IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals

Human Journals
Review Article
March 2020 Vol.:15, Issue:1

Skimmed Research Methodology

IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals

Renata Klafke*[1], Josué Flores[2], Krishna Devulapalli[3], Abhishek Nirjar[4]

[1]FTEC University Center Group – Brazil & Universidade Federal do Paraná (UFPR).

[3]Data Scientist, TapChief Expert in Data Science and Guest Faculty in BioStatistics, NIPER, Guwahati, India.

[4]Academic Administration: Director, IFIM Business School Bangalore.

Submission: 20 February 2020
Accepted: 27 February 2020
Published: 30 March 2020

HUMAN JOURNALS
www.ijsrm.humanjournals.com

Keywords: Methodology, Dissertation; Skimming Methodology, Text Mining, Word Cloud, Document Term Matrix

ABSTRACT

In this paper, we present a new methodology that can assist scholars in their academic post-graduation course. Our method is based on personal and practical teaching experience, as well as using this methodology in our own research in papers that have been published in academic journals. apart from self-methodology used to write and have published papers. We describe and explain in 7 Stages (steps), how it can be scientifically applied. This methodology utilizes as its foundation the principle of "skimming", therefore delimiting and sorting, the relevant articles published on a specific topic, which may be particularly beneficial to those in search of exploratory research questions and various orientations. During the last decade, text mining has become a widely used discipline utilizing statistical and Learning Analytics (LA). Unfortunately, it has been applied only for analysis of results and not as a mechanism to organize the state of the art of a subject or even develop the literature review. The Skimmed Research Methodology is an important technique for scholars and can deepen understanding of different dimensions and variables generating new insights and understanding. Additionally, it allows for the capturing of key data elements and concepts to be considered when conceptualizing the scope of the paper.

INTRODUCTION

It has always been a challenge for teaching and lecturing on methodology topics in master's and Doctoral courses. Most of the students arrive at this level with limited knowledge of research methodology. This is probably an outcome of lack of awareness of the critical importance that methodology has in research or sometimes it is because of the dull classes [14] or it may be because methodology courses require more effort of the intellect than the subject matter courses due to the complex abstract principles and processes[3]. In many ways, it includes a many of the bewildering principles of statistical analysis in the field of Sociology and Social Work in which by means of a formula which makes no sense whatever to the student, he or she has the capability of determining whether there is a "statistically significant correlation" between the data employed in the research.

At this academic level, it is not advisable for students to rely on seeking their professor´s assistance like they used to in undergraduate courses. After all, they have been accepted on a postgraduate course that has expectations and a natural demand for more autonomy and self-learning. In this sense, the knowledge of knowing how to elaborate a scientific research is taken for granted. Students must follow their own path in formulating the research problem and identifying a gap in knowledge that they would endeavour to fill or develop and propose a new theory. The research problem in scientific research represents the researcher´s study target. A proper proposition of a research problem requires critical analysis of the current state of knowledge in his/her area of interest. Researchers take a particular view or a perspective when they pursue research which may lead to there being some gap that may be left. Now, theory gap is "something" unexplained; not so clear among researchers that needs further attention[14]. Comparing and contrasting approaches is relevant and a nice way of getting started, which gives an opportunity to know the different point of views and gaps that deserve clarification. But the key question is: how to get started?

According to Early (2014), there is a great deal of literature pertaining to teaching strategies that increase students' engagement with research methods courses (it includes active learning, problem based learning, experiential learning, among others), but most of them ignore the learning goals and what is really "behind" a research. Narrow attention has given to research problem and theory gap.

A doctoral research is not simply any research and a publication. For it aims to either delineate, integrate, debate, refute and is therefore complex if we compare natural, social sciences investigation procedures - the discussion about procedures is always partial, incomplete, and strictly according to the author´s point of view[3]. And here is where epistemology takes central focus.

The production of scientific research must be, *a priori*, preceded by the resolution of philosophical and theoretical presuppositions that are indispensable to the researcher. It is not just a question of having a certified methodology or of ethical legitimacy for the development of the research itself. Therefore, we invite the reader to understand a few of the many of epistemological assumptions from the social sciences that, in our view, the foundational sediment on which the natural and exact sciences will be anchored.

