Associated conference: “Yes we can!” - Digital Education for Better Futures (EDEN 2023 Annual Conference)

Conference location: Dublin City University (DCU), Dublin, Ireland

Conference date: 18-20 June 2023


Published on: 27 October 2023

Copyright: © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
Abstract
Open, distance, and digital education (ODDE) has a long history concerning dropout. With the new reality of the digital age, the terminologies and concepts that have been long adopted are inevitably subject to change. To understand dropout in the digital transformation, it is important to take an evidence-based approach and examine the literature. Therefore, this study applies the steps of a systematic literature review to retrieve relevant articles on dropout since 2010 and conducts a content analysis via text mining. The results indicate that student dropout and retention declined in publication during the pandemic despite remaining one of the critical issues in higher education. The main concepts indicate that most of the dropout studies focus on comparisons between face-to-face and online modalities; the learning environment and engagement; and support and motivation.

Keywords:
Dropout, student retention, distance education, open education, online education

Introduction
Student attrition has been a major area of concern in higher education (HE) throughout the years and various theoretical models have explored the dimensions and factors that result in student retention or attrition (Rovai, 2003; Spady, 1970; Tinto, 1975). However, open and distance education, and subsequently online education, made the dropout phenomenon a central point because student-attrition rates have long been reported as higher than in face-to-face, on-campus education (Simpson, 2013; Smith, 2010). According to the traditional definition, dropout is the failure of an individual to continue their engagement in the registered programme and disengagement from the academic institution as a result of socially and/or academically unfavourable experiences (Tinto, 1975).

As the advancement of technology-driven education and the demand for HE continue to grow (Arnhold & Bassett, 2021), the context is subject to change, too. The long-established binary position of brick-and-mortar institutions versus distance education is becoming blurred as technology-driven education continues to advance. The rapidly increasing speed of technological development has made digital education, either online or technology-enhanced, a preferred option in higher education (Allen & Seeman, 2017). Online or distance learning has become a viable option for HE institutions looking for prevalence in more and wider student retention (Nakamura, 2017). The Covid-19 pandemic has further propelled this trend and proved that teaching through online modalities is a resilient option amidst the crisis. Although there are numerous publications that report the difficulties and limits of online education, digital learning and new developments around generative AI technologies continue to reshape education.

Within this transformation, it is important to sustain the quality of higher education and improve student retention, which is one of the quality criteria of HE. However, it is questionable if the old definitions of retention and dropout are still valid in the current technological landscape. Facilitating up-to-date and valid terminology for the field of ODDE is highly important to explore current practices. To achieve this goal, it is crucial to attain an in-depth understanding of the dropout phenomenon first and map the phenomenon with evidence from the literature. This study serves as a preliminary analysis of a literature review corpus from 2010 onwards. The year 2010 was taken as a starting point for the spread of online education to wider communities through various advancements such as massive open, online courses (MOOCs) and open education resources (OER; Zawacki-Richter & Naidu, 2016).
Within the given frame, this study aims to map the dropout terrain according to the thematic clusters to answer the following questions:

- How are publications on dropout between 2010 and 2022 distributed according to years and journals?
- What is the thematic scope in the field of drop-out research and how do the themes relate to each other?

This study focuses on the reasons, factors, or results that affect dropout or student retention among undergraduate or master's level students. The exploration and mapping of how the dropout phenomenon is addressed in the literature is formed according to the PICO framework (Boland et al., 2017). The dropout context is framed with HE including any ODDE setting. MOOCs are excluded as long as they are not offered as a part of formal education.

Methodology

Method and Sample

Analyses of trends and conceptual themes are important to picture the progress of a certain phenomenon and make evidence-supported projections about the field of the studied subject (Lee et al., 2004). A good example of this is a content analysis by Zawacki-Richter and Naidu (2016), in which they map three decades of distance-education literature. Exploring the thematic clusters based on large body of texts enables a broad understanding of trends and themes within the field under investigation (Krippendorff, 2013). A crucial point of the method is the need for well-informed interpretation of the data. With this aim, this paper adopts content analysis of systematically selected literature on student dropout and retention between January 2010 and December 2022. Content analysis is a labour-intensive process as it requires the validation of codebooks, glossaries, and inter-rater reliability as these factors increase the risk of bias of the authors. Therefore, an automated content analysis tool, Leximancer, is used as the text mining tool. This text-mining approach has been used to decrease interpreter coding bias, generate automated semantic relations and outline the most-discussed themes in connection to the phenomenon under study.

As the reproducibility and replicability of such methods are crucial in research (Smith & Humphreys, 2006, p. 262) and as the articles assessed here are not limited to a single journal, the methods of a systematic review were adopted to form the data set. Systematic reviews have three main, consecutive steps, namely: a) identification of the major research publications, b) critical appraisal of the obtained studies, c) synthesis of findings in the final corpus based on the evidence (Gough et al., 2012). Such analyses can enrich the literature by providing information on the selected study field and guidance for future studies and directions (Petticrew & Roberts, 2008).

