JOURNAL OF E-LEARNING AND KNOWLEDGE SOCIETY
Vol. 20, No. 1 (2024), pp. 47-55

# Trends in the use of Multivariate Analysis in Educational Research: a review of methods and applications in 2018-2022

Annamaria De Santis[1], Katia Sannicandro, Claudia Bellini, Tommaso Minerva

*University of Modena and Reggio Emilia, Reggio Emilia (Italy)*

*(submitted: 18/5/2024; accepted: 25/6/2024; published: 30/6/2024)*

**Abstract**

Multivariate analysis is a statistical solution effectively used to investigate educational phenomena. It operates simultaneously on many variables and data and allows the development of classifications and models by returning data-driven understandings.

How does the international educational research community make use of multivariate analysis techniques?

We conducted a methodological review to identify trends in applying these methods in education. We extracted only papers written in English, indexed in Scopus, and published from 2018 to 2022 in journals in the Education category. Our review included bibliometrics such as years of publications, leading journals, and most cited articles.

We detected an increase in papers using multivariate analysis in the educational research in Scopus publications over the past five years, particularly in journals in quartiles Q1 and Q2. MANOVA represents the main method used for the analysis along with regression methods; the former may be overestimated due to the overlap of names with terms searched in the string. University students represent the preferred experimental subjects for investigation; the administration of surveys and questionnaires is the most practiced way to collect data; preferred analysis tools among those declared are non-free. Based on the topics, some research categories emerged: Teaching, Medical Education, STEM, Digital Education, Professional Development, Inclusion, Wellbeing.

However, the number of citations is low (less than 8) for three-quarters of the articles in our selection.

To increase the effective use, confidence, and understanding of multivariate analysis processes, appropriate skills in education, statistical analysis, and interpretation of results need to be strengthened.

**KEYWORDS:** Multivariate Analysis, Educational Research, Methodological Review, DIKW Hierarchy.

**DOI**
https://doi.org/10.20368/1971-8829/1135946

**CITE AS**
De Santis, A., Sannicandro, K., Bellini, C., & Minerva, T. (2024). Trends in the use of Multivariate Analysis in Educational Research: a review of methods and applications in 2018-2022. *Journal of e-Learning and Knowledge Society*, *20*(1), 47-55. https://doi.org/10.20368/1971-8829/1135946

## 1. Introduction

Multivariate analysis is a collection of statistical techniques that analyze multiple variables that simultaneously change and whose effects cannot be understood separately (Hair et al., 2014). These methods allow to develop models (i.e., regressions), classifications, and groupings (i.e., cluster analysis) and to reduce dimensionality and identify factors and latent traits (i.e., factorial analysis) by returning data-driven understandings of phenomena also in social sciences (Bartholomew et al., 2008; Hair et al., 2014; de Lillo et al., 2007; De Santis, 2022).

In education research, scholars use multivariate analysis to investigate students' opinions, perceptions and assessments, teachers' training and skills, efficacy of teaching methods and digital tools, educational poverties, organization of training institutions, and more just like: flipped class (Sointu et al., 2023) or university climate (Felini & Zobbi, 2022) through factorial analysis; serious games (Iten & Petko, 2016) using linear regression; teachers' role (León-Jariego et al., 2020) through cluster analysis and multidimensional scaling; students' activities in blended courses (Stites et al., 2019) using cluster

[1] corresponding author - email: annamaria.desantis@unimore.it

analysis; disabilities and ethnic disparities (Morgan et al., 2017) or lifelong learning (Narushima et al., 2016) using logistic regression.

A large amount of data is stored in platforms and contexts related to the educational sector and can represent a relevant source to deepen the procedures related to learning and teaching and the management of education systems. Data are produced in digital learning environments that host large numbers of learners in online learning (Reich, 2022) but also in physical environments where online surveys or assessments and mobile apps are commonly used (Sannicandro, 2023).

However, storing data is not enough to obtain new understandings. Looking at information philosophy and knowledge management, in the well-known DIKW pyramid (also called "information hierarchy", "knowledge hierarchy", or "wisdom hierarchy"), data are defined as symbols, objective observations, basic elements that have no meaning or value until processed. Classification, selection, sorting, and calculation procedures allow data transformation into information with meaning and significance; the combination of information, experiences, human understandings, skills, and values generates knowledge (Ackoff, 1989; Rowley, 2007).

