A Concept-driven Approach to Measurement:

The Lexical Scale

John Gerring
Department of Political Science
Boston University
jgerring@bu.edu

Svend-Erik Skaaning
Department of Political Science
Aarhus University
skaaning@ps.au.dk

Daniel Pemstein
Department of Political Science
North Dakota State University
daniel.pemstein@ndsu.edu

Comments welcome!

Draft: 10-21-14
ABSTRACT

This paper introduces a new method of scale construction dubbed the *lexical scale* (following Rawls 1971). This scale assumes the form of a series of necessary-and-sufficient conditions arrayed in an ordinal scale, generating a cumulative ordering of attributes. (Condition $A$ is necessary and sufficient for *Level 1*; conditions $A&B$ are necessary and sufficient for *Level 2*; and so forth.) The relative priority granted to each condition is a reflection of a priori conceptual and theoretical properties rather than the distribution of data. In this respect, the construction of a lexical scale is deterministic rather than probabilistic. After introducing the scaling procedure, we offer several examples of lexical indices that might be developed to measure prominent social science concepts. We proceed to contrast the lexical scale with other scaling methods, and conclude with an evaluation of the overall strengths and weaknesses of the proposed approach.
The theoretical burden of social science is carried by highly abstract concepts such as democracy, state, inequality, and rule of law. We require such concepts in order to articulate high-order theories. Yet, they are difficult to operationalize, even when agreement can be reached on a general definition for a term.

A key obstacle is aggregation. Faced with a number of indicators that seem relevant to a concept the researcher must decide how to combine them into a single index. Factor analysis, structural equation models, and item response theory (IRT) models are three generic approaches to this problem (DeVellis 2011). Other approaches, such as those employed by Polity IV (Marshall, Jaggers 2007) and Freedom House (2013) to measure facets of political regimes, are more idiosyncratic but perhaps more influential. Where attributes co-vary in a predictable fashion the problem of aggregation is easily solved. But where they do not, as is generally the case with multivalent concepts, researchers are at pains to solve the aggregation problem in a non-arbitrary fashion.

This paper introduces a novel approach to scale construction that builds on the properties of concepts. Any attempt at measurement must be oriented around a concept, the idea that is purportedly being measured. Indeed, problems of validity often stem from inattention to the core concept and a consequent lack of fit between the term, its definition, and the indicators employed to measure it (Adcock, Collier 2001; Goertz 2006; Saylor 2013). This much is generally recognized (even if not always achieved).

In this study, we look to concepts to do more than identify relevant indicators. Specifically, we enlist the structure of concepts as an alternative method of solving the aggregation problem. This is accomplished by regarding conceptual attributes as necessary-and-sufficient conditions arrayed in an ordinal scale. Following Rawls (1971), we refer to this as a lexical scale.

We begin by reviewing concept-led approaches to measurement, which provide the background for the present initiative. Next, we lay out the core properties of a lexical scale. In the third section, we offer several examples of how lexical scaling might be applied to social science concepts. The fourth section situates the lexical scale in the measurement literature, with special reference to Guttman scales and IRT. The paper concludes with general observations about the strengths and weaknesses of the lexical scale relative to other approaches to scale construction.

A few notes on terminology will be helpful before we begin. Concept formation refers to the construction of a concept, including the choice of terms, defining characteristics, and referents. Defining properties of a concept refer to attributes that provide its formal definition. Associated properties are attributes that are thought to be associated with the defining properties, but not definitional. Measurement refers to concept operationalization, i.e., the instructions or instruments required to identify membership, or degrees of membership, in the extension of a concept. This involves the construction of an indicator or index (a group of indicators combined in some fashion). A scale is a generic type of indicator or index.

1. Concept-Driven Approaches

In choosing how to operationalize a term social scientists often rely on the empirical properties of things “out there” to construct an index. For example, the measurement model may combine variables by granting equal weight to each. Alternatively, the measurement model may employ factor analysis, IRT, or structural equations in order to combine information from a vector of indicators. In any case, the researcher seeks to understand latent qualities of a concept from observable qualities
contained in extant indicators. In this sense, concept formation is subordinated to measurement, as Sartori (1970: 1038) lamented some time ago (see also Borsboom et al. 2004; Saylor 2013).

Concept-led approaches to measurement are comparatively underdeveloped. With the exception of a small body of work on concept formation and typologies (Bailey 1994; Collier, LaPorte, Seawright 2012; Collier, Gerring 2009; Elman 2005; Gerring 2012: ch 5; Goertz 2006; Sartori 1984), and a few studies focused explicitly on the connection between conceptualization and measurement (e.g., Adcock, Collier 2001; Munck 2009; Saylor 2013), no literature addresses the question of how one might construct a scale based on the properties of a concept. Despite efforts to overcome the qualitative/quantitative gap (Brady, Collier 2010), the twin tasks of conceptualization and measurement still seem far apart.

Insofar as a tradition of concept-driven measurement exists it is contained in the binary scale, where a crisp-set concept is operationalized as a matter of membership (in/out). Commonly, membership criteria consist of one or more necessary conditions, jointly understood as necessary-and-sufficient. Occasionally, sufficient conditions are invoked (Goertz 2006). In either case, binary scales are generally constructed with a view to represent ordinary meanings and/or important theoretical properties of a concept. This is central to the “classical” tradition of concept formation (Collier, Gerring 2009; Sartori 1984) and to set-theoretic approaches to social science (Goertz 2006; Goertz, Mahoney 2012; Schneider, Wagemann 2012). It is also implicit in experimental and quasi-experimental studies, where treatments are usually understood in a binary fashion and are derived from a priori research hypotheses (Shadish, Cook, Campbell 2002).

As an example, let us consider the influential binary index of democracy known as Democracy-Dictatorship (DD). According to Przeworski and colleagues (Przeworski, Alvarez, Cheibub, Limongi 2000; Cheibub, Gandhi, Vreeland 2010: 69), a regime is a democracy if leaders are selected through contested elections. To operationalize this conception of democracy the authors identify four criteria:

1. The chief executive must be chosen by popular election or by a body that was itself popularly elected.
2. The legislature must be popularly elected.
3. There must be more than one party competing in the elections.
4. An alternation in power under electoral rules identical to the ones that brought the incumbent to office must have taken place.

Like many binary scales, the DD index adopts a minimal definition of democracy and operationalizes it with one or more necessary conditions (in this case, four), all of which must be satisfied in order to receive a score of 1 (=democracy).

