

UNIVERSITY OF SZEGED
DOCTORAL SCHOOL OF EDUCATION
INFORMATION AND COMMUNICATION TECHNOLOGIES IN
EDUCATION PROGRAMME

GYULA NAGY

SCIENTOMETRIC AND CONTENT ANALYSIS OF MAGYAR PEDAGÓGIA

Summary of the PhD dissertation

Supervisor:

Gyöngyvér Molnár DSc

Professor of Education



Szeged, 2019

THE TOPIC AND THE STRUCTURE OF THE DISSERTATION

This research interest emerged due to the fact that Hungarian scholars dealing with humanities and social sciences – or specifically, educational research – are poorly represented in citation databases. For progressive academicians and scientists today it is unavoidable to use international citation databases such as Web of Science or Scopus. However, there are numerous disciplines in Hungary which fail to benefit from the use of such databases because of being only marginally represented therein. Nevertheless, the representatives of such disciplines would also be interested in understanding the inner structure of their own fields of study. Filling this gap in scientific evaluation is possible by means of finding alternative solutions, i.e. studying and analysing the networks of scientific cooperation or representing the citation network of a specific journal or a whole discipline.

Due to the changes taking place in the last couple of decades, in the evaluation of one's scientific output the number of academic papers published in journals is increasingly gaining importance. Compared to other means of scientific communication, the importance of journal papers has risen significantly, which has affected social sciences as well. Owing to such changes, it has proved to be important to study journals in detail: to detect their strengths, weaknesses and the on-going structural changes.

Due to its long history, the well-known Hungarian academic journal, *Magyar Pedagógia* [Hungarian Pedagogy], which was established in 1892, is a unique journal in the field of educational research. In recognition of this fact, the papers published between 1892 and 2014 were analysed in detail in order to reveal certain trends and phenomena.

Although the results of the present research are related to one educational research journal only, all the research questions and the applied methodology are universal. This means that the key findings of the research are not the results presented in the dissertation but the methods that were developed and applied. The main research question is the following: how can those scientific journals be measured in an objective way which are outside the scope of the major international scientometric providers? On the basis of this, our task was to reproduce the techniques applied by the most prominent international content providers on a smaller scale. The dissertation focuses on the description of this process, and the applied methodology integrates certain techniques of educational research, scientometrics, citation and text mining analysis. Moreover, in order to get a comprehensive analysis of the chosen journal, a detailed content analysis was also conducted by means of text-mining techniques too.

Chapters 1-3 of the dissertation cover the problem statement and the literature review. The theoretical background focuses on the transformation of scientific communication on a national and international level, interdisciplinary techniques related thereto (scientometrics, text mining and network science) and the history and importance of the Hungarian academic journal *Magyar Pedagógia*. Chapters 4-5 then investigate the aims and objectives of the research, including the research questions and the hypotheses. The same two chapters contain the description of the applied methods too. The results of the three key structural parts (the examination of metadata, citation and content analyses) are discussed in Chapters 6-8. Finally, Chapter 9 summarizes the findings of the research, the limitations of generalisation and other possible future research directions.

There are three “pillars” underlying the research project: the first one is the analysis of metadata related to the journal articles which enables us to identify the main impacts on the journal, the most significant tendencies that are detectable and the patterns characterising authorship and co-authorship. The second pillar then includes detailed citation analyses based on the bibliographies found at the end of each paper, which were conducted following the automatic detection and extraction of citations to study the types, the distribution and the freshness of references. These steps also included the generation and the examination of full

citation networks based on the extracted citations as well as a reference graph built to depict which international and national sources and authors are referred to most frequently by the authors of the journal. Network analysis makes it possible to explore which authors are the most cited scholars and the key figures in the citation graphs. Finally, as the third “pillar” of this study, the comprehensive content analysis of the journal papers was also conducted. By means of such analysis, thematic changes, key discourses and main thematic groups characterising certain periods could become observable.

THEORETICAL BACKGROUND

Scientific publishing and the dissemination of scientific findings have undergone significant changes in the last decades. Scientific discourse started to take place in journals instead of books (monographs, edited and synthesized volumes) (Larsen & Von Ins, 2010). The phenomenon that first emerged in the sphere of natural sciences spread to the field of social sciences and even arts and humanities (Csapó, 2011). Furthermore, increased scientific collaboration was also observable in every field of science (Glänzel, Schubert, & Czerwon, 1999). The number of publications written in collaboration has risen and the number of publications written by one author only has decreased. At the same time, scientific competition has increased significantly (Van Raan, 2001; Wong & Hodson, 2010). To underline the existence of this phenomenon, there is a detectable preference for citing the works written by multiple authors (Wuchty, Jones, & Uzzi, 2007). Fierce competition and on-going transformation gave way to the birth of different fields of meta-research and phenomena, such as scientometrics (Hood & Wilson, 2001) or the Open Access phenomenon (Hajjem, Harnad, & Gingras, 2006). The field of educational research has also undergone changes as new trends affected this field significantly too (Csíkos, 2013).

Scientific interest in education emerged in Hungary at least one hundred and forty years ago (Nagy, 2011). Institutionalised research into education in Hungary started with the establishment of the Hungarian Pedagogical Society in 1891 (Klamarik & Suppán, 1892). The Society’s own journal, which was called *Magyar Paedagogia*, was started the next year. One hundred years later, after the change of the regime, from the 90’s, there were four academic journals in Hungary dominating the field of educational research. The situation is approximately the same today. According to the Hungarian Academy of Sciences’ Committee on Pedagogy, these prominent journals are the following: *Magyar Pedagógia* (Hungarian Pedagogy), *Iskolakultúra* (School Culture), *Educatio* and *Új Pedagógiai Szemle* (New Pedagogical Review) (Biró, 2009a). The ranking of the leading journals was also confirmed by different empirical citation analyses (Tóth, Toman, & Cserpes, 2008; Pap, 2009). Due to its long history and importance, *Magyar Pedagógia* was chosen to be studied in the present research. Mészáros (1992) published a detailed and valuable article on the history of this one-hundred-year-old journal. The journal was renewed after its hundredth year of existence, and the new principles were introduced at that time regarding its publishing activity, which still prevail today. Csapó (2005) summarizes these principles in his paper.

