A MACHINE LEARNING, ARTIFICIAL INTELLIGENCE APPROACH TO INSTITUTIONAL EFFECTIVENESS IN HIGHER EDUCATION

This page intentionally left blank

A MACHINE LEARNING, ARTIFICIAL INTELLIGENCE APPROACH TO INSTITUTIONAL EFFECTIVENESS IN HIGHER EDUCATION

BY

JOHN N. MOYE

Reprints and permissions service
Contact: permissions@emeraldinsight.com

ISOQAR certified
Management System,
awarded to Emerald
for adherence to
Environmental
standard
ISO 14001:2004.

ISOQAR REGISTERED

Certificate Number 1985
ISO 14001

INVESTOR IN PEOPLE

Contents

This page intentionally left blank

List of Tables

This page intentionally left blank

Author Biography

John Moye is a native of Jacksonville, FL, where he attended Jacksonville University. During his undergraduate and master's degree experience, he studied with a series of forward-looking thought leaders in education from which he developed an interest and belief in the science of learning and the power and importance of education for all learners. These interests have accompanied him throughout his career and led to a focus on the performance, effectiveness, and responsibilities of higher education.

Dr Moye continued his pursuit of the science of learning through his Ph.D. studies at Florida State University, where he focused his research in the field of psychophysics. Heavily impacted by the burgeoning field of neuroscience, he examined the response of the human perceptual systems to sensory stimuli as a model for understanding learning as a psychophysical process in individuals and organizations. The conceptual frameworks contained in this text are based on the evidence of the psychophysics of learning that are still emerging in the academic learning literature.

Dr. Moye has held effectiveness positions with numerous institutions of higher learning in the United States, including Saint Mary's University of Minnesota, Capella University, and De Paul University, Chicago, IL, in which he has researched, developed, and applied these approaches to the development of relevant, innovative, and effective learning environments. In addition, he has contributed to a wide array of other institutions of higher learning as a consultant, which has provided a comprehensive perspective on the science of measurement and assessment in complex organizations.

Dr Moye believes the leverage point in the system of institutional improvement to be the availability of authentic, credible, and trustworthy information to make sense of institutional performance for effectiveness improvement efforts. The research and development of systematic assessments to measure the effectiveness of unique institutions is a focus of his ongoing professional efforts.

This page intentionally left blank

Foreword

This work seeks to catalyze discussion and thinking about the information required to measure, assess, and make sense of institutional performance with credible and trustworthy data. To those who believe it is possible to improve the performance of our institutions this work offers a method to improve service to students and society through data-informed problem-solving and decision-making.

To achieve this outcome requires data that objectively describe the "actual" performance of the institution, which the faculty and staff use to understand current performance and improve the future performance of their programs and institutions (Tadesse, Manathynga, & Gillies, 2018). The result is a system in which the principles of machine learning define the data processing functions and create a credible and trustworthy artificial intelligence for institutional effectiveness (Yousef, Allmer, Baştanlar, Özuysal, & Walker, 2013). The purpose of this work is to offer a fully aligned system of authentic assessments, which provide faculty and staff with credible and trustworthy information to monitor, demonstrate, and enhance institutional performance (Swaggerty & Broemmel, 2017).

The processes and procedures in this work adapt recent and current strategies of performance measurement, assessment, and sensemaking in the discipline of organizational effectiveness into a science-based approach to the assessment and sensemaking of institutional effectiveness in higher education (Cameron & Whetten, 2013). The principles of organizational assessment and the sciences of educational and psychological measurement and assessment define the content and structure of the information collected in this system (Knight, McLaughlin, & Howard, 2012). As such, this approach is a "best science" approach to institutional assessment and effectiveness. In this work, the goal is to present a fully-aligned system of assessments for institutional effectiveness, which are disciplined by appropriate technologies.

The methods and instruments employed in this assessment system emerged from research, design, development, and testing the results of their use as institutional effectiveness assessments for a cross-section of higher education institutions. These instruments have consistently yielded stable, statistically powerful, credible, and trustworthy data about the performance of the institution. These data inform authentic assessments, data modeling, and sensemaking functions to evaluate the effectiveness of the institution (Swaggerty & Broemmel, 2017).