The main complaint about binary scales when imposed on complex concepts is that they reduce all aspects of that concept to two categories, converting a plethora of information into a series of 0’s and 1’s (Elkins 2000). Of course, this procedure is perfectly reasonable (a) if the associated properties of a concept are highly correlated with the binary division or (b) if the associated properties are randomly distributed across the two groups. However, with non-experimental data the likelihood of either (a) or (b) is slight. Associated properties of democracy are not likely to be correlated perfectly with the binary scale; nor are they likely to be distributed in a random fashion. Likely as not, they will be somewhere in between. This means that the resulting scale is difficult to interpret. It is neither a proxy for democracy at-large nor a uniform treatment,
and is apt to be associated with confounders, i.e., other properties of democracy that do not align with the binary index.¹

Binary scales play a vital role in social science and also constitute an important link to the classical tradition of concept formation. However, they cannot carry all the freight that is sometimes assigned to them.

II. A Lexical Scale

We propose to preserve the virtues of a conceptually driven approach to measurement while honoring the need for greater differentiation than is provided by binary scales. This, in brief, is the strategy of the lexical scale, which incorporates necessary-and-sufficient conditions as distinct levels of an ordinal scale. While the binary scale treats all conditions as necessary for establishing membership in a concept’s extension, the lexical scale enlists conditions to establish levels of membership in that concept. In this sense, it is fair to regard the lexical scale as an extension of the classical binary scale.

Prior to constructing a lexical scale it is important to conduct a thorough survey of potential attributes attached to the concept of theoretical interest. A concept-led approach to measurement must take seriously the task of definition, for it is this task that sets the framework for scale construction. In order to make sure that all possible attributes for Concept X are considered, and none arbitrarily excluded, one is well-advised to survey definitions and usage patterns of a concept in ordinary language and in whatever specialized language region may be relevant to the research. The initial culling of attributes should be as comprehensive as possible, excluding only idiosyncratic features.²

Next, one must arrange these attributes so that each serves as a necessary-and-sufficient condition within an ordered scale. That is, each successive level is comprised of an additional condition, which defines the scale in a cumulative fashion. Condition A is necessary and sufficient for L₁; conditions A&B are necessary and sufficient for L₂; and so forth, as illustrated in Table 1. If there are five levels to an index, five necessary conditions must be satisfied in order to justify a score of 5. This means that each level in a lexical scale is defined by a set of conditions that are both necessary and sufficient, fulfilling a goal of the classical concept. Note that the structure of a lexical scale presupposes that there is a true zero, representing phenomena that do not meet the first condition (~A).

¹ Consider the DD index, as defined above, and consider an attribute that is probably associated with its defining features such as executive constraints (EC). Let us suppose that EC can be conceptualized in a binary fashion, and let us suppose that it correlates partially with the DD index: countries that are scored 1 on the DD index are more likely to score 1 on a binary index of EC, but the association is not perfect. It follows that any causal analysis that places DD on the right side of the model runs into a problem of omitted variable bias if EC is unmeasured and a problem of collinearity (and possible endogeneity) if it is included in the model. (We are presuming that EC may be affected by DD, but is not entirely the product of DD.) Thus, even when binary coding is transparent and consistent with a common reading of the concept, as with DD, the resulting index may engender confusion.

² Examples of this sort of semantic surveying can be found in Collier, Gerring (2009) and Sartori (1984).
In achieving these desiderata four criteria must be satisfied: (1) binary values for each condition, (2) unidimensionality, (3) qualitative differences, and (4) centrality or dependence. These may be briefly reviewed.

First, each level in the scale must be measurable in a binary fashion without recourse to arbitrary distinctions. It is either satisfied or it is not. (The construction of a binary condition may be the product of a set of necessary and/or sufficient conditions. Collectively, however, these conditions must be regarded as necessary and sufficient.)

Second, levels in a lexical scale must be understood as elements of a single (probably unobserved) concept. Empirical multidimensionality may persist, as discussed below. However, conceptual multidimensionality must be eliminated, either by dropping the offending attribute and/or by re-defining the concept in a clearer and perhaps more restrictive fashion.

Third, each level must demarcate a distinct step or threshold in a concept, not simply a matter of degrees. Levels in a lexical concept identify qualitative differences. A “3” on a lexical scale is not simply a midway station between “2” and “4.” Indeed, each level may be viewed as a subtype of the larger concept. Note that these sub-types are defined by cumulative combinations of the attributes possessed by the full concept – $A$, $A&B$, $A&B&C$, and so forth – fulfilling the criterion of a classical concept.

The most challenging aspect of lexical scale construction is the ordering of attributes, which follows a conceptual (rather than empirical) logic. One attribute may be considered prior to another if it is more central to the concept of theoretical interest (from some theoretical vantage point). This follows a constitutive approach to measurement, where attributes are the defining elements of a concept (Goertz 2006: 15). Alternatively, one attribute may be considered prior if it is a logical, functional, or causal pre-requisite of another. The dependence of $B$ on $A$ is what mandates that $A$ assume a lower level on a scale.

Whether responding to considerations of centrality or dependence, the levels of a lexical scale bear an asymmetric relationship to each other; some are more central (to the concept) than others. This is the most distinctive feature of a lexical scale, clearly differentiating it from Guttman scales (discussed below). While traditional measurement models usually consider attributes in an independent fashion, the lexical scale presumes that attributes can be understood only according to their inter-relationship with one another. A concept is thus defined by specific configurations of attributes.

Where lexical ordering is unclear a priori (according to considerations of centrality and dependence) one may wish to consider the shape of the empirical universe. Specifically, if $A$ is always (or almost always) present where $B$ is present there may be grounds for considering $A$ as more central or more fundamental than $B$. However, any conclusions reached on the basis of an exploration of empirical properties must be theoretically justified as a matter of centrality or dependence. The relative prevalence of attributes offers a clue to asymmetric relationships among the properties of a concept; it is not, itself, a desideratum. In constructing a lexical scale, deductive considerations trump data distributions.

In principle, there is no limit to the number of levels in a lexical scale. In practice, we anticipate that the number of levels will rarely surpass ten. This is because some concepts do not have a great many (truly distinct) attributes and because, where a great many attributes are present, they are unlikely to satisfy the criteria required for a lexical scale, as specified above.

