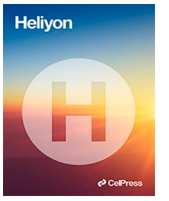
Heliyon 9 (2023) e19683

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



# Improving the quality of reporting findings using computer data analysis applications in educational research in context

Patrick Ngulube

*School of Interdisciplinary Research and Postgraduate Studies, University of South Africa, South Africa*

ARTICLE INFO

*Keywords:*
Computer packages for data analysis
Computer applications for data analysis
Educational research
Data analysis methods
Doctoral research
Methodological transparency

ABSTRACT

Data analysis is an important step in the research process as it influences the quality and standard of reporting research findings. Based on a review of the content of 255 doctorate theses, the use of computer applications for data analysis in educational research was assessed. It was feasible to assess how extensively used and accepted computer packages had become in educational research using an aspect of the diffusion of innovations theory as part of the conceptual framework. The results showed that the use of computer applications to analyse data was more prevalent among researchers using quantitative and mixed-methods research methodologies than among qualitative educational researchers. Educational researchers have not yet fully adopted innovative computer data analysis techniques in their research. It is evident that they use traditional technologies more than computer applications in their research. Name dropping of the computer applications used without employing the language or visualisations features provided by the applications was rife. This article bridges the gap between methodological scholarship and the use of computer applications in data analysis. It illuminates the potential of computer software to enhance the quality of the reporting of findings. The article aims to contribute to improvements in the standard of research reporting and the attributes of the graduates. The practical methodological advice in this article is aimed at guiding researchers who consider using computer packages in data analysis, irrespective of their methodological orientation. It stimulates debate on the use of computer applications in data analysis.

## 1. Introduction and background

Data analysis is at the heart of the interpretive work that researchers do, even though it is usually misunderstood and rarely discussed [1]. To strengthen the validity of an empirical study, data must be analysed properly [2]. The quality of the interpretation and reporting of research findings is influenced by data analysis. The nature of the data analysis being conducted determines the choice of the data analysis tools employed. That implies that the shortcomings of a research design cannot be made up for by data analysis technologies. The data analysis tools for qualitative, quantitative and mixed methods research data are different because they are based on different theoretical assumptions. The methodological perspective of a study dictates the approach to data analysis. Data can either be analysed manually or computationally.

Computer software packages are gradually replacing manual approaches to data analysis. Researchers should migrate from the traditional technologies such as notebook, pens, files, and file folders to computer applications. Data analysis approaches should be transformed to embrace technology since technology offers several advantages [3]. Researchers should adopt new technological

*E-mail address:* ngulup@unisa.ac.za.

https://doi.org/10.1016/j.heliyon.2023.e19683
Received 29 May 2023; Received in revised form 15 August 2023; Accepted 30 August 2023
Available online 31 August 2023

innovations to improve their practice. However, computers cannot perform miracles – they can help researchers to analyse their data. Computers cannot analyse data themselves [4]. Neither can they "replicate the complex capabilities of the human brain for conceptualisation of latent patterns of social behaviour" [5, p. 287], nor unravel the complexity of social phenomena. Computer packages are merely an "enabling technology" [6, p. 237). Researchers are essential to the data analysis process across all stages and oversee the establishment of all the relationships throughout the analytical process [7]. As Patton [8] points out: "Whether you do or do not use software, the real analytical work takes place in your head" (p. 531).

This article presents a study that employed content analysis to examine how and to what extent educational researchers use data analysis software. Several studies have looked at some aspect of data analysis in educational research. Hsu [9] examined the research procedures and data analysis approaches used in mathematics and science education research by conducting a content analysis of journal articles. Hsu [9], in contrast to the current study, solely focused on one aspect of the educational topic and did not take computer software packages into account when analysing the data. White, Judd and Poliandri [10] analysed journal articles to determine the use of qualitative data analysis software in the social sciences between 2000 and 2010. Disman, Ali and Barliana [11] assessed the use of quantitative research method and statistical data analysis in 106 dissertations, including 59 studies from seven subfields in education.

The current study examined the usage of computer packages across all major research methodological traditions. The study was more comprehensive than previous studies because it investigated research in many education subdisciplines, including chemistry education, curriculum studies, comparative education, education management, inclusive education, mathematics and technology education, philosophy of education and psychology of education. The study contributes to the theoretical and practical foundation that will ground the utilisation of innovative data analysis software packages to improve the standard of reporting research results, and to uplift the quality of education in line with the ideals of the Sustainable Development Goals of the United Nations.

Lastly, this study recognises that there are two levels of data analysis in line with Creswell and Creswell [18]. Organising, classifying, and summarising the acquired data at the first level entails applying descriptive techniques like coding, classifying, and sorting the data to find patterns and themes that characterise the dataset. The second level applying conceptual or theoretical frameworks on the data, finding connections, and explaining the results to gain a deeper understanding of the data, develop theories and produce new knowledge. This study focuses on the first level of data analysis in line with the conceptual framework of the study. Many studies do not seem to make that distinction [See, 9,11].

## 2. Statement of problem

The use of research methods generates data that must be analysed at two levels before it can be interpreted. The two levels of data analysis can be conducted manually or with the assistance of computer packages. Computer data analysis packages influence how qualitative, quantitative and mixed methods researchers analyse data. Computer software enables researchers to perform things that they could not previously do (or could not do as easily) without software [3,12,13]. White et al. [10], called for research to be conducted using content analysis to establish the extent to which computer software applications were used to analyse data. It is apparent that not much has been done in that regard. For instance, Hsu [9] examined research methods and data analysis procedures used by educational researchers. Their study focused on journal articles, but they did not assess researchers' (i.e., the authors of those journal articles) use of technology to analyse data. To date, limited empirical research has been carried out to understand the extent to which some educational researchers in postgraduate contexts exploit the advantages offered by computer packages in the analysis of their data.

The purpose of this study was to assess the use of computers for data analysis in educational research. To achieve this, the content of doctoral theses was analysed. The overarching concern was to gauge the extent to which the computer data analysis innovation was pervasive in educational research in line with the conceptual framework that informed this study. More specifically, this study aimed to answer the following questions:

- What are the prevalent research data analysis techniques used in educational research?
- How often are computer programs for data analysis employed in studies on education?
- To what extent have quantitative educational researchers used computer programs to analyse data?
- How often are CAQDAS programs used by qualitative educational researchers for data analysis?
- How commonplace are computer analysis programs in mixed-methods research on education?

