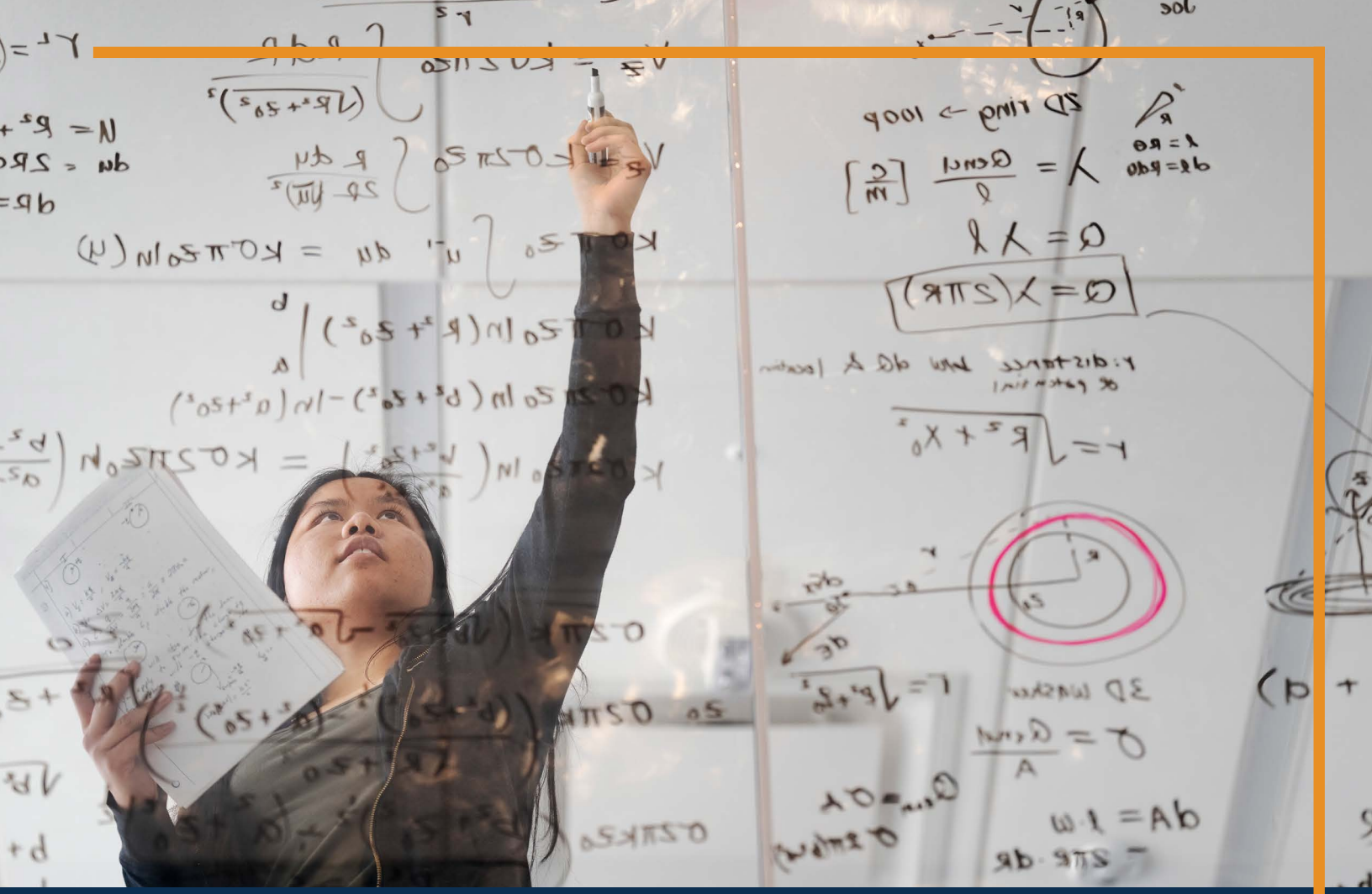


Organizational Classification: Approaches, Methods, and Applications to Higher Education Organizations

Kyle Whitman

ACE® American Council on Education®

Carnegie Foundation for the Advancement of Teaching

About the Carnegie Classifications White Papers

The Carnegie Foundation for the Advancement of Teaching and the American Council on Education (ACE) partnered in February 2022 to reimagine the future of the Carnegie Classifications. As part of this collaboration, the Carnegie Foundation for the Advancement of Teaching and ACE are working to develop new and refined versions of the classifications that better reflect the public purpose, mission, focus, and impact of higher education.