---

3 This echoes the approach of qualitative comparative analysis (QCA) to empirical relationships more generally (Schneider, Wagemann 2012).
Deliberation and Disagreement

A product of deliberation, the lexical scale follows the procedure by which John Rawls (1971) orders the three core principles that establish his theory of justice – (1) the Liberty principle, (2) the Fair Equality of Opportunity principle, and (3) the Difference principle. These are arranged in order of lexical (short for lexicographical) priority. That is, one should not consider 2 or 3 until 1 has been fully satisfied, nor 3 until 1 and 2 are fully satisfied. Thus, each principle serves as a necessary condition of the next, creating a lexical scale with four levels. The force of this argument hinges, in large part, on a conceptual argument – that the core meaning of justice is reflected in this particular ordering of attributes, with category (1) understood as the most basic or essential (Moldau 1992).

Of course, Rawls was interested in defining the terms by which institutions within a society could be established and justified. He was not interested in measuring the presence/absence or degrees of justice in a society, and it is not clear whether he would have applied the same rules to such a measurement instrument. Even so, his approach is remarkably similar to that which we envision for empirical concepts in the social sciences. The proposed lexical scale depends upon reaching a reflective equilibrium with respect to the defining attributes of a concept and their lexical ordering. Insofar as this solution is persuasive, the scale will be useful. Insofar as it strains the meaning of a concept or theory it will seem arbitrary and forced, and is on that account unlikely to perform any useful function in social science. A lexical scale must resonate with everyday usage of a word as well as with considered judgments about what a concept should mean in a given theoretical context.

Just as there are disagreements over the meaning of justice, so there will be disagreements over how to scale empirical concepts (concepts whose purpose is to capture something specific and measurable “out there”). These may be placed into three categories.

First, we must consider disagreements over the definition of a term. Evidently, many high-order concepts in the social-science lexicon are contested (Collier, Hidalgo, Maciuceanu 2006), and this sort of contestation necessarily affects a concept-driven approach to measurement (though it also affects any other approach to measurement). In particular, it may affect the attributes included and excluded as conditions in a scale.

Yet, because the lexical scale is developed by reference to a concept’s core meaning we anticipate that disagreements over attribute inclusion are likely to affect positions on the periphery of the scale. As such, they will impact only those cases whose value is determined by features at the high end of the scale. For example, if two 7-point scales for the same concept differ in the chosen attributes these differences are most likely to be located at levels 6 and 7 and least likely to be located at levels 1 and 2. As such, only cases with the highest scores, i.e., those whose score is affected by the 6th or 7th conditions, will be affected. It follows that measurement error arising from errors in the hierarchy are more likely at the high end of the scale than at the low end.

The second sort of disagreement concerns the number of levels assigned to a lexical scale. Evidently, scholars working on the same concept may produce scales of differing lengths. Indeed, the decision about when to aggregate and when to disaggregate attributes (to form conditions) is somewhat arbitrary, hinging on contextual matters such as the sort of data that is at-hand and the use envisioned for the scale. For example, a scale developed for use on the right side of a causal model ($X$) is likely to be more concise than a scale developed for use on the left side of a causal

---

4 On the concept of reflective equilibrium see Daniels (1996).
model \((Y)\), as discrimination is vital in measuring outcomes whereas in constructing independent variables one usually strives to limit the number of treatments.

In any case, disagreement over scale length is not critical, as differently-sized scales will co-vary so long as the identifiable elements are ordered in the same fashion. Compare two hypothetical scales: \(I (A-B-C)\) and \(II (A_1-A_2-B_1-B_2-C_1-C_2)\), where the latter disaggregates each element of the former into two components. These alternate scales for the same concept are different insofar as one has more levels than the other; but they are not in conflict with each other.

A third, more damaging, sort of disagreement concerns the *lexical priority* of different elements. If researchers cannot agree on how to prioritize the attributes of a concept there is little hope of arriving at a useful lexical scale – or, to put the matter differently, each scale will be useful only in a very narrow context and may seem idiosyncratic. This sort of basic-level disagreement is encountered whenever the attributes of a concept bear no apparent relationship – functional or logical – to each other or where multiple attributes are judged equally important to the core concept.

While there is no simple solution to this situation, one strategy is to redefine the boundaries of the concept in a narrower fashion so as to exclude elements that cannot easily be integrated. This may be understood as a shift from a background concept to a systematized concept (Adcock, Collier 2001), and causes no damage so long as the re-definition is plausible since it does not strain the meaning of the core concept. Such a redefinition can be communicated by a compound noun that makes clear how the narrower concept relates to the parent concept. Accordingly, we eschew democracy in favor of a diminished subtype, electoral democracy, in one of the examples discussed below.

### III. Examples

As with most things, it is easier to grasp the workings of a method when specific examples are brought into view. In this section we provide a cursory exploration of possible lexical indices for the following well-known concepts: (1) civil liberty, (2) party strength, (3) rule of law, and (4) electoral democracy. Each discussion begins with a brief definition, clarifying how we understand the concept. This is followed by a proposed index, following the principles of lexical scaling. While the discussion is terse we hope that it fulfills a heuristic function, i.e., showing how the lexical approach to measurement might be applied to a range of key social science concepts.

*Civil liberty* is a human right, as well as a key component of democracy. Here, we shall understand civil liberty as a property of a government (including parties, civil society groups, and paramilitary groups that are closely associated with that government). Thus, we intend to measure the extent to which governments respect civil liberties *not* the extent to which civil liberties exist in a society (which may depend on things other than government action). Another important caveat is that we are concerned with the actions of government relative to the citizens of a polity. Its actions towards non-citizens (foreign nationals) lie outside the boundaries of our concept. With these qualifications, the following lexical index of civil liberty may be considered:

---

5. For further examples of concept operationalizations that seem to follow the logic of a lexical scale one might consider constitutionalism (Nino 1998: 3-4), human security (Tadjbakhsh & Chenoy 2007: Ch.2), peasants (Kurtz 2000: 96), and liberal democracy (Howard, Roessler 2006; Møller, Skaaning 2013).

6. Of course, one might argue that it is the responsibility of governments to protect civil liberties, regardless of who is infringing upon them. Nonetheless, it seems important to distinguish between the actions of governmental and non-governmental actors.
1. Political and extra-judicial murder.

2. No political and extra-judicial murder. The government does not organize or condone arbitrary killings or the killing of dissidents or of citizens based on their ascriptive characteristics (e.g., ethnic minorities).