## 3. Conceptual background

This section presents the conceptual framework and the literature review on the data analysis techniques and computer data analysis packages used by researchers employing quantitative, qualitative and mixed methods research methodologies.

### *3.1. Conceptual framework*

The variables that this study sought to examine could not be fully covered by any theory. Consequently, the study opted to use a conceptual framework as suggested by Ngulube [14]. The conceptual framework was formulated by combining ideas from the literature, which are discussed in the following sections, with one of the four dimensions of the diffusion of innovations theory [15]: the innovation itself. One sign of how diffusion happens is the rate at which an innovation is adopted in a community. The rate of

adoption will grow if many people adopt the innovation. People choose to adopt and implement a new idea or technology when it is introduced, or to reject it. The tipping point at which a novel concept achieves widespread acceptance and adoption is at the centre of the diffusion of innovations theory [16].

The study of the patterns of technological diffusion can provide insights into the advancement of using data analysis software in educational research. The methodology that was adopted outlined how the innovation had been diffused and how much progress that had been made. If it were not for the constraints of the research methods used in this study, all four domains that make up the theoretical foundation of the diffusion of innovations theory, and theories such as the technology adoption model employed by Salmona and Kaczynski [17], for example, could have been applied to the current research.

### 3.2. Data analysis methods used in research

Empirical studies are mainly based on quantitative, qualitative and mixed methods research methodologies [18]. These research traditions are dominant in the field of educational research [4]. Some of the data analysis techniques associated with each research methodology are illustrated in Fig. 1 [18,19].

These data analysis techniques can be conducted computationally or manually. The use of computers data analysis applications leads to improved data analysis. But there is still much to be done in terms of devising, adapting and systematically using sound and state-of-the-art approaches, techniques and tools for data analysis. Consequently, the general and systematic use of advanced methodologies for data analysis is certainly the 'new frontier' for educational research today [61 p. 364]. The same sentiments were expressed in the field of psychological research [62]. The statements still ring true. The following paragraphs give a brief description of software applications used in each research traditions to give a framework for analysis of the findings.

### 3.3. Software applications for quantitative data analysis

Quantitative researchers were the earlier adopters of computational analysis because their data analysis procedures were mainly formulaic, and computers could easily be programmed to perform such analyses. Qualitative researchers, on the other hand, require data analyses that are fluid by nature and demand interactive thinking [20]. Data analysis is a key methodological component of quantitative education studies because it affects "the validity of inferences: quality of constructed measures, proper handling of missing data, proper level of measurement of a dependent variable, and model checking" [21, p. 59). Statistical tests can be used as a tool to draw reliable conclusions from data. While the analysis can be done either manually or computationally, computational analysis can easily organise and transform the data to reveal the underlying patterns and relationships in relation to the development of a certain phenomenon [22].

There are several statistical analysis software packages that can be used by quantitative researchers to conduct specific statistical analyses, including Statistical Package for Social Sciences (SPSS®), SYSTAT®, NCSS®, STATGRAPHICS Centurion®, R® software, Statistical Analysis System (SAS®), STATA®, Statistica®, eViews®, MPlus®, UNISTAT®, XLSTAT®, Microsoft Excel® and AMOS® [23]. The thre[25][24]

In contrast to SPSS® and STATA®, SAS® is operated mostly by programming terminology rather than point-and-click menus, therefore researchers need some familiarity with the programming language to use it effectively [25].

### 3.4. Computer-assisted qualitative data analysis

Before the 1960s, qualitative researchers frequently used manual procedures to analyse their data [13,26]. It was not until the 1980s, when personal computers became ubiquitous, that qualitative researchers gradually embraced qualitative data analysis software [27]. Although researchers have adopted computer-assisted qualitative data analysis software (CAQDAS) packages as technologies advanced, their use is still subject to debate.

Considering the influence of the paradigm wars and the need to establish the legitimacy of qualitative research, qualitative researchers were apprehensive about the new technology-driven analytic procedures that seemed alien to their logic [20]. The perceived

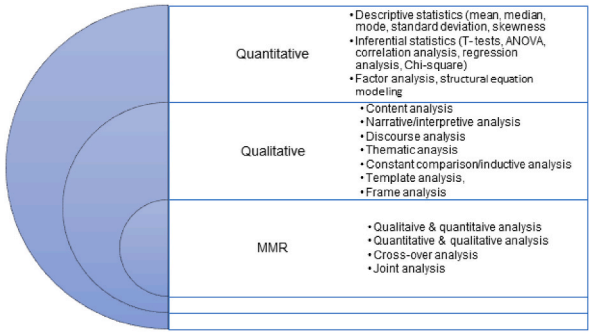



**Fig. 1.** Some data analysis methods associated with each research method.

over-reliance on software tools, particularly auto-coding tools, in the analysis and theorising processes caused critics to condemn qualitative data analysis software as mechanistic [28]. Some scholars posit that the use of qualitative data analysis software tools separates or distances researchers from their data [12,29]. In a nutshell, technology-driven analysis was perceived to lack complexity as it separated data from theory and was inflexible, constrictive and reductive [30].

It must be noted that software cannot take over the role of the researcher [29], and any fears that this would happen were caused by the "slight paranoia about technology" in the circles of some qualitative researchers [20, p. 1163). Whenever conducting qualitative research, it is crucial to remember that the researcher is the primary tool for analysis [31].

Despite persistent myths about CAQDAS in the literature, qualitative data analysis software offers a lot of potential advantages and is increasingly becoming popular [12,29]). CAQDAS enables researchers to perform things that they could not do (or could not do as easily) without software (12, 3, 20, 7). CAQDAS programs make it easier for qualitative researchers to gather, arrange, analyse, visualise and present their data [12]. In other words, CAQDAS can do the following:

- Make systematic and reliable data management possible.
- Eliminate manual tasks by automating some operations.
- Integrate multiple data types, and support multimedia and web-based sources.

CAQDAS packages include Atlas.ti®, Cassandre®, Dedoose®, f4analyse®, HyperRESEARCH®, MAXQDA®, MonkeyLearn®, NVivo®, Provalis Research Text Analytics Software®, QDA Miner®, Quirkos®, Taguette® and Transana® [13,32].

### *3.5. Software packages for mixed methods research data analysis*

Mixed methods research (MMR) has grown in popularity [18]. However, software design has taken a while to catch up with the expansion of the field, and the subject of data analysis of mixed methods data has been neglected [32–34]. Consequently, many MMR researchers use various types of computer applications in a single study. For example, they use SPSS® and NVivo® to conduct MMR data analysis "concurrently and consecutively" [35]. Given that integration is the lifeblood of MMR, researchers run the danger of not integrating the data adequately by utilising a variety of computer applications [12,33,36]. General-purpose spreadsheets such as Excel® and Access® can also be used to combine qualitative and quantitative data [32].

