

Article

The Validity Gap: A Measurement Model of Country-Level Institutions Using Structural Equation Modelling

Mariam Alsabah ^{1,*}  and Ahmad Alshehabi ^{2,3} ¹ College of Business and Entrepreneurship, Abdullah Al Salem University, Kuwait City 13001, Kuwait² Centre for Research in Accounting, Accountability and Governance (CRAAG), Department of Accounting, Southampton Business School, University of Southampton, Southampton SO16 7QF, UK; a.alshehabi@soton.ac.uk³ Department of Accounting, Faculty of Economics, Damascus University, Damascus 11191, Syria

* Correspondence: mariam.alsabah@aasu.edu.kw

Abstract

We examine the measurement of country-level institutions in international accounting research. Despite a large body of literature linking national institutions to financial reporting outcomes, the construct validity and reliability of the institutional measures used in cross-country studies have rarely been assessed. We show that widely cited measures of investor protection, legal quality, and equity market development have been adopted without validity or reliability assessment and used inconsistently across published studies, with the same indicator labelled as a measure of different constructs in different papers. Using forty-eight candidate indicators from six international databases for seventy countries, we conduct an exploratory factor analysis followed by a confirmatory factor analysis and identify three latent constructs that pass conventional thresholds for convergent reliability and discriminant validity. The constructs are empirically distinct, indicating that measures routinely treated as interchangeable in published work measure different latent dimensions. We offer a validated three-factor measurement model that researchers can use to operationalise country-level institutions in future cross-country studies and discuss the implications for the interpretation of existing studies.

Keywords: country-level institutions; construct validity; confirmatory factor analysis; structural equation modelling; AMOS

1. Introduction

International accounting research has grown rapidly over the past two decades. A central tenet is the proposition that the country in which a firm reports affects the quality of the information conveyed by its financial statements (Leuz et al., 2003; Leuz, 2010; Wysocki, 2011; Isidro et al., 2020). This proposition is supported by substantial empirical literature that has used country-level measures of investor protection, legal quality, and market development to explain cross-country variations in earnings management, disclosure quality, value relevance, and a range of other reporting outcomes (e.g., Halabi et al., 2019; Alshehabi et al., 2021, 2024; Fargher & Zhang, 2024). The continued growth of this body of literature has placed increasing weight on the country-level measures it relies upon.

However, the construct validity and reliability of those measures have rarely been assessed. Cross-country studies routinely operationalise multidimensional institutional constructs through a single indicator drawn from one of the international databases that have proliferated since the late 1990s: the Worldwide Governance Indicators (Kaufmann



Academic Editor: Thanasis Stengos

Received: 11 May 2026

Revised: 29 June 2026

Accepted: 30 June 2026

Published: 3 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

et al., 2010; Kaufmann & Kraay, 2024), the Doing Business reports, the Global Competitiveness Report, the Economic Freedom of the World index, the Global Financial Development database, and the indices constructed by La Porta et al. (1997, 1998, 2006) and Djankov et al. (2008). The single indicator is then assigned a construct label—investor protection, rule of law, regulatory quality, or stock market development—and entered into a regression as if the construct content of the indicator and the construct content of the label have been established. However, this is generally not the case. The research question that motivates this study is therefore methodological: do the country-level institutional indicators routinely used in international accounting research cluster into empirically distinct latent constructs that satisfy established criteria for construct validity and reliability? In answering it, we address three gaps in the existing literature: the absence of construct validation for the country-level institutional measures in widespread use, the inconsistent classification of the same indicators across published studies, and the lack of a validated measurement model for country-level institutions.

In this paper, we document the consequences of this practice and develop a validated alternative. Our contribution is threefold: First, we provide systematic evidence that the same widely cited indicators have been labelled as measures of different constructs across published studies. Drawing on fourteen cross-country accounting research papers published between 2003 and 2012, we show that the anti-director rights index has been used as a measure of investor protection in seven studies and of legal enforcement in two; that the disclosure requirements index has been treated as a measure of investor protection in three studies and of securities regulation in four; and that the public-enforcement index has been variously labelled as a measure of investor protection, securities regulation, and legal enforcement. We argue that this pattern is too systematic to be the product of substantive disagreement among researchers and that it reflects the absence of an underlying measurement model.

Second, we develop and validate a three-factor measurement model for country-level institutions that addresses the concerns regarding construct validity. Using forty-eight candidate indicators from six international databases for a sample of seventy countries (the same set used by Djankov et al. (2008) less Taiwan and Tunisia) and performing exploratory factor analysis, we identify three latent constructs—investor protection, quality of legality, and equity market development—that jointly explain 81.84% of the variance in the original indicators. Confirmatory factor analysis on the same sample produces a three-factor measurement model with a satisfactory overall fit (CMIN/DF = 1.429; CFI = 0.970; RMSEA = 0.079). All three constructs satisfy conventional thresholds for composite reliability and average variance extracted; the squared inter-construct correlations are below the lowest AVE in the model, satisfying the Fornell–Larcker criterion for discriminant validity (Fornell & Larcker, 1981).

Third, we document the implications of our measurement model for the interpretation of single-item proxies that have come to dominate the literature. The four Worldwide Governance Indicators components that load on the quality of legality factor are statistically near-interchangeable, with loadings between 0.957 and 0.963 and communalities between 0.924 and 0.936. The four indicators that load on the investor protection factor, by contrast, are not interchangeable, with loadings ranging from 0.742 to 0.903; the original anti-director rights index carries the lowest loading and communality among the retained indicators. We argue that, for some constructs, the single-item proxy that a researcher selects makes little difference to what is measured, whereas for others, the choice is decisive, determining which dimension of the construct the study actually captures.

Our paper builds on and complements Isidro et al. (2020), who surveyed more than a hundred international accounting studies and reported that the levels of seventy-two

country attributes used in the literature were largely captured by four latent factors, and that once those factors were included, none of the seventy-two individual variables retained explanatory power for reporting quality. These authors proposed confirmatory factor analysis and structural equation modelling as the path forward but did not develop a CFA-based measurement model for the constructs used by the field. We take that step. The two papers are complementary: [Isidro et al. \(2020\)](#) document the scale of the co-dependencies among country attributes; we develop a validated measurement model for the institutional constructs that the current literature claims to study.

The rest of our paper is organised as follows: Section 2 discusses the institutional system and the conceptual definition of institutions that we adopt. Section 3 documents the two principal methodological difficulties facing the existing literature—the absence of standardised conceptual definitions and the absence of validity and reliability assessment—and presents our cross-study evidence on inconsistent indicator labelling. Section 4 discusses two further issues: item selection from a larger pool of candidate indicators and interpretational confounding. Section 5 describes our data and methodology. Section 6 reports our exploratory factor analysis and Section 7 our confirmatory factor analysis. Section 8 concludes.

2. The Institutional System

2.1. What Is the Institutional System?

National institutions can be described as a workable combination of arrangements whose functionalities are integrated into a coherent whole. The identification of all relevant institutions is difficult as a result of institutional interdependence, both conceptual and statistical: each institution affects and is affected by other institutions, and a change in one tends to lead to changes in others ([Brinkerhoff et al., 2007](#)). As [Dignam and Galanis \(2009, p. 51\)](#) put it, institutions do not exist and operate in isolation from each other. They coexist as components or subsystems of a coherent system; they may complement or reinforce one another, but they do not substitute for one another in any straightforward sense. The empirical task that follows is to identify the salient institutional dimensions, each standing in its own right and capturing a unique facet of the country's institutional system.

In practice, a great deal of accounting scholarship proceeds as if a country's institutions were independent of one another and, accordingly, treats them as exogenous in cross-country models, even when the same scholars are convinced on theoretical grounds that institutions arise endogenously and evolve jointly over time ([Choi & Meek, 2011](#); [Glaum et al., 2013](#); [Leuz, 2010](#); [Leuz et al., 2003](#); [Wysocki, 2011](#)). [Wysocki \(2004\)](#), for example, suggests that tax compliance is an endogenous outcome of other institutions: better institutions provide incentives for higher-quality reporting and thereby for greater tax compliance. One reason such assumptions persist is that identifying the structure of the institutional system is not easy, particularly at a theoretical level, given the ambiguities inherent in the literature concerning the conceptual meanings of its components. Ambiguity at the conceptual level translates into difficulty at the measurement level: it is not straightforward to develop reliable and valid measures of constructs whose definitions are unstable across studies.

Our empirical analysis is not designed to settle the causal architecture of the institutional system; this is beyond our scope. Instead, we design our analysis to identify, from a large pool of candidate indicators, the latent dimensions on which they cluster and to assess whether these dimensions are empirically distinct from one another. We identify three latent dimensions that are correlated but empirically separable.

2.2. What Are Institutions?

We are concerned not with the underlying causes of institutions or with institutional change over time, but with the impact that existing institutions have on the sets of choices available to individuals and organisations. This is consistent with the historical-institutionalist concern of how institutions affect individual behaviour, as opposed to the functional-institutionalist concern of where institutions come from (Zysman, 1994). Consistent with this primary goal, we adopt a definition of institutions that focuses on the role of institutions in monitoring and restricting the ability of those who have incentives to tilt the rules to their own benefit. We follow North's (1990, p. 3) definition:

Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction. In consequence, they structure incentives in human exchange, whether political, social, or economic. In the jargon of the economist, institutions define and limit the set of choices of individuals.