Starting from the general idea that knowledge (in its broadest sense) is socially conditioned, the sociology of knowledge has three main questions: the definition of the social conditioning factor; the type of conditioning; the extent of conditioning according to the types of knowledge. The treatment of these questions, and especially of the latter, has quite often resulted in the inclusion of bias and social conditioning, not only of the theoretical contents of science but also of the theoretical and methodological conditions and criteria of validity inherent in the scientific process[2].

Throughout modern history, scientific answers had never been so questioned as they are at present in a world of “fake news” and total disregard of the facts. As it was the case with religion (and therefore with theology), which, after centuries of hegemony in the field of responses to social problems, came into question during the Enlightenment, due its existence in the spiritual rather than the material world, lacking credible explanations explaining the new reality and discoveries in science, religion itself became marginalized in many areas in relation to science as we know it, given to a new obscurantism, has been relativized, deconstructed with exotic arguments, and, finally, rejected by numerious segments globally. This puts us (as researchers) in front of problem to be faced, and which may be related to our own formation. Popper[9], in his critique of Kuhn's epistemology, stated:

In my view, the ‘normal’ scientist is a person one ought to be sorry for. (According to Kuhn’s views many great scientists must have been ‘normal’; The ‘normal’ scientist has been taught badly). I believe, and so do many others, that all teaching on the University level (and below)

should be training and encouragement in critical thinking. He has been taught in a dogmatic spirit: he is a victim of indoctrination. He has learned a technique which can be applied without asking for the reason why (especially in quantum mechanics). Consequently, he has become what may be called an applied scientist, in contradistinction to what I should call a pure scientist. He is, as Kuhn puts it, content to solve 'puzzles[9].

His refutation of the idea of a normal scientist was founded precisely on the basis of what we wish to anchor as the underlying argument for the researcher's role in contemporary times, namely, critical thinking. Contrary to what obscurantism claims as a herald of a new time, of which critical thinking is indoctrination, for Popper it was just the opposite. For him, the normal is an indoctrinated, who applies what he has learned without confronting minimally with his reality, a dogmatic. The researcher, above of all, is the ontological possessor of a social place. This ontological data must be segmented in his 'view' of the world.

The researcher, within his life/reality, will produce responses to society from the real problems he faces. The more displaced from this reality, the more alienated the research will be and the more abstracted (surfing in the transcendence of metaphysics) will be, without the commitment to the true genesis of science, that is, it must fundamentally serve the dilemmas and questions of the human being and not the other way around. Consequently, in social research, the phenomena are always moving on as philosopher Feyerabend states, therefore, it is crucial to understand the evolution of it. Historical research entails the systematic collection and analysis of data with the aim of understanding some entity or entities through time. The insights can clarify extant theory, resolve controversies, and contribute to the development of new theory[4].

This article sets out to emphasize two stages of the research process: formulation of the research problem and theory gap, which can be considered, without exaggeration, two essential stages of the process to structure any research. Therefore, this paper will also help students to select and develop an appropriate literature review.

Why is this Paper not only Important for the Literature Review?

The answer is quite simple. To organize any literature review, students must know, or at least have some idea, of what he/she is "looking for". Which database to use, what has been published and so on.

Most of the literature review help scholars to understand what has been researched from what perspective and using which theoretical standpoint has it been done. Thereby helping the scholar identify and finalise the gap in literature and thereafter decide on his/ her own research questions. This also helps to develop and justify their hypotheses. Due to the infinite number of scientific publications and appearance of new journals, researchers need to identify and systematically select the most relevant literature, which is time consuming apart from being complex[8]. Therefore, some methods have been developed to periodically summarise and draw out the future research opportunities. Some of these techniques are meta-analysis and integrative reviews.

Meta-analysis is a research method that combines the evidence of multiple primary studies by employing statistical methods. According to Johnson[6], although meta-analysis is at the top of the evidence for literature review, pitfalls exist. The most common is inclusion of low evidence studies, which dilutes the overall rigor and reliability. Positive studies, especially small series and case reports are published more often than negative studies.

On the other hand, integrative review method is the only approach that allows the combination of diverse methodologies (for example, experimental and non-experimental research), and has the potential to play a greater role in evidence-based practice areas[18]. The objective is to summarize past empirical or theoretical literature to provide a more comprehensive understanding of a particular phenomenon[1], but the problem with it, is that systematic bias and error can occur at any stage of the review, specially if data from primary sources are misunderstood or interpreted[18].

So, a method that will "skim" the literature in an objective manner is needed; a method that focusses on relevant information and skip the irrelevant ones.