An aspect of this work involves learning from experts about key topics that can inform future methodological and data decisions. The Carnegie Classifications White Papers series aims to contribute to the body of knowledge and research about the impact of the historic Basic Classification, areas of consideration for a new Social and Economic Mobility Classification, and the role of classification systems. The analyses and takeaways from these papers provide guidance for potential updates. All released white papers can be found at **carnegieclassifications.acenet.edu**.

Reimagining the Carnegie Classifications is made possible by a cohort of funders that are dedicated to utilizing the classifications to help postsecondary education advance students' social and economic mobility through learner-centered outcomes. Partners include ECMC Foundation, the Bill & Melinda Gates Foundation, Imaginable Futures, the Kresge Foundation, Lumina Foundation, Mellon Foundation, the Alfred P. Sloan Foundation, and Strada Education Foundation, as well as a donor who wishes to remain anonymous.

American Council on Education
One Dupont Circle NW
Washington, DC 20036

Abstract

This paper describes and evaluates the approaches and methods available for classifying organizations. It begins by reviewing the history and fundamental assumptions of the essentialist, cognitive, and empirical approaches to classification, particularly within an organizational classification context. The paper then describes the analytical and statistical techniques available to create classifications, including definition-and-assignment, decision tree classifier, k-means cluster analysis, latent profile analysis, and multi-label classification. The paper includes a discussion of opportunities and limitations of applying these empirical approaches to the population of higher education institutions in the United States.

Introduction

Classification, in the broadest sense, occurs when an actor organizes objects from a large, heterogenous population into smaller, more homogenous groupings (Sneath and Sokal 1962). This activity is a fundamental cognitive process. The ability to meaningfully group together similar entities—such as animals or plants or clouds in the sky—helps actors reduce the complexity of the world, facilitate decision-making, and increase their likelihood of survival.

Classification aids in communication and the development of shared knowledge. It is important to note that while classification represents systematic knowledge, it is most productively understood as a prerequisite to the scientific process rather than an output of the scientific process. The scientific method requires the formulation of testable hypotheses on the relationships between observed phenomena (Lawson 2015). A classification, which allows investigators to specify and analyze a more homogenous grouping of observations than they otherwise would be able to, increases the ability of analytical methods to identify relationships among phenomena in the collected data. Classifications also assist other investigators in replicating results in other datasets.

Formally constructed classifications can be deductive or inductive in nature. Classifiers operating in a deductive mode often center existing theory and knowledge to specify the features used in the classification or the number of groupings present in an overall population. Classifiers operating in an inductive fashion center data and often use methods that allow for analysis without needing strong prior assumptions about what features should form the basis of groupings or how many groupings may be present in the data. These modes of operation exist on a spectrum, with classifications incorporating existing knowledge in various ways and degrees.

Classification is a complex process that involves iterating through a web of questions and considerations. These include specifying the purpose of the classification, understanding the relationship between the classification and the entities that are classified, defining what similarity and difference mean within and between groupings, and identifying how attributes will be selected for inclusion in the classification. These considerations also include specifying the overall population and defining individual entities—sometimes referred to as Operational Taxonomic Units. Enduring and useful classifications often have internally consistent and thoughtful positions on these issues.

