JOURNAL OF ASIAN SOCIOLOGY
Volume 48 | Number 3 | September 2019, 343-358
DOI 10.21588/jas/2019.48.3.003 **Special Issue**

New Opportunities for Sociological Research: A Discussion of the Usefulness of Mixed Methods with Data Science

Yeunchul Kim | Seoul National University of Science and Technology

Traditionally, research methods in sociology have been divided into quantitative and qualitative methods, and these methodologies have been carried out separately as if they were different academic disciplines. The divide between these 'two cultures' of methodology has served to engender a crisis in the field of sociology. As an alternative to this, an emphasis on mixed methods is emerging. The mixed methods approach combines quantitative and qualitative methods, allowing researchers to conduct research in an integrated way. It provides some utility in that it reveals things which cannot be achieved with a traditional single research method, and it enables researchers to overcome the 'two cultures' bifurcation which plagues contemporary sociological research. Here we propose that the combination of recently developed data science and mixed methods approaches will bring new innovations and opportunities to sociological research by enhancing the efficiency of research and enabling more integrated analysis and interpretation of social phenomena.

Keywords: *Mixed Methods, Data Science, Sociological Methodology, Crisis of Sociology, Two Cultures*

Introduction

Sociology is fundamentally concerned with the way people live in society. The lives of individuals are greatly influenced by the structure and institutions of society, but technological changes also have considerable impact on them. The most dramatic change in technology over the past two decades is the information technology most represented by the Internet. With the advent of the Internet, people's lives have been separated into off-line and on-line domains.

In the early days of the Internet, the on-line world was regarded as an extra area with little relation to the off-line, and its connectivity with the off-line world did not receive much attention. However, nowadays many people are spending more time and doing more work in the on-line world than the off-line one. The on-line world is now more important than the off-line world. How then should sociological research change to accommodate this phenomenon?

The usefulness of the on-line world for sociological research is its traceable trail. In the on-line world, an individual's life leaves traces, called logs, at every step, and these logs are easily collected in digital format. A new research object, completely different from off-line data that can disappear from memory or be forgotten in the brain over time, has emerged, and presents enormous potential for analysis. As sociological analysis of the on-line world becomes more important and studies dealing with data in digital form are being widely introduced, a new sub-field so-called 'digital sociology' (Lupton 2014; Marres 2017; Orton-Johnson and Prior 2013) is emerging. However, the vision of 'digital sociology' is still different for different scholars, and it lacks a solid foundation that is commonly recognized among those working in the field.

The proliferation of new research fields facilitates the emergence of new research agendas. However, how precisely to conduct the research remains vague. While many of the latest techniques for dealing with digital data are in development, there is not yet an established protocol for how to utilize this new wealth of data for sociological research. In this age of vague transitions, the traditional division of quantitative and qualitative methods in sociology has become a hindrance rather than an advantage to new innovative research.

The recent emergence of digital data as a new source for understanding the on-line world has been promising, but analysis of it alone is not enough for sociological research, which endeavors to explore broader aspects of the

lives of individuals as situated within society.[1] Therefore, the research method for digital sociology should encompass both on-line and off-line worlds, and researchers in this field should be skilled in techniques for handling data in digital formats. That is, deciphering digital data requires a completely different set of skills, and the emerging field of data science represents these kinds of skills.

In addition, research methodologies must also change in order to study a society and behaviors which have transformed as a result of the on-line world. To analyze the totality of social phenomena, it is necessary to handle dissimilar types of data and to establish research designs from an integrated perspective. Because of these conditions, mixed methods are emerging as a new alternative and opportunity for sociologists (Edmonds and Kennedy 2017; Leavy 2017; Fox and Alldred 2018). This article aims to discuss how mixed methods, along with data science, can bring new innovations to sociological research.