The Skimming Research Method

The objective here is to develop and propose a methodology to select and collect scientific papers published in top journals, in such a way that will provide students with indications of research problems and theoretical gaps.

Initially, the method will help the student to delimit the research topic. The first four steps of the proposed methodology will address these concerns. Subsequently if will offer objective insights to the students about the "theory gap" and "research question/problem" formulation.

Stage 1 – Top Journals Selection: The first stage is to select three or four Journals that are well known and that have international credibility in the scholar community. For that, there are basically two easy ways: student can check with his/her professor which are the most indicated journals for his/her research, depending on his "research line", for example; or search at Scopus website[16] or search other website and use the filters available for a more precise journal research. This would require reading the aims and objectives of the journals to understanding the scope of the research article.

For information: Normally, journals display their Impact Factor and CiteScore on their "About Page". However, one must pay attention because a few journals with questionable reliability post fake Impact Factors too. Therefore, Scopus and SCImagojr website[17] are safe choices.

Important: No one wants to spend 3 or 4 years of the post-graduation course attached to a subject that is not of his/her interest. So, basically, one of the most important decision to be made is to choose a topic of personal interest; a topic that stirs passion and one would be happy to commit to it for a long haul.

Stage 2 – Most Downloaded Papers or/and Most Cited Papers: The selection of these "sections" may seem to lack objective justification, even then it makes sense to look at these. Most downloaded papers indicate the papers that are of most interest to readers, and here we do not only have the academic community, but also managers, consultants and others, while "Most Cited Paper" is an indicator of citation and for that (article writing) mainly scholars and researches do. However, most cited also shed light on what issues and topics have been found to be important over the years and in current times. There must be a reason why so many have cited it that it is one of the most cited papers!

Important: The exploratory reading aims to investigate: What has been the most interesting object for the readers? (Verifiable by most downloaded papers). Also, other prospective insights might be drawn from that subsequently helping in identifying some new dimensions for research. These could be like: Are the phenomena from specific or generalized market? Which methods have been applied?

Stage 3: Data Tabulation: While reading the selected papers from Stage 2, it is advisable for the student to tabulate the main information as given in Table. It is a very simple process, which will help him/her to then find out some patterns, such as: What are the most theories applied/published in these Journals? How is the theory gap indicated on each selected paper? What are the interesting insights given in each of them?

Basically, an Excel table is set up to record:

• Article Title and Authors;

• Year of Publication;

• Main Topics (and not only the Keywords);

• Research Question (It states the rationale or hypothesis);

• Theoretical Gap (Some papers will not really have one. It will depend the objective of the study);

• Purpose of the Study;

• Theory Chosen (Some studies will not apply any theory to explain a phenomenon. That happens quite often with Commentaires or Reviews);

• Methodology (It is interesting to learn different type of methodologies. Depending on the area of research, some studies will make use of more qualitative data and discussion, while others quantitative);

• Conclusions (it is not a copy and paste activity. Instead, you are supposed to choose paper´s points adequadamente. If possible, try to link research objectives with the main results. It is interesting how some studies are not able to proper do so. Some objectives will even remain unanswered!);

• Interesting Insights (Here students are supposed to add information from the text that they haven't read or learned before. Maybe they can add new concepts, new frameworks developed or any significant message that could be written in their work. The data generated help to develop the understanding of patterns of scientific productivity and their implications);

• Future Research (It offers new research directions, which may be even considered. The problem with it, is that generally the respective author or his students are already working on it);

All this reading and notes will provide structured and organised information about the various research topics as well as guide the researcher towards a theoretical gap. At this Stage students will have more doubts than answers. They possible will state their own research questions, which will be confirmed as good (or appropriated) or not in the next steps.

Stage 4 – Quantify the Most Read or/and Most Downloaded Topics: This makes visible what has been interesting in the eyes of readers and researchers, as it will indicate the quantity of work in related fields of interest.t also helps to justify the reasons for choosing a specific theme to write. At this time of the skimming process, the research design is already beginning to take shape.

Stage 5 – Database Selection for Historical Analysis: Historical analysis is particularly important to understand how, and in which circumstances, the concept or phenomena was studied.

The search threads can be constructed through Scopus or Web of Science databases. In Scopus, more than half of the applicable publications came from Europe, Latin America, and Asia[13], while the Web of Science database yield several American and Western Europe publications[7].