3. No torture. The government does not organize or condone the torture of dissidents or of citizens based on their ascriptive characteristics (e.g., ethnic minorities).

4. Due process. The government does not arbitrarily arrest, imprison, or harass its citizens.

5. Free movement. The government does not restrict movement and residence within the polity.

6. Free discussion. The government does not restrict discussion in private arenas (among family, friends).

7. Free public speech. The government does not restrict speech in public arenas including the media.

8. Free association. The government does not restrict association, including political parties, labor unions, religious organizations, and other civil society organizations.

The ordering of attributes follows the logic of centrality. Specifically, we suppose that freedom from murder should be prioritized over freedom from torture, freedom from torture over due process, and so forth. Insofar as civil liberty guarantees basic liberties, those liberties that are more fundamental should be granted priority over liberties that are less fundamental.

Political parties may be defined minimally as organizations that nominate officials for public office, a key function in most theories of representative democracy (Schumpeter 1950). Within this context, the relative strength of these organizations may be regarded as an important component of democracy and good governance (Hicken 2009; Mainwaring, Scully 1995; Ranney 1962; Schattschneider 1942). (The strength of parties within an authoritarian context may also be important, e.g., for regime stability. However, measuring party strength in this context would require a different sort of scale since parties function quite differently in authoritarian contexts.) Party strength is understood here as the mean strength of all political parties that gain entrance into the legislature, and is thus differentiated from party system strength (the durability of a set of parties within a polity). With these clarifications, we propose the following lexical index of party strength:

1. Not allowed. Parties are not allowed to organize.

2. Allowed. Parties are allowed to organize. If the system is minimally democratic, the state may restrict entry to small parties judged to be hostile to democratic principles.

3. Independence. Parties are independent of the state (e.g., the bureaucracy, the military) and independent of each other (though naturally members of a coalition will be to some extent constrained by coalition agreements).

4. Defections rare. Party officials rarely leave their party voluntarily (to join another party or to continue their political career as an independent). Expulsions and retirements are not counted as defections.

5. Legislative cohesion. Members of a party usually vote together in the legislature.

7 The index does not include a consideration of party nationalization. In our view, parties may be strong while also being rooted in particular regions, as in the United Kingdom.
6. Centralization. Parties do not have strong factions or regional strongholds with distinct organizational structures; important decisions over policy and candidate selection are made at the center, or can be overturned by central party leadership.

7. Programmatic. Parties publically embrace policies and ideologies that are relatively distinct.

In prioritizing these attributes we rely, first of all, on considerations of dependence. The character of political parties cannot be considered unless and until parties are allowed to exist. Likewise, one can scarcely imagine a situation in which a party is allied with the state (not independent) and where defections from the party are tolerated. We rely, secondly, on considerations of centrality. For example, we regard a party’s independence from the state as more central to the concept of party strength than its level of centralization or programmatic orientation. Likewise, defections (members who leave one party and join another) from a party are judged more consequential than cohesion (voting behavior). The programmatic nature of a party is regarded as the least central element of our proposed index. This is consistent with the view that parties need not be ideological in order to be regarded as strong, and ideological parties are not necessarily strong. The concepts of party strength and party ideology should not be conflated. That said, a party possessing all other attributes of party strength that also has a clearly differentiated ideology should be considered stronger than a party whose philosophy and issue-positions are indistinguishable from others in the same party system.

The rule of law is a virtually universal political ideal which has in recent decades been identified as crucial for economic and human development (Tamanaha 2004: 1-4). Among the varying definitions of this concept, most of the attributes may be understood along a continuum of “thin/thick” conceptions (Bedner 2010; Møller, Skaaning 2012; Tamanaha 2004; Trebilcock, Daniels 2008: 12-13). Incorporating these various properties into a single graded scale, we suggest a lexical index of rule of law:

1. No rule by law.
2. Rule by law. Law is used as instrument for government action.
3. Formal legality. Laws are general, clear, prospective, certain, and consistently applied.
4. Institutional checks. An institutionalized system of government characterized by checks and balances, including an independent judiciary and penalties for misconduct.
5. Civil liberties. Liberal (negative) rights in the form of physical integrity rights and First Amendment-type rights are safeguarded.
6. Democratic consent determines laws. The citizens, through their elected representatives, are the ultimate source of laws.

In arriving at an ordering of these attributes we are once again cognizant of relationships of dependence. Attributes 2-5 are logically impossible to implement if law is not a principal instrument of governmental action. Institutional checks are generally inoperable without a system of formal legality. Civil liberties cannot be instituted unless there is a system of formal legality and institutional checks, including an independent judiciary.

The final condition, democratic consent, is not a product of dependence (it does not depend upon the others, at least not in any strict sense). However, it is an element of rule insofar as the fullest, most complete realization of the principle cannot be achieved if the citizens of a state are not sovereign (Habermas 1996). Any time an unelected individual or group of individuals are capable of altering the law in fundamental ways without recourse to democratic approbation the rule-of-law ideal is violated. Likewise, in a situation where this is possible the structure of law is, by definition,
ad hoc and unpredictable since it is privy to the whims of whoever happens to be serving as head of state.

Electoral democracy refers to the idea that democracy is achieved through competition among leadership groups which vie for the electorate’s approval during periodic elections before a broad electorate. This narrow – but nonetheless extremely influential – conception may also be referred to as a competitive, elite, minimal, realist, or Schumpeterian model of democracy, or as polyarchy (Dahl 1971; Przeworski, Alvarez, Cheibub, Limongi 2000; Schumpeter 1950). In order to operationalize this concept, we propose the following lexical index:

1. No elections. Elections are not held for any national-level policymaking offices. This includes situations in which elections are postponed indefinitely or the constitutional timing of elections is violated in a more than marginal fashion.

2. No-party or one-party elections. There are regular national elections but they are non-partisan or only a single party or party grouping is allowed to participate.

3. Multi-party elections for legislature. Opposition parties are allowed to participate in legislative elections and to take office.

4. Multi-party elections for executive. The executive is chosen directly or indirectly (by an elected legislature) through multi-party elections.

5. Minimally competitive elections. The chief executive offices and the seats in the effective legislative body are – directly or indirectly – filled by elections characterized by uncertainty, meaning that the elections are, in principle, sufficiently free to enable the opposition to win government power.

