Received Date: 26 April 2025 I Accepted Date: 06 May 2025

OPUS
Journal of Society Research

RESEARCH ARTICLE

From Hearing to Doing: Tracing the Gaps in Familiarity, Competency, and Use of Big Data Among Türkiye's Social Scientists*

Yaser Koyuncu[1], İlknur Yüksel-Kaptanoğlu[2]

[1] Research Assistant, Hacettepe University Institute of Population Studies, Department of Social Research Methodology, Ankara/Türkiye
ORCID: 0000-0002-5379-3804
E-Mail: yaserko@hacettepe.edu.tr

[2] Assoc. Prof. Dr, Hacettepe University Institute of Population Studies, Department of Social Research Methodology, Ankara/Türkiye
ORCID: 0000-0001-5969-1238
E-Mail: ilknury@hacettepe.edu.tr

Corresponding Author: Yaser Koyuncu

May 2025
Volume:22
Issue:3
DOI: 10.26466/opusjsr.1684496

Citation:
Koyuncu, Y. & Yüksel-Kaptanoğlu, İ. (2025). From hearing to doing: Tracing the gaps in familiarity, competency, and use of big data among Türkiye's social scientists . *OPUS– Journal of Society Research*, *22*(3), 368-390.

Abstract

This study investigates how academics in Türkiye's social sciences engage with big data by examining their familiarity, competency, and actual use. Drawing on a large-scale web survey of 3,606 academics, we analyze how individual backgrounds, methodological orientations, and institutional environments relate to engagement with big data. Descriptive and logistic regression analyses reveal that methodological proximity to quantitative research, openness to innovation, and institutional exposure through departmental curricula are key drivers of competency and use. Conversely, structural divides, such as inequalities between universities and socio-economic development levels, appear to affect familiarity more than deeper engagement. Theoretically, the study integrates frameworks including epistemic cultures, pedagogic device theory, diffusion of innovations, and technology acceptance models to contextualize the findings. This research contributes to the limited empirical literature on big data adaptation among social scientists outside the Global North and address structural, curricular, and attitudinal barriers and approaches for broader adoption in Türkiye's social research methodology.

Keywords: *Big data engagement, computational social sciences, social research methodology, higher education in Türkiye, innovation adoption*

Öz

Bu çalışma, Türkiye'deki sosyal bilimler alanında çalışan akademisyenlerin büyük veri ile nasıl ilişkilendiğini, aşinalık, yeterlik ve kullanım düzeyleri üzerinden incelemektedir. 3.606 akademisyeni kapsayan büyük ölçekli bir web anketine dayanarak, bireysel tecrübelerin, metodolojik yönelimlerin ve kurumsal etkilerin büyük veri ile etkileşimle nasıl ilişkilendiği analiz edilmiştir. Betimleyici analizler ve lojistik regresyon sonuçları, nicel araştırmaya metodolojik yakınlık, yeniliğe açıklık ve bölüm müfredatları aracılığıyla kurumsal düzeyde maruz kalmanın, büyük veri yeterliği ve kullanımının temel belirleyicileri olduğunu ortaya koymaktadır. Buna karşılık, üniversiteler arası eşitsizlikler ve sosyoekonomik gelişmişlik düzeyleri gibi yapısal ayrımların, daha derin düzeydeki etkileşimden ziyade aşinalığı etkilediği görülmektedir. Çalışma, bulguları bağlamlaştırmak amacıyla bilgi üretim kültürleri, pedagojik düzenek kuramı, yeniliklerin yayılımı kuramı ve teknoloji kabul modelleri gibi kuramsal çerçeveleri bir araya getirmektedir. Bu araştırma, Küresel Kuzey dışındaki sosyal bilimcilerin büyük veriye adaptasyonu konusundaki sınırlı ampirik literatüre katkı sağlamakta ve Türkiye'de sosyal araştırma metodolojisinde daha geniş çaplı bir benimsemenin önündeki yapısal, müfredatsal ve tutumsal engellere ve yaklaşımlara dikkat çekmektedir.

Anahtar Kelimeler: *Büyük veriyle etkileşim, hesaplamalı sosyal bilimler, sosyal araştırma yöntemleri, Türkiye'de yüksek öğretin, yenilikleri benimseme*

* *This article is derived from the PhD thesis prepared by Yaser Koyuncu at Hacettepe University Institute of Population Studies, Department of Social Research Methodology*

Introduction

The advancements in computer technology, digitalization, the increasing use of the internet, and the emergence of data science led to the establishment and expansion of computational social sciences, which was declared in a paper as a new discipline by nine scholars in 2009 (Lazer et al., 2009). This period is characterized by the pervasive trend of quantification, known as the data revolution, which posits that no domain will be untouched and that all facets will be transformed into data (King, 2011).

Big data, as a central concept of these advancements, is defined by its volume, velocity, and variety and presents an unprecedented opportunity to provide insights into social trends, behaviors, and issues. It is possible to view the data as "found" and "organic" in digital sources, and social scientists increasingly encounter "dynamic data" as opposed to the "static data" gathered by conventional methods (Veltri, 2020). That being said, from a methodological standpoint, integrating big data into social and humanitarian sciences has sparked significant debate on its potential benefits and inherent challenges. Big data has been cited as providing several advantages, including the capacity to evaluate extensive information in real-time, reveal concealed patterns, and produce insights that were previously inaccessible using conventional research methods. Additionally, it is argued that big data can enhance the depth of information generated by traditional methods, offering a "thick description" (Bjerre-Nielsen & Glavind, 2022) Also, utilizing computational science tools can enable social scientists to benefit from the new data landscape (Halford & Savage, 2017). Having an abundance of data at hand, *prima facie*, might be hype for researchers, considering that it reduces concerns for sampling and provides higher inclusiveness to analysis (Mazzocchi, 2015). However, some scholars have emphasized the extensive problems related to bias, population, inference, and representations compared to traditional methods (Japec et al., 2015; Marres, 2017). Further to that is the theory building challenges with big data, which require approaches that handle complex systems, multi-level phenomena and interdisciplinary data (Astleitner, 2024; Cabrera-Álvarez, 2022) as well as unique ethical problems stemming from using tools of STEM fields with social science methodologies that lead to ethical ambiguities (Hosseini et al., 2022; Stegenga et al., 2024).

One of the central concerns is how social researchers will adapt to these technologies, particularly given that working with such data necessitates additional skills, posing a significant obstacle for those without, *inter alia*, a strong foundation in mathematics, statistics, and programming languages. In line with this, Brady (2019) underscores the essential skills for handling large datasets, asserting that political scientists and other social researchers need to familiarize themselves with techniques like machine learning and data visualization. More particularly, qualitative researchers are argued to require the acquisition of these skills (Beuving, 2020; Sarkar, 2021). Although limited, several studies have investigated how researchers view and adapt to the changes brought about by big data. SAGE's white paper revealed that nearly half of 9142 researchers want to engage in big data research in the future, with access to commercial and proprietary data being a significant concern (Metzler et al., 2016). Moreover, Faveratto and colleagues (2020) reported feelings of "uncertainty and uneasiness" towards the big data phenomenon amongst Swiss and American researchers.

That being the case, the incorporation of big data into Türkiye's social sciences is predominantly underexamined, as this topic has just lately gained prominence. For instance, the Presidency of Strategy and Budget (PoSB) of Türkiye has recently referred to big data-driven policies and the need for increased capacity for detailed data and big data analysis in the last two strategic plans (PoSB, 2018; 2023). Although the recognition of big data and data mining in public institution papers is increasing, a comprehensive strategy has yet to be established, and practical implementation is constrained (Karaca, 2024; Köseoğlu & Demirci, 2017). While Koç University has established Türkiye's first and only graduate degree program in Computational Social Sciences (CSS), the number of programs and departments in this field is still limited. The results of Şallı's survey (2021) indicate minimal engagement among researchers in Türkiye, as 36% of them were unfamiliar with the notion of CSS, whereas 33.5% were aware of the concept but had not had the opportunity to learn these methods. Aytaç and Bilge (2021) reported that academics working in computer science-related fields had low interest in big data. Similarly, Bölükbaş (2021) examined master's and PhD theses and found that 73% of studies on big data were concentrated in the life sciences domain.

Against this background, we aimed to investigate how academics in the social sciences perceive big data, their level of adaptation to social research methodologies, and their perspectives on the resulting changes. Specifically, we analyzed academics' familiarity with big data, their competency in using it, and their use of it in research, considering how these factors alter based on their backgrounds and attributes.

1.1 Background

1.1.1 Big data's definition

As Veltri (2020) suggests, big data is rather an umbrella term, and research in this domain typically pertains to managing, manipulating, and generating insights from extensive datasets in the digital realm. The term was first flagged by NASA researchers M. Cox and D. Ellsworth (1997), and was later expanded conceptually through the three V's; volume, velocity, and variety, introduced by D'Laney (2001). Some scholars trace back the epistemic origins of big data to the social physics movement of 19th-century Europe, which focused on the large-scale statistical measurement of social variables that are believed to explain both the social and natural world, which is grounded in a larger epistemological standpoint (Barnes & Wilson, 2014). Today, big data is understood as more than just large-scale datasets. It generally is defined in terms of the "three V's", by large datasets that are continuously fed in diverse digital sources (Callegaro & Yang, 2018). Foster (2017) succinctly described big data as "anything too big to fit onto your computer." These extend to news stories, streaming video, and e-commerce platforms that have "trace data" of people (Ignatow, 2020).

Some scholars have expanded the original V's by adding new dimensions. By way of example, Lukoianova and Rubin (2014) introduced veracity to describe inconsistency and uncertainty. Kitchin and Ardle (2016) also proposed seven characteristics of big data: volume, velocity, variety, exhaustivity, resolution, indexicality, extensionality, and scalability, arguing that volume and variety are not essential and that the three V's rules are erroneous. Instead, they emphasized that velocity and exhaustivity are the key boundary markers of big data. Despite the absence of a universally accepted definition of big data (Connelly et al., 2016; Veltri, 2020) and the reluctance of some scholars to employ the term due to the evolving nature of what constitutes "big" (Davenport, 2014), it can be posited that the most comprehensive interpretation of big data is that it is "organic" and "found" within the digital landscape, generated continuously without direct intervention.

