Alumni.in

Iberoamerican Journal of
Competitive Intelligence
Revista de Inteligência Competitiva

GUEST PAPER

Check for updates

POST-PANDEMIC REFLECTIONS ON CHALLENGES AND OPPORTUNITIES FOR MARKETING RESEARCH IN THE 21ST CENTURY

Joe F. Hair Jr.[1] | Dana E. Harrison[2] | Jeffrey J. Risher[3]

[1]University of South Alabama, Alabama, EUA.
[2] East Tennessee State University, Tennessee, EUA.
[3]Southeastern Oklahoma State University, Durant, EUA.

Corresponding Author:
Joe F. Hair Jr.
E-mail: jhair@southalabama.edu

How to cite this article:

ABSTRACT

The role of marketing is evolving rapidly, and design and analysis methods used by marketing researchers are also changing. These changes are emerging from transformations in management skills, technological innovations, continuously evolving customer behavior, and most recently the Covid-19 pandemic. But perhaps the most substantial driver of these changes is the emergence of big data and the analytical methods used to examine and understand the data. To continue being relevant, marketing research must remain as dynamic as the markets themselves and adapt accordingly to the following: data will continue increasing exponentially; data quality will improve; analytics will be more powerful, easier to use, and more widely used; management and customer decisions will increasingly be knowledge-based; privacy issues and challenges will be both a problem and an opportunity as organizations develop their analytics skills; data analytics will become firmly established as a competitive advantage, both in the marketing research industry and in academics; and for the foreseeable future, the demand for highly trained data scientists will exceed the supply.

Keywords: big data, data analytics, data quality, marketing analytics, marketing research.

Check for updates

Open access

1 | INTRODUTION

The role of marketing for consumers and businesses is rapidly evolving (Ferrell, Hair, Marshall & Tamilia 2015). Indeed, the marketing function is fundamentally changing as a result of digital transformation, data, data analytics and most recently, personal mobile devices, such as infusion software, which track time as individuals work or play, can search for local businesses or customers while in a coffee shop, send an invoice or quote from a client's office, and much more (Shah et al 2014). Marketing research design and analysis methods are also changing rapidly (Hair, Black, Babin & Anderson 2019). These changes are emerging from transformations in management skills (Henke, Levine & McInerney 2018; Davenport 2018), technological innovation, quite often in the digital marketing space (Davenport 2018), and continuously evolving customer behavior, particularly resulting from changing needs and expectations (Wedel & Kannan 2016), most recently impacted by the Covid-19 pandemic. To continue being relevant and effective, marketing research must remain as dynamic as the markets themselves.

At the same time, academic research is increasingly applying advanced analytic tools and algorithms (Erevelles, Fukawa & Swayne 2016; Syam & Sharma 2017, Wedel & Kannan 2016; Petrescu & Krishen, 2017), and even artificial intelligence in combination with neural networking (Ababukar et al. 2017). Just as advanced analytics is providing key competitive advantages in industry practices, it also is forcing rapid changes in social sciences academic research (Krishen & Petrescu 2018).

Directions for research, particularly marketing and related social sciences, include applications of advanced analytics techniques and technologies. Wedel and Kannan (2016), for example, suggest how machine learning methods and cognitive computing technologies can be applied to better understand marketing problems and opportunities. Similarly, Erevelles, Fukawa and Swayne (2016) identify propositions that examine big data solutions to various marketing activities, as well as how these findings can be applied to develop sustainable competitive advantages. Syam and Sharma (2017) highlight the importance of understanding how machine learning (ML) and artificial intelligence (AI) can advance marketing research, and Abadular et al. 2017) propose ways to apply AI and neural networking. Thus, as these and other parallel developments are applied to expand our knowledge of marketing and consumer behavior trends, businesses can more effectively meet consumer needs and reduce information overload.

What was not anticipated when we initially shared our thinking on the emerging role of marketing research in the 21st century (Hair, Harrison & Risher., 2018) was the subsequent influence of the Covid 19 pandemic and the related economic and social developments. This updated version of our previous article includes comments on more recent developments in marketing research as well as initial impressions of how the pandemic is impacting marketing research.

The purpose of this article, therefore, is to suggest how emerging market trends, and particularly those in marketing, are likely to impact marketing research. First, we

explore the emerging role of the marketing function that is developing from new business models. Second, we will examine the implications of this change on academic marketing research. Third, we will address transformations that are likely to continue evolving in the future. Finally, we will include reflections on more recent developments impacting marketing research such as the Covid-19 pandemic and the resulting supply chain disruptions. Our framework for these comments is a previous article we published prior to these events that did not anticipate these recent developments (Hair, Harrison & Risher, 2018).

2| THEORETICAL BACKGROUND

2.1 | Emerging Role of Marketing and Marketing Research

Numerous developments are influencing marketing, and thereby marketing research. Among the most important developments are the following:

2.1.1 |Modern Markets

Perhaps one of the most important trends impacting marketing research is internet-based markets. Business models such as the sharing economy or digital matching firms have emerged from advanced technology platforms and become highly successful, digitally connected marketplaces (Harrison & Hair 2017). Examples include internet-based companies such as Uber, Amazon, Airbnb, DoorDash, Netflix, as well as cloud storage and computing services available from Salesforce.com, Microsoft Azure, IBM, and Amazon Web Services, that have forged successful new business models, ultimately revolutionizing industries. While these developments have been gradually materializing over the last 20 years, their full impact is now becoming evident.

Customers and companies have collectively become less averse to risks often associated with online transactions, and this has fueled growth among these emerging business models. Furthermore, dynamic customer behavior and market changes have contributed to the need for greater convenience and more personalized experiences than initially offered through online markets. For example, faster moving trends in clothing fashions and fads is one of many market developments that can be attributed to the capabilities of internet-based products and services. But there are many other examples. In essence, innovative companies have developed and mastered value propositions that more effectively respond to market opportunities, and more closely fulfill customer wants and needs.

A critical component that has emerged along with these new value propositions is digital data, which has enabled both businesses and consumers to not only be more informed than ever before, but also to sell and seek almost all products and services completely online. For example, if you have a My Google account it is tracking your activities 24/7. One individual checked the number of items tracked over a period of time and found that on average, more than 200 activities were added to their My Google account per day, which is more than 73,000 per year. Who would have thought a few years ago individuals would not only be sharing their personal data, but bicycles and cars instead of buying them, not to mention being willing to give organizations such as Google

or Facebook access to huge amounts of their personal data, or to use their mobile phone to search for and purchase groceries, or almost any other product or service they want to be delivered to their door.

2.1.2 | Emergence of Big Data

Companies and customers are producing large amounts of data from various touchpoints. On the customer side data is captured through heavily promoted omni-channel experiences. As examples, consider Target's buy online and immediate pickup in store, Amazon's prime membership and associated sales and shipping relationships, or FedEx's use of blockchain methodology for managing customer dispute data and as an efficient method for tracking packages, all are producing mountains of useful information for both business and consumer sides of transactions. Perhaps the most pervasive data collector is Google, which has trackers on almost 80 percent of all websites (except China, where Google withdrew because they refused to permit censoring of information). Similarly, on the B-to-B side General Electric and Boeing are installing sensors in jet engines, robots are being coordinated to manufacture many products, such as autos and airplanes, sensors are enabling vehicle brands such as Volvo, Tesla, Mercedes, BMW, and others to market self-driving cars, and many other distribution and manufacturing innovations, all of which are also collecting data across the entire manufacturing and distribution ecosystem to generate actionable and valuable insights. As a result, companies can better manage production, understand and respond to customer requirements, and add value. With this emerging resource of data, they can more intelligently manage their businesses, improve response times, promote innovation, reduce costs, and boost revenues and profits.

