2023

EJBRM Volume 21, Issue 1

Editor

Ann Brown

Published by Academic Publishing International Limited
Curtis Farm, Kidmore End, Nr Reading,
RG4 9AY, United Kingdom
karen.harris@academic-publishing.org

eISSN: 1477-7029

EJBRM Volume 21, Issue 1

Contents

Towards Delphi Rigor: An Investigation in the Context of Maturity Model Development

Samuel Reeb
Business Information Systems - Information Management, TU Dresden, Dresden, Germany

samuel.reeb@tu-dresden.de

Abstract: The maturity model (MM) and Delphi research areas are extensive and diverse, leading to numerous approaches. This study addresses the Delphi method regarding its rigor requirements within the IS literature. To this end, the example of maturity model development is investigated. Hence, Delphi studies for MM development are identified and analyzed regarding their rigorous application and design. The examination focuses on the connections between maturity model aspects and the Delphi methodology. Hence, relevant aspects of maturity model and Delphi literature are elaborated, and criteria for methodological rigor are derived. A key challenge is linking the method to the specific design objective (in this study example, the development of a maturity model). After conducting a literature search to identify studies that use the Delphi method for MM development, these criteria are used as a basis for deductive content analysis. The results indicate a lack of clarity regarding the methodology, as different aspects are reported, although the general demands, starting points, and goals were similar. A need for design guidelines for planning and conducting Delphi studies is emphasized and addressed in this paper. Hence, guidelines for designing Delphi studies are developed and presented, considering relevant aspects for ensuring rigorous implementation. The focus lies on the linkage between study design and intended model elements, as this demonstrates the complexity of the study design and the relevance of the design decisions through an example. The guidelines integrate the different methodological aspects of maturity model development and the Delphi methodology, providing an orientation framework for the design process of such research projects. Therefore, this study contributes to existing research by proposing design guidelines for Delphi studies to foster rigor in the specific context of maturity model development. Although the presented guidelines focus on the maturity model context, the general design approach and decisions are transferable and applicable to other domains. Hence, this research contributes to the Delphi literature by providing insights into how relevant elements should be addressed in the designing process of a Delphi Study. Scholars should investigate how the presented guidelines must be adapted for other domains in future research.

Keywords: Maturity model, Rigor, Delphi, Guideline, Model development, Research design

1. Introduction

The Delphi method is a well-established procedure in information systems (IS) research (Gallego and Bueno, 2014; Paré et al., 2013), and multiple variations of the approach exist (Paré et al., 2013; Skinner et al., 2015). The method's strengths lie in identifying current topics, consolidating different experts' opinions (de Bruin et al., 2005), and adaptability to specific research conditions (de Bruin et al., 2005; Gallego and Bueno, 2014). A methodologically rigorous application is essential to realize these benefits adequately. As the method allows adapting to study specifics (Gallego and Bueno, 2014), its design process is rather complex, and it risks losing methodological rigor (McKenna, 1994; Paré et al., 2013; Rauch, 1979; Skinner et al., 2015; Strasser, 2017). Consequently, the method is suitable for investigating complex issues, as comprehension questions and uncertainties are eliminated through its iterative process, and an adaptation to the concrete objective is possible (Skinner et al., 2015). However, it is questionable whether existing discussions about the methodological approach are sufficient and specific enough to ensure a clear understanding and accurate implementation.

The complex maturity model (MM) domain, comprising different development steps, is a possible application area for the Delphi method (Becker et al., 2009; de Bruin et al., 2005). In a narrower sense, model development is carried out in the "design model" step, for which various methods (e.g., case study interview, focus groups) can be applied (Mettler, 2011). Several scholars suggest that using the Delphi method is appropriate and beneficial for the "design model" step, as model deficiencies are already addressed at an early stage (Becker et al., 2009; de Bruin et al., 2005; Lasrado et al., 2015; Pereira and Serrano, 2020). Hence, the Delphi method achieves a more practical result in a shorter time during model development and tackles problems such as "limited exposure to relevant context" (Lasrado et al., 2017) or "competition for attention" (Lasrado et al., 2017) and should arguably be applied more often.

However, despite the Delphi method's advantages, MMs are rarely developed using this approach (Pereira and Serrano, 2020). The development of MMs using the Delphi method is barely addressed in current research, even though the documentation of this research process is essential as it is complex and diversely applied. A rigorous approach is vital because the Delphi method's diversity for MM development allows for various arbitrary options

Reference this paper: Reeb, S., 2023. Towards Delphi Rigor: An Investigation in the Context of Maturity Model Development *The Electronic Journal of Business Research Methods*, 21(1), pp. 01-13, available online at www.ejbrm.com

in the procedure and the associated risk of quality loss. Accordingly, the research design for the MM and the Delphi study must be aligned considering the different methodological aspects to ensure a rigorous approach. To ensure the quality and reliability of such a study and to counteract arbitrariness in the conduction, it is necessary to know how a Delphi-based MM development study can be designed. Consequently, the question arises as to whether researchers using a Delphi study to develop a MM follow a methodologically rigorous approach, which is not addressed in the literature. Therefore, MMs represent a suitable application area to study the Delphi method from a rigorous point of view, as the complexity of both approaches requires a deep understanding of the connections and a rigorous application of the Delphi method.

Therefore, this paper investigates the methodological rigor of Delphi studies in the IS domain using the MM development process as an example. Typical design demands for both methods are elaborated and used to analyze identified studies of maturity model (MM) development with the Delphi method. Accordingly, the following research questions emerge.

- To what extent do existing studies on maturity model development utilizing the Delphi method follow the methodological rigor?
- How can maturity model development studies using the Delphi method be designed rigorously?

2. Theoretical Background on Criteria for Methodological Rigor

The work on MMs in IS research is extensive (Carvalho et al., 2019; Mettler, 2011) and applied in various domains (Lasrado et al., 2015; Mettler, 2011). MMs depict how capabilities for a domain develop through discrete, successive maturity levels (Becker et al., 2009; Stelzl et al., 2020). The most prominent representatives of a methodological framework for MM development in the IS literature are de Bruin et al. (2005), Becker et al. (2009), and Mettler (2010). Comparing the approaches shows that the research designs follow similar steps (Mettler, 2011). Preeminent authors of the literature body repeatedly describe the Delphi method as a suitable and beneficial method for designing a MM (Becker et al., 2009; de Bruin et al., 2005; Lahrmann et al., 2011; Pereira and Serrano, 2020).

The Delphi method represents a structured, iterative procedure of an anonymous expert survey, in which expert knowledge is collected with the help of questionnaires, condensed, and controlled feedback of a statistically processed group response is given to the experts (Gallego and Bueno, 2014; Paré et al., 2013; Rowe and Wright, 1999). Delphi studies must have the following four generic characteristics to count as a Delphi study (Gallego and Bueno, 2014; Rowe and Wright, 1999):

- *Iterative process:* the study represents a round-based process, and in each round, the experts can communicate and adjust their opinions through a questionnaire and have the group position reflected to them (Paré et al., 2013; Rowe and Wright, 1999).
- *Anonymity*: the experts give their opinion independently and individually through questionnaires, without social pressure, and have the opportunity to change their opinion through the iterative process without losing face (Rowe and Wright, 1999; Skinner et al., 2015). Furthermore, it ensures that experts cannot trace the origin of provided information back to a single expert (Rowe and Wright, 1999).
- *Controlled feedback*: feedback on the group's responses is provided to the participating experts between rounds of questioning for comments and/or as a starting point for the follow-up round (Rowe and Wright, 1999; Skinner et al., 2015). The research team eliminates irrelevant information for the study in advance (Strasser, 2017).
- *Statistically processed group response*: The answers can be processed quantitatively and statistically to provide an anonymized group response to the experts (Rowe and Wright, 1999; Strasser, 2017).

The Delphi method is beneficial for designing a MM, as weaknesses can be identified and addressed in the initial development phase (Pereira and Serrano, 2020). An adequate development through a literature analysis is unlikely beyond the structural depth of the component layer (de Bruin et al., 2005). Thereby, a Delphi study is beneficial in the following points, which are relevant in the development of a MM within a complex domain (de Bruin et al., 2005):

- exploring and presenting complex topics (Okoli and Pawlowski, 2004)
- bringing opinions together for an improved result (de Bruin et al., 2005)
- making a scientific contribution to knowledge in a domain (Delbecq et al., 1975)
- lack of empirical evidence (Murphy et al., 1998).

Furthermore, the general objective of concept development of Delphi studies and the intention of the MM development coincide (de Bruin et al., 2005). Another advantage of the Delphi method is its adaptability to the specifics of a study (McKenna, 1994; Strasser, 2017).

Gallego and Bueno state that the majority of scholars use a modified Delphi method, as the versatility and flexibility to change each variable allow one to meet the specific needs of the research (Gallego and Bueno, 2014). Adjustments in the methodology can be helpful but reduce the quality and credibility of the method and thus the research results (Gupta and Clarke, 1996). Accordingly, methodological modifications risk losing methodological rigor, regardless of whether this is done intentionally (McKenna, 1994; Rauch, 1979; Skinner et al., 2015; Strasser, 2017). Several research contributions address the issue of rigor in Delphi studies and emphasize the need to consider general design elements and specifics of Delphi variants (Gallego and Bueno, 2014; Paré et al., 2013; Skinner et al., 2015; Strasser, 2017).

The remainder of this paper examines whether existing studies of MM development applying the Delphi method meet methodological rigor. Hence, this section elaborates on methodological requirements for Delphi studies to develop MMs, focusing on the interaction of both approaches. Table 1 shows an overview of the design elements of a Delphi study that are connected to the design elements of a MM. These requirements are later used to analyze existing studies regarding their methodological rigor.

Table 1: Overview of Design Elements and Their Relations to Delphi and MM

Delphi design element		**Relating MM design elements**
Delphi variant		Central constructs
Expert Panel	Expert Criteria	Model scope (e.g., function, auditorium), especially domain Central constructs
	Panel composition	Model scope (e.g., function, model complexity)
Number of rounds and Statistical Processing		Central constructs Development Basis
First-round design		Development Basis
Modifications		Research objective Model characteristics (model scope)

The Delphi method is used for forecasting, decision making, and concept development (Okoli and Pawlowski, 2004; Paré et al., 2013). Since the introduction of the classical Delphi method by Dalkey and Helmer (1963), multiple variations of the method have been introduced and adapted to different research problems (Okoli and Pawlowski, 2004; Paré et al., 2013; Strasser, 2017). The key characteristic of a modified Delphi study is difficult to define (Hasson and Keeney, 2011), causing no clear and uniform distinction in the literature on what a (genuine) Delphi variant is. Strasser (2017) identified seven genuine Delphi variants: Classic Delphi, Policy Delphi, Decision Delphi, estimate-feedback-talk-estimate (EFTE) Delphi, Ranking-Type Delphi, Argument Delphi, and Disaggregative Policy Delphi, which differ in their focus and objective. Furthermore, Delphi variants representing implementation variations of genuine variants, like Online Delphi, Real-Time Delphi, and Mini-Delphi, exist (Hasson and Keeney, 2011; Strasser, 2017). A central decision scholar must make is the selection of an appropriate Delphi variant according to the research objective, as Delphi variants can vary in their characteristics (e.g., terminology, goals, procedure),.

As this decision depends on the research objective (development of a MM), scholars need to specify this by elaborating on the intended MM´s central construct. MMs are structured into the central constructs of maturity levels, process area (components, sub-components), goals, assessment elements, development path, and improvement actions. A clear understanding of this aspect and decision is crucial for the design process. A short explanation of the different constructs and alternative terms is provided in the following, as no consistent terminology for the different structural elements exists in current literature. Hence, scholars should specify which central construct(s) is/are elaborated on through the Delphi study and define the terminology used regarding the existing literature on MMs.

Maturity levels (also maturity stages) represent the highest level of abstraction of a MM (Lasrado et al., 2015) and describe an archetypal representation of characteristics and conditions of a developmental level (de Bruin

et al., 2005; Object Management Group, 2008; Stelzl et al., 2020). During development, the levels should be given a short name and a short description of the essential elements of the stage (de Bruin et al., 2005).

Each stage is described by a collection of process areas (also capability area, dimension, focus area) (Team CMMI Product, 2006), which can be divided into Components (also process area threads) and Sub-Components (also process area, factor) for more complex domains (de Bruin et al., 2005). A process area represents different facets of a content cohesive area (Hasson and Keeney, 2011). A process area represents a collection of content-related goals (also factor specification (Object Management Group, 2008)). A goal describes the intended condition within a process area that significantly improves this area (Team CMMI Product, 2006). The goal layer is the link between the abstract layers of the MM and the operational assessment (Object Management Group, 2008; Team CMMI Product, 2006). The goals represent a measurable basis and can include practices that exemplify what should be done to achieve the process area and reach the targeted level of maturity (Object Management Group, 2008). Consequently, the assessment elements form the lowest level of a MM (Stelzl et al., 2020).

These hierarchical elements are related to each other through the development path (also maturation path), mapping the goals and practices to the maturity levels. Accordingly, this defines what a mature development state, respectively the status perfect, characterizes and how the path is designed (Mettler and Rohner, 2009). Improvement actions are needed to move along the maturation path, indicating how process areas, goals, and maturity can be achieved (Reeb and Pinnecke, 2021).

After selecting a Delphi variant, a vital aspect of the Delphi study is selecting a suitable expert panel. First, it is crucial to define what constitutes an expert status (Hasson and Keeney, 2011; Skinner et al., 2015) and disclose the criteria for replicability (Skinner et al., 2015). The expert criteria depend on the research objective (Hasson and Keeney, 2011). In the context of MM development, these are related to the MM´s scope, and scholars should relate the expert selection to relevant criteria like the domain (e.g., software development, business process management), audience/ user (e.g., higher management, executives), or function (e.g., descriptive, prescriptive) of the intended MM. Although these criteria are examples, this author argues that the domain must be defined and related to the expert selection criteria, as should the intended central constructs. Furthermore, aspects like the function or the complexity of the intended MM can influence the panel's composition. Thus, scholars need to define whether the panel consists of homogeneous or heterogeneous experts (Gallego and Bueno, 2014). Overall, it is relevant that the selection and composition of the expert panel are reported and related to the research´s objective and relevant characteristics of the MM.

Another element is the number of executed iterations/rounds. The fundamental goal is to reach a consensus, which theoretically can lead to an unlimited number of rounds (Hsu and Sandford, 2007; Torrecilla-Salinas et al., 2019). An absolute consensus may not be achievable, for example, due to opinion divergence among experts (Paré et al., 2013). Furthermore, too many rounds can lead to experts dropping out or a not substantively change of opinion towards the group opinion (Dransfeld et al., 2000; Gallego and Bueno, 2014). Accordingly, the researcher needs to know when to stop the data collection (Schmidt, 1997). Scholars need to define rules for reaching an adequate level of consensus and stopping the iterations (Okoli and Pawlowski, 2004; Paré et al., 2013). These rules must be related to the intended central constructs, what specific characteristics are investigated, and how their quality is measured. Measuring the quality is further related to the central Delphi design element of a statistically processed group response (Rowe and Wright, 1999; Strasser, 2017). The answers of each round must be statistically analyzed, and the results reported to the panel to indicate the panel's opinion. Thus, suitable measurements must be defined for the central constructs elaborated through the Delphi study. The development of a MM is connected to related MMs (new development, further development, model combination, structure transfer, content transfer) (Becker et al., 2009). Hence, it is beneficial to investigate relating MMs, as insights for quality criteria can be found in relating models.

As an additional element, the design of the first round represents a distinctive feature within the Delphi study. Following the Classic Delphi design, the first round starts with a qualitative survey using open-ended questions (Hsu and Sandford, 2007; Skinner et al., 2015). The use of structured questionnaires with pre-selected items is an accepted modification of the Delphi method (Hsu and Sandford, 2007; Skinner et al., 2015). In this case, the questionnaire should be based on an extensive literature review and a rigorous grounding of the included items (Hsu and Sandford, 2007) and the relating MMs in the MM context. Hence, scholars should indicate whether they ground their Delphi study on existing work and how it influenced the first round.

An advantage of the Delphi method is its modifiability, which at the same time presents a crucial challenge to its rigor. Hence, it is relevant that researchers report and justify modifications. Thereby, modifications must be

related to the research objective and the corresponding characteristics of the MM, which should be reported in the MM´s scope.

3. Research Design

Following vom Brocke et al. (2009), a systematic literature analysis was conducted to identify relevant literature. First, the review scope was defined following Paré et al. (2015). The literature search and selection process was designed according to Webster and Watson (2002). Finally, the identified literature was analyzed using deductive content analysis, as suggested by Elo and Kyngäs (2008), based on the above-defined criteria for a rigorous Delphi and MM development methodology.

Defining the scope of a review "is a necessary first step of clarification in any literature review, which bears implications for the later search process" (vom Brocke et al., 2009). Following the categorization of Paré et al. (2015), a critical review is conducted. Hence, representative studies for applying the Delphi method in the context of MM development are analyzed and critically held up against the criteria for a rigorous study.

According to Webster and Watson (2002), the literature search was conducted as a commonly applied approach in the IS domain. There was no restriction regarding the period of the investigation. To identify relevant and representative literature, four central academic databases (Business Source Complete, Emerald Insight, Science Direct, SpringerLink) and two high-quality conference bases (AIS Electronic Library, IEEE Xplore), typical for the IS domain, were accessed. The database SpringerLink was limited to the disciplines "Computer Science" and "Business Management" to include only papers relevant to the IS domain. The query string "maturity model*" AND "Delphi" was used, searched in title, abstract, and attributed keywords. The title, abstract, and complete text of the results were screened regarding their relevance to the defined scope. Furthermore, according to Webster and Watson (2002), the identified literature was scanned by a reference backward and forward search with one iteration to identify further relevant articles. During this step, no other literature was identified. The results of the literature search are summarized in Table 2. Overall, 61 papers were scanned regarding relevance, with 14 suitable for this research's objective.

Table 2: Overview of Search and Selection Process

Database/search step	Number of results	Number of relevant results
Business Source Complete	8	5
Emerald Insight	0	0
Science Direct	12	2
SpringerLink	35	2
AIS Electronic Library	6	3
IEEE Xplore	10	4
Duplicates	n/a	2
Forward & Backward search	n/a	0
Total	61	14

A deductive content analysis following Elo and Kyngäs (2008) was conducted to analyze the literature. A structured coding scheme was developed representing the presented characteristics of the previous section. MAXQDA was used to code the literature following the coding scheme. Finally, the codes were analyzed and compared to the rigorous demands, focusing on the relation of Delphi and MM aspects.

4. Findings

In this section, the results of the analysis are discussed. Firstly, each paper's overall research objective and research design were analyzed. Every study stated the MM´s domain, and it is indicated that preliminary content was used as an input for the Delphi study either directly or in the context of the paper. Furthermore, three of the 14 papers evaluate the Delphi results in the form of a subsequent practical demonstration or evaluation.

In the context of preliminary work, 13 papers used literature data, and one paper used project data, and a distinction can be made between the two goals of Delphi studies, "refine" and "extend". Seven Delphi studies aimed to "refine" a central construct, developed or derived in the preliminary work, to improve it. Six Delphi studies aimed to "extend" the MMs by elaborating on a deeper level of central constructs (e.g., elaborate on

what sub-components are relevant for identified components). The study by Labaka et al. (2019) had both goals by first "refining" a central construct and then "extending" it.

The studies were further analyzed for the four defining characteristics of iterative, anonym, controlled feedback, and statistical aggregation of a Delphi study (Strasser, 2017), regardless of the context of MM development. Only five of the 14 studies reported the four requirements sufficiently. An additional three addressed each of the four requirements at least to a limited extent by presenting the criteria as requirements for the rigor of a Delphi study from a theoretical perspective but not eliciting how they were accomplished. Scholars in Delphi research have called for a more rigorous and variant differentiated approach in the last ten years (Gallego and Bueno, 2014; Paré et al., 2013; Skinner et al., 2015; Strasser, 2017). As eleven of the 14 examined papers were published since 2016, they should have acknowledged this requirement for rigor in scientific research and incorporated it into their research. Nevertheless, only two of these studies met the mandatory requirements for a rigorous Delphi study, and another three reported at least all requirements from a theoretical perspective. Based on these findings, researchers need further guidance on applying the scientific methodological requirements and the design process of a rigorous Delphi study. The requirements of anonymity and feedback are most often not considered, although these two are decisive characteristics of a Delphi study, which raises the question of whether a Delphi method or another approach was used. For example, Schriek et al. (2016) can be stated, as they apply a modification by using face-to-face meetings and not distinguishing their methodological design, e.g., focus group interviews.

As stated in section 2, the author argues that the criteria of the intended central construct(s) and the development basis of the MM must be defined for an adequate Delphi design. These two criteria were focused on regarding the MM development process during the coding process. Although nearly all scholars mention the intended central construct(s), the results confirm that no consistent wording exists throughout the research area, as mentioned in previous studies (Lasrado et al., 2015; Stelzl et al., 2020). Hence, information about the study's central construct(s) must be provided. It is crucial to provide this information as it gives researchers and practitioners an overview of the intended MM structure, how the study contributes to this model design and its useability for their context. Ten of the twelve papers that mention the intended central construct(s) provide enough information to do so. Nevertheless, four of the ten studies leave room for improvement, as it is quite complex to collect the relevant information, and they allow interpretation to some extent (e.g., Labaka et al. (2019)). All papers state the development basis of their MM by presenting preliminary work for the Delphi study. A distinction can be made between new development (use of self-developed MM), further development (use one existing MM), model combination (use a fusion of different MM), structural transfer (use structure of a MM), and content transfer (use content related central construct(s) of a MM) for the use of the preliminary work as a basis for the Delphi study. Although the intended central constructs and developmental basis are mostly reported, only Reyes and Giachetti (2010) (developmental basis) and Bruin and Rosemann (2007), as well as Kerpedzhiev et al. (2021) (central construct), relate them to design decisions of the Delphi study. Furthermore, additional aspects of the MM scope, focusing on a relation to the Delphi study design, were coded and analyzed. However, no paper relates further scope elements to the Delphi study design, and they are thus not further discussed.

Lastly, design aspects of a Delphi study, which are interrelated with the MM are analyzed and an overview is presented in Table 3. Although the focus is on how the MM aspects relate to the design decisions for the study, each aspect is first analyzed independently. Ten studies specify the first-round design, with two using a quantitative design, five using a qualitative design, and three a combination of both designs, whereas none describes the relation to the MM characteristics. Twelve of the fourteen studies report the number of iterations, with five studies stating a consensus criterion. Only de Bruin and Rosemann (2007) relate the iterations and consensus to MM characteristics, particularly the central construct. Furthermore, eleven scholars report the expert criteria and panel composition, with four relating it to Delphi´s research goal(s) and domain. Hence, the relevant design aspects are usually described to some extent but are only occasionally related to the MM characteristics. None of the papers directly reported a Delphi variant. The best examples are:

- George et al. (2020) present that a modified Delphi is used, and from the context, the application of the classical Delphi can be concluded.
- Schriek et al. (2016) indirectly report using a ranking-type Delphi.
- Serral et al. (2020) designed their study similarly to referenced other Delphi studies and, drawn from the context, using a classic Delphi design.

Table 3: Literature Analysis Results on Delphi and MM Aspects Relation

Study	Delphi Criteria	MM Aspects that are related to…				
		Delphi Variant	Expert Panel	Consensus Criterion	First-round Design	Modifications
de Bruin and Rosemann (2007)	sufficiently addressed	not reported	domain & research goal	central construct	reported but not related	central construct
Reyes and Giachetti (2010)	not sufficiently addressed	not reported	reported but not related	reported but not related	reported but not related	/
Stojanov et al. (2015)	sufficiently addressed	not reported	domain	reported but not related	reported but not related	/
Schriek et al. (2016)	not sufficiently addressed	not reported	reported but not related	reported but not related	reported but not related	central construct
Mens and Ravesteyn (2016)	addressed to a limited extent	not reported	reported but not related	reported but not related	reported but not related	/
Radosavljevic et al. (2016)	not sufficiently addressed	not reported	reported but not related	not reported	not reported	/
Karabacak et al. (2016)	sufficiently addressed	not reported	reported but not related	reported but not related	not reported	/
Smits and van Hillegersberg (2017)	not sufficiently addressed	not reported	domain & research goal	reported but not related	reported but not related	/
Nurcahyo et al. (2018)	addressed to a limited extent	not reported	domain	not reported	not reported	/
Labaka et al. (2019)	addressed to a limited extent	not reported	reported but not related	reported but not related	reported but not related	/
Özturan et al. (2019)	not sufficiently addressed	not reported	not reported	reported but not related	not reported	/
George et al. (2020)	sufficiently addressed	not reported	reported but not related	reported but not related	reported but not related	reported but not related
Serral et al. (2020)	not sufficiently addressed	not reported	reported but not related	reported but not related	reported but not related	/
Kerpedzhiev et al. (2021)	sufficiently addressed	not reported	domain	reported but not related	reported but not related	central construct

Nevertheless, nine papers provide more details on the structure and procedure of the Delphi, from which the respective Delphi variant can be presumed. Identifying modifications is problematic, as the Delphi variant is usually not presented and must be based on a presumed variant. Still, some modifications were deemed evident and were therefore coded and analyzed. Modifications were identified within four studies, three of which related these to MM aspects. Schriek et al. (2016) and George et al. (2020) apply face-to-face meetings at some point in their Delphi study. While Schriek et al. (2016) relate this decision to the central construct to gain clarity and a shared understanding of the research objectives and constructs, George et al. (2020) do not justify and

relate this modification. Another critical point that must be considered is that neither paper examines this modification in terms of its impact on rigor (e.g., the impact of face-to-face meetings on anonymity). With the same intention as Schriek et al. (2016), justify Kerpedzhiev et al. (2021) their modification of a two-phase approach with a first phase to develop a common language between experts of different domains. De Bruin and Rosemann (2007) use a different modification, who relate their design decision to the complexity of the central construct and split their Delphi study into several ones with the same design to get a deep insight into each process area. Nevertheless, modifications are not enough supported from a scientific perspective. Thus, the danger of rigor loss in Delphi studies and the associated danger of quality loss and credibility exists in the field of MM development, and future research requires a more rigorous approach.