Florescu and colleagues (2014) provide a valuable approach for differentiating big data from survey and administrative data. They assert that the main distinction lies in their aims; surveys and administrative data are explicitly designed for research and monitoring, whereas big data is "organic" and not generated with research objectives in mind. Moreover, classical statistical methods are frequently insufficient for big data, which is predominantly unstructured.

It is also essential to note that 'big data' may have different meanings across various disciplines. Faveratto and colleagues (2020) suggest a tentative approach to broad generalizations, positing that understanding its constituents is more important.

1.1.2 Methodological Implications of Big Data

On the one hand, the growing use of big data in social sciences offers advantages over traditional labor-intensive methods; on the other hand, it also brings forth specific challenges. It introduces obstacles such as issues related to data complexity, representativeness, the epistemological divide, ethics and legal gaps, and the need for new skills.

First, the massive amounts of data, which are often unstructured, diverse, and highly varied, pose significant challenges for researchers. Analyzing this data requires advanced technological tools and additional skills. While big data research is often less expensive than surveys at the outset, the costs of cleaning and processing become significant at the back end (Cai & Zhu, 2015; Japec et al., 2015).

Additionally, a significant concern is the representativeness of big data; conventional research methodologies rely on established theories for population inference, whereas big data provides insights into users of specific platforms. This constraint is associated with the concept of digital bias, which refers to the selective characteristics of digital environments that appeal to specific individuals or groups, thereby posing issues with the validity and generalizability of research outcomes (Marres, 2017). Research on social media encounters the same methodological challenges, particularly in terms of representativeness, the presence of trolls, and fraudulent accounts. Ruths and Pfeffer (2014) highlight these biases, which can skew data and thereby complicate researchers' ability to derive accurate results. In this environment, researchers must also address the accumulation of noise and the potential for misleading correlations, where unrelated variables may appear connected due to the extensive volume of data.

Furthermore, there might be discord between social and computer scientists, which stems from the divergent approaches of social science and computational models. Although computational tools are sometimes likened to telescopes or electron microscopes for societal analysis, this analogy may not be entirely appropriate for the social sciences, given the variability of human behavior across different temporal and spatial contexts (Lazer et al., 2021). Social scientists focus on elucidating behavior, whereas computer scientists emphasize the precision of prediction (Hofman et al., 2021). This illustrates the dichotomy between prediction and explanation, with computer scientists prioritizing model complexity for enhanced forecasts while social scientists uphold substantial theories (Agrawal et al., 2020).

Another essential aspect of the matter is the infrastructure and access to big data. There is a divide and inequality in infrastructure, skills, and networks for utilizing and employing big data, particularly in the developing and underdeveloped world (Serra, 2016). Further, there are limited opportunites for social researchers to access big data, reaching funds from private and public institutions that serve data structures equivalent to that of technological conglomerates such as Google or Facebook; thus, access to big data is limited to using text mining, web scraping, and API for social researchers (Diaz-Bone et al., 2020). Specific sites, such as X (formerly Twitter), have significantly restricted access to their data, making it more difficult than previously. At the same time, some scholars addressed an urgent need for action in view of the disappearing access to platform data (Parry, 2024).

Moreover, big data presents a multitude of ethical issues, as it may compromise people's privacy and have surveillance implications (Howe III & Elenberg, 2020). Tech players like Google and Facebook have been involved in scandals due to privacy breaches (Chen & Quan-Haase, 2020). On top of that, there is no global legal framework or common law enforcement mechanism for big data and its analytics (Japec et al., 2015). The potential for privacy violations is significant, as large-scale data collection may infringe on individuals' rights without their consent (Howe III & Elenberg, 2020).

As previously indicated, social researchers may require new skills to comprehend and interpret various data sources, which can present a significant challenge, especially for individuals without a quantitative or statistical background. In that vein, qualitative researchers, as some scholars have pointed out, should also acquire skills in big data analytics and statistics, including proficiency in Python or R, as integrating computational models with qualitative methods in mixed-methods frameworks is likely to become a prevalent trend in the future (Beuving, 2020; Sarkar, 2021). For illustration, the Towards Data Science platform identifies the ten essential skills for data scientists seeking employment as probability and statistics, multivariate calculus and linear algebra, data wrangling, data visualization, machine learning/deep learning, cloud computing, Microsoft Excel, DevOps, and programming languages such as Python, R, SQL, Java, and MATLAB.

The aforementioned methodological discussions, encompassing epistemological tensions, issues of representativeness, ethical infrastructure, and access problems, as well as skill-related obstacles, highlight essential elements that influence how scholars adapt to big data methodologies. Nevertheless, little empirical attention has been given to academics' actual familiarity, competencies, and usage patterns, particularly within the Turkish academic community. Our study specifically addresses the variations associated with their individual background and attributes.

Methodology

2.1 Survey Design

This study was conducted as part of mixed-methods research within a PhD thesis. In this paper, we will discuss the results of the quantitative phase of the study. For that, through a web survey, we aimed to gather data from scholars working in the social, humanities, and administrative sciences in Türkiye. The questionnaire, formed using the SurveyMonkey platform, consisted of 10 sections and 48 questions. These sections aimed to explore respondents' attitudes, experiences, and views on big data, covering demographics, research background, methodological skills, big data competency, integration into curriculum, plans, and perceived pros and cons.

2.2 Target Frame, Ethics, and Data Collection

The survey is applied between November to December 2022. The principal source for gathering contact lists was the Turkish Council of Higher Education's (YÖK) Academic database, an open-access website that offers contact information for academics in Türkiye. We opted not to utilize sampling and instead distributed the survey to all individuals on our contact list, comprising academics in the Social, Administrative, and Humanities fields, due to our concern regarding low response rates in web surveys (Daikeler et al., 2020). The Hacettepe University Social and Humanitarian Sciences Ethics Committee approved this study.

The survey is sent using Hacettepe University's bulk e-mail system with an invitation text directed to the web survey link. 31,067 academics received the survey with two reminder rounds, and 3,606 individuals completed the questionnaire, with a 11,6 percent of the response rate. Figure 1 shows the profile of respondents and target frame based on their academic title and type of university, which is to say, they are broadly similar, except Professors having higher share (16.5% vs. 12.5%) and Associated Professors are underrepresented (13.7% vs. 18.4%) while the university type breakdown closely mirrors the target frame both in public and private university type. Although this study did not aim to generate a fully probabilistic representative sample in the statistical sense, these figures suggest a reasonable degree of alignment based on academic title and university type, and the fact that academics have internet access can help to mitigate concerns about undercoverage in web surveys addressed by Bethlehem (2010).

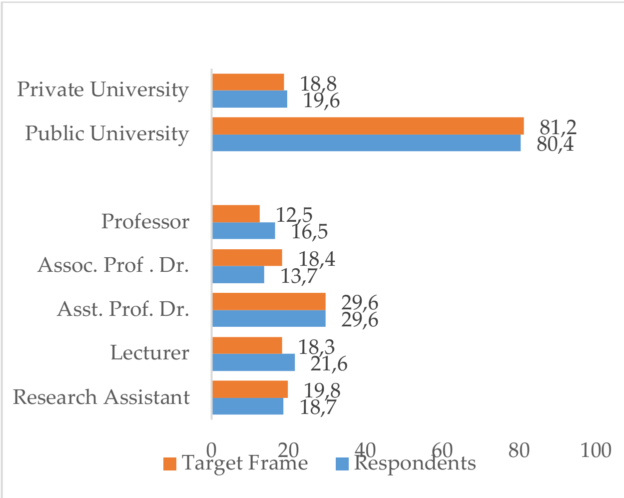

Figure 1. Title of respondents versus academics in the target frame (%)

2.3 Data Preparation and Variables

In the data analysis preparation phase, we recoded the data, focusing on open-ended responses and creating new variables. This process involved excluding unreliable cases based on conflicting information, such as mismatches between age, academic title, and years of experience. We also grouped the academic disciplines into categories using YÖK references. Additionally, we created new variables, such as a regional development index based on Socio-Economic Development (SEDP), and then employed a factor analysis using the attitude towards big data questions. The analysis was based on 12 Likert-scale statements with response options ranging from "strongly agree" (5) to "strongly disagree" (1), addressing the influence of big data in social research methodology, adaptation to it, and its integration with research methods. To ensure consistent directionality, all items were re-coded so that higher values reflected more positive attitudes or a stronger belief in change. We opted for the Direct Oblimin rotation method to account for potential correlations between factors, as the statements were conceptually related and not wholly independent. Sampling adequacy was confirmed through the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity[1], ensuring suitability for factor analysis, while the final model yielded three components with Eigenvalues exceeding 1, accounting for 51.6% of the total variance. As presented in Appendix - Table 3, the first five variables are associated with Factor 1, and the next four are negatively associated with Factor 2. We labelled the first group as *Openness to Big Data* and the second group as *Belief in the Impact of Big Data* (Openness). The remaining two variables are *Efficiency* and *Qualitative Research's Integration* (Efficiency and Integration). The Openness factor captures receptiveness to integrating big data into research, Impact factor measures perceptions of big data's transformative potential—ranging from acceptance to skepticism (with negative factor loadings, lower negative factor loadings reflect retention of traditional methods rather than adopting big data-driven changes), and the Efficiency and Integration factor highlights beliefs about big data's role in streamlining data collection, particularly in terms of time and cost, as well as the essentiality of qualitative research in big data research. Factor scores were calculated for each respondent and categorized into low, medium, and high (strong) levels to denote differing degrees of the categories. These categories were formed by distributing values equally across tertiles to ensure an equitable presentation of respondents at each level.

2.4 Descriptive and Multivariate Analyses

For the analysis, we first conducted a descriptive examination and ran chi-square tests to determine whether the observed differences were statistically significant at $p<0.05$. In this paper, we mainly focus on the statistically significant differences.

Alongside the descriptive analysis, we performed three regression analyses. Specifically, we utilized a binary logistic regression to examine familiarity with big data, a multinomial regression analysis to assess competency in big data, and a binary logistic regression to explore the use of big data. To eliminate highly correlated variables, we also ran a multicollinearity test for each model; only one variable's VIF value was over 5 in the multinominal logistic regression model, leading to its exclusion (Big Data's involvement in master's degree Curriculum). Reference values are determined based on the descriptive results, with those having the lowest familiarity, competency, and use across characteristics.

