READING RESEARCH AND DIGITAL READING RESEARCH: an overview of the current scientific scenario

Taisa Dantas*
Almudena Mangas-Vega**
Raquel Goméz-Díaz***
José Antonio Cordón-García****

ABSTRACT: Historically, Reading Research has always been a broad and multidisciplinary research field, motivated by the complexity of the act of reading. Based on a review of the literature, the aim of this paper is to provide a current overview of Reading Research, starting from the study of new factors that have influenced the growth of this field. To this end, a bibliometric analysis was carried out, focusing on four main points: areas of knowledge; the most prolific authors; the contrast between Reading Research and Digital Reading Research; and the main topic areas being studied. The methodology used involves an extensive collection of data originating from different sources, data analyses, the calculation of efficiency rates and the solutions used to address disparity in the amount and type of data found. This study confirms that the onset of the use of technological devices for reading stimulated the growth of Reading Research. Since 2009, when the use of the Kindle e-reader (Amazon) stabilized, preceding the launch of the iPad tablet (Apple) in 2010, digital reading played an important role in Reading Research, attracting new topics and research prospects.

Keywords: Reading, Reading Research, Digital reading, Scientific publication, Reading devices

I INTRODUCTION

Reading Research is a field of study comprised of multiple focus points and perspectives. The multidisciplinary collection of studies available on reading is directly related to the fact that reading is a complex act, a neurological and cognitive process with social implications.

The interest in the phenomenon of reading is not recent, and entails, to a certain point, the history of the book industry, but more specifically the expansion of educational policies.

Hence, the history of Reading Research consists of important frameworks that range from the actions carried out by the Humanist movement in Italy in the 16th century in relation to literacy to the influence of psychological components associated with the actual act of reading.

These frameworks involve four different disciplines that have acted as the precursors of Reading Research, which have contributed to the multidisciplinary nature of the field. These disciplines are: Education, Ophthalmology, Psychology, and Bibliology. (CORDÓN-GARCÍA E LÓPEZ-CÓZAR, 1990; POULIOT, 2009; RAYNER, 1998; VENEZKY, 1984; WADE E TATLER, 2009)
Since Reading Research is multidisciplinary, it is often encompasses different types of objectives and perspectives, which have led to many different types of scientific studies being carried out. Therefore, to organize this plethora of information in a more targeted way, the field was divided into three main categories: process, practice, and instruction.

The early studies focused on the process of reading and as a result received more visibility and prestige (VENEZKY, 1984). Consequently, at this time, most knowledge concerning reading did not affect the actual act of reading, but rather contributed to other scientific advances and served to understand the fundamental basis of the process, mainly by groups associated with the field of Psychology.

However, during the last decades, new research groups have approached this subject in a more meaningful way and have started to investigate other aspects of reading.

By the early 1990s “reading researchers began to draw from a variety of social-science disciplines – most noticeably, sociology and anthropology. In the process, reading took on social, cultural, and multicultural dimensions” (KAMIL et al., 2000). The development of data collection methods, instruments of analysis, and the perception of the importance of reading from a social perspective, both cultural and inclusive of different types of people, have generated a considerable increase in the number of disciplines focusing on Reading Research in the last 20 years.
This growth was more or less stable up to 2009, until the introduction of a new technological factor which caused the number of published studies to gain a different momentum.

This technological factor involved the appearance of mediation devices used in the practice of reading such as computers, e-readers, tablets, phones, smartphones, Phablets, among others. The growing use of electronic devices has produced a new phenomenon in relation to the reading process. Nowadays, the printed book is not an exclusive medium and new media require greater skills on the part of the reader. These skills go beyond traditional literacy. In addition to decoding the code by reading, the reader must also learn how to handle a mediation device. In the case of a book, the reader handles the object in an almost intuitive way.

Furthermore, Reading Research has also started to include the analysis of the appropriation process of electronic devices, the applications used for reading, and the characteristics of digital environments that affect the act of reading.

The transformation experimented by the editorial industry, the information systems, and communication in the last years, with an intense migration from analog to digital, has forced a rethinking of the way that we conceive documents, reading, writing, creation, and investigation (CORDÓN-GARCÍA et al., 2013).

After analyzing how the areas of knowledge dedicated to the study of reading have evolved during 1980 and 2016 (Figure 2), it is possible to observe a significant increase in the last 7 years. This increase coincides with the appearance of new gadgets used to read written text, and in particular with the time points which coincide with the stabilization of the use of the Kindle e-reader (Amazon) (the Kindle was released only in the US in 2007 and worldwide in 2009). Moreover, an additional increase can be observed between 2010 and 2013, which coincides with the launch of Apple’s iPad (2010).