As the American Council on Education reconceptualizes the Basic Classification of the Carnegie Classification of Institutions of Higher Education and creates a new Social and Economic Mobility Classification, it may be useful to examine how various intellectual traditions have approached the process of classification. To assist in this process, this paper will describe the origin, assumptions, and uses of essentialist, empirical, and nominalist approaches to classification. It will then describe some common analytical and statistical methods for creating classifications and how these relate to the various approaches to classification.

Approaches to Classification

Essentialism

The essentialist approach to classification is definition-focused and stems from an Aristotelian logic that all entities possess essential, unchangeable characteristics that exist independent of human observation. The purpose of an essentialist classification is to define these immutable characteristics and to assign entities into groups based on the presence of these characteristics (Rehder 2007). In this way, essentialist classifications create perfectly homogenous groupings but only with respect to the identified properties. Since immutable properties are unchanging, essentialist classifications are static; entities have little to no agency within them. Categories do not change over time, and entities cannot change their category membership.

Essentialist classifications were not useful in building systematic knowledge in the natural sciences. For example, an essentialist approach to a classification of trees might involve the specification and definition of "Maple-ness." This concept might involve references to leaf shape, trunk proportion, and tree sap chemistry. Any tree that possessed "Maple-ness" would be classified as a maple tree. This circular logic, which rests on a "common sense" understanding of entities, is a critical weakness of essentialist approaches. The creators of essentialist classifications often misinterpret what features are unique and differentiable among entities (McKelvey 1982) and end up producing something of little analytical or practical value.

Essentialist approaches persist in the social sciences. The concept of sector is an example of an essentialist classification. There are definitions of public, private, and for-profit organizations based on organizational legal status, and organizations that fit those definitions are classified as either public, private, or for-profit organizations. Essentialist classifications are appealing in their simplicity, but the approach quickly encounters issues when definitions are either unavailable or contested or the variation that users are interested in is unrelated to the definitions used to create categories.

Empiricism

Empirical classification approaches reject the idea that existing theory or definitions should exclusively guide classification; they instead allow classification results to emerge from the systematic analysis of data. Empirical approaches include phenetic, phyletic, and economic approaches to empirical classification.

Biologists developed the phenetic approach to classify plants and animals. The name of the approach refers to phenotype, which is a term for an organism's physical expression within its environment. Since there is variation in the phenotype of entities in the same category or species, this approach assumes that many attributes of organisms are needed to adequately describe the phenotype as well as to understand what variation is normal within and across categories (Sneath and Sokal 1962). In this way, phenetic approaches use many observed variables rather than pre-existing definitions to form groupings and assign entities to them. Phenetic classifications produce polythetic groupings—entities in the same grouping share many but not necessarily all characteristics—that are static over time. Phenetic classifications do not advance explanations of how groupings are formed or related to each other.

Phenetic approaches have been used in the organizational sciences. Many scholars have observed that they are particularly well suited to capture the multidimensional nature of organizations in which no single attribute can be necessary or sufficient to ensure an organization is assigned to a particular grouping (McKelvey 1982). Researchers creating and using phenetic classifications of organizations have disagreed on the role of theory in their work; some argue that classifications should be unconstrained by existing theory (Rich 1992), and others argue that a purely inductive approach is not possible (Doty and Glick 1994). The current methodology of the Carnegie Basic Classification shares assumptions with this approach.

The phyletic—or evolutionary—approach to classification attempts to identify what categories presently exist and describe how current categories came to be. To do this, phyletic classifications posit some sort of evolutionary process occurring within a population of entities. Because of this, groupings are conceptualized as inherently dynamic and containing a diverse range of entities.

Although the phyletic approach is currently the dominant approach to the classification of plants and animals in the field of life sciences, it has found few productive applications in the social or organizational sciences. The complicated nature and diverse life-courses of organizations—which are significantly affected at the individual and group level by broad social and technological trends; regulatory and governance systems; personalities of employees and founders; serendipity; and other factors—make the theorization of a "speciation" process extremely difficult. One interesting example of a phyletic classification is Crow and Dabars's (2020) Fifth Wave theory, which describes the social and political processes that evolved five organizational designs in American higher education from colonial times to present day.