Data changes power relationships within the organization. Today only about ten percent of the U.S. economy is the data economy – small now but the percent is expanding exponentially. In the past, data was the responsibility of the Chief Information Officer (CIO), a staff position. Last year (2021) in the U.S., the trend of a larger proportion of information technology budgets being spent by line managers than CIOs continued. For CIOs, the security of data systems continues to be a high priority. But CMOs and CFOs focus more on business execution. The Equifax data breach several years ago was likely a result of this transition in responsibility for data security. At the same time, for individuals trained in the business college it is likely surprising that over 15 percent of CEOs are now data scientists – they do not have an MBA degree!

Data and data analytics have the potential to change the economic order of the world in ways that will be disturbing to many companies and people. Moreover, knowledge-based data will likely become a source of power that is difficult to challenge for countries and companies lacking this knowledge and these skills. In short, data and analytics have become to the 21st century economies what oil was to the 20th – a valuable asset essential to a better economic life. This was evidenced with the addition of Salesforce.com (CRM) to the Dow Jones Industrial Index in 2020, not only by replacing Exon-Mobile, but by marking the first time in the history of the New York Stock Exchange that a CRM/Marketing Research company was deemed relevant to represent a substantial portion of all industrial activity. It appears we are on the brink of a new data world, which will definitely change both marketing and marketing research.

On the data analytics side, advanced technology and lower costs are now capable of storing and processing the large amounts of data from these various interactions. Storage costs of one megabyte of information, for example, have dropped from $15.00 U.S. in 1992, to less than 10 cents U.S. for one megabyte in 2021. But, as more data is produced and stored, almost all of the data located in cloud storage, there is a greater need to apply advanced marketing analytics methods. These emerging methods enable marketing researchers to gain a more sophisticated understanding of the data, not only related to what is selling and through which supply chain channels, but to produce improved measures of return on marketing investments, customer lifetime value, drivers of customer loyalty, and so forth. Clearly, without these emerging analytics tools the supply chain disruptions associated with the pandemic would have been much worse for consumers and businesses than was experienced.

2.1.3 | Transformative Marketing

As the amount, type (structured and unstructured), and complexity of data changes, so do the methods of research necessary to support these developments. The variety, velocity and volume of data emerges from changing marketing environments, emerging technology, and the lower cost of collecting and storing data (Erevelles, Fukawa & Swayne 2015; Kumar 2018; Varadarajan 2018; Davenport 2018). Examples of drivers of data creation include evolving customer purchase behavior, the internet of things, digital wearables, the rise of artificial intelligence, enhanced supply chain technology, and the ability to embed analytics into sales, distribution, and production systems, to mention a few (Kumar 2018). Facebook alone has over 200 billion photos stored in the Cloud, and more than 350 million new photos are uploaded every day. Moreover, Facebook users upload more than 4 petabytes of data per day, which represents more than a million gigabytes. In response to these drivers, marketing strategies, organizational decisions and performance are being revolutionized by information assets, analysis capabilities, and increased customer knowledge.

2.1.4 | Changing Role of Marketing Research

There is value in both traditional research approaches that examine, for example, hypotheses derived from practice and theory, as well as predictive analytics that commonly emphasize solving business problems with data (Babin, Griffin & Hair 2015; Delen & Zolbanin 2018). Indeed, tools and techniques are increasingly available that inform different levels of sophisticated intelligence (Davenport & Harris 2017). These tools and techniques range from simple spreadsheet analyses used by companies to understand how many, how often or why something occurred, predictive tools that help to explain and forecast what is likely to happen, automated machine learning tools applied to obtain insights from data, and most recently a college textbook on Essentials of Marketing Analytics, McGraw-Hill Education (Hair, Harrison & Ajjan, 2021). In short, all levels of sophistication offer the potential for competitive advantages (Davenport & Harris, 2017), but increasingly companies must adopt and rely on the most sophisticated tools available. What is often overlooked, however, is that machine learning, increasingly coupled with artificial intelligence, almost always requires human intervention to execute and complete the analyses. Thus, the role of marketing researchers will both expand as well as increase in complexity, requiring much more advanced training, thus creating additional job opportunities for college graduates studying relevant analytics

fields. The following paragraphs provide an overview of how three marketing research approaches – descriptive, predictive, and prescriptive analytics – are evolving and changing.

2.1.5 | Traditional Descriptive Analytics

Traditional research often uses descriptive and diagnostic methods. This type of research is more explanatory and confirmatory in nature (Sivarajah, Kamal, Irani & Weerakkody 2017). For example, key performance indicators (KPIs) provide real-time dashboard visualizations for understanding how effective and efficient the company is in achieving business objectives such as sales performance, inventory control management or supply chain responsiveness. Although valuable as historical metrics, this confirmatory feedback offers only limited predictive insights as to whether future results will be similar. But it will continue to be very useful for marketers as a means of monitoring day-to-day operations, and marketing researchers will need to provide support for better understanding and explaining market developments, as well as suggesting innovative ways to apply this information.

Data used in descriptive analytics can be found within and external to the organization. Companies often collect data on a regular basis regarding customers and competitors, but also use data from external sources such as data.com and salesforce.com. Although the amount of internal data is much easier to collect at this point, it is likely that the future will bring advanced technology resources enabling organizations to more efficiently collect data from external sources.

2.1.6 | Predictive Analytics

Advanced technology combined with complex algorithms facilitates predictive analytics approaches. These emerging techniques examine mountains of data to promote narratives for more effectively assessing opportunities as well as predicting future developments. In short, predictive analytics applies modeling tools to data to predict future market developments and trends (Davenport & Harris 2017). Researchers have categorized predictive analytics into two categories: regression techniques and machine learning techniques (Gandomi & Haider 2015), both of which are increasingly core elements of marketing research programs.

While predictive analytics typically requires some level of human intervention, there is often flexibility in the level of involvement (e.g., selecting target variables, developing theoretically derived models, permitting the data to tell a story). Big data obtained from many sources, potentially containing thousands and even millions of driver variables, is analyzed using hundreds of analytical models. But this type of modeling presents substantial challenges compared to traditional methods (Davenport & Harris 2017) and requires more knowledgeable data analysts than in the past. For example, with many sources and types of data (both structured and unstructured) integrated into a single database, it is virtually impossible to hypothesize every possible relationship. According to Peter Norvig (2009), Google's Research Director, the large amount of data and resulting complexity of the relationships often make it advantageous for computers to derive models from the data, versus having humans initially spend time using scientific

approaches to identify possible models. Computers, therefore, explore the data to locate hidden relationships that otherwise often would not be identified, and then develop predictive models that support tentative conclusions and facilitate strategic decision-making.