Unlike other software packages that support one or two methods, MAXQDA® supports mixed methods research [35,36]. It also uses artificial intelligence to assist users with audio transcription [38]. MAXQDA® uses typology tables to combine numerical variables with thematic codes that have been converted into a variable [37]. Typology tables are useful in joint displays [39], which are essential to MMR data analysis.

Some other software packages are also currently being tested to gauge their usefulness as MMR data analysis tools. Dedoose®, a web-based MMR data analysis software package for data management, enables researchers to analyse qualitative and quantitative data sets together rather than as two separate components that need to be put together to achieve meaningful data integration and a better understanding of the relationships between the two aspects of the data sets [34,40]. Mixed methods researchers can also use Nvivo® to analyse MMR data [41].

## 4. Methodology

This study used qualitative content analysis (QCA) [42]. By systematically classifying the content through coding and identifying themes and trends, QCA aligns the coding frame to the content [42]. Concept-driven (inductive) and data-driven (deductive) categories are combined in QCA [43]. The current study followed an inductive approach as the concepts described in the preceding sections guided the coding of the content that had been analysed. In QCA, the design process starts by deciding what kind of content would be evaluated and what the place of theory in the research would be [44]. Many data sources are used for QCA, including books, journals, newspapers, magazines, radio broadcasts, interviews and speeches, conference proceedings, photographs, videos, and a broad range of textual, visual and audio materials [43].

### *4.1. Data sources and search strategy*

Journal articles seem to be a popular target for conducting QCA in the field of education. For instance, Hsu [9] and White et al. [10], used content analysis to investigate various aspects of scholarly communication in educational research. Journal articles were also employed by Merchant et al. [45], to analyse the research questions used in mixed-methods research in instructional design. Admittedly, other documents, such as theses and dissertations, might have contained material necessary to comprehend the phenomenon fully.

In contrast to some previous studies that mainly used journal articles, the objectives of the current study were achieved using doctorate theses produced in the South African context. Unlike theses, journal articles may have limitations regarding the extent to which research procedures are covered, depending on editorial restrictions. Doctoral theses were chosen as samples for this study in acknowledgment that doctoral education is the primary source of research productivity and innovation and is at the core of a university's research capabilities [46]. Munoz-Najar Galvez, Heiberger and McFarland [47], Karsli, Karabey, Cagiltay and Goktas [48] and Disman et al. [11], also used theses to assess scholarly communication patterns in the field of education.

The education discipline was chosen for the study because Wilkinson, Van Jaarsveldt, Grimsley and Seoka [49] revealed that

master's and doctoral studies in South Africa focused primarily on this field between 2006 and 2016. It was logical to begin the analysis in 2010 since that was the year in which the academic institution that serves as the context for this study established a separate college for the education field [50]. The cut-off date of 2016 was selected arbitrarily, but it exceeds the minimum specified five-year timeframe for identifying trends and patterns in scholarly communication [51].

Various strategies were employed to identify samples for the study. Phase one involved conducting searches in the *Directory of Open Access Repositories* [52]. There was no need to triangulate data sources as Ngulube [50] had established that data from the institutional repository were more reliable than those of national datasets such as the NEXUS database of the National Research Foundation. Following the identification of the repository, a corpus of 1240 dissertations were retrieved and extracted after browsing the collection of the educational community in the relevant institutional repository. Notably, the database that served as the data source did not distinguish between master's and doctoral studies.

### 4.2. Screening

While phase one involved planning and reviewing, phase two involved screening the studies that passed the initial assessment using six inclusion and exclusion criteria based on Kgoroeadira [53] (see Table 1). Each study was scanned to shortlist the dissertations that met the criteria outlined in Table 1. A total of 255 dissertations made the final list.

### 4.3. Analysis procedure

The 255 theses that advanced to the final stage of examination were coded. The theses were coded based on three descriptive characteristics and the indicators stated in Table 2. Two postdoctoral fellows coded a third of the shortlisted dissertations to ensure intercoder reliability. Any significant discrepancies in the coding were discussed and resolved. The coders initially agreed on their ratings by 75%. Later, the coders met to talk about and resolve their disagreements. No statistical analysis was performed to confirm intercoder reliability. Excel® version 10.0 was used to input the coded data. Excel® also facilitated the running of frequencies. A total of 79 theses were identified as having used some type of computer software package. These 79 theses were further subjected to analysis to identify the names and versions of the software applications used as well as the data analysis methods followed. The results were presented using frequencies and percentages.

## 5. Results

Results are organised according to the five research sub-questions that guided the study.

### 5.1. Data analysis in educational research

Table 3 shows that data were analysed in accordance with three methodological traditions. The quantitative methodology was less dominant. Data were analysed manually (176; 69%) and computationally (79; 31%). In contrast to other research traditions, qualitative researchers frequently utilised manual data analysis techniques. Fig. 2 illustrates that thematic analysis was the dominant technique of data analysis, followed by descriptive statistics. A total of 15 of quantitative (3) and MMR (12) studies used inferential statistics. Two quantitative theses indicated that the researcher had received technical support from a statistician in addition to guidance from a supervisor.

### 5.2. Using computer software packages to analyse data in educational research

Table 3 shows that a total of 79 theses used a total of 90 computer software packages. Table 4 shows a total of 60 (67%) instances of using SPSS®, followed by 11 (12%) incidences of ATLAS/ti®. The use of SAS®, EXCEL®, STATA®, NVIVO® and HyperResearch® was below double digits, as illustrated in Tables 3 and 4 Education management (37 theses or 41%) and curriculum studies (28 theses or

**Table 1**
Inclusion and exclusion criteria.

| Criteria | Exclusion | Inclusion |
|---|---|---|
| Topic and duplicates | Dissertations that were not in the field of education and duplicates were excluded. | Dissertations that were in the field of education were included (n = 1240). |
| Publishing date | Dissertations that did not indicate year of publication were excluded (n = 12). | Dissertations indicating year of publication were included (n = 1228). |
| Date of publication | Dissertations published before 2010 or after 2016 were excluded (n = 608). | Dissertations published between 2010 and 2016 were included (n = 620). |
| Level of study | Dissertations that were not at doctoral level were excluded (n = 333). | Dissertations at doctoral level were included (n = 287). |
| Language | Non-English language dissertations were not included (n = 32). | Dissertations published in English were included (n = 255). |
| Specification of research methods | Dissertations not indicating research methods were excluded (n = 0). | Dissertations indicating research methods were included (n = 255). |