The field of economics has grappled with concepts of classification and categorization to the extent that they implicate the definition of markets, a central analytical concept in neoclassical economic theory. In contrast to contemporary biological classification approaches developed to classify individual animals into coherent groupings based on observable characteristics, the economic approach to classification uses market transactions to create and describe categories (Wang and Archer 2007). Because market transactions are used to derive groupings and categories, they are considered dynamic, diverse, and sometimes overlapping.

There are few examples of economic-based classifications of higher education institutions, yet it is easy to imagine how they might be created. For example, a study that examined undergraduate student application patterns across schools could describe the contours of the student application market. Alternatively, a study that examined the employment histories of faculty across multiple schools could construct segmentations of the faculty job market. Consistently overlapping schools in the analysis of student or faculty transactions in the examples above would form the basis for groupings. When considered together, these groupings would form the overall market as well as the classification of the market.

Nominalism

The nominalist approach differs from the essentialist and the empirical approaches by rejecting the idea that categories exist independent of human observation and that data can be used to objectively identify them (Simons 2013). Nominalism insists that categories are socially constructed, reflect the power and agency of actors, and are often used for strategic purposes. In this way, nominalist classifications are both individual and instrumental.

Many nominalist classifications of organizations are rooted in cognitive psychology and strategic management and have been developed in the context of explaining how strategists categorize their own or other organizations based on attributes (Porac and Thomas 1990). This activity often assists organizational leaders in navigating competitive environments (Zuckerman 1999; Cattani, Porac, and Thomas 2017).

Brint, Riddle, and Hanneman's (2006) study of school president "reference sets" is an example of a nominalist higher education classification. The authors surveyed college and university presidents and asked them to identify a set of peer institutions to their institutions. They then used nomination networks to create groupings of similar institutions.

TABLE 1: BROAD APPROACHES TO CLASSIFICATION

	ESSENTIALISM	EMPIRICISM			NOMINALISM	
		Phenetic	Evolutionary/ Phyletic	Economic	Micro Approach	Macro Approach
Disciplinary origin	Natural Philosophy	Biology	Biology	Economics	Psychology	Sociology
Purpose of classification	Define natural categories	Reveal natural categories	Reveal natural categories	Determine boundaries for analytic purposes	Create categories for strategic or other purposes	Reveal socially constructed categories
Source of classification relative to entities	Exogenous	Endogenous	Endogenous	Interactive, market transactions	Endogenous	Exogenous
Category dynamism	Static, existing in nature	Static, existing in nature	Dynamic, created through natural mechanisms	Emergent, created through market exchanges	Created by individuals for purposes	Created by groups for purposes
Category heterogeneity	Monothetic relative to characters	Polythetic	Polythetic	Polythetic	Polythetic	Polythetic
Source of entity variation	Inherent to entity	Unspecified	Various, but must be specified	Innovation	Dependent on use	
Agency of entity	None	Possible but likely constrained	Possible but likely constrained	Depends	High	Constrained
Change of category membership	Impossible	Possible but likely constrained	Possible but likely constrained	Possible	Possible	Possible
Source of attributes	Underlying theory	Structure/ morphology of org	Structure/ morphology of org	Market transactions	Organizational offerings	Organizational field
Use of characters / object attributes	Few, selective	Many, comprehensive	Many, comprehensive		Few, selective	
Hierarchical or flat	Hierarchical	Either	Hierarchical	Either	Either	Either
Utility in theoretical work	Limited/None	High	High	High	Limited but high practical/ strategic use	Classification itself is object of study
Example in higher education context	"Sector"	Carnegie Basic	Crow and Dabars's Fifth Wave Framework	College co-application patterns	Brint, Riddle, and Hanneman 2006	

Analytical Methods of Classifying Organizations

There are a wide range of analytical and statistical methods used to reduce diverse populations into more homogeneous groupings. These methods are associated with essentialist, empirical, and nominalist approaches. This section will provide an overview of the definition-and-assignment, decision tree, k-means, latent profile analysis, and multi-dimensional classification methods.