North differentiated between formal constraints, such as written rules, and informal constraints, such as unwritten codes of conduct. In cross-country accounting research, such informal constraints—including culture and language—have been shown to shape reporting outcomes under IFRS (Alshehbi et al., 2023; Alsabab & Alshehbi, 2026; Al-dousari & Alsabab, 2025). Institutions can be created (e.g., the United States Constitution), or they can evolve over time (e.g., the common law). North understood institutions as being like the rules of sport, defining the way the game is played, and organisations as the players, whose objective is to win the game. Whether the game produces the outcomes its rules seem to imply depends on how the rules are enforced and how violations are sanctioned. Two implications follow from this. First, institutions are shaped by the way in which they are enforced. Second, different enforcement levels can produce different patterns of outcomes; the same rule is not necessarily the same institution in countries whose enforcement environments differ. The implication for measurement is direct: a multidimensional institutional construct is not adequately captured by a single indicator, and the choice of indicator carries assumptions about which dimension of the construct is most relevant for the analysis.

3. Methodological Difficulties in the Existing Literature

There are two methodological difficulties that have not been systematically addressed in the cross-country empirical accounting literature. The first concerns the absence of standardised conceptual definitions and the second the absence of validity and reliability assessment.

3.1. The Absence of Standardised Conceptual Definitions

The first difficulty is the lack of clearly specified, standardised conceptual definitions of country-level institutions. This concern is not new, nor is it confined to international accounting research. As Kollat et al. (1970, p. 329) put it, “definitions are only means, rather than ends... So many definitions make it difficult and hazardous to compare, synthesize, and accumulate findings”.

Without standardised conceptual definitions, individual studies cannot be readily compared regarding the following question: Are the studies measuring the same construct? The reader is left to inspect the indicator in each study and infer, from its label and its formula, whether the construct it operationalises is the same construct that another study claims to measure. Such inferences are not always successful, because labels are not stable, and formulas vary in ways that labels do not always signal.

This point can be illustrated empirically. Table 1 reports how each of fourteen widely cited cross-country accounting studies published between 2003 and 2012 labels seven

indicators drawn from La Porta et al. (1997, 1998, 2006) and Djankov et al. (2008). Each cell records whether the study used the indicator as a measure of investor protection (V), securities regulation (S), or legal enforcement (F); a dash indicates that the study did not use the indicator at all. The pattern is notable: the same anti-director rights index is treated as a measure of investor protection in seven of the studies and of legal enforcement in two; the same disclosure requirements index is treated as a measure of investor protection in three studies and of securities regulation in four and as a hybrid of the two in a fifth; the public-enforcement index of La Porta et al. (2006) is variously treated as a measure of investor protection, of securities regulation, and of legal enforcement. This pattern reflects, in part, the absence of an underlying measurement model that would inform the assignment of indicators to constructs.

Table 1. Demonstration of the ad hoc measurement of country-level institutions in published cross-country accounting research.

Author(s)	Anti-Director Rights	Disclosure Requirements	Liability Standard	Anti-Director Rights (Rev.)	Anti-Self-Dealing	Public Enforcement
Hope (2003)	F	—	—	—	—	—
Leuz et al. (2003)	V	—	—	—	—	—
Burgstahler et al. (2006)	V	S	S	—	—	—
La Porta et al. (2006)	V	V/S	V/S	—	—	V/S/F
Bushman and Piotroski (2006)	—	S	S	—	—	S
Djankov et al. (2008)	—	V	V	V	V	F
Francis and Wang (2008)	V	V	V	—	—	V
Reverte (2008)	V	S	S	—	—	—
Chen et al. (2009)	V	V	V	V	V	—
Ahmed et al. (2010)	—	—	—	—	—	—
Leuz (2010)	—	S	S	V	V	S
Christensen et al. (2011)	F	—	—	—	—	F
André et al. (2015)	V	—	—	—	—	—

Notes. V = the indicator is treated as a measure of investor protection; S = the indicator is treated as a measure of securities regulation; F = the indicator is treated as a measure of legal enforcement; — = the indicator is not used in the study. The seven indicators are drawn from La Porta et al. (1997, 1998, 2006) and Djankov et al. (2008). The pattern across the table reflects the absence of an underlying measurement model that would inform the assignment of indicators to constructs.

International accounting research is a comparative enterprise: studies typically examine many countries rather than a single country, and the questions they address are framed as comparisons across countries. This requires comparable measures. The evidence in Table 1 suggests that within the existing research literature, measures that appear to be comparable can in fact be operationalising different constructs. If measurement decisions are made on case-by-case grounds, three situations are possible: the measurement may be invalid in all countries; it may be valid in some countries but not others, in which case comparisons of findings across countries are misleading; or it may be valid in all countries, in which case substantive results and comparisons hold. Without empirical assessment, it is not always possible to determine which of the three applies in a given case.

Isidro et al. (2020) reach a similar conclusion through a different approach. Their survey of more than a hundred international accounting studies reports that the same country-level constructs are operationalised differently across studies. The two sets of evidence are complementary: Isidro et al. (2020) document the scale of the issue across the literature; we develop and validate a measurement model for the institutional constructs the literature uses. In Sections 5–7, our methodological response is developed—a multi-indicator measurement model that has been subjected to validity and reliability assessment.

3.2. The Absence of Validity and Reliability Assessment

The second difficulty is that the validity and reliability of country-level institutional measures used in international accounting research have rarely been assessed using confirmatory factor analysis (CFA), which permits direct estimation of construct validity and construct reliability. In much of the literature, measures of institutions have been adopted on the grounds of face validity and on the strength of their use in prior studies, rather than on the basis of formal validation. As [Marshall \(2009\)](#) notes, this practice is common when researchers wish to operationalise a concept that they cannot measure directly, and a measure is then assigned to the concept on conventional grounds. The line between manifest and latent variables can become difficult to discern ([van der Eijk, 2007](#)), and the form of measurement employed depends on presumed rather than demonstrated relationships between the indicators and the underlying construct ([Torgerson, 1958](#), p. 22).

The investor protection construct illustrates the issue. The construct was initially measured using the anti-director rights index of [La Porta et al. \(1997, 1998\)](#), a composite of six items that focuses primarily on the voting rights of minority shareholders. [Djankov et al. \(2008\)](#) note three concerns with the index: it was constructed by selecting items on a case-by-case basis; some of its components are conceptually ambiguous; and it contains coding errors. These observations led to the construction of the revised anti-director rights index ([Djankov et al., 2008](#)) and to alternative measures of investor protection, including the disclosure requirements index, the liability standard index of [La Porta et al. \(2006\)](#), and the anti-self-dealing index of [Djankov et al. \(2008\)](#). A substantial portion of the empirical accounting literature, however, continues to use the original anti-director rights index, suggesting that it takes time for newer measures to replace established ones.

3.3. Validity, Reliability, and Construct Measurement

Addressing the two difficulties requires both clearer conceptual definitions and explicit validity and reliability assessment. [van der Eijk \(2007, p. 2\)](#) notes the following:

A number of single items should not be combined into a composite measure without further ado. Relying only on the theorising (or gut-feeling) of the analyst is insufficient to avoid the risk of combining items that do not in fact reflect the same thing, but different things instead. Such measurement-by-fiat can be avoided by the use of measurement models, which explicitly test whether or not (or to which extent) separate items do reflect the same empirical phenomenon.

Validity, as defined by [Carmines and Zeller \(1979\)](#), requires that the instrument adequately measures the underlying theoretical construct, not that the instrument is treated as a measure of that construct simply because the analyst designates it so. Reliability requires that the instrument produces consistent results when applied across settings or over time. Reliability is necessary but not sufficient for validity ([Churchill, 1979, p. 65](#)), while valid measures are also reliable to some degree.

The wider methodological literature on construct measurement, developed in political science, management, marketing, medical research, and psychology, sets out the framework for assessing validity and reliability ([Adcock, 2001](#); [Churchill, 1979](#); [MacKenzie et al., 2011](#); [Schwab, 2005](#)). International accounting research has drawn less from this body of literature, which has limited the application of formal validation techniques in the field. [MacKenzie et al. \(2011, p. 294\)](#) note that engaging this body of literature productively requires familiarity with structural equation modelling procedures. A recent methodological commentary stresses that failing to assess and transparently report the validity and reliability of measures is itself a questionable measurement practice that undermines the validity of study conclusions ([Flake & Fried, 2020](#)).

Cross-country accounting research can fail to integrate findings across studies when the method of measuring the same nominal construct varies. Three empirical possibilities follow when measurement practice is not validated. First, items hypothesised to measure the same construct may be combined into a non-homogeneous composite, in which case the estimated relationships between the combined measure and other variables are contaminated and misleading (Scheuer, 2006, p. 5). Second, items treated as conceptually distinct may in fact reflect a single homogeneous construct (van der Eijk, 2007, p. 2). Third, the items may indeed measure the same construct, in which case the operationalisation is sound. Without empirical assessment, the field cannot determine which of the three applies in a given study (Scheuer, 2006).

