

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Virtual Choreographies of Learning: AI-Driven Discovery in K–12 Online American Schools

Eduardo Salé Areias

WORKING VERSION



Mestrado em Engenharia Informática e Computação

Supervisor: Prof. Fernando Cassola Marques

Co-Supervisor: Prof. Dennis Beck

December 1, 2025

Resumo

As *Connections Academy*, apoiadas pela Pearson Online and Blended Learning, oferecem escolas totalmente online do Jardim-de-Infância ao 12.º ano em 32 estados dos EUA, servindo mais de 100 000 alunos, muitos dos quais transitaram do ensino presencial devido a problemas anteriores. Relatórios nacionais indicam piores resultados médios em exames padronizados para alunos de escolas online a tempo inteiro, embora estas escolas matriculem mais estudantes em risco, dificultando comparações diretas. Apesar da escala, existe pouca investigação sobre o que os alunos efectivamente fazem nestes ambientes e como os seus padrões de estudo e interação com docentes/colegas se relacionam com o progresso. Esta dissertação analisa dados das 32 escolas usando o enquadramento de *coreografias virtuais* do INESC TEC para identificar sequências recorrentes de ações de aprendizagem online e a sua ligação ao sucesso. Para tal, explora técnicas de IA como *k*-means, t-SNE/UMAP e modelos supervisionados (árvores de decisão, redes neuronais leves, *random forests*) para descobrir *clusters* e trajetórias comportamentais menos óbvias e enriquecer a compreensão da dinâmica de aprendizagem em escolas K–12 totalmente online.

Abstract

Connections Academy schools, supported by Pearson Online and Blended Learning, provide fully online K–12 education across 32 U.S. states and currently serve over 100,000 students. Many families choose these schools because their children faced specific problems in prior in-person settings. National reports indicate that students enrolled in full-time online schools often perform worse on standardized exams than peers in traditional public schools; however, fully online schools enroll a higher proportion of at-risk students, which complicates fair comparisons.

Despite the scale of this model, there is limited research on what actually happens in these learning environments: Which patterns of study do students follow? How do they engage with teachers and peers? And how are these patterns connected to learning progress? This dissertation addresses these questions by analyzing data from the 32 Connections Academy schools using the *virtual choreographies* framework developed at INESC TEC. The goal is to identify recurring sequences of online learning actions and examine their relevance for student success.

To that end, the work explores Artificial Intelligence techniques to detect and characterize behavior patterns and trajectories, including unsupervised methods such as k -means and dimensionality-reduction techniques (t-SNE/UMAP), alongside supervised models such as decision trees, lightweight neural networks, and random forests. These approaches are expected to surface non-obvious configurations of behavior and to enrich our understanding of student learning dynamics in fully online K–12 schools.

Chapter 1

Literature Review

1.1 Overview

This chapter describes the methodology used to identify and select the scientific literature relevant to this dissertation. It explains the type of literature review adopted, the search strategy, the inclusion and exclusion criteria, and the process followed to screen and select the final set of papers that compose the document corpus. In line with the thesis manifest, the review focuses on research about K–12 online learning and full-time virtual schools in the United States, with particular emphasis on their effectiveness, challenges and implications for students, teachers and school systems. This corpus provides the conceptual and empirical background needed to frame the later analysis of virtual choreographies of learning in Connections Academy schools.

1.2 Type of review

Given the objectives of this dissertation, a structured literature review was conducted, following principles of a systematic literature review while allowing some flexibility that is closer to a narrative synthesis. The goal of the review is to identify, select and document existing research on K–12 online learning and virtual schools, so as to support the definition of the dissertation research questions and to provide a solid basis for the later State of the Art chapter and for the empirical analysis of student activity data.

The corpus prioritizes contemporary systematic reviews and empirical studies that reflect the current state of K–12 online learning technologies. Instead of early foundational work, the review relies on recent comprehensive analyses, such as the systematic reviews by Martin et al. [9] and Johnson et al. [8]. It also incorporates critical studies on student outcomes, engagement and school-level implementation of K–12 online programmes, specifically those addressing at-risk populations as explored by Beck [3, 4] and Toppin and Toppin [11].

1.3 Search strategy

The literature search was carried out in 2025, in the months leading up to this deliverable. The primary search engine used was Google Scholar, which offers broad coverage of peer-reviewed publications in education and related fields. In addition, two discovery tools: Litmaps and ResearchRabbit were used to expand the initial set of articles through citation-based exploration and visualisation of related work.

The search targeted peer-reviewed journal articles, conference papers and book chapters written in English. Given the rapid evolution of educational technology and the impact of the COVID-19 pandemic, the search prioritized studies published between **2015 and 2025**. This timeframe ensures that the analyzed studies reflect the current technological infrastructure (modern LMS and learning analytics capabilities) and the post-pandemic reality of K-12 online education.