Definition-and-Assignment

The definition-and-assignment method is useful for essentialist classification and is likely the simplest to use. It involves using existing theory and knowledge to create definitions of groupings and assigning organizations to groupings based on them meeting those definitions. Although this method produces very precise groupings, the groupings are only precise relative to the definitions used, and the method is limited to what is already known about these organizations. Definition-and-assignment may be particularly useful for special-purpose classifications created for narrowly defined use case but is likely less useful for classifications where results are needed to explain or account for a wide range of factors that are outside of the narrow definitions used to create the classification.

TABLE 2: COMMON ANALYTICAL AND STATISTICAL METHODS FOR CLASSIFICATION

	Definition-and-Assignment	Decision Tree	K-means Cluster Analysis	Latent Profile Analysis / Mixture Modeling
Connection to classification approach	Essentialist	Essentialist or Nominalist	Empirical or Nominalist	Empirical or Nominalist
Method of grouping	Create definition and assign entities	Identify variables and sequentially sort entities based on variable values	Partition observations into optimal number of clusters	Specifies group membership as unobserved variable and uses mixture modeling to determine group memberships
Determination of number of clusters/groupings	Definition/theory	Number of variables and variable splits used	Mathematical criteria	Mathematical criteria
Precision of groupings	High, perfectly homogenous with respect to definition	High, perfectly homogenous with respect to variables	Groupings have variation but membership is binary	Groups have variation and membership is probabilistic
Use	Common	Common	Common	Uncommon
Ability for lay person to understand method	High	High/Medium	Low	Low
Examples	Sector of institution	Basic Classification of Carnegie Classification	Used in academic studies	Used in academic studies

Decision Tree/Classification Chart

The decision tree or classification chart method is useful in empirical phenetic or phyletic classifications. It involves identifying characteristics that are relevant for the identification of categories and then sequentially sorting entities into increasingly smaller groupings based on the presence of these properties. The specification of nodes (branching points) on the trees can be informed by a wide range of techniques, including using existing theory or sophisticated machine learning methods (Song and Lu 2015).

Decision trees are highly precise; they create groupings that are perfectly homogenous with respect to the variables used in the sorting. Although decision trees produce dendritic structures that visualize similarities between categories, they are not necessarily evolutionary. For a decision tree to be phyletic, the variables used must correspond to the features, properties, or characteristics that entities shared in common before they evolved into increasingly separate categories.

K-means

K-means clustering is a common method used in empirical phenetic classifications. The method uses patterns of variation across identified variables to partition observations into a mathematically optimal number of clusters (Likas, Vlassis, and Verbeek 2003). Researchers have used k-means in a wide range of domains, including machine learning, data science, social science, health and medicine, and natural science.

As an inductive approach, k-means cluster analysis allows for the analysis of multivariate datasets without strong prior theoretical expectations of classification outcomes, such as the number of groupings or the assignment of particular observations to particular groupings.

Since results are based on patterns of similarity rather than clear definitions or meeting thresholds on variables in a particular sequence, k-means produces diverse and imprecise groupings. In k-means, no variable value is necessary or sufficient to determine membership in a particular group. Depending on the use of the classification, this may be a strength or a weakness.

Latent Profile Analysis

Latent profile analysis (LPA) is also an inductive approach. It involves specifying group membership as an unobserved latent variable and then using mixture modeling on observed data to estimate the latent variable of group membership. Like k-means, LPA uses statistical criteria to determine the optimal number of groupings present in the data and the assignment of analyzed entities to those groupings (Masyn 2013).

Although LPA is a computationally intense method and difficult to explain to lay users, the method produces useful results that better approximate real-world observed phenomena. One way it accomplishes this is by allowing for various specifications of the relationship between observed variables within and between each latent profile (changing the variance-covariance matrix). LPA also produces probabilistic results that facilitate detailed analysis of unusual classification cases.