The transformation of data into useful information and knowledge, therefore, requires operations of analysis and interpretation. To do this, the following are necessary at least:

- *an interdisciplinary set of skills*: educational knowledge should be developed/combined alongside statistical and computing skills;
- *rigorous practices*: multivariate analysis can represent one of the solutions to work on multiple variables and plenty of data;
- *proper tools*: statistical softwares simplify the mathematical and computing procedures behind statistical techniques.

Our study aims to identify trends in applying multivariate analysis methods in the education field over the past five years, showing common uses and areas of application that can expand.

We realized a review of analysis methods used in the papers published between 2018-2022 and indexed on Scopus in journals in Education category.

Section 2 describes the method used for the investigation; Section 3 contains the results achieved. The last Section presents conclusions and future developments of the research.

## 2. Materials and Methods

Based on the previous considerations, the analysis in this paper starts from the research question: how does the international educational research community make use of multivariate analysis techniques?

We carried out a review of methods to reply to our question.

*Methodological reviews*, also found under other nomenclature in scientific literature (e.g. "methodological survey", "systematic review", "meta-research", "research-on-research"), increased over the last few years. They examine the methodological issues and choices related to design, conduct, analysis, or reporting adopted in research, describing and comparing research practices and appropriate structured procedures. The aim is to detect and enhance the quality, accuracy, and consistency of research, showing, in some cases, gaps and needs to improve the investigation methodology (Mbuagbaw et al., 2020; Aguinis et al., 2023).

The research was structured as a *descriptive review* (Paré et al., 2015) that aims to summarize a topic of interest and provide a comprehensive scenario of an area of study. By proposing a replicable process of selecting the papers through structured search methods, in descriptive review, researchers collect the features of the analyzed studies and provide quantitative results about frequency of themes, methods, authors, and so on. The main goal is to identify trends and patterns and draw conclusions on the investigation topic.

We have conducted a review that delves into the application of multivariate analysis techniques in the educational field. By investigating the most recent scientific literature, we can gain awareness of how multivariate analysis techniques are being used in educational research, the topics for which the use of such techniques is well established, and the areas in which further actions are needed.

The review covers five years between 2018-2022 and consists of two phases. In the first one, we conducted a more general investigation by inserting the term "multivariate analysis" in the search string. In the second phase, aware that multivariate analysis includes numerous techniques, we'll refine our search based on the results of the first phase, replacing the most popular multivariate analysis techniques in the strings.

This paper presents the results of the first phase of our study.

We used as search strings: "multivariate analysis" AND education; "multivariate analysis" AND learning; "multivariate analysis" AND teaching. We decided to set the search string without using boolean operators because we were interested in how the single terms combined with the others.

We defined the following eligibility criteria for the papers:

1) Period: 2018-2022
2) Category Journal on Scopus: Education (1470)
3) Language: English
4) Document type: article
5) Source type: journal
6) Publication stage: final
7) Distribution: open access
8) Topic: compliant with the themes

The research focus is not on the quality of the studies or their results but on how the techniques were used.

In the analysis, we provided the topics, kinds of methods, data features, sample sizes, analysis tools, and the countries where the studies were conducted. Our analysis included bibliometrics such as years of publications, leading journals, and most cited articles.

## 3. Results

Figure 1 contains the selection process for the identification of the papers to include in the review.

| | "multivariate analysis" AND education | "multivariate analysis" AND learning | "multivariate analysis" AND teaching |
|---|---|---|---|
| IDENTIFICATION | 20,007 doc | 6,000 doc | 3,715 doc |
| **APPLYING THE ELIGIBILITY CRITERIA** | | | |
| SCREENING | 111 doc | 82 doc | 42 doc |
| **ELIMINATING DUPLICATION** | | | |
| INCLUDED | 157 doc | | |