The results indicate a lack of clarity regarding the methodology, as different aspects are reported, although the general demands, starting points, and goals were similar. In summary, it can be said that a Delphi study for MM development should take various aspects into account to meet the requirement of rigor. Since modifications and adaptability are a significant advantage of the Delphi methodology and will remain relevant in future research, guidelines are needed for rigorously designing, conducting, and presenting the corresponding methodology.

5. Research Design Guidelines

According to the need for rigorously designing a Delphi study for MM development, this section introduces a first framework for a rigorous approach. This framework proposes guidelines, based on the preceding considerations, for researchers to ensure rigor in the design and, consequently, ensure the quality and credibility of the study's results. This paper proposes the general steps (illustrated in Figure 1) that scholars should consider when designing and conducting a Delphi study for MM development. The steps are discussed in detail below, focusing on the Delphi design step concerning the scope of the MM.

Figure 1: Design Steps of a Delphi Study for MM Development

5.1 Define Maturity Model Scope

As the overarching research goal is developing a MM or contributing to a specific part of a MM, scholars must first elaborate on the scope of the intended MM. This paper proposes using an overview of MM aspects for designing and developing constructs by Lasrado et al. (2015) to define relevant aspects for the development process. Although not all aspects of this overview may be relevant to every study, and no single recommendation can be made, it provides an overview of aspects researchers should look at and establishes a common vocabulary. Furthermore, this paper is based on the perception that all necessary aspects are included and, accordingly, recommendations are proposed as to which of these aspects should mandatorily be addressed. The recommendations are based on the analysis results and the theoretical work presented in section 2. First, the focus of the model should be defined. It should be determined whether a general or domain-specific model is to be developed, and the corresponding domain should be specified through practice or theory-based problem presentation. Domain and problem-specific requirements for the intended MM should be identified using literature and empirical data to ensure theoretical and practical relevance. Furthermore, the target audience and level of abstraction should be defined. These should be examined and presented, as they set specific knowledge and characteristics of experts that are needed for adequate development. The aspect constructs is considered crucial, as defining the central construct and its relation to the research goal is central for the later decision of a Delphi variant. Although Lasrado et al. (2015) provide a basis for a common vocabulary of the central construct, scholars should define their understanding of the construct and its position in the architecture of a MM. This ensures a clear understanding for scholars and provides a quick overview of the architectural understanding of the intended MM. This paper proposes utilizing the maturity metamodel by Bley et al. (2020) to define the central construct within a MM architecture. Furthermore, it is recommended to present information on every relevant aspect related to the later design process.

5.2 Define Development Base

As Becker et al. (2009) called for, one step in the MM development process is analyzing related MMs. The domain should be examined concerning related MMs, as all reviewed studies develop/derive initial aspects from the literature. Weaknesses in related MMs can be identified, and initial insights into structure, content and development strategy can be provided. Hence, it is recommended to elaborate elements of/for the central construct and define whether these are validated enough to use as a basis to extend the structural depth of the construct/ MM (e.g., develop sub-components for components) or if they should be refined. When developing initial aspects, one should keep in mind that sufficient information can hardly be achieved deeper than the structural layer of components (de Bruin et al., 2005). This aspect defines which content-related considerations and preliminary work from the literature are used as input for the Delphi study.

5.3 Design Delphi Study

The third step addresses the design of the Delphi study. A Delphi study's four generic and mandatory characteristics must be described. Additional elements must be considered, for which using Strasser´s (2017) taxonomy for Delphi studies is suggested. As stated in section 2 and Table 1, five design elements are identified as critical for MM development and are subsequently discussed.

The choice of the *Delphi variant* is seen as a significant decision (Skinner et al., 2015; Strasser, 2017). This paper perceives the epistemological objective of every Delphi study for MM development as concept development. Thus, relevant Delphi variants are the classical and ranking-type Delphi. In Table 4, a mapping of the two variants regarding their suitability for developing the central constructs is proposed, with "X" indicating a fit. A "(X)" indicates a possible but less probable fit, which was added, as the exact research target may vary within a central construct. The constructs, maturity level, (sub-)component, and goal are primarily analyzed regarding what characterizes them and less regarding their relative importance. In contrast, elaborating the development path is about the relative importance of the subcomponents or goals within a maturity level. As assessment elements and improvement actions were of less interest in the analyzed studies, a claim can only be made from a theoretical perspective. Both variants are of equal relevance, as various assessment elements and improvement measures need to be identified and their characteristics and relative importance explained in more detail.

Table 4: Allocation of Relevant Delphi Variants Regarding Their Suitability for the Development of Different Central Constructs

Delphi Variant	Maturity Level	(Sub-) Component	Goal	Development Path	Assessment Elements	Improvement Actions
Classic Delphi	X	X	X	-	X	X
Ranking-Type Delphi	(X)	(X)	(X)	X	X	X

Additionally, the expert panel must be defined. The author proposes using a knowledge resource nomination worksheet (Delbecq et al., 1975; Okoli and Pawlowski, 2004) to identify and select adequate experts for the panel. Although this process may seem resource-intensive, it is beneficial because selecting appropriate experts is crucial for successful study implementation and reliable and valid results. For complex and diverse domains, a heterogeneous disciplinary panel and different groups of experts for specific components can be helpful (e.g., for multi-dimensional MMs). This may result in several simultaneous Delphi Studies for the intended constructs as modification, as in the study of de Bruin and Rosemann (2007). Independently of the multidisciplinary aspect, it is recommended to use a heterogeneous panel of academics and practitioners. It is vital to define the experts based on criteria relevant to the field and present these criteria and the differences between each group of experts. For defining the expert criteria, domain specifics should be included in this process and the intended auditorium of the MM. Hence, the panel should be related to the domain and complexity of the intended central construct(s). When selecting the experts for the panel, the number of participating experts must be defined (Gallego and Bueno, 2014). No precise information exists in the literature on how many experts are adequate (Paré et al., 2013). Regarding the panel size, various suggestions can be found in the literature, such as not less than seven (Dalkey and Helmer, 1963), between ten and 15 (Linstone et al., 1975), or around 30 (Delbecq et al., 1975). We follow the reommandation by Gallego and Bueno (2014) of ten to 15 experts for homogeneous panels and the recommendation by Linstone et al. (1975) of four to five for each expert group for heterogeneous panels are recommended.

Furthermore, the number of iterations is vital for a Delphi study. It is appropriate to establish rules for reaching an adequate level of consensus and thus end the data collection (Okoli and Pawlowski, 2004; Paré et al., 2013). Defining a consensus criterion as a stopping rule should be related to the central construct and depends on the Delphi type (e.g., typical criteria for ranking type Delphi is Kendell`s W). Additionally, as the consensus criterion is usually defined as a threshold of a statistical value, it can be beneficial to relate this decision to findings of the development base. Furthermore, as the number of rounds increases, the speed of observed convergence between experts decreases, and proceeding over more than three rounds is not appropriate (Linstone et al., 1975). Nevertheless, various recommendations for the number of iterations exist in the literature, ranging from two to ten rounds (Lang, 1995). Recommendations of three rounds (Brooks, 1979; Mulligan, 2002), more than two rounds (Dransfeld et al., 2000; Murphy et al., 1998; Rowe and Wright, 1999) or up to four rounds (Erffmeyer et al., 1986; Hsu and Sandford, 2007) occur most frequently in the literature. Accordingly, a general recommendation of two to four rounds seems to represent the consensus among scholars in the field of Delphi methodology. This discussion leads to further design decisions as researchers must explain which and how aspects can influence the number of iterations. Relating to the research objective (for which construct is consensus intended), scholars must address the issue of what information can be developed and provided in advance to reach consensus with a high degree of probability without providing too much information to overwhelm the experts or bias the results. For example, a greenfield approach may lead to a high number of rounds until a satisfactory consensus is reached, whereas showing all sub-components with elaborated descriptions may influence opinions too much or discourage experts from further participation. Thus, studies need to be set on a reasoned compromise between these two design decisions, as it allows to minimize the Delphi rounds without constraining the experts too much. Another aspect that significantly influences the number of iterations is the overall extent of elements that need to be developed. As previously mentioned, this can be tackled by modifying the study into several ones with the same design.

As mentioned before, the research aim of the Delphi study is related to the development basis, and it can be distinguished between extending and refining. This primarily affects the design of the first round, as the development basis is primarily used to evaluate and refine the theoretical findings (resulting in a predominantly quantitative first round) or obtain the experts' opinions and then synthesize them to compare or expand the theory-based model. As the findings of the analyzed studies show, both approaches can be combined in the first round.

An advantage of the Delphi method is its modifiability, which at the same time presents a crucial challenge to its rigor. Hence, it is of relevance that researchers report and justify modifications. Thereby, modifications must be related to the research objective and the corresponding characteristics of the MM, which should be reported in the MM´s scope. Due to the diversity of applications, further adjustments in the method may be necessary. Paré et al. (2013) advise following at first the recommendations from corresponding literature on the Delphi variant, followed by checking whether corresponding modifications have already been validated in the literature under comparable conditions and if the modification is novel, the impact on the results, the rigor and the study design as well as possibilities to increase validity need to be explored and presented (Paré et al., 2013).

5.4 Conduct Delphi and Define Round Adjustments

Step four defines conducting the Delphi study. First, the entire study design should be reviewed regarding the relations between the research goal(s), the intended MM, the Delphi methodology, and its correct application. The Delphi study can be conducted according to the predetermined design if no adjustment is required. During this step, detailed documentation of feedback or recommendations from experts and taken actions are required. Only the first round can be one hundred percent predesigned. In each subsequent round, the exact design and layout of the questionnaire will depend on the previous round(s) results. Each iteration can focus on a different goal concerning the central construct, distinguishing brainstorming, narrowing-down, validation, and ranking (Paré et al., 2013). Hence, scholars should report what specific goal was followed each round, what results were achieved, and how they relate to the design of the next iteration. Besides designing the specific rounds, it is essential to provide the panel with feedback on the results in every iteration, resulting in the presented process in Figure 2 for each iteration.

Figure 2: Process of a Delphi Round

5.5 Report Results

The last step comprises presenting results to researchers and practitioners for dissemination. Scholars should indicate how and what should be done in the further development process of the MM. As the proceeding steps of the model development are about the practical transfer and application (Becker et al., 2009; de Bruin et al., 2005; Mettler, 2010), action recommendations and possible hurdles concerning the study's central construct should be indicated. Furthermore, the used Delphi procedure, including modifications, should be reflected and insights for future applications given.

6. Conclusion

This paper examines the rigor of the MM development process using a Delphi approach in the IS domain. First, both methodological approaches were examined in their approach and rigorous demands focusing on their interaction. A systematic literature analysis was conducted and analyzed with a deductive content analysis based on these elaborations. The findings show no clarity on the design aspects, and their methodological interactions exist, and a clarification is needed. Hence, guidelines to develop a Delphi study for MM development were elaborated and presented. This is the first attempt towards a rigorous approach but does not provide a step-by-step approach, as Delphi studies are characterized by their modifiability, and different Delphi types and a variety of MM constructs exist that can be a goal of the development process.

The developed guidelines integrate the different methodological aspects of the MM development and the Delphi methodology. It represents an orientation framework for researchers to design a Delphi study for MM development from a rigorous perspective. To this end, it provides scholars with an overview of aspects that must be considered and in which order they should be addressed. Furthermore, it is explained how the aspects are linked to one another and what decisions regarding opposing aspects in the design process can occur. Hence, scholars can use the guidelines to design a rigorous Delphi study for MM development. Although this paper focuses on the MM context, the general design approach and decisions are transferable and applicable to other domains. Hence, this research contributes to the Delphi literature by providing insights into how relevant elements should be addressed in the designing process of a Delphi Study. Thus, scholars should investigate how the presented guidelines must be adapted for other domains.

The research areas of MM and Delphi are extensive and diverse, leading to numerous existing and possible approaches. It should be noted that this paper only analyzes literature that indicates that a Delphi method to develop MM aspects was applied. Scholars that might have applied this method and did not indicate the methodology could not be included and might have applied it rigorously. This paper does not intend to give a solution that fits every possible scenario of combining these two methods but instead suggests aspects and guidelines to maneuver through the design process. This research gives a foundation and orientation for researchers that want to use the Delphi method for MM development. Furthermore, researchers are encouraged to expand this work on missing aspects and more concrete application scenarios, such as the sole application of a ranking-type Delphi to develop a maturation path.

Acknowledgements

The Article Processing Charges (APC) were funded by the joint publication fund of the TU Dresden, the Medical Faculty Carl Gustav Carus, and the SLUB Dresden.

References

Becker, J., Knackstedt, R. and Pöppelbuß, J. (2009) 'Developing Maturity Models for IT Management', Business & Information Systems Engineering, vol. 1, no. 3, pp. 213–222.

Brooks, K. W. (1979) 'Delphi technique: Expanding applications', North Central Association Quarterly, vol. 53, no. 3, pp. 377–385.

Carvalho, J. V., Rocha, Á., van de Wetering, R. and Abreu, A. (2019) 'A Maturity model for hospital information systems', Journal of Business Research, vol. 94, pp. 388–399.

Dalkey, N. and Helmer, O. (1963) 'An experimental application of the Delphi method to the use of experts', Management science, vol. 9, no. 3, pp. 458–467.

de Bruin, T. and Rosemann, M. (2007) 'Using the Delphi technique to identify BPM capability areas', ACIS 2007 Proceedings, 42 [Online]. Available at http://aisel.aisnet.org/acis2007/42.

de Bruin, T., Rosemann, M., Freeze, R. and Kaulkarni, U. (2005) 'Understanding the main phases of developing a maturity assessment model', Australasian Conference on Information Systems (ACIS), pp. 8–19.

Delbecq, A. L., van de Ven, A. H. and Gustafson, D. H. (1975) Group techniques for program planning: A guide to nominal group and Delphi processes, Scott, Foresman.

Dransfeld, H., Pemberton, J. and Jacobs, G. (2000) 'Quantifying Weighted Expert Opinion', Technological Forecasting and Social Change, vol. 63, no. 1, pp. 81–90.

Elo, S. and Kyngäs, H. (2008) 'The qualitative content analysis process', Journal of advanced nursing, vol. 62, no. 1, pp. 107–115.

Erffmeyer, R. C., Erffmeyer, E. S. and Lane, I. M. (1986) 'The Delphi Technique: An Empirical Evaluation of the Optimal Number of Rounds', Group & Organization Studies, vol. 11, 1-2, pp. 120–128.

Gallego, D. and Bueno, S. (2014) 'Exploring the application of the Delphi method as a forecasting tool in Information Systems and Technologies research', Technology Analysis & Strategic Management, vol. 26, no. 9, pp. 987–999.

George, A., Schmitz, K. and Storey, V. C. (2020) 'A Framework for Building Mature Business Intelligence and Analytics in Organizations', Journal of Database Management, vol. 31, no. 3, pp. 14–39.

Gupta, U. G. and Clarke, R. E. (1996) 'Theory and applications of the Delphi technique: A bibliography (1975–1994)', Technological Forecasting and Social Change, vol. 53, no. 2, pp. 185–211.

Hasson, F. and Keeney, S. (2011) 'Enhancing rigour in the Delphi technique research', Technological Forecasting and Social Change, vol. 78, no. 9, pp. 1695–1704.

Hsu, C.-C. and Sandford, B. A. (2007) 'The Delphi Technique: Making Sense of Consensus', Practical Assessment, Research, and Evaluation, no. 12.

Karabacak, B., Yildirim, S. O. and Baykal, N. (2016) 'A vulnerability-driven cyber security maturity model for measuring national critical infrastructure protection preparedness', International Journal of Critical Infrastructure Protection, vol. 15, pp. 47–59.

Kerpedzhiev, G. D., König, U. M., Röglinger, M. and Rosemann, M. (2021) 'An Exploration into Future Business Process Management Capabilities in View of Digitalization', Business & Information Systems Engineering, vol. 63, no. 2, pp. 83–96.

Labaka, L., Maraña, P., Giménez, R. and Hernantes, J. (2019) 'Defining the roadmap towards city resilience', Technological Forecasting and Social Change, vol. 146, pp. 281–296.

Lahrmann, G., Marx, F., Mettler, T., Winter, R. and Wortmann, F. (2011) 'Inductive Design of Maturity Models: Applying the Rasch Algorithm for Design Science Research', in Jain, H., Sinha, A. P. and Vitharana, P. (eds) Service-Oriented Perspectives in Design Science Research, Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 176–191.

Lang, T. (1995) 'An overview of four futures methodologies', Manoa Journal of Fried and Half-Fried Ideas, vol. 7, pp. 1–43.

Lasrado, L., Vatrapu, R., Karsgaard, H. B. and Kjaer, J. F. (2017) 'Towards sustainable design for maturity measurement marketplace', International SERIES on Information Systems and Management in Creative eMedia (CreMedia), 2016/2, pp. 25–33.

Lasrado, L. A., Vatrapu, R. and Andersen, K. N. (2015) 'Maturity Models Development in IS Research: A Literature Review', Selected Papers of the IRIS, no. 6 [Online]. DOI: 10.13140/RG.2.1.3046.3209.

Linstone, H. A., Turoff, M. and others (1975) The delphi method, Addison-Wesley Reading, MA.

McKenna, H. P. (1994) 'The Delphi technique: a worthwhile research approach for nursing?', Journal of advanced nursing, vol. 19, no. 6, pp. 1221–1225.

Mens, J. and Ravesteyn, P. (2016) 'Using the Delphi Method to Identify Hospital-Specific Business Process Management Capabilities in The Netherlands', Bled eConference, 22.

Mettler, T. (2010) Supply management im Krankenhaus: Konstruktion und Evaluation eines konfigurierbaren Reifegradmodells zur zielgerichteten Gestaltung, Sierke Göttingen.

Mettler, T. (2011) 'Maturity assessment models: a design science research approach', International Journal of Society Systems Science, vol. 3, 1-2, pp. 81–98.

Mettler, T. and Rohner, P. (2009) 'Situational maturity models as instrumental artifacts for organizational design', Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST '09. Philadelphia, Pennsylvania, 07.05.2009 - 08.05.2009. New York, New York, USA, ACM Press, pp. 1–9.

Mulligan, P. (2002) 'Specification of a capability-based IT classification framework', Information & Management, vol. 39, no. 8, pp. 647–658.

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F., Askham, J. and Marteau, T. (1998) 'Consensus development methods, and their use in clinical guideline development', Health technology assessment (Winchester, England), vol. 2, no. 3, i-iv, 1-88.

Nurcahyo, R., Darmawan, D., Jannis, Y., Kurniati, A. and Habiburrahman, M. (2018) 'Maintenance Planning Key Process Area: Case Study at Oil Gas Industry in Indonesia', 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). Bangkok, 16.12.2018 - 19.12.2018, IEEE, pp. 1704–1708.

Object Management Group (2008) Business Process Modeling Notation (BPMN): Version 1.0 [Online]. Available at https://www.omg.org/spec/BPMM/1.0/PDF (Accessed 18 September 2021).

Okoli, C. and Pawlowski, S. D. (2004) 'The Delphi method as a research tool: an example, design considerations and applications', Information & Management, vol. 42, no. 1, pp. 15–29.

Özturan, M., Mutlutürk, M., Çeken, B. and Sarı, B. (2019) 'Evaluating the information systems integration maturity level of travel agencies', Information Technology & Tourism, vol. 21, no. 2, pp. 237–257.

Paré, G., Cameron, A.-F., Poba-Nzaou, P. and Templier, M. (2013) 'A systematic assessment of rigor in information systems ranking-type Delphi studies', Information & Management, vol. 50, no. 5, pp. 207–217.

Pereira, R. and Serrano, J. (2020) 'A review of methods used on IT maturity models development: A systematic literature review and a critical analysis', Journal of Information Technology, vol. 35, no. 2, pp. 161–178.

Radosavljevic, M., Barac, N., Jankovic-Milic, V. and Andjelkovic, A. (2016) 'Supply chain management maturity assessment: challenges of the enterprises in Serbia', Journal of Business Economics and Management, vol. 17, no. 6, pp. 848–864.

Rauch, W. (1979) 'The decision Delphi', Technological Forecasting and Social Change, vol. 15, no. 3, pp. 159–169.

Reeb, S. and Pinnecke, M. (2021) 'Continuous Improvement with Maturity Models – A Procedural Model for Action Derivation', 25th Pacific Asia Conference on Information Systems, PACIS 2020, Dubai, UAE, July 12-14, 2021.

Reyes, H. G. and Giachetti, R. (2010) 'Using experts to develop a supply chain maturity model in Mexico', Supply Chain Management: An International Journal, vol. 15, no. 6, pp. 415–424.

Rowe, G. and Wright, G. (1999) 'The Delphi technique as a forecasting tool: issues and analysis', International Journal of Forecasting, vol. 15, no. 4, pp. 353–375.

Schmidt, R. C. (1997) 'Managing Delphi Surveys Using Nonparametric Statistical Techniques', Decision Sciences, vol. 28, no. 3, pp. 763–774.

Schriek, M., Türetken, O. and Kaymak, U. (2016) 'A Maturity Model for Care pathways', 24th European Conference on Information Systems (ECIS). Istanbul, Turkey, June 12-15, 127.

Serral, E., Stede, C. V. and Hasic, F. (2020) 'Leveraging IoT in Retail Industry: A Maturity Model', 2020 IEEE 22nd Conference on Business Informatics (CBI). Antwerp, Belgium, 22.06.2020 - 24.06.2020, IEEE, pp. 114–123.

Skinner, R., Nelson, R. R., Chin, W. W. and Land, L. (2015) 'The Delphi Method Research Strategy in Studies of Information Systems', Communications of the Association for Information Systems, vol. 37, no. 1, pp. 31–63.

Smits, D. and van Hillegersberg, J. (2017) 'The development of a hard and soft IT governance assessment instrument', Procedia Computer Science, vol. 121, pp. 47–54.

Stelzl, K., Röglinger, M. and Wyrtki, K. (2020) 'Building an ambidextrous organization: a maturity model for organizational ambidexterity', Business Research, vol. 13, no. 3, pp. 1203–1230.

Stojanov, I., Turetken, O. and Trienekens, J. J. (2015) 'A Maturity Model for Scaling Agile Development', in 41st Euromicro Conference, pp. 446–453.

Strasser, A. (2017) 'Delphi method variants in information systems research: Taxonomy development and application', Electronic Journal of Business Research Methods, vol. 15, no. 2, pp120-133.

Team CMMI Product (2006) 'CMMI for Development, version 1.2'.

Torrecilla-Salinas, C. J., Troyer, O. de, Escalona, M. J. and Mejías, M. (2019) 'A Delphi-based expert judgment method applied to the validation of a mature Agile framework for Web development projects', Information Technology and Management, vol. 20, no. 1, pp. 9–40.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R. and Cleven, A. (2009) 'Reconstructing the giant: On the importance of rigour in documenting the literature search process', ECIS.

Natural Language Processing and Machine Learning Approach to Teaching and Learning Research Philosophies and Paradigms

Marcia Mkansi and Ntombi Mkalipi
University of South Africa, South Africa

Marcia.mkansi@gmail.com (corresponding author)
36898570@mylife.unisa.ac.za

Abstract: The paper cogitates on the critical advent of 4th IR focusing on the concept of machine learning (ML) underpinned by natural language processing (NLP) to demonstrate how research philosophies and paradigms can be better taught and learned for students' benefit. A systematic literature review was earmarked for its depth and textual inquiry from scholarly arguments. The main purpose of this paper is to aver a progressive technological approach towards better comprehensive research paradigms and philosophies, which are complex domains with diverse variety in higher education and a cause of discomfort for students at post-graduate levels. Using quantitative algorithm, the natural language processing and machine learning-inspired digital model poses questions that place students in a reflexive mode and draws their articulated responses as inputs that model their worldviews against a host of philosophies in the database. The paper revealed that, discordant with previous scholars who advocated for a single philosophical assumption for a field, subject or researcher, such as the existence of a pure positivist and/ or pure interpretivist, purist philosophical assumptions should be challenged to benefit students and academics. It means that the digital discovery of research philosophies and paradigms extends the work of previous theorists to the technologically inspired discovery of episteme, ontology and axiology. By its nature, the use of NLP becomes an advanced channel on how we know what we know and the nature of the reality and values being displayed. The paper contributes to the evocation of deep learning arising from new philosophies and methods. The inquiry-based teaching approach transforms learning from the generic push-teaching method that assumes universality to the fostering of a reflexive approach that helps resolve the deep ideological approaches that caused the polarisation. The manner in which NLP and ML are able to extract information relevant to knowledge or philosophical discovery paves the way for approaches that can lead to the depolarisation and decolonisation of research philosophies, which can ultimately boost the development of research students.

Keywords: Research philosophies and paradigms, Natural language processing, Teaching and learning, Research methodology, Machine learning

1. Introduction

Research paradigm advocates (Alvesson and Skoldberg, 2009; Guba and Lincoln, 1994; Killam, 2013; Polit and Beck, 2008; Saunders et al., 2012; Steen and Roberts, 2011) can attest that to understand research, one must examine the philosophies that underlie the researcher's paradigm. This chain of evidence encourages academics to observe the same phenomena in different ways, and subsequently, to derive different kinds of knowledge from the various different philosophical perspectives (Blaikie, 1993, 2000; Guarino, 1998). However, the knowledge-creation process is replete with vast array of data that is hard to capture in a single mind, book, journal or encyclopaedia. Furthermore, the dichotomies that exist because of the preferential philosophical approaches of supervisors can also impact students' research development (Acheampong et al., 2015; Silverman, 2010). The dichotomies and data of the research approaches are amplified by the parallels and dilemmas of the concepts, which make it hard for students to comprehend the research philosophies, paradigms and their role in knowledge generation (Mkansi and Acheampong, 2012; Scotland, 2012; Sefotho, 2015). The dichotomies and dilemmas are especially applicable to research at the post-graduate level, and research has revealed the undesirable picture of a near-total absence of understanding among the students in relation to the philosophies and paradigms (Hesse-Biber, 2015; Scotland, 2012).

