



## Possibilities and Potentialities of Coalescing Processed Produced Survey data and Elicited data in Sociology of Deviance

Khumo Motshwari

PhD Student, Faculty of social sciences; Department of Sociology, Universitat Augsburg, Germany

E-mail: [motshwarikp@gmail.com](mailto:motshwarikp@gmail.com)

<http://dx.doi.org/10.47814/ijssrr.v6i8.1384>

---

### **Abstract**

Abstract: Processed produced data has always been criticized for its limitations and inability to address certain kinds of research questions because it is data not produced for research purposes. Process produced data such as administrative data collected by government ministries and organizations, constitute an important data source that can be useful in research, yet often not sufficient alone to answer different kinds of research questions, especially when they are used in isolation. Whilst these limitations are evident and indisputable, it is equally clear that such data can be effectively harnessed to answer specific types of research questions, and this paper offers an example of the usefulness of process produced data in studying juvenile delinquency. This paper intends to use the Botswana Youth Risk Behavioral and Biological Surveillance Survey (BYRBBSS II) by the Ministry of Basic Education as a form of process produced data set, to analyze the factors associated with juvenile delinquency in Botswana. Whilst this report provides a basis from which to launch the study, it cannot answer other uniquely qualitative research questions that would require thorough descriptions from the juveniles involved in these acts of deviance. This paper, therefore, uses this argument as a starting point to explore the possibilities and potentialities associated with combining process produced data with research elicited data, with the ultimate aim being to demonstrate how doing so offers more satisfying research results. Mixing Processed Produced Data and Research Elicited Data (interviews) is a methodological approach that has not yet been extensively applied in the context of Botswana, and therefore the paper will open discussions on the possibilities of conducting research in this manner, especially where process produced data sets are publicly accessible and available.

**Keywords:** *Processed Produced Data; Survey Data; Research Elicited Data; Deviance*

### **Introduction**

Early sociologists such as Max Weber naturally triangulated different data sources, modern sociologists have mainly based their research and methodological discussion on research elicited data

(Baur, 2009). Researchers are collecting their own data which addresses their research questions. Currently, with the development of data infrastructure, there are vast amounts of data available for researchers to use in their own studies. These data are mostly collected and stored by organizations, governments, and companies for the purpose of their daily use in the workplace. Examples of these data include data sets, customer data bases, newsletters and memorandums, photos etc. The availability of these data served a good purpose in the research sphere during the Covid 19 pandemic. The new normal which included among other things temporary lockdowns, keeping a social distance, working from home, entry and travel restrictions which caused a major challenge and concern to researchers. These challenges, rapid technological advances in data infrastructure and availability as a result creates new opportunities for gathering and combining data for methodological research. These offers the potential of using available data collected and stored by organizations, governments, and companies to researchers.

According to Baur (2019), these data is termed processed produced data. When this data is collected, it is initially not collected for the purpose of conducting research, but for everyday use within the workspace. This poses a limitation and criticism by researchers for its inability to address the research questions especially when the data is used in research. This paper, therefore, asserts that processed produced data must be complemented with research elicited data because the central argument is that doing so may enhance data quality for specific research and gives richer analyses of social phenomena. This paper uses this as a starting point to explore the possibilities and potentialities associated with combining process produced data with research elicited data, with the aim being to demonstrate how doing so offers more satisfying research results. This a methodological approach that has not yet been extensively applied in the context of Botswana especially in juvenile delinquency studies, hence this paper will open discussions on the possibilities of conducting research in this manner, particularly where process produced data sets are publicly accessible and available. Lastly, this paper contributes to the debate of complementing process produced data and with elicited data as a research method for sociology of deviance.

## ***Definitions***

Processed produced data are data that are generated through the very process of living, working, interacting in the society (...) from plain material evidence through all kinds of artifacts to the varieties of symbolic representations of ideas, activities, and events whether drawings, tales, messages, or documents (Baur, 2009: p.7). This data is basically anything available that researchers can use as a social science data source.