The transition from printed to digital text has affected many levels of society, and
of course, has also affected the field of Reading Research. This field of study is now approached in a different way, which involves the presence of a transversal category known as appropriation process analysis.

This study aims to analyze other factors that are new to Reading Research, such as authors, subjects and understanding the importance of Digital Reading Research, and these are the areas that will be addressed in order to construct a general overview of the current scientific scenario regarding this field.

2 MAIN PURPOSE

Due to great changes in the reading environment, motivated by the frequent use of digital devices, the field of Reading Research currently includes new research areas. The main purpose of this paper is to provide a current overview of Reading Research and focuses on four main points:

• The areas of knowledge devoted to reading studies.
• The most prolific authors.
• The contrast between Reading Research and Digital Reading Research.
• The main subjects being studied.

To understand the complexity of this field (mainly due to being multidisciplinary), this study proposes an initial analysis, a general overview, as a means to justify the decision to analyze these four main points, which will require further investigation.

3 METHODOLOGY

The first step was to search various literature databases and other bibliographic sources to identify publications related to the study of reading:

• Scientific Databases: Web of Science – WOS (Thomson Reuters); SCOPUS (Elsevier); LISA and MLA (Proquest); LISTA and ERIC (Ebsco)
• Another information sources: Biblioteca Nacional de España (BNE); British Library; IFLA; ISBN Database; Library of Congress; ResearchGate; Mendeley; UNESCO.

During data selection two difficulties emerged: the amount of data on the subject and the data fields provided by each of the information sources. The main challenge was the problem of data gaps: the different databases did not provide the same results and the other information sources used different organizing systems.

Therefore, it was necessary to design a way to achieve the maximum possible objectivity for each of the information sources.

As previously stated, the field of Reading Research is extremely broad; thus, deciding which terms to use in the search was considerably important. Since the terms act as keywords, they consequently act as filters, sometimes excluding relevant papers or including studies that could be relevant, but related to other fields of research (FINELLI, BORREGO, RASOULIFAR, 2015).

However, at this point, the goal was for the search to be as extensive as possible, with the least amount of noise, and included the following search string:

• To research about “reading”: ["reading research" OR "reading investigation") OR ("investigación" Y "lectura")]
• To research about “digital reading”: ["digital reading" OR “digital reading research”)
• Special filters were applied:
  • The search carried out using the WOS was done using core collections. The data provided was more detailed, and its accuracy was confirmed after comparing the results with those obtained using the other databases.
  • The filter “scientific journals only” was selected in LISA and MLA.
  • The filter “peer reviewed journals only” was selected in LISTA and ERIC.
  • Manual filtering was used in all searches with the various databases to eliminate the results irrelevant to this study.

The time frame used as a search filter was from 2009 (an important year for digital reading, related to the stabilization of the use of the Kindle [Amazon] and prior to the release of Apple’s iPad [2010]) to January 2016. The data collection took place in January 2016. The time frame was also used when analyzing the progress
of areas of knowledge (Figure 2) and confirmed the importance of this period with respect to the multidisciplinary aspect of Reading Research.

3.1 Data processing and statistical analysis

After the data was selected, the results were processed using two different reference managers, Mendeley and RefWorks, and source code editors, such as Notepad ++, in order to standardize the data and to eliminate duplicated results. Then, the data were entered into an Excel table to be analyzed.

Once the results were included within the Excel table, useful indicators were selected and the data was standardized. These data fields were:

<table>
<thead>
<tr>
<th>Table 1: Data fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field</td>
</tr>
<tr>
<td>Details</td>
</tr>
<tr>
<td>Title</td>
</tr>
<tr>
<td>Primary Author</td>
</tr>
<tr>
<td>Other authors</td>
</tr>
<tr>
<td>If needed</td>
</tr>
<tr>
<td>Territory 1</td>
</tr>
<tr>
<td>In cases where the result specifies a territorial area, in most cases the data displayed is the place of publication of the article.</td>
</tr>
<tr>
<td>Territory 2</td>
</tr>
<tr>
<td>Expansion of data from the previous field</td>
</tr>
<tr>
<td>Keywords</td>
</tr>
<tr>
<td>Given by the author and / or collected by the information source</td>
</tr>
<tr>
<td>Area Knowledge</td>
</tr>
<tr>
<td>Area of Knowledge that the publication belongs</td>
</tr>
<tr>
<td>Funding</td>
</tr>
<tr>
<td>Information about the research funding: public / private.</td>
</tr>
<tr>
<td>Resource</td>
</tr>
<tr>
<td>Name of the source of information from which is extracted that result.</td>
</tr>
<tr>
<td>Publisher</td>
</tr>
<tr>
<td>Reading / Digital</td>
</tr>
<tr>
<td>Refers to the search that has been localized the publication, “reading research” or “research in digital reading”</td>
</tr>
<tr>
<td>Reference Type</td>
</tr>
<tr>
<td>Document type (article, book, ...)</td>
</tr>
<tr>
<td>Journal</td>
</tr>
<tr>
<td>Name of the journal that published the article</td>
</tr>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>Contains the summary of the publication, in cases where such data is available.</td>
</tr>
</tbody>
</table>