**Figure 1** - Process of identification of the papers to include in the review.

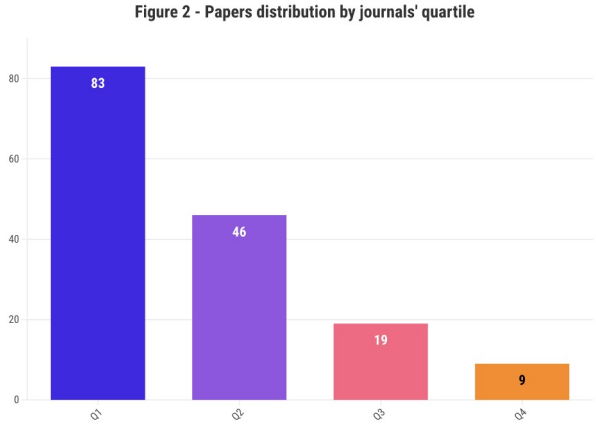
We obtained the highest number of papers using the term "education" in the search string and the lowest using the term "teaching". This result brings forward the findings on the topics focused on by the studies we present in the next.

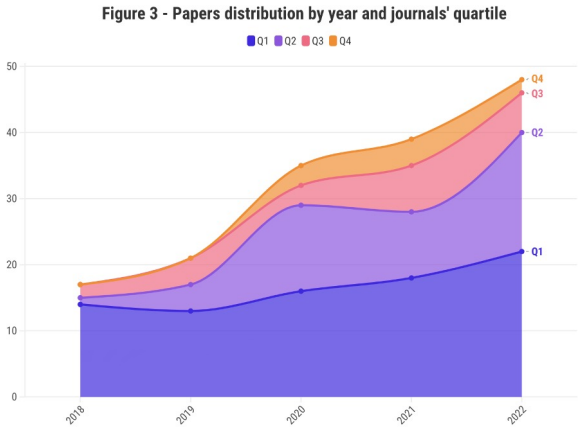
The described process yielded 157 documents.

Figures 2 and 3 describe the distribution of papers among journals by journals' quartile and year.

Around half of the selected papers (83) belong to Q1 journals.

In the last five years, the number of papers in Q1 published every year went from 14 to 22, delivering a relevant increase. However, the quartile where the changes are more important is the second one, where the papers published every year went from 1 to 18 over the five years, reaching almost the amount for Q1 journals.




**Figures 2 and 3** - Paper distribution by year and journals' quartile.

The papers published in Q1 belong to 39 journals. 41 of them (49%) are published in journals of the medical field. The Q1 journals that host the higher number of papers that deal with education and multivariate analysis are not surprisingly related to the health sector: "BMC Medical Education" (22); "Academic Medicine" (4); "Nurse Education Today" (4); "Medical Education Online" (3).

The number of papers belonging to the medical field in the journals of other quartiles is smaller: 5/46 in Q2, 2/19 in Q3, 2/9 in Q4.

Table 1 contains the four papers with the highest citations on Scopus. We also report the citation number on Scholar as a confirmation. These papers cover a comprehensive range of topics: the role of parents in school education and the life quality of university students during the COVID-19 pandemic, the skills of nurses and midwives, and teaching in science courses.

The article with the highest number of citations is not in the health field but is published in a strictly educational journal and, as we shall see below, it involves parents who are an unusual type of statistical unit with respect to the mainly implemented practices.

50% of papers in our selection have less than three citations on Scopus, and 75% less than 8, indicating that they have little impact on the scientific discussion.

Figures 4 and 5 represent the countries in which the studies were conducted (from now on, the sum can be higher than the number of papers because more techniques, statistical units, software, and tools can be used in each research).

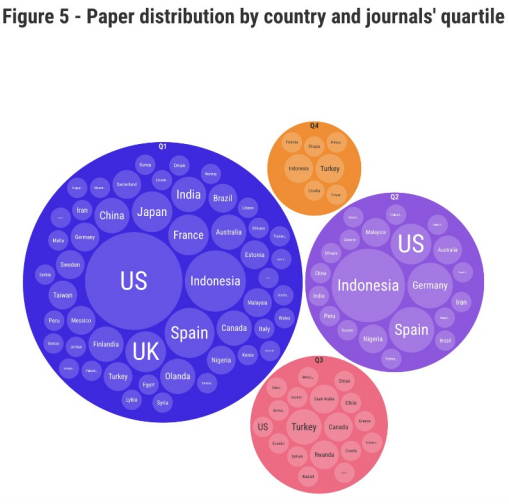
There is a distribution throughout the world.

A third of the studies analyzed were conducted in the United States (28) and Indonesia (25). Other countries that more consistently are the setting for the research on education through the use of multivariate analysis techniques are Spain (11), Germany (8), Turkey (7), Australia (6), and France (5).