Data were retrieved from three databases of the educational technology field, namely Education Source, Scopus, and Web of Science, using the search string given in Table 1 and the inclusion and exclusion criteria given in Table 2.

Table 1

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Search terms with Boolean operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dropout</td>
<td>(dropout OR drop-out OR retention OR persistence OR attrition OR disengagement* ) N5 student*</td>
</tr>
<tr>
<td>Digital context</td>
<td>AND (online* OR distance* OR blend* OR mobile OR technology-enhance*) N3 (learn* OR teach* OR study* OR study* OR degree)</td>
</tr>
<tr>
<td>Higher education</td>
<td>AND &quot;higher education&quot; OR universit* OR college* OR &quot;postsecondary education&quot; OR &quot;tertiary education&quot; OR undergard* OR postgrad*</td>
</tr>
</tbody>
</table>
Table 2
Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Not in English</td>
</tr>
<tr>
<td>Journal articles</td>
<td>Editorials, books, book chapters, conference proceedings, introductions, reviews, grey literature</td>
</tr>
<tr>
<td>Empirical research</td>
<td>Literature reviews, conceptional papers</td>
</tr>
<tr>
<td>Digital higher education, distance, open, online, hybrid, or blended learning</td>
<td>face to face only, MOOCs that are not a part of a formal university programme</td>
</tr>
<tr>
<td>Undergraduate students, master’s students</td>
<td>K-12, PhD students, adult learners in lifelong learning programmes</td>
</tr>
<tr>
<td>Papers report on student retention, attrition, persistence, dropout</td>
<td>Papers report a sub-dimension of retention or dropout phenomenon without direct link to dropout or retention</td>
</tr>
</tbody>
</table>

The initial corpus after elimination of duplicates \((N = 2426)\) was screened based on the abstracts. Before the abstract screening, inter-rater reliability between three raters was calculated \((k = .70)\), resulting in good agreement. Of these abstracts screened, 669 articles were selected for full-text screening. Of these, the studies that were not retrieved for full texts \((n = 33)\) were removed. After the final exclusion and inclusion assessment and further exclusions \((n = 352)\), the abstracts of 287 studies were found eligible for this study and included for further analysis.

The titles and abstracts of the 287 included articles were used for text mining as they present the purpose, population, and main results of the studies in detail and in condensed form. As the data was extracted automatically from the original CSV file of the included studies, they were checked for and cleaned by removing stopwords as well as words, lexicons and/or expressions irrelevant to the analysis (e.g., automatic indication of copyright).

**Limitations**

This study is limited to journal articles, three databases, and English only publications. Although the selection of data was subject to wide inclusion of three big databases that include high ranking journals, the publication of journal articles is limited by the editorial processes and the scope of the journals. In addition, conference papers and book chapters were not included. MOOCs, unless a part of a formal degree programmes are also excluded from this study as the reasons for dropping out from MOOCs are diverse and may not show the same pattern of disengagement from institutional, formal-learning environments. Although Leximancer has been found to produce valid evidence for content analysis (e.g., Zawacki-Richter & Naidu, 2016), it "...is not a panacea and requires analytical sensitivity and judgment in its interpretation." (Harwood et al., 2015, p. 1041).

**Findings and Discussion**

**Descriptive results**

Figure 1 shows the trend of publications from 2010 to 2022:
The publication trend per year reveals interesting results regarding its slightly fluctuating but generally increasing line until 2020. From the year 2020 on, there is a notable decline, which may be due to several reasons. The publication interest during the pandemic (2020-2022) may have shifted to more acute problems due to emergency remote teaching practices such as equality, inclusive education, and other constraints being experienced.

The included articles came from 133 different journals, with those contributing more articles from the field of ODDE. Online Journal of Distance Learning Administration ranks first (n = 12), followed by Online Learning (n = 10), Journal of Educators Online (n = 9), International Review of Research in Open & Distance Learning (n = 8), and Distance Education (n = 8). The list of the top ten journals is presented in Table 3.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online Journal of Distance Learning Administration</td>
<td>12</td>
</tr>
<tr>
<td>Online Learning</td>
<td>10</td>
</tr>
<tr>
<td>Journal of Educators Online</td>
<td>9</td>
</tr>
<tr>
<td>International Review of Research in Open &amp; Distance Learning</td>
<td>8</td>
</tr>
<tr>
<td>Distance Education</td>
<td>8</td>
</tr>
<tr>
<td>Computers &amp; Education</td>
<td>7</td>
</tr>
<tr>
<td>Open Learning</td>
<td>7</td>
</tr>
<tr>
<td>American Journal of Distance Education</td>
<td>6</td>
</tr>
<tr>
<td>Internet &amp; Higher Education</td>
<td>6</td>
</tr>
<tr>
<td>Journal of College Student Retention: Research, Theory &amp; Practice</td>
<td>6</td>
</tr>
</tbody>
</table>

The majority of articles on drop-out and retention was published in journals with a management and institution focus. Based on the 3M-Framework, which groups research themes in three layers of micro, meso, and macro levels (Zawacki-Richter, 2009, Zawacki-Richter & Bozkurt, 2023), the dropout issue is considered on the meso-level of HE institutions. Drop-out and student retention rates are widely considered connected to institutional quality assurance and management strategies, which is also a meso-level issue.