We added certain variables in addition to those indicated in the descriptive results. To examine predictors of familiarity with big data, apart from variables used in the descriptive analysis, we included the binary variable "Having a PhD." Additionally, based on responses regarding quantitative and qualitative data collection practices, we established binary variables indicating whether data were gathered from digital environments (e.g., social media or communication apps) for each research type. From questions on analysis tools, we derived "Used software based on programming languages (e.g., R, Python) for quantitative analysis" and a composite variable "Use of computer for qualitative analysis," which included all digital tool use except manual analysis.

We incorporated independent variables addressing curricular and attitudinal factors for the competency model. These included the importance attributed to big data in the departmental curriculum and the presence of big data–related courses at the undergraduate, master's, and PhD levels. Attitudinal predictors encompassed the factor analysis variables: "Openness", "Impact" and "Efficiency and Integration."

In the model examining the use of big data, we incorporated all predictors from previous models. We added "Competency in Big Data" as an inde-

[1] *Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy = 0.761; Bartlett's test of sphericity: $\chi^2(66)$ = 6368.11, p < .001.*

pendent variable to assess whether higher competency levels predict actual application in research. (See Table 1 below.)

Table 1. Regression Models, Variables Used, and R^2 Values

Variable Type	Model 1: Binary Logistic (Familiarity)	Model 2: Multinomial Logistic (Competency)	Model 3: Binary Logistic (Use)
Dependent Variable	Familiarity (0 = Not Familiar (Reference), 1 = Familiar)	Competency (0 = Not at all (Reference), 1 = Low, 2 = Medium–High)	Use (0 = No (Reference), 1 = Yes)
Sociodemographic	Sex, Age, Title, PhD, SEDP	✓	✓
Institutional/Field	Academic Field, University Type	✓	✓
Methodological Practice	QUAN / QUAL / MM usage + exclusives	✓	✓
Digital Practice	Digital data collection, Software use	✓	✓
Self-Reported Competency	QUAN, QUAL, MM competencies	✓	✓
Curriculum Background	No. of RM courses (UG / Master's / PhD)	✓	✓
Big Data in Curriculum	–	Importance + Courses (UG / Master's / PhD)	✓
Attitudes Toward Big Data	–	Openness, Impact, Efficiency and Integration	✓
Competency in Big Data	–	(DV)	✓
Nagelkerke R^2	0.313	0.477	0.429

3. Analysis

3.1 Overview of Familiarity, Competency, and Use

To assess familiarity with big data, academics were asked whether they had heard of the concept, and 77% responded affirmatively. These 2,712 respondents were then asked to rate their competency in conducting big data analyses: 37.6% reported no competency, 32.5% low, 25.4% medium, and 4.5% high. The third measure focused on actual use. Those who rated themselves as at least minimally competent (n = 1,692) were asked whether they had conducted analyses using big data in their research. Of these, 35.8% reported using big data, while 64.2% did not (See Appendices, Table 4).

3.2 Differences by Sex, Age, and Academic Title

Sex: Male academics report slightly higher familiarity and self-assessed competency in big data. Specifically, they are more likely to rate themselves at high (5.2%) and medium (28.3%) competency levels compared to females (3.5% and 22.4%, respectively). Female academics, on the other hand, are more likely to rate themselves in the "Not at all" category (43.4% vs. 32.1%).

These differences suggest a potential gender gap in engagement and confidence related to big data.

Age: Familiarity with big data peaks among academics aged 40–49 (79%) and declines in the 60+ group (66%). However, older respondents are more likely to report higher competency levels. For example, 8.7% of academics aged 60+ rate themselves as highly competent, compared to just 2.7% among those aged 18–29. Similarly, the proportion reporting "Not at all" competency is highest in the youngest group (41%) and lowest among the oldest (30%).

Academic Title: Familiarity and competency also vary by academic title. Professors and Associate Professors report the highest familiarity (nearly 80%), while Lecturers have the lowest (69%). Higher titles are also associated with increased self-assessed competency. Full Professors have the highest proportion of "High" competency (8.7%) and the lowest proportion of "Not at all" (30.1%), while Research Assistants show the inverse pattern (3.3% and 41%, respectively) which may reflect accumulated experience, exposure to recent trends, or greater confidence among senior academics (See Appendices, Table 4).

3.3 Variations in Disciplines, Regions, and Types of Universities

Disciplinary Differences in Familiarity and Competency: The disciplines that are most familiar with the big data concept are Marketing and Advertisement (95%), Communication (91%), and Management and Administration (85%) (Table 2). In contrast, academics primarily interested in History and Archeology (32%) have the lowest percentage of those who encountered the concept, followed by Educational Sciences (60%). Political Science (71%), and Sociology and Cultural Studies (75%). Generally, fields closely tied to quantitative data and business have more familiarity among their participants, while more conventional disciplines exhibit less awareness, suggesting that big data-related practices are possibly less incorporated in those fields (See Appendices, Table 5).

Institutional Differences and SEDP: Academics from private universities are more likely to report their familiarity with big data (83%) than their counterparts in public universities (76%). This might be attributable to private universities' higher financing or technological resources supporting data-driven studies. Moreover, it may be posited that academics in more socio-economically developed provinces are more familiar with big data since 82 percent of the individuals living in Level 1 provinces are acquainted with the concept, whereas the percentage declines to 63 percent for Level 6. This may indicate inequalities in access to training, technology, and educational materials or inadequacies of the curricula for different levels of socioeconomic development.

Awareness–Competency Gaps in Data-Rich Disciplines: Some disciplines show a noticeable gap between familiarity with big data and feeling confident using it. For example, although nearly all academics in Marketing and Advertisement (95.4%) have heard of the concept, three out of four report low or no competency. This could suggest that while big data is widely discussed or used in the field, perhaps through tools built into marketing platforms, many academics might not have hands-on experience or formal training in how it actually works. A similar pattern appears in fields like Communication and Management, where professionals are often exposed to data in practice but may not be as equipped with the technical skills needed to work with it more deeply. That being said, the differences between the type of university (public-private) and the socioeconomic development of the province are not statistically significant according to the chi-square test for competency in big data.

3.4 Methodological Approaches and Big Data Engagement

To explore the relationship between methodological orientation and big data engagement, we asked academics which research approaches they had used—quantitative, qualitative, or mixed methods—with the option to select more than one. Familiarity with big data was significantly higher among those using quantitative methods (83%) than those who did not (64%) (Table 3). This difference was smaller for qualitative approaches (79% vs. 74%), while mixed methods users reported familiarity levels (80%) similar to those of quantitative users. However, interpretations of "mixed methods" likely varied, as no definition was provided, which is also valid for other methodological approaches (See Appendices, Table 6).

Focusing on respondents who used only one approach reveals sharper contrasts, as just 18% of those using only quantitative methods were unfamiliar with big data, compared to 34% among qualitative-only users. Mixed methods users again fell between (30%), suggesting greater exposure among those trained in quantitative traditions.

This pattern extends to competency and use. Researchers using only quantitative approaches reported the highest levels of competency (7.1% high, 27.3% medium), while qualitative-only users had the lowest, with over half (50.6%) indicating "not at all." Big data use followed a similar trend: 34.4% of quantitative-only respondents had used big data in their work, compared to 26.5% of qualitative-only academics. Mixed methods users again occupied an intermediate position.

We also examined how self-rated methodological competency aligns with big data familiarity and use. A clear positive trend emerged for quantitative skills: 51% of those with "very low" quantitative competency were familiar with big data, rising to 93% among those with "very high" skills (See Appendices, Table 7). This relationship was less consistent for qualitative competency, with familiarity dipping at medium and high levels.

Quantitative competency also appeared to drive actual use. Only 8% of those with very low quantitative skills used big data, compared to nearly 20% with very high skills. In contrast, researchers with low qualitative competency reported higher big data use (around 29%) than their quantitative counterparts, which may reflect differing self-assessments, where researchers perceive qualitative skills as more intuitive and rate themselves more generously, while viewing quantitative skills as more technical and more challenging to master.

Overall, the findings highlight the central role of quantitative training in big data engagement. One can argue that while familiarity is widespread, meaningful use and competency remain firmly tied to methodological background, particularly in quantitative methods.

3.5 Self-Assessed Competency and Actual Use

We also examined the relationship between self-assessment of competency in big data and the actual use of big data. Among those who regard themselves as highly competent in analysis using big data, 86.8 percent have used big data, while almost half of the medium-competent group did not use it (See Appendices, Table 7). Notably, 14.4% of those who rate themselves as low-competent have still used big data, suggesting that some academics engage with big data, albeit with limited self-perceived expertise.

3.6 Curricular and Institutional Influences

To explore the relationship between departmental curricula and big data engagement, we asked respondents how many courses related to research methodology were offered in their departments—at the undergraduate, master's, and doctoral levels. Familiarity with big data increased consistently with the number of available courses. For example, at the undergraduate level, familiarity rose from 74% among those with no methods courses to 80% among those reporting multiple courses. This pattern was even stronger at the master's level (61% vs. 82%), and similarly present at the doctoral level (68% vs. 78–82%) (See Appendices, Table 8).

Competency in big data also rose with curricular emphasis on methodology, particularly at the undergraduate and master's levels. However, this association was not statistically significant at the PhD level. In contrast, the number of methodology courses alone did not significantly affect the actual use of big data (See Appendices, Table 8).

When asked to what extent their department's curriculum emphasizes big data analysis, a strong association emerged: over half of those in departments placing "high" or "very high" importance on big data reported having used it (50.5% and 61.5%, respectively).

We also assessed whether big data is addressed in the curriculum through a four-tier classification: no coverage at all, no course but coverage in other courses, one dedicated course, or multiple dedicated courses. Results indicate that 73.2% of academics in departments with multiple dedicated undergraduate-level big data courses reported at least medium competency. Even minimal curricular reference to big data was associated with lower "Not at all" or "Low" competency rates (See Appendices, Table 9).