Research elicited data (also known as research induced data) is data produced by social scientists for research purposes (Baur et al., 2020) while Salmons (2016) defined it as data drew or produced from participants in response to the researchers' questions.

Deviance are the behaviors that violate significant social norms and expectations and are negatively valued by many people (Nalah & Ishaya, 2013).

## ***Processed Produced Data***

Processed produced data is administrative data collected by government ministries and organizations and are not sufficient alone if they are to be used in research studies. The reason for this is because the data is usually not produced for research purposes but are a side product of social processes (Baur, 2019). Social process in this case means data collected on a day to day, monthly or yearly basis for various activities within an organization. One can imagine, information collected on day-to-day affairs of

an organization can be complex in its form and the variations are based on the intention of the data collected. In Baur et al. (2020)'s views, processed produced data usually consist of both qualitative and quantitative components from a variety of data types (e.g., numerical, verbal, and visual data) and data sources. However, there are also non-mixed processed produced data which contain only quantitative or qualitative component. It is important to highlight that even though processed produced data can be in form of quantitative (surveys) or qualitative, they serve as processed produced data because they were not produced for research. Processed produced data have in common with secondary data that they were originally collected at an earlier time by a different person from the current researcher (Johnson and Turner, 2003). This creates problems for researcher, because the variables the researcher needs due to their theoretical approach cannot be found in the data.

The availability of processed produced data is redefining the role played by processed produced data in social science research at such a time as this. In Germany and other European countries, processed produced data have been used by scientists and non-scientists for more than 200 years ago (Baur et al., 2020). This is becoming prominent in Botswana because there are several data that the government of Botswana has invested in collecting, and therefore researchers can use these available data. At least in Botswana there are other few significant reasons why sociologists and other researchers have recently started using processed produced data in a form of survey data sets. A few of them are their increased availability, permission by the government to analyze data and lack of funds to do research or PhD studies. The use of processed produced data is important because it saves resources, time and allows the researchers to make use of the available processed produced data. This is a methodological advancement that may help regarding research in Botswana.

### ***Limitations of Processed Produced Data***

It is evident from the literature that there are limitations of using processed produced data. However, these limitations foster the integration of research elicited data. One of the main disadvantages of processed produced data is that the researcher cannot not control the process of data production and therefore must live with whatever blind spots the data has (Baur, 2009). This means that scholars using these data must use whatever information is collected since they were not part of the process of data production. Blind spots are things or aspects which the method, definition or theoretical approach do not allow to be seen or said (Maunganidze, 2020), or the researcher did not consider because they did not think or know about (Matthews, 2019). Though scholars consider blind spot as errors, see Bador (1994), Bick and Muller (1984), argued that the blind spots on processed produced data are not errors because they are not data producing institutions and it is not their fault that their needs might not fit the needs of researchers. The other limitation is that processed produced data usually has large volumes of data, which some of it cannot be used to address the research questions and theoretical questions. Data of any sort are a source for building and testing theories (Baur, 2009). It is important to note that processed produced data in its nature cannot address all the theoretical questions and this may point to a possibility that this may to some level affect possible links to data. That is why it is obvious and always the case that when tackling the relationship between theory and data, one must start from data to check if there is no good fit between the processed produced data and the theory, otherwise it will be impossible to assess the validity of the research. In addition, it is important to highlight that most of processed produced data may not be guided by theory, and if it so the researcher does not know why the questions were asked.

### ***Elicited Data***

The limitations of processed produced data to some extent have led to researchers dismissing it for its limitations, but this should not be the case because there are ways of overcoming these challenges

by introducing elicited data to complement the processed produced data that is being used. Elicited data is when the researcher uses verbal or written questions to elicit responses to focus groups, interviews, and participant observation (Salmons, 2016). Here it is important to note that the researcher has control of the process of data production and that there is a lot of interaction between the researcher and participants. It is during this stage that the researcher collects qualitative or quantitative data to address the questions that process produced data could not address. This study advocates for the use of elicited data as a solution to the obvious drawbacks and limitations of processed produced data. This therefore leads to this study's contribution to the current debate on methodological discourse that processed produced data is relevant and sufficient to be complemented by elicited data.