Although the above-mentioned scholars have revealed the breadth of the issues being experienced as challenges in the teaching, learning, and understanding of research philosophies at post-graduate level, little, if any, efforts have been made to minimise the challenges by exploring some of the latest technologies. In light of the observed gap, how can technology help lecturers and students at post-graduate level to teach and learn a complex subject with different types of philosophies? Also, how can technology contribute to Te'eni et al.'s (2015) genre of knowledge discovery-based research methodologies, and Davison and Martisons' (2016) call for an inclusive approach to the development of new theories?

Reference this paper: Mkansi, M. and Mkalipi, N., 2023. Natural Language Processing and Machine Learning Approach to Teaching and Learning Research Philosophies and Paradigms *The Electronic Journal of Business Research Methods*, 21(1), pp. 14-30, available online at www.ejbrm.com

In an effort to address the above two questions, and also to improve the knowledge and understanding of research philosophies and paradigms at post-graduate level, this study use natural language processing (NLP) and machine learning (ML) (which are technological approaches) to introduce a way of simplifying the teaching and learning of research philosophies. NLP present opportunities for the comprehension and processing of human language and its translation into machine text whilst ML provides opportunities for comprehension and processing, in great detail, of research information and data (big data) that is beyond the command of a lecturer, book, journal, and/or student. It is the latter capabilities that have foregrounded the quantitative algorithm underlying the digital-based model in this study. This digital-based model processes a range of qualitative questions based on epistemological, ontological and axiological assumptions derived from consolidated literature. The systematic literature review became a compelling method for analysis, and the NLP and ML demonstrated the possibilities of achieving the objectives of the research. The manner in which NLP and ML are able to extract information relevant to knowledge or philosophical discovery paves the way for approaches that can lead to the depolarisation and decolonisation of research philosophies, which can ultimately boost the development of research students. This study underpins research methodology in business and management through the capability of technology in augmenting human knowledge using algorithms that have the capabilities to extract and process information from a large database, transforming it into deep learning that can guide decision making and knowledge production for scholars and students engaging in research for knowledge generation output. The advancement of the business and management research methodology is aligned with modern technology advancement and reduces the stereotyping and over-compensated positivist and interpretivist philosophies towards broader reflexive modes and plenitudes of philosophies and paradigms - including newly developed philosophies and paradigms most suitable for business and management.

The following section presents a background to research philosophies and paradigm challenges vis-a-vis and contemporary views on teaching and learning to dovetail with NLP and ML-based philosophical and paradigm positioning. Section 3 presents the methods of dataset collection and the analysis and annotation process that underpins the digital-based model architecture's performance, input and output data. The final section provides insights into the implication of the model and future research.

2. Philosophy and Paradigm Challenges

The theoretical background to research philosophies and paradigms gained momentum in the 13th century (Jensen, 2000; Lincoln, 1990). Eminent scholars of the time, such as John Locke, David Hume, René Descarte, Immanuel Kant and Hans Reichenback were some of the pioneers of philosophical assumptions; they also acted as advocates for knowledge generation and played a role in the development of research methods (Denzin and Lincoln, 2003; Killam, 2013; Miller and Grimwood, 2015; Saunders et al., 2019). However, since the origination of the concepts, there has been little progress towards technological advancement. This applies especially to those concepts that are reportedly difficult, and that consist of large amounts of data with many variations, such as research philosophies and paradigms, which causes, in effect, a living gulf between theory and practice.

In an effort to expose the gap, several scholars (Hesse-Biber, 2015; Scotland, 2012; Sefotho, 2015) revealed a series of issues related to understanding, and they aimed to raise awareness about the concepts and the challenges being experienced in connecting the relationship between research philosophical assumptions, methods and knowledge. For example, Acheampong et al. (2015) and Silverman (2010) revealed dichotomies in research methods that are due to the preferential philosophical approaches of supervisors, and which highlight the potential impact on students' ability to do sound research. A study by Hesse-Biber (2015) on the problems and prospects in the teaching of research revealed the pedagogical challenges students faced in understanding how research methods connect to philosophical assumptions, and at worst, some of the students did not know that they have philosophical assumptions.

Makombe's (2017) investigation into the relationship between research philosophies, methods and design reported that a good percentage of researchers avoid the discussion of their guiding paradigm due to a lack of knowledge related to the topic. Mkansi and Acheampong (2012), and Mackenzie and Knipe (2006) reported on the students' dilemmas in understanding the research philosophies and paradigms; problems that are caused and amplified by the incomprehensible classifications and contradictory terminology of philosophies. In emphasising the importance of research paradigms and philosophies, Sefotho (2015) suggested that the philosophy must be the guiding force that drives the researcher in developing a thesis, and if absent, the researcher's investigation is directionless. Mkansi (2018), similarly revealed a lack of application or acknowledgement of research philosophies in high impact knowledge hubs.

In a review of the gaps in technology aimed at improving the teaching and learning of research methods, this study found several technological advancements that complement qualitative and quantitative data analysis techniques such as SPSS (Field, 2009), Analysis of Moment Structures (Amos) (Arbuckle (2008), Maxqda (Saillard, 2011; Faherty, 2010) and Nvivo (Faherty, 2010). There are few exceptions that explored and studied the NLP and ML application to research methods. For example, Chen et al. (2021) demonstrates how and why machine learning can advance social science from analysing causality and correlations to prediction. Whereas Chang et al. (2021) use NLP to illustrate its strength in rapid analysis of big mixed methods data in times of catastrophic change. In addition, Javed et al. (2021) evaluates the objectivity of ML as a field of study. Although these studies make significant headways in their application of NLP and ML to research methodologies, their focus is mostly limited to data analysis of primary qualitative, quantitative and mixed methods, rather than a diagnostic inquiry-based learning of research philosophies and paradigms.

Beyond application to research methods, Kanchan and Yadav's (2022) systematic review of NLP research conducted from 1997 to 2021, revealed the application of sentiment analysis in many business spheres that includes product reputations to customers reviews etc., but none on the aspect of research philosophies and paradigms. Similarly, Zhao et al.'s (2021) report of 404 studies that used NLP for requirements engineering excludes matters of research philosophies and paradigms. Viewed differently, although contemporary applications of NLP to teaching and learning of research methods exist, they serve as forms of advanced alternatives or compliments of other technology-based analysis tools. Thus, research philosophies remain complex domains in higher education (Muhaise et al., 2020; Groeesler, 2017) and a cause of discomfort for students at post-graduate levels. While research competencies are becoming increasingly invaluable for employability, there is insufficient research on innovative pedagogical approaches to support students in this aspect (Daniel, 2018). Hence, this study's consideration of NLP and ML techniques. A review of the defined NLP and ML is presented in subsequent sections and its application in the study.

2.1 Natural Language Processing

Research is a field that is characterised by an unprecedented amount of data, ranging from 'episteme' or knowledge, and 'ontologia' as an inner cognitive state that is connected to, among others, observations, experiments, abstracts, narratives, interpretations of experiences, and conceptualised practice (case studies). Some of the research concepts, such as philosophies and paradigms, consist of a large, diverse amount of data and variations that have blurred the similarities and differences (see example screenshot of some of the philosophes in Figure 2). There is such a vast number of variations and large amount of data related to research philosophies and paradigms, and to complicate matters it is not available in a single book, journal or review, which leads to the polarisation and limited development of research students.

Natural language processing presents opportunities for the comprehension and processing, in great detail, of research information and great amounts of data that is beyond the comprehension of a lecturer, book, journal, and/or student. The relevance of NLP lies in its ability to analyse and help machines understand the nuances in human languages (Marr, 2016). NLP is concerned with the processing of written and spoken human language through computational techniques, and is equally dependent on machine learning (ML) in its pursuit of helping machines comprehend the human language (Marr, 2016). In this regard, it is necessary to capture the user's expression of their worldview through a series of questions that serve to present a specific customised representation of the user's research philosophical assumptions beyond the cluster-level results from ML.

There are various types or concepts of NLP, for example, there is natural language understanding (NLU), or natural language translation, which are both concerned with the comprehension or understanding of human language and its translation into machine text (Bonaccorso, 2017; Kaminski, 2017). Furthermore, natural language generation (NLG) generates human language text from machine text, or numbers, by making choices based on the grammatical content, correctness and readability of the generated language (Bonaccorso, 2017; Kaminski, 2017). Information extraction (IE) has also become popular with the rise of social media, and is mainly used for sentiment analysis (Derczynski et al., 2014; Jiang, 2012). This study adopted the NLU concept for its strength in capturing an unstructured understanding of users' inputs or texts of their worldview, and its translation into a supervised ML algorithm.

A search of the literature did not reveal any previous work done in relation to the use of NLP for the prediction or classification of user ideas and concepts into research philosophies and paradigms. However, it is noted that NLP has been used in marketing research (Leeson et al., 2019; Yu and Kwok, 2011), aviation (Kumar and Zymbler, 2019, and education (Waters et al., 2017;). The evident lack of NLP application in research philosophies provided

the impetus for addressing Davison and Martinson's call for new ways of developing theories, including those emanating from indigenous backgrounds.

The NLP procedure involves the following procedures: tokenising, part of speech tagging (PoS Tagging), word embeddings, stemming and lemmatisation, and named entity recognition (NER) (see screenshot of Figure 3 in the methodology section).

- Tokenising uses a lexer (lexical analysis) to identify and separate instances of a sequence of characters or words, referred to as tokens, in a given sentence (Zhao and Kit, 2011). These tokens are used as input for the process of parsing or text mining. Parsing defines the grammatical rules for the tokens and the relationship between them in an abstract form by producing a dependency tree (Zeroual and Lakhouaja, 2018).
- Part of speech tagging (PoS) assigns morpho-syntactical features to words based on their context, thus enabling simple syntactic searches.
- Lemmatisation reduces words to their dictionary form, taking into account the meaning of words in sentences or nearby sentences, whereas stemming establishes relationships between words by reducing them to their basic or root form (Pedrycz and Chen, 2016).
- Named entity recognition involves the extraction of specific words or entities from within text. These entities are identified and linked to a category instance in a knowledge base to resolve their contextual meaning (Chang et al., 2016).
- Word embeddings, which is a distributed representation of text, is often used to overcome the weakness computers have in processing natural language by mapping words to vectors of numerical values, and also to map the relationships between words by creating similar representations for words with similar meaning (Li and Sha, 2017).

2.2 Machine Learning

Machine learning presents opportunities for the comprehension and processing, in great detail, of research information and data (big data) that is beyond the comprehension of a lecturer, book, journal, and/or student. The strength of ML is in augmenting human knowledge, using algorithms that have the capabilities to extract and process information from a large database (Carpenter, 2019), transforming it into deep learning that can guide decision-making and knowledge production. Yet, such ML opportunities remain unexplored in the research methods' space.

Machine learning can be classified into unsupervised, supervised, and semi-supervised learning (Brunton et al., 2019). Supervised learning algorithms (for example, naïve Bayes; LR; SVM; regressions such as linear, and the gaussian process; classifications, such as decision trees, and random forests; optimisation and control techniques, such as genetic algorithms, and deep model predictive controls) build models by learning relationships between descriptive features (input) and target features (output) based on historic datasets (Kelleher et al., 2015). The algorithm is trained by supplying it with known inputs and their matching responses, and from the learned relationship it can predict responses for unknown inputs (Shouval et al., 2013). Unsupervised learning machines require no supervision as they are capable of discovering information independently, through the use of, for example, clustering techniques such as k-means, spectral clustering, and dimensionality reductions such as autoencoders and diffusion maps (Brownlee, 2018). Semi-supervised techniques, such as reinforcement learning (Q-learning, Markov decision process, etc.), and generative models, such as generative adversarial networks, learn with partially labelled data (Brunton et al., 2019). This study adopted the supervised ML concept to complement NLP.

3. Methodology

This study used NLP and ML to discover research philosophies and paradigms. The class of techniques in use belongs to computer science and is highly quantitative in nature. Quantitative research deals with numeric data (Creswell, 2013, 2014; Saunders et al., 2016), which for the purposes of this study were mainly NLP algorithms. The study combined systematic literature and algorithm methodology to investigate how NLP can help lecturers and students at post-graduate level to teach and learn a complex subject (namely, the subject of research philosophies and paradigms) which has more than 180 different types of philosophies collected from various sources of literature.

The methodology embodies the research philosophies and paradigms artefacts developed for this study, which contributes to Peffers et al.'s (2018) quest for the acknowledgement of design science research in information

systems. In addition, it demonstrates the approaches that can lead to a depolarisation of research philosophies that can boost the development of research students and knowledge production. The systematic literature review provided contextual data (for the database), and the variables crucial for gathering input data from users based on the multidimensional philosophical assumptions offered by Saunders et al. (2016) and Guba and Lincoln (1994).

As such, the next section presents the methods for data collection, followed by a discussion of the data annotation process. The systematic literature review represents the foundation of the model architecture developed. Put differently, the research philosophies and paradigms (RPP) model's contextual database is grounded on scientific knowledge that stretches from the early centuries to date.

3.1 Data set and Collection

In gathering and preparing the data set to serve as contextual data (database) and which would also be used for deriving the input questions, the present study followed Denyer and Transfield's (2000) principles of a systematic literature review (see Figure 1). The principle of a systematic literature review involves scoping the subject matter by using keywords (for example, research, philosophies, paradigms, interpretivism, positivism, etc.) to search for relevant subject matter across a host of information hubs (journals, reviews, encyclopaedia, conferences, etc.). Information relevant to the literature review was published in the International Scientific Indexing (ISI), Scopus list, Science Electronic Library Online (SciELO) and the International Bibliography of Social Sciences (IBSS), dating back from the early centuries up to the year 2020. In addition, the researchers consulted research methods textbooks available from various publishing houses and google scholars, including those by popular scholars, such as Creswell (2014), and Saunders et al. (2016). The in-depth search of the information hubs produced a total of 180 research philosophies. However, most of the philosophies did not have enough reviews and data, which proved difficult for the NLP algorithm. As such, the research philosophies with limited information were excluded, leaving a total of 180 usable research philosophies for input to the system.

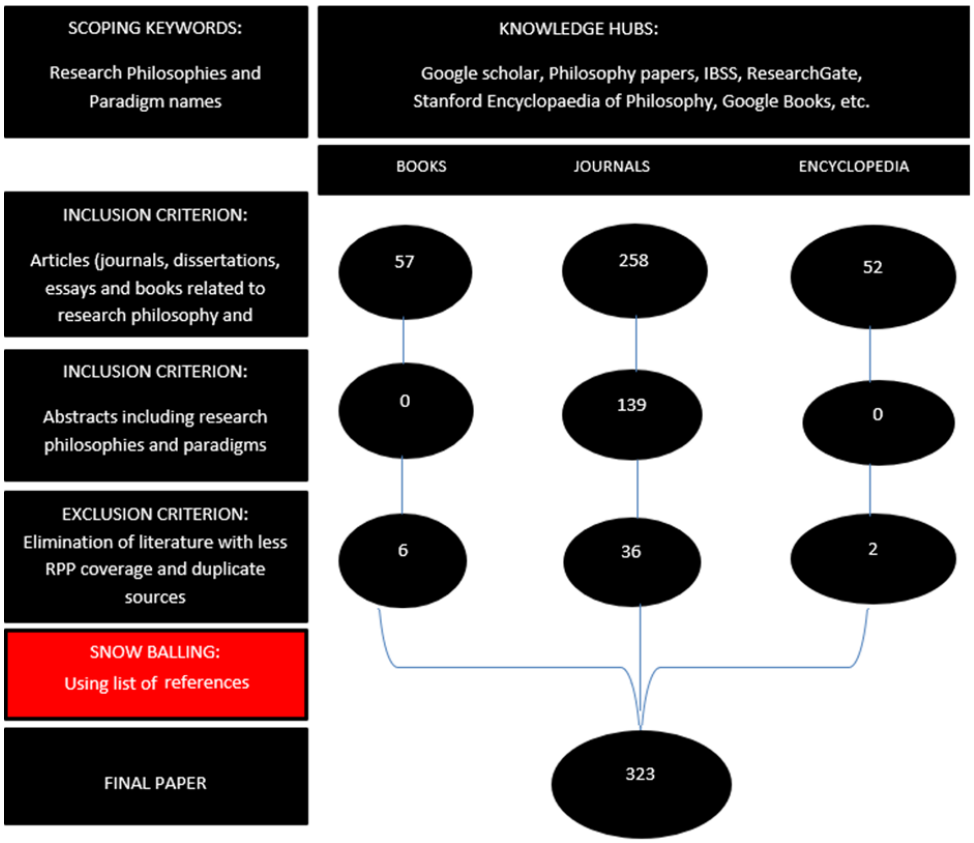


Figure 1: Process of Systematic Data Collection of Research Philosophies and Paradigms

(Adapted from Denyer and Transfield, 2000)

The literature findings supported the development of the model in terms of the contextual data (database) and the design of questions that reflect the multidimensional structure of a research philosophy based on the various

ontological, epistemological, and axiological stances. In particular, data derived from the literature sources was crucial in grounding both the RPP model and the study's objectives on credible scientific knowledge, and for identifying relevant factors from the different sources of data. The relevance of data sources in achieving conformability and credibility, whilst limiting subjectivity is highly endorsed by Saunders et al. (2019) and Yin et al. (2018) (see Appendix 1 for a sample corroboration of research philosophies' data against different sources of literature).

3.2 Data Analysis and Annotation

The data was first analysed using a summative content analysis which involved counting the number of scholarly articles that included research philosophies' keywords, including the interpretation of their respective underlying context, as endorsed by Hsieh and Shannon (2005). The analysis focused on the meaning, characteristics, and the options emanating from the meaning, which was useful for the NLP analysis. A Microsoft Excel spreadsheet was used to organise each philosophy within the epistemological, ontological, and axiological lenses or framework, as offered by Saunders et al. (2016), Seddon and Scheepers (2012), and Guba and Lincoln (1994) (see example screenshot of Figure 2). This firstly, involved the Microsoft Excel spreadsheet with the content analysis of the streamlined 180 research philosophies to create a corpus of RPPs (see Figure 2).

	Variables	Construct		Definitions		
2	Paradigm/Philosophy	Component	Characteristics	Meaning	Options	References
3						
4	Philosophical skepticism	Ontology	No knowledge certainity, unknown reality,	The truth cannot be known.	The truth cannot be known.	Chakravartty, A. 2015, international journal for the study of skepticism 5 (2015) 73-79. Narboux, J.P 2014, international journal for the study of skepticism 4 (2014)
5				Nothing is known with certainity	Nothing is known with certainity	
6		Epistemology	Impossible to know or prove anything, Impossible to have knowledge, unreliable arguments, unattainable	Impossible to know or prove anything,	Impossible to know or prove anything,	
7				Impossible to have knowledge, unreliable arguments, evidence	Unreliable arguments and evidence	
8	Mathematical Platonism	Ontology	Independent, real mathematical world, mathematics exists independent of our thoughts	Real mathematical world exists	Real mathematical world exists	Encyclopedia Britannica (2018). Burov, A. 2017, Mathematical Platonism As a Necessity of Reason. Ramal, L. 2012, Platonism in Modern Mathematics
9				Mathematical world independent of human thoughts	Mathematical world independent of human thoughts	
10		Epistemology	Accessible through our intelligence, incompletly perceived, the mind is knowledge	Mathematical knowledge is accessible through our intelligence	Mathematical knowledge is accessible through our intelligence	
11				The mind provides and holds knowledge	The mind provides and holds knowledge	
12		Axilology	Value-free	Not infuenced by personal values	Not infuenced by personal values	
13	Pluralism (philosophy)	Ontology	Multiple reality, reality consists of many different substances	More than one version of the truth	More than one version of the truth	The basics of Philosophy, 2018. Shaheen, J. 2017, Explanatory Pluralism and Philosophy of Language: Explications and concepts. Robbins, J. 2013, Monism, pluralism, and the structure of
14				Many and differing world views and opinions	Many and differing world views and opinions	
15		Epistemology	Various or many means of approaching truths about the world, subjective	Various ways of approaching truths about the world	Various ways of approaching truths about the world	
16		Axilology	Many independent sources of value and that there is no single truth, even in moral matters.	Different sources of truth and values	Different sources of truth and values	
17				Differering views on morality	Differering views on morality	

Figure 2*:* Sample Multidimensional Philosophical Assumptions Summative Content Analysis

Following the presentation of research philosophies and paradigm and their associated meaning, the study carried out data annotation. Data annotation commonly involves including parts of speech (POS), or tags, which label each word in terms of its grammatical category (Reppen, 2010). Data tags are crucial for addressing data inputs from users of RPP model and helps to eliminate issues associated with searching for polysemous words (i.e. can be used as a modal verb or a noun) to disambiguate and focus the search results (see Figure 3) below for a sample screenshot of data annotation). The labels or annotated data are crucial for training data using the BoW (bag-of-words) and machine learning algorithm.

				Natural LanguageProcessing				
1								
2	**Variables**	**Nature**	**Meaning**	**Tokenizing**	**PoS Tagging**	**Word Embeddings**	**Stemming**	**NER**
3	Absolute idealism	Single	Truth exists independent of us, whether we know it or not Physical world is only	'Truth', 'exists', 'independent', 'of', 'us,', 'whether', 'we',	Truth/NNP exists/NNS independent/JJ of/IN us/PRP whether/IN we/PRP know/VBP it/PRP or/CC	knowledge understanding skills practice information scientific teaching influence development process communication	Truth exist independ	Truth (exists independent/RL) of us, whether we know it or not Physical world is only an (appearance/BL) to our expression of mind/ (Only one/RL) reality or world view in a (well balanced manner/TR)
4	Absolute idealism		Knowledge can be seen as mental or spiritual in nature Knowledge can be	'Knowledge', 'can', 'be', 'seen', 'as', 'mental', 'or', 'spiritual',	Knowledge/NNP can/MD be/VB seen/VBN as/IN mental/JJ or/CC spiritual/JJ in/IN nature/NN		Knowledg can be obtain through pure uniform	Knowledge can be seen as (mental or spiritual/VL) in nature Knowledge can be obtained through (pure uniform spiritual consciousness/MT) All views come together in a state of harmony
5	Absolute idealism		Values and morals are representation of the truth, not the truth itself	'Values', 'and', 'morals', 'are', 'representation', 'of', 'the', 'truth,',	Values/NNS and/CC morals/NNS are/VBP representation/NN of/IN the/DT truth/NN not/RB		Valu and moral are represent of the truth	Values and morals are (representation /AE)of the truth, not the truth itself
6	Absurdism	Unknown	The existence or non-existence of anything is meaningless A world of contradictions where	'The', 'existence', 'or', 'non-existence', 'of', 'anything', 'is', 'meaningless.',	NN The/DT existence/NN or/CC non/NN -/: existence/NN of/IN anything/NN is/VBZ meaningless/NN A/DT	meaningless however thus being therefore cause often both that whole sometimes	A world of contradict where noth is signific	The existence or non-existence of anything is (meaningless/RL) A world of contradictions where (nothing is significant/RL) Individuals and their (existence are not important/BL)
7	Absurdism		The world is full of contradictions and not easy to understand Impossible to obtain	'The', 'world', 'is', 'full', 'of', 'contradictions', 'and', 'not',	NN The/DT world/NN is/VBZ full/JJ of/IN contradictions/NNS and/CC not/RB easy/JJ to/TO		The world is full of contradict and not easi to	The world is full of contradictions and not easy to understand (Impossible to obtain full knowledge/SC) in an unreasonable world Better to rely on one's (experience/MT) than phenomena
8	Absurdism		No hereafter, or life after death Nothing needs to be taken seriously	'No', 'hereafter,', 'or', 'life', 'after', 'death.', 'Nothing',	NN No/DT hereafter/NN or/CC life/NN after/IN death/NN Nothing/NN needs/VBZ to/TO be/VB		No hereaft , or life after death	(No hereafter/ET), or life after death (Nothing needs to be taken seriously/AE)
9	Accidentalism (philosophy)	Multiple	Things have only accidental properties, no essential properties, or no common nature.	'Things', 'have', 'only', 'accidental', 'properties,',	NN Things/NNS have/VBP only/RB accidental/JJ properties/NNS no/DT essential/JJ properties/NNS	properties values current system specific levels	All event and properti are accident	Things have only (accidental properties/TR), (no essential properties/TR), or no common nature. All events and properties are (accidental/BL). The occurrence of some events is either (not necessitated/RL) or (not

Figure 3: Example of Data Annotation Using NLP Methodology Application to RPP

3.3 RPP Model Architecture

The architecture of the RPP model consists of four major components, namely: input data (client side), RPP data from literature or context data (database on the server), bag of words (BoW) and ML algorithm (server), and output data detailed in the corresponding sections (accessed from client side, retrieved from the database). Figure 4 provides an understanding of the behavior of the system and how it performs the classification of a researcher's sentiment into research philosophy and paradigm categories.