The methodology applied in the research is based on a number of scientometric analyses. It was Vassily V. Nalimov who coined the term scientometrics in the 1960s. Scientometrics enables us to carry out the complex task of analysing and measuring scientific performance. Derek J. de Solla Price (Price, 1979), Robert K. Merton (Merton & Garfield, 1986) and Eugene Garfield (Garfield & Merton, 1979; Garfield, 2009) are considered to be the most respected authorities in this field of science. Scientometric analyses focusing on the evaluation of the scientific impact of various journals are mainly based on international citation databases (Moed, 2009). Web of Science, Scopus and Google Scholar have proved to be the three most significant and most widely consulted databases in this field (Bornmann, Thor, Marx, & Schier, 2016). On recent advancements concerning the measurement of the outputs of academic research, it is

advisable to read the initiative called the Declaration on Research Assessment (DORA) (San Francisco Declaration, 2013) and the article written by Hicks, Wouters, Waltman, De Rijcke and Rafols (2015) titled Leiden Manifesto. In addition, the articles of David and Frangopol (2015) and Schubert (2018; 2015b) on the difficulties related to the objective measurement of scientific performance are also worth considering.

As far as the connection point between scientometrics and educational research is concerned, it can be said that there are several analyses in international literature related in some ways to the aforementioned databases, but there are evaluations with different approaches as well. Ivanović and Ho (2017), for example, studied the most highly cited articles in the 'Educational Research' category of the Social Science Citation Index, which is the most relevant database in social sciences, and they considered the following aspects in 2091 publications: the distribution of such articles per publication year, journals, institutions and countries. Youtie, Solomon, Carley, Kwon and Porter (2017) went beyond the scope of educational research by studying the relationship between cognitive science and educational research on the basis of analysing citation networks. It is important to emphasize that the period studied by them (from 1994 until 2014) is similarly as long as the period chosen to be studied in the present research (from 1991 until 2014). Weller, Jordan, DeVries and Rolfe (2018), in their article, made an attempt to map the area of open education by means of social network analysis, which goes beyond the study of citation databases (Borgatti, Mehra, Brass, & Labianca, 2009).

There are two Hungarian studies, which were conducted approximately ten years ago, affording a detailed insight into the outputs of scholarly research on a national level in the field of educational research. In these, the most influential scientific media, institutions of universities and the leading scholars in educational research were identified. The first study was conducted by Tóth, Toman and Cserpes (2008) under the name of Wargo Institute of Economic Research, with the aim of measuring the scientific performance characterising Hungarian educational research by means of empirical methods. The activities of the main Hungarian research groups, leading scholars and most prominent journals of the field were examined in their research. As scientometrics is concerned with the analysis of citations (King, 1987; Kostoff, 1998), the so-called Wargo study also relies heavily on a number of citation analyses, which are presented in detail in the paper. The second study that must be mentioned describes the investigation conducted by the Research Group of Historical Pedagogy of the Doctoral School of Educational Science, Eötvös Loránd University, between the periods of 2007-2008 and 2012-2015, under the supervision of András Németh, financed by OTKA. Their findings reflect the main tendencies and conditions characterising the field. The results are available in several edited volumes (Németh & Biró, 2009; Németh, Biró, & Garai, 2015) and a number of articles (Németh, Biró, & Varga, 2009; Biró, 2009a; Biró, 2009b; Pap, 2009) as well. The publication activities of the most prestigious Hungarian educational research journals between 1997 and 2006 were examined by the researchers (Németh, Biró, & Varga, 2009). At the beginning of the 1990's, Schriewer and Keiner (1993) made an attempt to compare and contrast the communication which characterised educational science in France and Germany over a period ranging from 1955 to 1985.

Following 2010, one of the most comprehensive studies on the relationship between educational research and scientometrics was published by Csíkos (2013). Then the study by Nagy (2016), from a similar perspective, discusses a number of new phenomena such as the importance of the transformation of scientific communication, the Open Access, the emergence of repositories and digital object identifiers. Csapó (2016) finds it especially important to encourage Hungarian scholars to participate in international research when examining the scientific background of teacher training and educational development. It is crucial to note that the conscious application of the new methods of scientometrics is necessary for such a change

of perspective on the part of academicians. When the scientific performance of different higher education institutions is considered, it is vital to highlight the findings of Fehér and Aknai (2016), who compared the scientific results, performance and the output requirements of different doctoral schools in the field of educational research in Hungary. Finally, Molnár, Pintér and Tóth conducted research into co-authorship networks in educational research by means of social network analysis (Molnár, 2018; Molnár, Pintér, & Tóth, 2018a, 2018b; Molnár, Tóth, & Pintér, 2018a, 2018b, 2018c).

For accomplishing the main goals of the research (the analysis of metadata as well as citation and content analyses), there was an essential need for the using of text mining tools. Text mining is also often referred to as text data mining by scholars. According to the definition by Tikk (2007), text mining is the method of processing and analysing unstructured or only partially structured textual data with the aim of identifying and studying latent information in different texts. Text mining is an application-oriented interdisciplinary field, and the typical tasks of text mining analysis are the following: categorization, clustering, summarising, extracting information, detecting trends, semantic analysis and visualization (Feldman & Sanger, 2007). Modern scientometrics, which is based on information technology, could not exist without the automated solutions such as text mining or artificial intelligence, which make it possible to process enormous amounts of data (Börner, 2010; 2015). Both text mining and its predecessor data mining are used by scholars conducting educational research (Hung, 2012; Romero, Ventura, Pechenizkiy, & Baker, 2010; Ueno, 2004). International literature refers to this sub-discipline as 'Educational Data Mining' (EDM). EDM is an emerging discipline with the aim of developing relevant methods for exploring large-scale and specific data and textual information coming from educational settings to understand students and their circumstances (Baker & Yacef, 2009).