Source: Own elaboration

Due to the lack of data in many fields, some were eliminated and therefore not included within the study. In all aspects studied, it was necessary to calculate an effective rate, which was achieved using the following formula (PORTAL, 2005):

\[
Er = \left( \frac{Dc}{Td} \right) \times 100
\]

Where \( Er = \) effectiveness rate, \( Dc = \) data collected and \( Td = \) total data.

Even after allowing for these corrections, it is necessary to note that making the adjustments automatically using the reference manager software also produced some errors: the elimination of duplicated data due to how the database processed special characters or errors caused by the information sources when assigning data. For these reasons, a manual review was conducted in some specific cases where these types of errors could distort the final results (such as calculations related to the authors field). In total, 2096 publications were selected once all the data were organized.
4 RESULTS AND DISCUSSION

The results of the analysis were organized into five main sections: areas of knowledge; distribution of the publication over the years; authors; reading vs digital reading; keywords. These sections were defined according to the objectives previously indicated.

4.1 Evolution of areas of knowledge

Numerous problems related to the disparity in the data collected from the different information sources were found, making it difficult to measure the progress of the areas of knowledge in which this subject had been studied over the years. Following the lead of other authors studying different areas of knowledge (BELMONT et al., 2013), only the data provided by the Core Collection of the WOS was used.

Furthermore, it should be noted that although the WOS refers to these data as areas of knowledge their classification does not follow the traditional UNESCO or OECD classification, but follows a more specific classification, commonly used by Thomson Reuters.

In order to make the approach more understandable and manageable, this section will only discuss the areas of knowledge derived from the WOS database.

As previously mentioned, the time frame analyzed was from 1980 to January 2016 to be able to emphasize the progression of the areas of knowledge, and to show which years had the most growth regarding this field.

Thus, a method was designed to evaluate the evolution of the areas of knowledge over the years (increase or decrease in the number of different areas of knowledge, typology, etc.). The development of the areas of knowledge in Reading Research was:

![Figure 3 The number of publications and the distribution of the areas of knowledge per year](source: Web of Science (WOS). Own Elaboration (2016))
Regarding the publications found that were related to the various areas of knowledge, it was necessary to perform a manual filtering to make working with them more suitable for the focus of this study.

As a result, the data was structured in a pie chart representing the percentages of each area associated with Reading Research.

**Figure 4** Areas of knowledge of publications filtered

![Pie chart showing the distribution of areas of knowledge for publications filtered.](image)

*Source: Web of Science (WOS). Own elaboration (2016)*

It must be noted that, nowadays, Reading Research is a research field that consists of many more diverse areas of knowledge than before, where the two main disciplines dedicated to this type of research are still among the precursor disciplines related to the study of reading (Figure 1) which are Education and Psychology.

Although it is possible to find publications from other areas throughout the 80s and 90s, such as Information Science and Library Science, it is only after the year 2000 that other disciplines begin to occupy important positions, with the highest number of areas of knowledge found within 2012.

### 4.2 Productivity during the analyzed period: effective rate: 99,6%  

After carrying out the search for publications between 2009 and 2016 (JANUARY), other publications prior to this time point were found and included within the study because they were considered as studies that had had an impact on the field.

Among the data collected, 151 publications were registered during the first year. During the second year this number increased to 222. Despite some irregularities, as can be observed mainly in 2015, the general trend was that the number of publications on the subject increased.