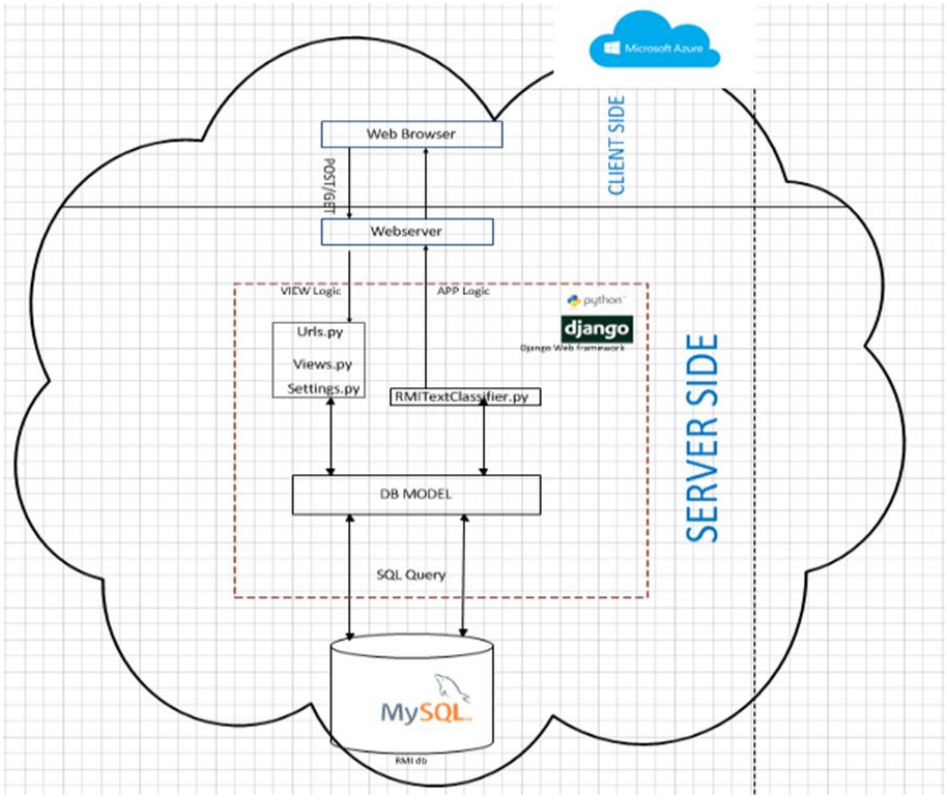


Figure 4: Research Methods Index NLP Architecture Diagram

The NLP application is hosted on the Microsoft Azure public cloud computing platform, as shown in Figure 4, for ease of access from multiple locations. The study adhered to the architectural style referred to as Representational State Transfer (RESTful) client-server design in deploying the NLP application. Python was used to write NLP code that is used for the classification of input. The Django Web framework was used for constructing the NLP web API, and was also used for the passing of data between the user interface and the back-end. The Django Web framework was also used for the development of the user interface. The Django Web framework supports the Model View Template (MVT) pattern, where the developer provides the model and Django uses the template and views to map the model to the Uniform Resource Locator (URL). The Django Web framework then renders the URL to the user interface.

The MySQL database is used to store information on research philosophies and paradigms, user account information, user input, system roles and reports. A user will be provided with a link to the system which will allow them to register an account, answer the questionnaire and then view their report.

3.4 Input Data (Client or User Side)

Input data for the NLP RPP model is derived from users' response to questions that represents the different variables of the multidimensional philosophical assumptions within the epistemological, ontological, and axiological lenses or framework, as offered by Saunders et al. (2016) and Guba and Lincoln (1994). The users' responses describe the specific worldview needed to run the RPP model. For example, the following questions are representative of the philosophical structure of the multidimensional assumptions by former scholars and are asked to gather inputs:

1. Questions based on the premise of ontology. The premise of these questions is to explore the user's ontological stance.

- How many versions of the truth can there be in a given situation?
- How can truth be influenced?

2. Questions based on the premise of epistemology. The premise of these questions is to explore the user's epistemological stance.

- How is knowledge acquired, that is, how do we know what we know?
- What influences what we know?
- How many sources of knowledge are there?
- *How can knowledge be advanced?*

3. Questions based on the premise of axiology. The premise of these questions is to explore the user's axiological stance.

- What is the importance of values and ethics?
- How can personal values influence the truth?
- What determines our values and ethics?

The user enters input data through the web browser that resides on their specific personal computer that is connected to the internet through an ISP or other means. The inputs are crucial for tokenisation and for performing the BoW and machine learning algorithm against the list of common words in the database, and to possibly identify the words or emerging worldviews that fall outside the purview of the database. A researcher uses the web browser to connect to the RMI application. The web browser is linked to the webserver through a POST/GET method.

3.5 RPP Context Data From Literature (Database on the Server)

The context data was derived from the 180 usable items of information on the different types of research philosophies collected from the various sources of literature, as previously discussed. The annotated context data is presented in the RPP model following the NLP methodology or procedure (see Figure 3). Appendix 1 provides an example of how data pertaining to each philosophy was collected and corroborated from different sources for conformability and credibility. The different sources helped not only with the study's credibility and conformability, but also provided coverage of the different emphases related to the phenomenon that was necessary to produce a rich and adequate volume of text for algorithm performance purposes. Further, corroboration of the different emphases of all the philosophies was essential for drawing more reliable and meaningful conclusions about each philosophy.

The strength of a small targeted database or (corpora) for the investigation of special uses, such as the RPP model, has been greatly discussed by Reppen (2010). The MySQL database stores all the RPP's data, participant data, information and responses which are then used to generate a report recommending research philosophy and paradigm categories to participants. The communication between the client and the webserver is through the Hypertext Transfer Protocol (HTTP). The function of the server is mainly to store, process ML algorithm discussed in the next section, and deliver web content (html pages) as requested by the client.

3.6 Classification Model: Bag of Words and Machine Learning Algorithm (Server)

The data that was sourced from journals, encyclopaedias, books, and so forth, and stored in the RPP's corpus, was split into train-test sets (at 90-10 ratio) for the purpose of training and testing the classification algorithms. According to Sebastiani (2002), ML algorithms cannot process text data directly, therefore the corpus data needed to be represented in a numerical form. The Bag-of-Words (BoW) model was used to pre-process and represent the RPP's corpus data in numeric vectors. This model is based on a vocabulary and measure of the frequency of all known words in a corpus (Aggarwal and Zhai, 2012). The Term Frequency – Inverse Document Frequency, or TF-IDF, was then used to rescale word frequencies by computing how many times the words appear in each of the RPPs, with the most frequently appearing words, such as 'there and that', being ignored.

This then produced a final score which was used as a weight for each of the RPPs. These weights or frequencies of known words in a category are considered as features to be used as equivalent fixed-length numerical vectors (as shown in Table 1). The numerical vectors or features are used to train classifiers such as the naïve Bayes, SVM and Logistic Regression to determine which one produces the highest level of confidence in classifying text into different RPP categories. The example of BoW model presented in Table 1 is also used to represent input text in numeric vectors to be used to determine the RPP category it belongs to.

Table 1: Vector Representation of RPPs in BoW

	World	Compose	Fundamental	substance	External	nature	kind	Independent	term	Language	Object	exist	function	relate
RPP 1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
RPP 2	0	0	0	0	1	1	1	1	1	1	0	0	0	0
RPP 3	1	0	0	0	0	0	0	0	0	0	1	1	1	1

RPP1: The world is composed of two fundamental substances

RPP2: External natural kind terms independent of language

RPP3: Objects exist and function only as relational in the world

Once the user has completed the questionnaire, the input is concatenated into a string that undergoes pre-processing before being vectorised. A comparison of the features derived from user input against the corpus category features yields the three topmost RPPs that are closely linked to the input, which in turn, represents a researcher's worldview. Furthermore, the NLP is linked to a lexical database, such as Wordnet, to find synonyms for other complex philosophical terms, and the emerging constructs necessary for deep learning. The output results are presented in a chart showing the degree to which a researcher is aligned with a particular RPP.

The server side refers to the NLP application that processes input, classifies it into RPP categories and then issues a report for the participant. This report is based on the information request from the MySQL database. The study ensures the validation of annotation or labels using supervised learning which helps with label data, or semi-structured supervised learning (taking the labels we have to leverage on data without labels). The latter labels are used to infer labels to emerging data that is outside the purview of the database.

On running the Naïve Bayes, SVM and LR from the same dataset represented by the BoW model it can be observed that the Naïve Bayes outperforms all the other algorithms (see Table 2). This is due to the Naïve Bayes' ability to be linearly scalable with the data points and the number of predictor features (Keogh, 2015; Sebastiani, 2002) and its ability to provide a probabilistic explanation for classifications based on the Bayes theorem.

Table 2: Accuracy and Precision of NLP Algorithms

Algorithm	Accuracy	Precision rate
Naïve Bayes	70%	85%
SVM	76%	78%
LR	77%	80%

Further, the Naïve Bayes algorithm is suitable for this study due to (1) its ability to learn, even if there are small amounts of training data, thereby significantly reducing the time it takes to train, (2) it yields parameter estimates that are good, and (3) with independent predictors it performs better than the discriminative LR and SVM algorithms (Keogh, 2015; Sebasitani, 2002; Uddin et al., 2012).

As for accuracy levels, Gaizauskas and Wilks (1998) attest that the precision of the information extraction (IE) system ranges around 50%. The natural language component of the RMI reported in this study produces a 70% accuracy and 85% precision level, which lends and builds support for the previous study by Gaizauskas and Wilks (1998).

In terms of confidence level, Gandrabur et al. (2006) argue that the confidence level related to NLP is subject to data, purpose of the study, and the variations in language. According to Gaizauskas and Wilks (1998), IE is as a subset of knowledge discovery and data mining research that extracts useful aspects of textual information from natural language texts, such as the one provided by users in the study. Confidence levels are further reflected based on precision, recall, and F1-score, as shown in the comparative view of different algorithms in Figure 5.

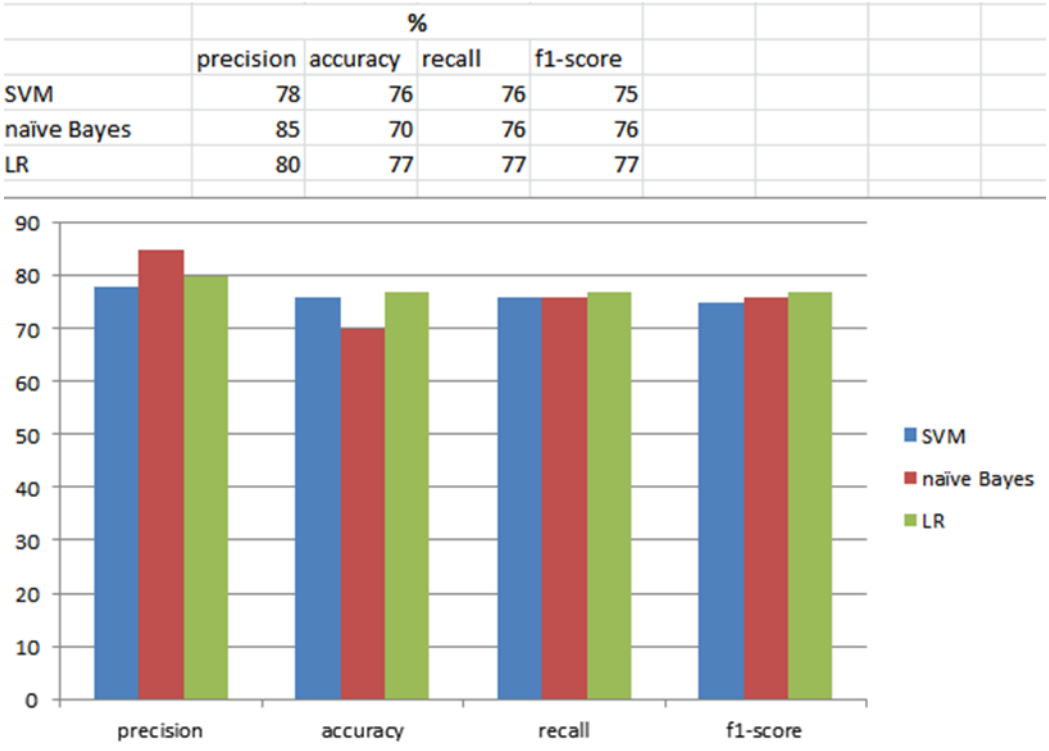


Figure 5: Comparative View of the Algorithms

3.7 Model Assumptions

The RPP model is not without assumptions. The acknowledgement of assumptions is a practice greatly endorsed by Seddon and Scheepers (2012) who argue for boundary conditions beyond which the findings of the study might not apply. In this study, there is an assumption in the analysis that the respondents had the same level of

cognition, use of English and expertise in research philosophies. There is also an inherent risk in text analysis because the same phrase can mean different things in different cultures. Put differently, text analysis is subject to imbalance and linguistic variations issues.

In addition, the considered research philosophies are (at this point of the model) limited relatively to the contextual data in the RPP database. Put differently, the RPP model at the time of writing this paper has not identified emerging or indigenous worldviews outside the database, but rather calculates the users' Bow against the stored contextual data. This is mainly because although the system is built to recognise emerging concepts beyond the database, the system only considers words outside the purview of the RPP model as an emerging philosophy, if there is a theme or high pattern from more people. In its current form, the model only highlights words that might be new to the RPP database.

The RPP model relies mainly on text, and excludes pictures, audio, video, social media and any other form of data. The selection of feature vectors based on text reliance or keywords limits the classification to text extracted from the users' inputs, thus limiting an understanding of the relationship between words. The latter is mitigated, to a certain extent, by the link between the RMI model and WordNet to try and explore the relationships between words. The domain focus of NLP to research philosophies limits its scalability.

3.8 Output Data (Accessed From Client Side, Retrieved From the Database)

This section presents the output of the realistic context of the RPP model application using NLP, the technological techniques used to present a way of making the teaching and learning of research philosophies and paradigms easier. The system was tested firstly, for performance against its ability to capture a user's exact inputs regarding his/her worldviews to align, correlate with, identify and reveal specific philosophies espoused by the user, against those collected from literature in the RPP database. Secondly, to demonstrate percentages and areas where a user's worldview correlates with more than one research philosophy. Lastly, for precision and accuracy.

Natural language processing is the most important indicator, as it gives the user freedom of expression through open-ended questions. Using the exact user's expression, the NLP processes the unstructured data using an algorithm based on the methodology previously discussed in the section on the NLP process (of tokenising, lemmetisation, and named entity recognition). Using ML algorithm, the NLP model dives deep to correlate the user's worldview beyond the preferential philosophical approaches of supervisors, subject, and field of study. By so doing, NLP provides an interesting approach to resolving the deep ideological approaches which result in polarisation.

The output data presents the following four categories: details of the user's customised worldview or research philosophies, a presentation of how the user's inputs were tokenised, lemmatised, and the name recognition entity process.

In the first instance, the output data reveals the type of variations in the philosophical assumptions that the user espouses within a cluster or clusters. Obviously, in the first instance, the observation is linked to the cluster output(s) of the NLP. For example, the pie chart in Figure 6 shows that the user's philosophical stance is skewed more to the realism cluster (i.e. cluster with different philosophies with assumptions closer to realism's assumptions).

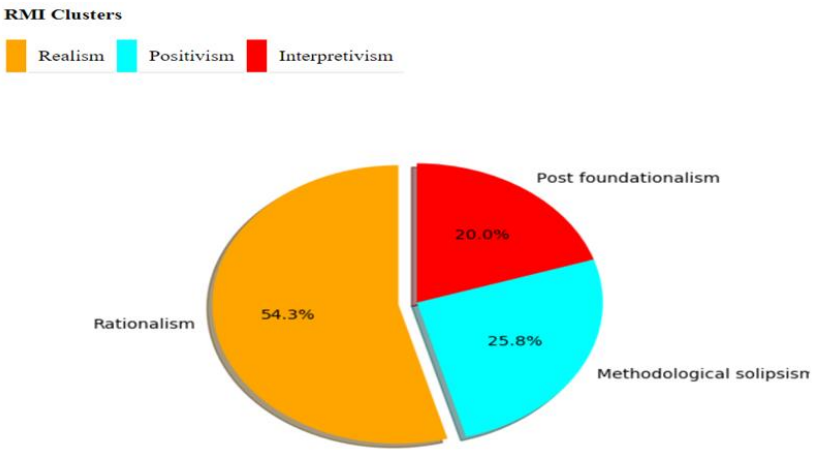
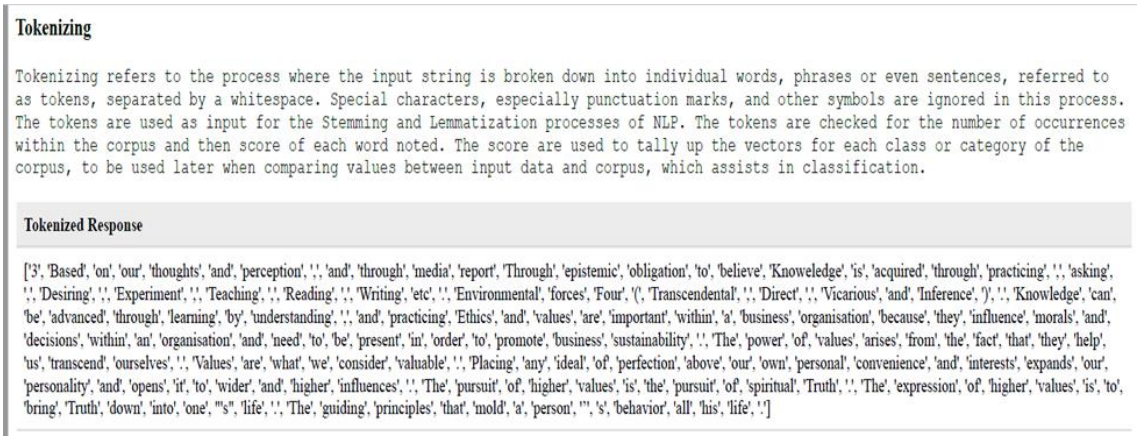


Figure 6: NLP Output Results

Most importantly, the pie chart reveals the specific philosophical assumption within the realist cluster with clear percentages of the most aligned philosophy. Notably, the user's expression reveals an alignment to philosophies in multiple clusters (i.e. realism, interpretivism, and positivists clusters), which indicates that a user possesses multiple philosophical assumptions.

Beyond the revelation of the cluster and actual philosophy(ies), the NLP results also present a transparent process of how the user's unstructured data was tokenised and lemmetised, across different name entities. For example, Figures 7, 8 and 9 below show how the user's unstructured data was broken down into different tokens, lemmetised, and the name entity recognition (NER) patterns. Of particular importance about NER, is that it reveals details of the actual philosophies with hyperlinks that crawl to the database of research philosophies and paradigm literature that provides the user with more information about each correlated philosophy.



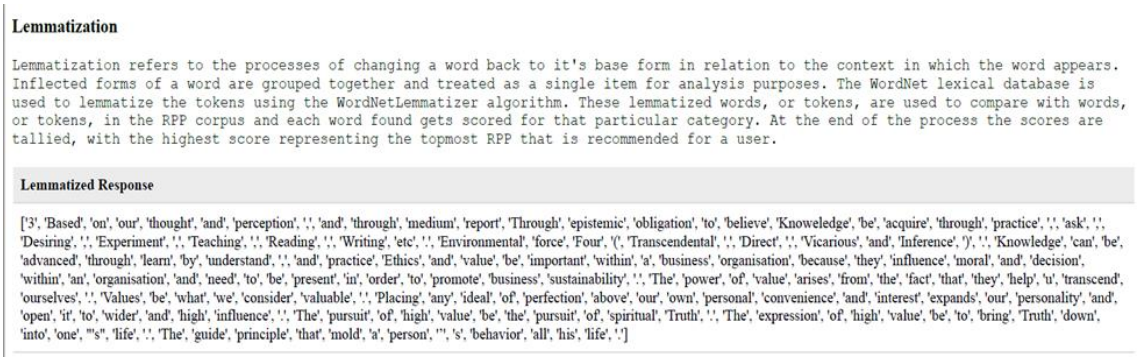
Tokenizing

Tokenizing refers to the process where the input string is broken down into individual words, phrases or even sentences, referred to as tokens, separated by a whitespace. Special characters, especially punctuation marks, and other symbols are ignored in this process. The tokens are used as input for the Stemming and Lemmatization processes of NLP. The tokens are checked for the number of occurrences within the corpus and then score of each word noted. The score are used to tally up the vectors for each class or category of the corpus, to be used later when comparing values between input data and corpus, which assists in classification.

Tokenized Response

['3', 'Based', 'on', 'our', 'thoughts', 'and', 'perception', ',', 'and', 'through', 'media', 'report', 'Through', 'epistemic', 'obligation', 'to', 'believe', 'Knoweledge', 'is', 'acquired', 'through', 'practicing', ',', 'asking', ',', 'Desiring', ',', 'Experiment', ',', 'Teaching', ',', 'Reading', ',', 'Writing', 'etc', '.', 'Environmental', 'forces', 'Four', '(', 'Transcendental', ',', 'Direct', ',', 'Vicarious', 'and', 'Inference', ')', '.', 'Knowledge', 'can', 'be', 'advanced', 'through', 'learning', 'by', 'understanding', ',', 'and', 'practicing', 'Ethics', 'and', 'values', 'are', 'important', 'within', 'a', 'business', 'organisation', 'because', 'they', 'influence', 'morals', 'and', 'decisions', 'within', 'an', 'organisation', 'and', 'need', 'to', 'be', 'present', 'in', 'order', 'to', 'promote', 'business', 'sustainability', '.', 'The', 'power', 'of', 'values', 'arises', 'from', 'the', 'fact', 'that', 'they', 'help', 'us', 'transcend', 'ourselves', '.', 'Values', 'are', 'what', 'we', 'consider', 'valuable', '.', 'Placing', 'any', 'ideal', 'of', 'perfection', 'above', 'our', 'own', 'personal', 'convenience', 'and', 'interests', 'expands', 'our', 'personality', 'and', 'opens', 'it', 'to', 'wider', 'and', 'higher', 'influences', '.', 'The', 'pursuit', 'of', 'higher', 'values', 'is', 'the', 'pursuit', 'of', 'spiritual', 'Truth', '.', 'The', 'expression', 'of', 'higher', 'values', 'is', 'to', 'bring', 'Truth', 'down', 'into', 'one', "'s", 'life', '.', 'The', 'guiding', 'principles', 'that', 'mold', 'a', 'person', '', 's', 'behavior', 'all', 'his', 'life', '.']

Figure 7: Tokenisation of the User's Unstructured Data



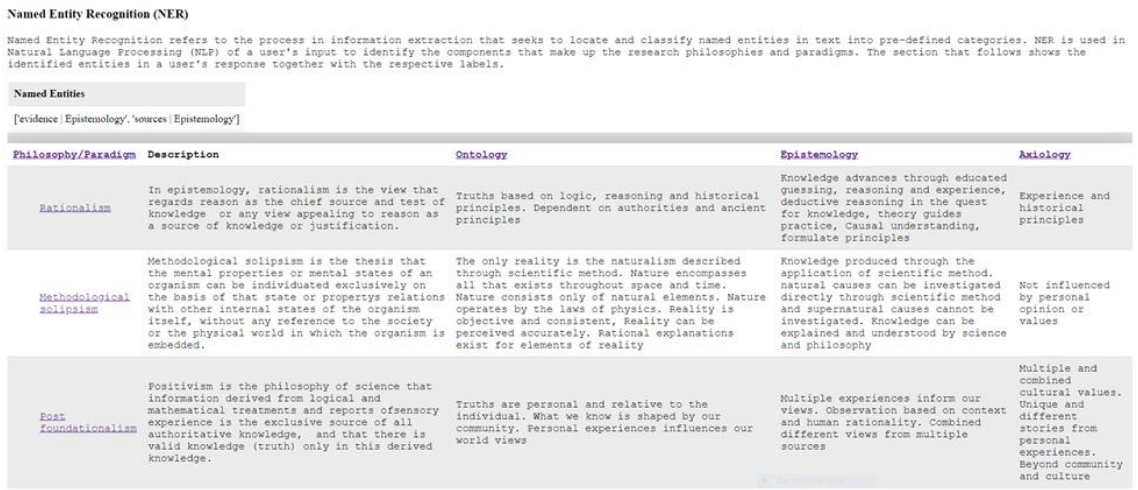
Lemmatization

Lemmatization refers to the processes of changing a word back to it's base form in relation to the context in which the word appears. Inflected forms of a word are grouped together and treated as a single item for analysis purposes. The WordNet lexical database is used to lemmatize the tokens using the WordNetLemmatizer algorithm. These lemmatized words, or tokens, are used to compare with words, or tokens, in the RPP corpus and each word found gets scored for that particular category. At the end of the process the scores are tallied, with the highest score representing the topmost RPP that is recommended for a user.

Lemmatized Response

['3', 'Based', 'on', 'our', 'thought', 'and', 'perception', ',', 'and', 'through', 'medium', 'report', 'Through', 'epistemic', 'obligation', 'to', 'believe', 'Knoweledge', 'be', 'acquire', 'through', 'practice', ',', 'ask', ',', 'Desiring', ',', 'Experiment', ',', 'Teaching', ',', 'Reading', ',', 'Writing', 'etc', '.', 'Environmental', 'force', 'Four', '(', 'Transcendental', ',', 'Direct', ',', 'Vicarious', 'and', 'Inference', ')', '.', 'Knowledge', 'can', 'be', 'advanced', 'through', 'learn', 'by', 'understand', ',', 'and', 'practice', 'Ethics', 'and', 'value', 'be', 'important', 'within', 'a', 'business', 'organisation', 'because', 'they', 'influence', 'moral', 'and', 'decision', 'within', 'an', 'organisation', 'and', 'need', 'to', 'be', 'present', 'in', 'order', 'to', 'promote', 'business', 'sustainability', '.', 'The', 'power', 'of', 'value', 'arises', 'from', 'the', 'fact', 'that', 'they', 'help', 'u', 'transcend', 'ourselves', '.', 'Values', 'be', 'what', 'we', 'consider', 'valuable', '.', 'Placing', 'any', 'ideal', 'of', 'perfection', 'above', 'our', 'own', 'personal', 'convenience', 'and', 'interest', 'expands', 'our', 'personality', 'and', 'open', 'it', 'to', 'wider', 'and', 'high', 'influence', '.', 'The', 'pursuit', 'of', 'high', 'value', 'be', 'the', 'pursuit', 'of', 'spiritual', 'Truth', '.', 'The', 'expression', 'of', 'high', 'value', 'be', 'to', 'bring', 'Truth', 'down', 'into', 'one', "'s", 'life', '.', 'The', 'guide', 'principle', 'that', 'mold', 'a', 'person', '', 's', 'behavior', 'all', 'his', 'life', '.']

Figure 8: User's Unstructured Data Using



Named Entity Recognition (NER)

Named Entity Recognition refers to the process in information extraction that seeks to locate and classify named entities in text into pre-defined categories. NER is used in Natural Language Processing (NLP) of a user's input to identify the components that make up the research philosophies and paradigms. The section that follows shows the identified entities in a user's response together with the respective labels.