In the distribution by journals' quartile, we can observe that in the first one, US (23), Indonesia (9), Spain (6) are the most productive countries. In the second one, Indonesia (13), Spain (5) and Germany (5). In the third and fourth, Turkey (respectively 3 and 2 papers) rises up.

**Table 1** - Most cited papers in journals of the education category.

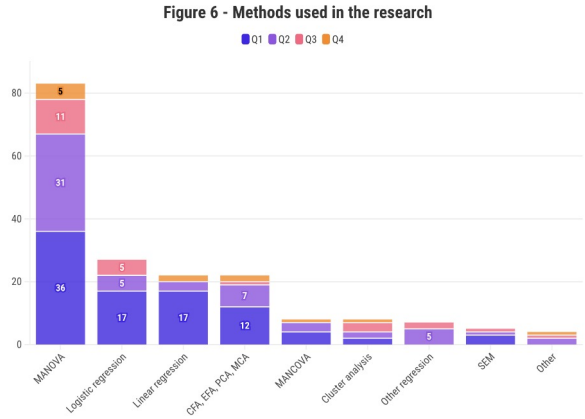
| *Q* | *Paper title* | *Scopus citations* | *Scholar citations* |
|---|---|---|---|
| Q1 | Lee, S. J., Ward, K. P., Chang, O. D., & Downing, K. M. (2021). Parenting activities and the transition to home-based education during the COVID-19 pandemic. *Children and Youth Services Review*, *122*, 105585. | 197 | 420 |
| Q1 | Ross, L., McSherry, W., Giske, T., van Leeuwen, R., Schep-Akkerman, A., Koslander, T., Hall, J., Steenfeldt, V.Ø., & Jarvis, P. (2018). Nursing and midwifery students' perceptions of spirituality, spiritual care, and spiritual care competency: A prospective, longitudinal, correlational European study. *Nurse education today*, *67*, 64-71. | 77 | 139 |
| Q1 | Cavanagh, A. J., Chen, X., Bathgate, M., Frederick, J., Hanauer, D. I., & Graham, M. J. (2018). Trust, growth mindset, and student commitment to active learning in a college science course. *CBE - Life Sciences Education*, *17*(1), ar10. | 68 | 168 |
| Q2 | Silva, P.G.D.B., de Oliveira, C.A.L., Borges, M.M.F., Moreira, D.M., Alencar, P.N.B., Avelar, R.L., Sousa, R.M.R.B., & Sousa, F.B. (2021). Distance learning during social seclusion by COVID-19: improving the quality of life of undergraduate dentistry students. *European Journal of Dental Education*, *25*(1), 124-134. | 45 | 107 |

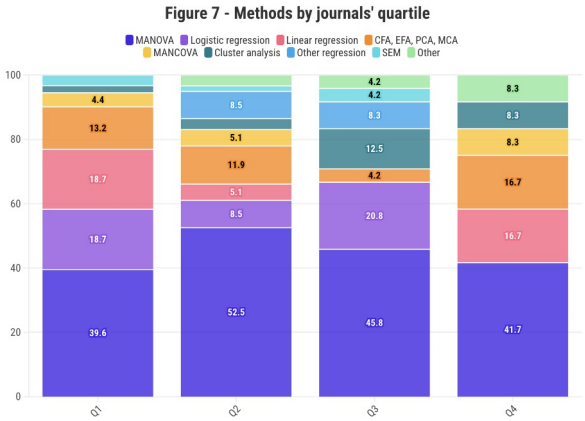




**Figures 4 and 5** - Paper distribution by country and journals' quartile.

Figures 6 and 7 represent more elements related to the methods used. In particular, around half of the research used multivariate analysis methods such as MANOVA (multivariate analysis of variance) followed by regression, data reduction, and classification techniques. MANOVA is used to compare group differences in metric dependent variables simultaneously so that it can be instrumental in experimental designs but also in non-experimental designs to assess the statistical significance of differences among groups built based on particular features (Hair et al., 2010). However, we are aware that the results can be distorted because the full name of the technique MANOVA contains the words in the search string.