Big data use followed a similar trend. Among those in departments offering multiple undergraduate-level big data courses, 65.5% had used big data in their work—compared to only one-third of those in departments with no relevant courses. Still, 35–38% of those with access to multiple courses across levels had not used big data, suggesting that while departmental curriculum plays an important enabling role, it is not sufficient on its own.

3.7 Attitude Factors

In addition to structural and methodological variables, we examined how individual attitudes toward big data relate to competency and use, where a clear pattern emerged regarding openness to big data. Academics who expressed high openness were significantly more likely to report both greater competency since 41% rated themselves at medium or high levels, and actual usage, with nearly half (46.4%) having engaged with big data in their research which suggests that openness may facilitate both the willingness and confidence to engage with emerging data practices. Conversely, the perceived impact of big data on social research methodology yielded less variation, given that familiarity, competency, and usage rates were relatively stable across low, medium, and high belief levels, and these differences were not statistically significant. This may indicate that simply recognizing the broader relevance of big data does not inherently translate into personal engagement or skill development (See Appendices, Table 9).

Attitudes about efficiency, including cost and data collection processes, and the compatibility of big data with qualitative research (also the need for a qualitative approach for big data) demonstrate a more gradual trend. We may argue that academics who strongly agreed that big data could be effectively integrated with qualitative approaches, which are a component of this factor, reported slightly higher competency and use rates. This pattern, thus, may reflect a growing recognition that big data is not limited to quantitative paradigms alone, but can also complement qualitative traditions. Besides, a positive outlook on big data's impact on research procedures is somewhat associated with higher competency and use.

Taken together, these findings underscore the role of attitudinal openness as a meaningful predictor of engagement with big data, while other belief dimensions may shape perceptions but are less directly tied to action.

3.8 Logistic Regression Results

Whereas descriptive analyses provided an overview of familiarity, competency, and use of big data and its relation to attributes like demographics, experience, and methods used, the multivariate analysis discussed in this section, which includes three models, will help clarify how these effects are retained while others are controlled.

3.8.1 Predictors of Familiarity with Big Data

Our binary logistic regression model for familiarity with big data, demonstrates that demographic and structural characteristics are at play when retaining the other variables. Firstly, male academics were significantly more likely to report familiarity with big data than their female counterparts (OR=1.56). Besides, younger academics, such as the age 18-29 age group, were much more likely to say that they know the concept of big data (OR=3.50) compared to the 60+ group (For the full summary, see Table 2).

Senior academics also appear to have more potential to have familiarity with the concept. Assistant professors, for instance, are 1.7 times more likely to hear the concept while professors had 3,25 times the odds of familiarity compared to their lecturer counterparts. However, it is important to note that the confidence intervals widen in these higher categories, which suggests greater variability and possibly smaller subgroup sizes.

Another apparent gradient increase can be observed in the province's socio-economic development (SEDP), given that academics working in more socio-economically developed provinces are more likely to report familiarity compared to the lowest SEDP variable (OR=3.07). The odds of familiarity were exceptionally high in fields such as Marketing and Advertisement (OR=28.98) and Communication (OR=20.94) compared to History and Archeology, which is the reference category. Although these values are statistically robust, their broader confidence intervals necessitate cautious interpretation, since they may indicate variation within fields or result from small sample sizes. Besides, fields like Sociology (OR=3.47) and Political Science (OR=3.50), Economics (OR=5.98) also have a greater likelihood of knowing the concept.

The survey had questions about data sources used separately for quantitative and qualitative research. Academics engaged in digital data for quantitative research, such as gathering data from social media, are markedly more likely to be familiar with big data (OR=1.36) whereas results do not show a significant association for digital data in qualitative research. We also focused on the association of familiarity and computer use in qualitative analysis, considering the arguable distance of qualitative researchers and the digital environment. Our model suggests that those who used computers/programmes for qualitative analysis (including software like Excel, Word, or NVivo, MaxQDA) are 1.5 times more likely to hear about big data (OR=1.45) compared to researchers who just manually analyzed qualitative data. In the quantitative analysis, an additional predictor was added to the model: using programming languages (e.g., Python, R) for quantitative analysis. This inclusion is grounded in the assumption that quantitative researchers typically maintain greater proximity to digital environments. Individuals who reported using these tools exhibited a fourfold likelihood of familiarity (OR=4.31) as opposed to non-users.

The model shows that self-assessed competency also matters to a meaningful degree. For the competency in quantitative methods, where "Very Low" is the reference category, those who rated themselves highly competent were approximately twice as likely to know the concept, while individuals who assessed themselves very highly competent were six times more likely to know the concept (OR=2.20, OR= 6.11) In contrast, medium and low competency levels did not yield statistically significant associations. Competency in qualitative methods, on the other hand, showed a weaker overall association: only the very highly competent group has a significantly greater likelihood to be familiar with the big data concept, relative to the very low competency group in qualitative research (OR=2.23)

That being the case, variables about the type of university (public vs. private), research approaches used (quantitative, qualitative, or mixed methods), and the presence of research method courses in the curricula did not exhibit significant associations with familiarity with big data.

3.8.2 Predictors of Competency in Big Data

The multinomial logistics regression helps to have a broader picture about the competency, as it is particularly worthwhile examining the predictors of incremental shifts in competency, both from the lowest (not at all) to a slightly higher level (low) and from the lowest to a more advanced level (medium/high).

For the medium/high category, male academics are significantly more likely to report themselves as medium/high competent than female academics (OR=1.62) where the reference is "not at all" competency. Unlike the model for the familiarity, those who used digital data for qualitative purposes were more likely to report medium/high competency (OR=1.81) than those who did not, while using digital data for quantitative research did not show a significant association. This pattern may suggest that engaging with digital data in qualitative contexts, where it is less prevalent, plays a more substantial role in how academics assess their own competency. In contrast, at this level, using digital data for quantitative research appears to no longer be a distinguishing factor (See Table 2).

Programming language use and quantitative competency have the strongest associations for the medium/high category which is to say, programming language users have 6.5 times more likelihood (OR=6.46) and those who view themselves as very highly competent in quantitative methods were associated with an odds ratio of 10.62, despite with a broad confidence interval. Interestingly, at this level alone across all models, medium, high, and very high self-assessed competency in mixed methods show strong and significant associations. This suggests that academics who consider themselves proficient in both quantitative and qualitative approaches, as mixed methods typically are thought to require, depending on how it is defined, are more likely to report higher competency in big data.

Institutional and curricular influences also appear important. The importance given to big data in the departmental curricula (very low, low, medium, high, and very high) has a solid association with the likelihood of reporting medium/high competency. For instance, academics who perceive the importance of big data as high in their departments are nearly eight times more likely to report themselves as medium/high competent than those in the "very low" category (OR=7.62) Although the number of courses related to big data in undergraduate degrees does not exhibit a significant association, and the master's degree variable was omitted due to high multicollinearity, the presence of PhD-level courses in the department is strongly associated with higher competency. Specifically, academics in departments offering several PhD courses on big data are over 14 times more likely to report medium/high competency (OR = 14.06) albeit with a relatively wide confidence interval. Even coverage of the topic without a direct course has a robust association (OR=1.73)

Attitudinal factors also play a role, as openness at medium and high levels increases the likelihood of competency in big data compared to the group with the lowest openness (e.g. High: OR=2.94). It is also worth noting, however, that several expected predictors including age groups, academic titles, most of the socio-economic development, academic fields, competency in qualitative research, research method courses, and attitudinal factors of impact and efficiency and qualitative integration did not have a significant association with the medium/high competency when other variables are controlled for.

Examining what differs from the medium/high level may be plausible for the low competency category. The results show that using programming languages for quantitative analysis, while significantly associated with competency, is less robust in distinguishing those with low competency from those without competency (OR = 2.31). This suggests that using such tools may not be as critical a differentiator at lower levels of self-assessed competency. In a similar vein, medium competency in quantitative methods has a significant association as opposed to the medium/high level (OR=2.42) meaning that individuals with moderate quantitative skills were still more likely to rate themselves in the category than "not at all".

Research method courses in the undergraduate curriculum have a relationship unlike other models, as having no course appears to decrease the chance of having low competency compared to no competency at all (OR=0.62) Also, having big data coverage without a direct course increases the likelihood by 1.5 times compared to no course (OR=1.47). Similarly, having one (OR=3.11) or multiple courses (OR=5.06) at the PhD level has a significant association, although not as much as the medium/high category. The presence of big data in certain curricula remained a consistent predictor, even in lower competency levels, suggesting that exposure to content in the departmental environment increases the likelihood regardless of its depth.

3.8.3 Predictors of Big Data Use

Finally, the model on actual use of big data highlights that, not surprisingly, competency in big data is the most decisive factor, as academics with high cometency had 40 times the odds of having used big data (OR=40.84). Even the medium competency group had nearly six times the likelihood (OR=5.75). Programming language use, using digital data for qualitative research, and having a receptive attitude to what big data brings to social research methodology, increases the likelihood of big data use. (See Table 2)

Notably; sex, age, academic title, academic field, type of university, having a PhD, research methods used, socio-economic development of the province does not show significance, which indicates that methodological and other factors that are not visible here may have more influence actual usage than background characteristics. Besides and intriguingly, Level 6 has almost 2.5 times of more odds compared to the Level 2 which are the only significant results, peripheralizing the importance of socio-economic development.

Furthermore, the results reveal that academics with medium (OR = 12.86) and low (OR = 15.43) competency in quantitative research methods are significantly more likely to report having used big data compared to those with very low competency.