### **Use of Processed Produced Data and Elicited Data in Juvenile Delinquency Studies**

The Botswana Youth Risk Behavioral and Biological Surveillance Survey (BYRBBSS II) which was collected by the Ministry of Basic Education is composed of quantitative data that has over 150 items. The purpose of the BYRBBSS II was to guide the design of appropriate prevention programs and monitoring the influence and success of programs in place while providing trend data of high-risk behaviors in schools (BYRBBSS, 2016). Through the BYRBBSS II, the Ministry of Basic Education, Centers for Disease Control and Prevention Botswana, Ministry of Health etc. wanted to monitor the success of national efforts. It is from these arguments that the researcher supposes the data collection was not guided by theory, hypothesis and that it is not scientific research. However, it is important to highlight that the report followed scientific procedures. The use of BYRBBSS II which is form of processed produced data plays a vital role in supporting this paper's purpose of demonstrating that for juvenile delinquency studies, data sets in survey form are not sufficient on their own and that a single data type cannot answer all the theoretical aims. An integrated research design (Kelle, 2008) can combine qualitative and quantitative research, and can combine the advantages of both data types and diminish the methodological disadvantages of qualitative and quantitative research respectively. Therefore, to enhance the study findings and yielding satisficing results, the researcher will collect qualitative elicited data through in-depth interviews to address the questions that processed produced data could not answer and to provide a better analysis of factors associated with juvenile delinquency in Botswana.

### **Benefits of Coalescing Botswana Youth Risk Behavioural and Biological Surveillance Survey (BYRBBSS II) with Elicited data in Conducting Research**

There were various benefits that were identified to be associated with coalescing Botswana Youth Risk Behavioral and Biological Surveillance Survey (BYRBBSS II) with elicited data and hence providing a better understanding of juvenile delinquency. It is imperative to highpoint that juvenile delinquency particularly the nature, the causes and outcomes can be better understood by using different data types. Coalescing BYRBBSS II and elicited data allows for combination of variety of data types and resulting in development of a new method that could yield potentials and benefits to the body of research. This then redefines the role played by processed produced data and elicited data in social research, as well as bringing to light a continued value and their importance as a tool for understanding the society. Furthermore, it then creates the possibility of coalescing processed produced data and elicited data and this development can result in new opportunities for methodological research aimed at identifying the best practice principles and providing recommendations to researchers who work with both processed produced data and elicited data.

The use of BYRBBSS II nor elicited data alone was not sufficient to yield satisficing results on discussions about factors associated with juvenile delinquency. The coalescing is then needed to minimize the individual limitations from each data type, and to provide a better understanding of deviant behavior interactions. There is evidence that processed produced data has limitations and has always been criticized for its inability to address all theoretical questions. Thus, coalescing of processed produced data

and elicited data remains valuable because one data type cannot answer all the theoretical questions. This statement has steered a methodological question this paper is trying to answer of; can one data give answers to all theoretical questions? Depending on researcher's theoretical viewpoint, the answer is "no" especially in juvenile delinquency studies. The coalescing of BYRBBSS II and elicited data was a clear demonstration that supports the answer to this question. The next paragraph provides clear examples that also tried to support the answer to this question.