**Table 2**
A guide to thesis analysis.

| Characteristics | Indicator | Question item |
|---|---|---|
| Methodology | Specify research methodology according to Creswell and Creswell [18] in one or all the following areas: abstract, chapter 1 and the methodology and data presentation chapters. | Research methodology stated (e.g., qualitative, quantitative and MMR) |
| Research procedures | Data analysis technique stated in one or all the following areas: abstract, chapter 1 and the methodology and data presentation chapters based on Fig. 1. | Data analysis procedures mentioned and explained (e.g., content analysis, text analysis, thematic analysis, types of descriptive or inferential statistics and joint displays) |
| Computer applications | Use of computer software package to analyse data in one or all the following areas: abstract, chapter 1 and the methodology and data presentation chapters. | (i) Stated use of computer software<br>(ii) Version of software package stated<br>(iii) Mention of how-to-do resources to understand the nature of software package<br>(iv) Data integration during analysis in MMR studies |

**Table 3**
The total number of times a software package is mentioned at least once in individual theses (n = 79).

| Methodology | Number | Software application | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | | SPSS® | SAS® | EXCEL® | STATA® | ATLAS/ti® | NVIVO® | HyperResearch® | |
| Quantitative | 29 (11%[a]) | 14 | 2 | 4 | – | 1 | – | – | 21 [**16**][b] (72%) |
| Qualitative | 142 (56%) | – | – | – | – | 8 | 1 | 2 | 11 [**11**] (8%) |
| MMR | 84 (33%) | 46 | 5 | 2 | 1 | 2 | 2 | – | 58 [**52**][b] (69%) |
| Total | 255 | 60 | 7 | 6 | 1 | 11 | 3 | 2 | 90 [**79**] (36%) |

[a] Percentages rounded to the nearest decimal place; bold figures in square brackets represent the actual number of theses.
[b] More than one software package could be used per thesis if more than two data analysis methods were used.

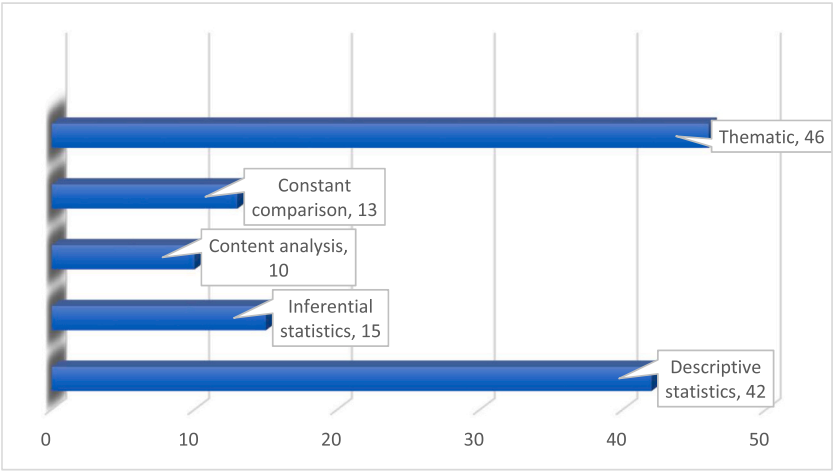



**Fig. 2.** Data analysis techniques.

31%) used computer packages more than other fields in education. A total of 74 theses described the type(s) of analyses that were done using a computer application without mentioning the specific software version that had been used. The version of the software programs that were used was indicated in three MMR theses, but only once in the two theses that used qualitative and quantitative research approaches. Only five theses used the language of the software applications they used and incorporated the visualisations features from the applications. Fig. 3 shows that researchers in education management used computer software packages more than researchers in any other subdiscipline of education.

### *5.3. Software packages used by quantitative educational researchers*

The findings reveal that 72% of quantitative educational researchers used computer applications for their data analysis. Tables 3 and 4 show that educational researchers mainly used SPSS® (14) to analyse quantitative data, followed by EXCEL® (4) and SAS® (2). Table 4 illustrates that education management (6) and inclusive education researchers (6) used quantitative computer applications more than any other subdiscipline in education. The use of EXCEL® was mentioned in the fields of curriculum studies (1), education management (1) and inclusive education (2).

**Table 4**
Number of theses per subdiscipline mentioning a software package at least once (n = 79).

| Sub-discipline | Methodology | Software application | | | | | | | Sub-Totals | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | | SPSS® | SAS® | STRATA® | EXCEL® | ATLAS/ti® | NVIVO® | HyperResearch® | | |
| Curriculum Studies | Quantitative | 3 | – | – | 1 | – | – | – | 4 | 28 |
| | Qualitative | | | – | – | 3 | – | – | 3 | (31%) |
| | MMR | 17 | 1 | 1 | – | 1 | 1 | – | 21 | |
| Chemistry Education | Quantitative | – | – | – | – | – | – | – | – | 1 (1%) |
| | Qualitative | | – | – | – | – | – | – | – | |
| | MMR | 1 | – | – | – | – | – | – | 1 | |
| Comparative Education | Quantitative | – | – | – | – | – | – | – | – | 2(2%) |
| | Qualitative | | – | – | – | – | – | – | – | |
| | MMR | 1 | 1 | – | – | – | – | – | 2 | |
| Education Management | Quantitative | 3 | 2 | – | 1 | – | – | – | 6 | 37 |
| | Qualitative | | – | – | – | 3 | 1 | 1 | 5 | (41%) |
| | MMR | 20 | 3 | – | 1 | 1 | 1 | – | 26 | |
| Mathematics Education | Quantitative | 1 | – | – | – | – | – | – | 1 | 4 (5%) |
| | Qualitative | | – | – | – | – | – | – | | |
| | MMR | 3 | – | – | – | – | – | – | 3 | |
| Psychology of Education | Quantitative | 3 | – | – | – | | – | – | 3 | 9 (10%) |
| | Qualitative | | – | – | – | 1 | – | 1 | 2 | |
| | MMR | 3 | – | – | 1 | – | – | – | 4 | |
| Inclusive Education | Quantitative | 4 | – | – | 2 | – | – | – | 6 | 9 (10%) |
| | Qualitative | | – | – | – | 2 | – | – | 2 | |
| | MMR | 1 | – | – | – | – | – | – | 1 | |
| Total | | **60 (67%)** | **7 (8%)** | **1(1%)** | **6 (7%)** | **11 (12%)** | **3 (3%)** | **2(2%)** | **90** | |

