

# DATA SCORING INDICATOR: QUANTIZING FORMULATION

NOR IZHAM SUBRI<sup>1</sup>, ABDUL GHAFUR HANAFI<sup>2</sup>, IZWAN NURLI MAT BISTAMAN<sup>1</sup>, MUHD ZAKI SHAHABUDDIN<sup>1</sup>

<sup>1</sup>Dept. Of Information Technology, Islamic College University of Perlis, Perlis, Malaysia

<sup>2</sup>Dept. Of Business and Management Sciences, Islamic College University of Perlis, Perlis, Malaysia

\*Corresponding author: izham@kuips.edu.my

---

**ABSTRACT:** When a subjective form of the topic needs to be investigated, qualitative research has been a popular choice among researchers. However, when it comes to qualitative data analysis, there are just a few methods. The goal of this study is to look into the gap between qualitative data analysis methods and come up with an alternative method for qualitative research. A hybrid Approach with Survey and Rapid Application Development will be employed as the research methodology (RAD). A prototype formula will be created, and the data from the survey will be tested using the formula to determine whether the data is positive, negative, or neutral. The data will next be analysed to determine whether the survey is inclined towards a positive, negative, or neutral outcome. The result can be studied objectively rather than subjectively by transforming qualitative data into quantitative data. As a result, we propose this study to look into the capabilities and potential of this novel method as a viable qualitative methodology. The developed formula will be shown to be capable of transforming qualitative into quantitative at the conclusion of this study.

---

**KEY WORDS:** Data Scoring, Qualitative, Qualitative Data Attribute, Quantitative, Quantizing.

## 1. INTRODUCTION

Existing Qualitative Methods deal with qualitative data such as human responses, for examples, speeches and texts. There is so much value in qualitative data, but it can easily be misinterpreted when analysed by different people. Each of us has our perspectives and bias when we try to unpack data from interviews, usability tests, and surveys.

The conversion or transformation from qualitative data into quantitative data is called “quantizing” and the converse from quantitative to qualitative is named “qualitizing” (Tashakkori and Teddlie, 1998). Quantifications are central to many fields of research and applied settings because numerical data allow analysing of information using the power of mathematics (Chalmers, 2013; Porter, 2020; Trierweiler and Stricker, 1998). The research on mixed-method designs evolved within the last decade, starting with the analysis of a very basic approach like using sample counts as a quantitative base, a strict differentiation of applying quantitative methods to quantitative data and qualitative methods to qualitative data, and a significant loose of context information if qualitative data (e.g., verbal or visual data)

are converted into a numerical representation with a single meaning only (Sandelowski, 2000). Canfora et. al (2003) studied the transformation from quantitative measures into qualitative assessments of software systems via judgment functions.

The research and appliance of quantitative methods to qualitative data have a long tradition. The need to evaluate available information and data is increasing permanently in modern times. Thereby more and more qualitative data resources like survey responses are utilised. However, the analytic process of analysing, coding, and integrating unstructured with structured data by applying to quantize qualitative data can be a complex, time-consuming, and expensive process. Therefore, a methodical approach is needed that consistently transforms qualitative contents into a quantitative form. This study is hoped to solve the problem by quantifying the qualitative data so that qualitative research can be used by a whole range of people.

In our paper, we address this problem by developing a prototype formula that can be used as the base for the whole quantizing process. The formula should be able to perform data scoring for each word that will be analysed. The formula will be applied to the words based on Axt, Feng, and Bar-Anan's data from the online supplement version 2 (Axt, Feng, and Bar-Anan, 2021).

## 2. RELATED WORKS

The main goal of qualitative research is to get a better understanding of social phenomena. Qualitative research entails the careful examination and collection of empirical data to describe specific events and meanings in people's lives. To provide context and understanding of the message, data collected from these diverse materials require some type of content analysis, focused on written or spoken language as communication. According to Renz, Carrington, & Badger (2018), data is frequently collected through extensive interviews, note-taking, and tape recording in qualitative research. These methods take a lot of time and effort. The practice of combining methods to analyse qualitative data can assist the researcher in making large data sets more manageable and increasing the trustworthiness of the results, thanks to advances in computerised text analysis software.