Overall, MANOVA and MANCOVA cover 48.9% (92 papers) of the research; regression methods 30.1% (56 papers), data reduction 11.8% (22 papers) and cluster analysis 4.3% (8 papers).




**Figures 6 and 7** - Paper distribution by methods and journals' quartile.

This odd is about similar also within each quartile.

However, in papers in Q1 journals, there is a higher percentage of logistic and linear regression techniques equally. In Q2, on the contrary, there is a greater representation of MANOVA and MANCOVA at the expense of regressions. Although the number of papers is lower in Q3 and Q4 journals, we observe that in Q3, 5 researches use logistic regression and none the linear one; in contrast, in Q4, the regression is used only in 2 papers as linear.

One third of the papers do not state the statistical analysis software used (54, 34.4%); SPSS is used in half of the papers (80, 51.0%). STATA, SAS, R, and Jamovi are chosen in the remaining papers, mainly from Q1 journals (Figure 8).

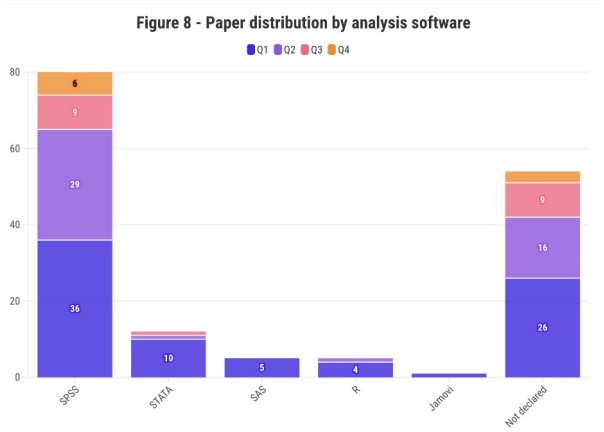



**Figure 8** - Paper distribution by analysis software.

Three-quarters of the articles analyzed have a sample/ population size below 500 units. Papers with larger samples were published in Q1 and Q2 journals (Figures 9 and 10).

This result should be read in conjunction with the data collection techniques used and the statistical units chosen.

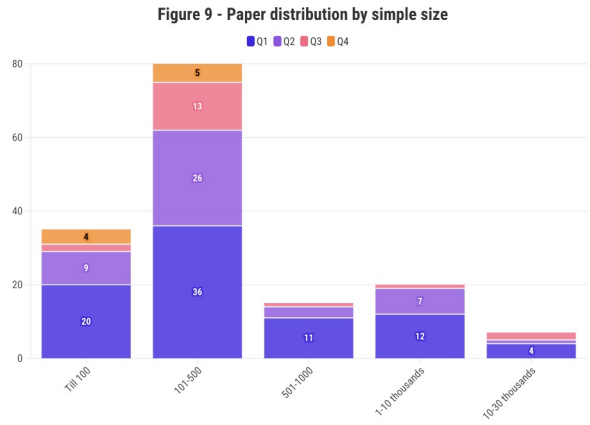
In almost all cases (87.6%), the techniques for detecting data are questionnaires, scales, inventories, and assessments. In only 10 cases, the analyses originate from national or regional datasets, and a few others involve the analysis of audio, documents, computer metrics, and medical records.

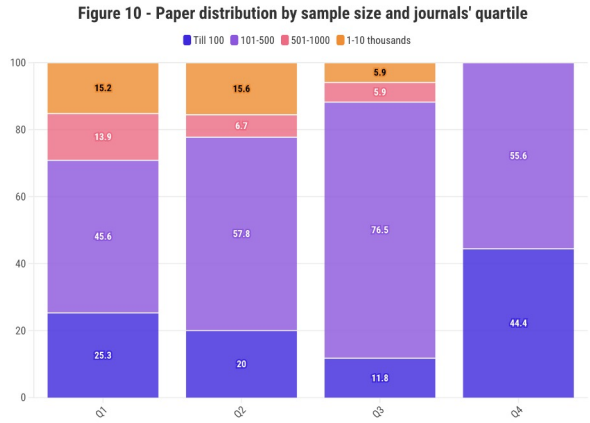
The main statistical units considered are students, teachers, and other professionals.

In details:

- 35.4%, university students;
- 24.4%, students from kindergarten, primary and secondary schools;
- 15.9%, preservice teachers, teachers, and professors;
- 10.9%, residents and health professionals.