Applications of machine learning (ML), natural language processing (NLP), and neural networking (NN) have all increased substantially in both academic marketing research and practice. Machine learning algorithms are designed to gather knowledge from available databases and combine it with automated learning behavior processes, with the objective of improving our understanding of complex relationships and enhancing prediction of desired outcomes. Most of these methods were inherent in traditional multivariate data analysis and are now being extended and adapted to solve additional problems with improved accuracy. Examples include combining these methods with artificial intelligence (AI) and the increasingly widespread digital data sources (Davenport et al. 2020) to develop algorithms and computer software that develop intelligent solutions (Shankar 2018). The solutions improve market researchers' ability to more effectively interpret both internal and external data, develop solutions using the data, and facilitate more flexible adaptations (Kaplan & Haenlein, 2019). Examples of related methods and analytical extensions include metaheuristics such as genetic algorithms and probabilistic methods such as Kalman filters.

Predictive analytics is increasingly being applied in the corporate sector as well. Not only by the largest companies but now some middle-sized ones are using predictive analytics. Many machine learning algorithms are publicly available through open-source programs such as R and Python. For example, both R and Python offer the ability perform a market basket analysis by applying simple association rules using the FP-Growth algorithm.   For unstructured data, natural language processing algorithms may be employed to determine sort, segment, and count the number of times words appear in corporate feedback sites, social media posts, and consumer-related blogs. When joined with customer transaction data, more probabilistic techniques such as linear discriminant analysis or naïve bayes may be useful in predicting which Google AdWords will provide the best Return on Marketing Investment. One drawback to using R and/or Python is they are not user-friendly. This has led to the emergence of niche companies such as Creatio and SugarCRM that cater to midrange companies and promote affordable low code/no code CRM platforms driven by AI. For regression-based predictive analytics, the SmartPLS software is a relatively inexpensive graphical interface which is easy to use and performs well even with limited data.

Many larger companies and advanced researchers employ automatic applications of machine learning. Automated Machine Learning (Auto ML) algorithms gain information from past data and use subsequent data entries to reduce error and improve predictive ability (Hair, Harrison, & Ajjan, 2021). Advanced data software such as Databot, Microsoft Asure ML, and Rapidminer offer users the ability to employ Application Programming Interfaces (API) to import data automatically into virtual repositories within the software, or through repository hosting services such as Github, which may then be immediately accessed by the software. This allows virtual updates of customer and business data from transactional data generated by customer touchpoints, web browser cookies, third-party data, web-scraping, econometric data, and information delivered through mobile applications which may deliver personal information such as

social media activity, geospatial data, and even biometric data (Hair & Sarstedt 2021). These Auto ML applications, coupled with continuous data updating, allows real-time updates to predictive models and outcomes, which may improve business relationships through better forecasting and consumer relationships through hyper-personalization (Ma & Sun 2020). A simple application of real-time hyper-personalization would be geofencing, in which a business may send personalized ads to customers who are identified as walking or driving within a set radius of its retail location. A much more complex example can be seen in the PROS platform which is employed by many airlines and online travel agencies to automatically adjust airfare based by using a host of inputs such as consumer search history, cookie data, and past purchase behavior, external market data, seasonality, and availability to predict the price which best matches capacity with demand.

The advancement of marketing research, both in academia and in practice, requires researchers to extend the goal of machine learning from predictive outputs to that of explainable artificial intelligence, which occurs when predictive outputs can be understood and explained using existing theory, reason, and logic (Hair & Sarstedt 2021). An increase in explainable AI would not only improve machine learning adoption in practice but could also be used to extend existing theory and create and testable hypotheses which could be validated through traditional research methods (Ma & Sun 2020). To do so, however, still involves considerable human intervention from marketing and data analysts. Competent individuals must be able to select the appropriate analytical method, and when solutions are identified other qualified analysts must interpret them in terms of logic, relevance to business problems, and ultimately strategic value. Ultimately, human researchers will continue be necessary to facilitate the "why" in response to the "what" and "when" findings generated by machine learning outcomes. The problem is there are not enough qualified data scientists to make these decisions. McKinsey estimates the U.S. alone faces a shortage of 140,000 to 190,000 people with analytical expertise, and 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of big data. In response to this need, U.S. universities have added more than 100 data analytics bachelor's degree programs in the past five years, with more being added every year.

2.1.7 | Prescriptive Analytics

Prescriptive analytics is an advanced level of analytics that examines what should be done or what can be done to make something happen. Prescriptive analytics enables marketing researchers to determine optimal behaviors and evaluate the eventual business impact (Davenport & Harris 2017; Sivarajah et al. 2017). To accomplish this, prescriptive analytics experiments with and optimizes various scenarios at an accelerated rate and is at less risk for subjective human interpretation (Davenport & Harris 2017). This may be good if the analytics work well but may not be so good otherwise. Thus, prescriptive analytics while not completely autonomous at this point can produce implications and generate recommendations. As one example, FleetPride, a provider of truck and trailer parts for customers across many industries, found success in implementing prescriptive analytics. Using solutions offered by IBM, results provided supply chain managers with insights to operate seamlessly; allowing the company to optimize distribution networks, increase the speed of inventory flow and reduce

warehouse packing errors. Although the number of organizations using prescriptive analytics is growing, it is difficult to know the level of adoption. Unfortunately, it isn't uncommon where adoption has occurred in departmental or unit silos, that others throughout the organization are often unaware that prescriptive analytics has been implemented (Rossi 2015). Of course, the key to successful prescriptive analytics relies, as with predictive analytics, on the skills of the data scientists executing it, and there is a global shortage of trained individuals in this area as well.

2.1.7.1 | *Where does Marketing Research go from here?*

Sophisticated tools and techniques are being readily adopted in market research practice, with budgets increasingly dedicated to the purchase of advanced analytics solutions (Mela & Moorman 2018). With the constantly evolving data inputs (e.g., technologies), data types (e.g., unstructured data) and changing customer behavior, it's important for marketing research to remain managerial relevant and timely. Several relevant questions are to what extent should marketing researchers and data scientists allow these new analytical tools to develop models and outputs without first developing theoretically derived hypotheses? Where is the place in marketing research for the integration of predictive and prescriptive tools? How could this method develop new theories and add value to the field? How can marketing research apply software such as SAS/STAT for predictive modeling, LIWK for text analysis, social media network analysis tools such as NodeXL or machine learning tools such as DataRobot? How are artificial intelligence (AI) solutions be applied, such as Salesforce Einstein that not only offers predictions but also makes recommendations on how to respond to customers, impacting salesforce capabilities? Marketing research must remain cognizant of the evolution in big data and analytics, understand how and when to use these methods in research, and maintain pace with managerial priorities in practice.

3| Changes in Methodology and Research Design

Parallel to the evolution in advanced techniques and tools, we are also witnessing changes in research designs and methodologies.