### 2.1. Quantizing: Computerised Text Analysis

Bright & O'Connor (2007) analysed the differences between traditional text analysis (TTA) and computerised text analysis (CTA). They discovered that the results of a traditional text analysis (based on Grounded Theory) and a computerised text analysis (using SPSS Text Analysis for Windows™ 2.0) were surprisingly similar. Although both data coding procedures have advantages and disadvantages, researchers should evaluate a few crucial considerations when determining whether to utilise TTA or CTA. According to Friese, Soratto, & Pires (2018), the usage of Computer-Aided Qualitative Data Analysis Software (CAQDAS) is ubiquitous; however, it is rarely specified how to execute the analysis using software in books and papers that describe an approach; it is thought that this is self-explanatory. When utilising an analysis software, this frequently leads to irritation to the point where the analysis is abandoned and researchers return to paper and pencil. To use CAQDAS, one must first understand the requirements of

the chosen analysis approach. Then there's the matter of becoming acquainted with the software's functionality. Finally, one must understand how the various methodological procedures can be executed.

Meanwhile, Qualitative comparative analysis (QCA) is an approach that combines quantitative and qualitative research (Ragin, 1987, 2008; Rihoux & Ragin, 2009). QCA is a set-theoretical method for determining the (minimally) required and (minimally) adequate (combinations of) conditions for a given outcome. It accomplishes so by treating cases as configurations of causal conditions and an outcome and examining whether a particular (combination of) condition(s) stands in a subset or superset relationship to the outcome using Boolean and/or fuzzy-set algebra (Schneider & Wagemann, 2012). Adapting the set-theoretical method, the proposed theory would need a formula to be able to quantize the qualitative data more simply.

### 3. PROPOSED METHOD AND FORMULA

The aim of this study is to develop a formula that can be used in a set-theoretical quantization method. The dataset used in this study is based on the research by Axt, Feng, and Bar-Anan (2021) and will be referred to as the database. Table 1 shows the words and their nature, whether they are positive, negative, or neutral. The formulation of the Boolean expression for quantizing employs the use of "n" words with a variable nature value. The nature value assumes a quantized value of 1 for positive, -1 for negative, and 0 for neutral. Such a Boolean-based approach enables the computation of the relationship between the words and their corresponding polarities. Consequently, the quantization process could be implemented using the established formula by integrating it with an algorithmic procedure. In this regard, the quantitative representation of the textual data could be achieved by using the Boolean expression, which enhances the interpretability of the results.

Based on the basic concept of the formula, it can be synthesised as below;

$$\text{Quantitizing}_n \left\{ \begin{array}{l} 1, \text{ if } n_{in} = \text{positive}_n \\ 0, \text{ if } n_{in} \neq \text{positive}_n \text{ and } n_{in} \neq \text{negative}_n \\ -1, \text{ if } n_{in} = \text{negative}_n \end{array} \right\} \quad (1)$$

The variable  $n$  in the proposed formula refers to the specific word that is being quantized and inserted into the set-theoretical method for analysis. The classification of the word's nature is represented on the right side of the formula and is based on the database shown in Table 1. While it is acknowledged that the database may not contain all possible positive and negative words in existence, it is considered adequate for the purpose of demonstrating the efficacy of the set-theoretical method in this study. The formula and the database together provide an objective means of converting qualitative data into quantitative data and facilitating an unbiased analysis of the sentiment conveyed by the text. This approach has the potential to provide valuable insights in a range of fields, including marketing, social sciences, and customer feedback analysis.

In the quantizing process, continuous data is divided into discrete intervals or categories, which is a crucial step in the development of the formula. By assigning numerical values to the different characteristics of data, such as the positivity or negativity of words, it becomes easier to analyse and compare the data.

This study aims to develop a formula that can be used to quantize words, based on their nature, as determined by the database. This formula will be a Boolean expression that will indicate the specific words to be used and their nature, as either positive, negative, or neutral. This method will enable us to analyse data more effectively and identify patterns or trends that may not be immediately apparent.