The principles of machine learning and artificial intelligence frame the data modeling and sensemaking strategies to visualize actual institutional performance from multiple perspectives. The output of this approach is a system that provides credible, trustworthy, and meaningful data for the evaluation of effectiveness, which the human intelligence in the institution evaluates.

This page intentionally left blank

Chapter 1

Defining, Measuring, and Assessing Effectiveness

Overview

The purpose of this discussion is to articulate the assumptions and clarify the terminology used to design a systematic assessment method to measure (monitor), assess (demonstrate), and make sense of institutional effectiveness. This approach synthesizes the sciences of assessment and related concepts into the design and development of each part of a practical and powerful assessment system, which produces authentic, reliable, and valid data for institutional sensemaking.

Extended empirical discussions of assessment theory and practice are beyond the scope of this work and only employed to create an understanding of the assumptions that underpin the design decisions. The Reference List of this work includes an extensive collection of research, which the reader may consult to pursue these theories and the research that underpins these design decisions in greater detail.

Defining Effectiveness Versus Ineffectiveness

In organizational terms, an effective process produces the intended result (Daft, 2006; Harrison & Shirom, 1999). Institutional effectiveness is the ability of the institution to do what it says it does, in the way it says it does it. Effectiveness includes the assessment of both products (what) and processes (how) to understand the performance of an institution.

A process or procedure that produces some outcome other than the intended outcome is ineffective. In organizational terms, ineffectiveness is a lack of effectiveness. In other words, the system is producing a different outcome than desired. From a systems perspective, a system always produces the result its design produces, and human interventions have negligible impact. Converting an ineffective system into an effective system needs a change to the design of the system. Hence, this work focuses on measuring and improving the design of the institutional systems instead of people.

Assessing Effectiveness

In the model of systematic assessment under discussion, institutional effectiveness is the synergistic interaction of all functional processes to produce an outcome that is greater than the sum of the functional processes. To evaluate effectiveness, the measurements (data) of functional performance are collected, analyzed, and modeled to show these functional interactions. However, individual measurement data are insufficient to understand institutional effectiveness (Badia, 2014; Glidden, 1996; Macfadyen & Dawson, 2012).

The credible assessment of institutional effectiveness measures the performance of each function as defined by the mission, vision, and values (MVVs) of the institution (Camelia & Marius, 2013; Mouritsen, 1986). These organizational statements define the intended outcomes of each functional system and the measurements collected describe the degree to which the process is delivering its intended outcome, which is the traditional definition of effectiveness: the system produces the intended outcome. The data collected to measure the performance of each function describe the interactions between the stakeholders and their interactions with the functional system (Sullivan & Wilds, 2001).

Authentic interactional data exist in the environment in which they occur and not a laboratory, which controls for threats to trustworthiness and credibility. These data contain measurement error and imprecision that the measurement, assessment, and sensemaking processes in this systematic process respect (Bowman, Kibria & Banik, 2013).

Terminology

Assessment is a well-defined and adaptable group of sciences codified through many years of application and testing (Banta, 2009). These sciences articulate the methods adapted in this work to design, develop, and administer authentic measurement instruments, summarize, and aggregate those data into meaningful assessments of functional performance, and model those same data into tools to facilitate sensemaking of institutional effectiveness. However, the adaptability and multiple applications of research methods have created an inconsistent use of terminology by assessment scientists as well as practitioners.

It is the responsibility of each assessment designer to articulate the terminology in use and to employ that terminology consistently to communicate with diverse populations and engage the diversity of those populations into the creation of a synergistic result. The discussions in this work follow the definitions and explanations articulated below without suggesting that this usage is the "correct" definition.

Measurement

For purposes of these discussions, measurement refers to the process of collecting data from multiple individual sources to quantify functional performance. Measurement data *describe* the performance between institutional processes and procedures (functions) and the interactions with the users of the system. They measure the performance of institutional functions with individuals (Harrison, 2015).

The size of the correlations in the resulting individual data reveals the key performance indicators (KPI) (Badawy, El-Aziz, Idress, Hefny, & Hossam, 2016; Chan, 2015). It is critical that these measurements produce credible and trustworthy data, which objectively describe the performance of each function with each user and enable the modeling of institutional effectiveness from these same data. These data describe the intended performance framed by the intentions articulated in the MVVs.