Named Entities

['evidence | Epistemology', 'sources | Epistemology']

Philosophy/Paradigm	Description	Ontology	Epistemology	Axiology
Rationalism	In epistemology, rationalism is the view that regards reason as the chief source and test of knowledge or any view appealing to reason as a source of knowledge or justification.	Truths based on logic, reasoning and historical principles. Dependent on authorities and ancient principles	Knowledge advances through educated guessing, reasoning and experience, deductive reasoning in the quest for knowledge, theory guides practice, Causal understanding, formulate principles	Experience and historical principles
Methodological solipsism	Methodological solipsism is the thesis that the mental properties or mental states of an organism can be individuated exclusively on the basis of that state or propertys relations with other internal states of the organism itself, without any reference to the society or the physical world in which the organism is embedded.	The only reality is the naturalism described through scientific method. Nature encompasses all that exists throughout space and time. Nature consists only of natural elements. Nature operates by the laws of physics. Reality is objective and consistent, Reality can be perceived accurately. Rational explanations exist for elements of reality	Knowledge produced through the application of scientific method. natural causes can be investigated directly through scientific method and supernatural causes cannot be investigated. Knowledge can be explained and understood by science and philosophy	Not influenced by personal opinion or values
Post foundationalism	Positivism is the philosophy of science that information derived from logical and mathematical treatments and reports ofsensory experience is the exclusive source of all authoritative knowledge, and that there is valid knowledge (truth) only in this derived knowledge.	Truths are personal and relative to the individual. What we know is shaped by our community. Personal experiences influences our world views	Multiple experiences inform our views. Observation based on context and human rationality. Combined different views from multiple sources	Multiple and combined cultural values. Unique and different stories from personal experiences. Beyond community and culture

Figure 9: NER of User's Philosophical Assumption

4. Conclusion and Recommendations

This study offers a progressive technological approach towards understanding research paradigms and philosophies. In particular, the study uses NLP and machine learning to introduce users to a wealth of research paradigms and philosophies that are aligned to their worldview, rather than those that are most familiar to their teacher or supervisor. The natural language processing and machine learning inspired digital model poses questions that place students in a reflexive mode and draws their articulated responses as inputs that model their worldviews against a host of philosophies in the database. The inquiry-based teaching approach transforms learning from the generic push-teaching method that assumes universality, to the fostering of a reflexive approach that helps resolve the deep ideological approaches that caused the polarisation. The polarisation and dichotomies that Silverman (2010) and Acheampong et al. (2015) identified in research methods may be largely eradicated, leading to a new discovery of deep learning, and an awareness and understanding of the concepts that can boost the development of research students and knowledge production.

Beyond providing insight into how technology can complement the teaching and learning of a complex subject that is characterised by high variety, the findings of this study contribute to the evoking of deep learning arising from new philosophies and methods. The application of NLP to the subject of research philosophies extends its relevance beyond the subjects of marketing, aviation, and customer relationship management that were quite evident in literature.

On consideration of the rate of automation in the fourth industrial revolution, and how technology permeates every industry, it is worthwhile applying the novelty of NLP to investigate research philosophies for teaching and learning research philosophies. At best, the NLP's insight into knowledge production can add to the discussion of how knowledge production can potentially be transformed. The latter elevates scientific knowledge beyond what might normally be expected of library of research methods, and provides an alternative direct response to Davison and Martisons' (2016) call for new ways of developing theories. It also reveals how NLP can help to facilitate a responsive interactive teaching and learning process that has the potential to improve understanding of the concepts.

Furthermore, this study revealed developments that are discordant with previous scholars who advocated for a single philosophical assumption for a field, subject, or researcher, such as the existence of pure positivist by Mach, and/ or pure interpretivist (Banks, 2013). Whether the discovery of users' multiple philosophical assumptions will help to curb the parallels, debates, and dilemmas documented by previous scholars, remains an interesting question (Mkansi and Acheampong, 2012; Scotland, 2012; Sefotho, 2015). Future research can engage scholars with extreme right or left philosophical assumptions to use the research method index (RMI) system to assess whether they are who they say they are. And if not, what can change moving forward?

The digital discovery of research philosophies and paradigms extend the work of previous theorists to the technologically-inspired discovery of episteme, ontology, and axiology. By its nature, the use of NLP becomes an advanced channel to how we know what we know, and what is the nature of the reality and values being displayed? Further studies could formulate several hypotheses and theorems that arise from technological foundations. It further reinforces the assertions of Habermas (1972) and James and Vinnicombe (2002) that commonly emphasise the unneutrality of knowledge, and how every researcher has deeply embedded worldviews that differently shape knowledge production and research methods.

Moreover, the findings present an attempt towards reducing the gulf between theory and practice, especially the lack of knowledge and awareness of the concepts. Future studies can explore whether the technological discovery inspires the application and acknowledgement of the concepts in major knowledge outlets, such as journals with high impact factors and by students in different fields.

Lastly, the study provides an alternative method of analysing and cross-validating data through the demonstration of NLP to shine some light on Aram and Salipanter's (2003) call for new communities of relevance and rigorous knowledge production. The NLP approach reveals how scholars can apply various methods to data while employing different research paradigms that yield lean insights for mass customisation in teaching and learning.

The study is not without limitation, as its first demonstration of NLP does not measure the impact of the software on improving users' understanding of the concepts and how it can shape research methods. Hence, the recommendation for interested parties (academics and post-graduate students) to explore the second and third phase of the RMI project which deals with determining how the system can improve understanding of the concepts, how research philosophies shape research methods, and help generate knowledge. Future studies can

use the first phase of the software project (www.rmi.unisa.ac.za) to conduct social experiments and report on the effects and impact of the software on their teaching and learning of research philosophies and paradigms at third-year (undergraduate) or post-graduate levels.

Reference

Acheampong, E, Mkansi, M, Kondadi, K, and Qi, B (2015) Polarization in research methods. In J. Mendy, & S. D. Geringer (Eds.), *Leading Issue in Business Research Methods Volume 2*. Academic Conferences and Publishing International Limited.

Acheampong, E, Mkansi, M, Kondadi, KR, and Qi, B (2012) Polarization in research methods application: Examining the examiner. *11th European conference on Research Methodology for Business and Management Studies*. Bolton, UK.

Aggarwal, CC, and Zhai, C (2012). *Mining Text Data*. Springer.

Alvesson, M, and Skoldberg, K (2009) *Reflexive methodology: new vistas for qualitative research*. Sage.

Aram, JD, and Salipanter, PF Jr (2003) Bridging scholarship in management: Epistemological reflections. *British Journal of Management, 14*(3), 189-205.

Arbuckle, JL (2008) *AmosTM 17.0 User's Guide*. Available at http://www.jou.ufl.edu/research/lab/pdf/Amos-17.0-User's-Guide.pdf (accessed 20 June 2011)

Banks, EC (2013) Metaphysics for positivists: Mach versus the Vienna Circle. *Discipline Filosophiche, 23*, 57-77.

Bertoni, M, Rondini, A, and Pezzotta, G (2017) A systematic review of value metrics for PSS design. *Procedia CIRP*, *64,* 289-294.

Blaikie, N (1993) *Approaches to social enquiry*. Polity Press.

Blaikie, N (2000) *Designing social research*. Polity Press.

Bonaccorso, G (2017) *Machine learning algorithms: Reference guide for popular algorithms for data science and machine learning.* Packt.

Brownlee, J (2018) *Supervised and Unsupervised Machine Learning Algorithms*. Available at https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/ (accessed 30 January 2020)

Brunton, SL, Noack, B R, and Koumoutsakos, P (2019) Machine learning for fluid mechanics. *Annual Reviews of Fluid Mechanics*, *52,* 477-508.

Carpenter, S (2019) Contemporary digital systems. In M. Mkansi, N. McLennan, and G. de Villiers (Eds.), *Contemporary Issues in Operations and Supply Chain Management* (pp. 146-175). Van Schaik.

Chang, T, DeJonckheere, M, Vydiswaran, VGV, Li, J, Buis, LR, Guetterman, TC (2021) Accelerating Mixed Methods Research with Natural Language Processing of Big Text Data. *Journal of Mixed Methods Research*, 15(3), 398-412.

Chen, Y, Wu, X, Hu, A, He, G, Ju, G (2021) Social Prediction: a new research paradigm based on machine learning. *The Journal of Chinese Sociology*, 8(1), 1-21.

Creswell, JW (2013) *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage.

Creswell, JW (2014) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). Sage.

Chang, AX, Spitkovsky, VI, Manning, CD, and Agirre, E (2016) *A comparison of named-entity disambiguation and word sense disambiguation.* Available at https://nlp.stanford.edu/pubs/chang2016entity.pdf (accessed 30 January 2020)

Daniel, B K (2018) Empirical verification of the "TACT" framework for teaching rigour in qualitative research methodology. *Qualitative Research*. Advance online publication.https://doi.org/10.1108/QRJ-D-17-00012.

Davison, R M, and Martinsons, M. G (2016) Context is king! Considering particularism in research design and reporting. *Journal of Information Technology*, *31*(3), 241-249.

Denyer, D, and Tranfield, D (2009) Producing a systematic review. In D. Buchanan, & A. Bryman (Eds.), *The Sage Handbook of Organizational Research Methods* (pp. 671–689). Sage Publications.

Denzin, N, and Lincoln, Y (2003) The discipline and practice of qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (2nd ed.). Sage.

Derczynski, L, Maynard, D, Rizzo, G, van Erp, M, Gorrell, G, Troncy, R, Petrak. J. & Bontcheva, K (2015). Analysis of named entity recognition and linking for tweets. *Information Processing and Management, 51*(2), 32-49.

Faherty, VE (2010) *Wordcraft: Applied Qualitative Data Analysis (QDA) Tools for Public and Voluntary Services*. Sage Publications.

Field, AP (2009) *Discovering Statistics Using SPSS*. Sage Publications.

Groeesler, A (2017) Teaching Research Methods: An Occasional Paper. *Institute of Teaching and Learning Innovation (ITaLi)*. The University of Queensland.

Gaizauskas, R, and Wilks, Y (1998) Information Extraction: beyond document retrieval, *Journal of Documentation*, 54(1), 70-105.

Gandrabur, S, Foster, G, and Lapalme, G (2006) Confidence estimation for NLP applications. *ACM Transactions on Speech and Language Processing, 3*(3), 1-29.

Guarino, N (1998) Formal Ontology in Information Systems. Paper published in *Proceedings of FOIS'98*, Trento, Italy, 6–8 June 1998 (pp. 3-15). Amsterdam: IOS Press.

Guba, EG, and Lincoln, YS (1994) Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 105–117). Sage.

Habermas, J (1972) *Knowledge and human interest*. Beacon Press.

Hesse-Biber, S (2015) The problems and prospects in the teaching of mixed methods research. *International Journal of Social Research Methodology*, *15*(5), 463-477.

Hsieh, H F, and Shannon, SE (2005) Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288.

James, K, and Vinnicombe, S (2002) Acknowledging the individual in the researcher. In D. Partington (Ed.), *Essential skills for management research* (pp. 84-98). Sage.

Javed, S, Adewumi, TP, Liwicki, FS, Liwicki, M (2021) Understanding the Role of Objectivity in Machine Learning and Research Evaluation. *Philosophies*, 6(22), 1-8.

Jensen, HS (2000) A history of the concept of knowledge. *Zagreb International Review of Economics and Business*, *3*(2), 1-16.

Jiang, J (2012) Mining text data: Information extraction from text. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 11-36). Springer.

Kaminski, J, Jiang, Y, Piller, F, and Hopp, C (2017) Do user entrepreneurs speak different? Applying natural language processing to crowdfunding videos. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2683-89). Denver.

Kanchan, N, and Yadav, PR (2022) Realisation of natural language processing and machine learning approaches for text-based sentiment analysis. *Expert Systesms*, DOI: 10.1111/exsy.13114.

Kelleher, J D, Namee, B M, and D'Arcy, A (2015) *Fundamentals of Machine Learning for Predictive Data Analytics.* Massachusetts Institute of Technology.

Keogh, E (2015) *Naïve Bayes classifier.* Available at: http://www.cs.ucr.edu/~eamonn/CE/Bayesian%20Classification%20withInsect_examples.pdf (accessed 01 October 2020)

Killam, L (2013) *Research terminology simplified: Paradigms, axiology, ontology, epistemology and methodology.* World Bank Publishers.

Kumar, S, and Zymbler, MA (2019) A machine learning approach to analyse customer satisfaction from airline tweets. *Journal of Big Data,* 6(62), 1-16.

Leeson, W, Resnick, A, Alexander, D, and Rovers, J (2019) Natural Language Processing (NLP) in qualitative public health research: A proof of concept study. *International Journal of Qualitative Methods,* 18, 1-9.

Li, Q, and Sha, S (2017) Learning stock market sentiment lexicon and sentiment-oriented word vector from StockTwits. *Proceedings of the 21st Conference on Computation Natural Language Learning* (pp. 301-310). Vancouver, Canada. Association for Computational Linguistics.

Lincoln, Y (1990) The making of a constructivist: A remembrance of transformations past. In E. Guba (Ed.), *The Paradigm dialog* (pp. 67-87). Sage Publications.

Loh, WY (2011) Classification and Regression Trees. *Wiley Interdisciplinary Review, Data Mining and Knowledge* Disc, 1, (pp. 14-23).

Mackenzie, N, and Knipe, S (2006) Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research*, *16*(2), 193-205.

Makombe, G (2017) An expose of the relationship between paradigm, method and design in research. *The Qualitative Report*, *22*(12), 3363-3382.

Marr, B (2016) *What is the difference between artificial intelligence and machine learning?* Available at https://www.google.co.za/amp/s/www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/amp/ (accessed 30 January 2020)

Miller, PK, and Grimwood, T (2015) Mountains, cones, and dilemmas of context: The case of "ordinary language" in philosophy and scientific method. *Philosophy of the Social Sciences*, *45*(3), 331-355.

Mkansi, M (2020) Research methods index. In D. Remenyi (Ed.), *Innovation in Teaching of Research Methodology Excellence Awards: An Anthology of Case Histories 2020*. Academic Conferences and Publishing International Limited.

Mkansi, M, and Acheampong, EA (2012) Research philosophical debates and classifications: Students' dilemma. *Electronic Journal of Business Research Methods*, *10*(2), 132-140.

Mkansi, M (2018) Research paradigm and philosophies swept under the carpet: A summative content analysis. *17th European Conference on Research Methods in Business and Management*, 12-13 July, Universita Roma Tre, Rome, Italy. Academic Conferences and publishing limited.

Muhaise, H, EJiri, AH, Muwanga-Zake, JWF, and Kareyo, M (2020) The Research Philosophy Dilemma for Postgraduate Student Researchers. *International Journal of Research and Scientific Innovation* Vol 7, No 4, pp. 2321–2705.

Pedrycz, W, and Chen, S (2016) *Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence*. Springer International Publishing.

Peffers, K, Tuunaneh, T, and Niehaves, B (2018) Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. *European Journal of Information Systems*, *27*(2), 129-139.

Polit, D, and Beck, C T (2008) *Nursing research: Generating and assessing evidence for nursing practice*. Lippincott Williams and Wilkins.

Reppen, R (2010) Building a corpus: What are the key considerations? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics.* Routledge Handbooks.

Saillard, KE (2011) Systems Versus Interpretive with Two CAQDAS Packages: Nvivo and MAXQDA. *Forum Qualitative Social Research*, *12*(1), Art. 34.
Saunders, M, Lewis, P, and Thornhill, A (2016) *Research methods for business students* (5th ed.). Pearson Education.
Saunders, M, Lewis, P, and Thornhill, A (2019) *Research methods for business students.* Pearson Education.
Sebastiani, F (2002) Machine learning in automated text categorization. *ACM Computing Surveys*, 34, 1–47.
Scotland, J (2012) Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *English Language Teaching*, *5*(9), 9–16.
Seddon, P, and Scheepers, R (2012) Towards the improved treatment of generalisation of knowledge claims in IS research: Drawing general conclusions from samples. *European Journal of Information Systems*, *21*(1), 6-21.
Sefotho, MM (2015) A researcher's dilemma: Philosophy in crafting dissertations and theses. *Journal of Social Science*, *42*(12), 23–36.
Shouval, R, Bondi, O, Mishan, H, Shimoni, A, Unger, R, and Nagler, A (2014) Application of machine learning algorithms for clinical predictive modeling: A data-mining approach in SCT. *Bone Marrow Transplantation*, *49*(3), 332-337.
Silverman, D (2010) *Doing qualitative research* (3rd ed.). Sage.
Steen, M, and Roberts, T (2011) *Handbook of midwifery research*. Wiley-Blackwell.
Te'eni, D, Rowe, F, Agerfalk, PJ, and Lee, JS (2015) Publishing and getting published in *EJIS*: marshalling contributions for a diversity of genres. *European Journal of Information Systems*, *24*(6), 559-568.
Waters, A, Grimaldi, P, Lan, A, Baraniuk, R (2017) Short-Answers Responses to STEM exercises: Measuring Response Validity and its Impact on Learning. *Proceedings of the 10th International Conference on Data Mining*, 374-375.
Yin, RK (2018) *Case study research and applications: Design and methods* (6th ed.). Sage.
Zeroual, I, and Lakhouaja, A (2018) Data science in light of Natural Language Processing: An overview. *Procedia Computer Science*, *127*(2018), 82-91.
Zhao, H, and Kit, C (2011) Integrating unsupervised and supervised word segmentation: The role of goodness measures. *Information Sciences*, *181*(1), 163-183.
Zhao, L, Alhoshan, W, Ferrari, A, Letsholo, KJ, Ajagbe, MA, Chioasca, E, Batista-Navarro, RT (2021) Natural Language Processing for Requirements Engineering: A systematic Mapping Study. ACM Comput. Surv, 54(3), 1-41.

Appendix 1: Corroboration of Research Philosophies Data

Corroboration of research philosophies data				**Data Sources**			
Paradigm/ philosophy	**Year**	**References**	**Title of data sources**	**Journals**	**Encyclopaedia**	**Books**	**Dissertations**
Infinitism	2018	Aikin, Scott.	Epistemic infinitism, 2018, doi:10.4324/0123456789-P077-1. Routledge *Encyclopedia of Philosophy*.		●		
	2013	Turri, J., & Klein, P.	Infinitism in epistemology	●			
Innatism	2010	Hill, J	The Synthesis of Empiricism and Innatism in Berkeley's Doctrine of Notions. *Berkeley Studies*. 21. 3–15.	●			
	2015	Winch C	Innatism, Concept Formation, Concept Mastery and Formal Education. *J Philos Educ* [Internet]	●			
Internalism	2017	Pappas, George	"Internalist vs. Externalist Conceptions of Epistemic Justification", *The Stanford Encyclopedia of Philosophy*		●		
	2017	Wilson, R.A.	Externalism and Internalism in the Philosophy of Mind			●	
Interpretivism	2015	Aliyu et al.	Positivist and Non-Positivist Paradigm in Social Science Research: Conflicting Paradigms or Perfect Partners?	●			

Corroboration of research philosophies data				**Data Sources**			
Paradigm/ philosophy	**Year**	**References**	**Title of data sources**	**Journals**	**Encyclopaedia**	**Books**	**Dissertations**
	2014	Vosloo, J.J	A Sport management programme for educator training in accordance with the diverse needs of South African Schools				●
	1967	Meyers	Peirce on Cartesian Doubt. Transactions of the Charles S. Peirce Society				●
	2012	Wahyuni	The Research Design Maze: Understanding Paradigms, Cases, Methods and Methodologies				●
	2012	Saunders et al.	*Research Methods for Business Students*			●	
Irrealism (philosophy)	2016	Cohnitz, Daniel and Rossberg, Marcus	"Nelson Goodman", *The Stanford Encyclopedia of Philosophy*		●		
	2014	Cohnitz	Nelson Goodman		●		
	2013	Bhaskar	A_Realist_Theory_of_Science			●	
Liberal naturalism	2018	Macarthur	Liberal naturalism and the scientific image of the world, Inquiry	●			
	2017	Crespo	Economics and Other Disciplines: Assessing New Economic Currents Routledge Advances in Social Economics			●	

Q-Sort: A Blended Methodology Applied to a Personality Inventory

Rosa Rodrigues[1,2,3], Paula Lopes[1,4] and Miguel Varela[1,2]
[1]Instituto Superior de Gestão – Business and Economics School, Lisboa, Portugal
[2]CEFAGE – Center for Advanced Studies in Management and Economics, Évora, Portugal
[3]The Transdisciplinary Research Center of Innovation & Entrepreneurship Ecosystems (TRIE), Lisboa, Portugal
[4]Lusofona University / CICANT - Centre for Research in Applied Communication, Culture, and New Technologies, Lisboa, Portugal

paula.rosa.lopes@ulusofona.pt
rosa.rodrigues@isg.pt
miguel.varela@isg.pt

Abstract: The instruments that are constituted by inquiries that intend to investigate the opinions, behaviors and attitudes, instead of putting the person to the test, intend to find out how they would act in a given situation. Although there are no right or wrong answers, there is a tendency to respond in a socially acceptable way, even if the answer does not correspond to reality. This problem can be overcome through the Q-sort methodology that combines quantitative and qualitative data and analytical techniques that are not present in other methods. In this way, it consists of presenting the participants with a set of statements on a given topic and asking them to classify them according to their opinion, according to a predefined distribution, which is generally approximately normal. This methodology forces participants to distribute the score among the items on the scale, thus avoiding the constraints associated with social desirability and the tendency to respond in the same way or always through the midpoint to different questions. Another advantage is that it provides linearity and near-normality to the scale, which makes it possible to compare subjects more easily. Nevertheless, its advantages, Q-sort methodology also has negative points because forced-choice measures produce ipsative data that lead to distorted scales and problematic psychometric properties. As the data are obtained by ordering a set of items or by forcefully choosing one item over another, it is impossible to achieve very high or very low values on all scales, which gives rise to a large number of negative values that, in turn, result in an average correlation between the scales, which is also negative. In view of the above, it was considered relevant to apply the Q-sort methodology to a personality inventory, whose data were collected from 175 university students attending the Portuguese higher education institution which specializes in the area of economic and business sciences. The Q-sort methodology plays a crucial role in personality inventories by offering a subjective and personalized approach to assessing personality traits. It enables a more thorough and contextual analysis of individual traits, thereby contributing to a deeper and more comprehensive understanding of the human personality. The results of the empirical study showed that despite the mean values being negative or very close to zero, they allowed the grouping of respondents according to their similarities in terms of their personality traits depending on the course they attend.

Keywords: Q-sort methodology, Data analysis methodology, Social research, Human perspectives, Personality

1. Introduction

Q-methodology was presented by Stephenson (1935) and later refined in Brown (1996) for the areas of social sciences and humanities for studying subjectivity through a combination of qualitative and quantitative techniques (Zabala, Sandbrook and Mukherjee, 2018). It is a methodology that aims to ascertain subjective perspectives, beliefs and concepts that characterize human mental states and behavioral manifestations (Akhtar-Danesh, 2018). May, Luke and James (2018) adds that the Q-sort methodology has been frequently used in research dedicated to the study of personality, since they are usually carried out with self-report instruments, which implies that in most cases the answers are given in line with social norms and with the representations that people construct of reality and not according to reality itself (Escobar Cabello and Sánchez Soto, 2019).

Research on social desirability gains importance when dealing with practical issues faced by organizations that use personality inventories to improve the decision-making process (Pechorro et al., 2012). As they are self-descriptive tests, they can be influenced by the motivation to present the expected result, leading the subject to purposely distort their answers to make a good impression. Many organizations consider this situation to be very difficult to overcome, which significantly reduces the validity of personality measures as useful tools in the work context, namely during the recruitment and selection process (Lee, Joo and Lee, 2019).

According to Fortunato, Tanzilli, Lingiardi and Speranza (2022) when a person fills out a personality inventory using a Likert-type scale, he or she becomes a passive subject who undergoes an assessment. Through Q-sort, people are genuinely active, because they give their point of view through a set of items that holistically portrays

Reference this paper: Rodrigues, R., Lopes, P. and Varela, M., 2023. Q-Sort: A Blended Methodology Applied to a Personality Inventory *The Electronic Journal of Business Research Methods*, 21(1), pp. 31-42, available online at www.ejbrm.com

their personal characteristics. Following this idea, Lundberg (2019) states that the added value of this method is that, unlike the traditional method (Likert-type scale) that assigns a value to each question, the Q-sort induces participants to consider the characteristics that the participants have. define as a whole.

Thus, it is desirable that the data follow a normal distribution, as it means that the values of the variable are grouped around the mean, in a symmetrical pattern, which reveals that the responses of the participants are diversified, a fundamental aspect for the realization of statistical inferences (Kamperman, Kooiman, Lorenzini, Aleknaviciute, Allen and Fonagy, 2022).

The Q-sort methodology is a widely employed technique in personality inventories, aiding in the assessment and categorization of individual traits and characteristics. Its significance in personality inventories stems from its provision of a more personalized approach to evaluating personality traits. Unlike other methods that utilize questionnaires with predetermined responses, the Q-sort enables participants to assess and rate their own personality traits on a graded scale (Block, 1961; Miller and Ozer, 2022).

This qualitative approach allows individuals to express their opinions and perceptions about themselves in a more individualized manner, considering the complexity and nuances of personality traits. Additionally, Q-sort enables comparison between different individuals, facilitating the identification of patterns and individual differences. Another advantage of the Q-sort methodology is its flexibility and adaptability (Rost, 2021). It can be applied in both clinical situations and academic research, allowing for the analysis of various aspects of personality (e.g., temperament, social interaction styles, personal preferences; Lutfallah and Buchanan, 2019).

However, it is important to note that the Q-sort methodology is not a standalone method for personality assessment but can be used in conjunction with other assessment instruments and techniques. The combination of different methods can provide a more comprehensive and accurate understanding of an individual's personality. Q-sort is a measure designed to describe reality and relate it to psychological attributes (Big Five traits, affect, and well-being; Miller and Ozer, 2022).

The purpose of this study is part of this theme and aims to present the Q-sort methodology, namely its stages and the advantages and disadvantages of its use. A comparison of this methodology with the Likert-type scales is also made and its application to a personality inventory based on the Big Five model is demonstrated (Lynn, 2021). Taking into consideration the information provided, the following research question has been formulated: What disparities can be identified in the outcomes of a personality inventory when comparing the Q-sort methodology with Likert-type scales?