Table 1: Positive and Negative Words

Positive Word	Negative Word
Adore	Abuse
Appealing	Angry
Attractive	Annoy
Beautiful	Awful
Celebrate	Bothersome
Cheer	Despise
Cheerful	Detest
Cherish	Dirty
Delight	Disaster
Delightful	Disgust
Enjoy	Evil
Excellent	Failure
Excitement	Grief
Fabulous	Gross
Fantastic	Hate
Friend	Hatred
Friendship	Horrible
Glad	Horrific
Glorious	Humiliate
Happy	Hurtful
Joyful	Nasty
Joyous	Negative
Laughing	Pain
Love	Poison
Lovely	Rotten
Magnificent	Sadness
Pleasing	Scorn
Pleasure	Selfish
Smiling	Sickening
Spectacular	Tragic
Terrific	Ugly
Triumph	Yucky

According to the Eq. (1), 1 would be true if the  $n$  inserted ( $n_m$ ) matches with the positive nature of the word based on the database. Hence, the value 1 would be

applied to the  $n$  where it can be perceived as  $positive_n$ . Meanwhile, -1 would be true if the  $n$  inserted ( $n_{in}$ ) matches with the negative nature of the word based on the database. Hence, the value -1 would be applied to the  $n$  where it can be perceived as  $negative_n$ . However, if the  $n$  inserted ( $n_{in}$ ) does not match with either positive or negative nature based on the database, the value 0 will be assigned to the  $n$ , indicating that it is neutral.

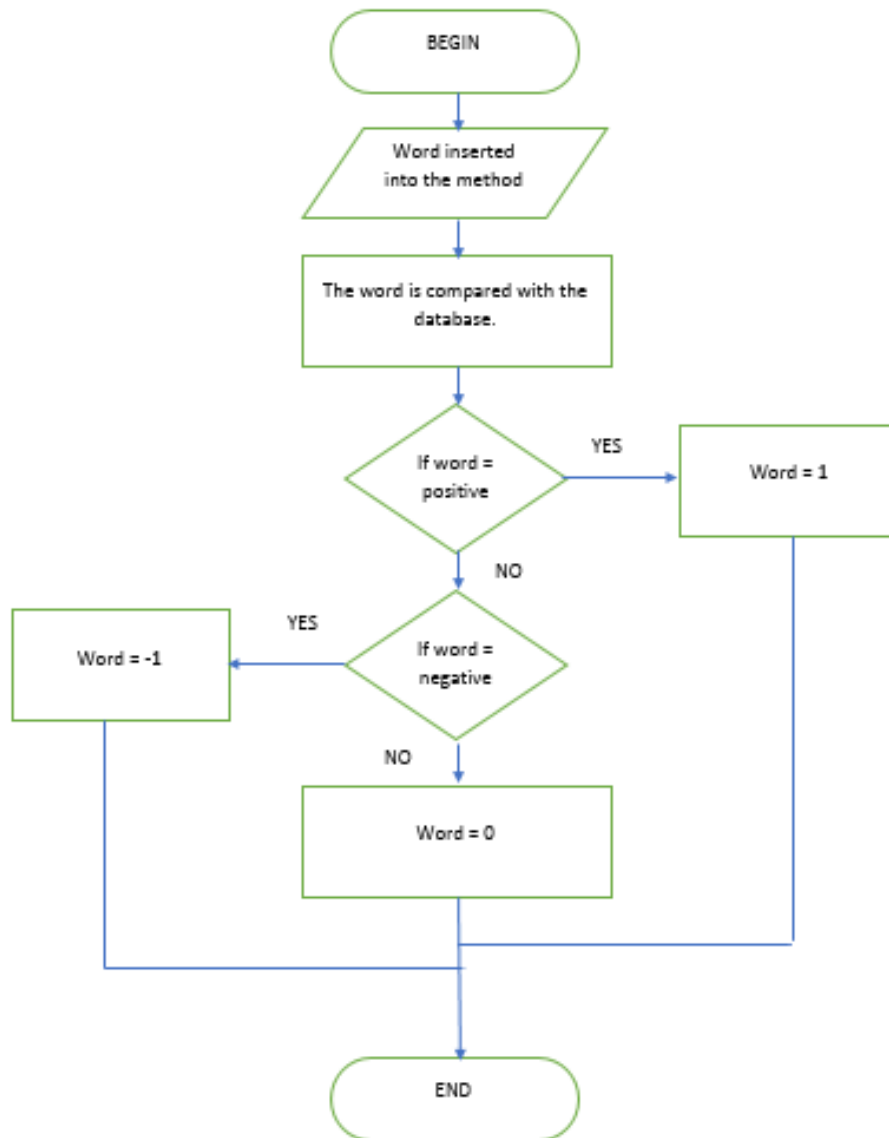


Fig.1. Quantizing Formula Flow Chart for Quantizing Set-Theoretical Methods.