The choice of statistical units "other than humans" is more unusual: schools, lessons, and courses, together with objects such as metrics, books, claims, hymns, and so on, represent the statistical units of 8.5% of papers.




**Figures 9 and 10** - Paper distribution by sample size and journals' quartile.

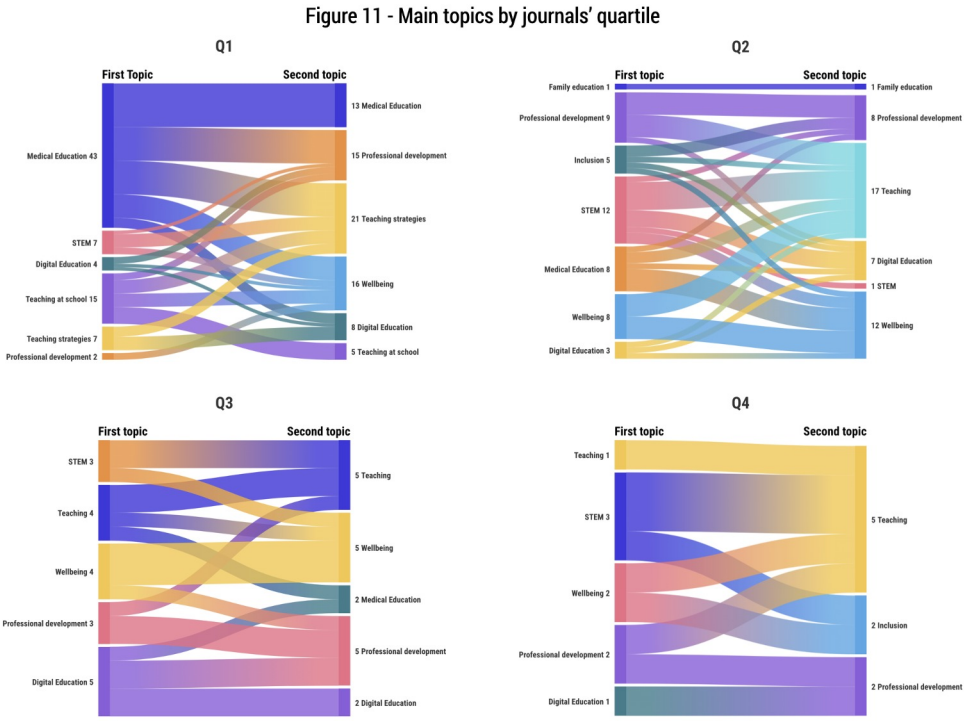
We identified one or more main investigation fields for each paper and connected them to the kinds of statistical units.

Figure 11 shows the main topics explored in the papers distinguished for journals' quartile.

Medical Education occupies a significant role in papers from Q1 journals; the topics of Medical Education are intertwined with Professional development (including faculty development, teachers' training, and career skills), Teaching strategies, Wellbeing, and Digital Education. Teaching strategies are also a theme related to schools, STEM, and Digital Education. Wellbeing, considered as a macro-area including motivation, anxiety, stress, and mental health, connects with all the previously mentioned themes. Papers dealing with Medical Education are in the same number as those dealing with Teaching strategies in general and in schools.

In the papers belonging to Q2, Q3, and Q4 journals, the number of journals dealing with health and the number of papers dealing with Medical Education decreases, leaving room mainly for topics related to Teaching, Professional development, and Wellbeing. Papers dealing with STEM and Digital Education are in a good number in the Q2 papers.

For papers in the Q2 and Q4 journals, it seemed important to add the topic of Inclusion for research focused on pre-service teachers' opinions on it, reasons for dropout, and EFL students' skills.




**Figures 11 and 12** - Main topics and statistical units by journals' quartile.

Finally, Figure 12 highlights the relationship between the main topics and the statistical units selected by the researchers.

University students represent the preferred experimental subjects for investigation together with secondary school students, teachers and pre-service teachers, and children in all journal quartiles.

Considering all the quartiles and only one main topic for the papers, we observed that residentials/ practitioners and health professionals are more frequently the units in papers in the Q1 sector focused on Medical Education; studies on Digital Education and Professional development more frequently involved teachers and pre-service teachers; research on Teaching and Wellbeing focus on children and secondary school students; the use of statistical units different from students or professionals is diffused among papers on Medical Education and Teaching.