The criteria for the search can be determined and applied such as in the example below:

a. In the Scopus website, select "*Document Search*" as the search feature;

b. In the "*Search for*" field, the topic (s) can be added as keywords;

c. Later, in "*Search Tips*", choose "*Article Title, Abstract, Keywords*". This is a more specific filter.

d. Date range must be chosen according to research purpose, but in the case of historical research, placing a wide range can give a good idea about the evolution of the concepts.

e. In the "*Document Type*" field, choose "*Article or Review*";

f. Repeat the exercise with one concept, with two or more than two concepts linking them. This would yield many papers that may appear depending on the area and topic, so another possibility is to extend it to another filter on "within" results. It is necessary to specify other Keywords to skim the results with even more specific works. A practical example of this methodology was published and can be read at Klafke et. al (2016).

Stage 6 - Abstract + Introduction + Conclusions Reading: At this moment, academics should judge what papers could possibly offer good insight and knowledge. This "fast reading screening" helps in saving time and gives a broad picture of Primarily the Abstract alone should eb able to help in selecting or not selecting papers for further use. The next step to gain clarity would be to read the introduction and conclusions sections.

Stage 7 – Text Mining Methodology: Seminal Studies and Reviews.

It has become increasingly difficult to conduct thorough reviews due to the vast quantities of sources and information. So, at this moment, after all the accomplished previous steps, the most important thing is to define which documents and words will be taken in consideration for extracting content with the assistance of an algorithm. In this case, it is suggested that academics choose seminal studies and reviews. These two are of central importance, often because they report a major breakthrough, insight and a generative synthesis of ideas through Reviews compare, contrast, and correlate various books, research articles, and other relevant sources. Now classic papers are in a sense of order and continuity related the academic topics because the episteme survive in the classic studies. Also, they are cited many times in the literature, which reiterates and reemphasises their relevance.

Text mining aims to analyse language text with the usage of a software (such as "R", "Aika", "GATE") in a search for rules, patterns, trends, clusters (Cluster analysis is the process of grouping data points together based on their relative similarities) or even evolution of terms,

in order to extract useful, relevant data from the submitted texts. It is a Bayesian contextual analysis algorithm known as CTM[5]. Although this methodology has been used in engineering and medicine, little has been used in management studies. Text Mining has two main and subsequent phases that are: extraction of information and information mining.

The data analysis is largely dependent on the pre-processing and the data representation model that was chosen in pre-processing.

So, the first thing to do is to take a look at the "packages" and choose the most appropriated for the dissertation:

tm: Text Mining, framework for text mining applications with R. (We will use this package for the example below).

In order to practically demonstrate the proposed methodology, we selected 29 papers about value and value creation in Marketing and Business Administration. The papers were collected from the Scopus database and represent the most cited works or seminal studies in the field.

Initially, all the pdf files are converted to text documents by Windows utility and a corpus was created containing all the text files. Later 'R' software was utilized to carry out the text mining, especially the R packages *tm, word cloud and SnowballC*. The text files are cleaned by tm_map () function to perform various operations on the text files as given below:

I. Converting foreign and special characters like: /@\â€u etc. to spaces

II. Converting all characters to lower case

III. Removing Numbers

IV. Removing Punctuation characters

V. Removing Stop Words of English language like "the", "we" etc.

VI. Stripping whitespaces

VII. Remove Number Punctuations

VIII. Removing general words like "can" etc.

Later a Document Term Matrix is generated from the text data so obtained after cleaning. The Document Term Matrix (DTM) is essentially a matrix in which rows correspond to the Documents and Columns correspond to the Terms/Words and the cells contain the frequencies/counts. By taking the column sums of the DTM and by arranging the column sums in the decreasing order, it is possible to find the Top Terms, whose frequency is more than a threshold value, say n. In this study, the value of n is taken as 25 and found out the top 25 words having the highest frequency. These results are displayed in Table 1.

Table 1. Top 25 words having the highest frequency.

Top 25 Frequently used Terms

word	freq
value	3423
marketing	1278
customer	1071
customers	895
journal	873
research	731
new	679
service	673
relationship	662
product	648
management	630
business	626
creation	581
services	550
values	532
consumer	494
social	489
vol	489
consumers	426
study	426
use	424
products	421
cocreation	391
process	380
relationships	375

Source: R output (2020).

Using frequency, for example, students can find words that are characteristic for one document within a collection of documents. Exploring term frequency on its own can give insights into how language is used in a collection of natural language. Other terms/codes like count () and rank () help to reason about term frequency.