6. Male or female suffrage. Virtually all adult male or female citizens are allowed to vote in national elections.

7. Universal suffrage. Virtually all adult citizens are allowed to vote in national elections.

This ordering of attributes is sensitive to relationships of dependence. Evidently, the existence of elections (L1) is a pre-condition for attributes that describe the quality and purview of elections. Likewise, the existence of multi-party elections is a pre-condition for elections that are minimally competitive. In other respects, the ordering of attributes rests on considerations of centrality. We regard legislative elections as more central to the concept of electoral democracy than executive elections. (Accordingly, a country that has elections for the executive only would be regarded as less democratic than a country with elections for the legislature only.) Likewise, the competitiveness of elections is judged more central than the extent of suffrage. (Accordingly, we regard a country with competitive elections and a narrow suffrage such as Britain in the nineteenth century as more democratic than a country with universal suffrage and non-competitive elections such as North Korea.)

In this brief and necessarily schematic discussion we hope to have demonstrated the potential applicability of a lexical approach to measurement for a broad range of social science concepts. (For a more extended treatment of the proposed lexical index of electoral democracy, including various applications, the reader may refer to Gerring, Skaaning, and Bartusevičius [2014].) Granted, the lexical approach may not be practicable for all concepts. And where it applies, the lexical scale constitutes just one of many possible approaches to measurement. Choices among these options are likely to hinge on the uses to which the concept is being put in a particular research context, as discussed below. We turn now to a discussion of how the lexical scale compares and contrasts with other measurement strategies.
IV. Situating the Lexical Scale

Within the literature on concept formation, the lexical scale may be viewed as an attempt to reconcile minimal (thin) and maximal (thick, ideal-type) strategies (Coppedge 1999; Gerring 2012: Ch. 5). Note that the first condition (or first several conditions) establishes a minimal definition while the last condition in the scale completes what might be viewed as an ideal-type concept. Granted, ideal-type definitions are often more expansive than those envisioned by the lexical scale, primarily because the requirements of an ideal-type are less restrictive (anything that coheres with the concept is admissible). Even so, the lexical scale serves as a bridge between minimal and maximal concepts.

The lexical scale is also closely linked to a long intellectual tradition focused on typologies and taxonomies (Bailey 1994; Collier et al. 2012; Elman 2005; Gerring 2012: Ch. 5; Lazarsfeld 1937; Lazarsfeld, Barton 1951). Note that each level of a lexical scale corresponds to a distinctive category or type and is defined by all attributes contained in the superordinate category plus one. Accordingly, the levels of a lexical scale may be represented in a tree diagram, as depicted in Figure 1.

[Figure 1 about here]

Within the literature on measurement the lexical scale is similar in structure to Guttman scaling (Coppedge, Reinicke 1990; Guttman 1950). Guttman scales are based on a series of observable binary attributes—typically survey items—that, for each case, are rank-ordered with respect to some unobservable latent trait. In its idealized form, observing a positive value for any item on a Guttman scale implies positive values for all lower ranked, or less “difficult,” items. For example, one might assume that students learn to count before they tackle addition, and learn to add before they master multiplication. Thus, one could consider a test containing three questions—one each testing counting, addition, and multiplication—a Guttman scale for a subset of mathematical ability (Abdi 2010); students who correctly answer the multiplication question would, by assumption, also provide correct answers to the other two questions. The original Guttman model is deterministic, and idealized Guttman scales are built using deductive reasoning. The lexical scale is also deterministic and deductive, and both scales are cumulative and unidimensional. Nonetheless, the cumulative relationships modeled by lexical and Guttman scales differ fundamentally.

To clarify this distinction, return to Table 1 and consider a case for which conditions A, C, D, and E—but not B—are met. Such a case would score as a 1 on the generic lexical scale depicted in Table 1 because B is a necessary condition for classification at ordinal level 2 and above. As a consequence of this structure, knowing a case's lexical score provides information only about those conditions which are necessary and jointly sufficient to achieve a score one above that recorded for the case. That is, knowing that a case scores 1 on the generic lexical scale in Table 1 tells us only that A is satisfied and B is not, but provides no clue about the values of conditions C-E. On the other hand, knowing the Guttman score for a case provides perfect information about which conditions that case satisfies.

Table 2 displays a hypothetical Guttman scale that mirrors Table 1 in terms of conditions (items) and scale levels. Note that, in contrast to Table 1, Table 2 has no missing cells, or undefined relationships between scale level and observable traits. Thus, a case scoring 1 on this generic scale must satisfy condition A, but not conditions B-E, a case scoring 2 would satisfy A and B but not C-E, and so on. Similarly, knowing that a case meets a particular condition on a Guttman scale allows one to infer that the case likewise meets all of the “easier” conditions belonging to the scale. For example, if we know that a case meets condition D on the generic scale in Table 2, we know that it must also meet conditions A-C. In contrast, when developing a lexical scale, the researcher need make no assumptions about empirical patterns of correspondence between attributes. Therefore,
when considering attributes that form a lexical scale, one cannot necessarily\(^8\) use observations of certain case condition values to make inferences about other attributes of that case.

Table 2 about here

Guttman scales are a form of generative model that predict observable characteristics based on an unobservable trait. The assumptions inherent in a Guttman scale are testable. In its idealized, deterministic, form, one can falsify a Guttman scale simply by finding a contradiction. For example, referring again to Table 2, a case sporting condition \(D\) but not \(B\) would contradict the assumptions in our generic Guttman scale. Because real-world conditions rarely match the strict conditions of an idealized Guttman scale, applied work typically applies Guttman scaling in a probabilistic manner. Generally, modern applications of Guttman scaling fall under the broad umbrella of IRT (Fox 2010; Johnson, Albert 1999; Lazarsfeld, Henry 1968; Lord 1980; Mokken 1971; Rasch 1960). These techniques relax the strict determinism inherent in idealized Guttman scales, allowing researchers to estimate latent traits based on a probabilistic understanding of the relationships between latent variables and observed characteristics. For example, Mokken scaling techniques (Cingranelli, Richards 1999; Mokken 1971; Sijtsma, Molenaar 2002; Sijtsma, Debets, Molenaar 1990; van Schuur 2003, 2011) and other IRT tools (e.g. Rasch 1960) allow researchers to test how well a collection of observed attributes fit the Guttman scale assumptions, while allowing for error in the generative process that maps latent traits to observables. In terms of the generic scale in Table 2, for example, knowing that a case meets condition \(D\) might tell us that it is likely—rather than certain—to meet conditions \(A - C\). In general, whenever one assumes that the Guttman assumptions hold in a probabilistic fashion, one can derive probability distributions for observable attributes as a function of latent traits. Given that certain assumptions are met,\(^9\) one can use these distributions to estimate latent traits from observable attributes, even when the empirical distribution of those attributes does not perfectly conform to the ordering of a pure Guttman scale. Thus, scales built using IRT are often inductively constructed;\(^10\) thus, one needn’t specify the degree of difficulty (rank order) of observable traits a priori when building such a scale.

