an administrative innovation? In E. Vos (Ed.), *European risk governance: Its science, its inclusiveness and its effectiveness* (pp. 257–294). Connex.

Schröer, C., Kruse, F., & Marx Gómez, J. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, *181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Senthilkumar, G., Madhusudhana, S., Flitcroft, M., Sheriff, S., Thalji, S., Merrill, J., Clarke, C. N., Maduekwe, U. N., Tsai, S., Christians, K. K., Gamblin, T. C., & Kothari, A. N. (2023). Automated machine learning (AutoML) can predict 90-day mortality after gastrectomy for cancer. *Scientific Reports*, *13*(11051), 1–13. https://doi.org/10.1038/s41598-023-37396-3

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 13–22.

Simon, H. A. (2013). *Administrative behavior: a study of decision-making process in administrative organization* (4th ed.). Free Press.

Sirikulviriya, N., & Sinthupinyo, S. (2011). Integration of rules from a random forest. *2011 International Conference on Information and Electronics Engineering IPCSIT, 6*. IACSIT Press.

Sønderby, K. (2016). *How can aviation regulators retain effectiveness in a performance-based environment?* [Master's thesis, Lund University]. https://www.lunduniversity.lu.se/lup/publication/8309632

Soper, D. S. (2024). *A-priori Sample Size Calculator for Structural Equation Models* [Software]. https://www.danielsoper.com/statcalc/calculator.aspx?id=89

Sparrow, M. K. (2000). *The regulatory craft: Controlling risks, solving problems, and managing compliance*. Brookings.

Stimpson, E. W., Babbitt, J. R., McCabe, W. O., Sparrow, M. K., & Vogt, C. W. (2008). *Managing risks in civil aviation: A review of FAA's approach to safety*. U.S. Department of Transportation. https://rosap.ntl.bts.gov/view/dot/61302

Stoker, G. (2018). Governance as theory: Five propositions. *International Social Science Journal*, *68*(227–228), 15–24. https://doi.org/10.1111/ISSJ.12189

Stolzer, A. J., & Goglia, J. J. (2015). *Safety management systems in aviation* (2nd ed.). Ashgate. https://doi.org/10.4324/9781315607504

Sudders, M., & Nahem, J. (2007). *Governance indicators: A user's guide*. United Nations Development Programme. https://anfrel.org/wp-content/uploads/2012/02/2007_UNDP_goveranceindicators.pdf

Sunstein, C. R. (2002). *Risk and reason: Safety, law, and the environment*. Cambridge University Press.

Swedavia AB, & McGregor & Company. (1988). *Review of civil aviation safety regulations and the resources, structure and functions of the New Zealand Ministry of Transport Civil Aviation Division*. New Zealand Department of Transport, The Treasury, State Services Commission, and the Department of Trade and Industry. https://www.aviation.govt.nz/assets/about-us/strategy/swedavia-mcgregor-report.pdf

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, *58*(1), 134–146. https://doi.org/10.2307/1130296

Taylor, F. W. (1919). *The principles of scientific management*. Harper & Brothers.

Thomas, M. A. (2010). What do the Worldwide Governance Indicators measure? *The European Journal of Development Research*, *22*(1), 31–54. https://doi.org/10.1057/ejdr.2009.32

Truong, D., Friend, M. A., & Chen, H. (2018). Applications of business analytics in predicting flight on-time performance in a complex and dynamic system. *Transportation Journal*, *57*(1), 24–52. https://doi.org/10.5325/transportationj.57.1.0024

Truong, D., Pan, J. Y., & Buaphiban, T. (2020). Low cost carriers in Southeast Asia: How does ticket price change the way passengers make their airline selection? *Journal of Air Transport Management*, *86*, 101836. https://doi.org/10.1016/j.jairtraman.2020.101836

van de Walle, S. (2007). The state of the world's bureaucracies. *Journal of Comparative Policy Analysis: Research and Practice*, *8*(4), 437–448. https://doi.org/10.1080/13876980600971409

van Rossum, G. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Vignieri, V. (2020). *Changes and current challenges in public administration* [Slides]. Retrieved from https://www.vincenzovignieri.com/wp-content/uploads/2020/05/Changes-and-challenges-in-Public-Administration.pdf

Wan, L., Zeiler, M., Zhang, S., Lecun, Y., & Fergus, R. (2013). *Regularization of neural networks using DropConnect*. *Proceedings of the 30th International Conference on Machine Learning*, *28*(3),1058-1066. https://proceedings.mlr.press/v28/wan13.html

Westland, J.C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications, 9*(6), 476-487. https://doi.org/10.1016/j.elerap.2010.07.003

Willing, T. K., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J. A., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., & Abdalla, S. (2016). Jupyter Notebooks -- A publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87–90). IOS Press. https://doi.org/10.3233/978-1-61499-649-1-87

Wilson, W. (1887). The study of administration. *Political Science Quarterly*, *2*(2), 197–222. https://doi.org/10.2307/2139277

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, (pp. 29–39). http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf

World Bank. (2023). *World Development Indicators*. https://datatopics.worldbank.org/world-development-indicators/

Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, *2*(3), 249–262. https://doi.org/10.1007/S41664-018-0068-2

Yadav, D. K., & Nikraz, H. (2014). Implications of evolving civil aviation safety regulations on the safety outcomes of air transport industry and airports. *Aviation*, *18*(2), 94–103. https://doi.org/10.3846/16487788.2014.926641

Zhang, A., & Moffat, K. (2015). A balancing act: The role of benefits, impacts and confidence in governance in predicting acceptance of mining in Australia. *Resources Policy*, *44*, 25–34. https://doi.org/10.1016/J.RESOURPOL.2015.01.001

Appendix A

Distribution of Demographics Among Datasets

Figure 16

Kernel Density Estimates Plots for the Six Variables (GDP, GDP per capita, land area, population, air passengers carried, air freight carried) in the Datasets Used in the Analyses