THE GOAL OF THE RESEARCH

In the first phase of the present research, the digitization of all the volumes of the journal *Magyar Pedagógia* and the bibliographic description thereof were considered to be the main goals. It was also important to make the full text of all the volumes digitized freely available for the scientific community. And at the same time, the improvement of the journal's visibility on the web proved to be essential too. Only after taking all these initial steps, was it possible to continue with the central aim of the research: the comprehensive scientometric analysis of the journal on the basis of national and international trends. The research had multiple aims: (1) to build up a co-authorship graph covering the whole period from the foundation of the journal (1892) until today. (2) The second aim was to build up the journal's complete inner reference graph covering the period from 1991, which was the year when standard reference style was first applied by the journal, until 2014, which was the year when the present research commenced. Content analysis was conducted with the following main goals: (3) the thematic exploration of *Magyar Pedagógia*, mostly by means of quantitative content and textual analyses and data mining; (4) the detection and identification of certain trends, discourses characterising the different decades of the existence of the journal and the changes related thereto. This journal with its hundred-year history provided a great opportunity to explore the main research directions in Hungarian educational research. The methods applied, the period covered and the quantity of analysed textual data in this study all led to novel results in social science research.

RESEARCH QUESTIONS AND HYPOTHESES

Analyses of metadata

- 1) How did the number of publications and the average length of articles change over time?
- 2) Who were the authors publishing the largest number of articles? Who had the most powerful impact on other authors?
- 3) What was the institutional background of these authors like?
- 4) How did the proportion of articles written by multiple authors change? Which authors usually published articles in cooperation?
- 5) How did the proportion of female and male authors change with time?

H₁: It is hypothesized that there has been a rise in the number and the length of publications since the change of the regime as a result of the massification of higher education (Oktatási Hivatal, 2017, Polónyi, 2008) and the spreading of doctoral training in social sciences (Inzelt & Csonka, 2018), which must have intensified the interest in educational research.

H₂: On the basis of Biró's (2009b) findings, it is hypothesized that the scholars, who work within a cognitive psychological framework, belonging to the Institute of Education of the University of Szeged, are among the most influential and productive authors; and their distinct group will be detectable.

H₃: When studying the scholars' institutional background, it is hypothesized that distinct groups will be found in the co-authorship graph (Molnár, Tóth, & Pintér, 2018a; 2018b).

H₄: Regarding co-authorship in publications, it is hypothesized, on the basis of international tendencies, that there was a rise in the number of articles written by multiple authors compared to the number of publication written by one person only (Fanelli & Larivière, 2016; Tóth, 2014; Inzelt, Schubert, & Schubert, 2008).

H₅: It is hypothesized that the ratio of female authors rose in the chosen journal in the studied period on the basis of the patterns of international trends affecting scientific performance (West, Jacquet, King, Correll, & Bergstrom, 2013; Mauleón, Hillán, Moreno, Gómez, & Bordons, 2013) and national statistics on higher education (Oktatási Hivatal, 2017; Eurostat, 2017; Polónyi, 2008).

Citation analyses

- 6) Which are the most cited references in the field of study? Does the distribution of references justify the significance of the four most prestigious educational research journals, and what is their order of importance?
- 7) On the basis of the references in Magyar Pedagógia, which books can be considered as reference works in educational research, and how are conference presentations and dissertations represented in references?
- 8) Who are the most influential and most cited authors? What is their international reputation like? Which institutions do the most cited authors belong to?
- 9) On the basis of the use of references, how did the distribution of such qualitative indicators as the freshness and the quality of references, the rate of international references and the interdisciplinary nature of the these references change with time? Is the number of references to journals with an impact factor growing or decreasing?

H₆: It is hypothesized that authors started to cite more articles from journals than other scientific genres (Bornmann & Mutz, 2015; Larsen & Von Ins, 2010), and the ratio of such cited articles was growing with time, while the ratio of other genres was falling continuously. According to the literature (Tóth, Toman, & Cserpes, 2008; Pap, 2009) and our expectations,

the four most cited Hungarian journals will be Magyar Pedagógia, Iskolakultúra, Educatio and Új Pedagógiai Szemle.

H₇: It is hypothesized that there are some especially influential edited volumes and monographs in the field of educational research, and such publications will hold prominent places in the list of the most cited references. Furthermore, a number of conference presentations and dissertations will also appear in citations (Mahood, Van Eerd, & Irvin, 2014); however, their role will be less significant.

H₈: The list of the most cited authors of Magyar Pedagógia will correlate with the list of authors having the highest number of publications in the journal and the most prominent scholars in educational research (members of the Hungarian Academy of Sciences, leading researchers and university professors etc.). In the light of the findings of Csapó (2016), the international reputation of the leading Hungarian researchers shall also be examined. Concerning the international significance of the most cited authors of Magyar Pedagógia, correlating results are expected with their reputation at the national level. Just like in the case of co-authorship graphs (Nagy & Molnár, 2017; Molnár, Tóth, & Pintér, 2018b), it is hypothesized that distinct groups will be found in the citation networks too as far as the researchers' institutional background is concerned.

H₉: Based on the results of Ware and Mabe (2015) and Bornmann and Mutz (2015), it is presumed that there was a rise in the number of references per article over the period studied, and the authors were likely to rely on fresher references. On the basis of the results of Csíkos, it was found worthwhile to consider whether there was a rise in the number of impact factor and non-impact factor international publications cited by the authors with the advancement of time. Moreover, a steady growth in the ratio of interdisciplinary sources in the journal was also expected after the findings published by Bárd (2009).

Content analyses

- 10) Which topics seemed to be the most central ones over the studied period and how did such topics change from decade to decade? Were the authors of the journal following the latest international research directions?
- 11) On the basis of the articles published in the journal, which periods can be identified in the history of the journal? When did marked thematic changes take place? As the consequences of prevailing political ideology, did gradual or dramatic thematic changes occur?
- 12) Did the distinct periods in the history of the journal overlap with the different periods of Hungarian educational research?