Factor analysis is a technique for assessing item homogeneity and identifying the latent structure underlying a set of indicators. It groups highly inter-correlated indicators into internally consistent sets, known as factors, each capturing a single underlying construct. The resulting groups display high internal homogeneity and high external heterogeneity: indicators of a single construct correlate highly with each other and weakly with indicators of different constructs (Hair et al., 2010; Ruland et al., 2007; van der Eijk, 2007).

Construct validation is, however, not exclusively a quantitative matter. As Hofstede (2011, p. 9) emphasises, equally important is the qualitative interpretation of what differences in the dimensions mean for each of the societies studied, which calls for an emic approach to each society, supporting the etic of the dimensional data. Convergent and discriminant validity, often regarded as a purely quantitative matter, should not be employed atheoretically. Scholars should have specific conceptual reasons for expecting convergence if it is to constitute evidence for validity (Adcock, 2001, p. 540). Conversely, content validity, focused predominantly on conceptual matters, can be assessed both qualitatively (through researcher judgement and expert opinion) and statistically (through homogeneity and convergent validity tests) (Hair et al., 2010).

In addition, construct validation requires the specification of two kinds of relationship: the theoretical relationship between the constructs themselves, and the statistical relationship between their measures (Collins et al., 2012). As Churchill (1979, p. 72) states, the analyst tries to establish the construct validity of a measure by relating it to a number of other constructs and not simply one. Validating a measure of a concept is not easily accomplished without a theoretical or nomological network surrounding that concept (Carmines & Zeller, 1979); the richer the network, the greater the analyst's confidence that the measure is capturing variance that is construct-valid (Schwab, 2005). Podsakoff et al. (2012) argue that determining the statistical relationships between the measures of constructs without first establishing the validity of the measures carries a particular risk, a view consistent with Campbell and Fiske (1959) and Schoenfeldt (1984, p. 78), who argue that construct validity is not only more fundamental than nomological or criterion validity but is the most important part of any study, as well as with Churchill (1979, p. 70), who holds that construct validity lies at the heart of the scientific process.

4. Item Selection and Interpretational Confounding

The conceptualisation and operationalisation of country-level institutions is further shaped by two inter-related difficulties.

4.1. Selecting Items from a Larger Pool

Many international organisations now conduct cross-national surveys that provide time series for a wide range of items asked in similar form across most countries: the World Development Indicators, the Worldwide Governance Indicators, the Global Financial Development database, the Doing Business reports, the Economic Freedom of the

World index, the Global Competitiveness Report, and others. Each of these databases offers researchers rich material for analysing the relationship between country-specific characteristics and accounting practices. The databases also overlap: a number of items are repeated, sometimes in identical form, sometimes with small differences in wording across studies and waves. Some items are combined into single composite measures; others are differently rescaled. The Economic Freedom of the World index, for example, has adopted multiple items (e.g., judicial independence) from the Global Competitiveness Report but rescaled them from their original 1-to-7 scale to a 0-to-10 scale using the formula $EFW = (GCR - 1)/6 \times 10$. Other items are combined into single variables, such as the impartial-courts measure and the extra-payments-and-favouritism measure.

A more substantive concern is the selection of particular items from a much larger pool. The pool of available items is not a shopping list from which a researcher can select what to include while ignoring everything else. The selection of items may benefit from a stated rationale, rather than being driven by convenience: based on what the items measure, not based on which items are most readily available and most familiar to readers.

A representative illustration of the selection pattern can be drawn from [Houque et al. \(2012\)](#). Recognising that investor protection is a multidimensional concept, these authors operationalise it through six alternative single-item measures applied one at a time: board independence, enforcement of securities laws, protection of minority shareholder rights, enforcement of accounting and auditing standards, judicial independence, and freedom of the press. Each measure is taken to reflect, in part, the conceptual domain of investor protection. The methodological observation here is general rather than specific to that paper: when a multidimensional construct is operationalised by entering several alternative single-item proxies into the regression model one at a time, the conceptual content of the construct moves with the content of whichever proxy is selected for a given specification. [Goldberg \(2001, p. 205\)](#) argues a related point about substitutable terms: “If one word is synonymous with another, each could be substituted for the other . . . each could be viewed as a definition of the other.” A useful follow-up question, then, is what the selected items, taken collectively, in fact represent.

A more general point follows: When alternative single-item proxies of a multidimensional construct are used interchangeably across model specifications, two methodological conditions are typically not jointly addressed. The first is content validity: each proxy may benefit from being conceptually associated with the construct of interest, and items used to operationalise the construct can be drawn from the same conceptual domain ([MacKenzie et al., 2011](#)). The second is the role of multicollinearity: where alternative proxies are entered one at a time on the grounds that their pairwise Pearson correlations are high, the high correlations are sometimes taken as evidence that the proxies capture the same underlying construct. As [Hair et al. \(2010\)](#) note, however, some degree of multicollinearity is desirable in measurement; it is what permits the analyst to identify inter-related sets of variables. [Kimberlin and Winterstein \(2008, p. 2277\)](#) note that items measuring the same construct should be correlated; [Churchill \(1979, p. 68\)](#) similarly argues that if all the items in a measure are drawn from the domain of a single construct, responses to those items should be highly intercorrelated. Multicollinearity is necessary but not sufficient for construct validity. High correlations among items are evidence that the items may belong together but are not, on their own, evidence that the items measure the construct that the analyst claims. More recent cross-country research reflects the same concern, measuring investor protection as a multidimensional construct—distinguishing shareholder from creditor protection through time-varying legal indices rather than a single cross-sectional proxy ([Kabbach de Castro et al., 2023](#)).

Isidro et al. (2020) provide complementary evidence on the selection issue. Using a sample of 35 countries, they regressed a country-level reporting-quality factor on each of the 72 candidate country-level attributes one at a time and found 45 variables to be individually significant, a result that is broadly consistent with the pattern observed in the prior cross-country accounting literature. When each individual variable was conditioned on four latent factors extracted from the remaining 71 country-level variables, only eight remained significant; once Bonferroni correction was applied, only one survived. The implication is that individual country-level variables can appear significant in cross-country regressions while the explanatory power of these individual variables is largely absorbed by a small number of latent factors.

A composite measurement approach addresses both content validity and the multicollinearity issue. Combining items into a composite measure, on condition that the items are both theoretically homogeneous (content valid) and empirically homogeneous (highly correlated), reduces non-systematic measurement error (Churchill, 1979) and permits direct assessment of validity and reliability (van der Eijk, 2007). As Churchill (1979, p. 68) notes, no single item is likely to provide a perfect representation of the concept, just as no single word can be used to test for differences in individual spelling abilities.

4.2. Interpretational Confounding

A second concern is interpretational confounding (Burt, 1976; MacKenzie et al., 2011), which arises when there is a discrepancy between the nominal meaning of a construct (the meaning the analyst attributes to it) and the empirical meaning produced by the relationship between the latent construct and the observed variables used to measure it. The discrepancy is absorbed into the regression coefficient, where it is difficult to detect without an independent measurement model.

This concern is particularly relevant in cross-country settings, where survey-based items can carry different connotations across languages and cultural contexts. By incorporating slightly different nuances of meaning into the items in a pool, an analyst can construct a stronger foundation for the eventual measure (Churchill, 1979); this is more difficult when the analyst relies entirely on items prepared by international organisations. The assumption that items, having been used in prior studies, mean what they appear to mean is more fragile in cross-country than in single-country research. Several of the indicators that load on the quality-of-legality construct are perception-based survey items, drawn for instance from the Global Competitiveness Report, and such items may carry cultural, linguistic, or respondent-specific biases across the seventy countries in the sample. A latent-variable framework partially mitigates this concern: the construct is identified from the variance shared across indicators that originate from different databases and different measurement methodologies, rather than from any single perception-based item, so idiosyncratic survey bias that is not common to the other indicators is relegated to the error term. It does not, however, eliminate biases that are common across perception-based sources—a concern that the producers of these indicators themselves analyse under the heading of correlated perception errors (Kaufmann & Kraay, 2024)—and the relationship between perceived and actual legal execution remains an assumption rather than a demonstrated equivalence.

Isidro et al. (2020) address a related concern empirically. They examined 21 country-level variables across the staggered mandatory adoption of IFRS in 30 countries and found an increase in the joint co-movement of changes during IFRS adoption windows: the average maximum adjusted R^2 from regressing each change variable on four other change variables rises from 0.49 in the pre-adoption window to 0.60 during the adoption window, then falls to 0.45 in the post-adoption window. Once the 21 changing country variables were added to a regression of changes in reporting quality on IFRS adoption, the IFRS-adoption

indicator lost its statistical significance. This pattern is consistent with the methodological literature on what occurs when an outcome is regressed on a proxy that co-moves with other unmodelled correlates. Although [Isidro et al. \(2020\)](#) develop this argument in the changes-and-shocks setting, the underlying mechanism—high causal density and co-dependencies among institutional variables—is also relevant in the levels analyses on which most of the literature relies.