Based upon the number of publications registered in January 2016, it can be predicted that this pattern of growth will continue in the coming years, confirming the continued growth of this field.
In the case of the most prolific authors, there was no distinction between first and second authors. After setting the limit to 7 items, the most prolific authors were:

Table 2 Most prolific authors and number of publications

<table>
<thead>
<tr>
<th>Authors</th>
<th>Nº of publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cordón-García, José Antonio</td>
<td>30</td>
</tr>
<tr>
<td>Alonso-Arévalo, Julio</td>
<td>19</td>
</tr>
<tr>
<td>Gómez-Díaz, Raquel</td>
<td>18</td>
</tr>
<tr>
<td>Chen, Chih-Ming</td>
<td>14</td>
</tr>
<tr>
<td>Buchanan, George</td>
<td>13</td>
</tr>
<tr>
<td>Pearson, Jennifer</td>
<td>9</td>
</tr>
<tr>
<td>Thimbleby, Harold</td>
<td>9</td>
</tr>
<tr>
<td>Linder, Daniel</td>
<td>9</td>
</tr>
<tr>
<td>Larson, Lotta C.</td>
<td>8</td>
</tr>
<tr>
<td>Zhang, Liyi</td>
<td>8</td>
</tr>
<tr>
<td>Allington, Richard L.</td>
<td>8</td>
</tr>
<tr>
<td>Ardoin, Scott P.</td>
<td>7</td>
</tr>
<tr>
<td>Baccino, Thierry</td>
<td>7</td>
</tr>
<tr>
<td>Conradi, Kristin</td>
<td>7</td>
</tr>
<tr>
<td>Nicholas, David</td>
<td>7</td>
</tr>
<tr>
<td>Binder, Katherine S.</td>
<td>7</td>
</tr>
</tbody>
</table>

Source: Own elaboration

In addition, the following groups of authors were established: [CORDÓN-GARCÍA, J.A.; ALONSO-ARÉVALO, J.; GÓMEZ-DÍAZ, R.; LINDER, D.], [PEARSON, J.; BUCHANAN, G.; THIMBLEBY, H] and [ARDOIN, SCOTT P.; BINDER, KATHERINE, S.]. These results indicate that research on reading and digital reading is not as widespread as originally thought. Also, sixteen of the researchers belonged to four different disciplines: Information Sciences; Computer Science; Education; and Psychology.

When comparing the disciplines that represented the areas where most of the reading studies took place between the 16th and 20th centuries (Figure 1) with the groups that were generated of the most productive authors (Table 1) and the areas of knowledge of publications (Figure 4), it can be seen that the introduction of technological supports as mediators was able to promote the consolidation of two recent disciplines among the most productive, which were Information Sciences and Computer Science.

The following dispersion graph of the total data shows that there are only a few authors who are continuously working or working in depth on this subject and that the vast majority of authors have only published 1 or 2 articles.
4.2.1 Co-authorship

Regarding collaboration between authors, the results were as follows:

Co-authorship Index: \( Ic = \frac{Caf}{Cd} \) where \( Caf \) indicates the number of authors receiving authorship and \( Cd \) the numbers of documents (PORTAL, 2005)

A manual review was previously performed to remove duplicate entries of the data to be used with the following calculation:

\[ Ic = \frac{3470}{1748} = 1.98 \]

This data is well below the average of the two areas that appear most within the results - Education (3.5) or Psychology (4.7) - according to data provided by the “Coauthor Index” tool 2014.  

These results seem to indicate that there is not a great deal of collaboration between authors regarding this subject. The reasons for this may be quite numerous, and may in fact be a subject that should be studied in the future; that is to say, collaborations within science. Although this issue may be difficult to measure, the study of collaborations could provide interesting data about how people work together with respect to this field (BARBA, 2003).

4.2.2 Contemporaneity of the subject depending on the number of authors

Contemporaneity of the subject is a calculation obtained as a derived conclusion from the Law of exponential growth of scientific information (PRICE, 1963). According to this, the number of scientists currently working on a subject must be much greater than those of the past, where past publications make up an almost nonsignificant proportion of the total number.

In this study, the calculation of contemporaneity of the subject was performed using the data collected from 2009 to January 2016, although, as mentioned earlier, data from 2007 and 2008 was also included. It was observed that 76% of the total number of authors that published material on the subject did so during the last three years (2013-2015).

Despite the short period of time measured, as can be seen in the graph in Figure 9, there is a sharp growth in the number of authors, indicating an increase in the popularity of the subject.
4.3 Reading vs digital reading

For a comparative analysis between the number of publications that refer to reading and the number of publications that refer to digital reading, the data was maintained with minimal processing to prevent bias. Despite this, there may be some distortion caused by certain automatic functions within the reference managers.

In the analysis, it was found that 45% of the publications referred to research on digital reading. This is a very high percentage of publications considering that the diffusion of the research on this topic is still low and more recent than the studies on reading in general.