Table 2. Logistic Regression Results (Summary Table) (Part 1)

Predictor	Familiarity Exp (B) [Confidence Interval]; Sig	Competency Medium/High Exp (B) [Confidence Interval]; Sig	Competency Low Exp (B) [Confidence Interval]; Sig	Use of Big Data Exp (B) [Confidence Interval]
Sex (Female ref)				
Male	1.56 [1.23–1.98]; p=0.000	1.62 [1.16–2.25]; p=0.004	1.56 [1.19–2.05]; p=0.002	n.s
Age Groups				
Age 18-29	3.50 [1.66–7.38]; p=0.001	ref	ref	ref
Age 30-39	2.73 [1.51–4.92]; p=0.001	n.s	n.s	n.s
Age 40-49	3.12 [1.78–5.48]; p=0.000	n.s	n.s	n.s
Age 60+	ref	n.s	n.s	n.s
Academic Title (Lecturer ref)				
Research Assistant	1.59 [1.09–2.31]; p=0.016	ref	ref	ref
Asst Prof Dr.	1.71 [1.13–2.59]; p=0.011	n.s	n.s	n.s
Assoc Prof Dr.	2.42 [1.52–3.86]; p=0.000	n.s	n.s	n.s
Prof Dr.	3.25 [1.87–5.67]; p=0.000	n.s	n.s	n.s
Socio-Economic Development				
SEDP L1	3.07 [1.95–4.85]; p=0.000	n.s	n.s	n.s
SEDP L2	1.84 [1.15–2.94]; p=0.011	ref	ref	ref
SEDP L3	2.03 [1.23–3.37]; p=0.006	n.s	n.s	n.s
SEDP L4	1.86 [1.12–3.11]; p= 0.018	2.72 [1.46–5.08]; p=0.002	n.s	n.s
SEDP L5	n.s	n.s	n.s	n.s
SEDP L6	ref	n.s	n.s	2.43 [1.11–5.35]; p=0.027
Academic Field				
History and Archeology	ref	3.25 [1.071–9.872]; p=0.038	n.s	
Educational Sciences	2.63 [1.24–5.58]; p=0.012	n.s	n.s	ref
Sociology & Cultural	3.47 [2.10–5.74]; p=0.000	n.s	n.s	n.s
Political Science	3.50 [2.21–5.52]; p=0.000	n.s	n.s	n.s
Finance/Banking	6.81 [3.84–12.06]; p=0.000	n.s	n.s	n.s
Economics	5.98 [3.45–10.38]; p=0.000	n.s	n.s	n.s
Communication	20.94 [10.07–43.53]; p=0.000	n.s	n.s	n.s
Management & Administration	7.82 [4.66–13.13]; p=0.000	n.s	n.s	n.s
Marketing & Advertisement	28.98 [10.64–78.93]; p=0.000	n.s	n.s	n.s
Tourism	3.46 [1.90–6.29]; p=0.000	ref	ref	n.s
Other Fields	2.82 [1.67–4.76]; p=0.000	n.s	n.s	n.s
Digital Data Engagement ("Did not use digital data" ref)				
Digital Data for QUAN	1.36 [1.01–1.83]; p=0.045	n.s	n.s	n.s
Digital Data for QUAL	n.s	1.81 [1.18–2.76]; p=0.006	n.s	1.55 [1.02–2.36]; p=0.038
Using Programming Languages ("Did not use prog lang" ref)				
Prog Lang for QUAN Analysis	4.31 [2.26–8.21]; p=0.000	6.46 [3.93–10.62]; p=0.000	2.31 [1.44–3.69]; p=0.000	2.40 [1.59–3.64]; p=<.001
Using Computers for QUAL Analysis (Manual analysis ref)				
Computers for QUAL Analysis	1.45 [1.00–2.10]; p=0.047	n.s	n.s	n.s
Competency in QUAN Methods (Very Low ref)				
Comp QUAN High	2.20 [1.22–3.96]; p=0.009	5.94 [1.82–19.39]; p=0.003	4.29 [1.80–10.22]; p=0.001	n.s.
Comp QUAN Very High	6.11 [2.57–14.53]; p=0.000	10.62 [2.99–37.77]; p=0.000	4.04 [1.53–10.70]; p=0.005	n.s.
Comp QUAN Medium	n.s	n.s	2.42 [1.05–5.54]; p=0.037	12.86 [1.23–134.56]; p=0.033
Comp QUAN Low	n.s	n.s	n.s	15.43 [1.38–172.12]; p=0.026
Note. Odds ratios with 95% CIs; n.s. = p>.05; — = not in model.				

Table 2 (continued). Logistic Regression Results (Part 2)

Predictor	Familiarity Exp (B) [Confidence Interval]; Sig	Competency Medium-High Exp (B) [Confidence Interval]; Sig	Competency Low Exp (B) [Confidence Interval]; Sig	Use Exp (B) [Confidence Interval]; Sig
Competency in QUAL Methods				
Comp QUAL Very Low	ref	ref	ref	n.s
Comp QUAL Medium	n.s	n.s	n.s	ref
Comp QUAL Very High	2.23 [1.05–4.74]; p=0.037	n.s	n.s	n.s
Competency in MM Methods (Very Low Ref)				
Comp MM Medium	n.s	2.66 [1.17–6.08]; p=0.020	n.s	n.s
Comp MM High	n.s	3.65 [1.48–8.97]; p=0.005	n.s	n.s
Comp MM Very High	n.s	6.74 [1.89–24.06]; p=0.003	n.s	n.s
Competency in Big Data Use (Low ref)				
Comp BD Medium	—	—	—	5.75 [4.04–8.18]; p=<.001
Comp BD High	—	—	—	40.84 [16.35–102.02]; p=<.001
Importance Given to Big Data (Very Low ref)				
Imp BD Low	—	2.09 [1.36–3.22]; p=0.001	2.08 [1.49–2.90]; p=0.000	n.s
Imp BD Medium	—	4.73 [2.97–7.54]; p=0.000	2.70 [1.83–4.00]; p=0.000	n.s.
Imp BD High	—	7.62 [3.78–15.36]; p=0.000	3.04 [1.59–5.80]; p=0.001	n.s
Imp BD Very High	—	4.15 [1.71–10.06]; p=0.002	n.s	n.s.
Research Method Courses in the Undergrad Curriculum				
RM Course None	ref	n.s	0.62 [0.40–0.96]; p=0.031	ref
RM Course One	n.s	ref	ref	n.s.
RM Course Multiple	n.s	n.s	n.s	n.s.
Big Data in Curricula (Undergrad)				
BD UG No course	—	ref	ref	ref
BD UG No course but coverage	—	n.s	1.47 [1.02–2.13]; p=0.040	n.s.
BD UG One course	—	n.s	n.s	n.s.
BD UG Multiple courses	—	n.s	n.s	n.s.
Big Data in Curricula (PhD)				
BD PhD No course	—	ref	ref	ref
BD PhD No course but coverage	—	1.73 [1.10–2.72]; p=0.017	n.s	n.s.
BD PhD One Course	—	8.34 [4.12–16.88]; p=0.000	3.11 [1.62–5.94]; p=0.001	n.s.
BD PhD Multiple Courses	—	14.06 [4.75–41.60]; p=0.000	5.06 [1.77–14.48]; p=0.003	n.s.
Attitude Factors (Low ref)				
Openness Med	—	1.83 [1.23–2.72]; p=0.003	1.55 [1.14–2.13]; p=0.006	—
Openness High	—	2.93 [1.96–4.37]; p=0.000	1.42 [1.02–1.99]; p=0.038	1.64 [1.09–2.49]; p=0.018
Note. Odds ratios with 95% CIs; n.s. = p>.05; — = not in model, ref=reference value				

However, those who rate themselves as having high or very high competency do not show a statistically significant association. This appears to be an unexpected pattern, and wide confidence intervals imply a degree of uncertainty and variability that should be interpreted carefully. It may be because individuals with low or moderate skills are more inclined to use automated tools than the very low-skilled groups, while academics with higher skills are more selective in conducting these types of studies.

4. Discussion and Conclusion

Our study attempted to examine how scholars in Türkiye's social, administrative sciences, and humanities perceive, receive, and engage with big data. The findings indicate the significance of not only demographic and practical factors but also epistemological and institutional dimensions, as revealed through descriptive and multivariate analysis.

Firstly, it can be argued that methodological proximity, particularly the competency in quantitative research, strongly predicts the familiarity, competency, and use of big data, besides descriptives that illustrate the closeness of quantitative research users vis-à-vis qualitative researchers, especially for those who employ only quantitative frameworks. The big data field appears more readily aligned with the quantitative paradigm, as the data revolution is characterized by the "march of quantification", wherein big data is expected to permeate all sectors (King, 2011). What is also related to that is the skills needed for big data, which are closely tied to the quantitative approach, as several scholar addressed, mostly about their suspicions about integration of it to the social sciences. In that context, Manovich (2011) raised concerns about the data-driven social sciences and humanities, given the lack of expertise in questions that can arise in computer science, statistics, and data mining. More to the point, Kitchin has addressed the quantitative techniques, inferential statistics, modeling, and simulation inherent in big data approaches, while advocating for a critical social science outlook for the field (2014). It may be posited that academics within a "quantitative" epistemic culture, within the context of what Knorr-Cetina (1999) defines as having field-specific knowledge-making practices in the scientific realms, are more predisposed to engage with big data methodologies, beyond the skills they possess. In line with that, our descriptive data further indicates that scholars in "more" quantitative disciplines like finance, banking, economics, management, and administration exhibit greater familiarity and self-assessed competency than their counterparts in sociology, cultural studies, history, archaeology, and political sciences which may often prioritize contextual understanding and subjectivity which may not align with the scale and abstraction of big data. However, multivariate analysis demonstrated that the academic field's predictive power remained significant only for familiarity when controlling for other variables, suggesting that other factors mediate the competency and use. In this context, building on the epistemic culture framework, the robustness of the openness factor that captures receptiveness to the methodological implications of big data, as a predictor of both competency and actual use, may imply that discipline-embedded cultural dispositions beyond training and access shape the tendency to engage in computational models of inquiry.

Furthermore, the datafication era, leading to important innovations in data gathering and analysis, can be interpreted via Rogers' Innovation Theory (Rogers, 2003) with respect to how individuals adapt to the novelties. The spread of innovations is not only influenced by the technical advantages of an innovation but also its perceived ease, relative advantage, compatibility with existing practices, trialability, and observability. Our findings suggest that methodological proximity, particularly the competency in quantitative methodology and programming tools like Python and R, increases the likelihood of interacting with big data, which aligns with the compatibility requirement with existing practices. Besides the openness factor, a strong predictor for both competency and use corresponds to the early adopters' model of the theory, for which we suggest that academics who are more open and receptive to innovation have a pivotal role in the use and engagement with big data. In parallel, the observed nexus between receptiveness to big data's utilization and training provides evidence for the Technology Acceptance Model (Davis, 1989) wherein perceived usefulness and perceived ease of use are central concepts. It can be contended that academics who perceive big data's involvement in research as applicable for quality and visibility are more likely to engage with big data, as our multivariate analysis demonstrates for both competency and use. Self-rated competency in quantitative methods and use of programming languages, in addition, supports the concept of perceived ease of use, as those with stronger quantitative and programming backgrounds have an increased level of adoption, both in competency and use of big data.