5. Data and Methodology

In this section, we describe our data, the candidate institutional indicators, and the factor-analytic procedures we use to develop and validate our measurement model. Our empirical strategy proceeds in two stages. In the first stage, we use exploratory factor analysis (EFA) to identify the latent dimensions on which the candidate indicators cluster. In the second stage, we use confirmatory factor analysis (CFA) to assess the construct validity, convergent reliability, and discriminant validity of the measurement model identified by the EFA.

5.1. Sample and Indicators

Our sample comprises 70 countries, the same set used by [Djankov et al. \(2008\)](#) excluding Taiwan and Tunisia, for which several of the candidate indicators are unavailable. The countries cover the major legal-origin families (English, French, German, Scandinavian, and socialist) and span the range of equity market development, from frontier markets to the deepest exchanges in the world. We derive country-level institutional indicators from six databases: the World Bank's World Development Indicators; the Worldwide Governance Indicators ([Kaufmann et al., 2010](#)); the World Bank's Global Financial Development database; the World Bank's Doing Business reports; the Fraser Institute's Economic Freedom of the World index; and the World Economic Forum's Global Competitiveness Report. Where annual values vary, we use the mean over 2006–2010 to dampen short-term fluctuations.

We begin our analysis with 48 candidate indicators selected on face validity grounds from prior empirical work ([La Porta et al., 1997, 1998](#); [Leuz et al., 2003](#); [Djankov et al., 2008](#); [Leuz, 2010](#); [Glaum et al., 2013](#)) and from the methodological literature on institutional measurement ([Kaufmann et al., 2010](#)). The 48 indicators include rights-based measures (anti-director rights, anti-self-dealing, strength of investor protection, business extent of disclosure, ex-ante and ex-post private control of self-dealing); governance-perception measures (six Worldwide Governance Indicators components); survey-based regulatory measures (eight Global Competitiveness Report items); legal system measures (eight Economic Freedom of the World components); and equity market measures (three Global Financial Development items). We report the full list of candidate indicators and their sources in Appendix A.

5.2. Exploratory Factor Analysis

We use exploratory factor analysis (EFA) when a small number of latent constructs is hypothesised to underlie a larger number of observed indicators ([Hair et al., 2010](#)). We use principal component extraction with oblique (Promax) rotation because the latent constructs of interest are expected to be correlated for substantive reasons: countries with stronger legal institutions tend, on average, to attract larger and more liquid equity markets, and stronger legal protection has been shown to be more effective when the institutions enforcing those protections are themselves strong. [Isidro et al. \(2020\)](#), whose EFA on 72 country attributes uses varimax rotation, note that orthogonal and oblique solutions converge when inter-factor correlations are below approximately 0.32 and report a highest pairwise factor correlation of 0.19 in their data. In our sample, the corresponding

correlations are 0.380, 0.384, and 0.517, which makes oblique rotation the appropriate choice. We standardise indicators before extraction so that no indicator dominates the analysis through scale alone. This two-stage exploratory-then-confirmatory design is the approach recommended in the measurement and structural equation modelling literature when a large pool of candidate indicators is expected to reflect a smaller number of latent constructs: EFA identifies the underlying factor structure without imposing prior restrictions, and CFA then tests that structure on the same data and yields direct estimates of construct validity, convergent reliability, and discriminant validity (Churchill, 1979; MacKenzie et al., 2011; Cheung et al., 2024).

Three diagnostic decisions govern our analysis. First, we assess the sampling adequacy of the correlation matrix using the Kaiser–Meyer–Olkin (KMO) measure (Kaiser, 1974) and Bartlett’s (1954) test of sphericity. Second, we retain factors if their eigenvalues exceed unity, the standard Kaiser criterion. Third, we retain an indicator as part of a factor only if its loading on that factor exceeds 0.65, the threshold Hair et al. (2010) recommend for samples of about 70 observations, and only if it does not cross-load above 0.40 on any other factor. Our threshold is conservative: we discard indicators that are theoretically attractive but empirically ambiguous on the principle that an indicator that fails to load cleanly is more likely to introduce measurement error than to reduce it.

5.3. Confirmatory Factor Analysis

EFA can produce a measurement model that fits the data on which it was estimated; but that model will not necessarily survive a fresh sample (Everitt & Hothorn, 2011). We split our 70-country sample at random into two equal sub-samples of 35 countries and re-estimate the factor structure on each half. We then compare the loadings and communalities in the second sub-sample with those in the first to confirm that the factor structure is not an artefact of the EFA sample.

We then perform confirmatory factor analysis (CFA) on the full sample using maximum likelihood in AMOS 22, with each indicator constrained to load on the factor identified by the EFA and on no other factor. We evaluate model fit using the chi-square to degrees-of-freedom ratio (CMIN/DF; values below 5 acceptable), the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the root mean square error of approximation (RMSEA; values below 0.10 acceptable, below 0.08 good), and the comparative fit index (CFI; values above 0.90 acceptable) (Hu & Bentler, 1999).

We assess convergent reliability using composite reliability (CR; threshold 0.70) and average variance extracted (AVE; threshold 0.50) (Fornell & Larcker, 1981; Cheung et al., 2024; Hair et al., 2022). We assess discriminant validity by comparing each construct’s AVE with the squared correlations between that construct and every other construct (Fornell & Larcker, 1981); the AVE should exceed every squared inter-construct correlation. As a complementary check, we follow Brown (2012) in ensuring that no inter-factor correlation exceeds 0.85, the threshold above which discriminant validity is compromised, consistent with recent guidance on assessment of discriminant validity (Rönkkö & Cho, 2022).

6. Exploratory Factor Analysis: Results

6.1. Sampling Adequacy and the Three-Factor Solution

Bartlett’s test of sphericity rejects the null hypothesis that the correlation matrix is the same as the identity matrix at the 0.001 level, indicating that the indicators are sufficiently inter-correlated to support factor extraction. The KMO measure of sampling adequacy is 0.885, well above the 0.65 threshold that Kaiser (1974) considers to be necessary and within the range that Hair et al. (2010) describe as meritorious. Three factors emerge with eigenvalues above unity: the first with an eigenvalue of 10.19, the second with an

eigenvalue of 2.38, and the third with an eigenvalue of 1.35. Together, they explain 81.84% of the variance in the original data, with each explaining 59.91%, 14.00%, and 7.93%, respectively. The Kaiser criterion and the Cattell scree plot agree on three factors.

Of the original 48 indicators, 16 survive the dual screening rule that loadings on the assigned factor exceed 0.65 and that no cross-loadings exceed 0.40. The remaining 32 indicators are dropped: most fail because they cross-load on two or three factors, and a few fail because their loadings on every factor are below the threshold. Each of the three retained factors is supported by a balanced set of indicators rather than by one or two dominant items.

6.2. The Structure Matrix

Table 2 presents the rotated structure matrix for the 16 retained indicators. Loadings below 0.30 are suppressed for clarity.

Table 2. Structure matrix from exploratory factor analysis (n = 70 countries).

Indicator	Communality	F1: Quality of Legality	F2: Investor Protection	F3: Equity Market Dev.
Government Effectiveness	0.932	0.963		
Control of Corruption	0.936	0.961		
Regulatory Quality Index	0.924	0.960		
Rule of Law	0.925	0.957		
Ethical Behaviour of Firms	0.915	0.955		
Strength of Auditing and Reporting Standards	0.865	0.922		
Regulatory Quality (perception)	0.825	0.907		
Protection of Minority Shareholders' Interests	0.804	0.875		
Efficacy of Corporate Boards	0.763	0.851		
Regulation of Securities Exchanges	0.724	0.845		
Anti-Self-Dealing Index	0.830		0.903	
Strength of Investor Protection Index	0.808		0.886	
Business Extent of Disclosure Index	0.757		0.868	
Revised Anti-Director Rights Index	0.552		0.742	
Stock Market Capitalisation/GDP	0.862			0.926
Stock Market Value Traded/GDP	0.822			0.905
Number of Listed Companies (per million)	0.670			0.813
Eigenvalue		10.190	2.380	1.350
Explained Variance (%)		59.91	14.00	7.93

Notes. Loadings below 0.30 are suppressed. Extraction is performed via principal component analysis with Promax rotation. The Kaiser–Meyer–Olkin measure of sampling adequacy is 0.885; Bartlett's test of sphericity is significant at the 0.001 level. All retained indicators have communalities exceeding 0.50 and primary loadings exceeding 0.65, the threshold that [Hair et al. \(2010\)](#) recommend for a sample of 70 observations.

6.3. Interpreting the Three Factors

Each factor admits a clear substantive interpretation. Naming the factors is largely a matter of letting the indicators speak.