When the relationship between a concept and observable attributes theoretically matches Guttman assumptions, at least probabilistically it makes sense to model that concept with a Guttman scale. Indeed, because Guttman and subsequent IRT models generate predictions about patterns of covariation between observable attributes, allowing one to test the plausibility of core scaling assumptions, they provide a powerful tool for modeling cumulative concepts and justifying scaling decisions. Yet, many concepts fall outside the scope of Guttman models. Moreover, it is important to note that the choice to adopt a particular scaling framework should be guided by theory, not by empirical patterns in the data. Crucially, while IRT techniques can be used to falsify the Guttman ordering constraints, they cannot confirm that a set of variables validly measure a concept of interest. In other words, the fact that a set of observed features covary across cases in a way that is consistent with a Guttman scale does not imply that the ordering assumption is consistent with what

---

8 That is, the fact that the attributes make up a lexical scale does not allow for such inferences. Other factors might allow an analyst to draw such conclusions.

9 Most important is the assumption of local independence: items are locally independent if they are independent from each other, conditional on the latent trait. In other words, the components that one aggregates into an IRT based scale should correlate with one another only through the influence of the latent trait.

10 Although IRT models in the Rasch (1960) tradition are generally deployed in a confirmatory manner and thus rely on both deductive and inductive components. In general, one can use IRT techniques to test how well deductively constructed scales fit Guttman-style assumptions or one can use IRT techniques inductively, examining empirical patterns to choose items to include in a scale.
the analyst wants to measure, or that latent variables inferred from observables using IRT methods will provide valid measures of particular theoretical concepts.

Lexical scales are appropriate for a class of concepts that do not fit into the Guttman framework. In pursuing this issue it may be helpful to compare concepts such as civil liberty, party strength, rule of law, and electoral democracy (presented above) with concepts that have been central to the development of Guttman/Mokken/IRT models in the fields of education and psychology such as intelligence and aptitude. Note that while the first class of concepts describes features of institutions, the second class of concepts describes features of individuals. (This is not the only aspect in which they differ, but it may be a critical one.) Four important differences across these two classes of concepts may be identified.

First, with individual-level concepts it is often possible to identify “outcome” indicators that measure a latent concept of interest. A subject’s performance on a test offers a good indication of their aptitude – if not of their overall intelligence at least of their knowledge in a subject area. With institutional concepts the outcome-based approach is often difficult to apply. Outcome measures of democracy might include the closeness of the vote between parties in an election or the frequency of turnover as the result an election (Gerring, Teorell, Zarecki 2014; Vanhanen 2000). While informative, such indicators obviously do not capture the entirety of the concept and can be misleading. (A large margin of victory, and infrequent turnover, may be a sign of citizen satisfaction rather than of authoritarian tendencies.) Thus, for institutional concepts the items on a scale are likely to consist of substantive attributes that define the concept of interest. In measuring electoral democracy, for example, an item might consist of the question “Are there competitive elections?” or “Is suffrage universal?” This is roughly equivalent to an aptitude test in geography that includes the question “Are you good at identifying place-names?” In this sense, lexical scales attempt to measure concepts that are not a good fit for latent variable modeling techniques.

Second, with individual-level concepts that can be measured with outcome indicators like answers to an aptitude test, it is fair to assume that all questions that relate in some way to the subject are relevant and none are more important than others, except insofar as they might serve a more useful function in discriminating between subjects, or tap different “levels of difficulty” on the underlying dimension. A question’s utility derives from its function within the overall schedule of items on a test, and one judges that utility by the extent to which the item improves the overall capacity of that test to discriminate among subjects. By contrast, when institutional qualities are measured by the attributes that define them, it follows that some items will be more important than others in defining the latent concept of interest. Consider the following two indicators of electoral democracy: “Are there competitive elections?” and “Are 18-21-year-olds allowed to vote?” Evidently, the first is more central to the concept (as normally understood) than the latter. Inductive methods of scale construction offer no way of sorting this out. (Granted, our lexical ordering of attributes results in an ordinal scale whose levels are probably not equidistant from each other. Nonetheless, the construction of the scale presumes that each level is qualitatively different and theoretically significant, guarding against the introduction of trivial conditions.)

Third, with individual-level concepts like intelligence it is often reasonable to construct a scale inductively by reference to the pervasiveness of different responses. This is because responses to items on a questionnaire often possess a cumulative quality. All subjects who answer a hard question correctly will also answer easier questions correctly. However, the empirical distribution of institutional concepts does not always follow this cumulative logic. For example, polities that are undemocratic may thwart the will of the people in different ways, meaning that they will not receive the same score on some questionnaire items even when they are equally authoritarian. IRT models
do not provide a means for identifying functional substitutes in the form of several individually sufficient conditions.

Fourth, within individual-level concepts like intelligence the meaning of items on a scale are independent of each other. A subject’s answer to question #4 does not affect the meaning of her answer to question #7. However, in a survey purporting to measure democracy the meaning of “Is there universal suffrage?” changes depending upon the answer to the question “Are there competitive elections?” Specifically, it is not clear that universal suffrage has much significance to electoral democracy unless and until there are minimally competitive elections, as discussed. This is not an empirical relationship. Indeed, many countries hold elections without competition. It is, rather, about the meaning of the attributes relative to the underlying concept. This issue of non-independence does not arise in most individual-level concepts. Accordingly, IRT models do not usually take into account relationships of logical necessity.11

Finally, it is important to stress that the validity of a lexical scale is independent of the empirical distribution of the attributes that make up that scale. This point stems from the fact that lexical scales imply no deterministic or probabilistic relationship between components. Thus, it is impossible to test the validity of a lexical scale using the empirical attribute distribution. Furthermore, this means that it is impossible to inductively construct a lexical scale from a set of observed attributes because no pattern of observations is more or less consistent with any particular lexical ordering of observed traits. On one hand, this is a weakness of the lexical scale, because researchers cannot justify their conceptual choices by appealing to the empirical record. On the other hand, this line of logic demonstrates that inductively constructed measures of concepts that are lexically structured may often be invalid, precisely because such concepts need not imply and particular relationship between observables. Fundamentally, lexical scales are not generative; they do not predict data based on latent traits.12 Rather, they describe a particular class of cumulative concept that is logically defined by a specific aggregation of sub-components.