3.1 | The Use of Survey Data

As recently as the 1990s more than 70% of scholarly published papers were survey or interview based. But by 2013, the number of survey-based published papers in the three top marketing journals [Journal of the Academy of Marketing Science (JAMS), Journal of Marketing (JM), and Journal of Marketing Research (JMR)] was only about 35% (Hulland, Baumgartner & Smith 2018). To collect data, researchers increasingly have relied on do it yourself, online survey methods such as Qualtrics, Google Survey, Mechanical Turk, and Survey Monkey (Hulland & Miller 2018). Companies and researchers alike are capitalizing on the digital transformation to reach survey participants via online platforms. Online platforms have provided convenient access to business and consumer samples that were difficult, costly, and often impossible to reach. Unfortunately, populations are being over surveyed resulting in refusals to respond, and the proportion of survey-based papers has started to decline (Hulland, Baumgartner & Smith 2018). Another issue noted by numerous social sciences disciplines, and particularly management and supply chain groups, is single respondent survey designs. In this type of

design, the same survey respondent answers questions related to both the independent and dependent variables, producing what is referred to a common methods bias (Krause, Luzzini & Lawson 2018; Kull, Kotlar & Spring 2018; Flynn, Pagell & Fugate 2018; Roh, Whippe & Boyer 2013). Other disciplines, including marketing, have raised similar concerns, and are suggesting designs that combine secondary data with primary data as a solution. Unfortunately, to date no method of precisely measuring the extent of common methods bias in survey designs has been developed.

3.2 | Shift Towards Objective Data

With the emergence of digital data, and lots of it, researchers have shifted toward collecting objective, secondary data. Archival or proprietary information, such as historical data, is readily available to researchers (Verma, Agarwal, Kachroo & Krishen 2017). Using historical data, event studies (Sorescu, Warren & Ertekin 2017) can examine the long-term impact of a particular decision on the firm. Businesses and researchers alike are focused on predictive and prescriptive strategies, enhanced by artificial intelligence (Huang & Rust 2018) and machine learning (Antons & Breidbach 2018). To some extent, these techniques are designed to be capable of producing self-correcting analyses, while others such as mathematical optimization, "determine the best solution to mathematically defined problems" (Snyman 2005). In consumer markets, for example, digitalization has facilitated the availability of physiological data. In fact, across the board more technical research methods include biometric and neurological focused data (Chan, Boksem & Smidts 2018; Harris, Ciorciara & Gountas 2017). These techniques, initially emerging in economics, finance, computer science and medical research, are now becoming more prevalent and infused in marketing research, and the social sciences in general. As we look to the future, these evolving research methods will become a critical resource in the dynamic and ever-changing business-to-business and consumer markets.

3.3 | Visualization Software to Communicate Results

It is critical to clearly report research findings in a manner that quickly and easily communicates meaningful results. Information visualization has increasingly been adopted as the leading approach to facilitate communication of the outcome of complex statistical analyses focusing on data exploration, discovery, and display of marketing research findings. When using visualization methods, researchers can create nontechnical reports for decision makers to understand regardless of their background. There are several popular data visualization tools such as Tableau and Microsoft BI that include advanced computer graphics, charts, and mapping. These tools enable researchers to quickly identify visual patterns, present information in real-time settings, create interactive dashboards to reveal filtered information, ultimately improving marketing decision making.

3.4 | The Rise of Unstructured Data

Digital marketing through areas such as social media has produced substantial unstructured content (Balducci & Marinova 2018) that can be submitted to text analysis (Humphreys & Wang 2017), and increasingly coded for quantitative modeling. In fact, the

conversion of unstructured data for quantitative analysis is expanding rapidly since 90 percent of all data is unstructured. Unstructured data includes tweets, photos, service call logs, blogs, customer complaint and comment data, customer purchase behavior, and so forth, all of which has to be analyzed and coded before being submitted to quantitative data analytics software. Programs such as NodeXL and LIWC are used to enhance quantitative modeling (Harrison, Ajjan & Coughlan 2018). Through the collection and analysis of unstructured data, such as digital text or comments, companies and researchers extend their expertise in segmentation, customers' level of engagement and experience, social networking, and sentiment analysis, which also creates expanded opportunities for market researchers.

3.5 | Greater Focus on Relevance, Rigor and Quality

As data and analytical methods become more complex, marketing research is also experiencing a shift toward more rigorous techniques. Researchers should use every reasonable approach to disprove their hypotheses versus feeling it necessary to statistically support hypotheses – so they make a contribution (Babin, Griffin & Hair 2015). For example, endogeneity, statistical power (effect size), external validity and predictive relevance play an important role in the accurate assessment of relationships between exogenous (independent) and endogenous (dependent) variables.

For many years, marketing researchers have examined observed heterogeneity. Observed heterogeneity is based on the knowledge that subgroups of populations likely exhibit differences, and these differences are known when the research is designed and conducted. For example, different age groups of consumers often exhibit different search and shopping patterns, and ultimately purchase behavior. For consumers, observed heterogeneity typically has been based on demographics. For organizations, observed heterogeneity has been based on company size, product or service offerings, convenience, and so forth. More recently, technology has enabled data analysts to examine the possibility of unobserved heterogeneity, which is the notion that populations consist of subgroups of the population that are not easily observable. Knowledge regarding the presence of unobserved heterogeneity is critical in today's empirical studies, and if not explored it potentially threatens the validity of measurement and structural models examined when applying both descriptive and predictive analytics (Becker, Rai, Ringle & Volckner 2013; Sarstedt & Ringle 2010; Hair, Matthews, Matthews & Sarstedt 2017; Hair et al., 2022a). Both industry and academic researchers are rapidly making strides to develop advanced statistical analyses that effectively address potential reliability and validity issues.

Although these issues have long been a concern in marketing research, enhanced statistical methods enable researchers to address them more rigorously (Hult, Hair et al. 2018; Worm, Bharadwaj, Ulaga & Reinartz 2017). Regression based analyses function on the assumption of minimal multicollinearity. At the same time, however, moderation occurs when two or more factors covary enough to alter the impact on the outcome. If the moderation is known or suspected, it can be tested with either continuous or categorical variables. This is particularly useful for identifying hygiene variables or boundary conditions.

Academicians and practitioners are increasingly using sophisticated software to execute predictive analytics. One example, the SmartPLS software (www.SmartPLS.de)

is cutting edge in terms of analytics, and a preferred option for many scholars and practitioners because it is very user friendly. The software executes multi-group analysis in SmartPLS and can be used to examine observed heterogeneity, in which known groups are proposed to behave differently (e.g., female/male comparisons; large/small businesses), as well as unobserved heterogeneity (Sarstedt et al., 2017; Hair et al., 2016; Matthews et al. 2016, Hair et al., 2022a). Multi-group analyses (MGA) assist researchers in discovering if differences exist in the parameter estimates (e.g., outer weights, outer loadings, and parameter estimates) of pre-defined data groups (Matthews, Hair & Matthews 2018). In PLS-SEM, MGA is beneficial in efficiently examining moderation across multiple relationships (Hair, Sarstedt, Ringle & Mena 2012). Another increasingly popular software is R, which is widely used among statisticians and data scientists to execute data analysis. The software is open source and free, and increasingly applied (Hair et al., 2022b).