It was identified that there is a good fit between BYRBBSS II and the research questions, meaning that there were some research questions that could be addressed by the BYRBBSS II; what is the level of juvenile delinquency among young people? What are the types of juvenile delinquency? Which demographic variables are associated with juvenile delinquency? Which family background variables are associated with juvenile delinquency? The answers to these questions were statistical tests which to some points were not sufficient to explain juvenile delinquency fully. These provided a description, for example, what happened and who did it? For this, the researcher can use process produced data. However, there are some problems with the theoretical guided analysis, for example, why did they do it? Which to some extent these descriptions cannot provide. In addition, due to the insufficiency of the statistical tests, there were some research questions that the BYRBBSS II could not address. For example, which socio-economic variables are associated with juvenile delinquency? How can the statistical results that will be obtained in phase 1 be explained? Why the selected factors identified in phase one influence young people to be juvenile delinquents? A statistical correlation cannot tell us how a factor sexual harassment influences a young person in a way that he/she becomes delinquent in the end. Typically, we find that there is more than one important factor, for example, different stages of decision making which at the end of this process there is a delinquent action. To get to the answer, some hints from qualitative research are needed. What are other additional factors from the viewpoint of the interviewee? These qualitative questions require thorough descriptions from the juveniles that are involved in various acts of delinquency.

From the previous paragraph, it is evident that BYRBBSS II alone could not answer all the research questions. This is mainly because juvenile delinquency as a social problem is complex in nature, and it originates from personal motivations hence making it difficult to understand it by relying on BYRBBSS II alone. Besides that, due to the nature of the research problem, it is hard to collect very good research elicited data because in most cases, respondents who are delinquent usually hesitate to admit their way ward behavior in surveys (Baur et al., 2020). Based on these arguments, coalescing is relevant and suitable to provide a better measure of juvenile delinquency.

Coalescing BYRBBSS II with elicited data within one study saves time and resources because it has revealed to reduce time spent in the research field while collecting. The use of BYRBBSS II meant that the researcher will spend a short time collecting elicited data in the research field because BYRBBSS II was already collected in 2016. This has also proven to be beneficial to the research community in Botswana. As a result, it then become advisable that since there are vast amounts of data are being collected and archived by researchers and organizations all over the world, and the practicality of utilizing existing data for research should become more prevalent (Johnston, 2013).

Processed produced data face several problems, severest being that from a statistical point of view, they are biased in format, purposes, and institutional embeddedness of the pertinent data type (Baur, 2009). This means that the data was collected for a different purpose to that of the researcher and comes in a certain format or type, which forces the researcher to deal with whatever information was collected. BYRBBSS II was collected in a survey form with the Ministry's purpose of monitoring the trends of high-risk behaviors in schools which makes the BYRBBSS II bias. Apparently, the biasness comes from the fact that the Ministry of Basic Education collected only information that will benefit their mandate and inform their policies, leaving out important items that could be of importance to researchers. This

then makes researchers to deal with whatever information was collected. The other limitation noted was the measurement precision of the items from the BYRBBSS II was low. For example: in the data set, respondents were asked whether they engaged in sexual activity? The response categories for this question were 'yes' and 'no'. Since there was no theory behind this question, there will be no possibility to make a test for construct validity. Instead of this question being asked the way it was, respondents could have been asked how often they engaged in sexual activities. This would have been a better measure for engagement in sexual activity because ordinal scales are necessary for a hypothesis testing using MPLUS. The introduction of elicited data gives the researcher an opportunity to collect qualitative data that addressed the biasness resulting from the BYRBBSS II format and data type. At the same time, it also allows the researcher to control the entire process of data collection.

### **The Challenges of coalescing Botswana Youth Risk Behavioural and Biological Surveillance Survey (BYRBBSS II) with Elicited Data in Conducting research**

Processed produced data is collected during a certain period before elicited data is collected. The time difference raises questionability to 1. Compatibility – do they talk the same thing? 2. Validity- The researcher must check how results correspond to established theories and other measures of the same concept. 3. How do you link a qualitative data which is from a different sample with BYRBBSS II? The answer to this is that the researcher must make sure that choice of respondents for qualitative interviews is a sample of the same characteristics as the respondents for BYRBBSS II, as it is not about the individuals but the issues or the problems.