Factor 1: Quality of Legality. Ten indicators load on this factor, with primary loadings between 0.845 and 0.963. The cluster combines the four Worldwide Governance Indicators components (government effectiveness, control of corruption, regulatory quality, rule of law) with six survey-based items from the Global Competitiveness Report and the Doing Business reports (ethical behaviour of firms, strength of auditing and reporting standards, regulatory quality perception, protection of minority shareholders' interests, efficacy of corporate boards, regulation of securities exchanges). The unifying construct is the effectiveness of the institutions that translate written rules into practice: the legal system, the regulatory apparatus, and the auditing infrastructure. The loadings indicate that these indicators, often used as alternatives in single-item designs, in fact measure the same latent construct. Consistent with [North's \(1990\)](#) emphasis on enforcement, several of

these indicators—rule of law, control of corruption, government effectiveness, regulatory quality, the regulation of securities exchanges, and the protection of minority shareholders' interests—capture enforcement-related dimensions rather than statutory rules alone, so the construct reflects the effectiveness of the broader legal and regulatory environment in translating written rules into practice. It does not, however, directly incorporate narrower measures of judicial enforcement such as court rulings, specialised commercial courts, or case-level contract-enforcement outcomes; augmenting the construct with such judicial-enforcement variables, in the spirit of recent country-level composite indicators that combine governance, audit, and enforcement activity (Quagli et al., 2021), would be a valuable refinement of the measurement model.

Factor 2: Investor Protection. Four indicators load on this factor, with primary loadings between 0.742 and 0.903. Three are rights-based—the anti-self-dealing index, the strength of investor protection index, and the business extent of disclosure index—and the fourth is the revised anti-director rights index of Djankov et al. (2008). All four were developed explicitly to measure the legal protection enjoyed by minority outside shareholders against expropriation by insiders. The clean loading pattern indicates that these are measures of the same underlying construct, but the moderate communality of the revised anti-director rights index (0.552) indicates that, if used alone, this index would absorb idiosyncratic variation that the other three indicators miss.

Factor 3: Equity Market Development. Three indicators load on this factor, with primary loadings between 0.813 and 0.926: stock market capitalisation scaled by GDP, value traded scaled by GDP, and the number of domestic listed companies scaled by population. The cluster captures the size, liquidity, and breadth of the domestic equity market, i.e., the economic depth of the institution rather than the rules that govern it.

6.4. Discriminant Validity at the EFA Stage

The Promax rotation produces a component correlation matrix that is informative for discriminant validity. The correlations between the three factors are 0.380 (investor protection with quality of legality), 0.517 (quality of legality with equity market development), and 0.384 (investor protection with equity market development). All three correlations are well below the 0.85 threshold above which Brown (2012) treats discriminant validity as compromised. The factors are correlated—countries with strong legal institutions also tend to have strong investor protection and deep equity markets—but not so strongly as to be conflated.

The structure matrix in Table 2 displays no indicator with cross-loadings above 0.40 on a non-primary factor (Table 3). The loading pattern is itself evidence of discriminant validity: each indicator is empirically associated with one and only one of the three constructs, despite the moderate inter-factor correlations.

Table 3. Component correlation matrix from exploratory factor analysis.

Component	1	2	3
1 (Quality of Legality)	1.000	0.380	0.517
2 (Investor Protection)	0.380	1.000	0.384
3 (Equity Market Dev.)	0.517	0.384	1.000

Notes. Component 1 = Quality of Legality; Component 2 = Investor Protection; Component 3 = Equity Market Development. Inter-component correlations from the Promax-rotated principal component solution. All correlations are well below the 0.85 threshold, above which discriminant validity is compromised (Brown, 2012).

6.5. Implications for Commonly Used Single-Item Proxies

The factor structure permits an evaluation of the single-item proxies most often used in the cross-country accounting literature. Three observations follow. First, the four

Worldwide Governance Indicators components that load on the quality-of-legality factor (government effectiveness, control of corruption, regulatory quality, rule of law) carry similar loadings (0.957–0.963) and similar communalities (0.924–0.936). The four indicators are statistically close to being interchangeable: studies that select one rather than the others are, for measurement purposes, working with very similar variables. Second, the four indicators on the investor protection factor are not similarly close. The revised anti-director rights index has the lowest loading (0.742) and the lowest communality (0.552) among the retained indicators on this factor, so studies that rely on this index alone are operationalising investor protection with a measure that captures less of the underlying construct than the alternatives. Third, among the three equity market development indicators, market capitalisation and value traded have similar loadings (0.926 and 0.905), while the number of listed companies has a lower loading (0.813) and the lowest communality (0.670) of the three.

These observations are consistent with the methodological literature on construct measurement (Churchill, 1979; Hair et al., 2010) but have not, to our knowledge, been reported jointly for the institutional indicators most widely used in cross-country accounting research. The implication for empirical practice is that the choice of single-item proxy carries little measurement consequence for some constructs and substantial measurement consequence for others and can be presented with that distinction in mind.

A complementary observation arises on the dependent-variable side. Isidro et al. (2020) apply factor analysis to six commonly used country-level reporting-quality measures (transparency, disclosure quality, asymmetric timeliness, abnormal return, abnormal volume, and return synchronicity) and find that a single underlying factor explains 88.9% of the variance, with all six measures loading strongly on it. Their finding parallels the finding here for the explanatory side: what the literature treats as multiple distinct properties under different labels is, on examination, a smaller number of latent properties measured under many labels. Taken together, the two sets of findings suggest that the construct content of cross-country regressions is concentrated on a small number of latent dimensions on both the explanatory and the dependent-variable sides.

7. Confirmatory Factor Analysis: Validation

7.1. Cross-Sample Stability

Following Everitt and Hothorn (2011), we split the 70-country sample into two halves of 35 countries each at random and re-estimate the factor structure on each half. We find that the loadings and communalities are quantitatively very close across the two sub-samples; no indicator changes its primary factor, and no factor admits a new indicator. We conclude that the three-factor structure is not an artefact of the EFA sample.

7.2. Model Fit

Our CFA specifies a three-factor measurement model in which each retained indicator loads on the factor identified by the EFA and on no other factor. We allow the three latent factors to correlate freely. Maximum likelihood estimation in AMOS 22 produces an overall model fit that meets conventional thresholds. The chi-square to degrees-of-freedom ratio (CMIN/DF) is 1.429, well below the upper bound of 5. The goodness-of-fit index (GFI) is 0.861, and the adjusted goodness-of-fit index (AGFI) is 0.788; both are below the 0.90 cut-off but within the range that Hu and Bentler (1999) describe as acceptable for small samples. The root mean square error of approximation (RMSEA) is 0.079, below the acceptable threshold of 0.10 and close to the boundary of 0.08 that is considered good. The comparative fit index (CFI) is 0.970, and the normed fit index (NFI) is 0.908, both above the

0.90 threshold. Figure 1 displays the full measurement model with standardised loadings, error variances, and inter-construct correlations.

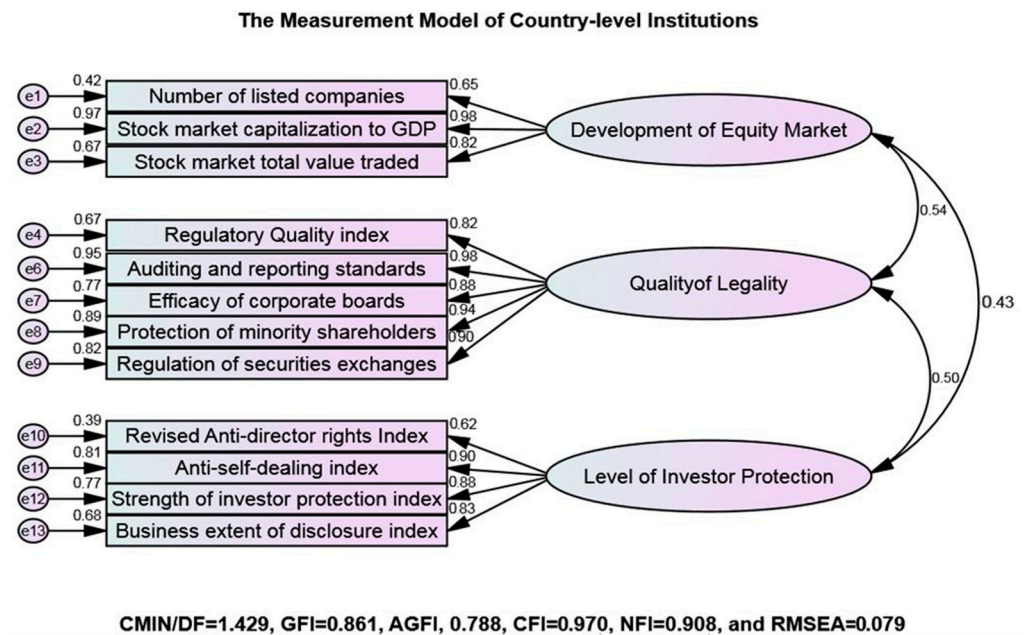


Figure 1. The measurement model of country-level institutions. Notes. Confirmatory factor analysis (CFA) path diagram for the three-factor measurement model of country-level institutions, estimated by maximum likelihood in AMOS 22 on the sample of 70 countries. Rectangles denote observed indicators; ellipses denote latent constructs; circles labelled e1–e13 denote error terms. Numbers next to single-headed arrows from latent constructs to indicators are standardised loadings. Numbers above the error terms are squared multiple correlations. Numbers on the curved double-headed arrows linking the three latent constructs are inter-construct correlations. Overall fit indices: CMIN/DF = 1.429; GFI = 0.861; AGFI = 0.788; CFI = 0.970; NFI = 0.908; RMSEA = 0.079.

The fit indices indicate that the three-factor measurement model provides a satisfactory description of the observed correlations among the 12 retained indicators. The model is parsimonious, with each construct identified by a small number of high-loading indicators, and each construct is empirically distinct from the others.