The 'Two Cultures' of Sociological Research Methodology

Sociology has long maintained the strong distinction between quantitative and qualitative methodologies. These two camps are based on different principles and practical experiences, and as such they do not allow for the accumulation of common research experience. As a result, these two camps are in a state of serious discord and often fail to understand one another. This is one of the major causes of the 'crisis of sociology' (Savage and Burrows 2007) that many scholars are discussing.

From the very process of becoming trained in sociological methodology, the path to accumulate experience by choosing either quantitative or qualitative methodology has become common, and the long-term outcome of this practice is a rift between the 'two cultures' of sociological methodology. 'Two cultures' is a term which was proposed by C. P. Snow ([1956] 2012) in 1956 and which symbolizes a criticism of the segregation between the humanities and the natural sciences. The key to the criticism

[1] The most worrying aspect of sociological research examining an individual's life through on-line data analysis is that information may not be found in the on-line world because it has not been recorded. Although the activities individuals partake in in the on-line world have become more important than ever, such activities are still determined by the individual's choice. We should keep in mind that information about people who have not made that choice cannot be found on-line at all, and that sometimes that information is more important than one might expect.

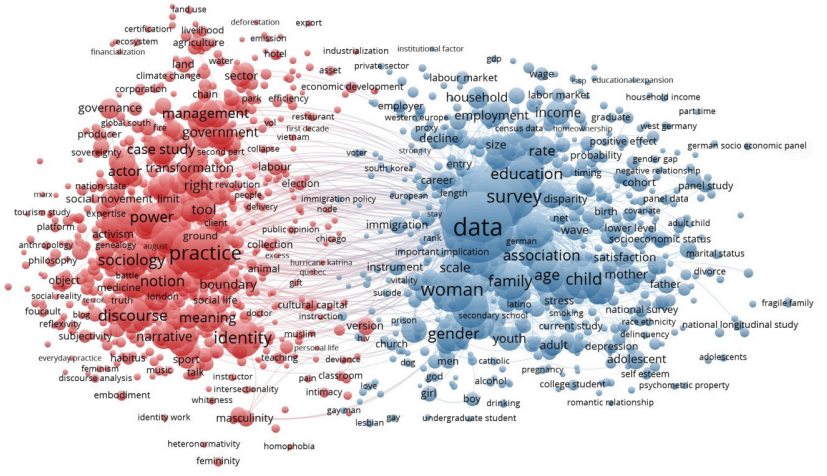

Fig. 1.—Divided Sociology in 'Two Cultures'[2] (Tragg and Franssen 2016)

that the term 'two cultures' symbolizes is not a state in which the two camps are separated from each other, but rather the intellectual laziness that takes for granted the ignorance of the other's knowledge. The separation of quantitative and qualitative methodologies within the field of sociology is also in accord with this point. Researchers who are accustomed to quantitative methodologies are not interested in qualitative research methodologies and this is not a point of embarrassment nor a sign of inadequacy. The opposite is also true.

The data on social phenomena manifests in very complex and diverse forms, and if a sociologist who wishes to investigate and analyze it only deals with a certain format of data, his or her conclusions are bound to be partial and imperfect sociology. Thus, the methodological 'two cultures' rift is more likely to deepen the current crisis of sociology rather than overcome it.

[2] This figure analyzes a total of 14,613 sociological articles published in 146 journals during 2010-2015 using Vosviewer. There is a huge gap between the qualitative and quantitative methodology camps.

Principles and Necessities of Mixed Methods

In traditional sociology, quantitative and qualitative methodologies differ in the value of research orientation. As can be seen in figure 2 (Hamberg et al. 1996, p. 178) below, quantitative research focuses on generalizing hypotheses by examining internal validity, whereas qualitative research focuses on persuading the reliability of credibility to provide transferability. Even though the two are not in a relationship of mutual substitution, communication between the two methodologies is severely lacking due to the 'two culture' division. If we can overcome the gap between the two cultures, we can expect to establish an integrated sociological research methodology.