If moderators are unknown ahead of time, this type of analysis becomes more difficult. Often, clusters appear which have similar values for certain variables. These types of groups cannot be accurately defined by any particular demographic or descriptive data variable. This type of moderation can be difficult to predict or even identify. In these cases, prediction-based techniques should be used to identify segments that may be creating unexplained heterogeneity. In addition, several statistical software packages offer researchers the opportunity to further verify external validity and predictive relevance (www.smartpls.de, www.r-project.org; Schmueli, Ray et al. 2016).

Focusing on PLS-SEM, researchers have proposed the methodological application of prediction-oriented segmentation (PLS-POS) and finite mixture partial least squares (FIMIX-PLS) segmentation as methods to overcome prior limitations of identifying unobserved heterogeneity. Other heuristic techniques should be used when the heterogeneity found in these effects is sequential. This goes far beyond longitudinal analysis, and often this type of variance results from nested effects, due to the order of effects, seasonality, grouping, etc. Traditionally, HLM has been used to identify these in econometrics. In machine learning, Bayes theorem is used to overcome the cross-sectional nature of these situations. Bayes theorem assumes that error terms vary depending on the situation or order of the effects. Bayesian modeling uses what is known to predict the unknown. Furthermore, Markov Chain Monte Carlo (MCMC) modeling, such as Bayesian techniques, consists of algorithms focused on probability distributions for large multi-dimensional, hierarchical models and unknown parameters. The method is advantageous when a high level of uncertainty exists and historical information is scarce. But can also be used to investigate information across multiple sources of data (Rossi & Allenby 2003). Applications of Bayesian methods have increased in parallel with the digital transformation driven by internet use.

Marketing research, particularly academic, has welcomed the adoption of Bayesian techniques and applied the method to address many aspects of marketing (Rossi & Allenby 2003) such as advertising (Agarwal, Hosanager & Smith 2011), customer choice (Chung, Rust & Wedel 2009) and customer relationships (Netzer, Lattin & Srinivasan 2008). Although data is plentiful, digitalization has magnified the unknown in consistently evolving, dynamic environments.

Partial least squares structural equation modeling (PLS-SEM) is an essential element of marketing researchers' methodological toolbox. The PLS-SEM field has undergone extensive development in recent years, but marketing researchers need to continue adopting the most recent extentions (Sarstedt, Hair, Pick, Liengaard, Radomir, & Ringle, 2022). The developments include PLSpredict (Shmueli, Sarstedt, Hair, Cheah, Ting, Vaithilingam, & Ringle, 2019; Manley, Hair, Williams, & McDowell, 2021; Hair & Sarstedt, 2021b); confirmatory composite analysis (CCA) (Hair, Howard & Nitzl, 2020); mediation (Sarstedt, Hair, Nitzl, Ringle & Howard, 2020); and model comparison metrics (Hair 2021).

A parallel trend to the numerous advances in analytical techniques is an emphasis on quality (Hair, Moore & Harrison, 2022). Emerging analytical methods being applied in marketing research are useful for decision making only when the data used with the techniques is valid. Thus, data quality is a fundamental concern in their application. The absence of data quality manifests in several ways, including missing, incomplete, inconsistent, inaccurate, out of date, and duplicate data (Gudivada et al. 2017). For example, outliers in datasets can result in biased parameter estimates, missing data can easily lead to substantial prediction errors, and small datasets typically lower the power of a model (Bosu & McDonell 2019; Gudivada et al. 2017).

Another essential criterion to ensure the quality of research results is the training of the market research analysts. Consider, for example, a recommender system of an online retailer where a customer typically purchases for herself or himself, but occasionally also for third parties. Data analysts must have sufficient knowledge and experience to determine whether the automated statistical algorithms that determine what is normal and what is an anomaly are designed to identify valid patterns. If the analyzed data does not have a sufficient number of responses to represent each type of purchase situation, measures of the relative frequencies of the different types of purchases will not be valid, and the results will not provide an accurate understanding of what is "normal" and what is not. In addition, datasets often include excess zeros as part of the data collection process, referred to as zero-inflated count data models, and this can produce biased results when not correctly analyzed (Spriensma, Hajos, de Boer, Heymans, & Twisk, 2013; Favero, Hair, Souza, Albergaria, & Brugni, 2021).

The need to accurately monitor data quality becomes more challenging when combining multiple disparate sources of data, often a characteristic of secondary data, which need to be harmonized for a single analysis. Differences in data treatment, for example, in terms of identifying and dealing with outliers, imputation of missing values, or converting unstructured to structured data, can change analytical results when predicting outcome variables. Identifying such inconsistencies is very challenging when the data is obtained from various sources which use different procedures to ensure data quality. Moreover, the sampling method is of equal importance in the age of big data because sampling approaches with big data are seldom based on scientific criteria. Indeed, it is not uncommon for less knowledgeable analysts to not even apply sampling to their data. Rather, sampling is frequently based on social, political, economic, and technical factors that determine which data ends up in the final dataset (Leonelli, 2014).

While the "garbage in – garbage out" principle applies to all of data analysis, data quality concerns seem to be much less debated among marketing researchers using

machine learning and artificial intelligence. In contrast, research on data quality has a long history in computer science and management information systems. Data quality researchers in these fields have formed their own community (Madnick et al. 2009) with numerous intellectual contributions in bias pattern discovery, missing value treatment, and development of accuracy metrics, many of which are documented in the ACM Journal of Data and Information Quality. Insights from this community will play an increasingly important role in marketing research to ensure marketers recommendations are not based on data and analysis characterized by quality problems (Hair & Sarstedt, 2021).

Moore, Harrison, and Hair (2022) have proposed a six-step process to ensure data quality in marketing research. The steps include: (1) start with qualitative research to ensure rigorous research design, (2) design questionnaire and collect data, (3) screen and clean data, (4) apply correct analytical method, (5) interpret results to represent all perspectives, and (6) communicate the process undertaken to ensure data quality at all stages of the research process. Following a comprehensive data quality assurance process is particularly important since in many organizations the quality of data is often questionable, the presence of data quality problems is seldom known, and data quality frequently cannot be resolved after the data is collected. Since poor quality of data leads to incorrect analytical results, distorted or misrepresented interpretation of findings, and inaccurate predictions of the likely success or failure of marketing strategies and tactics, marketing researchers should consistently focus on minimizing its presence in their data.

4| Concluding Observations

As should be evident, much has changed in marketing and marketing research in the past few years. Many more changes are anticipated in the next few years. The following summarizes our thoughts about future developments. Predictions of the future are clearly risky. Many intelligent individuals have attempted to forecast the future – and often been proven wrong. Our thoughts are based on what we believe is sound reasoning, but like others in the past it is unlikely that our thinking will be entirely as it appears to us today. We do believe that we are mostly correct and ask that you view our thoughts as at least directional – as we are confident about the direction but not necessarily the specifics.

4.1 | Data Will Continue Increasing Exponentially

The first discussions of the rapid expansion of data began at least 70 years ago, well before the current widespread interest in big data and its impact. It was then when the first attempts to quantify the growth rate in data began, and for the next 50 or so years this trend was referred to as the "information explosion". At about the same time (1941), the Oxford English Dictionary for the first time included the term information explosion in its list, and many other data terms have been added since.