Thirdly, the pedagogic device framework (Bernstein, 2000) offers a useful lens through which to examine our findings about the curricula in the departments, the environment in which academics work, and how the field of big data is institutionalized. According to Bernstein, knowledge is not neutral and is shaped by institutions through the "device", which determines what is legitimate to teach and should be visible, acting as a gatekeeper of valid knowledge. One might argue that departments that incorporate (or recontextualize, as Bernstein calls it) knowledge from external fields like data science or computational sciences, big data signals the legitimacy of the big data training, and academics in such departments may feel more competent and internalize this field, as their environment encourages engaging with the domain. In this vein, the descriptive findings support the idea that more courses in the curricula and the overall importance given to big data give rise to self-assessed competency and use. However, they lose their significance as a predictor of the use of big data, which suggests that the exposition in the departmental environment fosters competency (or vice versa). However, that does not necessarily translate into applied practice and remains a symbolic change rather than a structural transformation.

Regarding demographics, the data descriptively point to younger academics' higher familiarity and older academics' higher competency, whereas age groups only significantly predicted familiarity based on the multivariate analysis. This might reflect the age dynamics in technology socialization as younger generations are more likely to exhibit surface-level acquaintance with concepts pertaining to the digital world. Along the same lines, initial findings show that full professors had the highest familiarity and total medium and high competency, while research assistants and lecturers had the lowest proportions of actual use. Still, regression findings indicate that academic titles are significant predictors for familiarity but not for competency/use. Positional power and advanced age may facilitate exposure, but not necessarily engagement, as skills related to quantitative methodology and programming languages, takes the lead. Further, sex is another sociodemographic variable that shapes the level of engagement, specifically in terms of familiarity and competency, as evidenced by both descriptive and multivariate analyses. However, when it comes to use, there are no significant differences between male and female academics, and sex does not appear as a significant predictor in the regression model. Male academics report more familiarity and competency, and their odd ratio of being familiar and competent is higher. The lack of significance in the use prompts the consideration of females being underrepresented in the science, technology, engineering, and mathematics (STEM) fields, which are linked with the big data domain, and the disparity between their actual abilities and self-perception, which is called the confidence gap (Kong et al., 2020; McKinnon & O'Connell, 2020).

We may also interpret our findings by the institutional and digital divide framework, as at the descriptive level, academics who work in Level 1 SEDP and private universities are more likely to be familiar with the big data concept, and multivariate analysis demonstrated that Level 1 academics have the highest odds of familiarity. However, most categories are insignificant and show some variety for the other models. In the Turkish context, the higher education system is significantly stratified. The top-tier universities, which are generally in İstanbul and Ankara, have more access to funding, collaborations, abroad networks, and faculty capacity, while there is a dramatic divide of quality amongst universities across different regions. Academics in private universities and developed provinces are more familiar with big data, most likely due to the contribution of high-status universities with better infrastructure, possible computational social sciences training, and networks. That being the case, even in the developed provinces, there are a lot of private and public universities, and institutional quality varies widely, and familiarity alone does not guarantee deeper engagement with the recent advancements in methodology. Van Dijk's (2006) account provides a useful framework here. He argued that the cause of digital inequalities mainly stems from social positions (in our case: top-tier universities/others or private/public universities) and should be interpreted not merely in access to technology.

Our study has certain limitations. As mentioned before, this analysis is built upon the quantitative results of a mixed methods study; qualitative findings may help explore nuanced motivations about adoption. Secondly, questions related to competency and use are based on self-assessment, which may not truly capture the phenomenon, while another type of measurement could determine a more objective finding. This also may have affected the results of models, such as the use of big data, as the meanings of big data use may alter across approaches or disciplines. Thirdly, by the nature of the survey, some questions, such as those related to the presence of big data courses, do not measure whether the academic directly delivers the course, which works somewhat as a proxy variable. Moreover, the data represents a snapshot in time, gathered in late 2022, which restricts evaluating, for instance, recent developments in AI and their effect on big data engagement. Lastly, this study is not a complete probabilistic representative study, and broad generalizations might be faulty.

This study contributes to the literature in a couple of ways. First, this is one of the first large-scale empirical studies, based on our knowledge, of how social science academics in Türkiye engage with big data. Second, we sought to go beyond descriptive reporting by incorporating relevant theoretical frameworks to interpret our results. Third, we believe this paper may stimulate a broader discussion on the evolving methodological landscape of social sciences in Türkiye by foregrounding structural and institutional dimensions.

References

Agrawal, M., Peterson, J. C., & Griffiths, T. L. (2020). Scaling up psychology via Scientific Regret Minimization. *Proceedings of the National Academy of Sciences*, *117*(16), 8825–8835. https://doi.org/10.1073/pnas.1915841117

Aytaç, Z., & Bilge, H. Ş. (2021). Öğretim elemanlarının büyük veri farkındalık ve eğitim beklentilerinin değerlendirilmesi.1. *İşletme*, *2*(2), 29–46.

Astleitner, H. (2024). We have big data, but do we need big theory? Review-based remarks on an emerging problem in the social sciences. *Philosophy of the Social Sciences*, *54*(1), 69–92. https://doi.org/10.1177/00483931231188825

Barnes, T. J., & Wilson, M. W. (2014). Big data, social physics, and spatial analysis: The early years. *Big Data & Society*, *1*(1), 205395171453536. https://doi.org/10.1177/2053951714535365

Bernstein, B. B. (2000). Pedagogy, symbolic control, and identity: Theory, research, critique (Rev. ed). Rowman & Littlefield Publishers.

Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, *78*(2), 161–188. https://doi.org/10.1111/j.1751-5823.2010.00112.x

Beuving, J. J. (2020). Ethnography's future in the big data era. *Information, Communication & Society*, *23*(11), 1625–1639. https://doi.org/10.1080/1369118X.2019.1602664

Bjerre-Nielsen, A., & Glavind, K. L. (2022). Ethnographic data in the age of big data: How to compare and combine. *Big Data & Society*, *9*(1), 20539517211069893. https://doi.org/10.1177/20539517211069893

Bölükbaş, K. (2021). Sosyal bilimlerde büyük veri üzerine yapılan akademik çalışmaların analizi: YÖK tez örneği (2010 - 2020). *Akdeniz Üniversitesi İletişim Fakültesi Dergisi*, 35, 158–173. https://doi.org/10.31123/akil.886708

Brady, H. E. (2019). The challenge of big data and data science. *Annual Review of Political Science*, *22*(1), 297–323. https://doi.org/10.1146/annurev-polisci-090216-023229

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data science journal*, 14, 2-2.

Cabrera-Álvarez, P. (2022). Survey research in times of big data. *Empiria. Revista de Metodología de*

Ciencias Sociales, 53. https://doi.org/10.5944/empiria.53.2022.32611

Callegaro, M., & Yang, Y. (2018). The role of surveys in the era of "Big Data." In *The Palgrave Handbook of Survey Research* (pp. 175–198).

Chakravorti, B., Bhalla, A., & Chuturverdi, R. S. (2019). Which countries are leading the data economy? *Harvard Business Review*. https://hbr.org/2019/01/which-countries-are-leading-the-data-economy

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*, 1–12. https://doi.org/10.1016/j.ssresearch.2016.04.015

Cox, M., & Ellsworth, D. (1997). *Managing big data for scientific visualization*. https://www.researchgate.net/publication/238704525_Managing_big_data_for_scientific_visualization

Daikeler, J., Bošnjak, M., & Lozar Manfreda, K. (2020). Web versus other survey modes: an updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology*, *8*(3), 513–539. https://doi.org/10.1093/jssam/smz008

Davenport, T. (2014). Stop using the term "big data." *Deloitte Insights*. https://www2.deloitte.com/us/en/insights/topics/analytics/big-data-buzzword.html

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Diaz-Bone, R., Horvath, K., & Cappel, V. (2020). Social research in times of big data. The challenges of new data worlds and the need for a sociology of social research. *Historical Social Research/Historische Sozialforschung*, *45*(3), 314-341.

Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLOS ONE*, *15*(2), e0228987. https://doi.org/10.1371/journal.pone.0228987

Florescu, D., Karlberg, M., Reis, F., Del Castillo, P. R., Skaliotis, M., & Wirthmann, A. (2014, June). Will 'big data'transform official statistics. In European Conference on the QualityofOfficial Statistics. Vienna, Austria (pp. 2-5).

Foster, I. (Ed.). (2017). *Big data and social science: A practical guide to methods and tools*. CRC Press Taylor & Francis Group.

Halford, S., & Savage, M. (2017). Speaking Sociologically with Big Data: Symphonic Social Science and the Future for Big Data Research. Sociology, 51(6), 1132–1148. https://doi.org/10.1177/0038038517698639

Hosseini, M., Wieczorek, M., & Gordijn, B. (2022). Ethical issues in social science research employing big data. *Science and Engineering Ethics, 28*(3), 29. https://doi.org/10.1007/s11948-022-00380-7

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., Vazire, S., Vespignani, A., & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595*(7866), 181–188. https://doi.org/10.1038/s41586-021-03659-0

Household Information Technologies Survey. (2024). Türkiye İstatistik Kurumu. https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2024-53492

Howe III, E. G., & Elenberg, F. (2020). Ethical challenges posed by big data. *Innovations in clinical neuroscience, 17*(10-12), 24.

Ignatow, G. (2020). *Sociological theory in the digital age* (1 Edition). Routledge.

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neil, C., & Usher, A. (2015). *AAPOR Report: Big Data*. American Association for Public Opinion Research. https://www.aapor.org/Education-Resources/Reports/Big-Data.aspx#8.%20Conclusions%20and%20Research%20Needs

Karaca, Y. (2024). Kamu *Politikalarında Büyük Veri Kullanımının İncelenmesi ve Türkiye Bağlamında Bir Analiz*. Muğla Sıtkı Koçman Üniversitesi.