The basic strategy for overcoming the 'two cultures' rift is simple: we will use both. Rather than distinguishing between quantitative and qualitative methods, we should pursue a strategy to integrate them into sociological research using mixed methods. Mixed methods is a strategy to conduct research using both established quantitative and qualitative methods of sociology. The principles of the mixed methods approach are as follows (Creswell and Clark 2017, p. 5).

- The researcher collects and analyzes both qualitative and quantitative data rigorously in response to research questions and hypotheses.
- The researcher integrates (or mixes or combines) the two forms of data and their results.
- The researcher organizes these procedures into specific research designs that provide the logic and procedures for conducting the study
- The researcher frames these procedures within theory and philosophy.

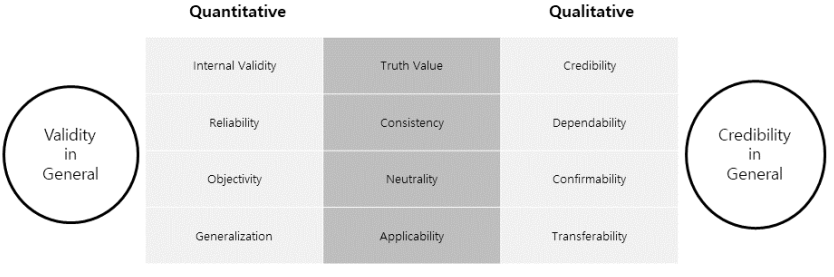

Fig. 2.—Principles of Quantitative and Qualitative Methodologies

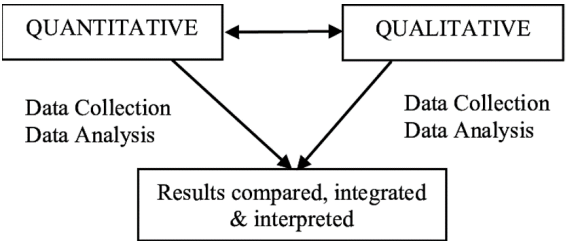

Fig. 3.—3-Axis Model of Mixed Methods (Atif, Richards, and Bilgin 2013)

In any discipline, the purpose of utilizing particular research methods is to obtain data on research subjects. Sociology is no different. If the data on social phenomena are only qualitative or quantitative in form, they can only be considered incomplete data about the research subject. Therefore, the most important reason to pursue mixed methodology is to obtain integrated data on the research subject, as the data is heterogeneous in form and type. In other words, a mixed methods approach in sociology is a research strategy that integrates both quantitative and qualitative data according to the researcher's research frame and design. Mixed methods are not a two-axis model that combines quantitative data and qualitative data in parallel, but a three-axis model in which a research frame and design can be considered together.

Therefore, merely combining quantitative data with qualitative data is not a proper interpretation of the mixed methodology approach. After the research design, or how to analyze the research subject, is first established, empirical research should be conducted to collect and analyze data that is consistent with the research design.

In regard to how to integrate and link quantitative and qualitative data, or which data to deal with first, mixed methodology can make procedural classifications. The following figure illustrates the procedural types of various mixed methods approaches.

As you can see from the figure above, mixed methods can be categorized into several strategies depending on how the researcher combines the heterogeneous data. The mixed method discussed in this article is the second, or the explanatory sequential method. This is because this mixed methods approach is the most fit for use in sociological research, in particular because of its potential to make progress in the emerging field of data science. While other mixed methods are also important, this article focuses singularly on the explanatory sequential method.