Among the first business authors to write about this phenomenon was Tom Peters in his 1991 book titled Thriving on Chaos. While his book predicted many broad economic changes occurring at that time, the phrase most relevant to big data was We are drowning in information and starved for knowledge. He went on to note that the difference in

information and knowledge is that information consists of only words and numbers that may be interesting, but they are not useful for decision-making. Moreover, the thinking in that day was that perhaps only five percent of all information was knowledge, and the challenge was to convert more of the available information to knowledge. Peters and others believed that would be possible if they created awareness of this problem, but they did not foresee the impact of the internet and particularly the emergence of digital data.

In the late 1990s the book titled How much information is there in the world?, by Michael Lesk (1997), concluded "There may be a few thousand petabytes of information all told; and . . . in only a few years . . . we will be able [to] save everything–no information will have to be thrown out, and . . . the typical piece of information will never be looked at by a human being." By this time scholars and practitioners were clearly aware not only of the huge amounts of data being produced, but also the importance of data analytics, and particularly the role of machine learning in analyzing the available data. Of course, the data scientists were also developing new terms in an effort to clarify how one could measure the large amounts of data, and began using words like terabyte, petabyte, exabyte, zettabyte, and finally yottabyte – with each succeeding term referring to an amount 1,000 times larger. According to a Forbes report, the total amount of information in 2013 was 4.4 zettabytes, and by 2020 the amount of data will reach 44 zettabytes (Kanellos, 2016). For the layman, 44 zettabytes has little meaning, but for the data scientist it clearly indicates a huge amount of data. Of course, this and other estimates are just informed predictions but the direction and amount of increase in data accumulation and storage is unquestionably correct.

As noted, we believe data will continue to increase exponentially, and this clearly will impact both market research practitioners and scholars. Practitioners must help organizations to understand and respond effectively to this glut of data. They must advise marketers on which data is relevant and which can be dismissed. Beyond that, marketing researchers will need to be more effective in informing marketing managers on how the data can be used, and what data will be most helpful. The bulk of data available now and in the future will be secondary and digital. This is in stark contrast to the past, in which marketing researchers collected primary data and controlled what type of data would be collected in hopes that it would be directly and immediately applicable to problems confronting their customers. With the increase in secondary data, marketing researchers must help organizations sift through the mountains of data and determine what is relevant and what is of little or no value. Moreover, going forward longitudinal data will be available to conduct retrospective analyses, as well as prepare more effective forecasts.

Marketing scholars must follow a similar path – determining and gaining access to secondary data that is relevant to their research and will enable them to publish in quality peer-reviewed journals. Academicians have thus far accessed and applied secondary and digital data in only a few instances. But there are increasingly substantial opportunities to access this type of data, particularly in marketing, and as data increases exponentially so will opportunities for academicians to pursue these opportunities. At the same time, marketing research texts must also undergo major revisions since most include only limited material on data analytics, or even on the explosion of secondary and digital data.

4.2 | Data Quality Will Improve

All aspects of both academic and practitioner research are being impacted by the availability of data and the advances in data analytics. One area impacted is data examination, and as data examination is implemented in organizations it will improve data quality and ultimately the success or failure of the research effort. While many researchers continue to operate with primary data, both qualitative and quantitative, many others increasingly are dealing with widely disparate data sources (e.g., customer-level data within multiple divisions of firms, social media and other digital data, locational data, etc.) that have different data structures and formats. Expanded data examination, often referred to as data cleaning, is necessary and will improve data quality and thus results.

Another evolving area, and one that is likely the most challenging of the big data era, is data management – organizing, cleaning and coding the data from multiple sources and in multiple formats to prepare for analysis. Many researchers think the primary question is which technique to use, and do not realize that the type of data often dictates the selection of the technology and analysis technique. In most instances, what precedes and dictates the appropriate analysis technique is the management of data from multiple sources, which can easily take 50 percent or more of the total time for completing a data analytics project.

Combining and managing data from multiple sources becomes complex very quickly. What may seem like a simple merging of data can be very difficult when the analyst must extract data from multiple databases, matching formats and timeframes and many other required adjustments. And since 90 percent of data is often unstructured, this task is clearly daunting. Therefore, academic and practitioner researchers in all fields will have to become "data managers" along with being data analysts.

4.3 | Data Analytics Will Improve

A primary force behind the widespread application of data analytics is the potential improvement in decision-making. Improved data analytics will benefit practitioners and academics, as well as governments and for-profit or non-for-profit organizations. But to realize these benefits analysts will need to be more involved in providing critical inputs on what data to collect, in what format, how to manage and combine the data, and ultimately selecting the best analytic technique.

Until recently academicians were slow in adopting and applying big data and data analytics in their teaching and research. Some disciplines in the sciences (biology, biomedicine, and neuroscience) are moving faster in adopting data analytics, but the social sciences are beginning to recognize the potential of more complex analytical methods. These enhanced analytics methods make it possible for additional types of research questions to be examined and new challenges to be overcome. As marketing research academics and practitioners become more aware of these opportunities, in both data and techniques, their utilization will increase, hopefully sooner rather than later.

4.4 | Decisions will Be More Knowledgeable

Both management and customer decisions will increasingly be knowledge-based. The era of big data has provided new and varied sources of data, and has placed additional requirements on the analytical techniques that must handle these data sources. There are, therefore, unique challenges facing today's analyst with the many issues in big data and analytics. We believe with an expanded focus on data challenges and data analytics, data scientists will be able to create awareness and motivate management to invest in these promising developments. If this happens, marketing research and business decisions in general will be based on better data and more effective analytical techniques, and thus be more knowledge based. Clearly, the organizations and individuals that move the quickest in these emerging areas will acquire and establish true competitive advantages in the marketplaces of tomorrow.

4.5 | Privacy Issues and Challenges

In his book The Assault on Privacy, Arthur Miller (1971) noted that "Too many information handlers seem to measure [an individual by] the number of bits of storage capacity [their] dossier will occupy." Today, in advanced economies data exists on everyone and has grown in exponential size due to digital records. Increasingly, however, government implemented policy changes are impacting the way we collect, access, store and use data. On the heels of the Cambridge Analytica's information security issues involving Facebook, two consumer privacy laws have been or are is in the process of being adopted – the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act, A.B. 375 were developed to provide greater control to customers of data that was collected on them. As a result, companies must now increase controls and transparency of how data is collected, what data is collected, why data is being collected, who has access to the data and how companies are using the collected data. Marketing researchers are also required to comply with these regulations as they face the same rules in collecting, securing, and distributing identifiable data.

4.6 | Data Analytics Capabilities Provide Substantial Competitive Advantages

Data is only as good as the intelligence we can glean from it, and that entails effective data analytics and a whole lot of computing power to cope with the exponential increase in the volume and type of data. Most all organizations (large, medium, and even small) can apply data analytics to improve manufacturing, supply chain, management, and marketing activities, and to be more efficient and effective in achieving organizational goals. While the application of data analytics is producing useful findings in the fields of agriculture, healthcare, urban design, crime reduction, and energy, along with business in general, much remains to be done. The field of analytics is evolving quickly and more opportunities are emerging to apply the new sources of data and the additional, more sophisticated methods of analysis. The speed of adoption and application to decision-making will clearly influence the acquisition of competitive advantages, but the leaders in this field will accrue the advantages, and their advantages could become insurmountable for the laggards.