Assessment

Assessments summarize the individual responses for the functional performance indicator they measure. These performance indicator summaries aggregate the data into the constructs that they define. This aggregation changes the focus of the indicators from the measurement of individual performance to the demonstration of the performance of the function.

The process of assessment aggregates multiple measures for each indicator into the constructs of the assessment, which demonstrate the performance of the function. The assessment function is another descriptive process meant to present a credible and trustworthy representation of performance.

Analysis

Analysis is the process of examining data to confirm the mathematical structure and relationships within it. In this work, analysis always refers to discovering the emergent "qualities" of the data through mathematical or statistical methods, rather than inferring the quality of the data. These methods reveal the underlying interdependencies in the data, which create practical models of actual performance (Tadesse, Manathynga, & Gillies, 2018). Actual performance describes what is occurring in the system without comparison with intentions or other assumptions (machine learning).

Evaluation

Evaluation is a process of interpreting the data in the assessments to understand the relationship between intended and actual performance. In this application, the comparison determines the degree to which each function is performing as intended. In institutional research, the lack of precision in the data dictates the use of descriptive methods instead of inferential methods (Heinze, Shapira, Rogers, & Senker, 2009; Knight, McLaughlin, & Howard, 2012). In this system of assessments, the evaluation phase of the process is objective (observational) and descriptive. It describes the actual performance of the functions within an institution.

Sensemaking

Institutional sensemaking facilitates "looking beyond the assessment data" to quantify the interactions (synergies) between the functions, which generate institutional effectiveness (Birnbau, 1991; Macfadyen & Dawson, 2012). Teresa

Amabile (1998) has described this intellectual process as "scanning the perceptual space" between and beyond the data. In institutional effectiveness research, sensemaking is a collaborative process that integrates a diversity of perspectives to generate innovative and plausible interpretations (Gokhale, 1995; Perry, 1999). The differences and disagreements that emerge from the amalgamation of these divergent perspectives drive the sensemaking process into new and innovative directions (Binswanger & Oechslin, 2015; Snehal & Kremer 2010; Woerkom & Sanders, 2010; Young, Alibali, & Kalish, 2012). By engaging and integrating the value of difference into institutional decision making, the sensemaking process creates a foundation for a culture of improvement and effectiveness (Drouin, Stewart, & Van Gorder, 2015; Goldman, 1999).

According to Weick (1995), the intellectual process of sensemaking in organizations has seven procedural characteristics. These characteristics are:

1. *Grounded in Identity Construction* – accomplished in this method by using the shared MVVs of the organization to collect and organize the data.
2. *Retrospective* – accomplished in this model by focusing on the objective measurement of real events that have occurred in the natural environment.
3. *Enactment of sensible environments* – accomplished in this method by assuming a connection between the conditions experienced and the outcomes of the participants, such as assuming the causal relationships between the components of the instructional system and student performance.
4. *Social* – the realization that sensemaking needs the collective intelligence of the organization to make sense of the data. The data give clues and hints, which the creative intelligence of the leaders synthesizes into a complete picture of performance and effectiveness (Fourie-Malherbe, 2015; Leithwood & Mascall, 2008).
5. *Ongoing* – the embedding and acculturalization of sensemaking into continuous activities and operations, accomplished in this method by the design of the systematic approach to assessment which monitors performance with credible and trustworthy data.
6. *Focused on and by extracted clues* – accomplished in this method by clustering the key performance indicators into models of actual performance and using those models as clues or hints (indicators) of institutional effectiveness.
7. *Driven by plausibility rather than precision* – accomplished in this model through the conceptualization and treatment of institutional data as indicators instead of data with absolute precision. That is, respecting the attributes of the data as they "are" rather than assigning or assuming attributes they do not demonstrate (i.e., parametric vs. non-parametric data). The search for plausibility is a distinguishing characteristic of this assessment process and one that distinguishes practical research from empirical research.

Collectively, the sensemaking activities create a systematic process for evaluating the effectiveness of institutional performance. The data from the measurements monitor performance at the individual level, the assessments aggregate those data into constructs to demonstrate performance at the program or functional level, and the sensemaking process uses the data to evaluate performance

at the institutional level, all within the limits of confidence (power) contained in the data.