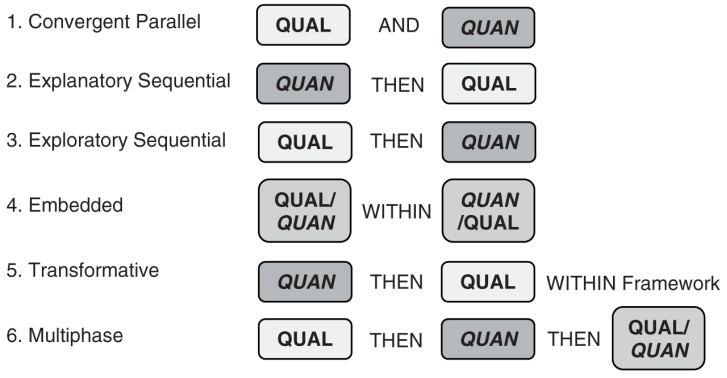

Fig. 4.—Mixed Methods Research Designs (Ozawa and Pongprirul 2014, p. 325)

The process of analyzing quantitative data and then confirming it through qualitative data is a significant help in obtaining a definitive sociological description of the research subject. The value of digital data, especially as a trail of on-line activity for individuals, is higher than ever before. This is because, firstly, advances in information technology make quantitative data in a digital form much easier and cheaper to collect than before, and second, the loci of peoples' lives have moved from off-line to on-line. Thus, analyzing highly accessible online digital data makes it easy to track people's social activities. But there are some problems here. First, digital data does not give us clues about off-line life, and it is not only impossible to know whether the result of digital data analysis is correct, but crucially, on-line digital data cannot inform researchers about the causes they may seek to identify. Therefore, we must inevitably also utilize qualitative research methods in order to determine whether the results of the preliminary quantitative data analysis are correct and to understand what the causes and drivers of researched phenomena are. In conclusion, the combination of quantitative data analysis based on data science and a qualitative research method thereby leads to a more thorough understanding of the research subject.

Another advantage of the mixed method of analyzing quantitative data prior to qualitative data is its efficiency. Data science has contributed significantly to lowering the cost of access and analysis of quantitative digital data. Therefore, if we can sketch a rough outline of the subject and evaluate the possibilities of research first, we can quickly determine whether to additionally apply qualitative methods that require a great deal of effort.

Data Science as a New Opportunity for Sociology

Since 2010, the term "Big Data" has exploded into global ubiquity. As information technology advances, information production increases exponentially, and the cost of information storage decreases dramatically, the information created in the on-line world is accumulating on a huge scale. This massive trove of data quickly began to be regarded as a resource of new value creation. Big data has become an enormous and irresistible tide in terms of its influence on research today.

However, big data was a relatively new kind of source that had not been experienced before, meaning that analyzing it with existing tools for analysis was not fruitful. This necessitated the need for new technical analysis tools, and in this way, a series of new analysis tools and techniques were created and utilized under the name of data science. This is the point at which the term data science became popular. As you can see in the figure above, after the explosion of the term big data beginning in 2010, it has become saturated since about 2014, when the term data science began to spread widely. The popularization of data science is only a matter of the past few years.

Data science is not merely a method for analyzing big data, but rather it signifies all the steps of searching, extracting, processing, visualizing, and conducting statistical analysis of data generated in the on-line world.

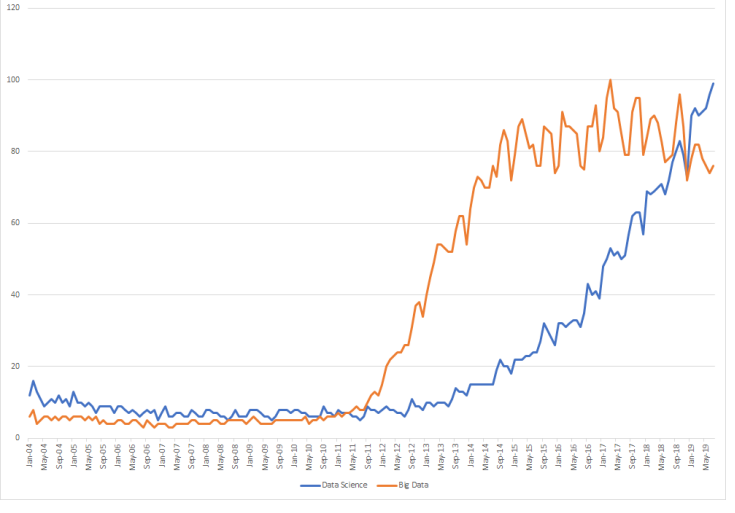

Fig. 5.—Google Trends Results for "Big Data" and "Data Science"

Discovering new meanings and values through this analysis is the goal of data science.