7.3. Convergent Reliability

Table 4 reports the standardised regression weights from the CFA. All loadings are positive and statistically significant at the 0.001 level. Ten of the twelve standardised loadings exceed the 0.70 threshold that Hair et al. (2010) recommend; the two that fall below the threshold—the revised anti-director rights index (0.622) and the number of listed companies (0.646)—are kept on theoretical grounds, as the only available indicators of one dimension of their respective constructs.

Composite reliability (CR), which adjusts coefficient alpha for unequal loadings, is 0.86 for equity market development, 0.96 for quality of legality, and 0.88 for investor protection—all above the 0.70 threshold (Fornell & Larcker, 1981). The constructs are reliable in the conventional sense: the indicators within each construct cohere as measures of one common latent factor.

7.4. Discriminant Validity

Table 5 reports the AVE for each construct on the diagonal and the inter-construct correlations off-diagonal. The AVEs are 0.69 (equity market development), 0.83 (quality of legality), and 0.66 (investor protection)—all above the 0.50 threshold (Fornell & Larcker, 1981). To establish discriminant validity, the AVE for each construct should exceed every

squared correlation between that construct and another. Table 6 reports the squared inter-construct correlations alongside their corresponding raw correlations: the squared values are 0.29, 0.18, and 0.25, all being below the lowest AVE in the model (0.66). The constructs share less variance with each other than each construct shares with its own indicators, meeting the formal Fornell–Larcker criterion for discriminant validity.

Table 4. Standardised CFA loadings.

Indicator	Std. Loading	Critical Ratio	p-Value	Factor
Regulatory Quality Index	0.820	—	—	F1
Strength of Auditing and Reporting Standards	0.976	12.081	<0.001	F1
Efficacy of Corporate Boards	0.880	9.823	<0.001	F1
Protection of Minority Shareholders' Interests	0.944	11.316	<0.001	F1
Regulation of Securities Exchanges	0.905	10.381	<0.001	F1
Business Extent of Disclosure Index	0.826	—	—	F2
Strength of Investor Protection Index	0.877	8.718	<0.001	F2
Anti-Self-Dealing Index	0.899	8.973	<0.001	F2
Revised Anti-Director Rights Index	0.622	5.491	<0.001	F2
Stock Market Value Traded/GDP	0.820	—	—	F3
Stock Market Capitalisation/GDP	0.983	8.629	<0.001	F3
Number of Listed Companies (per million)	0.646	5.947	<0.001	F3

Notes. Loadings reported for one indicator per factor are constrained to identify the model. F1 = Quality of Legality; F2 = Investor Protection; F3 = Equity Market Development. Estimation is by maximum likelihood in AMOS 22.

Table 5. Construct reliability, average variance extracted, and inter-construct correlations.

Construct	CR	AVE	F1	F2	F3
F1 (Quality of Legality)	0.96	0.83	0.91		
F2 (Investor Protection)	0.88	0.66	0.50	0.81	
F3 (Equity Market Development)	0.86	0.69	0.54	0.43	0.83

Notes. CR = composite reliability (threshold > 0.70). AVE = average variance extracted (threshold > 0.50; AVE > SIC for discriminant validity). Diagonal entries (in bold) are the square roots of the AVEs. Off-diagonal entries are inter-construct correlations from the CFA. For discriminant validity, each diagonal entry should exceed each off-diagonal entry in its row and column (Fornell & Larcker, 1981).

Table 6. Squared inter-construct correlations.

Inter-Construct Correlation	SIC
0.54	0.29
0.43	0.18
0.50	0.25

Notes. SIC = squared inter-construct correlation. The three pairwise CFA correlations from Table 5 are reported alongside their squared values. For discriminant validity, each construct's AVE (Table 5) should exceed every squared correlation involving that construct (Fornell & Larcker, 1981). The lowest AVE in the model is 0.66; the highest SIC is 0.29.

The three constructs are therefore not interchangeable, even when their inter-correlations are economically meaningful. A country that scores high on quality of legality is somewhat more likely to score highly on investor protection (correlation 0.50) and equity market development (correlation 0.54), but the overlap is far from complete. Substituting one construct for another in an empirical model will introduce systematic measurement error, not random error.

8. Conclusions

In this paper, we develop and validate a three-factor measurement model for the country-level institutions used in international accounting research. Our study makes three contributions to the existing literature. First, the same widely cited indicators have been used in published cross-country accounting studies as measures of different constructs, with the anti-director rights index, the disclosure requirements index, and the public enforcement index variously labelled as measures of investor protection, securities regulation, or legal enforcement across studies. Second, the construct validity and reliability of the country-level institutional measures used in the literature have rarely been assessed. Third, conditioning on a small number of latent factors substantially reduces the explanatory power of individual country-level variables in cross-country regressions, as documented by [Isidro et al. \(2020\)](#).

Our empirical analysis identifies three latent constructs—investor protection, quality of legality, and equity market development—that jointly explain 81.84% of the variance across 48 candidate indicators sourced from six international databases. We replicate the three-factor structure in two random sub-samples, which is supported by a confirmatory factor analysis that shows satisfactory overall fit ($CMIN/DF = 1.429$; $CFI = 0.970$; $RMSEA = 0.079$). All three constructs satisfy conventional thresholds for composite reliability and average variance extracted, and the squared inter-construct correlations are below the lowest AVE in the model, indicating discriminant validity in the formal Fornell–Larcker sense.

Our findings have three implications for empirical practice in cross-country accounting research. First, country-level institutions can be operationalised as composite measures derived from our validated three-factor model rather than through whichever single indicator is most readily available. We report country-level composite scores derived from the model in Appendix B, and they are available for use in future cross-country studies. Second, we argue that the choice of single-item proxy can be calibrated to the construct content of the proxy. The four Worldwide Governance Indicators components that load on the quality-of-legality factor are statistically near-interchangeable, so the choice between them carries little measurement consequence. The four indicators loaded on the investor protection factor are not interchangeable; in particular, the original anti-director rights index has the lowest loading and communality out of all the indicators loaded on that factor and would be a noisier measure than the alternatives. Third, we argue that the construct labels in published studies can be aligned with the construct content of the indicators that operationalise them. A study using the anti-director rights index alone is operationalising a rights-based dimension of investor protection rather than investor protection in the broader sense.

Our research has three limitations that suggest avenues for future research. First, we develop our measurement model on a single cross-section averaged over 2006 to 2010. Validation in earlier or later windows would strengthen confidence in the temporal stability of the three-factor structure. In particular, the sample pre-dates a number of significant post-2010 reforms to corporate governance, legal systems, and disclosure practices, and it spans the 2008 global financial crisis; averaging each indicator over 2006 to 2010 dampens the influence of short-term market shocks, including those of the crisis, on the equity-market indicators but does not remove them entirely. Re-estimating the exploratory and confirmatory analyses on a recent multi-year window (for example, 2019 to 2024) would test whether the same three constructs emerge with comparable fit under the current regulatory environment. Second, we do not directly demonstrate the consequences of measurement choice for the substantive inferences drawn in published studies. A re-analysis of an established cross-country result using our validated composite measures, alongside the single-item proxies it currently uses, would be a natural next step. In re-examining such

results, the validated composite constructs can replace the single-item proxies currently used to explain a wide range of accounting and finance outcomes—including earnings management, financial reporting quality, disclosure quality, value relevance, audit quality, and the cost of capital—so that the contribution of the measurement model is ultimately to improve the quality of the explanatory variables on which much international accounting research relies. Third, our analysis focuses on the explanatory-variable side of cross-country regressions. [Isidro et al. \(2020\)](#) document a similar measurement issue on the dependent-variable side, where six commonly used reporting-quality measures collapse onto a single factor. Integrating both sides into a single structural model would be a productive extension.

In sum, we contribute to the international accounting literature by providing a validated measurement model for the country-level institutions that the literature relies on and by documenting that the single-item proxies in widespread use are not, based on the evidence, interchangeable.

Author Contributions: Conceptualisation, M.A. and A.A.; methodology, A.A.; software, A.A.; validation, M.A. and A.A.; formal analysis, A.A.; investigation, A.A.; resources, A.A.; data curation, A.A.; writing—original draft preparation, M.A. and A.A.; writing—review and editing, M.A. and A.A.; visualisation, A.A.; supervision, A.A.; project administration, M.A.; funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The country-level institutional data analysed in this study were derived from publicly available sources: the World Bank’s World Development Indicators, the Worldwide Governance Indicators, the World Bank’s Global Financial Development database, the World Bank’s Doing Business reports, the Fraser Institute’s Economic Freedom of the World index, and the World Economic Forum’s Global Competitiveness Report. Composite scores derived from the measurement model are reported in Appendix B. Replication materials are available on request.

Acknowledgments: This publication was made possible by the support of the Abdullah al Salem University open access fund.

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

Appendix A. Description of the Institutional Variables

In this appendix, we set out the candidate institutional variables used in our analysis, organised under the three latent constructs identified by our EFA and validated by our CFA. We selected the 48 candidate indicators on face-validity grounds from prior empirical work and from the methodological literature on institutional measurement. Variable definitions and data sources follow the conventions of the source databases and of the comparative institutional literature ([La Porta et al., 1997, 1998](#); [Djankov et al., 2008](#); [Kaufmann et al., 2010](#)).