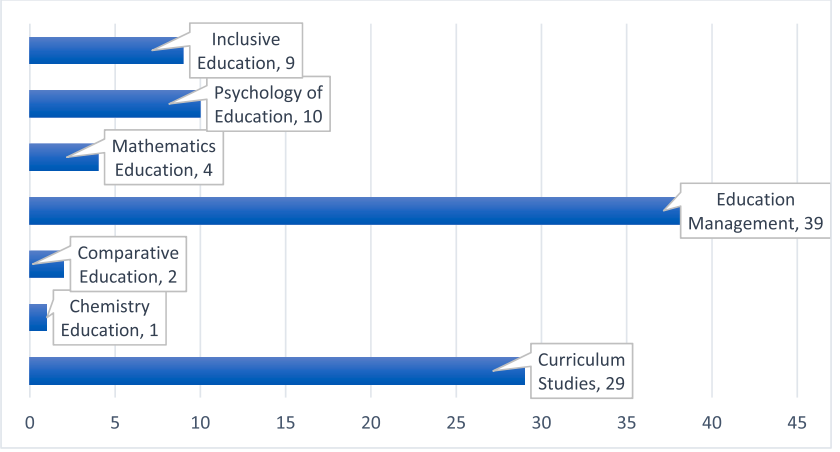



**Fig. 3.** Computer software packages used per subdiscipline.

### 5.4. Use of CAQDAS packages by educational researchers

Qualitative researchers in education use data analysis software programs less than their quantitative and MMR counterparts. Only eight per cent of the researchers used computer applications to analyse their data. Table 4 shows that education management researchers (5) are at the top of the pole in the use of CAQDAS, followed by curriculum studies (3). in contrast to subdisciplines such as curriculum studies and inclusive education, which only used ATLAS/ti®, education management researchers used a variety of CAQDAS, including ATLAS/ti®, NVIVO® and HyperResearch® (see Table 4). The use of HyperResearch® was also mentioned once in the field of psychology of education.

### 5.5. Software packages used by MMR educational researchers

Tables 3 and 4 show that MMR theses frequently used SPSS®, followed by SAS® (5) and EXCEL® (2). Most of the theses that used these three software packages were in education management (23), followed by curriculum studies (18). STRATA® was used once in curriculum studies. As shown in Table 4, 46 out of the 52 MMR studies (88%) employed quantitative software for analysis and

manually coded qualitative data. Only six theses used a software package to analyse the qualitative component of the data. Three were in education management, two in curriculum studies and one in educational psychology. Theses in curriculum studies (2) and education management (2) either employed ATLAS/ti® or NVIVO®, whereas both educational management and educational psychology employed EXCEL® once. Data were analysed separately in all the studies and integrated during the interpretation stage.

## 6. Discussion

Based on the findings, the discussion explains how deeply education researchers have embraced computer data analysis. The limitations of the study and areas for further research are given before the conclusion and recommendations of the study.

### 6.1. Data analysis in educational research

Data analysis was predominantly conducted manually. Educational researchers have not fully embraced the technological innovation of computer applications in data analysis. Data analysis techniques include thematic analysis, descriptive statistics, constant comparison, content analysis and inferential statistics, as indicated in Fig. 2. According to Creswell and Creswell [18], Creswell and Guetterman [4] and Lester et al. [19], these data analysis techniques are frequently used, along with the others that are depicted in Fig. 2.

Manual thematic analysis and descriptive analysis were dominant. Perhaps thematic analysis and descriptive analysis were prevalent because a study that triangulates qualitative and quantitative data can use both data analysis methods. The high frequency of thematic analysis and descriptive statistics may also have been explained by the incidence of MMR studies. A study conducted by Hsu [9] revealed that the most popular data analysis techniques used in education research included descriptive statistics and inferential statistics. The finding of the current study did not differ drastically from those of Hsu [9] because inferential statistics was also dominant.

The statistical analysis is an important component of students' research, and it is essential that students must undertake it themselves. Nevertheless, to prevent any biases, mistakes, or misinterpretations in the analytic process, Creswell and Creswell [18] advise involving an expert in data analysis. Their involvement requires adherence to a high ethical conduct so that the contribution of the doctoral candidate is not compromised. In that regard, there is a need for transparency on reporting the level of the involvement of an expert in the analysis of data.

### 6.2. Using computer software packages to analyse data in educational research

The use of computer applications was not prevalent in educational theses – the low levels of use were evident in qualitative research. The use of computers to analyse data shows some level of sophistication on the part of the researchers [3, p. xxv]. The results suggest that the use of innovative computer applications had not yet fully taken root in educational research. That implies that educational researchers did not fully take advantage of the standardised approach to data analysis offered by computer applications.

Many studies did not specify the version of the software application they had used. Specifying the software version would have given other researchers insight into how the limitations of the software and its features might have affected data analysis. As technology advances, updated software with new capabilities enhances system performance and job quality. It also gives researchers more options. For instance, using outdated software might result in poor user experience [54]. Neglecting to indicate the version of the software used has implications for methodological transparency as readers cannot determine which features and software capabilities were at the disposal of the researcher. Many theses resorted to name dropping of the computer applications without using the language or visualisations provided by the applications. Educational researchers can take advantage of the powerful visualization tools provided by computers to understand, explore and report their data.

### 6.3. Software packages used by quantitative educational researchers

Software and technological advances have influenced how research is conducted [12]. The fact that SPSS® was used in most of the studies confirms that it is a popular quantitative data analysis programme, as suggested in the literature [23,55]. Contrary to the assertion by Crossman [25], Excel® proved to be more popular than STATA® and (SAS)® in the studies that were analysed. Unlike researchers in educational management and curriculum studies, some educational researchers seemed not to fully exploit the potential advantages that the innovative computer data analysis packages present. The adoption levels for computer applications remain very slow considering that computer applications for quantitative data analysis have been used since the mid-1960s [33]. Levels of adoption should have been relatively higher considering the long history of computer applications for data analysis. The use of computer applications improves the standard of reporting and the level of inference. Computers applications facilitate the application of complex statistical and mathematical techniques that may not be feasible manually. Their use can also improve the standard of reporting of results in education research.

### 6.4. Use of CAQDAS packages by educational researchers

Findings revealed that educational researchers used CAQDAS packages such as ATLAS/ti®, NVIVO® and HyperResearch®. Their use confirms the popularity of these packages, as indicated by White et al. [10], and Wolff [38]. Based on the study of theses, the use of

CAQDAS does not seem to be widespread in education, although Friese, Soratto and Pires [56] stated that the use of CAQDAS packages was commonplace. This suggests that the researchers were unable to manipulate and query data in ways that went beyond traditional methods.