There are several reasons why data science could develop rapidly. First, the primary center of people's lives is moving increasingly on-line. All on-line activities leave digital traces at every step. Every time a person logs in, reads, writes, takes a bus, buys things, and sends or receives pictures with their friends, the traces automatically accumulate. All of this digital data is a valuable resource that allows us to analyze individuals' activities in detail. Second, tools for analyzing digital data have quickly evolved. Currently, the most common analysis tool for data science professionals is the programming language Python and the statistical package software R. The rapid development of these two tools has greatly reduced the barriers to entry for data science, and even sociologists are now able to benefit from this evolution. Third, the cost of data has dropped significantly. Surveys, which are the traditional method of obtaining quantitative data, generally require large costs. However, the quantitative data used by data science is often easily collected through the Internet, which greatly reduces the cost constraints. As interest in personal information produced in the business domain has increased and government regulations have strengthened, obtaining digital data is more difficult than ever before, but it remains much cheaper than traditional surveys, and opportunities for access are on the rise. In short, advances in information technology have greatly increased access to digital data, which will open new doors for sociological research.

Early data science was an arena in which only a few specialists participated. This is because it required all sorts of specializations: statistics, programming and hacking skills, and the ability to interpret data. Since the usefulness of data science has become widely acknowledged, and many have captured its potential, data science has become a basic tool and technique for analysis. But there is still a barrier to entry: knowledge of programming languages and hacking technologies. It is undeniable that engineering skills are very important in the process of finding, collecting, and organizing digital data, but interpreting the data obtained and finding meaning within it requires yet another set of expertise and experience. This is where the role of sociology becomes vital.

As you can see from the figure above, complete data science requires domain knowledge of the data finally collected. The data used by data science is largely traces of what people have left behind in the on-line world. No matter how well this data is collected, the interpretation of the traces cannot be done properly without knowledge of an individual's behavior and its social

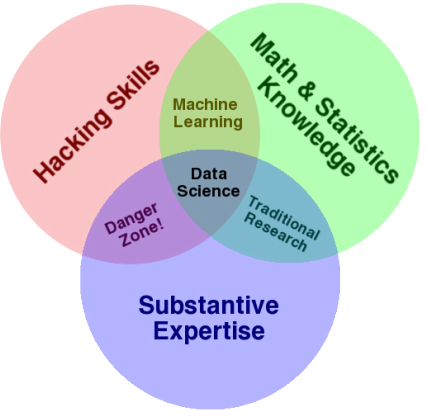

Fig. 6.—The Data Science Venn Diagram (Conway 2013)

context. Therefore, data science is not only a field for statisticians or programming engineers, but also a field apt for sociologists, whose role is to provide a thorough understanding of social structures and individual behaviors. That is why mixed methods using data science offer new breakthroughs to traditional sociological understandings of phenomena

Two Case Experiences

The direct utility of quantitative data analysis using data science is to increase the procedural efficiency of the research. Quantitative data analysis alone cannot provide enough sociological insight, which is why qualitative research is necessary. However, if we can easily determine in advance whether to conduct qualitative research, which requires a great deal of effort, it will greatly increase the efficiency and productivity of our research. This is where the high efficiency of the mixed method of conducting quantitative data analysis first exhibits its strength.

Below, I will introduce two firsthand experiences with using this particular mixed methodology approach. Both attempts were unsuccessful and went uncompleted. Despite these failures, however, I believe they are sufficient to illustrate the usefulness of pursuing quantitative data analysis using data science for exploratory purposes before delving into qualitative research.