Naturally, lexical and Guttman-like scales lead to different indices. Consider our proposed index of electoral democracy, introduced in the previous section. We stipulated that suffrage does not have much (if any) impact on the quality of electoral democracy unless and until prior conditions are satisfied, i.e., there are multiparty elections, the executive is electorally accountable, and elections are competitive. This sort of prioritization has a large impact on coding if one compares lexical and Mokken scales with the same general features. Because of the ubiquity of universal suffrage in the modern era, this criterion will be fundamental to any Mokken scale of electoral democracy. Thus, many countries will pass this minimal threshold, rendering a score that is, arguably, higher than a particular regime deserves. Similar incongruities affect other scaling procedures (e.g., factor analysis or IRT models) insofar as these indices rest on inductive features of the empirical landscape. (Naturally, theoretical priors may be introduced to these models, making them more deductive in character and closer in spirit to a lexical scale.)

To say that lexical and Guttman/IRT scaling procedures sometimes arrive at different indices for the same general concept is not to say that one is superior to the other. Both surely have

---

11 If they do, they are constructed in a highly deductive manner. Recall that our argument is not with technical methods of scale construction per se but rather with their use.

12 Technically, one could use knowledge of a lexical score to predict something about the distribution of observable attributes. So, knowing that a case scores 5 on the generic scale in Table 1 tells you that the case satisfies all of the conditions. But the logic of lexical scaling is not based on a data generating process that maps latent values into observed characteristics, or responses.
their uses. However, there are reasons to imagine that the solutions provided by Guttman/IRT models are less appropriate for some concepts than for others, as we have suggested. Our extended comparison indicates that, in some settings, the Guttman/IRT approach may result in problems of concept validity and, in these settings, a more concept-driven approach to measurement may be warranted.

V. Discussion

We shall now attempt to summarize our wide-ranging discussion pertaining to the strengths and weaknesses of the lexical scale. In principle, the recalcitrant aggregation problem is solved by treating defining attributes as necessary-and-sufficient conditions arrayed in an ordinal fashion.\(^{13}\) If the scale is true to its objectives, each level in the scale defines a stronger, more complete instantiation of the underlying concept. This is no mean feat, given that composite indices are often plagued by problems of aggregation (Goertz 2006; Munck 2009).

Because conceptualization is integrated into measurement there should, in principle, be less slippage between concept and indicator than is typically encountered with other methods of scale construction.\(^{14}\) Nonetheless, if the analyst drops attributes of a concept from an index – because they cannot be meaningfully arrayed in an ordinal scale – forces continuous phenomena into an arbitrary binary coding, or prioritizes conditions without some underlying rationale, the resulting index will depart from ordinary meanings implied by the concept. The lexical scale is by no means immune to problems of concept/construct validity.

Similarly, the strictures of the lexical scale are not universally applicable. They require that relevant attributes of a concept be coded in a binary fashion without undue distortion and that the chosen attributes be arrayed along a single dimension according to their centrality to the concept or relations of dependence. These are not easy requirements to satisfy.

We have noted that the deductive properties of a lexical scale require many judgments on the part of the analyst. Accordingly, different analysts may arrive at different scales for the same concept. This, by itself, does not differentiate the lexical scale from scales constructed in a more inductive fashion. After all, there are many moving parts to any scale, particularly when one is attempting to operationalize a highly abstract concept. One must choose an indicator or set indicators to represent a concept and, if more than one indicator is chosen, one must decide upon an aggregation technique(s) that combines those elements into a single scale. Accordingly, it is not the case that lexical scales are more “subjective” than other scales.

\(^{13}\) This presumes, of course, that each condition can be accurately measured in a binary fashion without too much loss of information.

\(^{14}\) Authors’ choices of indicators to include in an index are often somewhat arbitrary (Goertz 2006; Haig, Borsboom 2008; Munck 2009). For example, the World Development Indicators for “rule of law” (Kaufmann, Kraay, Mastruzzi 2007) primarily measures crime and property rights, downplaying or entirely excluding other attributes of the concept (Skaaning 2010). Likewise, the Freedom House Political Rights and Civil Liberties indices (a component of various latent-variable models of democracy) include indicators pertaining to corruption, civilian control of the police, the absence of widespread violent crime, willingness to grant political asylum, the right to buy and sell land, and the distribution of state enterprise profits (Freedom House 2013). Some observers might regard these features as elements of political rights and civil liberties; others might not. Since most abstract concepts can be defined in a variety of ways and do not possess sharp boundaries, it is no surprise to discover that one analyst’s bundle of indicators may be quite different from another’s, even when they purport to operationalize the same term.
Arguably, the assumptions employed in the construction of a lexical scale are more transparent than the assumptions used to construct many other composite indices, especially when a number of aggregation principles are embedded in a complex statistical model. On the other hand, because of the set-theoretic nature of the lexical scale it seems likely that alternative lexical scales for the same concept will be less highly correlated than varying inductive scales for the same concept. A small change in an ordinal scale generally has greater consequences than a small change in an interval scale.

Lexical scale construction is a highly deductive enterprise insofar as the resulting index is constructed to suit a priori requirements drawn from the concept rather than from the empirical distribution of the data. Yet, many concepts do not provide clear guidance with respect to the relative priority of their defining conditions, a limiting condition on the applicability of lexical scaling.

Where applicable, however, the deductive properties of the scaling procedure offer certain advantages. Note that insofar as the distribution of data is allowed to govern the construction of an index, the resulting variable is sample-dependent. If key properties of a sample change (e.g., when drawn from different populations or when drawn non-randomly from a single population), so does the resulting scale. Sometimes, sample bias can be corrected with an IRT model but this presumes considerable knowledge about the larger population of interest. In most circumstances (and especially where the population extends into the future), it is not possible to determine what the shape of a larger population looks like. In these situations, indices are biased – or, alternatively stated, they lack generalizability because they are sample-dependent. An IRT-based index of electoral democracy, for example, is likely to vary across sample periods for the very reason that the composition of regimes around the world (the basis for the index) has varied enormously over the past two centuries. A IRT-based index constructed with data for the nineteenth century will be different from a IRT-based index constructed on the basis of data for the twentieth century. Indeed, every time a new decade of data is added to a sample (assuming the sample is updated regularly), the resulting index of democracy could change. This sort of instability is problematic in a measurement model.