Seminal studies on the "value of things" began with Woodruff[19], Tzokas & Saren[12] and Doyle[2], who explored topics such as relationships and value creation. The first frequencies suggest that value is related to customer and service, which indicates relationship. About this, value will depend on the interaction between people - consumer vs. product/service[10].

To display the Top Terms, whose frequency is more than a threshold value, data visualization is done through Barplots. In this study, a bar plot of the top terms whose frequency is more than 100 is generated. Also, using the SnowballC package, the word cloud is generated for the Corpus, which displays the importance of the words.

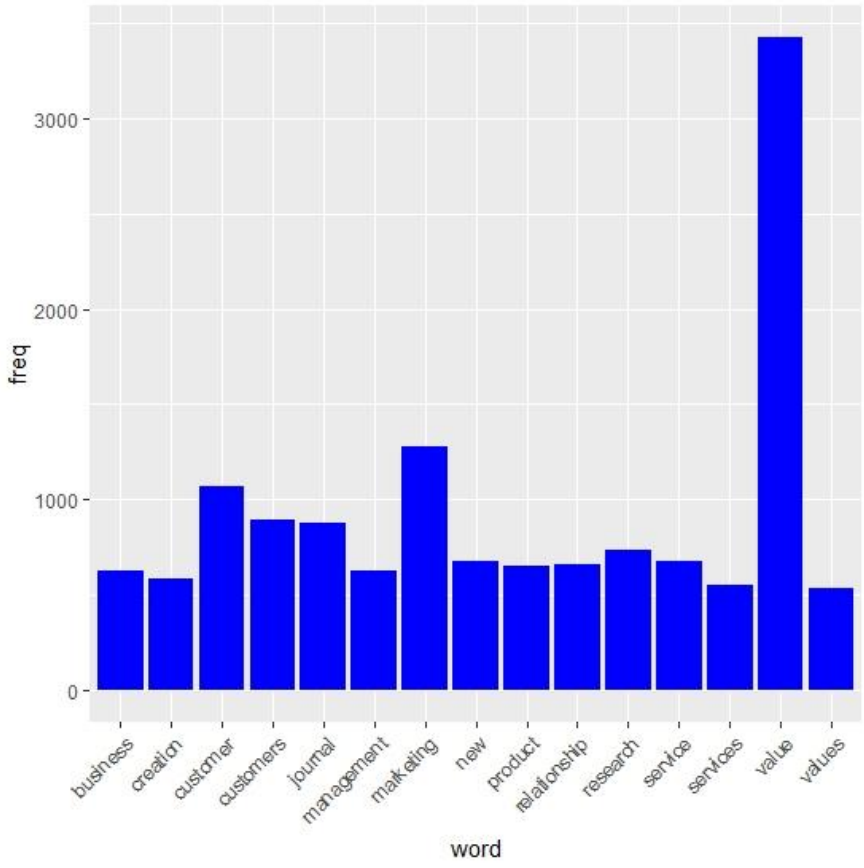

Figure 1. Bar Plot of top terms whose frequency is more than 100.

Source: R outcome (2020).

In marketing, value is related to product or service, although many do not even know exactly what value is or exactly how to measure it[12].

Figure 2 displays the word cloud depicting the importance of the words, whose frequency is more than 200.

Figure 2. Word Cloud of Top Terms whose Frequency is more than 200

Source: R outcome (2020).

Word clouds provide a better, although extensive, words and expressions that surrounds value in the seminal and most cited studies chosen for this brief analysis. It appears specific marketing and business basic dimensions, such as quality, information and culture. Variables that were not seen in the previous plot.

CONCLUSION

Our goal with this Skimming Research Method was to offer certain rationality for students who often do not know how to begin their research and complete their dissertation. Some methodologies, like systematic review, can take a long time to be accomplished, while the Skimming Method is both easy and fast. The text mining application in the selected papers from Stage 5, will guide to the theory gap offering greater efficacy and assertiveness when selecting the literature for literature review. Consequently, after the reading selection, the literature review will be basically accomplished and ready to be consulted for further action for completing the research.

The paper's contribution to business education derives from its focus on structured, postgraduate research processes. The insights this methodology can clarify extant theory and

contribute to the development of new ideas and research problems. Also, it details, *Stage pour Stage* how to find out the most published papers, the different types and most methodology used, among others information.