Keyword network of GMO News

GMOs are one of the great scientific achievements of our time but have an ambivalent meaning that can even be considered dangerous by the public. Indeed, in South Korea, a movement against GMOs has been led largely by civic organizations. Therefore, in order to infer the public's attitude towards GMOs, I extracted keywords from Korean newspaper articles containing the term "GMO" and tried to grasp how the network of keywords changed. The actual outline of my quantitative data analysis is as follows.

1. To find and collect articles containing the keyword "GMOs" for the following two periods in BigKinds,[3] which provides a service for searching newspaper articles.
   A. Period 1: November 1, 2010 – March 1, 2011 (5 months)
   B. Period 2: November 1, 2015 – March 1, 2016 (5 months)
2. To visualize[4] the keyword networks based on frequency in all article contents corresponding to each period.

The thickness of the line between the three keywords "Organic - Certification - Sign" is very thick. This means that this keyword connection is very powerful. This powerful connection, however, disappeared completely in the articles published during the second period of data collection, between 2015-2016. This shift could lead us to believe that there was a significant change in the public's attitude toward GMOs, and that the change lowered the public's interest in the institutional process of "certifying and labeling organic produce."

At this stage in the research, I had captured one important research topic. The next step was to determine whether these changes found in keyword networks were real or only coincidental. If I determined the change real, I would need to investigate what caused the change. Only then will it be possible to identify how the public's interest in GMOs had changed and what was driving such a change. To do this, it would be necessary to apply a qualitative methodology to find the appropriate focus groups.

Due to several practical constraints, I have not yet completed any subsequent qualitative research on this subject. However, the case presented above is sufficient to confirm that exploratory data science analysis can

[3] https://www.bigkinds.or.kr/

[4] Both data mining and visualization are utilized by R and qgraph package.

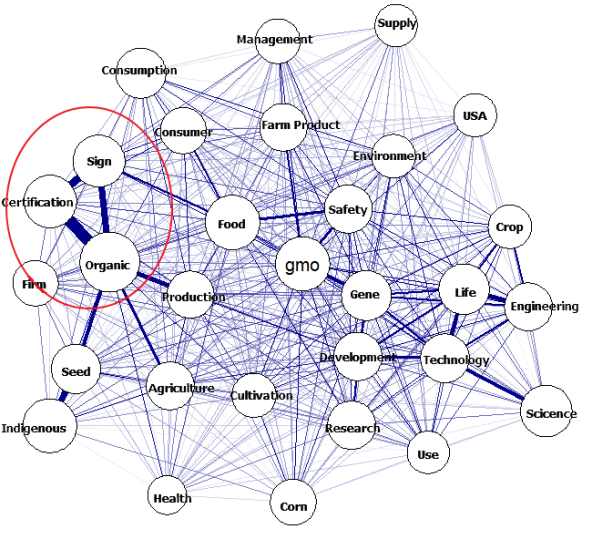

Fig. 7.—Keywords Network of GMO, Period 1 (2010-2011)

increase the efficiency of research by effectively narrowing the research focus before exploring qualitative research, which is time-consuming and expensive.

Public reaction to artificial intelligence through AlphaGo vs. Lee Sedol

In March 2016, there was a historic battle between Google's AI Go software, AlphaGo, and human professional Go player, Lee Sedol. The game was the subject of worldwide attention, especially in Korea. Of the five matches, Lee Sedol won only the fourth match, and AlphaGo won the overwhelming victory in all other games. This was a symbolic event that declared that AI had outpaced human abilities.