H₁₀: It is anticipated that there were noticeable thematic changes every decade throughout the history of the journal. It is hypothesized that topics in the journal followed international trends in educational research within a short period of time. This hypothesis is based on the information revealed by the study written by Mészáros (1992) discussing the history of the journal.

H₁₁: On the basis of the study of Mészáros (1992), it is also hypothesized that the impacts of the key questions dominating certain periods in educational research and current political ideology affected the content of publications from time to time, and such effects will be detectable, even as trends, during the analyses.

H₁₂: The distinct periods in the history of the journal that will be identified by content analysis overlapped with the different periods of Hungarian educational research (Németh, 2015, 2014, 2013; Németh, Biró, & Garai, 2015; Németh & Biró, 2009).

THE RESULTS OF THE RESEARCH

Results derived from the analysis of metadata

One of the findings of the assessment of the situation that was considered in the research is that the number of publications continually decreased in the Hungarian educational research journals, which refers to a fall in the productivity of Hungarian educational researchers in the past decades. This is not only true of Magyar Pedagógia but other journals as well. Taking into account Hungarian scholars' international publication practices too, it can be claimed that there is a general reduction in the number of publications in Hungary in the field of educational research. This particular result contradicts that part of the first hypothesis which proposes that the numbers of publications and authors were growing after the change of the regime. However, when studying other indicators of the articles published in Magyar Pedagógia, several positive tendencies are detectable: there was a growth in the length of articles published in the chosen journal; the quality of papers also improved; there was growing scientific cooperation between scholars, which can be proved by the increased proportion of the articles with multiple authorship when considered from a scientometric perspective. This tendency can also be identified at an international level, proving what was presupposed in the fourth hypothesis.

Another central aspect of the present research was to examine the authors of the journal in detail. A list was compiled of the names of the authors having the largest number of publications in the journal with their institutional affiliation and the numbers of occasions they were cited. Such detailed information has been lacking or sparse in the field of educational research so far. Besides listing the names of the most productive authors of Magyar Pedagógia, the names of the most central persons in the field of Hungarian educational research were also identified as regards the analysed journal. Relying on the information concerning the authors' institutional affiliation, the leading role of Eötvös Loránd University and the University of Szeged appears to emerge. These results also support the second hypothesis that the scholars who work within a cognitive psychological framework, belonging to the Institute of Education of the University of Szeged are among the most influential and productive authors. Furthermore, it was also hypothesized that there would be distinct groups of authors in the co-authorship graph. On the basis of the data gathered after the analysis of such graphs, it can be said that there are some distinct groups which can be identified easily, mainly according to the institutions the authors belong to. Moreover, several sub-networks could also be detected within certain institutions. However, these sub-networks are ultimately linked by central connecting authors, which factor usually binds the groups together. Therefore we can claim that strong institutional segmentation is not typical of the authors of the journal. The research results prove that there is real scientific cooperation among Hungarian educational researchers.

The following important tendencies, which affect the journal considerably, were also identified in the research: the number of foreign authors decreased whereas the number of female authors rose in the studied period. After considering the institutional background of the authors, it was discovered that the number of academicians is noticeably larger than the number of other authors belonging to organisations of educational administration or institutes of public education. It would be desirable in the future to have more authors from the latter group as it would reflect the fact that the number and importance of research teachers is growing.

The change in the proportion of articles written by multiple authors was also examined on the basis of metadata. In line with international tendencies, there was a considerable growth in the number of articles written by multiple authors, which proved that scientific cooperation had become increasingly important in Hungary as well. This was basically a positive tendency, and it might indicate the beginning of a process which will be more accurately detectable in the future. Such growth may also be considered as one reflecting a change in quality, because publications of this kind usually summarize the findings of more extensive and more

comprehensive research which has been conducted by larger research groups of scholars in partnership to succeed. Such groups sometimes even collaborate with interdisciplinary professionals, which can contribute to the development of the field of research and active cooperation between different institutions too.

Citation analyses

In the light of the findings of the study, it can be declared that scientometrics offers new and relevant tools for scholars conducting educational research. One of the aims of the present research was to help bridge the gap caused by the lack of representation of Hungarian educational research journals in international citation databases. In fact, only a small number of articles published in foreign journals, meeting the strict requirements of international scientometric databases, is included in such citation databases. Csapó (2016) pointed out that such scholarly output in Hungary is extremely low compared to the scientific performance in other European countries. This also implies that only a tiny section of scientific work and output can be examined by means of professional scientometric tools. The present research therefore offers a novel approach to investigate any journal in detail, despite the fact that it only focuses on one journal only.

On the basis of the results of the citation analyses carried out in the research project, the main changes identified took place throughout the past twenty-five years of existence of the journal. The results were in line with our expectations, as over time a growth in the number of references per article and a general decrease in the number of publications per year, which stopped at a point and stagnated thereafter, were identified. When studying the quality of references, it can be claimed that the scientific quality of the examined journal improved significantly with time. On the one hand, the amount of recent references grew. On the other hand, a growing number of journal citations, especially international ones, could be witnessed. On the basis of such results, a rise in the number citations of publications with an impact factor was anticipated too. Likewise, an increase in the rate of interdisciplinarity was also expected. However, concerning the latter phenomenon, fewer clear data were gathered. Although several references to interdisciplinary sources were detectable too, the authors of *Magyar Pedagógia* seemed more likely to cite publications from other journals of educational research.

As far as the most cited journals are concerned, the hypothesis of the study has proved to be correct, except for one item (*Educatio*). Moreover, the identification of the genres of publications was also accomplished in the study which revealed which were the most popular types of publications on the basis of the distribution of references per category. Among the thirty most cited publications, there were eight books, which confirmed the hypothesis. The assumption, however, regarding the proportion of conference presentations and dissertations could not be justified properly. Although these sources can be found in the list of references, their role seems to be much more marginal than expected.