Valid Data – Credibility

The term validity applies to the collected data and not the instrument. The content of the data demonstrates its validity, which is identical to the definition of the function to be valid. Data that are valid to one situation are only valid in another situation if both situations are identical, not similar.

The discussions in this work focus on design validity, which demonstrate construct (concepts) and content (processes) validity through the design assumptions for each instrument. Validity, however, is an attribute of the collected data, and not the instrument. Since institutional data contain copious amounts of variance and error, the term that describes valid data in this work is credibility, which assumes the data to be more qualitative than quantitative and respects its lack of precision.

Reliable Data – Trustworthiness

It is the structure of the instrument that generates the trustworthiness of the collected data as confirmed (calculated) in the collected data. Trustworthiness refers to the degree to which the data reveal the meaning that the respondents intended. For these purposes, a statistical power test measures the trustworthiness of the collected data (Fosnacht, Sarraf, Howe, & Peck, 2017). The power test, discussed in detail later in this chapter, directly measures the attributes of the collected data to determine its trustworthiness.

Authentic Data

Another characteristic of the data that is important to institutional assessment is the authenticity of the data. Authenticity describes data collected from an activity that aligns with the performance objective. If the objective is to demonstrate knowledge, then the data measure knowledge. If it demonstrates performance, then the data measure performance. For institutional research, the "best evidence" of performance is direct measurements of authentic activities. A distinguishing characteristic of this work is the attention to authenticity in the data collection phase.

Functional Data

Functional data are those data that affect and influence the performance of a specific institutional function. In the case of higher education, the functional data measure the components of teaching, learning, and support services. They measure the results of the functional system's interaction with each student or stakeholder.

Institutional Assessment of Effectiveness

Educational Institutions are complex organizations that employ complex functions to serve the needs of stakeholders living and performing in the twenty-first

century. There is little that is simple or dualistic about the MVVs of higher education. The assessment and evaluation of institutional effectiveness need systems of assessment that reflect and accommodate this complexity and employ methods that study complex phenomena. Such systems use multiple scales to measure the dimensions of functional performance and aggregate those measures into the constructs of performance as defined in the MVVs of the institution.

There are at least two broad models of information needed to assess institutional effectiveness comprehensively: political models and performance-based models. A political model of effectiveness measures the construct of satisfaction. It asks stakeholders to decide whether they are satisfied with the performance of the organization based on their opinions of satisfaction. Political models appropriately measure some of the more personal functions in higher education that are too complex to parse into practical assessments (Charteris-Black, 2005; Charteris-Black & Palgrave Connect, 2011). Basing the evaluation of the effectiveness of the entire organization on a political model does not give a complete demonstration of institutional effectiveness and may assess the characteristics of the students more than the institution.

A performance-based model of effectiveness collects objective measurements of the performance of institutional functions and compiles them into the constructs of functional performance to show institutional effectiveness. These data reveal the drivers and constraints within the functional data to describe institutional performance.

These models conceive the institution as a complex, multidimensional phenomenon, created by the synergistic interaction of the parts of the institution, its functions. This work assumes the performance of the institution to be a quantum or synergistic result of the interactions of those parts. Therefore, single performance indicators do not fully demonstrate institutional performance, and it is necessary to "look" beyond the performance indicators to make sense of institutional effectiveness (Bazeley, 2015; Bedeian, 2015). In this application, the sensemaking phase enables looking beyond the data for meaning (Sullivan & Wilds, 2001).

Machine Learning – Artificial Intelligence

Machine learning (ML) is the use of algorithms to model data to reveal the "actual" performance of a complex system (Alpaydin, 2014; Burger, 2018). ML uses algorithms to define the attributes of data and build relevant performance models, which describe current performance and predict the effects of changes on future performance (George, Osinga, Lavie, & Scott, 2016; Hastie, Tibshirani, & Friedman, 2009). The measurement of the internal interactions or processes within the system (machine) constitutes the ML data (Domingos, 2015). This approach creates the ability to infer causation between the data points and predict the results of changes to the system, a process known as predictive analytics (White & Breckenridge, 2014; Witten, 2017).