In fact, before the start of the game, the public predicted Lee's victory. This was because of Go software that had appeared in the past, a majority had performed poorly. But AlphaGo won a clear-cut victory over the world's best Go player, proving the public's predictions totally out of line. I was very interested in this point. On the very day that the 0:3 defeat of human representative Lee Sedol was confirmed, I anticipated that the public would have been surprised by the power of artificial intelligence. Therefore, I hypothesized that the public would have a vague fear of AI technology. Next,

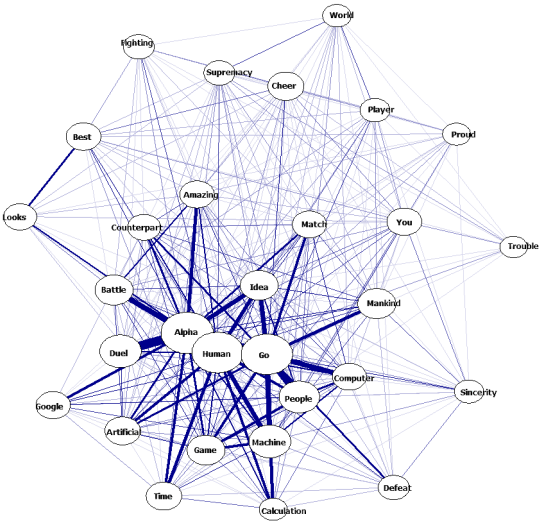

Fig. 8.—Keyword Network of AlphaGo

I found the news articles with the most comments among the relevant news published on March 12, 2016, the day when Lee's 0:3 defeat was confirmed, and extracted all the comments to conduct a keyword analysis. The steps of my analysis are as shown below:

1. To find five articles with a lot of comments in the news related to AlphaGo on March 12, 2016.
2. To extract all comments of relevant news articles: total 9,099 comments.
3. To extract keywords from the comments and visualize[5] the network.

The figure above shows the results of this visualization. This analysis confirmed that the assumptions I had made were completely wrong. With the overwhelming performance of artificial intelligence technology, I had expected the public to express fear, awe, or shock, but my prediction was completely overturned. The Korean public was much more interested in the defeated Korean representative Lee Sedol than in the AI technology, AlphaGo, and they were proud of him. No fear or awe could be found in their comments.

[5] Using the R and qgraph package as in the previous case.

A mere single quantitative data analysis does not indicate that this is the correct conclusion. However, the results clearly show that the research design needs to be re-established, or at least the data must be found somewhere else. Therefore, there is no need to conduct the next step, qualitative research. If I had been convinced of my assumptions, and had carefully established and carried out my research procedure, the research would be in big trouble. The fact that it is not necessary to carry out follow-up research after using data science for exploring a topic highlights the increase in the efficiency of the research in that it allows us to avoid wastes of time and resources.

Implications of Mixed Method with Data Science

This article presents the practical utility of mixed methods for performing qualitative analysis after quantitative analysis. In particular, the recent advances in analytical techniques for data science make this approach more useful. This mixed method approach presents several implications for sociological research.

First, sociologists must be able to handle diverse types of data. Sociology, which conducts research on social phenomena, encounters various forms of data during its research. The ability to integrate these diverse types of data into research designs is essential to sociologists. Therefore, training on this is also crucial. Second, sociologists should be capable of carrying out a whole series of processes from data production to processing and analysis. Sociologists who can perform quantitative analysis only with externally given data, already processed and neatly organized, can only conduct partial research. On the other hand, researchers conducting qualitative research have already trained themselves to find, process, and organize data on their own. However, the efficiency of their research and the productivity of their articles are also significantly lowered, which puts them at a disadvantage. Good sociological research requires a complete analysis and interpretation of the subject. To this end, there is a real need to combine the research strategy of qualitative methodologies with the efficiency of quantitative methodologies. In this sense, the importance of mixed methods has recently become more prominent, and all sociologists should be prepared and trained in its methods of analysis.