2. Literature Review

2.1 Q-sort Methodology

Q-sort is a mixed research method that uses quantitative results to confirm qualitative results in order to better understand the phenomenon under study (Santos et al., 2019). One of the strengths of this methodology is its exploratory approach and its potential to generate theories, since, generally, no hypothesis is formulated before the execution of the study (Lundberg, 2019). In addition, several authors (e.g., Moree, 2017; Thompson et al., 2013) describe the Q-sort methodology as a reliable approach to filling the gaps inherent in the R method, which only allows relating the differences between individuals, while Q-sort makes it possible to relate intra-individual differences.

Eyvindson et al. (2015) allude that the R method emphasizes the analysis of the relationships between the variables, while the Q method focuses on comparing the perception of each individual within the sample in which they belong. This method consists of a group of statements or objects about a topic predetermined by the researcher (e.g., The sustainability practices adopted by the organization comply with current legislation) and from which participants classify these statements into categories (Santos, Petrini, Lupion and Hepper, 2019).

The Q-sort table includes a rating scale that can range from minus three to plus three (-3 and +3) to minus six to plus six (-6 to +6), depending on the degree of agreement, the frequency with which certain behavior occurs or the importance attributed by the participants to each of the statements presented. There is no ideal range, because it depends on the number of statements. A greater number of statements generally requires a wider reach (Akhtar-Danesh, 2018).

When organizing the statements according to their personal opinions, each one occupies a place in the table according to the score given to each one, which varies from negative to positive (Figure 1). After completing each individual Q-sort, the statements receive the rating corresponding to their place in the table and in the

general Q-sort, made up of the responses of all participants. The score given to each statement is added up and generates the final score for each category (Lutfallah and Buchanan, 2019).

The number of columns and rows that make up the table depends on the number of statements developed/selected by the researcher, each of which must have a place in the table so that it can be sorted by the participant. If the number of statements to be evaluated is 25, the table must consist of 25 spaces (Santos, Petrini, Lupion and Hepper, 2019).

The table presented in Figure 1 is composed of nine columns, where the value minus four (-4) corresponds to total disagreement with the statement and four represents total agreement with it. The number of responses corresponding to each statement was limited in advance, which forces a forced distribution and encourages participants to carefully reflect on the ranking of statements according to their point of view on the topic (Lucinski, 2016). The Q-sort matrix facilitates the analysis and interpretation of results and increases the precision of the analysis performed (Dieteren, Patty, Reckers-Droog and Exel, 2023).

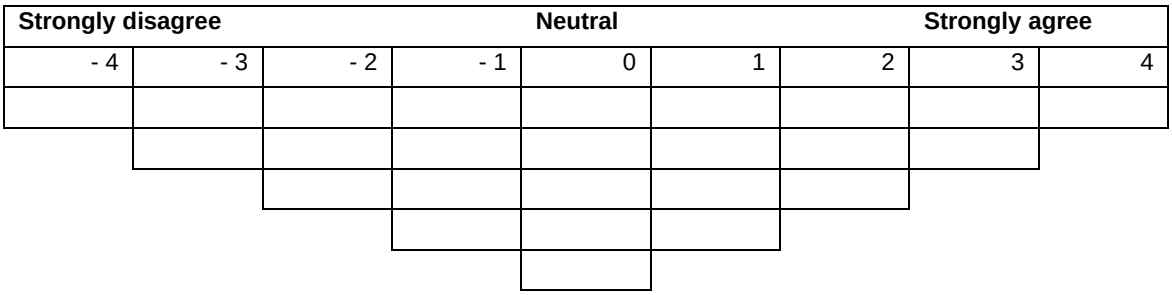


Figure 1: Q-sort Table With Extremes That Oscillate Between – 4 and 4

(Adapted from Santos, Petrini, Lupion and Hepper, 2019).

The Q-sort methodology makes it possible to establish correlations between people and not between measurement instruments, so that participants who order items in a similar way share the same point of view on the subject under analysis. Based on individual correlations, factors are extracted that identify people who reveal similar or different opinions regarding a given topic (Escobar Cabello and Sánchez Soto, 2019). The linearity and approximation of the normality of the data distribution, allows the subjects to be compared with each other more easily (Stenner and Capdevila, 2020).

2.2 Q-sort Methodology Phases

The effectiveness of the Q-sort methodology depends on the fulfillment of the steps recommended by Ferreira, Oliveira and Ferreira (2022), namely: (i) identification of the topic to be analyzed; (ii) gathering information from a literature review or through interviews with experts in the study area; (iii) selection/development of a representative set of statements; (iv) election of participants who meet all the inclusion criteria; (v) construction and application of the Q-sort with the statements considered most relevant to answer the research problem; (vi) statistically analyze individual and global results through factor analysis of Q-sorts, with the aim of revealing which individuals are part of each point of view and the "strength" of this connection; (vii) and qualitative interpretation of ratings to explore statements that differentiate one factor from others (Figure 2).

It should be noted that the Q-sort methodology does not require large samples, because after a certain number of statements, theoretical saturation is reached and no new information is introduced (Maia, Espindola and Veiga, 2018). In addition, statistical validity is not the main concern of this type of methodology, since the importance falls on the different opinions (subjectivity) about the topic being studied and not on the percentage of the population that adheres to each of these opinions (Ferreira, Oliveira and Ferreira, 2022).

This methodology allows analyzing whether there is agreement between the opinions of the participants, how and why it occurs. The comparison of convergent, complementary and/or contradictory opinions can be extremely useful in all areas of knowledge, in particular those dedicated to the study of subjective social phenomena that largely depend on the values and beliefs of those who evaluate them (Escobar Cabello and Sánchez Soto, 2019).

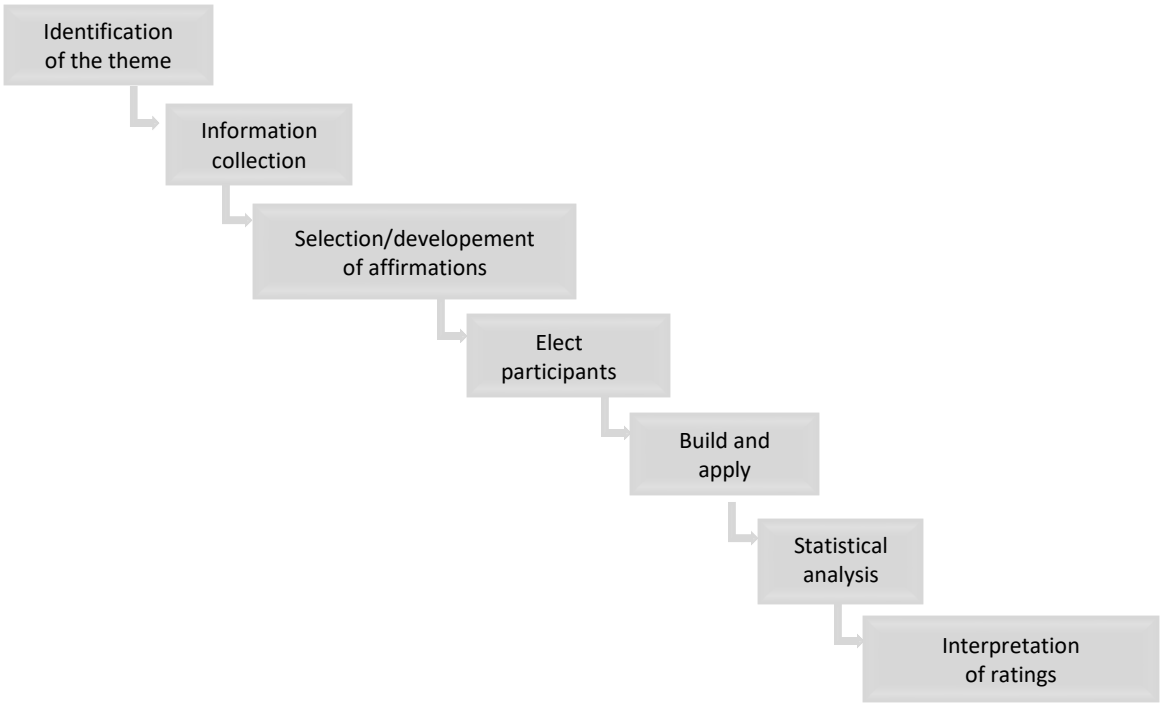


Figure 2: Q-sort Methodology Phases

(Adapted from Ferreira, Oliveira and Ferreira, 2022).

The use of the Q-sort methodology in self-report measures that assess human behaviors and characteristics (e.g., personality, skills) has also been growing in recent years (May, Luke and James, 2018). The use of this methodology makes it possible to define similar profiles of individuals who share the same ideas, attitudes and behaviors (Stenner and Capdevila, 2020) regardless of sociodemographic variables (e.g. sex, age, education level).

It is important to mention that like any methodology, this one also has advantages and disadvantages that we will now describe.

2.3 Q-sort Methodology's Advantages and Disadvantages

One of the main advantages of this methodology is related to the decrease in the number of answers attributed to the midpoint of the scale, because when participants answer different types of questions in the same way, validity is compromised due to the bias of the results (Karim, 2001).

According to Stenner and Capdevila (2020), the person always marks the same answer because he is afraid to take a position on the issue under analysis and decides to respond according to his perception that it is socially desirable. With Q-sort, participants are forced to distribute their answers across the various spaces of the table, which implies decision making and, as such, the effort to distort the answers is considerably less (Ramlo, 2021).

Unlike questionnaires, whose items can be answered using a Likert-type scale, which requires a single reading for a generally direct and quick response, Q-sort may require several decisions to assign an answer, because it requires the comparison of each affirmation with the others, which increases the validity of the evaluation process (Eyvindson et al., 2015).

Another advantage concerns the a priori criterion used by the researcher to develop the measurement instrument, since, as he is responsible for choosing theories and/or variables, he selects the response scales that may allow him to validate his hypotheses (Lundberg, 2019).

In the Q-sort, the answers are classified, by the participants, according to their references on the topic under analysis, that is, it is the respondent who decides the importance he attaches to each statement in relation to the others and their order in the table. Subsequently, this ranking is compared with that of the other participants to assess the similarities and differences regarding their points of view. Thus, it is not the researcher who prepares the Q-Sort who decides, a priori, the classifications, but the respondent, from his point of view (Stenner and Capdevila, 2020)

Despite its advantages, the Q-sort methodology also has negative points, because forced-choice measures produce ipsative[1] data that lead to distorted scales and problematic psychometric properties (Salgado, Anderson and Tauriz, 2015). In this context, Martínez, Moscoso and Lado (2021) refer that as the results are relative, it is impossible to obtain very high or very low values on all scales, which increase a large number of negative values which, in turn, result in an average correlation between the scales, also negative. And despite these averages approaching zero, when there are few scales involved, it becomes difficult to assess construct validity through the Classical Test Theory (Kleka and Soroko, 2018), which leads Walton, Cherkasova and Roberts (2020) stating that with less than 30 scales it is practically impossible to obtain psychometric parameters that can be interpreted.

Dieteren, Patty, Reckers-Droog and Exel (2023) also mention that often a normal distribution may not be appropriate for ipsative data, because as all scales are correlated, it is more likely that profiles with predominantly positive or negative values will emerge, which have asymmetry coefficients and kurtosis that deviate from the range -1.96 to 1.96, recommended in the literature (Marôco, 2021).

The interdependence present in the forced choice scales and in the observed results can change the psychometric properties of the instrument, because the selected item does not depend only on the level of latency that it is measuring, but also on the set of items to which it belongs, which makes that each observed result is influenced by the results of the set of items (Welter and Capitão, 2007). Santos, Petrini, Lupion and Hepper (2019), in turn, report that due to the low intercorrelation of the items, the subscales tend to have low internal consistency, with average values around 0.20, which is why it is not rare to reject them to the detriment of the subscales. Likert scales.

3. Q-sort Methodology Versus Likert-Type Scales

The choice of the Q-sort methodology instead of the Likert scale was due to the fact that the objective of the study was to order a set of behaviors/characteristics according to the frequency with which they occur (Stenner and Capdevila, 2020).

Likert-type scales have the inconvenience that the respondent looks at each variable individually and not as a whole composed of items that must be related (Costa, Orsini and Carneiro, 2018). When considering each variable separately, it becomes very difficult to weigh its importance in relation to the others. In addition, there is a tendency to select the midpoint of the scale, assign extreme values and/or respond according to what is considered socially desirable (Stenner and Capdevila, 2020). This strategy causes many repetitions between the variables, which tends to bias the results, since the objective is to produce a list ordered according to the frequency with which a certain behavior occurs, the degree of agreement or the relative importance of each item (Ramlo, 2021).

Added to this inconvenience is the propensity to generate asymmetric distributions and with Q-sort this problem is overcome because the participant has to look at the items as a whole and classify them according to a predefined quasi-normal distribution (Escobar Cabello and Sánchez Soto, 2019). The respondent has to reflect on each of the items and order them according to the degree of importance, frequency or agreement in relation to a given subject. In this way, a hierarchical list is obtained, without classification ambiguities and without any probability of having repeated variables in the same position (Santos, Petrini, Lupion and Hepper, 2019).

By forcing the creation of an approximately normal distribution, the Q methodology facilitates inferences and statistical comparisons between the various elements of the group and how each item is compared with the others, more realistically reflects the opinion of the participants on the subject under study, without it being socially desirable (Kamperman et al., 2022).

Despite the advantages of the Q-sort, when questions are formulated to measure attitudes and opinions, Likert-type scales are generally used because they allow the assessment of the respondents' opinion on a specific subject, in terms of agreement (agree versus disagree) or frequency (few versus often). However, since its appearance (Likert, 1932) validity and reliability have been discussed in terms of the number of points on the scale. There are authors (e.g., Dalmoro and Vieira, 2013; Lozano, García-Cueto and Muñiz, 2008; Weng, 2004) who argue that a seven-point scale gives better results than a five-point scale, because the internal consistency of the instrument increases as the number of response categories increases.

[1]Data obtained through the ranking of a set of items or through the forced choice of one item over another (Brown and Maydeu-Olivares, 2011).

However, the validity of Likert-type scales depends on the subject under study and the importance attributed to it by the respondents, because any question that is considered irrelevant will have a similar answer regardless of the number of points on the scale. to cluster around the center or at the extremes (Costa, Orsini and Carneiro, 2018). On the other hand, it appears that when the subject in question is relevant to the respondents, the greater the number of response options, the better the reliability and construct validity of the scale, because it reduces the ambiguity of the answers and becomes closer to the reality of respondents (Dalmoro and Vieira, 2013).

Both methodologies have advantages and disadvantages, and both require correlations, but while in Likert-type scales correlations are performed between items, in Q-sort methodology they are performed between people (Gao and Soranzo, 2020).

4. Application of the Q-sort Methodology to a Personality Inventory

Personality has become increasingly important in several areas of knowledge, as it is through its study that it becomes possible to identify patterns of behavior and attitudes that make each individual unique (Rodrigues and Gomes, 2022). In addition, there are several studies (e.g., Brandt et al., 2019; Graham et al., 2020; Rais and Chandgude, 2020; Smith et al., 2021) that over time have shown that the measures of personality predict academic and professional performance, learning ability, productivity, salary, career progression and other relevant criteria in both an educational and organizational context.

In this research, the application of the Q-sort methodology followed the steps recommended by Ferreira, Oliveira and Ferreira (2022). Thus, after identifying the model to be used, an extensive literature review was carried out around the Big Five model, which postulates the existence of five dimensions: (i) Extroversion, which characterizes warm, assertive people, who experience positive emotional states and they feel good about themselves and the world (Brandt et al.,2019); (ii) Neuroticism, which concerns the tendency to experience negative emotional states and to face the world and oneself as something harmful (Pechorro, 2019); (iii) Agreeableness, which represents cooperativism, altruism and team spirit and as such constitutes the core of interpersonal relationships (Tackett, Hernandez and Eisenberg, 2019); (iv) Conscientiousness, which is seen as a good predictor of professional performance, because it reflects how scrupulous, careful and persevering a person is (Smith et al., 2021); (v) and Openness to experience, which is directly related to the person's area of interests and their tendency to take calculated risks (Bergner, 2020).

The added value of the Q-sort methodology lies in its ability to enable a deeper understanding of individual perspectives and differences among participants. This is particularly useful in studies seeking to comprehend the diversity of opinions within a group and identify patterns based on the rankings (Ramlo, 2021).

5. Method

In accomplishing this work, a mixed methodology was employed, from which two studies were conducted: one of a qualitative nature and another anchored in a quantitative approach.

5.1 Participants

Study 1

Semi-structured interviews were conducted with 97 employees of a consulting company who received ratings of "Very good" and "Excellent" in their performance evaluations in the past year. The ages ranged from 28 to 51 years old (M = 35.2; SD = 1.03), and 56.7% were male. Regarding tenure, it was found that 35.2% of the interviewees had been with the organization for more than 10 years.

Study 2

The sample consisted of 175 students, aged between 18 and 31 years old (M = 22.40; SD = 3.15), who are studying at a Portuguese higher education institution specializing in the area of economic and business sciences. Of the students, 36.0% are enrolled in the Bachelor's program in Management, 29.7% are pursuing Economics, and 34.3% are studying Human Resources Management (HRM).

Procedures

The interviews were conducted individually in an appropriate location and lasted approximately 20 minutes. The questionnaires were administered in the classroom with prior authorization from the instructors. In both cases, the participants were informed about the study's objectives. It was also ensured that the guidelines of the

General Data Protection Regulation (GDPR) regarding the confidentiality and anonymity of the provided responses would be followed.

After analyzing the content of the interviews, the development of items was carried out, which were subsequently analyzed through spoken reflection of the items, involving ten experts dedicated to the study of personality (Rodrigues and Gomes, 2022). Next, the statements to be included in the measurement instrument were selected, and the Q-sort was constructed (Ferreira, Oliveira and Ferreira, 2022). It is worth mentioning that the research aimed to analyze the type of personality of university students, so only subjects who were enrolled in an undergraduate program were included in the study (inclusion criteria).

Results Before applying the Q-sort, participants were instructed to assign a score to the items, ranging from -3 (Never) to 3 (Always). Then, each respondent was asked to order the 23 items in the seven columns presented, and what was intended was explained in detail (Table 1).

Table 1: Q-sort Presented to Study Participants

This questionnaire consists of 23 statements that you should read carefully, and then sort them according to the frequency with which they occur. For each statement, there are seven response options ranging between -3 = Never and 3 = Always. In each cell of the table, place the number corresponding to the following statements.

1	I put others in first place.
2	I am always ready to help.
3	I put the interests of the group ahead of mine.
4	I consider myself a tolerant person.
5	I'm a humble person.
6	I am selfless.
7	I make friends easily.
8	I feel good when I am surrounded by people.
9	I like to meet new people.
10	I easily fit into any group
11	I am little impulsive.
12	I'm a relaxed person.
13	I'm a patient person
14	I consider myself a calm person.
15	I'm a punctual person
16	I perform all the tasks assigned to me with the same rigor.
17	I always do more than what is asked of me.
18	I consider myself a scrupulous person.
19	I think through all the pros and cons thoroughly before deciding.
20	I easily adapt to new contexts (e.g. cultures, environments).
21	I am not afraid to face any kind of challenge.
22	I am always willing to learn new things.
23	I can solve complex problems.

Never			**Neutral**			**Always**
- 3	- 3	-1	0	1	2	3

In the first column on the left (-3) the respondent placed the characteristics/behaviours he demonstrates less frequently and in the first column on the right (+3) he placed his most frequent behaviors/characteristics. The number of values in each column was defined in order to form a normal distribution.

The Q-sort technique aimed to identify the participants' opinion about their way of thinking, feeling and acting that characterize the way they interact with others and adjust to the demands of the environment in which they are inserted. This methodology forced respondents to distribute the statements among the various cells that make up the table, which prevented them from always responding in the same way or always using the midpoint to the different statements (Kamperman et al., 2020). In addition to avoiding social desirability, Q-sort facilitates the comparison of the various participants in relation to a given characteristic/behavior, because the correlations are determined between the various subjects and not between the measures that evaluate them (Gao & Soranzo, 2020).

The individual and global results analyzed through the factor analysis of the Q-sorts revealed that the internal consistency indexes are much lower using this methodology than when using Likert-type scales (Table 2). Conclusions that are congruent with those found by Havlíková (2016), according to which the Likert-type scale improves statistical relevance and facilitates the interpretation of results. In the same vein, Walton, Cherkasova and Roberts (2020) report that when the measures are composed of a small number of subscales, as is the case of the study presented here, it is very difficult to achieve adequate indices of validity and reliability.

Table 2: Internal Consistency Indices: Likert Scale Versus Q-sort Methodology

Personality Dimensions	Likert scale	Q-sort methodology
Neuroticism	0.83	0.53
Conscientiousness	0.84	0.48
Openness to experience	0.80	0.31
Extroversion	0.75	0.36
Agreableness	0.80	0.26

As previously mentioned, the ipsative data resulting from measures that force the choice of one item over another give rise to questionable psychometric indicators (Salgado, Anderson and Tauriz, 2015). Since the answers given to a given item are ranked, in relation to the others, they generate negative average values or very close to zero, which in turn translate into equally negative correlations, in most dimensions (Table 3).

Table 3: Correlation Between the Five Personality Dimensions

Personality Dimensions	M	SD	1	2	3	4
Neuroticism (1)	0.84	0.72	-			
Conscientiousness (2)	0.43	0.34	- 0.100	-		
Openness to experience (3)	0.72	0.92	- 0.315**	- 0.010	-	
Extroversion (4)	0.04	0.57	- 0.206**	- 0.353**	- 0.248**	-
Agreableness (5)	- 0.51	0.60	- 0.105	- 0.054	- 0.019	0.033

Note: M = Mean; SD = Standard deviation; $^{**}p < 0.001$

Despite its limitations, this study allowed the grouping of respondents according to their similarities in terms of their personality traits, depending on the course they attend. The results obtained were coded into two groups according to their midpoint: (i) low values and (ii) high values, resulting in the results shown in Table 4.

Table 4: Personality Traits of Participants Depending on the Degree They Attend

Personality traits	Economics	Management	HRM
Neuroticism			
Low values	42.3%	68.3%	53.3%
High values	**57.7%**	31.7%	46.7%

Personality traits	Economics	Management	HRM
Conscientiousness			
Low values	38.5%	46.0%	46.7%
High values	**61.5%**	54.0%	53.3%
Openness to experience			
Low values	42.9%	40.4%	43.3%
High values	57.1%	**59.6%**	56.7%
Extraversion			
Low values	68.3%	73.0%	67.3%
High values	31.7%	27.0%	**32.7%**
Agreeableness			
Low values	90.4%	63.5%	6.7%
High values	9.6%	36.5%	**93.3%**

The results obtained suggest that students who attend the degree in Economics are the most emotionally unstable, as they have the highest percentage of high values in the Neuroticism dimension (Pechorro, 2019); on the other hand, they are also the ones that stand out in the Conscientiousness dimension, which assesses the degree of organization, persistence and motivation that characterize goal-oriented behavior (Dietl and Kombeiz, 2020). High values in this domain differentiate people who are scrupulous from those who are lazy and careless. The conscientious subject is determined, confident, punctual, hardworking, self-disciplined and ambitious, so he is usually successful both academically and professionally (Smith et al., 2021).

Students in the Management course stand out in the Openness to experience dimension, which is expressed through the exploration of what is unfamiliar to them (Bergner, 2020). People with high values in this factor are curious about their inner and outer world, so their experiences are very rich and, as such, they are always willing to consider new ideas and unconventional values (Lynn, 2021).

Finally, it was found that students who attend the degree in HRM are the ones with the highest values in the Extraversion and Agreeableness dimensions, which is reflected in the amount and intensity of their interpersonal interactions (Tackett, Hernandez, Eisenberg, 2019). Extroverts are sociable people who, in addition to enjoying socializing with others, are optimistic, enjoy fun, are affectionate, active, and talkative (Brandt et al., 2019). Allied to these characteristics are the attributes of kindness, namely: altruism, sympathy, benevolence, and willingness to believe in others (Smith et al., 2021).

6. Conclusion

In conclusion, the Q-sort methodology offers a unique approach to interpreting results, focusing on the comparison of individuals rather than items, as observed in Likert-type scales. This method reveals the subjectivity of interrelationships and similarities among individuals within a sample (Gao and Soranzo, 2020).

A significant distinction between Likert-type scales and Q-sort is that the former yields average item scores based on participants' random choices, whereas Q-sort compares and hierarchically groups all items. Likert-type scales are prone to bias when participants tend to select the same answers, which is mitigated by the forced choice technique employed in Q-sort (Almiro, 2017; Escobar Cabello and Sánchez Soto, 2019).

It is essential to note that Q-sort is not a statistical method aimed at identifying correlated input variables or forming new variables for evaluating similar aspects. Instead, its purpose is to explore similarities and differences among participants (Stenner and Capdevila, 2020).

The Q-sort methodology finds utility in studying social phenomena characterized by debate, conflict, and contestation (e.g., sustainability, politics, religion; Lundberg, 2019), as well as in measuring attitudes, behaviors, and personal characteristics assessed through self-report instruments (e.g., personality, skills, aptitudes; May, Luke and James, 2018). Its objective is not to ascertain absolute truth but rather to collect and compare diverse opinions and perspectives (Wulff, 2019).

According to Santos, Petrini, Lupion and Hepper, 2019 (2019), the distribution utilized in Q-sort methodology presents advantages over traditional Likert-type scales. Notably, respondents are unable to agree with all questions or solely choose central answers; instead, all responses must be allocated within the corresponding table space assigned to each item (Ramlo, 2021).

Developing a Q-sort instrument involves presenting a set of statements to participants, who must then distribute them based on their degree of agreement, frequency, or importance. This classification relies on participants' own references and follows a pre-established nomenclature (Rampold et al., 2020).

The Q-sort methodology was devised to provide individuals with an opportunity to express their opinions on a given topic, without the researcher determining the variables necessary for validating their research hypotheses (Ferreira, Oliveira and Ferreira., 2022).