This systematic approach to institutional effectiveness applies the principles of ML to design the measurement, assessment, and data modeling processes (Alpaydin, 2014). The following ML strategies are employed to construct this systematic assessment.

Authentic Data

The data collected in an ML process authentically measure the performance of the system. In an institutional effectiveness (IE) system, this is accomplished by measuring the characteristics of the individual interactions between the system and the user and modeling those data for evaluation (Mohammed, Khan, & Bashier, 2017).

Algorithms as Measurement Strategies

The measurements of the individual interactions are framed by definitions of performance, which are extracted from theories of organizational effectiveness and adapted to institutional systems. These algorithms define and frame the interactions between the machine (institutional system) and the human being (user, constituent) to measure performance (Domingos, 2015; MacKay, 2003).

Internal Data

To reveal the performance of a system, the data describe the performance of the system. Data external to and outside the control of the system are not valid measures of system performance in an ML environment and interfere with the ability of the ML process to describe actual performance through a coherent systematic assessment process.

Acyclical Data

The measurement of the performance of an institution aggregates data from multiple performance cycles (sessions, semesters, quarters, years) to comprehensively understand the performance of the institution over time. Acyclical data increase the statistical power of the data and allow outliers and other anomalous data to emerge above the threshold of discrimination (Lerma, 2012; Liu, Yue, & Li, 2011; Wanner, Hajicová, Gerdes, & IOS Press, 2013).

Classification

In ML and statistics, classification is the process of categorizing or grouping data into interconnected subsets. In other words, all data that impact a function are grouped into a description of that function, even if they may be applied to more than one function (Medler, 1998). These classifications result from the summarization of the data (Anitha & Deisy, 2015; Balyan, McCarthy, & McNamara, 2018; Medler, 1998; Salles et al., 2017).

Statistical Modeling

The data summarization, aggregation, and modeling strategies employ valid statistical methods that are appropriate to the attributes of the data to analyze and model the data. An ML process aggregates, describes, and organizes the raw

measurement data into assessments that construct institutional performance (Hastie et al., 2009; Yan, Wang, & Lu, 2018).

Dimension Reduction and Modeling

Dimension reduction refers to the process of reducing the number of actions under evaluation by calculating the statistical strength of each interaction contained in the raw data. This process organizes the data into a smaller collection of data points to enhance human information processing limits (Goodfellow, McDaniel, & Papernot, 2018; Saraswati, Nguyen, Hagenbuchner, & Tsoi, 2018; Tangkaratt, Morimoto, & Sugiyama, 2016).

Association Learning

Association learning is an ML method for discovering interesting relations between variables in large databases. It reveals strong rules (relationships) that exist in the data using measures of interestingness (causality). As the data set grows, new rules emerge from the increased power within the data (acyclical), which reveal hidden "interests" (Piatetsky-Shapiro, 1991).

Outlier Detection

In ML, outlier (anomaly) detection is the process of identifying rare data, events, or observations, which differ significantly (interestingness) from the majority of the data (Piatetsky-Shapiro & Matheus, 1994). Typically, the anomalous items represent an attribute of the system that exists in a small sample of the population and, as a result, does not contain enough power to rise to the system level of performance (Angiulli, Ben-Eliyahu-Zohary, & Palopoli, 2014; Ha, Seok, & Lee, 2015; Yan, Wang, & Lu, 2018; Zimek & Schubert, 2017).

Structured Prediction

Structured prediction is an ML technique that involves predicting the structure of systems conceived as complex phenomena, as differentiated from discrete or single (linear) values. Structured prediction employs the interconnectedness in the data to determine the effects of a change in one data point on others (Du, 2017; Jiuqing, Xu, & Xianhang, 2017).

Artificial Intelligence

Organizing the ML data outputs into models that describe institutional performance as the human intelligence has defined it creates an artificial intelligence (AI) (Duda, Stork, & Hart, 2001; Gopnik, 2017; MacKay, 2003). There are many models of AI that create a comprehensive collection of flexible conceptual frameworks to model the attributes and interconnectedness of the data (Boden, 2015; Ngai, Tao, & Moon, 2015; Russell & Norvig, 2014). To visualize data as a series of