# Learning Analytics and Responsible Knowledge Discovery in Education

**Daniela Rotelli**

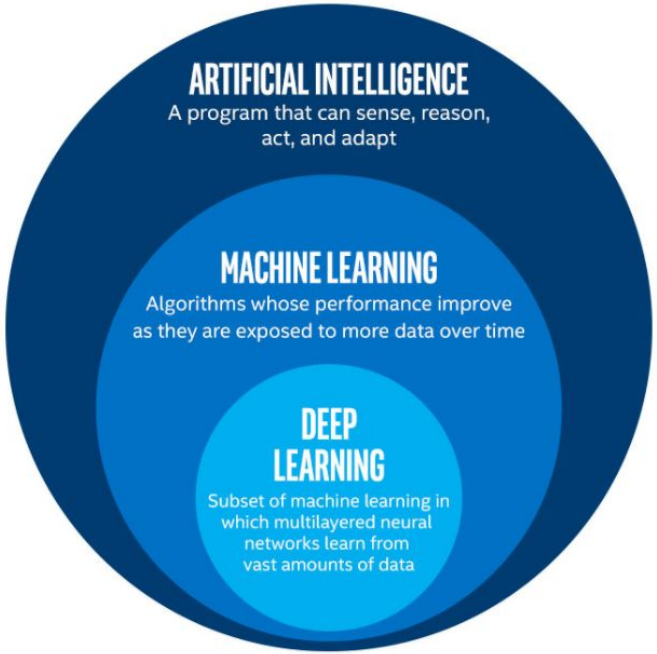

Department of Computer Science
University of Pisa
Pisa, Italy

# Daniela Rotelli

- Laurea in Lingue e Letterature Straniere (Università di Milano e Bergamo)
- Azienda
- Insegnamento
- Laurea in Informatica Umanistica (Università di Pisa)
- Dottorato di ricerca in Informatica (Università di Pisa)
- Assegnista di ricerca al Dipartimento di Informatica (Università di Pisa)

**daniela.rotelli@di.unipi.it**

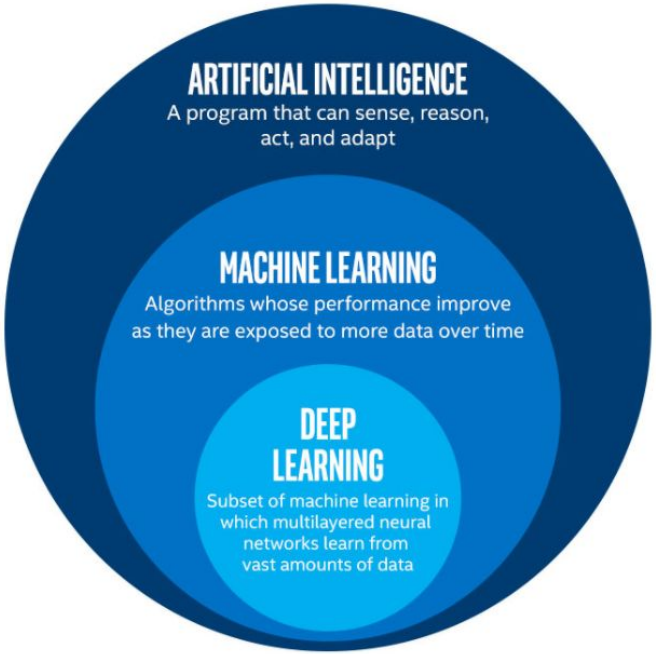
# AI, Machine Learning, Deep Learning, Data Science, Big Data

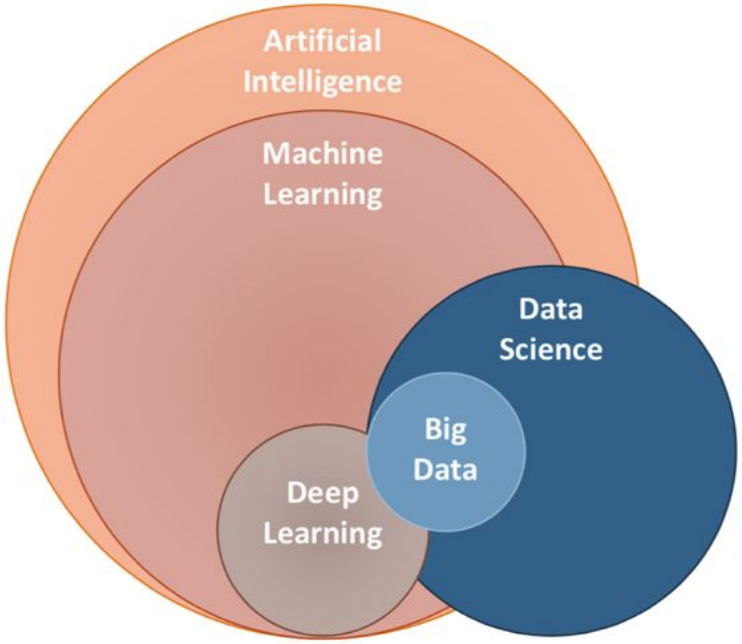



| | Acting | Thinking |
|---|---|---|
| **Humanly** | "The art of creating machines that perform functions that require intelligence when performed by people." —Kurzweil, 1990<br><br>"The study of how to make computers do things at which, at the moment, people are better." —Rich and Knight, 1991 | "The exciting new effort to make computers think [...] machines with minds, in the full and literal sense." —Haugeland, 1985<br><br>"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning [...]" —Bellman, 1978 |
| **Rationally** | "Computational Intelligence is the study of the design of intelligent agents." —Poole et al., 1998<br><br>"AI [...] is concerned with intelligent behavior in artifacts." —Nilsson, 1998 | "The study of mental faculties through the use of computational models." —Charniak and McDermott, 1985<br><br>"The study of the computations that make it possible to perceive, reason, and act." —Winston, 1992 |