References

Akhtar-Danesh, N. (2018) "Qfactor: A command for Q-methodology analysis", *The Stata Journal,* Vol. 18., No. 2, pp. 432-446. https://doi.org/10.1177/1536867X1801800209

Almiro, P. (2017) "Uma nota sobre a desejabilidade social e o enviesamento de respostas" [A note on social desirability and response bias], *Avaliação Psicológica*, Vol. 16, No. 3, pp. 1-4. http://dx.doi.org/10.15689/ap.2017.1603.ed

Bergner, R. (2020) "What is personality? Two myths and a definition", *New Ideas in Psychology*, Vol. 57, pp. 1-7. https://doi.org/10.1016/j.newideapsych.2019.100759

Block, J. (1961) The Q-sort method in personality assessment and psychiatric research. Illinois: Charles C Thomas Publisher. https://doi.org/10.1037/13141-000

Brandt, N., Lechner, C., Tetzner, J. and Rammstedt, B. (2019) "Personality, cognitive ability, and academic performance: Differential associations across school subjects and school tracks", *Journal of Personality,* Vol. 87, No. 1, pp. 1-17. https://doi.org/10.1111/jopy.12482

Brown, A. and Maydeu-Olivares, A. (2011) "Item Response Modeling of Forced-Choice Questionnaires", *Educational and Psychological Measurement*, Vol. 71, No. 3, pp. 460-502. https://doi.org/10.1177/0013164410375112

Brown, S. (1996) "Q methodology and qualitative research", *Qualitative health research*, Vol. 6, No. 4, pp. 561-567. https://doi.org/10.1177/104973239600600408

Costa, F., Orsini, A. and Carneiro, J. (2018) "Variações de mensuração por tipos de escalas de verificação: uma análise do construto de satisfação discente" [Dilemmas of the type likert scales construction: does the number of items and the disposition influence results?], *Gestão.Org: Revista Eletrônica de Gestão Organizacional,* Vol. 16, No. 2, pp. 132-144. http://dx.doi.org/10.21714/1679-18272018v16n2

Dalmoro, M. and Vieira, K. (2013) "Dilemmas of the type likert scales construction: does the number of items and the disposition influence results?", *Revista Gestão Organizacional*, Vol. 6, No. 3, pp. 161-174.

Dieteren, C., Patty, N., Reckers-Droog, V. and Exel, E. (2023) "Methodological choices in applications of Q methodology: A systematic literature review", Social Sciences & Humanities Open, Vol. 71, No 1, 1-26. https://doi.org/10.1016/j.ssaho.2023.100404

Dietl, E. and Kombeiz, O. (2020) "The interplay between extraversion, task significance and task variety at work", *Personality and Individual Differences*, Vol. 171, pp. 1-8. https://doi.org/10.1016/j.paid.2020.110471

Escobar-Cabello, M. and Sánchez-Soto, I. (2019) "Implementación del Q-SORT en la evaluación diagnóstica y formativa de logros competenciales en estudiantes de Kinesiología" [Implementation of the Q-SORT in the diagnostic and formative evaluation of competencies achievement in Kinesiology students], *Revista de investigación,* Vol. 43, No. 97, pp. 191-209.

Eyvindson, K., Kangas, A., Hujala, T. and Leskinen, P. (2015) "Likert versus Q-approaches in survey methodologies: discrepancies in results with same respondents", *Quality & Quantity*, Vol. 49, No. 2, pp. 509-522. https://doi.org/10.1007/s11135-014-0006-y

Ferreira, D., Oliveira, A., Ferreira, O. and Gomes, M. (2022) "Espiritualidade nexo qualidade de vida no trabalho: Um estudo subjetivo utilizando a metodologia Q" [Spirituality nexus quality of life at work: A subjective study using the Q methodology], *Research, Society and Development*, Vol. 11, No. 2, pp. 1-13. http://dx.doi.org/10.33448/rsd-v11i2.25648

Fortunato, A., Tanzilli, A., Lingiardi, V. and Speranza, A. (2022) "The CPAP-Q: A Q-Sort Assessment Procedure for assessing traits and emerging personality patterns in childhood", *Acta Psychologica*, Vol. 229, pp. 1-10. https://doi.org/10.1016/j.actpsy.2022.103688

Gao, J. and Soranzo, A. (2020) "Applying Q-Methodology to Investigate People' Preferences for Multivariate Stimuli", *Frontiers in Psychology* (published online). https://doi.org/10.3389/fpsyg.2020.556509

Graham, E., Weston, S., Gerstorf, D., Yoneda, T., Booth, T., Beam, C., Petkus, C., Drewelies, J., Hall, A., Bastarache, E., Estabrook, R., Katz, M., Turiano, N., Lindenberger, U., Smith, J., Wagner, G., Pedersen, N., Allemand, M., Spiro, A., Deeg, D. … Mroczek, D. (2020) "Trajectories of Big Five Personality traits: A coordinated analysis of 16 longitudinal samples", *European Journal of Personality*, Vol. 34, No. 3, pp. 301-321. https://doi.org/10.1002/per.2259

Havlíková, M. (2016) "Likert scale versus Q-table measures: a comparison of host community perceptions of a film festival", *Scandinavian Journal of Hospitality and Tourism*, Vol. 16, No. 2, pp. 196-207. https://doi.org/10.1080/15022250.2015.1114901

Kamperman, A., Kooiman, C., Lorenzini, N., Aleknaviciute, J., Allen, J. and Fonagy, P. (2020) "Using the attachment network Q-sort for profiling one's attachment style with different attachment-figures", *PloS one,* Vol. 15, No. 9, pp. 1-22. https://doi.org/10.1371/journal.pone.0237576

Karim, K. (2001) "Q-methodology: Advantages and disadvantages of this research method", *Journal of Community Nursing,* Vol. 15, pp. 8-10.

Kleka, P. and Soroko, E. (2018, August) "How to Abbreviate Questionnaires and Avoid the Sins?", *Survey Research Methods*, Vol. 12, No. 2, pp. 147-160. https://doi.org/10.18148/srm/2018.v12i2.7224

Lee, P., Joo, S. and Lee, S. (2019) "Examining stability of personality profile solutions between Likert-type and multidimensional forced choice measure", *Personality and Individual Differences*, Vol. 142, pp. 13-20. https://doi.org/10.1016/j.paid.2019.01.022

Likert, R. (1932) "A technique for the measurement of attitudes", *Archives of Psychology*, Vol. 22, pp. 140-155.

Lozano, L., García-Cueto, E. and Muñiz, J. (2008) "Effect of the number of response categories on the reliability and validity of rating scales", *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, Vol. 4, No. 2, pp. 73-79. https://doi.org/10.1027/1614-2241.4.2.73

Lucinski, M. (2016) "*Competências de gestores de projetos voltadas para uma liderança sustentável: um estudo subjetivo utilizando a metodologia Q"* [Competencies of project managers aimed at sustainable leadership: a subjective study using the Q methodology]. Escola Politécnica da Universidade de São Paulo

Lundberg, A. (2019) "Teachers' beliefs about multilingualism: findings from Q method research", *Current Issues in Language Planning*, Vol. 20, No. 3, pp. 266-283. https://doi.org/10.1080/14664208.2018.1495373

Lutfallah, S. and Buchanan, L. (2019) "Quantifying subjective data using online Q-methodology software", *The Mental Lexicon*, Vol. 14, No. 3, pp. 415-423. https://doi.org/10.1075/ml.20002.lut

Lynn, M. (2021) "Effects of the Big Five personality traits on tipping attitudes, motives, and behaviors", *International Journal of Hospitality Management*, Vol. 92, pp. 1-10. https://doi.org/10.1016/j.ijhm.2020.102722

Maia, L., Espindola, D. and Veiga, C. (2018) "Operations social practices in safety and health at work: scale validation and reliability through the Q-sort method", *Revista de Gestão*, Vol. 25, No. 1, pp. 119-139. https://doi.org/10.1108/REGE-11-2017-004

Marôco, J. (2021) *Análise estatística com o SPSS Statistics*. Pêro Pinheiro: Report Number.

Martínez, A., Moscoso, S. and Lado, M. (2021) "Faking Effects on the Factor Structure of a Quasi-Ipsative Forced-Choice Personality Inventory", *Journal of Work and Organizational Psychology*, Vol. 37. No. 1, pp. 1-10. https://doi.org/jwop2021a7

May, E., Luke, W. and James, C. (2018) "Q-Sort Personality Assessment." In: *The Star Gate Archives: Reports of the United States Government Sponsored Psi Program, 1972-1995,* Vol. 2, pp. 469-472. North Carolina: McFarland & Company, Inc.

Miller, T. and Ozer, D. (2022) "The Day Evaluation Q-Sort", *European Journal of Psychological Assessment*, Vol. 38 No. 3, pp. 198-209. https://doi.org/10.1027/1015-5759/a000654

Moree, W. (2017) "Q-Methodology Explained by Comparing Q-Sort Survey with Conventional R Sample Survey and Relating Factor Analysis Described", *Civil Engineering Research Journal,* Vol. 1, No. 2, pp. 1-5. https://doi.org/10.19080/CERJ.2017.01.555560

Pechorro, P., Nunes, C., Gonçalves, R., Simões, M. and Oliveira, J. (2019) "Estudo de Validação do Inventário de Personalidade Narcísica – 13 numa Amostra Escolar de Jovens Portugueses" [Validation study of the Narcissistic Personality Inventory – 13 among a school sample of Portuguese youths], *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, Vol. 50, No. 1, pp. 71-81. https://doi.org/10.21865/RIDEP50.1.06

Pechorro, P., Vieira, R., Poiares, C. and Marôco, J. (2012) "Contributos para a validação duma versão curta da Escala de Desejabilidade Social de Marlowe-Crowne com adolescentes portugueses" [contributions to the validation of a short version of the marlowe-crowne social desirability scale with portuguese adolescents], *Arquivos de Medicina,* Vol. 26, No. 1, pp. 11-17.

Rais, S. and Chandgude, A. (2020) "Personality prediction using HR Assessment Tool", *International Journal of Current Engineering and Technology,* Vol. 8, pp. 1-4.

Ramlo, S. (2021) "Q methodology as mixed Analysis". In: *The Routledge reviewer's guide for mixed methods research analysis*" (pp. 199-208). London: Routledge.

Rampold, S., Coleman, B., Bunch, J. and Roberts, R. (2020) "Exploring students' cultural competence development during a short-term international experience: A Q-sort study*", Advancements in Agricultural Development*, Vol. 1, No. 2, pp. 65-78. https://doi.org/10.37433/aad.v1i2.45)

Rodrigues, R. and Gomes, C. (2022) "Desenvolvimento e Validação de uma Versão Portuguesa do Inventário de Personalidade Big Five" [Development and Validation of a Portuguese Version of the Big Five Inventory]. *Revista Iberoamericana de Diagnóstico y Evaluación-e Avaliação Psicológica,* Vol. 2, No. 63, pp. 163-176. https://doi.org/10.21865/RIDEP63.2.12

Rost, F. (2021) "Q-sort methodology: Bridging the divide between qualitative and quantitative. An introduction to an innovative method for psychotherapy research", *Counselling and Psychotherapy Research*, Vol. 21, No. 1, pp. 98-106. https://doi.org/10.1002/capr.12367

Salgado, J., Anderson, N. and Tauriz, G. (2015) "The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis", *Journal of Occupational and Organizational Psychology*, Vol. 88, No. 4, pp. 797-934. https://doi.org/10.1111/joop.12098

Santos, A., Petrini, M., Lupion, R. and Hepper, E. (2019) "With your actions I discover who you are: Profiles of Sustainable Organizations", *Revista de Administração e Contabilidade da Unisinos,* Vol. 16, No. 3, pp. 464-487. Hepper, E. (2019) "https://doi.org/10.4013/base.2019.163.05

Smith, M., Hamplová, D., Kelley, K. and Evans, M. (2021) "Concise survey measures for the Big Five personality traits", *Research in Social Stratification and Mobility,* Vol. 73, pp. 1-11. https://doi.org/10.1016/j.rssm.2021.100595

Stenner, P. and Capdevila, R. (2020) *Q methodology*. Newcastle: Sage.

Stephenson, W. (1935) "Correlating persons instead of tests", *Journal of Personality*, Vol. 4, No. 1, pp. 17-24. https://doi.org/10.1111/j.1467-6494.1935.tb02022.x

Tackett, J., Hernandez, M. and Eisenberg, N. (2019) "Agreeableness". In: R. McAdams, J. Shiner and J. Tackett (Eds.), *Handbook of Personality Development* (pp. 171-184). New York: Guilford Press.

Thompson, A., Dumyahn, S., Prokopy, L., Amberg, S., Baumgart-Getz, A., Jackson-Tyree, J., Perry-Hill, R., Reimer, A., Robinson, K. and Mase, A. (2013) "Comparing random sample Q and R methods for understanding natural resource attitudes", *Field Methods*, Vol. 25, No. 1, pp. 25-46. https://doi.org/10.1177/1525822X12453516

Walton, K., Cherkasova, L., and Roberts, R. (2020) "On the validity of forced choice scores derived from the Thurstonian item response theory model", *Assessment*, Vol. 27, No. 4, pp. 706-718. https://doi.org/10.1177/1073191119843585

Welter, G. and Capitão, C. (2007) "Medidas ipsativas na avaliação psicológica" [Ipsative measures in psychological evaluation], *Avaliação Psicológica*, Vol. 6, No. 2, pp. 157-165.

Weng, L. (2004) "Impact of the number of response categories and anchor labels on coefficient Alpha and test-retest reliability", *Educational and Psychological Measurement*, Vol. 64, No. 6, pp. 956-972. https://doi.org/10.1177/0013164404268674

Wulff, D. (2019) "Prototypes of Faith: Findings with the Faith Q-Sort", *Journal for the Scientific Study of Religion*, Vol. 58, No. 3, pp. 643-665. https://doi.org/10.1111/jssr.12615

Zabala, A., Sandbrook, C. and Mukherjee, N. (2018) "When and how to use Q methodology to understand perspectives in conservation research", *Conservation Biology*, Vol. 32, No. 5, pp. 1185-1194. https://doi.org/10.1111/cobi.13123

Using Mixed Methods to Understand Tax Compliance Behaviour

Nompumelelo Monageng
University of Pretoria, South Africa

mpumi.monageng@up.ac.za

Abstract: Mixed methods research is not commonly adopted by researchers studying tax compliance behaviour despite the benefits that it can bring to research in this area. This is a method that is generally associated with social sciences however, this emergent methodology is being increasingly applied in disciplines that are traditionally associated with quantitative research, including in tax compliance research. Despite the growing trend in applying mixed methods to tax compliance research, there are no known studies that have summarised this methodological approach for researchers and provide guidance on how mixing research methods can allow for an in-depth view of tax compliance behaviour. The purpose of this article is first, to briefly explain mixed methods research for novice and established researchers unfamiliar with this methodological approach; second, provide an overview of the use of mixed methods in tax compliance research; third, provide an example of using mixed methods in order to illustrate a practical application of mixed methods; fourth, to discuss the value gained in applying mixed methods to examine and understand the effect of reciprocity nudges on tax compliance behaviour as well how challenges in applying mixed methods research can be faced. This article contributes to the business and management methodological literature by summarising the implementation of this approach and how studies aimed at understanding tax compliance behaviour could be enriched by embracing a mixed methods approach.

Keywords: Mixed methods, Tax compliance behaviour, Content analysis, Experiment

1. Introduction

Qualitative research approaches and quantitative research approaches are often seen as two separate and opposing ways of undertaking research. This view is due to the preconceived association of positivist paradigms with the quantitative methods and the interpretivist and constructivist paradigms with qualitative methods (Onwuegbuzie and Leech, 2005). The introduction of a mixed methods approach to research addresses this apparent divide between these two approaches. The mixed methods approach allows for a holistic understanding of complex research problems and encourages a greater focus on the research problem rather than the research method (Truong, Xiaoming Liu and Yu , 2020).

Although researchers have largely applied quantitative approaches in studying tax compliance, qualitative approaches have also been applied, albeit to a lesser extent (Dallyn, 2017; Hodgson, Walpole and McKerchar, 2009) and a limited number of studies have combined the two approaches (Angeles, 2021; Chen Loo, McKerchar and Hansford, 2009; Enachescu et al., 2019; Kurniawan, 2020; Mohdali and Pope, 2014; Saulitis and Chapkovski, 2023).

Reviews of tax compliance studies conducted by Jackson and Milliron (1986) prior to 1986; Richardson and Sawyer (2001) from 1986 to 1997 and Yong et al. (2019) from 1998 to 2017 reflect this dominance of quantitative approaches and the limited use of qualitative methods and mixed methods throughout the years. Jackson and Milliron (1986) review showed that the methodological approaches followed prior to 1986 were quantitative comprising of surveys, experiments, analytical models and regression analysis. Richardson and Sawyer (2001) study showed the emergence of qualitative approaches with the authors identifying the use of process tracing, which is a qualitative method, in one of the tax compliance studies. From 1998 to 2017 there appeared to be an acceptance of the use of qualitative methods and also an emergence of mixed methods in tax compliance research. Yong et al. (2019) study shows that out of the 713 studies reviewed, 34 of the studies were qualitative in nature and 20 were mixed methods studies. The balance of the studies was quantitative.

Although there has been a gradual increase in the utilisation of mixed methods, this method is still underutilised potentially due to resistance to deviate from the traditional manner of conducting tax compliance research which intuitively implies applying a quantitative approach, despite that some posed research questions may be better addressed by taking a mixed methods approach. Researchers' limited understanding of applying mixed methods could also explain the reluctance to take a mixed methods approach. Fortunately, in recent times the emergence of specialist journals such as the *Journal of Mixed Methods Research* reinforces the application of mixed methods to research as an acceptable and credible approach to research. Such journals also allow researchers to share their experiences and lessons gained from applying mixed methods to their research. The sharing of knowledge and lessons learned encourages methodological development and this article contributes

Reference this paper: Monageng, N., 2023. Using Mixed Methods to Understand Tax Compliance Behaviour, 21(1), pp.43-53, available online at www.ejbrm.com

to the development of mixed methods approach, to not only the tax compliance behaviour research but to business and management research as a whole, by sharing knowledge and lessons learned.

The complexity of understanding tax compliance behaviour may necessitate that researchers also explore mixed methods in order to address research questions that cannot be answered by only employing quantitative methods. A mixed methods approach provides an opportunity for a holistic view of tax compliance behaviour given, as noted by Alm et al. (2012), there are still major gaps in our understanding of tax compliance behaviour.

The remainder of this article is organised as follows: A brief review of mixed methods research is provided and also a discussion on the use of this approach in tax compliance research. An example of using mixed methods in tax compliance research is provided followed by an outline of the value gained in using mixed methods and how challenges in mixing methods can be faced. The last section sets out the conclusion.

2. What is Mixed Methods Research?

The mixed methods approach to research is a fairly new methodology, which originated in the late 1980s and 1990s. The approach is defined as:

> *"… the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration" (Johnson, Onwuegbuzie and Turner, 2007.p123).*

Taking a mixed method approach to research can assist in addressing inherent shortcomings in qualitative and quantitative approaches as often times the shortcomings of one may be addressed by the strength of the other. One shortcoming of the quantitative approach is that the results or information obtained may be too general to be applied to a specific setting. This shortcoming can be addressed by one of the strengths of applying a qualitative approach which is the approach's usefulness in providing in-depth understanding of the topic or specific setting being studied (Johnson and Onwuegbuzie, 2004). Some of the shortcomings of a qualitative approach, such as that the findings may not be generalised to the wider population or the difficulty of testing hypotheses using this approach (Johnson and Onwuegbuzie, 2004), can be addressed by applying a quantitative approach which allows a researcher to test hypotheses the results of which can often be generalised to a wider population.

The manner in which methods are mixed is also driven by the purpose of undertaking a mixed-methods approach. There are different purposes and advantages of mixing methods:

- Triangulation – use of one set of results with another in order to enhance validity
- Complementarity – elaboration or clarification of one set of results with findings from another method
- Development – use of one set of results to develop the use of another method
- Expansion – seeking to expand the range of inquiry by using different research methods for different inquiry components (Molina-Azorin, 2016)

In recent years mixed methods have been used in a few tax compliance behaviour studies. These studies adopted different approaches in combining qualitative and qualitative methods. Table 1 shows some of the recent tax compliance studies that have adopted a mixed methods approach.

Table 1: Tax Compliance Studies

	Data collection method		**Data analysis**	
Study	**Qualitative method**	**Quantitative method**	**Qualitative method**	**Quantitative method**
Saulitis and Chapkovski, 2023	Field experiment		Qualitative content analysis	Quantitative content analysis; Statistical analysis
Muharremi, Salé & Hoxhaj, 2022	Literature review	Survey		Statistical analysis
Angeles, 2021	Interviews	Survey	Thematic and joint-result analyses	Binary logistic regression

	Data collection method		Data analysis	
Study	**Qualitative method**	**Quantitative method**	**Qualitative method**	**Quantitative method**
Kurniawan, 2020	Interviews	Survey	Detail not provided in study	Statistical analysis
Sebele-Mpofu, 2020	Interviews and document review	Survey	Detail not provided in study	Statistical analysis
Bornman and Ramutumbu, 2019	Structured Questionnaires		Thematic analysis	Descriptive statistics
Enachescu et al., 2019	Focus groups	Experimental online survey	Thematic analysis	Statistical analysis
Sobhkhiz et al., 2019	Semi-structured interviews and Survey -containing open-ended questions	Survey – containing close-ended questions	Detail not provided in study	Partial least squares path modeling
Alleyne and Harris, 2017	Survey – containing close-ended and open-ended questions		Thematic analysis	Multivariate analyses
Rosid, Evans and Tran-Nam, 2016	Interviews	Survey	Thematic analysis	Structural equation modeling
Isa, 2014	Focus group	Survey	Thematic analysis	Statistical analysis
Mohdali and Pope, 2014	Interviews	Survey	Detail not provided in study	Statistical analysis
Lavermicocca and McKerchar, 2013	Interviews	Survey	Thematic analysis	Statistical analysis
Lozza et al., 2013	Focus groups	Survey	Thematic analysis and CAQDAS lexical analysis	Statistical analysis

In most of these studies data is collected using mixed methods and a combination of interviews and surveys appear to be a popular choice for researchers. The dominant method is often the quantitative method with limited focus on justifying the choice of method, providing detail on how data was analysed and presenting results of the qualitative method. The dominance of the quantitaive method is perhaps expected as researchers have largely applied quantitative approaches in studying tax compliance and are perhaps reluctant from moving away from this established practice due to some academic journals' unwillingness to publish tax compliance studies that do not present quantitative results. This stance often misses the point that it is not a matter of qualitative method versus quantitative method but rather whether the observed results of a study are a valid representation of the phenomenon being studied and whether the posed research question(s) can be better addressed by adopting a different research approach (Malina, Nørreklit and Selto, 2011).

Development of the quantitive method is often the reason for mixing methods (Enachescu et al., 2019; Isa, 2014; Lavermicocca and McKerchar, 2013; Muharremi, Salé & Hoxhaj, 2022; Rosid, Evans and Tran-Nam, 2016; Sobhkhiz et al., 2019). This perhaps also explains the disproportionate presentation of results, with the focus on presenting results from the quantitative method and a superficial presentation of the results from the qualitative method.

An important aspect of adopting a mixed-method approach is the intergration of the qualitative and quantitative methods (Molina-Azorin, 2016). The qualitative and quantitative results in tax compliance studies that adopted mixed methods are often presented separately and intergration is only evident in the discussion section of the study where an attempt is made to combine qualitative findings with quantative findings.

Mixed method research has a number of advantages, apart from the advantage of allowing a researcher address the shortcomings of one method with the strenght of another method, it can be seen as a a superior method in adequately addressing research questions compared to applying a single research method (Chen Loo, McKerchar

and Hansford, 2009). It also provides a comprehensive understanding of the research topic. This added value is noted in a number of these tax compliance studies (Enachescu et al., 2019; Kurniawan, 2020; Saulitis and Chapkovski, 2023; Sebele-Mpofu, 2020). There is also an understanding that "qualitative research in economics should not be treated as direct empirical evidence but rather offers a critical perspective on standard theoretical assumptions and presenting alternative viewpoints."(Saulitis and Chapkovski, 2023. p. 8).

Like any other approach to research conducting mixed method research has a number of challenges, Firstly, there is a requirement that the researcher should have a good understanding of both qualitative and quantitative approaches. Secondly, combining methods can be time consuming as each method has its individual processes that need to be followed. Thirdly, presenting the results of a mixed methods study can be difficult, particularly for researchers aiming to publish their research in an academic journal. The unbalanced reporting of results in tax compliance studies published in academic journals is evidence of this challenge. Lastly, there is challenge in ensuring meaningfully integrating of the qualitative and quantitative methods.

In spite of its challenges, however, there has been a growing appreciation for the fact that the mixed methods approach is a methodology that can offer the needed scope for researchers who are focused more on the purpose of the research question than on being confined to one paradigm (Creswell and Creswell, 2018; McKerchar, 2008).

The sections that follows provides an example of how mixed methods were applied in tax compliance research, with a focus on the added value that was gained from mixing methods and how methodological challenges can be faced.

3. An Example of Mixed Methods in Tax Compliance Research

The aims of this article are to highlight the added value of mixing methods and how methodological challenges can be faced. The study began in 2017 as part of a PhD research project. The principal objective of the project was to consider whether, and if so to what extent, personal taxpayers might be positively influenced so far as their tax compliance behaviour was concerned, by so-called 'nudging'. Nudge theory is a concept in behavioural economics that proposes positive reinforcement and indirect suggestions as ways to influence the behaviour and decision-making of groups or individuals. Nudging contrasts with other ways to achieve compliance, such as through education, legislation, regulation or enforcement.

The study was particularly concerned with the structural and content attributes of nudge messages and the timing of reciprocity nudges, and sought to determine the association between reciprocity nudges and tax compliance behaviour and to determine whether the effect of reciprocity nudges differs based on factors such as gender, income level, population group of the taxpayer, attitude towards tax and perception of corruption.

The data were collected in sequential stages using mixed methods. The first stage involved conducting a qualitative content analysis, followed by a quantitative content analysis and the last stage involved conducting a laboratory experiment (quantitative).

Therefore, the motive of mixing methods was for development and expansion purposes. For development purposes, the results from the qualitative content analysis phase were used for the purposes of conducting the quantitative content analysis. The data obtained from the quantitative content analysis were then embedded in the quantitative experimental design. Figure 1 illustrates the design.