4.7 | Demand for Trained Data Scientists Will Exceed Supply

More universities need to become involved in training new data scientists. Training and awareness of big data and data analytics will encourage analysts to define not only their research questions more broadly, but also the scope of their responsibility in leading the way to more effectively use data to solve business problems. The successful utilization of data analytics will be influenced substantially by the analysts and the decisions they make not only in informing senior managers of the potential of data analytics, but also in the selection and application of the techniques to use moving forward.

According to the Study Portals website, in the U.S. there are slightly more than100 bachelor's degree programs with a concentration in data analytics. This number needs to be five and perhaps ten times that many – and similar changes must follow globally. But an even greater need exists for master's degrees in data analytics (only about 40 are offered in the U.S.) since being a data scientist requires substantial training. Universities globally need to move quickly to train these data scientists to meet not only the current demand, but to continue producing what will be a substantial and expanding need for data scientists in the future, particularly in the field of marketing research. Institutions in the U.S. offering master's degrees in data analytics, https://www.hotcoursesabroad.com/study/training-degrees/us-usa/masters/data-analysis-courses/loc/211/slevel/57-3-3/cgory/cb.451-4/sin/ct/programs.html, accessed August 2018.

5| REFERENCES

Abubakar, A.M., Namin, B.H., Harazneh, I., Arasli, H., & Tunç, T. (2017). Does gender moderate the relationship between favoritism/nepotism, supervisor incivility, cynicism, and workplace withdrawal: A neural network and SEM approach. Tourism Management Perspectives, 23, 129-139.

Antons, D., & Breidbach, C. F. (2018). Big data, big insights? Advancing service innovation and design with machine learning. *Journal of Service Research*, *21*(1), 17-39.

Agarwal, A., Hosanagar, K., & Smith, M. D. (2011). Location, location, location: An analysis of profitability of position in online advertising markets. *Journal of marketing research*, *48*(6), 1057-1073.

Bachelor's Degrees in Data Science & Big Data in the United States, https://www.bachelorsportal.com/study-options/269779226/data-science-big-data-united-states.html, accessed August 2018.

Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 1-34.

Becker, J. M., Rai, A., Ringle, C. M., & Völckner, F. (2013). Discovering unobserved heterogeneity in structural equation models to avert validity threats. *MIS Quarterly*, *37*(3), 665-694.

Bosu, M. F., & MacDonell, S. G. (2019). Experience: Quality Benchmarking of Datasets Used in Software Effort Estimation. ACM Journal of Data and Information Quality, 11(4), Article No. 19.

Chan, H. Y., Boksem, M., & Smidts, A. (2018). Neural profiling of brands: Mapping brand image in consumers' brains with visual templates. *Journal of Marketing Research*. 55(4), 600-615.

Chung, T. S., Rust, R. T., & Wedel, M. (2009). My mobile music: An adaptive personalization system for digital audio players. *Marketing Science*, *28*(1), 52-68.

Davenport (2018) Will data scientist continue to be the sexiest job? Downloaded June 2018 at https://iianalytics.com/research/will-data-scientist-continue-to-be-the-sexiest-job.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48, 24-42.

Davenport, T., & Harris, J. (2017). *Competing on Analytics: Updated, with a New Introduction: The New Science of Winning*. Harvard Business Press.

Delen, D., & Zolbanin, H. M. (2018). The analytics paradigm in business research. *Journal of Business Research*, *90*, 186-195.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897-904.

Favero, Luiz Hair, J. F., Jr., Souza, Rafael de Freitas, Albergaria, Matheus, and Brugni, T. V. (2021). Zero-Inflated Generalized Linear Mixed Models: A Better Way to Understand Data Relationships, Mathematics, 9, 1100, https://doi.org/10.3390/math9101100.

Flynn, B., Pagell, M., & Fugate, B. (2018). Survey Research Design in Supply Chain Management: The Need for Evolution in Our Expectations. *Journal of Supply Chain Management*, *54*(1), 1-15.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137-144.

Gudivada, V., Apon, A., & Jing, J. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations". In: International Journal on Advances in Software 10.1 (2017), pp. 1 - 20.

Hair, J. F. (2021). Next Generation Prediction Metrics for Composite-based PLS-SEM. *Industrial Management & Data Systems*. 121(1), 5-11. https://doi.org/10.1108/IMDS-08-2020-0505 .

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning, U.K.

Hair, J. F., Harrison, D., & Ajjan, H. (2022). *Essentials of Marketing Analytics*, McGraw-Hill, N.Y., New York.

Hair, J. F., Howard, M., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research, 109*, 101-110. https://doi.org/10.1016/j.jbusres.2019.11.069 .

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2022a), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM),* 3rd ed., Sage, Thousand Oaks, CA.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., and Ray, S. (2022b). *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R*. Springer, New York.

Hair, J.H., Matthews, L., Matthews, R., & Sarstedt, M. (2017). PLS-SEM or CB-SEM: Updated Guidelines on Which Method To Use," (2017), *International Journal of Multivariate Data Analysis*, Vol. 1 (2), pp 107-123.

Hair, J.H., Matthews, L., Ringle, C., & Sarstedt, M. (2016). Identifying and Treating Unobserved Heterogeneity with FIMIX-PLS: Part I – Method, *European Business Review*, 28(1), 63-76.

Hair, J. F., Moore, Z., & Harrison, D. (2022). Data Quality Assurance Begins Before Data Collection and Never Ends: What Marketing Researchers Absolutely Need to Remember (2022). *International Journal of Market Research*, https://doi.org/10.1177/14707853211052183.

Hair, J., Risher, J., Sarstedt, M., & Ringle, C. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24.

Hair, J. F., & Sarstedt, M. (2021a). Data, Measurement, and Causal Inferences in Machine

Learning: Opportunities and Challenges for Marketing, *Journal of Marketing Theory and Practice*, 29(1), pp 65-77, doi.org/10.1080/10696679.2020.1860683.

Hair, J. F., & Sarstedt, M. (2021b). Explanation plus prediction – The logical focus of project management research. *Project Management Journal*. 52(4), 319-322. https://doi.org/10.1177%2F8756972821999945

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An Assessment of the use of Partial Least Squares Structural Equation Modeling in Marketing Research. *Journal of the Academy of Marketing Science*, 40(3), 414-433.

Harris, J. M., Ciorciari, J., & Gountas, J. (2018). Consumer neuroscience for marketing researchers. *Journal of Consumer Behaviour*, 17(3), 239-252.

Harrison, D.E., Ajjan, H. and Coughlan, A. (2018). Working paper: Understanding social media sentiment, positioning and engagement: the impact on direct selling sales performance. *Informs Society of Marketing Science Proceedings*.

Harrison, D. E., & Hair, J. F. (2017). The Use of Technology in Direct-Selling Marketing Channels: Digital Avenues for Dynamic Growth. *Journal of Marketing Channels*, 24(1-2), 39-50.