The low use of computer applications in qualitative studies might not be peculiar to the context of this study. For instance, a review of 72 research articles published in the *American Educational Research Journal* between 2014 and 2015 revealed that nine of the 24 articles that reported on qualitative research or mixed-method techniques used CAQDAS [57]. Leading qualitative methodologists are on record for being hesitant to use CAQDAS tools [27].

Qualitative data analysis tools allow for rigorous and reliable research [29]. CAQDAS programs also enhance the research process and provide researchers with access to insights they otherwise would not have gained by uncovering answers that remained "hidden in the data" [3, p. xxv). The limited use of qualitative data analysis software in the theses that were analysed implies that the rigor and reliability might be inadequate. Furthermore, the studies might also have been unable to keep a thorough audit trail of the process [12] a feature offered by CAQDAS programs. Failing to keep and provide a thorough record of the analysis process has implications for methodological transparency.

### 6.5. Software packages used by MMR education researchers

One of the most challenging aspects of MMR is data analysis [33,58]. Many of the software packages used to analyse MMR data in the sample studies did not have MMR assumptions. That has implications for data analysis and integration, the lifeblood of MMR studies. Quantitative and qualitative data were mainly examined independently using traditional analysis methods –only six theses used computer software to analyse the qualitative and quantitative data components. This implies that there was a lack of true data integration. The result of analysis was not usually "more than the sum of the individual quantitative and qualitative parts" [60, p. 8], and did not add any value equivalent to formula $1 + 1 = 3$ suggested by Fetters and Freshwater [59]. This confirms Bazeley's [32] assertion that "[p]ublished reports of studies that truly integrate qualitative and quantitative data sources in analysis are rare" (p. 72). Computer software that supports integration suggested by Bazeley [41] and Leech [55] include Dedoose®, R®, MAXQDA®, NVivo®, and QDA Miner®, but these had rarely been used in the theses that were analysed.

Computer applications can also assist in quantitizing qualitative data to facilitate quantitative analysis and non-parametric statistical analysis [55].[55][3]

### 6.6. Limitations of the study

The features of each software and how they work are beyond the scope of this research. There are manuals and texts that cover those aspects, and the aim was not address technical issues. Furthermore, software development is ongoing, therefore it would be counterproductive to include a tutorial in a piece of this nature. Software becomes obsolete as new versions with new features are developed. As Friese [3] points out: "Writing about software, however, is like football (soccer): after the game is before the game" (p. 297). This study looked at the use of computer applications in one discipline. Future research can consider other disciplines to determine the extent to which disciplinary norms affect the extent of the adoption of computer applications in research.

The methodological limitations of the use of one approach and a single case study are acknowledged. Therefore, the results cannot be generalised. There are obvious drawbacks to using only one research strategy. For instance, researchers' personal preferences based on their "history, prior research experiences", and "training programmes" [29, p.163) and available "how-to" manuals are determining factors when adopting data analysis technologies. This study was not able to establish whether these factors played a role in the low diffusion of data analysis software as the research methods used did not make provision for such findings. It was not possible to apply all the dimensions of the diffusion of innovations theory and other possible theories because the methodology was limited.

### 6.7. Areas for further research

The results should be regarded as preliminary as there is room to study more cases to develop a theory about the phenomena. The potential of mixed-methods research to lead to a deeper understanding of the phenomena should also be considered. The study did not examine long-term trends in the use of computer applications or the extent to which educational researchers had really used the analytical features. All these limitations are potential areas for further research. The study did not look at the second level of data analysis suggested by Creswell and Creswell [18]. The envisaged research will provide a deeper understanding of how the data is analysed to develop theories or explanations and generate new knowledge in educational research.

### 6.8. Recommendations

It is recommended that educational researchers should:

- Choose and employ appropriate data analysis tools and draw appropriate conclusions relating to their research questions.
- Exploit the potential of computer packages to support data analysis.
- Be cognisant of the versions of the software packages they use for data analysis for their capabilities to be appreciated.
- Avoid name dropping when it comes to the use of computer applications by employing the language and visualization functions provided by the relevant computer programmes

- Consider the potential advantages provided by computer packages, give reasons why they have not used such software to achieve methodological transparency and efficiency.
- Know that computers cannot replace them in data analysis as they will always be called upon to engage with data analysis as methodological choices cannot be made by technology.

## 7. Conclusions

Although software for qualitative data analysis appears to be becoming popular across many academic fields, educational researchers are not utilising it optimally. There was reliance on traditional data analysis methods. Less than half of the researchers used computer packages to analyse data. There was a reluctance among researchers in education to employ computerised data analysis tools. Studies rarely specified the versions of computer applications employed. In a nutshell, the innovation does not enjoy widespread acceptance in education research in the context of this study. It is evident that education researchers' practices are still rooted in traditional data analysis technologies. Educational management, quantitative and mixed-methods research studies paint an optimistic picture of the adoption of computer applications for data analysis innovation. The data analysis techniques highlighted in this article can help educational researchers to draw the best from empirical data, to provide well-founded and generalisable findings and recommendations, and to contribute to the accumulation of knowledge. This will serve as a practical guide for researchers who are beginning to use computer software in data analysis. The quality of reporting of findings from educational research will increase with the acceptance of the innovation of using computer applications.

## Author contribution statement

Patrick Ngulube, Ph.D: Conceived and designed the experiments; Performed the experiments; Analysed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

## Data availability statement

Data will be made available on request.