**The scope of dropout phenomenon (2010 - 2022)**

The first analysis was run for the whole corpus including all 287 papers based on titles and abstracts. Figure 2 displays the major themes and concepts covered in the articles published between 2010 and 2022.
The concept map shows five thematic regions that are named after the concept with the highest number of mentions in the corpus, namely: students (1238), online (968), learning (829), support (267), and motivation (125).

The thematic region of students covers three well-known facts in ODDE literature in general: 1) dropout rates are a major concern in distance education (see concept path distance – education – higher – attrition; Radovan, 2019; Xavier & Meneses, 2020); 2) as well as in online learning (completion – course – success – online); 3) often compared with face-to-face learning (Ashby et al., 2011; Ferguson, 2020). Many studies have compared and contrasted face-to-face education with online education (Clayton et al., 2010; Summers et al., 2005), despite the fact that the design quality of a course matters more than the mode of delivery (Matcha et al., 2020). The theme of online clearly indicates the link between success and course completion. This is a common approach to predicting dropout in online learning. When considering the development of higher education, the focus has tended to shift more towards graduation and course-completion rates as indicators of success, rather than the competencies that students achieve (Aljohani, 2016; Bodin & Orange, 2018).

The central thematic region of learning is the sole circle connecting all other four themes. Many factors of dropout based on Rovai’s (2013) internal factors or Tinto’s (1975) academic factors are highly dominant in this circle. For example, the relation of learning experience to academic strategies and engagement, and thus to satisfaction, is one highlight among the semantic connections (see concept path satisfaction – engagement – academic – experience). These are all linked to retention in online degrees. Whilst many studies discuss the vital mediator role of satisfaction and engagement for persistence in the corpus (e.g., Pechenkina, et al., 2017; Rahim, 2020), we must note the hazard of evaluating satisfaction based on course evaluations with superficial understanding of quality of online learning.

The concept map nicely underscores that student support is critical in ODDE (Brindley & Paul, 1996) with support as a thematic region of its own. The direct connection between institutional, support, and students indicates that researchers are dealing with institutional support systems in the publications. The learning – environment is connected to information systems through teaching in the support theme. Information systems include any
medium such as learning management systems or dashboards that provide information about teaching and learning activities in the online environment. The corpus covers several examples on early warning systems (e.g., Bañeres et al., 2020; Kustitskaya, et al., 2022), the use of dashboards (e.g., Marshall, 2016); and predictive learning analytics (e.g., Herodotou, et al., 2020; Yasmin, 2013).

Finally, the concept of motivation is related to persistence in learning. Motivation may be due to internal or external factors of the individual learners (Price & Kadi-Hanifi, 2011). Thus, it is not directly linked to institutional measures of support in the concept map. Still, motivational intervention programs are in use through predictive learning analytics (e.g., Herodotou, et al., 2020). More importantly, motivation demonstrates the social interaction in the circle, which is highly motivational (Waite & Davis, 2006). Technology-enhanced media and digital learning tools have long enabled rich social-learning environments (Barlow, 2008; Price & Kadi-Hanifi, 2011).

Conclusion and future directions

This study analyzed the content of titles and abstracts of systematically collected journal articles with a text-mining tool and provided an overview of the themes, concepts, and semantic relations among the concepts.

The findings demonstrate that the dropout and student-retention literature discusses the high attrition rates in distance education, the common understanding of success as course completion, the need for support at the institutional level (meso) through information gathered at the teaching and learning (micro) level. The map also displays engagement, satisfaction, and academic strategies as topics related to retention.

This study contributes to ODDE literature by mapping the last thirteen years of publications gathered through three large databases that host high-quality, peer-reviewed journals and filtering the results to retention and dropout studies with replicable, reproducible data curation methods. However, there is still need for further research.

Student retention is considered a quality indicator (Burke, 2019) and being financially better for the institutions and government (Simpson, 2005). In addition to global crisis, new advances in the digital world are forcing HE to transform. These issues include, generative AI, advanced access to the technology in the global North, a growing digital divide between the Global South and North, and the need for effective teaching and learning strategies through hybrid learning environments. Further analysis and comparison studies should be conducted to see whether the trends in dropout research differ or vary based on a) period of time of the publications, b) digital divide-related research, and c) digital transformation-based research. Thus, a new understanding and problematization of dropout phenomenon should be developed and discussed.

References