### **Conclusion**

In conclusion, the central argument is, complementing processed produced survey data with qualitative elicited data enhance data quality for specific research and doing so gives richer analyses of social phenomena. The quality of processed produced survey data for juvenile delinquency studies can be enhanced by introducing qualitative elicited data, as there is evidence that depending on processed produced survey data only may not yield satisfying results. No data in the form of processed produced data or elicited data can give satisfying results about juvenile delinquency. There are benefits associated with coalescing processed produced data with elicited data, and these benefits have proven to reduce individual limitations of each data type. This paper contributes to the methodological discourse by asserting that though the use of processed produced data is not new, what is new is the possibility of coalescing processed produced data with elicited data to obtain satisfying results. Lastly, the debate of coalescing processed produced data with elicited data has demonstrated to be a beneficial research method in sociology of deviance and has magnificently explored the possibilities and potentialities which have yielded virtuous outcomes.

### **References**

- Bador, E. (1994). *Controversies around the Use of Secondary Data*. London: Sage Publications.
- Baur, N., Graeff, P., Braunisch, L., & Schweia, M. (2020, July). The Quality of Big Data. Development, Problems, and Possibilities of Use of Process-Generated Data in the Digital Age. *Journal of Historical Social Research*, 45(3), 209-243. doi:10.12759/hsr.45.2020.3.209-243.
- Baur, N. (2019, June). Linearity vs. Circularity? On Some Common Misconceptions on the Differences in the Research Process in Qualitative and Quantitative Research. *Journal Frontiers in Education*, 4(53), 1-16. doi:10.3389/feduc.2019.00053.

- Baur, N. (2009). Problems of Linking Theory and Data in Historical Sociology and Longitudinal Research. *Journal of Historical Social Research*, 34(1), 7-21.
- Bick, W., & Müller, P. J. (1984). Sozialwissenschaftliche Datenkunde für prozeßproduzierte Daten: Entstehungsbedingungen und Indikatorenqualität. In W. Bick, R. Mann, & P. J. Müller (Hrsg.), *Sozialforschung und Verwaltungsdaten* (S. 123-159). Stuttgart: Klett-Cotta.  
<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-330744>.
- Ministry of Basic Education Botswana. (2016). *Second Botswana youth Risk Behavioural and Biological Surveillance Survey Report II*. Gaborone: Botswana.
- Johnson, R. B., Turner, L. A. (2003): Data Collection Strategies in Mixed Methods Research. In: Tashakkori, A./Teddlie, C. (Eds.): *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks et al.: Sage. 297-320.
- Johnston, M. (2013). Secondary Data Analysis: A Method of which the Time Has Come. *Qualitative and Quantitative Methods in Libraries*, 3, 619-626.
- Kelle, U. (2008). *The Integration of Qualitative and Quantitative Methods in Empirical Social Research*. VS Publishing House for Social Sciences. (DOI: 10.1177/1478088706070839).
- Matthews, K. (2019). How to Find and Resolve Blind Spots in Your Data. Retrieved from <https://www.smartdatacollective.com/how-to-find-resolve-blind-spots-in-data/>.
- Maunganidze, L. (2020). Excavating Blind Spots in Qualitative Research: Reflecting on a Lecturer's Experiential Past. *Mosenodi Journal*, 23(2), 41-62.
- Nalah, A. B., & Ishaya, L. D. (2013, December). A Conceptual Overview of Deviance and Its Implication to Mental Health: a bio psychosocial Perspective. *International Journal of Humanities and Social Science Invention*, 2(12), 1-9. doi:10.1080/15248372.2012.646613.
- Salmons, J. (2016). *Doing qualitative research online*. London: Sage Publications.
- Upton, S. (2020). The relating facts to Covid- 19 Transmission through paper surfaces. *Two sides*, 2(1), pp. 1, 2. Retrieved from <https://www.twosides.info/uk/covid-19-and-paper>.

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).