## Additional information

No additional information is available for this paper.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] H. Kara, Mixed methods data analysis, 2017. Retrieved 30 April fr, https://helenkara.com/2017/03/23/mixed-methods-data-analysis/.
[2] M.A.A. Abulela, M.M. Harwell, Data analysis: strengthening inferences in quantitative education studies conducted by novice researchers, Educ. Sci. Theor. Pract. 20 (1) (2020) 59–78, https://doi.org/10.12738/jestp.2020.1.005.
[3] S. Friese, Qualitative Data Analysis with Atlas.*ti*, third ed., Sage, 2019.
[4] J. Creswell, T. Guetterman, Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, sixth ed., Pearson, 2018.
[5] J.A. Holton, The coding process and its challenges, in: A. Bryant A, K. Charmaz (Eds.), The Sage Handbook of Grounded Theory, Sage, 2007, pp. 265–290.
[6] R. Lee, L. Esterhuizen, Computer software and qualitative analysis: trends, issues and resources, Int. J. Soc. Res. Methodol. 3 (3) (2000) 231–243, https://doi.org/10.1080/13645570050083715.
[7] J. Soratto, D.E.P. Pires, S. Friese, Thematic content analysis using Atlas.ti software: potentialities for research in health, Rev. Bras. Enfermagen 73 (3) (2019) 1–5.
[8] M.Q. Patton, Qualitative Evaluation and Research, fourth ed., Sage, 2015.
[9] T. Hsu, Research methods and data analysis procedures used by educational researchers, Int. J. Res. Method Educ. 28 (2) (2005) 109–133, https://doi.org/10.1080/01406720500256194.
[10] M.J. White, M.D. Judd, S. Poliandri, Illumination with a dim bulb? What do social scientists learn by employing qualitative data analysis software in the service of multimethod designs? Socio. Methodol. 42 (1) (2012) https://doi.org/10.1177/0081175012461233, 43-76.
[11] Disman, M. Ali, S.M. Barliana, The use of quantitative research method and statistical data analysis in dissertation: an evaluation study, Int. J. Educ. 10 (1) (2017) 46–52, https://doi.org/10.17509/ije.v10i1.5566.
[12] P. Bazeley, Qualitative Data Analysis: Practical Strategies, second ed., Sage, 2021.
[13] J.C. Evers, Current issues in qualitative data analysis software (QDAS): a user and developer perspective, Qual. Rep. 23 (13) (2018) 61–73, https://doi.org/10.46743/2160-3715/2018.3205.
[14] P. Ngulube, Theory and theorising in information science scholarship, in: Ngulube (Ed.), Handbook of Research on Connecting Research Methods for Information Science Research, IGI Global, 2020, pp. 18–39.
[15] E.M. Rogers, Diffusion of Innovations, fourth ed., Free Press, 1995.
[16] M. Conyers, D. Wilson, Smarter Teacher Leadership: Neuroscience and the Power of Purposeful Collaboration, Teachers College Press, 2015.
[17] M. Salmona, D. Kaczynski, Don't blame the software: using qualitative data analysis software successfully in doctoral research, Forum Qual. Sozialforschung/Forum: Qual. Soc. Res. 17 (3) (2016), https://doi.org/10.17169/FQS-17.3.2505. Art 11.
[18] J. Creswell, J.D. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approach, fifth ed., Sage, 2018.