Relatedly, a basic (and nearly universal) operating assumption of standard index construction is that one can combine information from observed variables by paying attention to their commonalities and discarding their differences as error. This is a reasonable set of assumptions in many circumstances, especially when the commonalities are great and the remaining differences do not seem to represent anything of substantive significance, i.e., they do not compose an identifiable dimension. However, it involves a considerable simplification of reality, especially when co-variations are modest. In such circumstances, the lexical scale offers a viable alternative.

With respect to discrimination, the lexical scale may be counted as modestly successful. It provides much more information than the classical concept, understood as a binary scale. It is on par with many ordinal indices, which generally incorporate a handful of levels. It is also on par with indices that purport to be interval scales but, in reality, are probably better understood as ordinal such as the Polity and Freedom House indices of democracy (Armstrong 2011; Cheibub, Gandhi, Vreeland 2010; Pemstein, Meserve, Melton 2010; Treier, Jackman 2008). To be sure, a lexical scale will discriminate less successfully than a scale whose construction is geared to detect small differences (e.g., IRT models).

While sensitivity to small differences is valuable, it is not the only factor of importance in constructing a scale. Note that some concepts in the social science universe are probably lumpy rather than continuous. This appears to be the case with electoral democracy. One is at pains to
describe the difference between a regime with popular elections and one without (the first condition of our proposed Lexical index) as a matter of degrees. The same point might be made with reference to the other examples discussed above.

Likewise, where a concept is being formulated as a right-side variable in a causal model it may be helpful to recognize distinct treatments, understood as a cumulative series of compound treatments – $A$, $A&B$, $A&B&C$, et al. These can be tested with (a) pairwise comparisons and matching algorithms, (b) dummy variables in a regression model, (c) generalized additive models (Beck, Jackman 1998), or (d) Bayesian shrinkage models (Alvarez, Bailey, Katz 2011). If used to achieve covariate balance in a matching analysis a categorical variable is generally more tractable than a continuous variable. In these respects, lexical scales are well-suited for causal inference.

By contrast, inductively derived indices often function awkwardly on the right side of a causal model. A useful treatment is uniform, imposing the same condition on all those within the treatment group. However, indices usually include heterogeneous elements – a little bit of this and little bit of that, in portions that are difficult to account for. Typically, there are many ways to obtain a score of “3” along a continuous scale. Consequently, it is difficult to say what the treatment consists of, what causal mechanisms might be at work, and whether the resulting relationship should be interpreted as causal.

Composite scales generally indicate differences of degree, but not of kind. A “4” on the Polity2 scale indicates that a regime is more democratic than a country receiving a “2.” But it offers no additional information about the qualities of these regimes. In this respect, the information contained in a standard composite index is “quantitative” (more/less) rather than “qualitative” (differences of type). Accordingly, a point on a composite index rarely has an obvious interpretation or meaning except in terms of standard deviations from the mean, and thresholds used to convert a continuous scale into a nominal or ordinal scale are apt to be highly arbitrary. This makes it difficult to evaluate concept validity, even if aggregation rules are perfectly transparent. And it makes it difficult to apply concepts to real-world situations, detracting from social science’s relevance to politics and policy.

By contrast, a lexical scale is relatively transparent. Researchers and reviewers know exactly what a shift from “2” to “3” or “3” to “4” means because each level in the scale is achieved by only one additional criterion. This eases the burden of ex ante coding and ex post interpretation. Likewise, insofar as levels correspond to distinctive types, membership in each category of an ordinal scale is meaningful. Units coded as “3” share various characteristics, which may signal important theoretical properties (e.g., as inputs or outputs of a causal model). Qualitative differences are sometimes more informative than quantitative differences.

We are not proposing that a Lexical index has any claim to ontological priority over other sorts of indices, each of which represent certain aspects of reality and each of which has its uses. Sometimes, relationships are continuous (and hence best measured with an interval scale) and sometimes they have only one threshold (and hence best measured with a binary scale). By the same token, sometimes causal relationships are ordinal in character, or they require an ordinal scale to test

---

15 With respect to the Freedom House indices, Cheibub, Gandhi, and Vreeland (2010: 75) note: “for each of the ten categories in the political rights checklist and the 15 categories of the civil liberties checklist, coders assign ratings from zero to four and the points are added so that a country can obtain a maximum score of 40 in political rights and 60 in civil rights. With five alternatives for each of ten and 15 categories, there are $5^{10} = 9,765,625$ possible ways to obtain a sum of scores between zero and 40 in political rights, and $5^{15} = 30,517,578,125$ possible ways to obtain a sum of scores between zero and 60 in civil liberties. All of these possible combinations are then distilled into the two seven-point scales of political rights and civil liberties.”
various threshold possibilities. In these settings, which surely apply to many theories, a lexical scale – where ordinal levels represent qualitatively different categories – may be appropriate. In this fashion, we propose to add another tool to the measurer’s toolkit.
VI. References


Table 1:

**Generic Lexical scale**

<table>
<thead>
<tr>
<th></th>
<th>~A</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>~A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>~B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>B</td>
<td>~C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>~D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>~E</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td></td>
</tr>
</tbody>
</table>

0-5 = ordinal scale. **A-E** = conditions that are satisfied. **~A-E** = conditions that are not satisfied. Relationships are deterministic except where cells are undefined (empty).
Figure 1:
Lexical Scale in Tree-Diagram (Taxonomic) Format
Table 2: Generic Mokken Scale

<table>
<thead>
<tr>
<th></th>
<th>~A</th>
<th>~B</th>
<th>~C</th>
<th>~D</th>
<th>~E</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>~B</td>
<td>~C</td>
<td>~D</td>
<td>~E</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>B</td>
<td>~C</td>
<td>~D</td>
<td>~E</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>~D</td>
<td>~E</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>~E</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
</tbody>
</table>

0-5 = ordinal scale. $A-E$ = conditions that are satisfied. $\sim A-E$ = conditions that are not satisfied. Relationships are probabilistic.