In the case of the most cited authors, it is crucial to mention that a comparison was made to examine the significance of these authors at an international level. As the study relies heavily on citation databases, the findings related to the reputation of the most cited authors were compared to the levels of their international reputation on the basis of data available in the database of Scopus. This analysis also confirmed that there is a correlation between the list of most cited authors and the list of those authors who had the largest number of publications in the journal. Mixed results were achieved when the international reputation of the authors of *Magyar Pedagógia* was examined; however, the results could be explained in each and every case after thorough consideration.

Furthermore, a difficulty arose during the graphic illustration of the inner and the full citation networks due to their huge size, which we tackled by selecting the most important parts of the graphs. All the key aspects of these graphs were summarised in both cases, which makes

it possible to avoid the inclusion of the original graphic illustration of the networks including every small detail in the dissertation. In addition, when examining the hubs of the graphs, valuable information about the institutional background of the authors was also collected. Just like in the case of co-authorship graphs, it was hypothesized that distinct groups would be found in the citation networks too. However, no completely separated groups were detected in the case of these networks either. Moreover, when studying the inner citation network in respect of the authors' institutional background, one can clearly see that the authors belonging to the same institution are familiar with and prefer to cite each other's publications.

Content analyses

During content analysis, it was especially important to apply a number of different methods to be able to study the various aspects of the long period of time (114 years) and the enormous amount of information (50,000 pages), which was particularly difficult. In an attempt to overcome this difficulty, various perspectives were adopted. There were certain methods used for experimentation only, which meant that their full potential was not exploited; however, it is important to discuss the importance of the application of such methods applied in the study as well. The following methods belong to this group: the detection of similar publications, cluster analysis and additional visualisation techniques. The potentially useful secondary sources and the points of reference were also identified, but only some of them were used later during the content analyses.

Frequency analysis, the content analysis of keywords, the detection of similar articles, content categorization, trend analysis and terminology analysis were techniques that completed each other and revealed valuable information about the same text from different perspectives. Frequency analysis is one of the most basic methods of content analysis to determine the relative and the absolute frequency of words in texts. An attempt was made to work out methods that are based on such an approach. To this end, frequency lists and heat maps were used for data visualization. Another bunch of results was obtained when the words and terminology typical of each decade were studied. In such analysis, we also relied on keywords related to articles extracted from various secondary sources. The visualization of automatically identified keywords led to novel results. The relevance of such findings is justified by the fact that they verify the results of the analysis of different external secondary sources. The different approaches applied in the research complement each other and strengthen the validity of the methodology.

The most important thematic areas were also identified in the history of the journal, and the changes in the ratio of one area to the other over time were also examined in detail. The beginning of the 1900's, World War II, (1961 – the relaunch of the journal) and the change of the regime were identified as dates or events each marking the beginning of a new period in the history of the journal. Furthermore, international influences were also detectable when the three main thematic changes were identified in the journal's history. Having ample knowledge of the historical and political contexts of the aforementioned times, it can be claimed that the main thematic changes took place as a result of ideological changes, except for the changes taking place at the beginning of the 1900's which were rather due to professional influences. Finally, as it was hypothesized, rather gradual thematic changes were detectable throughout this long period.

When the distinct decades were examined, noticeable changes were identified in the main topics of the journal. The change of topics in Magyar Pedagógia followed international trends and new research directions in educational research within a short period of time. Magyar Pedagógia must be considered as one of the leading educational research journals due to the fact that its authors were always given the opportunity to react to the newest international literature without delay on the pages of the journal, introducing the most up-to-date research

directions in the field. It was suspected that both the key questions dominating certain periods in educational research and the current political ideology had affected the content of publications in Magyar Pedagógia from time to time. It was hypothesized that these two perspectives must always have affected the way publications were written throughout the history of the journal. This assumption was justified in most of the cases. However, after the change of the regime, no publications with political content or articles discussing the questions of education management were published in the journal. Instead, it was the period in the history of Magyar Pedagógia when authors were expected to meet strict professional requirements. Finally, in line with the case of the hypothesis concerning the changes of the periods in the history of the journal, content analysis revealed that such periods coincided with the main periods of Hungarian educational research.

CONCLUSION

It would be essential to conduct scientometric analyses in each field of science. Through such examinations, the inner characteristics, conditions and relations of a certain field are possible to be studied and revealed. Such knowledge is crucial to the advancement of various fields of research. The same holds true for educational research too. The relevance of the present study is justified by the remarks and recommendations sent to us regarding the extension of this research to include the articles of leading Hungarian scientific journals, conference publications, edited volumes and monographs. Due to the fact that the study is based on three pillars, there are three further research directions as well. However, this goes beyond the scope of this research project, and what has been accomplished so far might be seen as the first steps taken on the road to revealing valuable information about the field of educational research as a whole.

This research can be considered as a pilot project in the field of educational data mining, with the professional and methodological limitations usually being typical of such projects. This study is a novelty in Hungarian educational research as no such a large corpus of scientific texts has been analysed by means of automated data mining, and an unconventional approach was also adopted to content analysis here, which makes the present investigation different from previous studies which usually applied more conventional, rather manual content analysis methods. To conclude, such an innovative approach might bring new opportunities in the field of educational research. Despite its limitations, one of the biggest benefits of this approach is that it has proved to be universally applicable. Moreover, the integrative analysis of different types of data makes the exploration innovative too as it makes it possible to analyse metadata, citations and various contents of articles at the same time.