Like many methodological papers, this one may present limitations. The first one is that some professors think it is best to focus on the literature review than working in parallel with the Gap Theory. Another one is that we did not develop the extent analysis that can be done with R software; hence, other research can further develop practical methodology using Text Mining and its functions.

Finally, we hope this article provides a perspective and suggestions to better manage the difficult and challenging initial steps of writing a dissertation. Sometimes the lack of resilience and the lack of supervisor´s support can be challenging! But remember: Never overcome the fear of seeing red ink all over the pages, use your supervisor's comments as an opportunity to improve!

REFERENCES

1. Broome M.E. (1993) Integrative literature reviews for the development of concepts. In Concept Development in Nursing, 2nd ed. (Rodgers B.L. & Knafl K.A., eds), W.B. Saunders Co., Philadelphia, PA, pp. 231–250.
2. Doyle, P. (2000). Value-based marketing: Marketing strategies for corporate growth and shareholder value. Chichester7 John Wiley & Sons.
3. Howard, C., & Brady, M. (2015). Teaching social research methods after the critical turn: Challenges and benefits of a constructivist pedagogy. International Journal of Social Research Methodology, 18(5), 511-525.
4. Hunt, S. D. (2011). On the intersection of marketing history and marketing theory. Marketing Theory, 11(4), 483-489.
5. Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence, 1(1), 60-76.
6. Johnson, T. M. (2008). Tips on how to write a paper. Journal of the American Academy of Dermatology, 59(6), 1064-1069.
7. Meho, LI and Yang, K. (2007), "Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar", Journal of the American Society for Information Science and Technology, Vol. 58 No. 13, pp. 2105–2125.
8. Pagani, R. N., Kovaleski, J. L., & Resende, L. M. (2015). Methodi Ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation and year of publication. Scientometrics, 105(3), 2109-2135.
9. Popper, K. Normal Science and its dangers. In: LAKATOS, Imre. MUSGRAVE, Alan (ed.). Criticism and the growth of knowledge. Cambridge: Cambridge University Press, 2004.
10. Prahalad, C. K., and Ramaswamy, V, (2004), Co-creation experiences: The next practice in value creation, Journal of interactive marketing, 18(3), pp 5-14.
11. Santos, Boaventura de Sousa. Introdução a uma ciência pós-moderna. Rio de Janeiro: Graal, 4a ed., 2003.
12. Tzokas, N., & Saren, M. (1999). Value transformation in relationship marketing. Australasian Marketing Journal, 7(1), 52 – 62.
13. Vieira, ES and Gomes, JANF. (2009), "A comparison of Scopus and Web of Science for a typical university". Scientometrics, Vol. 81 No. 2, pp. 587– 600.

14. Wagner, C., Garner, M., & Kawulich, B. (2011). The state of the art of teaching research methods in the social sciences: Towards a pedagogical culture. Studies in Higher Education, 36, 75 –88.
15. Wei, X., Croft, W. (2006). LDA-based document models for ad-hoc retrieval. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 178–185. ACM, 2006.
16. Scopus website : https://www.scopus.com/sources?zone=&origin=sbrowse
17. SCImagojr website : https://www.scimagojr.com/journalrank.php
18. Whittemore, R., & Knafl, K. (2005). The integrative review: updated methodology. Journal of advanced nursing, 52(5), 546-553.
19. Woodruff, R. B. (1997). Customer value: The next source of competitive advantage. Journal of the Academy of Marketing Science, 25(2), 139 – 153.

Image *Author -1*	***Renata Klafke – Corresponding Author*** *UFPR – Universidade Federal do Paraná* *Campus Jardim Botânico* *Avenida Pref. Lothário Meissner, 632 – Jardim Botânico, Curitiba- PR*
Image *Author -2*	***Josué S. Flores*** *UFPR – Universidade Federal do Paraná* *Campus Jardim Botânico* *Avenida Pref. Lothário Meissner, 632 – Jardim Botânico, Curitiba- PR*
Image *Author -3*	***Krishna Devulapalli*** *Indian Institute of Chemical Technology (IICT) – India* *Uppal Rd, IICT Colony, Tarnaka, Hyderabad, Telangana 500007, India*
Image *Author -4*	***Abhishek Nirja*** *Anil Surendra Modi School of Commerce (MMS)* *V. L, Pherozeshah Mehta Rd, Vile Parle, Mumbai, Maharashtra 400056, India*