Appendix A.1. Investor Protection

Investor protection is the principal component of (i) the revised anti-director rights index; (ii) the anti-self-dealing index; (iii) the strength of investor protection index; and (iv) the business extent of disclosure index.

Appendix A.1.1. Revised Anti-Director Rights Index

An aggregate measure of the legal protection of minority shareholders against expropriation by corporate insiders. Source: [Djankov et al. \(2008\)](#).

Appendix A.1.2. Anti-Self-Dealing Index (0–1)

The average of ex-ante and ex-post private control over self-dealing transactions. Source: [Djankov et al. \(2008\)](#).

Ex-ante private control of self-dealing: Captures the strength of private enforcement of provisions against self-dealing by insiders, focusing on ex-ante control (e.g., requiring approval by disinterested shareholders and ex-ante disclosures). Source: [Djankov et al. \(2008\)](#).

Ex-post private control of self-dealing: Captures the strength of private enforcement of provisions against self-dealing by insiders, focusing on ex-post control (e.g., periodic-filings requirements and ease of proving wrongdoing). Source: [Djankov et al. \(2008\)](#).

Appendix A.1.3. Strength of Investor Protection Index (0–10)

The average of (i) the extent of disclosure index, (ii) the extent of director liability index, and (iii) the ease of shareholder suits index. Source: Doing Business Indicators (World Bank Group).

Appendix A.1.4. Extent of Disclosure Index

Measures the extent to which investors are protected through disclosure of ownership and financial information. The index ranges from 0 to 10, with higher values indicating more disclosure. Source: World Bank/World Development Indicators.

Appendix A.2. Quality of Legality

Quality of legality is the principal component of (i) the regulatory quality index; (ii) strength of auditing and reporting standards; (iii) efficacy of corporate boards; (iv) protection of minority shareholders; and (v) regulation of securities exchanges.

Appendix A.2.1. Regulatory Quality Index

The simple average of (i) judicial independence; (ii) impartial courts; (iii) protection of property rights; (iv) military interference in the rule of law and politics; (v) integrity of the legal system; (vi) legal enforcement of contracts; (vii) extra payments/bribes/favouritism; and (viii) government enterprises and investment. Sources: Economic Freedom of the World; Global Competitiveness Report.

Judicial independence. From the Global Competitiveness Report; rescaled from 1–7 to 0–10 using $EFW = ((GCR - 1)/6) \times 10$. Source: World Economic Forum.

Impartial courts. From the Global Competitiveness Report; Rule of Law ratings from the WGI fill in country omissions. Source: World Economic Forum.

Protection of property rights. From the Global Competitiveness Report. Source: World Economic Forum.

Military interference in rule of law and politics. Based on the International Country Risk Guide Component G (Military in Politics); Political Stability and Absence of Violence ratings from the WGI fill in country omissions. Source: World Economic Forum.

Integrity of the legal system. Based on the International Country Risk Guide (Law and Order). Source: World Economic Forum.

Legal enforcement of contracts. Based on World Bank Doing Business estimates for time and money to collect a debt. Source: World Economic Forum.

Extra payments/bribes/favouritism. Based on the Global Competitiveness Report. Source: World Economic Forum.

Government enterprises and investment. Government investment as a share of total investment, rescaled to 0–10. Source: World Economic Forum.

Appendix A.2.2. Strength of Auditing and Reporting Standards

Global Competitiveness Report (1–7 scale). Source: Global Competitiveness Report.

Appendix A.2.3. Efficacy of Corporate Boards

Global Competitiveness Report (1–7 scale). Source: Global Competitiveness Report.

Appendix A.2.4. Protection of Minority Shareholders

Global Competitiveness Report (1–7 scale). Source: Global Competitiveness Report.

Appendix A.2.5. Regulation of Securities Exchanges

Global Competitiveness Report (1–7 scale). Source: Global Competitiveness Report.

Appendix A.3. Development of Equity Market

Development of equity market is the principal component of (i) the ratio of the number of domestic firms listed in a given country to its population; (ii) market capitalisation of listed companies (% of GDP); and (iii) stock market total value traded to GDP.

Appendix A.3.1. Average Ratio of Total Market Capitalisation to GDP for the Period 2006–2010

Source: World Bank/Global Financial Development (GFD).

Appendix A.3.2. Natural Logarithm of the Average Ratio of the Number of Domestic Firms Listed in a Given Country to Its Population (In Millions) for the Period 2006–2010

Source: World Bank/GFD.

Appendix A.3.3. Average Ratio of the Total Value of Shares Traded to GDP for the Period 2006–2010

Source: World Bank/GFD.

Appendix B. Data on Measures of Institutions for 70 Countries

Table A1. Composite scores on the three institutional constructs, by legal origin.

Country	Legal Origin	Investor Protection	Equity Market Dev.	Quality of Legality
Australia	English	8.70	141.03	11.91
Canada	English	9.65	140.98	11.91
Ghana	English	8.24	23.10	9.51
Hong Kong	English	11.87	510.38	11.75
India	English	8.02	101.66	10.46
Ireland	English	10.43	53.53	10.57
Israel	English	9.54	117.84	10.95
Jamaica	English	6.04	78.04	10.12
Kenya	English	4.72	49.73	8.93
Malaysia	English	11.31	143.33	10.91
New Zealand	English	11.62	56.13	12.10
Nigeria	English	6.55	36.38	8.32

Table A1. Cont.

Country	Legal Origin	Investor Protection	Equity Market Dev.	Quality of Legality
Pakistan	English	6.98	40.22	8.70
Singapore	English	11.93	188.57	12.11
South Africa	English	10.27	232.95	12.24
Sri Lanka	English	6.19	35.34	10.15
Thailand	English	9.70	75.11	9.93
Uganda	English	4.95	22.74	8.11
United Kingdom	English	11.04	144.36	11.50
United States	English	9.10	141.71	10.78
Zimbabwe	English	6.58	216.99	9.84
Argentina	French	5.54	31.81	7.89
Belgium	French	8.22	76.73	11.22
Bolivia	French	3.45	24.12	7.01
Brazil	French	6.28	76.32	9.83
Chile	French	8.13	121.70	10.83
Colombia	French	8.13	56.80	8.78
Ecuador	French	3.26	17.38	7.74
Egypt	French	5.85	73.23	9.18
El Salvador	French	4.47	33.91	8.88
France	French	7.27	95.69	11.12
Greece	French	3.60	56.89	9.52
Indonesia	French	8.36	47.91	9.50
Italy	French	6.68	42.95	8.28
Jordan	French	4.64	174.27	10.39
Kazakhstan	French	7.18	48.56	8.17
Lithuania	French	6.05	30.26	9.73
Luxembourg	French	5.56	204.41	11.51
Mexico	French	6.14	44.16	9.16
Morocco	French	5.80	83.34	8.86
Netherlands	French	4.71	100.78	11.54
Panama	French	4.74	40.54	9.80
Peru	French	7.65	71.52	9.43
Philippines	French	4.64	62.56	9.42
Portugal	French	6.82	53.44	10.07
Romania	French	7.67	36.92	8.73
Russia	French	6.28	81.61	7.61
Spain	French	6.35	110.52	9.74
Turkey	French	7.01	44.71	8.84
Ukraine	French	3.75	40.27	7.06
Uruguay	French	4.41	11.53	9.03
Venezuela	French	2.73	11.48	7.64
Austria	German	5.37	46.11	11.43
Bulgaria	German	8.29	38.65	8.17
China	German	8.01	110.89	8.81
Croatia	German	4.12	67.70	8.73
Czech Republic	German	5.41	38.89	9.73
Germany	German	5.82	60.67	11.46
Hungary	German	4.18	36.52	9.93
Japan	German	8.09	98.08	10.55
Korea (Rep.)	German	7.14	106.75	9.63
Latvia	German	6.17	20.35	9.24
Poland	German	6.26	46.02	9.49
Slovak Republic	German	5.06	18.97	9.34
Switzerland	German	3.97	243.12	11.37
Denmark	Scandinavian	7.63	83.35	11.78

Table A1. Cont.

Country	Legal Origin	Investor Protection	Equity Market Dev.	Quality of Legality
Finland	Scandinavian	7.08	100.71	12.18
Iceland	Scandinavian	5.94	108.34	10.39
Norway	Scandinavian	7.56	81.09	12.04
Sweden	Scandinavian	6.78	126.88	12.47

Notes. Composite scores are the principal component values from the EFA. Higher values indicate stronger investor protection, more developed equity markets, and higher quality of legality. Country values are averaged over 2006–2010. Legal-origin classification follows La Porta et al. (1998) and Djankov et al. (2008).