REFERENCES

- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
- Bárd, E. (2009). Referenciaszemélyek a magyar, a német és a francia neveléstudományi diskurzusokban. In Németh A. & Biró Zs. H. (Eds.), *A magyar neveléstudomány a 20. század második felében* (pp. 227–244). Budapest: Gondolat.
- Biró, Zs. H. (2009a). A magyar neveléstudományi kommunikáció szereplői. Tudományszociológiai elemzés a központi pedagógiai folyóiratok szerzőiről (1997–2006). In Németh, A. & Biró, Zs. H. (Eds.), *A magyar neveléstudomány a 20. század második felében* (pp. 168–209). Budapest: Gondolat.
- Biró, Zs. H. (2009b). A magyar neveléstudományi kommunikáció jellemzői (1997–2006), Összehasonlító tudományszociológiai elemzés, avagy kísérlet egy nemzetközi kutatás adaptációjára. *Magyar Pedagógia*, 109(1), 49–76.
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916), 892–895. <https://doi.org/10.1126/science.1165821>
- Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215–2222. <https://doi.org/10.1002/asi.23329>
- Bornmann, L., Thor, A., Marx, W., & Schier, H. (2016). The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute. *Journal of the Association for Information Science and Technology*, 67(11), 2778–2789. <https://doi.org/10.1002/asi.23627>
- Börner, K. (2010). *Atlas of science*. Cambridge, MA: MIT Press.
- Börner, K. (2015). *Atlas of knowledge: Anyone can map*. Cambridge, MA: MIT Press.
- Csapó, B. (2005). A Magyar Pedagógia feladata a neveléstudomány fejlesztésében. *Iskolakultúra*, 15(4), 3–10.
- Csapó, B. (2011). Az oktatás tudományos hátterének fejlődése. *Magyar Tudomány*, 172(9), 1065–1076.
- Csapó, B. (2016). A tanárképzés és az oktatás fejlesztésének tudományos háttere. *Iskolakultúra*, 26(2), 3–18. <https://doi.org/10.17543/iskkult.2015.11.3>
- Csíkos, Cs. (2013). Az impakt faktor a neveléstudományi kutatásokban. *Iskolakultúra*, 23(9), 3–16.
- David, D., & Frangopol, P. (2015). The lost paradise, the original sin, and the Dodo bird: a scientometrics Sapere Aude manifesto as a reply to the Leiden manifesto on scientometrics. *Scientometrics*, 105(3), 2255–2257. <https://doi.org/10.1007/s11192-015-1634-2>
- Eurostat. (2017). *Felsőfokú oktatási statisztikák*. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tertiary_education_statistics/hu
- Fanelli, D., & Larivière, V. (2016). Researchers' individual publication rate has not increased in a century. *PLoS One*, 11(3), e0149504. <https://doi.org/10.1371/journal.pone.0149504>
- Fehér, P., & Aknai, D. O. (2016). Aki "doktor" akar lenni..., avagy doktorjelöltek publikációs teljesítményének összehasonlítása a hazai neveléstudományi doktori iskolákban. In

- Molnár, Gy. & Bús, E. (Eds.), *PÉK 2016 = [CEA 2016]: XIV. Pedagógiai értékelési Konferencia = [14. Conference on Educational Assessment]: Program; Előadás-összefoglalók = [Program; Abstracts]*. (pp. 53). Szeged: SZTE BTK Neveléstudományi Doktori Iskola.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: University Press.
- Garfield, E. (2009). From the science of science to Scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics*, 3(3), 173–179. <https://doi.org/10.1016/j.joi.2009.03.009>
- Garfield, E., & Merton, R. K. (1979). *Citation indexing: Its theory and application in science, technology, and humanities*. New York: Wiley.
- Glänzel, W., Schubert, A., & Czerwon, H. J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985–1995). *Scientometrics*, 45(2), 185–202. <https://doi.org/10.1007/bf02458432>
- Hajjem, C., Harnad, S., & Gingras, Y. (2006). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *ArXiv preprint*, cs/0606079
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429. <https://doi.org/10.1038/520429a>
- Hood, W. W., & Wilson, C. S. (2001). The Literature of Bibliometrics, Scientometrics, and Informetrics. *Scientometrics*, 52(2), 291–314. <https://doi.org/10.1023/a:1017919924342>
- Hung, J. L. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology*, 43(1), 5–16. <https://doi.org/10.1111/j.1467-8535.2010.01144.x>
- Inzelt, A., Schubert, A., & Schubert, M. (2008). Incremental citation impact due to international co-authorship in Hungarian higher education institutions. *Scientometrics*, 78(1), 37–43. <https://doi.org/10.1007/s11192-007-1957-8>
- Inzelt, A., & Csonka, L. (2018). A PhD-fokozatot szerzettek karrierje a társadalom-és humán tudományokban. *Educatio*, 27(1), 30–45. <https://doi.org/10.1556/2063.27.2018.1.3>
- Ivanović, L., & Ho, Y. S. (2017). Highly cited articles in the Education and Educational Research category in the Social Science Citation Index: a bibliometric analysis. *Educational Review*, 1–10. <https://doi.org/10.1080/00131911.2017.1415297>
- King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of information science*, 13(5), 261–276. <https://doi.org/10.1177/016555158701300501>
- Klamarik, J., & Suppán, V. (1892). A Magyar Paedagogiai Társaság Alapszabálya. *Magyar Paedagogia*, 1, 124–128.
- Kostoff, R. N. (1998). The use and misuse of citation analysis in research evaluation. *Scientometrics*, 43(1), 27–43. <https://doi.org/10.1007/bf02458392>
- Larsen, P. O., & Von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3), 575–603. <https://doi.org/10.1007/s11192-010-0202-z>
- Mahood, Q., Van Eerd, D., & Irvin, E. (2014). Searching for grey literature for systematic reviews: challenges and benefits. *Research synthesis methods*, 5(3), 221–234. <https://doi.org/10.1002/jrsm.1106>