[19] J.N. Lester, Y. Cho, C.R. Lochmiller, Learning to do qualitative data analysis: a starting point, Hum. Resour. Dev. Rev. 19 (1) (2020) 94–106, https://doi.org/10.1177/1534484320903890.
[20] C. Seale, Using computers to analyse qualitative data, in: D. Silverman (Ed.), Doing Qualitative Research, second ed., Sage, 2005, pp. 154–172.
[21] M. Abulela, M. Harwell, Data analysis: strengthening inferences in quantitative education studies conducted by novice researchers, Educ. Sci. Theor. Pract. 20 (2020) 59–78, https://doi.org/10.12738/jestp.2020.1.005.
[22] M.J. Albers, Quantitative data analysis: in the graduate curriculum, J. Tech. Writ. Commun. 47 (2) (2017) 215–233, https://doi.org/10.1177/0047281617692067.
[23] R.W. Cooksey, Computer programs for analysing quantitative data, in: Illustrating Statistical Procedures: Finding Meaning in Quantitative Data, Springer, Singapore, 2020, pp. 33–49, https://doi.org/10.1007/978-981-15-2537-7_3.
[24] UNISTAT, UNISTAT Statistic Software, 2020. Retrieved 26 April 2023 from, https://www.unistat.com/download/.
[25] A. Crossman, A review of software tools for quantitative data analysis: how to get started with statistical analysis. https://www.thoughtco.com/quantitative-analysis-software-review-3026539, 2019.
[26] U. Wolski, The history of the development and propagation of QDA software, Qual. Rep. 23 (13) (2018) 6–20, https://doi.org/10.46743/2160-3715/2018.2984.
[27] J. Davidson, S. Di Gregorio, Qualitative research and technology: during a revolution, in: N.K. Denzin, Y.S. Lincoln (Eds.), Sage Handbook of Qualitative Research, fourth ed., Sage, 2011, pp. 627–643.
[28] L. Roberts, L. Breen, M. Symes, Teaching computer-assisted qualitative data analysis to a large cohort of undergraduate students, Int. J. Res. Method Educ. 36 (3) (2013) 279–294, https://doi.org/10.1080/1743727X.2013.804501.
[29] J. Tummons, Using software for qualitative data analysis: research outside paradigmatic boundaries, Big Data? Qualitative Approaches to Digital Research Studies in Qualitative Methodology 13 (2014) 155–177, https://doi.org/10.1108/S1042-319220140000013010.
[30] K. Bhattacharya, Coding is not a dirty word: theory-driven data analysis using NVivo, in: S. Hai-Jew (Ed.), Enhancing Qualitative and Mixed Methods Research with Technology, IGI Global, 2015, pp. 1–30.
[31] N.K. Denzin, Y.S. Lincoln, Introduction: the discipline and practice of qualitative research, in: N.K. Denzin, Y.S. Lincoln (Eds.), The Sage Handbook of Qualitative Research, third ed., Sage, 2005, pp. 1–32.
[32] P. Bazeley, The contribution of computer software to integrating qualitative and quantitative data analysis, Res. Sch. 13 (1) (2006) 64–74.
[33] U. Kuckartz, S. Rädiker, Using MAXQDA for integration in mixed methods research, in: J.H. Hitchcock, A.J. Onwuegbuzie (Eds.), The Routledge Handbook for Advancing Integration in Mixed Methods Research, Routledge, 2022, pp. 540–562.
[34] S.S. Taylor, A. Treacy, Just Dedoose it! Making Mixed Methods Data Analysis Manageable, , Paper presented at the annual meeting of the Northern Rocky Mountain Educational Research Association, Jackson Hole, WY, 2013.
[35] Z. Lysaght, G. Cherry, Using SPSS and Nvivo to conduct mixed-methods analysis collaboratively online: challenges, opportunities, and lessons learned, in: Sage Research Methods: Doing Research Online, Sage, 2022, https://doi.org/10.4135/9781529604023.
[36] E.G. Creamer, Adding to the understanding of fully integrated mixed methods research, in: J.H. Hitchcock, A.J. Onwuegbuzie (Eds.), The Routledge Handbook for Advancing Integration in Mixed Methods Research, Routledge, 2022, pp. 45–54.
[37] C. Silver, A. Lewins, Using Software in Qualitative Research: A Step-by-step Guide, second ed., Sage, 2020.
[38] R. Wolff, 8 Great Tools to Perform Qualitative Data Analysis in 2022, 2021. Retrieved 10 April 2023, from, https://monkeylearn.com/blog/qualitative-data-analysis-software/.
[39] T.C. Guetterman, J.W. Creswell, U. Kuckartz, Using joint displays and MAXQDA software to represent the results of mixed methods research, in: M. McCrudden, G. Schraw, C.W. Buckendahl (Eds.), Use of Visual Displays in Research and Testing: Coding, Interpreting, and Reporting Data, Information Age Publishing, 2015, pp. 145–176.
[40] M. Salmona, E. Lieber, D. Kaczynski, Qualitative and Mixed Methods Data Analysis Using Dedoose: A Practical Approach for Research across the Social Sciences, Sage, 2020.
[41] P. Bazeley, Using NVivo for mixed methods research, in: A. Onwuegbuzie, R. Burke Johnson (Eds.), The Routledge Reviewer's Guide to Mixed Methods Analysis, Routledge, 2021, pp. 343–354, https://doi.org/10.4324/9780203729434-29.
[42] B. Devi Prasad, Qualitative content analysis: why is it still a path less taken? Forum Qual. Sozialforschung/Forum Qual. Soc. Res. 20 (3) (2019) https://doi.org/10.17169/fqs-20.3.3392.
[43] M. Schreier, Qualitative Content Analysis in Practice, Sage, 2012.
[44] W.J. Potter, D. Levine-Donnerstein, Rethinking validity and reliability in content analysis, J. Appl. Commun. Res. 27 (3) (1999) 258–284, https://doi.org/10.1080/00909889909365539.
[45] Z.H. Merchant, A. Sadaf, L. Olesova, T. Wu, A systematic review of research questions in mixed methods studies in instructional design, Pedagogical Res. 6 (4) (2021), em0107, https://doi.org/10.29333/pr/11282.
[46] D. Teferra, Manufacturing – and exporting – excellence and 'mediocrity': doctoral education in South Africa, S. Afr. J. High Educ. 29 (5) (2015) 8–19.
[47] S. Munoz-Najar Galvez, R. Heiberger, D. McFarland, Paradigm wars revisited: a cartography of graduate research in the field of education (1980–2010), Am. Educ. Res. J. 57 (2) (2020) 612–652, https://doi.org/10.3102/0002831219860511.
[48] M.B. Karsli, S. Karabey, N.E. Cagiltay, Y. Goktas, Comparison of the discussion sections of PhD dissertations in educational technology: the case of Turkey and the USA, Scientometrics 117 (3) (2018) 1381–1403, https://doi.org/10.1007/s11192-018-2955-8.
[49] A. Wilkinson, D. Van Jaarsveldt, E. Grimsley, L. Seoka, Research trends in higher education studies at master's and doctoral level: 'Interdisciplinarity' and postgraduate supervision, in: M. Fourie-Malherbe, R. Albertyn, C. Aitchison, E. Bitzer (Eds.), Postgraduate Supervision: Future Foci for the Knowledge Society, SUN MeDIA, 2016, pp. 349–363.
[50] P. Ngulube, Postgraduate supervision practices in education research and the creation of opportunities for knowledge sharing, Probl. Educ. 21st Century 79 (2) (2021) 255–272, https://doi.org/10.33225/PEC/21.79.255.
[51] M.C. Stansbury, Problem statements in seven LIS journals: an application of the Hernon/Metoyer-Duran attributes, Libr. Inf. Sci. Res. 24 (2002) 157–168, https://doi.org/10.1016/S0740-8188(02)00110-X.
[52] OpenDOAR, Directory of Open Access Repositories, 2023. Retrieved 23 April 2023 from, https://v2.sherpa.ac.uk/opendoar/.
[53] R. Kgoroeadira, Promoting Entrepreneurship as a Means to Foster Economic Prosperity: A Review of Market Failure and Public Policy, 2010. Retrieved, https://dspace.lib.cranfield.ac.uk/bitstream/handle/1826/6901/Kgoroeadira_Reabetswe_Thesis_2010.pdf. (Accessed 12 May 2023).
[54] Thale Group, Software Versioning Best Practices, 2023. Retrieved on 22 April 2023, from, https://cpl.thalesgroup.com/software-monetization/software-versioning-basics.
[55] N.L. Leech, Using IBM SPSS statistics for integration in mixed methods research, in: J.H. Hitchcock, A.J. Onwuegbuzie (Eds.), The Routledge Handbook for Advancing Integration in Mixed Methods Research, Routledge, 2022, pp. 527–539.
[56] S. Friese, J. Soratto, D. Pires, Carrying out a computer-aided thematic content analysis with ATLAS. ti, MMG Working Paper 18 (2) (2018) 6–28. Retrieved 24 April 2023 from, https://www.mmg.mpg.de/62130/wp-18-02.
[57] P. Zhao, P. Li, K. Ross, B. Dennis, Methodological tool or methodology? Beyond instrumentality and efficiency with qualitative data analysis software, Forum Qual. Sozialforsch. Forum: Qual. Soc. Res. 17 (2) (2016), https://doi.org/10.17169/fqs-17.2.2597.
[58] A.J. Onwuegbuzie, J.P. Combs, Emergent data analysis techniques in mixed methods research, in: A. Tashakkori, C. Teddlie (Eds.), Sage Handbook of Mixed Methods in Social and Behavioral Research, second ed., Sage, 2010, pp. 398–430.
[59] M.D. Fetters, D. Freshwater, The 1 + 1 = 3 integration challenge, J. Mix. Methods Res. 9 (2) (2015) 115–117, https://doi.org/10.1177/1558689815581222.

[60] A. Bryman, Barriers to integrating quantitative and qualitative research, J. Mix. Methods Res. 1 (1) (2007) 8–22, https://doi.org/10.1177/2345678906290531.
[61] G. Tchibozo, Applications in data analysis for educational research, Pol. Futures Educ. Internet 7 (4) (2009) 364–367, https://doi.org/10.2304/pfie.2009.7.4.364.
[62] M.J. Blanca, R. Alarcón, R. Bono, Current practices in data analysis procedures in Psychology: what has changed? Front. Psychol. 9 (2018) 2558, https://doi.org/10.3389/fpsyg.2018.02558.