References

- Adcock, R. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, 95, 529–546. [\[CrossRef\]](#)
- Ahmed, A. S., Billings, B. K., Morton, R. M., & Stanford-Harris, M. (2010). Do positive earnings surprises lead to increased optimism in analyst forecasts? *Journal of Accounting Research*, 48(1), 1–35. [\[CrossRef\]](#)
- Aldousari, A., & Alsabah, M. (2025). ESG integration and technical efficiency: A comparative frontier analysis in kuwait financial sector. *Sustainability*, 17(22), 10231. [\[CrossRef\]](#)
- Alsabah, M., & Alshehabi, A. (2026). Beyond standards: How culture and institutions shape goodwill impairment under IFRS. *Journal of Financial Reporting and Accounting*. Forthcoming.
- Alshehabi, A., Georgiou, G., & Ala, A. (2021). Country-specific drivers of the value relevance of goodwill impairment losses. *Journal of International Accounting, Auditing and Taxation*, 43, 100384. [\[CrossRef\]](#)
- Alshehabi, A., Halabi, H., Adwan, S., & Boubaker, S. (2024). The impact of industry competition on the value relevance of goodwill impairments across different information environments. *Journal of International Accounting, Auditing and Taxation*, 56, 100639. [\[CrossRef\]](#)
- Alshehabi, A., Halabi, H., & Afrifa, G. (2023). The effect of time orientation in languages on the recognition of goodwill impairment losses. *The International Journal of Accounting*, 58, 2350011. [\[CrossRef\]](#)
- André, P., Filip, A., & Paugam, L. (2015). The effect of mandatory IFRS adoption on conditional conservatism in Europe. *Journal of Business Finance & Accounting*, 42(3–4), 482–514. [\[CrossRef\]](#)
- Bartlett, M. S. (1954). A note on the multiplying factors for various chi-square approximations. *Journal of the Royal Statistical Society, Series B*, 16, 296–298. [\[CrossRef\]](#)
- Brinkerhoff, D. B., White, L. K., & Ortega, S. T. (2007). *Essentials of sociology*. Wadsworth.
- Brown, T. A. (2012). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
- Burgstahler, D., Hail, L., & Leuz, C. (2006). The importance of laws, institutional factors, and enforcement for earnings management around the world. *Journal of Accounting Research*, 44(3), 511–565. [\[CrossRef\]](#)
- Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research*, 5, 3–52. [\[CrossRef\]](#)
- Bushman, R. M., & Piotroski, J. D. (2006). Financial reporting incentives for conservative accounting: The influence of legal and political institutions. *Journal of Accounting and Economics*, 42(1–2), 107–148. [\[CrossRef\]](#)
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, 56, 81–105. [\[CrossRef\]](#)
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage.
- Chen, F., Hope, O. K., Li, Q., & Wang, X. (2009). Financial reporting quality and investment efficiency of private firms in emerging markets. *The Accounting Review*, 84(4), 1125–1156. [\[CrossRef\]](#)
- Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, 41, 745–783. [\[CrossRef\]](#)
- Choi, F. D. S., & Meek, G. K. (2011). *International accounting* (7th ed.). Prentice Hall.
- Christensen, H. B., Hail, L., & Leuz, C. (2011). Capital-market effects of corporate disclosure regimes: Evidence from the European Union’s Market Abuse Directive. *Journal of Accounting Research*, 49(5), 1165–1203. [\[CrossRef\]](#)
- Churchill, G. A. J. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16, 64–73. [\[CrossRef\]](#)
- Collins, D., Jordan, C., & Coleman, H. (2012). *An introduction to family social work* (4th ed.). Brooks/Cole.
- Dignam, A. J., & Galanis, M. (2009). *The globalization of corporate governance*. Ashgate.

- Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2008). The law and economics of self-dealing. *Journal of Financial Economics*, 88, 430–465. [\[CrossRef\]](#)
- Everitt, B. S., & Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Springer.
- Fargher, N., & Zhang, L. (2024). Investor protection, aggregate changes in profit margins and forecasts of growth in GDP: International evidence. *Review of Accounting Studies*, 29, 3497–3533. [\[CrossRef\]](#)
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3, 456–465. [\[CrossRef\]](#)
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39–50. [\[CrossRef\]](#)
- Francis, J. R., & Wang, D. (2008). The joint effect of investor protection and big 4 audits on earnings quality around the world. *Contemporary Accounting Research*, 25(1), 157–191. [\[CrossRef\]](#)
- Glaum, M., Schmidt, P., Street, D. L., & Vogel, S. (2013). Compliance with IFRS 3- and IAS 36-required disclosures across 17 European countries: Company- and country-level determinants. *Accounting and Business Research*, 43, 163–204. [\[CrossRef\]](#)
- Goldberg, L. (2001). *A journey into accounting thought*. Routledge.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.
- Halabi, H., Alshehaby, A., & Zakaria, I. (2019). Informal institutions and managers' earnings management choices: Evidence from IFRS-adopting countries. *Journal of Contemporary Accounting & Economics*, 15, 100162. [\[CrossRef\]](#)
- Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, 2, 8. [\[CrossRef\]](#)
- Hope, O. K. (2003). Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study. *Journal of Accounting Research*, 41(2), 235–272. [\[CrossRef\]](#)
- Houque, M. N., van Zijl, T., Dunstan, K., & Karim, A. K. M. W. (2012). The effect of IFRS adoption and investor protection on earnings quality around the world. *The International Journal of Accounting*, 47, 333–355. [\[CrossRef\]](#)
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. *Structural Equation Modeling*, 6, 1–55. [\[CrossRef\]](#)
- Isidro, H., Nanda, D., & Wysocki, P. D. (2020). On the relation between financial reporting quality and country attributes: Research challenges and opportunities. *The Accounting Review*, 95, 279–314. [\[CrossRef\]](#)
- Kabbach de Castro, L. R., Martins, H. C., Schiehl, E., & Terra, P. R. S. (2023). Investment–cash flow sensitivity and investor protection. *Journal of Business Finance & Accounting*, 50, 1402–1438. [\[CrossRef\]](#)
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31–36. [\[CrossRef\]](#)
- Kaufmann, D., & Kraay, A. (2024). *The worldwide governance indicators: Methodology and 2024 update* (World Bank policy research working paper No. 10952). World Bank.
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The worldwide governance indicators: Methodology and analytical issues* (World Bank policy research working paper No. 5430). World Bank. [\[CrossRef\]](#)
- Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65, 2276–2284. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kollat, D. T., Engel, J. F., & Blackwell, R. D. (1970). Current problems in consumer behavior research. *Journal of Marketing Research*, 7, 327–332. [\[CrossRef\]](#)
- La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2006). What works in securities laws? *Journal of Finance*, 61, 1–32. [\[CrossRef\]](#)
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Legal determinants of external finance. *Journal of Finance*, 52, 1131–1150. [\[CrossRef\]](#)
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1998). Law and finance. *Journal of Political Economy*, 106, 1113–1155. [\[CrossRef\]](#) [\[PubMed\]](#)
- Leuz, C. (2010). Different approaches to corporate reporting regulation: How jurisdictions differ and why. *Accounting and Business Research*, 40, 229–256. [\[CrossRef\]](#)
- Leuz, C., Nanda, D., & Wysocki, P. D. (2003). Earnings management and investor protection: An international comparison. *Journal of Financial Economics*, 69, 505–527. [\[CrossRef\]](#)
- MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, 35, 293–334. [\[CrossRef\]](#)
- Marshall, G. (Ed.). (2009). *Oxford dictionary of sociology*. Oxford University Press.
- North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.
- Podsakoff, N. P., Podsakoff, P. M., MacKenzie, S. B., & Klinger, R. L. (2012). Are we really measuring what we say we are measuring? *Journal of Applied Psychology*, 97, 99.

- Quagli, A., Lagazio, C., & Ramassa, P. (2021). From enforcement to financial reporting controls (FRCs): A country-level composite indicator. *Journal of Management and Governance*, 25, 397–427. [\[CrossRef\]](#)
- Reverte, C. (2008). Do corporate governance structures affect the transparency of corporate financial reports? Evidence from Spain. *Corporate Governance: An International Review*, 16(1), 26–39. [\[CrossRef\]](#)
- Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, 25, 6–14. [\[CrossRef\]](#)
- Ruland, C. M., Bakken, S., & Røislien, J. (2007). Reliability and validity issues related to interactive tailored patient assessments: A case study. *Journal of Medical Internet Research*, 9, e22. [\[CrossRef\]](#) [\[PubMed\]](#)
- Scheuer, A. (2006). *How Europeans see Europe*. Amsterdam University Press.
- Schoenfeldt, L. F. (1984). Psychometric properties of organizational research instruments. In *Method and analysis in organizational research*. Reston Publishing.
- Schwab, D. P. (2005). *Research methods for organizational studies* (2nd ed.). Psychology Press.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. Wiley.
- van der Eijk, C. (2007). *Multiple-item measurement: Why factor analysis is often inappropriate for mapping latent structure and what to do instead*. Methods and Data Institute Working Paper. University of Nottingham.
- Wysocki, P. D. (2004). *Earnings management, tax compliance, and institutional factors* (Working Paper). Massachusetts Institute of Technology.
- Wysocki, P. D. (2011). New institutional accounting and IFRS. *Accounting and Business Research*, 41, 309–328. [\[CrossRef\]](#)
- Zysman, J. (1994). How institutions create historically rooted trajectories of growth. *Industrial and Corporate Change*, 3, 243–283. [\[CrossRef\]](#)

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