## 4. Conclusions

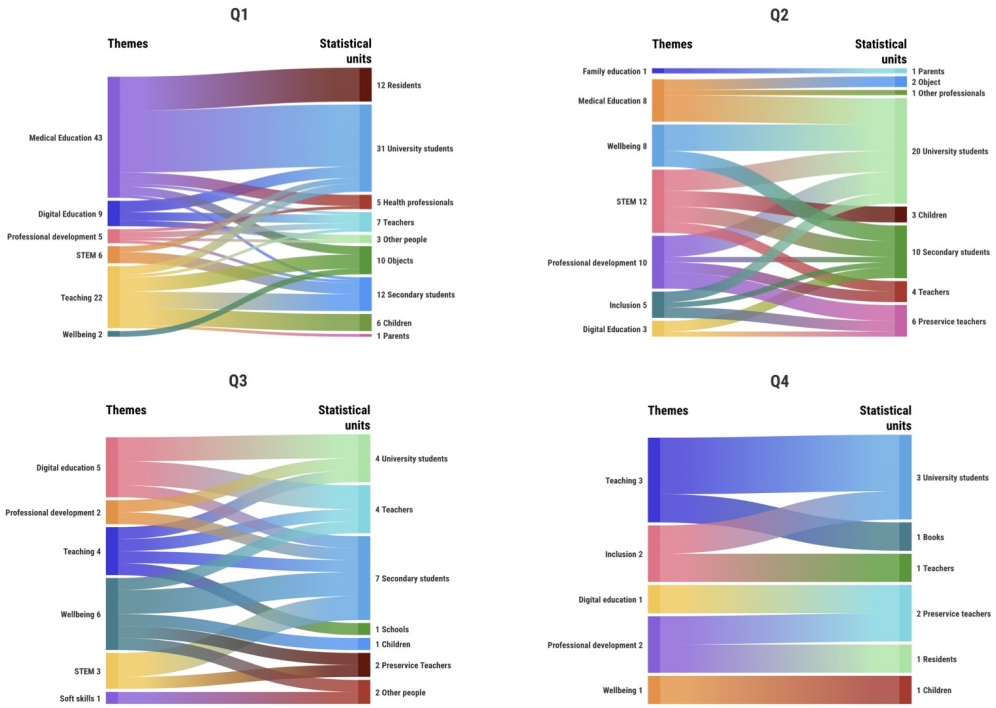
This first stage of review analysis leads to some considerations.

Our analysis shows that there has been an increase in papers using multivariate analysis in the educational field in Scopus publications over the past five years, particularly in Q1 and Q2 journals ranked in the Education category.

Synthesizing the main results of our analysis, we can say that:

- the number of citations is low (less than 8) for three-quarters of the articles in our selection;
- a large number of countries produced studies using these methods; however, one third took place in the United States and Indonesia;
- MANOVA is among the most commonly used multivariate analysis techniques, along with regression methods. The former may be overestimated due to the overlap of names with terms searched in the string;
- samples/populations less than 500 is the most commonly recorded choice;
- administration of surveys and questionnaires is the most practiced way to collect data;
- in about one-third of the papers, the tools used are not stated. Preferred analysis tools among those declared are non-free. SPSS is used in half of the studies, and few studies state the use of open source software;
- Teaching, Medical Education, STEM, Digital Education, Professional Development, Inclusion, and Wellbeing are among the many topics investigated using multivariate analysis;
- researches mainly use (university) students as statistical units.

Figure 12 - Main topics and statistical units by journals' quartile




**Figures 11 and 12** - Main topics and statistical units by journals' quartile.

Our research returns a scenario in which, as a hypothesis, poor citations could show a need for more confidence and understanding of multivariate analysis processes or, on the contrary, they could be a symptom of low relevance of results, which was not an element analyzed in our study.

Many other applications of multivariate analysis, using varying statistical units, topics, methods, and data collection tools, could be considered.

Databases in online learning environments or data from digital devices can replace those collected by surveys or similar methods, which represent the most frequent choice in the paper analyzed and could be affected by the perceptions of subjects involved in detection.

Data collection and analysis could be extended to other unrepresented fields of study by also addressing subjects not necessarily involved in formal training.