Multi-Label Classification

Although not a classification method in the technical sense, multi-label classification can be reasonably considered along with the methods described above as an empirical approach to classify organizations. Multi-label classification is a type of classification in which each entity is assigned multiple labels rather than a single label that corresponds to group membership (Tsoumakas and Katakis 2007). In multi-label classification, labels are non-exclusive. For example, a multi-label classification of human personality could assign the labels of introversion, intuition, thinking, and judging. A multi-label classification of a movie could assign the labels of action, crime, and science fiction. The labels used in a multi-label classification can be created by various classification methods, including the methods described above.

Synthesis and Outlook

Much classification work focuses on incremental improvements in statistical methods at the expense of engaging broader questions embedded in the analytical methods. In order to expand and deepen the conversation around organizational classification, particularly during the reimagination of the Carnegie Classification of Institutions of Higher Education, this paper began by reviewing three fundamental approaches to classification and describing the varying assumptions, perspectives, and goals that they have. By examining these broad ways of thinking about classifications, creators of organizational classifications can be mindful of what is either locked-in or foreclosed when they select an analytical or statistical method.

These approaches to classification have been presented as ideal types—in their pure, theoretical form. Classification is often a complex process, and many real-world classifications typically tweak, combine, or otherwise alter aspects of what is described here as is relevant to the entities they are classifying and the purpose of their classification. It is appropriate to be guided by pragmatism, as classification is complex work characterized by trade-offs.

None of the approaches or statistical methods for classification are inherently superior to the others, even in the narrow context of higher education organization classifications. Rather, some may be better suited to particular use cases or policy contexts. It is important to design an approach and select an analytical method that is congruent with the organizations being classified, the data available on them, and objectives of the project.

That said, the analysis in this white paper supports several conclusions and recommendations for classification work in the context of reimagining the Basic Classification and designing the Social and Economic Mobility Classification.

First, essentialism is highly unlikely to be a productive guiding approach. Although essentialist classifications and typologies have long existed throughout the field of higher education, they tend to be very limited in scope—labels such as land-grant universities or Historically Black Colleges and Universities can be thought of as essentialist—and ignore much of the interesting variation present in these organizations. The higher education institutional design literature is insufficiently developed to reliably provide more elaborate definitions of the essences of school types across the full range of institutions present in U.S. higher education. Relatedly, definition-and-assignment is unlikely to be a viable method to produce classification results. Obtaining and analyzing the detailed organizational data needed for a comprehensive essentialist classification on over 3,000 institutions would quickly prove impractical.

Second, a purely nominalist approach would be somewhat incompatible with the objectives of the Carnegie Classifications. Nominalism stresses that groupings are socially constructed. The Carnegie Classifications need to produce results that are useful to the field and are understood as representing something beyond the subjective judgment of project staff. Aspects of nominalism, however, are important to keep in mind as the project progresses. The perspective reminds us that colleges and universities exist in the social realm and that it is not possible to completely remove the human element from the process of classification. Instead, we should strive to be in continuous dialogue with ourselves, our colleagues, and the field about biases and assumptions that inform responses to design choices we confront.

Third, there are various viable analytical techniques available for the classification of higher education. Each method has strengths and weaknesses that speak to the need for a flexible approach. For example, the mathematics that power k-means and latent profile analysis may require schools to be pre-sorted into very broad categories, such as two-year schools and four-year schools, before analysis so that models converge to interpretable solutions. This pre-sorting cannot be done without the use of some amount of existing higher education theory and may introduce elements of essentialism.

Fourth, a very productive orientation for the project would be a pragmatic empirical approach. Such an approach would:

- Recognize that groupings of schools exist independently of subjective human experience and that analyzing school data can identify these groups
- Recognize that an amount of existing theory and knowledge would need to be incorporated into the classification order to guide the selection of variables, sort schools into analyzable segments, and/or interpret results
- Understand that design choices may be influenced by perspectives, training, and biases of project staff
- Understand the results of the classification will likely be interpreted not through the lens of the empirical data but rather by the perception of how they challenge or reinforce the power, status, and prestige structure of the field

References

Brint, Steven, Mark Riddle, and Robert A. Hanneman. 2006. "Reference Sets, Identities, and Aspirations in a Complex Organizational Field: The Case of American Four-Year Colleges and Universities." *Sociology of Education 79* (3): 229–252. https://doi.org/10.1177/003804070607900303.