The main search string used in Google Scholar was iteratively refined, but was based on combinations of the following keywords:

- **Core topic terms:** “K-12 online learning”, “virtual schools”, “virtual schooling”, “online teaching”, “online school”, “distance education”.
- **Outcome / focus terms:** “effectiveness”, “student achievement”, “engagement”, “school choice”, “best practices”, “teacher training”, “at-risk students”.
- **Context terms:** “K-12”, “secondary education”, “United States”, “high school”.

An example of a typical query used in Google Scholar is:

“K-12 online learning” OR “virtual schools” OR “virtual schooling” AND (effectiveness OR achievement OR engagement OR “distance education”)

The initial Google Scholar searches generated a broad set of potentially relevant results. These references were exported and then imported into Litmaps and ResearchRabbit. From this initial seed set, both tools were used to:

- identify additional papers that frequently cite or are cited by the seed articles;
- visualise clusters of research focusing on K–12 virtual schools and online programmes;
- surface key works such as large-scale reviews, empirical studies on student outcomes, and research on teacher practices in virtual schools.

Through this process, an initial pool of approximately 150 records was assembled before applying the inclusion and exclusion criteria described below.

1.4 Inclusion and exclusion criteria

To ensure the relevance and quality of the selected literature, explicit inclusion and exclusion criteria were defined and applied during the screening process.

1.4.1 Inclusion criteria

A study was included in the corpus if it met all of the following criteria:

- **IC1** – The study focuses on K–12 online learning, virtual schools or virtual schooling (fully online or primarily online programmes).
- **IC2** – The study is peer-reviewed (journal article, conference paper or book chapter).
- **IC3** – The study is written in English.
- **IC4** – The study was published between **2015 and 2025**.
- **IC5** – The study addresses at least one of the following aspects: effectiveness or outcomes of virtual schooling (e.g., student achievement, completion, equity), teaching and instructional practices in online K–12 settings, student engagement and experiences, or system-level issues such as school choice, policy and implementation of virtual schools.

1.4.2 Exclusion criteria

Studies were excluded if any of the following applied:

- **EC1** – The main focus is higher education, adult education or corporate training, rather than K–12 schooling.
- **EC2** – The publication type is a thesis, dissertation, report, blog post, poster, editorial or other non-peer-reviewed document.
- **EC3** – The study is not available in full text.
- **EC4** – The study deals with general educational technology or blended learning without a clear focus on fully or primarily online K–12 programmes.
- **EC5** – The paper is purely theoretical or opinion-based without sufficient connection to K–12 online or virtual schooling.

1.5 Screening and selection process

The selection process followed three main stages: (i) title and abstract screening, (ii) full-text assessment, and (iii) citation-based expansion and refinement.

1.5.1 Stage 1: Title and abstract screening

After removing duplicates from the combined Google Scholar, Litmaps and ResearchRabbit exports, approximately 150 unique records remained. Titles and abstracts were screened against the inclusion and exclusion criteria. At this stage, studies that clearly focused on higher education, general ICT in education or non-online contexts were removed. This first screening excluded around 90 records, leaving approximately 60 papers for full-text assessment.

1.5.2 Stage 2: Full-text assessment

The full text of the remaining studies was examined in more detail. Papers that mentioned online learning only superficially, or that did not present results or discussion directly related to K–12 virtual schools, were excluded. This stage reduced the set to a core group of studies that provide substantial evidence on K–12 online learning. This includes recent systematic reviews [8, 9] and research on student experiences in online K–12 environments, with a particular focus on support for at-risk students as emphasized by Beck [3, 5]. After full-text assessment, around 20 papers remained.

1.5.3 Stage 3: Citation-based refinement

The core set of papers was then used for backward and forward snowballing. Their reference lists and citation networks were explored using Litmaps and ResearchRabbit. This step led to the identification of additional key works, including highly cited recent reviews that had not appeared in the initial Google Scholar rankings. After applying the same inclusion and exclusion criteria to these additional articles and removing new duplicates, the final corpus consisted of **12 highly relevant papers**. This set represents the state-of-the-art in the field and includes:

- Systematic reviews of K–12 online learning effectiveness and teaching practices, such as those by Martin et al. [9], Johnson et al. [8], and Barbour and Hodges [2];
- Empirical studies on student engagement, particularly regarding at-risk populations and support structures, including recent work by Beck [3–5] and Curtis and Werth [7];
- Analyses of policy, school choice and the growth of K–12 online learning, as discussed by Barbour [1], Molnar et al. [10], and Borup and Kennedy [6].

The detailed bibliographic information for these papers is provided in the Bibliography at the end of this document.