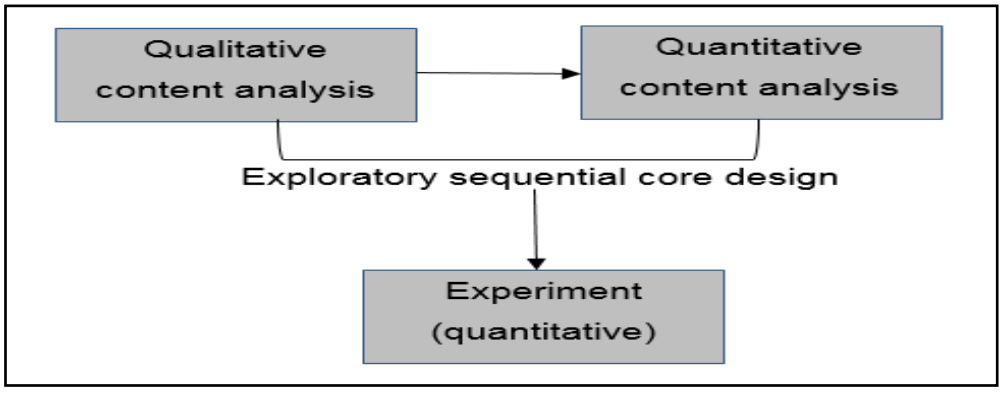


Figure 1: Mixed Methods Design Used in the Study

For expansion purposes, one of the research questions of the study was " what structural and content message attributes are appropriate when using reciprocity to encourage voluntary tax compliance?" This research question demanded that a qualitative analysis of academic literature be conducted in addition to a laboratory experiment.

The researcher adopted this design as a large majority of academic research conducted on the influence of nudge messages on tax compliance behaviour applied only quantitative methods in examining the association between these two variables: see Antinyan and Asatryan (2020) for a meta-analysis of studies on nudging and tax compliance. This leaves open questions on what role is played by the structural and content attributes of the messages in the effectiveness of the nudge and whether the observed results in relation to this research area are a true representation of the phenomenon being studied? As noted by Malina, Nørreklit and Selto (2011), quantitative accounting research could be examined meaningfully by applying complementary qualitative methods in the design. The same holds true for tax compliance research in that complementary, expansionary, developmental qualitative methods could assist in providing a complete picture of the phenomenon being studied.

The following paragraphs will briefly explain the process undertaken for each of the three methods applied (qualitative content analysis, quantitative content analysis and quantitative experiment). This is followed by a discussion on the value gained by mixing methods and how challenges in mixing methods can be faced.

3.1 Qualitative Content Analysis

A qualitative content analysis method was applied first in order to determine the structural and content attributes of effective nudge messages when using audio-visual media as a mode of delivery. We set out to determine this as prior research has indicated that structural and content attributes are important aspects of an effective message (Morgan et al., 2003) and similar research that has examined the relationship between tax nudges or messages and tax compliance behaviour has reduced the understanding of this relationship to numerical values rather than attempting to understand whether the structure and content of the messages are appropriately designed for the intended target audience.

The qualitative content analysis was conducted on academic literature. The aim of the qualitative content analysis was not to provide a comprehensive analysis of literature but rather to identify structural and content message attributes that have been found to be likely to draw a person's attention to the message being communicated and thus translate into change in behaviour.

An inductive content analysis process was followed by reading through the selected academic material to identify attributes from the data. Academic literature was initially identified by searching databases. The snowballing technique, this is a technique in which subsequent relevant literature is identified from the initial literature, was also used to identify other relevant literature. The attributes (results from this qualitative content analysis) identified through this process were then used to create a code book for the quantitative content analyses stage which later informed the design of the laboratory experiment. The qualitative content analysis method also contributed to the expansion of the inquiry by addressing the research question on what structural and content message attributes are appropriate when using reciprocity to encourage voluntary tax compliance. Incorporating this qualitative method provided a holistic view of potential moderating factors that could affect the strength of the relationship between reciprocity nudge messages and tax compliance behaviour.

3.2 Quantitative Content Analysis

After conducting a qualitative content analysis as explained above, a quantitative content analysis was conducted. The quantitative content analysis was conducted on 12 videos used by the South African tax authority in a "Touching Lives" campaign that was run during the period 2012-2013. Although the campaign was broadcast on television in the 2012-2013 period, at the time of conducting this research the videos were still publicly available, and therefore current, on the tax authority's YouTube channel.

The quantitative content analysis process entailed developing a coding scheme. The coding scheme was developed using the attributes identified through the qualitative content analysis. Each of the 12 attributes – which are: captures attention, new information, evokes emotion, credible source, goal framing, credible message, message clarity, message sidedness, message efficacy, and fits with prior knowledge – was assigned codes for coding purposes. Therefore, the data collected in the qualitative content analysis stage were used for the development of the quantitative analysis stage. Two attributes (directs attention and duration of exposure) were not included in the development of the coding scheme as they were difficult to measure in this study.

As a result of the quantitative content analysis process, two videos were selected for the purpose of designing the experimental stage of the research. We selected one video which contained most structural and content attributes of an effective message and one video which contained the least attributes of an effective message in order to examine the role played by structural and content attributes in nudge messages and test the hypothesis that in order to be more effective a reciprocity nudge message must contain most structural and content attributes of an effective message.

3.3 Experiment

The third and final stage was to design and conduct a laboratory experiment using the results from the quantitative content analysis process. A post-test laboratory experimental design was used as illustrated in Table 2.

Table 2: Experimental Design

Random assignment	Groups 1-4	Treatment	Observe
	Group 5	Control	Observe

Source: Adapted from (Leedy and Ormrod, 2014)

The experiment simulated the general environment of a voluntary tax reporting system. The participants in the experiment earned income by performing tasks. They then decided on how much (if any) of the income earned to declare; as a result, taxes were only paid on income which the participants had reported to the tax authority.

Watching the two videos selected in the quantitative content analysis stage formed part of the tasks that the participants in the treatment groups had to perform. Participants in the treatment groups were required to watch one of the two videos as part of a series of other videos. After watching each video, they were required to complete a task which entailed answering questions related to the videos watched and which would potentially earn the participants income (a performance-based payout or reward, based on their answers).

Figures 2 and Figure 3 show the steps followed by the participants in the treatment groups. Each of the videos selected through the quantitative content analysis stage is circled in the Figures 2 and 3 respectively to show how the videos were used in the experimental design.

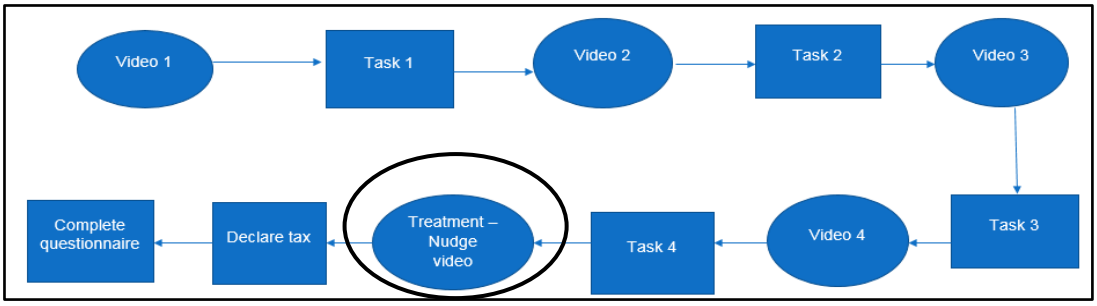


Figure 2: Treatment group A

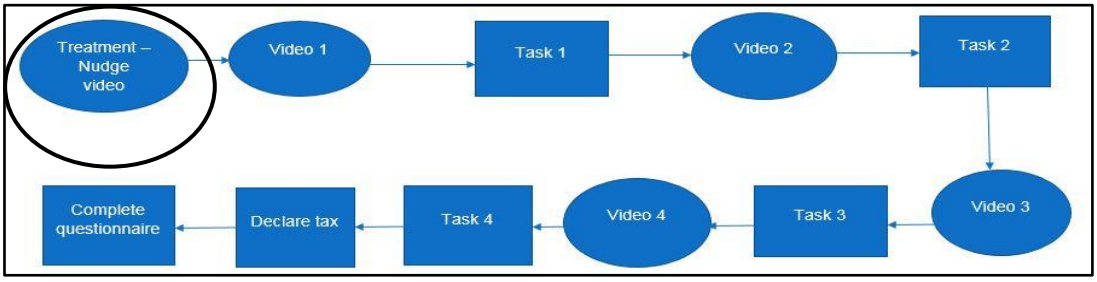


Figure 3: Treatment group B

4. Value Gained by Mixing Methods

4.1 Test and Build Theory

As noted by Allen (2017), although content analysis can be useful in describing messages, it cannot be used as a method to determine cause-and-effect of a message. However, this method can be combined successfully with other methods in order to determine cause-and-effect. In planning and executing the research project, advantage was taken of this strength. As shown in the previous sections, the researcher utilised a qualitative and quantitative content analysis to develop an experiment in which cause-and effect could be tested.

The qualitative content analysis process provided rich data which provided an understanding that in order to be more effective nudge messages should contain certain structural and content attributes. This informative insight could have not been obtained by applying a quantitative method. The results of the content analysis process enabled the researcher to conduct an in-depth interrogation of the 12 videos in order to determine whether they contained these identified attributes of an effective nudge message. Thus, the qualitative content analysis was not conducted merely as a means to select videos for the experiment but formed part of one of the objectives of the research which was to make recommendations about the structural and content attributes of a nudge message. By implementing a mixed method approach the researcher was able to both test theory and also build on theory. The theory was tested by conducting the experiment and testing the causal relationship between nudge messages and tax compliance behaviour; while conducting the qualitative content analyses allowed for building on theory by highlighting the structural and content attributes that ought to be incorporated into nudge messages delivered using audio-visual media.

4.2 Improved Data Collection Instrument

Another value gained by mixing methods is that the data from one method can be embedded in another method in order to improve a data collection instrument. The rich data obtained through the qualitative content analysis allowed a deeper understanding of the role of structural and content message attributes and that understanding informed and guided the selection of nudge messages that were used in the collection of data in the experiment stage. Evidence on the effect of nudge messages on tax compliance behaviour is often mixed with lack of effort invested in establishing the root cause of these differences. Mixing methods enables a researcher to attempt to establish the existence of moderating factors that could affect the relationship between variables. By incorporating the role of structural and content attributes the researcher was able to test the effect of this factor on the relationship between nudge messages and tax compliance behaviour thus increasing the validity of the quantitative results. A few tax compliance studies have acknowledged this value of improving data collection instruments by adopting a mixed methods approach (Enachescu et al., 2019; Isa, 2014; Lavermicocca and McKerchar, 2013; Muharremi, Salé and Hoxhaj, 2022; Rosid, Evans and Tran-Nam, 2016; Sobhkhiz et al., 2019).

4.3 Addressed Methodological Weaknesses

The advantage of a content analysis is that this method can be applied to various types of communications (Allen, 2017). This allowed the researcher to conduct a content analysis on two different types of communications (written and audio-visual). The data collected from the analysis of the written communication were then used in the analysis of the audio-visual communication. By undertaking this process, the researcher was able to address one of the research objectives which allowed for identification and description of specific attributes that are important in ensuring the effectiveness of nudge messages delivered using audio visual media. This in-depth understanding could not have been obtained by solely conducting an experiment.

Conducting the laboratory experiment then allowed the researcher to test whether the message that contained most of the structural and content attributes for an effective message, as identified through the qualitative content analysis process, was likely to be more effective message in influencing tax compliance behaviour compared to the message that was lacking in these attributes. The experiment allowed for testing of cause-and-effect which could not be done by merely conducting a content analysis. Mixing the methods thus permitted the researcher to address the weaknesses of the methods applied.

4.4 Comprehensive Understanding of Research Results

Adopting a mixed methods approach allowed for a comprehensive understanding of results observed from study. The researcher was able to explain the reason for one nudge message being more effective in positively influencing tax compliance behaviour compared to another nudge message. This allowed the researcher to make recommendations about the design of nudge messages aimed to change tax compliance behaviour. The ability

a of mixed method approach in comprehensively understanding of research findings is noted in other tax compliance studies:

> *"With the mixed method, research results can be explained comprehensively. The results of this study are explained using quantitative methods and corroborated by qualitative methods"(Kurniawan, 2020. p60)*
>
> *"Our mixed-methods approach delivers informative insights for a broader set of taxpayers"(Saulitis and Chapkovski, 2023.p18)*
>
> *"Taken together, both studies provide a detailed picture of the role of emotions in tax compliance."(Enachescu et al., 2019.p3)*
>
> *"This approach is useful when the aim is to comprehend cultural meanings, values, beliefs, implicit norms and behaviours of a target population for a topic that is very specific and circumscribed, such as tax compliance."(Lozza et al., 2013.p59).*

5. How Challenges in Mixing Methods can be Faced

The discussion so far has highlighted the value gained by mixing methods. However, despite these benefits the approach of mixing methods comes with challenges. Firstly, there is a requirement that the researcher should have a good understanding of both qualitative and quantitative approaches. This challenge was faced by investing time in obtaining a good understanding of the different methods. This was done by engaging methodology literature with a focus on the selected methods for the study. Upskilling through attending research courses is also beneficial. Another manner in which this challenge can be faced is by adopting a team approach to the research by inviting co-researchers that have the missing research method expertise for collaboration.

Secondly, mixing methods was time consuming as each individual method required time in designing the data collection instrument and collecting the data. The qualitative content analysis process required a significant time investment in searching through databases for relevant academic literature. The process of reading through the selected literature and identifying concepts/attributes demanded time. For the quantitative content analysis, majority of the time was spent on development of the coding scheme. The experiment was also particularly time consuming as the design of the experiment needed to be carefully done and tested. As this was a laboratory experiment there were a number of logistical considerations that needed to be organised. These challenges were faced by planning the research project to identify tasks that should be carried out by the researcher and those that can be delegated to research assistants. For mixed method studies that do not necessitate sequential data collection, qualitative and quantitative data can be collected concurrently to elevate time constraints .

Thirdly, presenting the results of a mixed methods study can be difficult, particularly for researchers aiming to publish their research in an academic journal. This is because of the word limitations of the journals which make it difficult for a researcher to comprehensively present their research. This leads to superficial presentation of results, particularly qualitative results, and lack of justification for methods adopted. This challenge also leads to lack of apparent integration of the qualitative and quantitative results. To overcome this challenge, researchers may decide to present their data from the different methods separately. However, identifying academic journals that embrace mixed-method research is preferable as it grants the researcher an opportunity to holistically present their research results. This also gives researcher an opportunity to clearly show the integration of results from the different methods applied in order to give understanding and context to the observed findings.

Lastly, adequately addressing inherent concerns related to validity and reliability for each of research methods is vital to mixed methods research. Failure to address these concerns will cast doubt on the validity of the results and conclusions drawn based on the results observed. This challenge was overcome by paying attention to the validity and reliability concerns of the individual methods. For the qualitative content analysis, generally, this method is known to suffer from issues of internal validity. This was addressed by analysing literature that was searched for in reliable databases. A detailed description of the structural and content attributes of an effective message was provided, in order to give other researchers an understanding of the attributes identified in the content analysis process, standard processes that apply when conducting a content analysis as a research method were outlined and followed. The content analysis of the academic literature in the study was performed by one person and even though the coding was rechecked to ensure that the data were correctly analysed,

recommendation is that more than one person should conduct the analysis as this is likely to enhance internal validity.

For the quantitative content analysis, in order to improve the coding scheme before final coding took place, a pilot test was conducted. The pilot test was conducted by the author and an independent coder, using coding forms. The intercoder reliability was determined in order to identify coding issues that needed to be rectified before commencing with the final coding process. This process was followed to ensure reliability.

In addition, the percentage agreement and the Krippendorff's alpha were calculated to determine the intercoder reliability of the final coding process. Percentage agreement is a common coefficient used to assess coder reliability in content analysis (Neuendorf, 2017). Krippendorff's alpha is also often used to assess reliability. Similar to the recommendation on the qualitative content analysis process, the use of more than one coder is recommended and is common practice when conducting a content analysis. Although there is a risk that different coders might interpret codes differently, having clearly defined codes will mitigate this risk (Rose, 2016). In addition, proper training of coders is essential as it will ensure intercoder reliability.

When conducting an experimental research study, internal and external validity is important. Without internal validity, a researcher will find it difficult to draw conclusions about the cause and effect relationship between the variables being studied (Leedy and Ormrod, 2014) and without external validity the researcher may find it difficult to generalise the findings of a study to settings outside the experimental settings. Laboratory experiments allow a researcher to have greater control over aspects of the experiment, such as selection of participants thus improving the internal validity of the research. However, in laboratory experiments external validity is more challenging to establish (Saunders, Lewis and Thornhill, 2012). Mixing methods allowed the research to address this issue of external validity as adopting a mixed method design in research may help with generalisability of a study (Saunders, Lewis and Thornhill, 2012). The experiment is preceded by a qualitative and a quantitative content analysis, which provides a more complete understanding of the influence of reciprocity nudge messages on tax compliance. To address the internal validity concerns, participants in the study were randomly allocated to the treatment groups and control group. Events occurring concurrently with the experiment may also pose a threat to internal validity. To reduce this threat, the experiment was conducted in one setting (computer labs at a university in South Africa).

6. Conclusion

This article highlighted the emerging use of mixed methods in tax compliance behaviour research and noted a few studies that have adopted different approaches in combining methods. Despite this emerging trend of applying mixed methods to compliance research, there are no known studies that have summarised this methodological approach for tax compliance researchers. Through the use of a study on the association between reciprocity nudge messages and tax compliance behaviour, this article assisted in illustrating how different research methods can be combined in order to address research objectives and provide an in-depth view of tax compliance behaviour.

The value obtained from mixing research methods is discussed. The first value identified is the ability to test and also build theory. In the research project example presented in this article, theory was tested by using a quantitative method to test the causal relationship between nudge messages and tax compliance behavior. Conducting a qualitative method also allowed for building on theory.

The second value identified is that one method can be used to inform and guide collection of data in another method. In the research project example, data obtained from a qualitative and a quantitative content analysis were used to inform and guide the collection of data in an experimental method.

The third value, a mixed methods approach can assist a researcher with addressing weaknesses in other methods. Finally, mixing methods provides a researcher with a deeper understanding of tax compliance behaviour. In the research project example discussed in this article, clearly identifying and understanding structural and content attributes of an effective message allowed for an appreciation of the role of these attributes as moderators in the relationship between nudge messages and tax compliance behaviour. This knowledge will better equip those tasked with implementing such strategies to be attentive to the appropriate structural and content attributes needed in a nudge message aimed at changing tax compliance behaviour.

As with any research method, mixing methods presents some challenges such as the requirement that the researcher should have a good understanding of both qualitative and quantitative methods, combining methods can be time consuming and researchers may have some challenges in presenting and publishing results from

mixed methods research. Notwithstanding these challenges, mixing methods provides tax researchers with an opportunity for an enriched understanding of tax compliance behavior.

References

Allen, M. 2017. *The SAGE Encyclopedia of Communication Research Methods*. Thousand Oaks, California: Sage.

Alleyne, P. and Harris, T., 2017. Antecedents of taxpayers' intentions to engage in tax evasion: Evidence from Barbados. *Journal of Financial Reporting and Accounting*, 15(1), pp. 2-21.

Alm, J., Kirchler, E., Muehlbacher, S., Gangl, K., Hofmann, E., Kogler, C., and Pollai, M., 2012. Rethinking the research paradigms for analysing tax compliance behaviour. *CESifo Forum*, 13(2), pp. 33-40.

Angeles, I. T., 2021. Categorizing Taxpayers: A Mixed-Method Study on Small Business Tax Compliance in the Philippines. *The South East Asian Journal of Management*, *15*(2),pp. 212-234.

Antinyan, A. and Asatryan, Z., 2020. 'Nudging for tax compliance: A meta-analysis', CESifo Working Paper No. 8500. Available through: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3680357 [Accessed: 6 June 2022].

Bornman, M. and Ramutumbu, P., 2019. A tax compliance risk profile of guesthouse owners in Soweto, South Africa. *The Southern African Journal of Entrepreneurship and Small Business Management*, *11*(1), pp. 1-14.

Chen Loo, E., Mckerchar, M. and Hansford, A., 2009. Understanding the compliance behaviour of Malaysian individual taxpayers using a mixed method approach. *Journal of the Australasian Tax Teachers Association,* 4(1), pp. 181-202.

Creswell, J. W. and Creswell, J. D., 2018. *Research design : qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, California: Sage.

Dallyn, S., 2017. An examination of the political salience of corporate tax avoidance: A case study of the Tax Justice Network. *Accounting Forum*, 41(4), pp. 336-352.

Enachescu, J., Olsen, J., Kogler, C., Zeelenberg, M., Breugelmans, S. M. and Kirchler, E., 2019. The role of emotions in tax compliance behavior: A mixed-methods approach. *Journal of Economic Psychology*, 74, pp. 1-16.

Hodgson, H., Walpole, M. and McKerchar, M., 2009. Understanding Australian small businesses and the drivers of compliance costs: A grounded theory approach. *Australian Tax Forum*, 24(2), pp. 151 -178.

Isa, K., 2014. Tax complexities in the Malaysian corporate tax system: minimise to maximise. *International Journal of Law and Management*, *56*(1), pp.50-65.

Jackson, B. R. and Milliron, V. C., 1986. Tax compliance research: Findings, problems, and prospects. *Journal of accounting literature*, 5(1), pp. 125-165.

Johnson, R., Onwuegbuzie, A. and Turner, L., 2007. Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 1(2), pp. 112-133.

Johnson, R. B. and Onwuegbuzie, A. J., 2004. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7), pp. 14-26.

Kurniawan, D., 2020. The influence of tax education during higher education on tax knowledge and its effect on personal tax compliance. *Journal of Indonesian Economy and Business*, 35(1), pp. 57-72.

Lavermicocca, C. and McKerchar, M., 2013. The impact of managing tax risk on the tax compliance behaviour of large Australian companies. *Australian Tax Forum*, 28(4), pp.707-724.

Leedy, P. D. and Ormrod, J. E., 2014. *Practical research : planning and design* (10th ed.).Harlow, UK: Pearson Education.

Lozza, E., Kastlunger, B., Tagliabue, S. and Kirchler, E., 2013. The relationship between political ideology and attitudes toward tax compliance: The case of Italian taxpayers. *Journal of Social and Political Psychology*, 1(1), pp. 51-73.

Malina, M. A., Nørreklit, H. S. O. and Selto, F. H., 2011. Lessons learned: Advantages and disadvantages of mixed method research. *Qualitative Research in Accounting and Management*, 8(1), pp. 59-71.

McKerchar, M.,2008. Philosophical paradigms, inquiry strategies and knowledge claims: Applying the principles of research design and conduct to taxation. *eJournal of Tax Research*, 6(1), pp. 5-22.

Mohdali, R. and Pope, J., 2014. The influence of religiosity on taxpayers' compliance attitudes. *Accounting Research Journal*, 27(1), pp. 71-91.

Molina-Azorin, F. J., 2016. Mixed methods research: An opportunity to improve our studies and our research skills. *European Journal of Management and Business Economics*, 25,pp. 37-38.

Morgan, S. E., Palmgreen, P., Stephenson, M. T., Hoyle, R. H. and Lorch, E. P., 2003. Associations between message features and subjective evaluations of the sensation value of antidrug public service announcements. *Journal of communication*, 53(3),pp. 512-526.

Muharremi, O., Salé, M. J. and Hoxhaj, M., 2022. A Mixed-Methods Study of the Influence of Demographic Factors on Albanian Individual Taxpayers' Ethical Beliefs Surrounding Tax Compliance. *Business Ethics and Leadership*, 6(1), pp. 47 - 66.

Neuendorf, K. A. 2017. *The content analysis guidebook* (2nd ed.). Thousand Oaks, California: Sage.

Onwuegbuzie, A. J. and Leech, N. L., 2005. Taking the "Q" Out of Research: Teaching Research Methodology Courses Without the Divide Between Quantitative and Qualitative Paradigms. *Quality and Quantity : International Journal of Methodology*, 39(3), pp. 267-295.

Richardson, M. and Sawyer, A. J., 2001. A taxonomy of the tax compliance literature: Further findings, problems and prospects. *Australian Tax Forum,* 16, pp. 137-322.

Rose, G. 2016. *Visual methodologies : an introduction to researching with visual materials* (4th ed.). Thousand Oaks:Sage.

Rosid, A., Evans, C. and Tran-Nam, B., 2016. Do perceptions of corruption influence personal income taxpayer reporting behaviour: Evidence from Indonesia. *eJournal of Tax Research* , 14(2*)*, pp. 387-425.

Saulitis, A. and Chapkovski, P., 2023. 'Investigating Tax Compliance with Mixed-Methods Approach: The Effect of Normative Appeals Among the Firms in Latvia', MPRA Paper No. 116560. Available through: https://mpra.ub.uni-muenchen.de/116560/1/MPRA_paper_116560.pdf [Accessed: 31 March 2023].

Saunders, M., Lewis, P. and Thornhill, A., 2012. Research methods for business students. 12th ed. Harlow, UK: Pearson.

Sebele-Mpofu, F. Y., 2020. Governance quality and tax morale and compliance in Zimbabwe's informal sector. *Cogent Business & Management*, *7*, pp. 1- 21.

Sobhkhiz, M. R., Mehdizadeh Ashrafi, A., Mojibi, T. and Otadi, M., 2019. Accountability, Legitimacy, and Voluntary Tax Compliance: A Case Study in Tax Offices in Mazandaran Province of Iran. *International Journal of Organizational Leadership*, *8*, pp. 38-55.

Truong, D., Xiaoming Liu, R. and Yu, J., 2020. Mixed methods research in tourism and hospitality journals. *International Journal of Contemporary Hospitality Management*, *32*(4), pp. 1563-1579.

Yong, S., Lo, K., Freudenberg, B. and Sawyer, A., 2019. Tax compliance in the new millennium: Understanding the variables. *Australian Tax Forum*, 34(4), pp. 766-808.