- Mauleón, E., Hillán, L., Moreno, L., Gómez, I., & Bordons, M. (2013). Assessing gender balance among journal authors and editorial board members. *Scientometrics*, 95(1), 87–114. <https://doi.org/10.1007/s11192-012-0824-4>
- Merton, R. K., & Garfield, E. (1986). Foreword to little science, big science and beyond. In D. J. de S. Price (Eds.), *Little Science, Big Science and Beyond* (pp. vii–xiii). New York: Columbia University Press.
- Mészáros, I. (1992). Magyar Paedagogia - Magyar Pedagógia. 1892-1992. *Magyar Pedagógia*, 92(1), 5–24.
- Moed, H. F. (2009). New developments in the use of citation analysis in research evaluation. *Archivum immunologiae et therapeuticae experimentalis*, 57(1), 13. <https://doi.org/10.1007/s00005-009-0001-5>
- Molnár, P. (2018). Társszerzői együttműködések hálózatai a bírálati rendszerű angol nyelvű neveléstudományi szakfolyóiratokban: európai helyzetkép. In Vidákovich T. & Füz N. (Eds.), *PÉK 2018 [CEA 2018] XVI. Pedagógiai Értékelési Konferencia [16th Conference on Educational Assessment]: Program és összefoglalók [Programme and abstracts]* (pp. 69.). Szeged: SZTE Neveléstudományi Doktori Iskola.
- Molnár, P., Pintér, H., & Tóth, E. (2018a). Co-authorship networks in Hungarian educational journals: 1991-2016: *Data Stories 2018 Research Visualization Exhibition at CEU*. Retrieved from http://j.mp/hunedunet_91_16_datastories18
- Molnár, P., Pintér, H., & Tóth, E. (2018b): A neveléstudomány társszerzői együttműködéseinek hálózatai hazai és nemzetközi folyóiratokban. *Magyar Pedagógia*, 118(4). Manuscript submitted for publication.
- Molnár, P., Tóth E., & Pintér, H. (2018a). A közös alkotás társszerzői mintázatai és dinamikája a neveléstudományi tudás megosztásának és gondozásának hazai formális fórumain, a vezető szakfolyóiratokban 1991 és 2016 között. In Veszelszki Á. & Cser N. (Eds.), *Tudománykommunikáció konferencia: Absztraktkötet* (pp. 22.) Budapest: Budapesti Corvinus Egyetem Magatartástudományi és Kommunikációelméleti Intézet.
- Molnár, P., Tóth, E., & Pintér, H. (2018b). Töredezettség és összekapcsoltság a hazai neveléstudományi folyóiratok társszerzői együttműködési hálózataiban. In Vidákovich T. & Füz N. (Eds.), *PÉK 2018 [CEA 2018] XVI. Pedagógiai Értékelési Konferencia [16th Conference on Educational Assessment]: Program és összefoglalók [Programme and abstracts]* (pp. 70.) Szeged: SZTE Neveléstudományi Doktori Iskola.
- Molnár, P., Tóth, E., & Pintér, H. (2018c): A neveléstudomány társszerzői együttműködéseinek hálózatai hazai és nemzetközi folyóiratokban. *Jel-Kép*, (4) 19–33. <https://doi.org/10.20520/JEL-KEP.2018.4.19>
- Nagy, P. T. (2011). Az oktatásról szóló tudomány kettős nyitottsága. *Magyar Tudomány*, 172(9), 1077–1089.
- Németh, A. (2013). A neveléstudomány főbb fejlődésmodelljei és tudományos irányzatai. *Neveléstudomány*, 1(1), 18–63.
- Németh, A. (2014). *A magyar neveléstudomány fejlődése. Nemzetközi tudományfejlődés és recepciós hatások, egyetemi tudományvá válás, középiskolai tanárképzés*. Budapest: Gondolat.
- Németh, A. (2015). A neveléstudomány nemzetközi modelljei és tudományos irányzatai. *Magyar Pedagógia*, 115(3), 255–294. <https://doi.org/10.17670/mped.2015.3.255>

- Németh, A., & Biró, Zs. H. (Eds.). (2009). *A magyar neveléstudomány a 20. század második felében*. Budapest: Gondolat.
- Németh, A., Biró, Zs. H., & Garai, I. (Eds.). (2015). *Neveléstudomány és tudományos elit a 20. század második felében*. Budapest: Gondolat.
- Németh, A., Biró, Zs. H., & Varga, K. (2009). Kísérlet egy nemzetközi kutatás adaptációjára. In Németh, A. & Biró, Zs. H. (Eds.), *A magyar neveléstudomány a 20. század második felében* (pp. 147–167). Budapest: Gondolat.
- Oktatási Hivatal. (2017). *Felsőoktatási statisztikák*. Retrieved from https://www.oktatas.hu/felsooktatas/kozerdeku_adatok/felsooktatasi_adatok_kozzetetele/felsooktatasi_statisztikak
- Pap, K. T. (2009). Folyóirat-hivatkozások a neveléstudományi folyóiratokban. In Németh A., & Biró Zs. H. (Eds.), *A magyar neveléstudomány a 20. század második felében* (pp. 210–226). Budapest: Gondolat.
- Polónyi István (2008): *Oktatás, oktatáspolitikai, oktatásgazdaság*. Budapest: Nemzeti Tankönyvkiadó.
- Price, D. de S. (1979). *Kis tudomány, nagy tudomány*. Budapest: Akadémiai Kiadó.
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (Eds.). (2010). *Handbook of educational data mining*. Boca Raton: CRC Press.
- San Francisco Declaration (2013). Putting Science into the Assessment of Research. *Microbe Magazine*, 8(12), 478–479. <https://doi.org/10.1128/microbe.8.478.1>
- Schriewer, J., & Keiner, E. (1993). Kommunikationsnetze und Theoriegestalt: Zur Binnenkonstitution der Erziehungswissenschaft in Frankreich und Deutschland. In Schriewer, J., Keiner, E., & Charle, Ch. (Eds.), *Sozialer Raum und akademische Kulturen / A la recherche de l'espace universitaire européen* (pp. 277–341). Frankfurt am Main: Peter Lang.
- Schubert, A. (2015b). Az impaktfaktor és akiknek nem kell. *Orvosi Hetilap*, 156(26), 1065–1069. <https://doi.org/10.1556/650.2015.30212>
- Schubert, A. (2018). Új utak és útvesztők a tudományos publikálásban és a tudományometriában. *Könyv és nevelés*, 20(1), 7–29.
- Szücs K., & Kiss A. (2015). Adatbányászati módszerek alkalmazása oktatási környezetben keletkezett adatokon. In Szlávi, P. & Zsakó, L. (Eds.), *INFODIDACT 2015 – Zamárdi, Magyarország, 2015. október 26*. Budapest: Webdidaktika.
- Tikk, D. (Eds.). (2007). *Szövegbányászat*. Budapest: Typotex.
- Tóth, I. J., Toman, N., & Cserpes, T. (2008). *A magyar neveléstudomány tudományos aktivitásának felmérése empirikus eszközökkel – Kutatási zárótanulmány*. Budapest: Wargo Közgazdasági Elemző- és Piackutató Intézet.
- Tóth, J. (2014). Szimulakratív szignifikáció és szimulakratív szinterek-egy deviáns publikációs gyakorlat vizsgálata. *Információs Társadalom*, 2014(2), 24–44.
- Ueno, M. (2004). Data mining and text mining technologies for collaborative learning in an ILMS. In *Advanced Learning Technologies, Proceedings*. IEEE International Conference (pp. 1052-1053). <https://doi.org/10.1109/icalt.2004.1357749>
- Van Raan, A. F. (2001). Competition amongst scientists for publication status: Toward a model of scientific publication and citation distributions. *Scientometrics*, 51(1), 347–357. <https://doi.org/10.1023/a:1010501820393>