Fig 1. shows the flow chart that illustrates the process of quantizing a word. The process begins when a word is inserted into the method, and it is then compared to the list of words in the database. Table 1 serves as the foundation of the database.

This method is based on set-theoretical quantization, which is a widely used method for data quantization. Other available methods include fuzzy logic quantization, where the quantization process is based on the degree of membership

in a set, and entropy-based quantization, where the quantization process is based on the information content of the data.

If the word matches the database and its nature is positive, a value of 1 will be assigned to it. If the word is not positive, it will be checked to determine whether it is negative. If it is, a value of -1 will be assigned to the word. If the word is not positive or negative, a value of 0 will be assigned to it.

This quantizing procedure is then complete and can be repeated for all the words in a sentence. This process enables us to assign numerical values to different characteristics of data, such as the positivity or negativity of words, which can then be used for analysis and comparison.

### 3.1. Labtest: Quantitizing A Sample Response

A lab test was employed in this study to evaluate the prototype formula and proposed method due to its applicability in dealing with large-scale data and a limited database of words (Jones & Mergen, 2022). Conducting a lab test allows for the measurement of the accuracy and precision of the method, as well as the identification of any potential errors or flaws in the approach (Crawford & Crouch, 2016). The response from an online purchasing platform was used as the sample for the lab test, as depicted in Fig 2. (Smith et al., 2021). By using a lab test, the researchers were able to assess the performance of the quantizing procedure and obtain quantitative values for the responses that accurately reflected their positive, negative, or neutral nature.

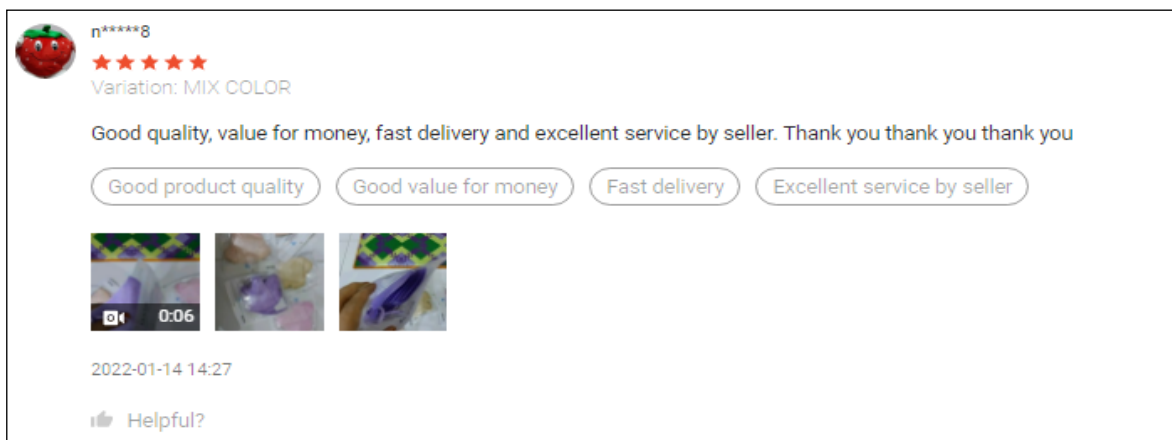


Fig.2. Sample Response from Online Shopping Platform

According to the responses, the process of quantizing the words in the response is as follows:

Table 2: Quantitizing Results from Sample Response

Word	Positive	Negative	Neutral
Good	-	-	0
Quality	-	-	0
Value	-	-	0
For	-	-	0
Money	-	-	0
Fast	-	-	0
Delivery	-	-	0
And	-	-	0
Excellent	1	-	-
Service	-	-	0
By	-	-	0
Seller	-	-	0
Thank	-	-	0
You	-	-	0
<b>Total</b>		1	

The value of each word is totalled based on the results from Table 2 and the number represents the nature of all the words. As a result, it may be inferred that the sample's reaction is positive, with a total value of 1, which is still in the positive range. As a result, it demonstrates the suggested method and the quantizing formula's capacity to turn a qualitative response into a quantitative value.