To make a more accurate approximation of this comparison, the percentage of both reading and digital reading found within each resource was calculated:

Figure 9 Evolution of the number of authors (2007-2015)

Source: Own elaboration
Another finding was that the percentage of publications found regarding Digital Reading Research decreased as the information sources became less scientific and more general. This may occur because Digital Reading Research is not only a recent subject but also one that is constantly changing. Scientific information resources usually have a greater capacity for constant updating.

There are also two other possibilities for these results: 1) the most general resources are not able to capture this type of research and 2) the material itself has not yet reached the level of development and stability required for its appearance in these types of databases.

4.4 Keywords

The effective rate of keywords assigned by the authors or information sources was around 67%. Also, the use of keywords extracted from the abstracts did not improve the effective rate, thus keywords assigned by the authors / resources were used for this analysis.

In this case, it is important to analyze the keywords because they can provide valuable information about the approach of the research. Therefore, to study the keywords assigned, code editors were used and the keywords were submitted to some degree of stemming owing to the presence of typographical errors.

Figure 10 Percentage of papers found on Reading vs Digital reading for each information source.

Source: Research data (2016)
There were a small number of keywords that often appeared and a huge number of words that did not. The statistical mode of the data was one, which meant that the most repeated variable was one (most keywords appear in a single article). In fact, 63% of the words only appeared once. This data indicated a severe lack of standardization of the allocation of keywords, which could present additional problems during information retrieval.

The number of terms utilized to specify a specific subject matter was too great, even for a multidisciplinary subject such as reading (proved by the number of different areas of knowledge where publications are included). This indicated a lack of normalization or standardization when assigning the subject matter, which suggested that the recuperation of the scientific publications was not exhaustive in this area.

Nevertheless, there is a small group of keywords that are repeated in numerous works:

![Figure 11 Distribution of the occurrences of keywords](source: Research data (2016))

![Figure 12 Most frequent keywords and the number of occurrences](source: Research data (2016))
As shown in the methodology, the search chains were reading research and digital reading. Hence, the keyword “reading” should have appeared as a high frequency word that was considered as empty within this particular context. However, this was not the case and the word “reading” appeared in second position.

A deeper analysis of the data, focusing on words related to new technologies, digital environments and the rest of the words showed that almost half of the most frequent keywords were related to the digital age and new technologies.

Figure 13 Percentage of most frequent keywords. Special marking of keywords related to the “digital environment

![Pie chart showing percentage of most frequent keywords]

Source: Research data (2016)

This result also demonstrates that the lack of standardization when assigning a keyword negatively affects the chances of recovering all relevant documents in the searches.

This is what happens to the small group of keywords that do frequently appear in articles. However, to analyze the percentage of keywords related to the digital age from the total number of keywords assigned, there should not be any kind of standardization that could bias the results. The results of this new analysis are clearly shown in the following pie chart:

Figure 14 Keywords related and unrelated to digital. [Unprocessed keywords]

![Pie chart showing keywords related and unrelated to digital]

Source: Research data (2016)
Only 15% of the keywords were related to digital terminology or new technologies. As the percentage of publications from the areas of knowledge of the digital field was much higher (Figure 10), mismanagement by the authors and resources when selecting and properly applying keywords could be considered.

In the case of papers centered round a digital theme, these types of mistakes make recovery difficult.

5 CONCLUSIONS

Starting from 2009, a notable increase in the number of publications in Research Reading and Digital Reading Research can be observed, involving the number of articles, authors and areas of knowledge studying this field.

This data support that the appearance of electronic devices and the stabilization of their use for reading are important factors that have led to enhancement of this subject.

It has been found that Digital Reading Research assumes an important role in the study of the field of reading, and almost half of the data analyzed specifically addresses this topic. However, the visibility of Digital Reading Research is still limited.

This may be due to where the article is published, but, perhaps more importantly, to some of the problems identified in this analysis; especially those related to assigned keywords:

• The bad practice by information sources that do not assign keywords to articles;
• The misallocation or misuse of keywords,
• The lack of standardized allocation of keywords, even within the same source.

The lack of standardization with respect to the data fields of different resources and the categorization of publications by area of knowledge or thematic area is another barrier for this kind of studies.

Additionally, it was also possible to observe a sharp growth in the number of authors. However, most of the authors only had one publication during the time frame studied. This could be an indicator that the subject is receiving more attention, but requires more detailed studies. The same was observed in relation to the rate of co-authorship, which was also very low.
REFERENCES