- Ware, M., & Mabe, M. (2015). *The STM report: An overview of scientific and scholarly journal publishing*. 4th edition, International Association of Scientific, Technical and Medical Publishers. Retrieved from https://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf
- Weller, M., Jordan, K., DeVries, I., & Rolfe, V. (2018). Mapping the open education landscape: citation network analysis of historical open and distance education research. *Open Praxis*, *10*(2), 109–126. <https://doi.org/10.5944/openpraxis.10.2.822>
- West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PloS one*, *8*(7), e66212. <https://doi.org/10.1371/journal.pone.0066212>
- Wong, S. L., & Hodson, D. (2010). More from the horse's mouth: What scientists say about science as a social practice. *International Journal of Science Education*, *32*(11), 1431-1463. <https://doi.org/10.1080/09500690903104465>
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827) 1036-1039. <https://doi.org/10.1126/science.1136099>
- Youtie, J., Solomon, G. E., Carley, S., Kwon, S., & Porter, A. L. (2017). Crossing borders: A citation analysis of connections between Cognitive Science and Educational Research and the fields in between. *Research Evaluation*, *26*(3), 242–255. <https://doi.org/10.1093/reseval/rvx020>

**PUBLICATIONS AND CONFERENCE PRESENTATIONS RELATED TO THE TOPIC OF THE
DISSERTATION**

- Nagy, G. (2014). Megy-e a digitalizálás által a világ elébb? Avagy mi végre digitalizálunk? *Információs Társadalom*, 14(3), 44–52.
- Nagy, G. (2016a). Tudománymetria és neveléstudomány. *Iskolakultúra*, 26(2), 50–62. <https://doi.org/10.17543/iskkult.2016.2.50>
- Nagy, G. (2016b). Tudománymetriai és tartalmi elemzések szövegbányászati módszerekkel. In Mikusné Sárvári, K. (Ed.), *Networkshop 2016, Debrecen, 2016.03.29–2016.04.01*. Budapest: NIIFI.
- Nagy, G. (2016c). A Magyar Pedagógia folyóirat tudománymetriai és tartalmi elemzése. In Molnár, G., & Bús, E. (Eds.), *PEK 2016. XIV. Pedagógiai értékelési Konferencia [CEA 2016. 14 TH Conference on Educational Assessment]* (pp. 52–52). Szeged: SZTE Neveléstudományi Doktori Iskola.
- Nagy, G. (2016d). A Magyar Pedagógia folyóirat tudománymetriai és tartalmi elemzése szövegbányászati módszerrel. In Zsolnai, A., & Kasik, L. (Eds.), *XVI. Országos Neveléstudományi Konferencia Program és absztrakt kötet* (p. 273). Szeged: MTA Pedagógiai Bizottság, SZTE BTK Neveléstudományi Intézet.
- Nagy, G. (2017). A neveléstudomány kutatási tendenciái a Magyar Pedagógia negyed évszázadának tükrében. In Kerülő, J., Jenei, T., & Gyarmati, I. (Eds.), *XVII. Országos Neveléstudományi Konferencia: Program és absztrakt kötet* (pp. 53). Nyíregyháza: MTA Pedagógiai Tudományos Bizottság, Nyíregyházi Egyetem.
- Nagy, G., & Molnár, G. (2017). A Magyar Pedagógia folyóirat tudománymetriai elemzése: tendenciák, szerzők, társszerzőségi együttműködés. *Magyar Pedagógia*, 117(1), 5–27. <https://doi.org/10.17670/mped.2017.1.5>
- Nagy, G., Molnár, S., & Kokas, K. (2017). A könyvtárak jövőjéről – Reflexiók egy előzetes kutatási jelentés kapcsán. *Tudományos és Műszaki Tájékoztatás*, 64(2), 64–82.
- Nagy, G. (2018a). Citation network analysis of an educational research journal. In Vidákovich, T., & Füz, N. (Eds.), *PÉK 2018 [CEA 2018] XVI. Pedagógiai Értékelési Konferencia [16th Conference on Educational Assessment]: Program és összefoglalók [Programme and abstracts]* (pp. 72–72). Szeged: SZTE Neveléstudományi Doktori Iskola.
- Nagy, G. (2018b). Text Mining-based Scientometric Analysis in Educational Research. In *The European Conference on Education 2018 Official Conference Proceedings* (pp. 129–142). Nagoya, Japán: International Academic Forum.
- Nagy, G., & Molnár, G. (2018). A Magyar Pedagógia folyóirat tudománymetriai elemzése a hivatkozási szokások és a hivatkozási hálózatok tükrében. *Magyar Pedagógia*, 118(3), 203–235.