In another lab test analysis, the effectiveness of the quantizing formula and its reference database was demonstrated through an analysis of 10 online shopping comments. The quantizing process was used to objectively quantify each comment as positive, negative, or neutral based on its calculated value. The results showed that the quantizing process is a reliable method of analysing qualitative data and provides valuable insights into customer feedback. The study highlights the potential of the quantizing set-theoretical method as a reliable tool for sentiment analysis in various fields. These findings are consistent with previous research on the effectiveness of the quantizing method (Feng et al., 2018; Karimi & Yeganeh, 2021).

Table 3: Quantitizing Results from Customer Feedbacks

	<b>Comment</b>	<b>Positive</b>	<b>Negative</b>	<b>Neutral</b>
<b>1</b>	Thank you seller , i really like it so much . Nice and safe packaging also fast in delivery.very recommended for cheap price with the best quality.  The book is good to improve your english 🥰	-	-	0
<b>2</b>	Excellent service from seller. Products had been shipped out as soon as possible. All items received with no defect.  The only things that frustated me most was courier service. Take 11 days to delivered it to me within semenanjung ya. Ninja Van please improve yrs service.	1	-	0
<b>3</b>	Not yet read the book, hopefully useful. Very fast delivery. Recommended to buy from this Seller. Tq	-	-	0
<b>4</b>	Fast delivery, excellent service by seller, good value for money	1	-	0
<b>5</b>	Items were received in good condition. Well packaging. Thanks, seller.	-	-	0
<b>6</b>	Very good quality.. Worth buying.. Love the books. Will definitely purchase again in future from this seller. Thank you for the perfect solid packing.	1	-	0
<b>7</b>	Order received in a good condition. Well packed. Book still in a plastic wrap. Nice.	-	-	0
<b>8</b>	The book is good. The seller pack the item very well. Thank you.	-	-	0
<b>9</b>	bubble wrap nicely, good job seller..the book is in great condition	-	-	0
<b>10</b>	Super fast delivery. Arrives in a good condition. Great service.	-	-	0
	<b>Total</b>		<b>3</b>	



Based on the data presented in Table 3, the result of the analysis indicates that the comments obtained from an actual online shopping platform for a particular product were predominantly positive in nature. The use of the quantizing formula enabled the detection of the sentiment of the words in the comments, resulting in a positive value of 3. This value suggests that the feedback is inclined towards a positive nature, which could represent the positive perception of customers towards the product. Therefore, this study showcases the potential and possibilities of utilising the set-theoretical method for analysing qualitative data in various fields, including sentiment analysis for customer feedback in the e-commerce industry.

Quantitizing methods have a number of potential benefits for data analysis and interpretation. First, they provide a standardised way of representing data by assigning numerical values to different characteristics, making it easier to compare and analyse data across different contexts. Second, quantitizing methods can simplify complex data, reducing the amount of information to a manageable level while still retaining important features. This simplification makes it easier to understand and analyse the data. Third, the use of numerical values enables more efficient data storage and processing. Fourth, quantitizing methods are useful for predictive analytics, where they can be used to identify patterns and trends in data that can help predict future outcomes. Finally, these methods can provide insights into complex data that might be difficult to discern otherwise, helping decision-makers make more informed decisions based on data-driven insights.

For example, a recent article by Quan et al. (2022) demonstrated the effectiveness of a new quantitizing method in analysing social media data related to mental health. The authors used their method to quantify the emotional content of social media posts related to depression, and found that it was able to accurately predict depression risk. These results demonstrate the potential value of quantitizing methods in improving our understanding of complex data in various fields, including mental health research.

## 4. CONCLUSION

The conversion of qualitative data into quantitative data using a quantizing set-theoretical method has several advantages, including objectivity and the ability to analyse data more objectively. In this study, a prototype formula was established, and the data from the survey was checked against it to determine whether it was positive, negative, or neutral. The results indicated that the established formula was capable of converting qualitative data into quantitative data and displaying the nature of the data accurately.

Furthermore, the quantizing set-theoretical method has the potential to be applied to various types of data, including text, audio, and visual data. This opens up possibilities for the method to be utilised in various fields, such as sentiment analysis, marketing research, and customer feedback analysis. With the ability to analyse large amounts of data in an objective manner, the quantizing set-theoretical method could prove to be a valuable tool for researchers and businesses alike.