King, G. (2011). *The* Social Science Data Revolution. Horizons in Political Science Talk, Harvard University. https://gking.harvard.edu/files/gking/files/evbase-horizonsp.pdf

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 205395171452848. https://doi.org/10.1177/2053951714528481

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 205395171663113. https://doi.org/-10.1177/2053951716631130

Knorr-Cetina, K. (1999). Epistemic cultures: How the sciences make knowledge. Harvard University Press.

Kong, S. M., Carroll, K. M., Lundberg, D. J., Omura, P., & Lepe, B. A. (2020). Reducing gender bias in STEM. MIT Science Policy Review, 1. https://doi.org/10.38105/spr.11kp6lqr0a

Köseoğlu, Ö., & Demirci, Y. (2017). Türkiye'de büyük veri ve veri madenciliğine ilişkin politika ve stratejiler: ulusal politika belgelerinin içerik analizi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *2017*, 2223–2239.

Laney, D. (2001). *3D Data management: Controlling data volume, velocity, and variety*. META Group. https://www.bibsonomy.org/bibtex/742811-cb00b303261f79a98e9b80bf49

Lazer, D., Hargittai, E., Freelon, D., Gonzalez-Bailon, S., Munger, K., Ognyanova, K., & Radford, J. (2021). Meaningful measures of human society in the twenty-first century. *Nature*, *595*(7866), 189–196. https://doi.org/10.1038/-s41586-021-03660-7

Lukoianova, T., & Rubin, V. L. (2014). Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online*, *24*(1), 4. https://doi.org/10.7152-/acro.v24i1.14671

Mann, A. (2016). Core concept: Computational social science. *Proceedings of the National Academy of Sciences*, *113*(3), 468–470. https://doi.org/10.-1073/pnas.1524881113

Manovich, L. (2011). The promises and the challenges of big social data. Debates in the digital humanities. *Debates in the Digital Humanities*, 2, 460–475.

McKinnon, M., & O'Connell, C. (2020). Perceptions of stereotypes applied to women who publicly communicate their STEM work. Humanities and Social Sciences Communications, 7(1), 160. https://doi.org/10.1057/s41599-020-00654-0

Marres, N. (2017). *Digital sociology*. Polity Press.

Mazzocchi, F. (2015). Could big data be the end of theory in science?: A few remarks on the epistemology of data-driven science. *EMBO Reports*, *16*(10), 1250–1255. https://doi.org/10.15252/embr.201541001

Metzler, K., Kim, D. A., Allum, N., & Denman, A. (2016). *Who is doing computational social science? Trends in big data research* (A SAGE White Paper). Sage Publication. https://static1.squarespace.com/static/5d5ad9e0100bdf0001af0f5e/t/5e15fca81845d424d285e22d/1578499246868/Who-is-doing-computational-social-science.pdf

Parry, D. A. (2024). Without access to social media platform data, we risk being left in the dark. *South African Journal of Science*, *120*(3/4). https://doi.org/10.17159/sajs.2024/17008

Rogers, E. M. (2003). Diffusion of innovations (Fifth edition). Free Press.

Şallı, Ş. (2021). *Büyük veri devrimi ve sosyal bilimler araştırma yöntemlerinde yeni paradigmalar*. Karabük University.

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346*(6213), 1063-1064.

Sarkar, S. (2021). Using qualitative approaches in the era of big data: A confessional tale of a behavioral researcher. *Journal of Information Technology Case and Application Research*, *23*(2), 139–144. https://doi.org/10.1080/15228053.2021.1916229

Strateji ve Bütçe Başkanlığı Stratejik Plan (2019-2023). (2018). Strateji ve Bütçe Başkanlığı. https://www.sbb.gov.tr/wp-content/uploads/2021/07/Strateji-ve-Butce-Baskanligi-2019-2023-Stratejik-Plani-28072021.pdf

Strateji ve Bütçe Başkanlığı Stratejik Plan (2024-2028). (2023). Strateji ve Bütçe Başkanlığı. https://www.sbb.gov.tr/wp-content/uploads/2024/04/STRATEJIK-PLAN-2024-2028_03042024.pdf

Stegenga, S. M., Steltenpohl, C. N., Lustick, H., Meyer, M. S., Renbarger, R., Standiford Reyes, L., & Lee, L. E. (2024). Qualitative research at the crossroads of open science and big data: Ethical considerations. *Social and Personality Psychology Compass*, *18*(1), e12912. https://doi.org/10.1111/spc3.12912

Van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. Poetics, 34(4–5), 221–235. https://doi.org/10.1016/j.poetic.2006.05.004

Veltri, G. (2020). *Digital Social Research*. John Wiley & Sons, Inc.& Society, 1(1), 205395171452848. https://doi.org/10.1177/2053951714528481

APPENDICES

Table 3. The Statements, Attitude Factors, and Factor Loadings

OPENNESS TO BIG DATA (Openness)	Factor Loadings
Research using traditional methods is more reliable than research using big data.	0,726
Programming languages and analysis methods related to the use of big data should be taught in undergraduate departments of social, humanities and administrative sciences.	0,657
Researchers using qualitative research approaches should receive training on the use of big data.	0,618
The use of big data has revolutionized social science methodology.	0,564
Big data can be utilized in research using qualitative research approaches.	0,426
BELIEF IN IMPACT OF BIG DATA (Impact)	
With big data, the use of traditional data collection/generation methods (such as surveys, interviews) in social sciences will end.	-0,774
With big data, the role of qualitative research in social sciences is diminishing.	-0,752
With the use of big data, sample surveys are losing their importance.	-0,722
Research using big data in social sciences yields better quality results than research using traditional methods.	-0,643
EFFICIENCY AND QUALITATIVE RESEARCH'S INTEGRATION (Efficiency and Integration)	
Compared to traditional research methods, big data facilitates the process of data collection and organization.	0,703
The use of big data reduces research costs.	0,678
Qualitative research approaches are essential for a better understanding of big data.	0,518

Table 4. Sex, Age Group, Academic Title by Familiarity with Big Data, Competency in Big Data

	Familiarity with Big Data			Competency in Big Data					Use of Big Data		
	Familiar %	**Not Familiar %**	**n**	**Not at all %**	**Low %**	**Medium %**	**High %**	**n**	**Used big data %**	**Did not use big data %**	**n**
Sex**											
Male	79.1	20.9	1778	32.1	34.3	28.3	5.3	1407	35.7*	64.3*	955
Female	74.9	25.1	1172	43.4	30.7	22.4	3.5	1289	35.8*	64.2*	730
Age Group											
18-29	77.6	22.4	286	41.0	33.8	22.5	2.7	222	30.5*	69.5*	131
30-39	77.5	22.5	1436	41.9	31.5	23.1	3.5	1113	34.8*	65.2*	647
40-49	79.3	20.7	1180	35.3	33.7	26.5	4.6	936	36.6*	63.4*	606
50-59	72.5	27.5	466	30.2	32.2	30.5	7.1	338	40.3*	59.7*	236
60 years and above	66.0	34.0	156	30.1	30.1	31.1	8.7	103	31.9*	68.1*	72
Academic Title***											
Research Asst.	77.1	22.9	717	45.6	32.9	18.3	3.3	553	30.9*	69.1*	301
Lecturer	68.8	31.2	673	40.2	31.7	24.4	3.7	463	31.4*	68.6*	277
Asst .Prof. Dr.	77.4	22.6	1048	36.9	32.1	26.6	4.4	811	38.3*	61.7*	512
Assoc. Prof. Dr.	81.3	18.7	646	33.0	35.2	28.6	3.2	525	38.6*	61.4*	352
Professor	82.2	17.8	433	30.3	30.1	30.9	8.7	356	37.1*	62.9*	248
Total	**77.0**	**23.0**	**3524**	**37.6**	**32.5**	**25.4**	**4.5**	**2712**	**35.8**	**64.2**	**1692**

**Categories do not have statistically significant difference according to Chi-Square Test at 0.05 significance level.*

***One of the answer categories was "I do not want to specify" which is not shown here as the frequency was below 25.*

****Other academic positions such as Post-Docs are excluded.*

Table 5. Academic Field, Type of University, SEDP)by Familiarity and Competency in Big Data

	Familiarity with Big Data			Competency in Big Data					Use of Big Data		
	Familiar %	Not Familiar %	n	Not at all %	Low %	Medium %	High %	n	Used big data %	Did not use big data %	n
Academic Field											
Educational Sciences	59.6	40.4	109	33.8	36.9	24.6	4.6	65	27,9*	72,1*	43
Sociology and Cultural Studies	74.7	25.3	372	42.4	29.9	21.9	5.8	278	36.9*	63.1*	160
History and Archeology	32.1	67.9	212	39.7	26.5	29.4	4.4	68	39.0*	61.0*	41
Political Science	70.5	29.5	495	45.0	31.2)	21.2	2.6	349	30.2*	69.8*	192
Finance, Banking, and Actuarial Sciences	84.2	15.8	279	26.4	34.0	35.3	4.3	235	38.2*	61.8*	173
Economy/Economics	84.0	16.0	325	32.6	29.3	28.2	9.9	273	42.9*	57.1*	184
Communication	90.7	9.3	322	31.8	41.1	25.3	1.7	292	32.2*	67.8*	199
Management and Administration	85.2	14.8	738	38.2	30.8	26.4	4.6	629	38.8*	61.2*	389
Marketing and Advertisement	95.4	4.6	197	43.1	32.4	22.9	1.6	188	31.8*	68.2*	107
Tourism	75.5	24.5	220	41.0	36.7	21.1	1.2	166	30.6*	69.4*	99
Other	66.3	33.7	255	37.3	30.2	24.3	8.3	169	34.0*	66.0*	106
Type of University											
Public University	75.6	24.4	2837	38.3*	32.6*	25.0*	4.1*	2144	35.3*	64.7*	1322
Private University	82.7	17.3	687	34.9*	32.0*	27.1*	6.0*	568	37.3*	62.7*	370
SEDP**											
Level 1	82.3	17.7	1495	38.1*	32.1*	24.7*	5.0*	1230	36.4*	63.6*	761
Level 2	75.3	24.7	632	39.3*	34.2*	23.7*	2.7*	476	32.9*	67.1*	289
Level 3	75.6	24.4	422	36.1*	33.2*	24.5*	6.3*	319	37.3*	62.7*	204
Level 4	73.2	26.8	336	32.9*	32.9*	30.5*	3.7*	246	33.9*	66.1*	165
Level 5	72.1	27.9	333	42.9*	28.3*	24.6*	4.2*	240	36.5*	63.5*	137
Level 6	62.8	37.2	261	34.8*	33.5*	28.7*	3.0*	164	35.5*	64.5*	107
Total	**77.0**	**23.0**	**3524**	**37.6**	**32.5**	**25.4**	**4.5**	**2712**	**35.8**	**64.2**	**1692**