Closing Remarks: Overcoming the "Two Cultures"

The discourse of a crisis in sociology has been persisted for many years, to the point that it has now become cliché. The problem outside of sociology, which brought about the crisis within sociology, certainly exists. However, it is not an excuse for the problems within the field of sociology. Some have pointed out that sociological research does not contribute to realistic problem solving (Burawoy 2005). The reason that sociology does not provide proper problem solving is that the problems it claims to shed light on have not been sufficiently analyzed. Sociological research that responds to social problems should be able to analyze the data thoroughly, but the actual sociologist doing this work is in fact only a half-sociologist when they deal only with either quantitative or qualitative methods and not both. Therefore, overcoming the methodological 'two cultures' of sociology is the most important challenge facing our field. In this regard, mixed methods should be established as the only method of sociological practice, not simply considered another academic fad.

In academia, the pressure to publish is always present and growing. Highly productive researchers are, of course, better than less-productive researchers, but high productivity does not guarantee the quality of research. In general, the quality of research tends to be proportional to the time and cost spent on the research in question. As can be seen in the cases presented earlier, the mixed method with data science offers new opportunities to dramatically change the trade-offs in research productivity and quality. I look forward to sociology making good use of this new opportunity and moving beyond the methodological 'two cultures' rift, opening up a new era of daring sociological inquiry.

(*Submitted*: July 31, 2019; *Accepted*: September 5, 2019)

References

Atif, Amara, Deborah Richards, and Ayse Bilgin. 2013. "A student Retention Model: Empirical, Theoretical and Pragmatic Considerations." 24th Australasian Conference on Information Systems, ACIS2013. RMIT University.

Burawoy, Michael. 2005. "For Public Sociology." *American Sociological Review* 70(1): 4-28.

Conway, Drew. 2013. "The Data Science Venn Diagram." Retreived July 30, 2019

(http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram)
Creswell, John W. and Vicki L. Plano Clark. 2017. *Designing and Conducting Mixed Methods Research*. Sage publications.
Edmonds, W. Alex and Thomas D. Kennedy. 2017. *An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods*. 2nd Edition. Sage Publications.
Fox, Nick J. and Pam Alldred. 2018. "Mixed Methods, Materialism and the Micropolitics of the Research-Assemblage." *International Journal of Social Research Methodology* 21(2): 191-204.
Google Trends Result of "Big Data" and "Data Science". Retreived July 30, 2019 (https://trends.google.com/trends/explore?date=all&q=big%20data,data%20science)
Hamberg, Katarina, Eva Johansson, Gerd Lingren, and Goran Westman. 1994. "Scientific Rigour in Qualitative Research: Examples from a Study of Women's Health in Family Practice." *Family Practice* 11(2):176-181.
Leavy, Patricia. 2017. *Research Design: Quantitative, Qualitative, Mixed Methods, Arts-based, and Community-based Participatory Research Approaches*. Guilford Publications.
Lupton, Deborah. 2014. *Digital Sociology*. Routledge.
Marres, Noortje. 2017. *Digital Sociology: The Reinvention of Social Research*. John Wiley & Sons.
Orton-Johnson, Kate and Nick Prior. 2013. eds. *Digital Sociology: Critical Perspectives*. Springer.
Ozawa, Sachiko and Krit Pongpirul. 2014. "10 Best Resources on … Mixed Methods Research in Health Systems." *Health Policy and Planning* 29(3): 323-327.
Savage, Mike and Roger Burrows. 2007. "The Coming Crisis of Empirical Sociology." *Sociology* 41(5): 885-899.
Snow, Charles Percy. [1956] 2012. *The Two Cultures*. Cambridge University Press.
Traag, Vincent and Thomas Franssen 2016. "Revealing the Quantitative-Qualitative Divide in Sociology using Bibliometric Visualization." February 09, 2016. Retrieved July 30, 2019 (https://www.cwts.nl/blog?article=n-q2v294)

YEUNCHUL KIM is Lecturer in Seoul National University of Science & Technology, South Korea. He received his Ph.D. in Sociology from Seoul National University. His main research themes are Sociology of Science, Information Technology and Innovation, and Sociological Theories. *Address*: 101-401 Gwanak-ro 263 Gwanak-gu Seoul South Korea. [*E-mail*: yeunchul.kim@gmail.com]