Additionally, the use of open source software could broaden the application of the techniques.

In general, as we have sustained from the beginning of this paper, appropriate skills in both education and statistical analysis and interpretation of results need to be strengthened.

Future studies may analyze papers belonging to other academic databases. The second phase of the research will involve using some of the most common multivariate analysis techniques in education and teaching in the search string.

## Authors' contribution

According to CRedit system, Annamaria De Santis: Conceptualization, Methodology, Investigation, Data Curation, Writing - Original Draft, Visualization. Katia Sannicandro: Methodology, Resources. Claudia Bellini: Resources. Tommaso Minerva: Supervision.

## Note

This paper describes the whole research partially introduced in the conference abstract "Trends in the use of multivariate analysis in educational research: a review of methods and applications in 2018-2022" presented at the Italian SYmposium on Digital Education (ISYDE) held in Reggio Emilia (Italy) on September 13-15th, 2023.

## References

Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, *16*(1), 3-9.

Aguinis, H., Ramani, R.S., & Alabduljader, N. (2023). Best-practice recommendations for producers, evaluators, and users of methodological literature reviews. *Organizational research methods*, *26*(1), 46-76.

Bartholomew, D.J., Steele, F., Moustaki, I., & Galbraith, J.I. (2008). *Analysis of multivariate social science data* (2nd ed.). Boca Raton (FL): CRC press, Taylor & Francis Group.

de Lillo, A., Argentin, G., Lucchini, M., Sarti, S., & Terraneo, M. (2007). *Analisi multivariata per le scienze sociali*. Milano: Pearson Education.

De Santis, A. (2022). *Analisi multivariata e learning analytics. Metodi e applicazioni*. Milano: Pearson.

Felini, D., & Zobbi, E. (2022). University climate in distance education contexts: developing an assessment instrument. *Journal of e-Learning and Knowldge Society*, *18*(1), 75-86.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2014). *Multivariate data analysis* (7th ed.). Edinburgh Gate, Harlow, Essex (GB): Pearson.

Iten, N., & Petko, D. (2016). Learning with serious games: Is fun playing the game a predictor of learning success?. *British Journal of Educational Technology*, *47*(1), 151-163.

León-Jariego, J.C., Rodríguez-Miranda, F.P., & Pozuelos-Estrada, F.J. (2020). Building the role of ICT coordinators in primary schools: A typology based on task prioritisation. *British Journal of Educational Technology*, *51*(3), 835-852.

Mbuagbaw, L., Lawson, D.O., Puljak, L., Allison, D.B., & Thabane, L. (2020). A tutorial on methodological studies: the what, when, how and why. *BMC Medical Research Methodology*, *20*, 1-12.

Morgan, P.L., Farkas, G., Hillemeier, M. & Maczuga, S. (2017). Replicated Evidence of Racial and Ethnic Disparities in Disability Identification in U.S. Schools. *Educational Researchers*, *46*(6), 305-22.

Narushima, M., Liu & J., Diestelkamp, N. (2016). Lifelong learning in active ageing discourse: its conserving effect on wellbeing, health and vulnerability. *Ageing & Society*, *38*, 615-675.

Paré, G., Trudel, M. C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & management*, *52*(2), 183-199.

Reich, J. (2022). Learning analytics and learning at scale. In C. Lang, G. Siemens, A.F. Wise, D. Gaševič, & A. Merceron (Eds.), *Handbook of*

*Learning Analytics* (2nd ed.). Vancouver, Canada: SOLAR.

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science*, *33*(2), 163-180.

Sannicandro, K. (2023). *Online Assessment. Valutazione degli apprendimenti nei corsi universitari a distanza*. Milano: FrancoAngeli.

Sointu, E., Hyypiä, M., Lambert, M.C., Hirsto, L., Saarelainen, M., & Valtonen, T. (2023). Preliminary evidence of key factors in successful flipping: Predicting positive student experiences in flipped classrooms. *Higher Education*, *85*(3), 503-520.

Stites, A.A., Berger, E., Deboer, J., & Rhoads, J.F. (2019). A Cluster-Based Approach to Understanding Students' Resource-Usage Patterns in an Active, Blended, and Collaborative Learning Environment. *International Journal of Engineering Education*, *35*(6/A), 1738-1757.