Henke, N., Levine, J., McInerney, P. (2018). You don't have to be a data scientist to fill this must have analytics role. *Harvard Business Review*. February 2005.

Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155-172.

Hulland, J., Baumgartner, H., & Smith, K. M. (2018). Marketing survey research best practices: evidence and recommendations from a review of JAMS articles. *Journal of the Academy of Marketing Science*, 46(1), 92-108.

Hulland, J., & Miller, J. (2018). "Keep on Turkin'"?. *Journal of the Academy of Marketing Science*, 46(5), 789-794.

Hult, T., Hair, J. F., Sarstedt, M., Ringle, C., Proksch, D., and Pinkwart, A. (2018). Addressing Endogeneity in International Marketing Applications of Partial Least Squares Structural Equation Modeling, *Journal of International Marketing*, forthcoming.

Humphreys, A., & Wang, R. J. H. (2017). Automated text analysis for consumer research. *Journal of Consumer Research*, 44(6), 1274-1306.

Kanellos, M. (2016). 152,000 smart devices every minute in 2025: IDC outlines the future of smart things. March 2016, retrieved from https://www.forbes.com/sites/michaelkanellos/2016/03/03/152000-smart-devices-every-minute-in-2025-idc-outlines-the-future-of-smart-things/.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15-25.

Krause, D., Luzzini, D., & Lawson, B. (2018). Building the Case for A Single Key Informant in Supply Chain Management Survey Research. *Journal of Supply Chain Management*, 54(1), 42-50.

Kull, T. J., Kotlar, J., & Spring, M. (2018). Small and Medium Enterprise Research in Supply Chain Management: The Case for Single-Respondent Research Designs. *Journal of Supply Chain Management*, 54(1), 23-34.

Kumar, V. (2018). Transformative Marketing: The Next 20 Years. *Journal of Marketing*. 82(4), 1-12.

Leonelli, S. (2014) What difference does quantity make? On the epistemology of Big Data in biology. *Big Data Soc* 1: 1–11.

Madnick, S. E., Wang, R. Y, Lee, Y. W., & Zhu, H. (2009) Overview and framework for data and information quality research. *ACM Journal of Data and Information Quality*, 1(1),

Article 2.

Manley, S. C., Hair, J. F., Williams, R. I., and McDowell, W. C. (2021). Essential new PLS-SEM analysis methods for your entrepreneurship analytical toolbox. *International Entrepreneurship and Management Journal*. 17(1), 1-21. DOI: 10.1007/s11365-020-00687-6

Matthews, L., Hair, J., Matthews, R. (2018). PLS-SEM: The holy grail for advanced analysis. *Marketing Management Journal*, *28*(1), 1-13.

Matthews, L., Hair, J.H., Sarstedt, M., and Ringle, C. (2016). Identifying and Treating Unobserved Heterogeneity with FIMIX-PLS: Part II – Case Study, European Business Review, 28(2).

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. *International Journal of Research in Marketing*, *37*(3), 481-504.

Mela, C.F. & Moorman, C. (2018). Why Marketing Analytics hasn't lived up to its promise. May 30, retrieved from https://hbr.org/2018/05/why-marketing-analytics-hasnt-lived-up-to-its-promise.

Moore, Z., Harrison, D., and Hair, J. F. (2021). Data Quality Assurance Begins Before Data Collection and Never Ends: What Marketing Researchers Absolutely Need to Remember (2022). *International Journal of Market Research*, 63 (6), https://doi.org/10.1177%2F14707853211052183.

Netzer, O., Lattin, J. M., & Srinivasan, V. (2008). A hidden Markov model of customer relationship dynamics. *Marketing science*, *27*(2), 185-204.

Roh, J. A., Whipple, J. M., & Boyer, K. K. (2013). The effect of single rater bias in multi-stakeholder research: A methodological evaluation of buyer-supplier relationships. *Production and Operations Management*, *22*(3), 711-725.

Rossi, B. (2015), From insight to action: why prescriptive analytics is the next big step for data. February 2015, retrieved from https://www.information-age.com/insight-action-why-prescriptive-analytics-next-big-step-big-data-123458977/.

Rossi, P. E., & Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science*, *22*(3), 304-328.

Sarstedt, M., Hair, J. F., Nitzl, C., Ringle, C. M. and Howard, M. C. (2020). Beyond a tandem analysis of SEM and PROCESS: Use PLS-SEM for mediation analyses! *International Journal of Market Research*, *62*(3), 288-299. https://doi.org/10.1177/1470785320915686.

Sarstedt, M., Hair, J. F., & Ringle, C. M. (2022). "PLS-SEM: Indeed a silver bullet" – Retrospective observations and recent advances*, Journal of Marketing Theory & Practice*, https://doi.org/10.1080/10696679.2022.2056488 .

Sarstedt, M., Hair, J. F., Pick, M., Liengaard, B., Radomir, L., &. Ringle, C. M. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade, *Psychology and Marketing*, 1-30.

Sarstedt, M., & Ringle, C. M. (2010). Treating unobserved heterogeneity in PLS path modeling: a comparison of FIMIX-PLS with different data analysis strategies. *Journal of Applied Statistics*, *37*(8), 1299-1318.

Sarstedt, M., Ringle, C., & Hair, J. (2017). Identifying and Treating Unobserved Heterogeneity in PLS: A State of the Art Review," chapter in Recent Developments on Partial Least Squares Structural Equation Modeling: Basic Concepts, Methodological Issues and Applications, H. Latan, and Richard Noonan, Editors, Springer International Publishing AG, Switzerland.

Shankar, V. (2018). How artificial intelligence (AI) is reshaping retailing. *Journal of retailing*, *94*(4), vi-xi.

Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models, *Journal of business Research*, 69, 4552-4564.

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J., Ting, H., Vaithilingam, S., & Ringle, C. M.

(2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*, *53*(11), 2322-2347. https://doi.org/10.1108/EJM-02-2019-0189.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263-286.

Sorescu, A., Warren, N. L., & Ertekin, L. (2017). Event study methodology in the marketing literature: an overview. *Journal of the Academy of Marketing Science*, *45*(2), 186-207.

Snyman, J. A. and Wilke, D. N. (2005). Practical mathematical optimization: basic optimization theory and gradient-based algorithms. Springer.

Varadarajan, R. (2018). A Commentary on "Transformative Marketing: The Next 20 Years". *Journal of Marketing*. 82(4), 15-18.

Verma, P., Agarwal, S., Kachroo, P., & Krishen, A. (2017). Declining transportation funding and need for analytical solutions: dynamics and control of VMT tax. *Journal of Marketing Analytics*, *5*(3-4), 131-140.

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, *80*(6), 97-121.

White, S., (2017). What is a data scientist? A key data analytics role and a lucrative career. https://www.cio.com/article/3217026/data-science/what-is-a-data-scientist-a-key-data-analytics-role-and-a-lucrative-career.html, accessed August 2018.

Worm, S., Bharadwaj, S. G., Ulaga, W., & Reinartz, W. J. (2017). When and why do customer solutions pay off in business markets?. *Journal of the Academy of Marketing Science*, *45*(4), 490-512.