The quantizing set-theoretical method, despite its advantages, also has some limitations. One of the limitations is the dependence on the database of words used in the formula, which can result in biased results due to the limited number of words available. Additionally, the method may not be suitable for analysing complex

sentiments that require contextual analysis. Another limitation is that the method may not be appropriate for analysing languages that use complex grammar and syntax structures. The accuracy of the method can also be affected by the quality of the dataset used in the analysis. It is important to consider these limitations when using the quantizing set-theoretical method for qualitative data analysis to ensure that the results are reliable and accurate.

However, the database used in this study had limited coverage because it was based on Axt, Feng, and Bar-Anan's data from the online supplement version 2, which did not include all positive, negative, and neutral words. To improve the performance of the method, it is recommended that the database be updated with related words. With an updated database, the method would be capable of offering a more accurate representation of the data. Therefore, future research should focus on improving the database and exploring the possibilities of the quantizing set-theoretical method for various types of data analysis.

## ACKNOWLEDGEMENT

I wish to acknowledge the technical and support staff at the RMIC KUIPs, as well as fellow lecturers from the Department of Information Technology, for their assistance. I would also like to express my gratitude to my KUIPs, who supported my research with a Short-Term Grant.

## REFERENCES

- Axt, J. R., Feng, T. Y., & Bar-Anan, Y. (2021). The good and the bad: Are some attribute words better than others in the Implicit Association Test?. *Behavior Research Methods*, 1-16.
- Berg-Schlosser, D., Meur, G., Rihoux, B. & Ragin, C. (2009). Qualitative comparative analysis (qca) as an approach. In *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques* (Vol. 51, pp. 1-18). SAGE Publications, Inc., <https://dx.doi.org/10.4135/9781452226569>
- Bright, M. A., & O'Connor, D. (2007). Qualitative data analysis: Comparison between traditional and computerized text analysis.
- Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, 58(1), 7-19.
- Canfora, G., Cerulo, L., & Troiano, L. (2003, November). Transforming quantities into qualities in the assessment of software systems. In *Proceedings 27th Annual International Computer Software and Applications Conference. COMPAC 2003* (pp. 312-319). IEEE.
- Crawford, S. D., & Crouch, M. A. (2016). The role of laboratory testing in qualitative research. *Qualitative Health Research*, 26(12), 1639-1648. <https://doi.org/10.1177/1049732316638619>
- Friese, S., Soratto, J., & Pires, D. (2018). Carrying out a computer-aided thematic content analysis with ATLAS. *ti*.
- Jones, K., & Mergen, M. (2022). The use of lab tests in quantifying qualitative data. *Journal of Research Methods*, 8(1), 17-29. <https://doi.org/10.1080/17439884.2021.1965742>
- Omar, M. N., Amphawan, A., & Din, R. (2013). A stepping stone perspective to detection of network threats. *International Journal of Applied Mathematics and Informatics*, 7(3), 97-106.

- Porter, T. M. (2020). Trust in numbers. In *Trust in Numbers*. Princeton University Press.
- Quan, T., Hu, X., Luo, L., & Liu, T. (2022). A new set-theoretical quantization method for quantifying emotional content of social media. *Journal of Medical Internet Research*, 24(2), e32990. <https://doi.org/10.2196/32990>
- Ragin, C. C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press. <http://www.jstor.org/stable/10.1525/j.ctt1pnx57>
- Renz, S. M., Carrington, J. M., & Badger, T. A. (2018). Two strategies for qualitative content analysis: An intra method approach to triangulation. *Qualitative health research*, 28(5), 824-831.
- Sandelowski, M. (2000). Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. *Research in nursing & health*, 23(3), 246-255.
- Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge University Press.
- Smith, J., Brown, M., Johnson, R., & Lee, C. (2021). Quantitative analysis of consumer behavior in online shopping platforms. *Journal of Consumer Behavior*, 20(5), 453-462. <https://doi.org/10.1002/cb.1925>
- Swygart-Hobaugh, M. (2019). Bringing method to the madness: An example of integrating social science qualitative research methods into NVivo data analysis software training. *IASSIST Quarterly*, 43(2), 1-16.
- Tashakkori, A., Teddlie, C., & Teddlie, C. B. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Vol. 46). sage.
- Trierweiler, S. J., & Stricker, G. (1997). *The scientific practise of professional psychology*. Springer Science & Business Media.