Cattani, Gino, Joseph F. Porac, and Howard Thomas. 2017. "Categories and Competition." *Strategic Management Journal* 38 (1): 64–92. https://doi.org/10.1002/smj.2591.

Crow, Michael M., and William B. Dabars. 2020. The Fifth Wave: *The Evolution of American Higher Education*. Baltimore: Johns Hopkins University Press.

Doty, D. Harold, and William H. Glick. 1994. "Typologies as a Unique Form of Theory Building: Toward Improved Understanding and Modeling." *The Academy of Management Review* 19 (2): 230–251. https://doi.org/10.2307/258704.

Lawson, Anton E. 2015. "Hypothetico-Deductive Method." In *Encyclopedia of Science Education*, edited by Richard Gunstone, 471–472. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2150-0_260.

Likas, Aristidis, Nikos Vlassis, and Jakob J. Verbeek. 2003. "The Global K-means Clustering Algorithm." *Pattern Recognition* 36, no. 2 (February): 451–461. https://doi.org/10.1016/S0031-3203(02)00060-2.

Masyn, Katherine E. 2013. "Latent Class Analysis and Finite Mixture Modeling." In *The Oxford Handbook of Quantitative Methods, Volume 2: Statistical Analysis*, edited by Todd D. Little, 551–611. Oxford Library of Psychology. New York: Oxford University Press.

McKelvey, Bill. 1982. *Organizational Systematics: Taxonomy, Evolution, Classification*. Berkeley: University of California Press.

Porac, Joseph F., and Howard Thomas. 1990. "Taxonomic Mental Models in Competitor Definition." *The Academy of Management Review* 15 (2): 224–240. https://doi.org/10.2307/258155.

Rehder, Bob. 2007. "Essentialism as a Generative Theory of Classification." In *Causal Learning: Psychology, Philosophy, and Computation*, edited by Alison Gopnik and Laura Schulz, 190–207. New York: Oxford University Press.

Rich, Philip. 1992. "The Organizational Taxonomy: Definition and Design." *The Academy of Management Review* 17 (4): 758–781. https://doi.org/10.2307/258807.

Simons, Peter. 2013. "Vague Kinds and Biological Nominalism." *Metaphysica* 14, no. 2 (October): 275–282.

Sneath, P. H. A., and Robert R. Sokal. 1962. "Numerical Taxonomy." *Nature* 193 (4818): 855–860. https://doi.org/10.1038/193855a0.

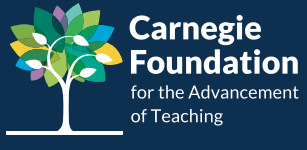
Song, Yan-yan, and Ying Lu. 2015. "Decision Tree Methods: Applications for Classification and Prediction." *Shanghai Archives of Psychiatry* 27, no. 2 (April): 130–135. https://doi.org/10.11919%2Fj.issn.1002-0829.215044.

Tsoumakas, Grigorios, and Ioannis Katakis. 2007. "Multi-Label Classification: An Overview." *International Journal of Data Warehousing and Mining* (IJDWM) 3 (3): 1–13. https://doi.org/10.4018/jdwm.2007070101.

Wang, Shan, and Norman P. Archer. 2007. "Electronic Marketplace Definition and Classification: Literature Review and Clarifications." *Enterprise Information Systems* 1 (1): 89–112. https://doi.org/10.1080/17517570601088380.

Zuckerman, Ezra W. 1999. "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." *American Journal of Sociology* 104 (5): 1398–1438. https://doi.org/10.1086/210178.


ACE®
American
Council on
Education®


Carnegie
Foundation
for the Advancement
of Teaching