1.6 Data management

All selected papers were organised using a combination of reference management and note-taking tools. Bibliographic information for each article (title, authors, year, venue, DOI and URL) was stored in a BibTeX file associated with the Overleaf project, ensuring consistency between the reference list and the citations used in this dissertation. In addition, the PDFs and summaries of the papers were uploaded to a NotebookLM notebook dedicated to the dissertation. For each paper, short notes were created including:

- a brief summary of the research questions, methods and main findings;
- tags indicating the main focus (e.g., effectiveness/outcomes, teaching practices, student engagement, policy and school choice);

- information about the context (e.g., U.S. state or region, subject area, grade levels).

This structure transforms the selected articles into a manageable “stack of documents” that can be systematically analysed in the next stages of the project and will support the development of the State of the Art chapter.

1.7 Limitations and threats to validity

The review process is subject to several limitations. First, although Google Scholar provides broad coverage, it does not index all relevant education databases, so some studies may have been missed. Second, the search was limited to publications in English, which may exclude research on K–12 online learning conducted in other languages and contexts. Third, the focus on K–12 virtual schools in the United States means that results from other countries were not systematically considered, even when they appeared in the search results. Finally, despite the use of explicit inclusion and exclusion criteria, subjective judgement was involved when assessing the relevance of titles, abstracts and full texts. These limitations will be taken into account when interpreting the findings of the selected studies in the subsequent State of the Art chapter.

References

- [1] M. Barbour. “K–12 Online Learning and School Choice: Growth and Expansion in the Absence of Evidence”. In: *Handbook of Research on K-12 Online and Blended Learning*. Ed. by K. Kennedy and R. E. Ferdig. ETC Press, 2017, pp. 433–456. URL: <https://press.etc.cmu.edu/index.php/product/handbook-of-research-on-k-12-online-and-blended-learning/>.
- [2] Michael K. Barbour and C. B. Hodges. “Preparing Teachers for Effective K-12 Online Learning in the Age of Disruptions”. In: *Open Praxis* 16.4 (2024), pp. 583–594. DOI: [10.55982/openpraxis.16.4.678](https://doi.org/10.55982/openpraxis.16.4.678).
- [3] Dennis Beck. “At-risk and online: Parent perceptions of at-risk learner’s supports in a fully online school”. In: *Journal of Research on Technology in Education* (2024). (Ahead-of-print). DOI: [10.1080/15391523.2023.2285497](https://doi.org/10.1080/15391523.2023.2285497).
- [4] Dennis Beck. “Be Prepared: Online School Experience and Student Achievement During the Pandemic”. In: *Frontiers in Education* 8 (2023), p. 1161003. DOI: [10.3389/feduc.2023.1161003](https://doi.org/10.3389/feduc.2023.1161003).
- [5] Dennis Beck and J. Levine. “The role of the advocate in cyber schools during the COVID-19 pandemic”. In: *Journal of Online Learning Research* 6.3 (2020), pp. 195–217. URL: <https://www.learntechlib.org/primary/p/217172/>.
- [6] J. Borup and Kathryn Kennedy. “The Case for K-12 Online Learning”. In: *Handbook of Research on K-12 Online and Blended Learning*. Ed. by K. Kennedy and R. E. Ferdig. ETC Press, 2017, pp. 405–432. URL: <https://press.etc.cmu.edu/index.php/product/handbook-of-research-on-k-12-online-and-blended-learning/>.
- [7] Heidi Curtis and Loredana Werth. “Fostering Student Success and Engagement in a K-12 Online School”. In: *Journal of Online Learning Research* 1.2 (2015), pp. 163–190. URL: <https://www.learntechlib.org/primary/p/151187/>.
- [8] Carla C. Johnson et al. “Online Teaching in K-12 Education in the United States: A Systematic Review”. In: *Review of Educational Research* 93.3 (2022), pp. 353–411. DOI: [10.3102/00346543221105550](https://doi.org/10.3102/00346543221105550).
- [9] Florence Martin, T. Sun, and C. D. Westine. “A systematic review of research on K-12 online teaching and learning: Comparison of research from two decades 2000 to 2019”. In: *Journal of Research on Technology in Education* 52.3 (2020), pp. 353–374. DOI: [10.1080/15391523.2019.1605825](https://doi.org/10.1080/15391523.2019.1605825).
- [10] Alex Molnar et al. *Virtual Schools in the U.S. 2023*. Boulder, CO: National Education Policy Center, 2023. URL: <https://nepc.colorado.edu/publication/virtual-schools-annual-2023>.

- [11] Ian N. Toppin and Sheila M. Toppin. “Virtual schools: The changing landscape of K-12 education in the U.S.” In: *Education and Information Technologies* 21 (2016), pp. 1571–1589. DOI: [10.1007/s10639-015-9402-8](https://doi.org/10.1007/s10639-015-9402-8).