**Categories do not have statistically significant differences according to the Chi-Square Test at a 0.05 significance level.*
***Academics who work abroad are excluded.*

Table 6. Research Method Approaches Used by Familiarity with Big Data and Competency in Big Data

	Familiarity with Big Data			Competency in Big Data					Use of Big Data		
	Familiar %	Not Familiar %	n	Not at all %	Low %	Medium %	High %	n	Used Big Data %	Did Not Use Big Data %	n
QUAN Research Approaches											
Used QUAN	83.0	17.0	2349	40.1	32.1	25.2	2.6	763	38.1	61.9	1235
Did not use QUAN	64.9	35.1	1175	36.6	32.6	25.6	5.2	1949	29.5	70.5	457
Used only QUAN**	82.1	17.9	767	35.9	29.7	27.3	7.1	630	34.4	65.6	1288
QUAL Research Approaches											
Used QUAL	78.8	21.2	2087	32.7	31.8	29.6	5.9	1067	38.7	61.3	718
Did not use QUAL	74.3	25.7	1437	40.8	32.9	22.7	3.5	1645	40.1	59.9	404
Used only QUAL**	66.0	34.0	497	50.6	28.4	18.9	2.1	328	26.5	73.5	162
MM Research Approaches											
Used MM	80.0	20.0	1431	43.2	30.6	22.2	4.0	1567	32.8	67.2	890
Did not use MM	74.9	25.1	2093	30.0	35.1	29.9	5.1	1145	39.0	61.0	802
Used only MM	69.6	30.4	398	25.5	34.9	36.0	3.6	275	34.0*	66.0*	206
Total	**77.0**	**23.0**	**3524**	**37.6**	**32.5**	**25.4**	**4.5**	**2712**	**35.8**	**64.2**	**1692**

**"Used only MM" category do not have a statistically significant difference compared to "Did not use only MM" category according to the Chi-Square Test at a 0.05 significance level.*
***Used only QUAN" and "Used only QUAL" categories have statistically significant differences (Chi-Square: 0.05) with the "Did not use only QUAN" and "Did not use only QUAL" categories, which are not shown here.*

Table 7. Competency in Research Method Approaches and Big Data by Familiarity, Competency and Use of Big Data

	Familiarity with Big Data			Competency in Big Data					Use of Big Data		
	Familiar %	**Not Famil-iar %**	**n**	**Not at all %**	**Low %**	**Medium %**	**High %**	**n**	**Used Big Data %**	**Did Not Use Big Data %**	**n**
Competency in QUAN											
Very low	51.1	48.9	186	73.7	18.9	7.4	0.0	95	8.0	92.0	25
Low	70.3	29.7	451	59.3	30.3	9.5	0.9	317	22.5	77.5	129
Medium	71.4	28.6	1302	43.3	35.6	20.0	1.1	930	31.5	68.5	527
High	84.5	15.5	1215	28.6	34.3	33.1	4.0	1027	35.9	64.1	733
Very high	92.7	7.3	370	19.0	24.5	37.0	19.5	343	52.2	47.8	278
Competency in QUAL											
Very low	67.7	32.3	201	50.7	26.5	19.1	3.7	136	29.1	70.9	79
Low	80.8	19.2	526	44.5)	32.0	18.8	4.7	425	29.2	70.8	243
Medium	74.3	25.7	1191	40.2	33.2	23.8	2.7	885	33.9	66.1	737
High	78.2	21.8	1264	32.9	34.1	28.5	4.6	989	39.5	60.5	547
Very high	81.0	19.0	342	29.2	28.2	32.9	9.7	277	52.3	47.7	86
Competency in MM											
Very low	68.3	31.7	306	62.2	23.4	13.4	1.0	209	29.1	70.9	79
Low	76.1	23.9	695	54.1	31.4	11.9	2.6	529	29.2	70.8	243
Medium	76.3	23.7	1537	37.1	35.2	25.0	2.6	1172	33.9	66.1	737
High	81.4	18.6	855	21.4	32.8	38.4	7.5	696	39.5	60.5	547
Very high	80.9	19.1	131	18.9	23.6	36.8	20.8	106	52.3	47.7	86
Competency in Big Data*											
Low	*	*	*	-	-	-	-	-	14.4	85.6	881
Medium	*	*	*	-	-	-	-	-	54.1	45.9	690
High	*	*	*	-	-	-		-	86.8	13.2	121
Total	**77.0**	**23.0**	**3524**	**37.6**	**32.5**	**25.4**	**4.5**	**2712**	**35.8**	**64.2**	**1692**

**Respondents who are not familiar with big data concept were not asked about their competency in big data. Respondents who said they are "not at all" competent in big data weren't asked if they have ever used big data.*

Table 8. Research Methods Courses in the Curricula and Importance Given to Big Data by Familiarity, Competency and Use of Big Data

	Familiarity with Big Data			Competency in Big Data					Use of Big Data		
	Familiar %	**Not Familiar %**	**n**	**Not at all %**	**Low %**	**Medium %**	**High %**	**n**	**Used Big Data %**	**Did not Use Big Data %**	**n**
Research Methods Courses***											
Undergraduate											
No course	73.5	26.5	431	43.8	27.4	24.6	4.1	317	33.1*	66.9*	178
One course	76.2	23.8	1574	36.8	36.3	24.0	2.9	1200	33.4*	66.6*	758
Multiple courses	80.0	20.0	1185	34.9	30.1	28.3	6.8	948	39.9*	60.1*	617
Master's											
No course	61.4	38.6	158	41.2	35.1	19.6	4.1	97	24.6*	75.4*	57
One course	76.8	23.2	1762	39.3	33.0	24.4	3.3	1353	35.2*	64.8*	821
Multiple courses	82.2	17.8	1094	32.8	32.6	28.4	6.2	899	38.9*	61.1*	604
PhD											
No course	67.6	32.4	262	39.5*	32.2*	24.9*	3.4*	177	29.9*	70.1*	107
One course	77.7	22.3	1249	37.5*	33.4*	25.7*	3.4*	971	35.9*	64.1*	607
Multiple courses	82.4	17.6	992	34.6*	32.7*	26.8*	5.9*	817	38.6*	61.4*	534
Importance Given to Big Data in the Curricula											
Very Low	**	**	**	56.8	26.1	15.4	1.8	969	29.1	70.9	419
Low	**	**	**	39.0	39.8	18.7	2.5	674	31.9	68.1	411
Medium	**	**	**	21.5	38.5	35.8	4.2	693	32.7	67.3	544
High	**	**	**	15.9	25.3	48.5	10.3	233	50.5	49.5	196
Very High	**	**	**	14.7	23.8	37.8	23.8	143	61.5	38.5	122
Total	**77.0**	**23.0**	**3524**	**37.6**	**32.5**	**25.4**	**4.5**	**2712**	**35.8**	**64.2**	**1692**

**Categories do not have statistically significant differences according to the Chi-Square Test at a 0.05 significance level.*

***Respondents who are not familiar with the big data concept were not asked about the importance of big data in their department's curricula.*

****Each degree had a category of "This question is not suitable for me/I don't know." whose values are not shown here.*

Table 9. Big Data's Presence in the Curricula and Attitude Factors by Competency in Big Data and Use of Big Data

	Competency in Big Data					Use of Big Data		
	Not at all %	**Low %**	**Medium %**	**High %**	**n**	**Used Big Data %**	**Did Not Use Big Data %**	**Total % (Count)**
Big Data's Presence								
Undergraduate								
No course, no mentions	49.0	31.9	16.9	2.2	1255	29.2	70.8	583
No course but mentions	25.2	40.8	30.3	3.7	779	34.0	66.0	583
One course	18.5	22.0	48.7	10.8	232	51.9	48.1	189
Multiple courses	8.1	18.7	48.0	25.2	123	65.5	34.5	113
Master's								
No course, no mention	52.9	30.3	14.6	2.2	944	29.2	70.8	640
No course but mentions	28.8	40.2	28.2	2.8	886	33.1	66.9	631
One course	15	26.5	50.0	8.5	294	50.0	50.0	250
Multiple courses	5.8	20.4	49.6	24.1	137	62.0	38.0	129
PhD								
No course, no mention	51.4	30.5	16.2	1.9	777	28.5	71.5	445
No course but mentions	31.5	39.2	26.8	2.4	783	31.5	68.5	536
One course	12.6	29.8	48.7	8.8	238	47.1	52.9	208
Multiple courses	7.1	23.0	46.8	23.0	126	62.4	37.6	117
Attitude Factors								
Openness	12.6	29.8	48.7	8.8				
Low	48.7	32.4	16.7	2.2	904	23.7	76.3	464
Medium	35.0	35.3	26.2	3.5	904	33.7	66.3	588
High	29.2	29.8	33.4	7.6	904)	46.4	53.6	640
Impact								
Low	39.6	34.3	21.2	4.9	904	33.0*	67.0*	546
Medium	37.1	31.8	26.9	4.2	903	37.1*	62.9*	568
High	36.1	31.4	28.2	4.3	905	37.0*	63.0*	578
Efficency and QUAL								
Low	41.5	33.0	21.5	4.0	903	28.6	71.4	528
Medium	38.1	31.4	25.5	5.0	905	35.2	64.8	560
Strong	33.2	33.1	29.3	4.4	904	37.0	63.0	578
Total	**37.6**	**32.5**	**25.4**	**4.5**	**2712**	**35.8**	**64.2**	**1692**

**Categories do not have statistically significant differences according to the Chi-Square Test at a 0.05 significance level.*