Stuart J. Russell and Peter Norvig. Artificial Intelligence: A Modern Approach (2010)

# AI, Machine Learning, Deep Learning, Data Science, Big Data




# Applied Data Science


Industry

Healthcare


Education

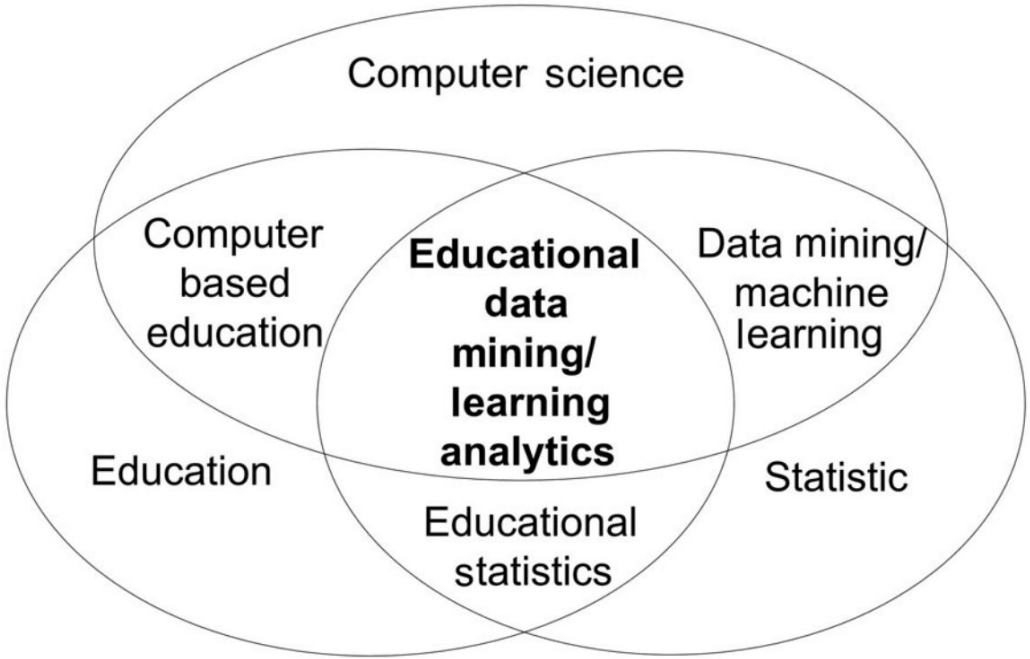
# Educational Data Science


Learning processes

Data-driven insights

Learning, teaching practices, educational systems

# Educational Data Science




Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

# Educational Data Mining

# Educational Data Mining

Educational Data Mining (EDM) is concerned with **developing methods for exploring the unique types of data** that come from educational environments (Bakhshinategh, Zaiane, ElAtia, & Ipperciel, 2018).

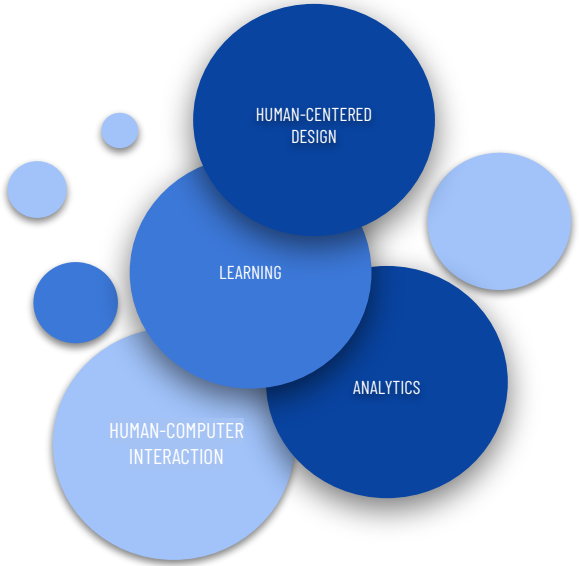
It can be also defined as the **application of data mining** (DM) techniques to this specific type of dataset that come from educational environments to address important educational questions (Romero & Ventura, 2013).

Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, *23*, 537-553.
Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, *3*(1), 12-27.

# Learning Analytics

# Learning Analytics

"The measurement, **collection**, **analysis** and reporting of data about learners and their contexts, for purposes of **understanding** and optimising learning and the **environments** in which it occurs" (Siemens, 2011)

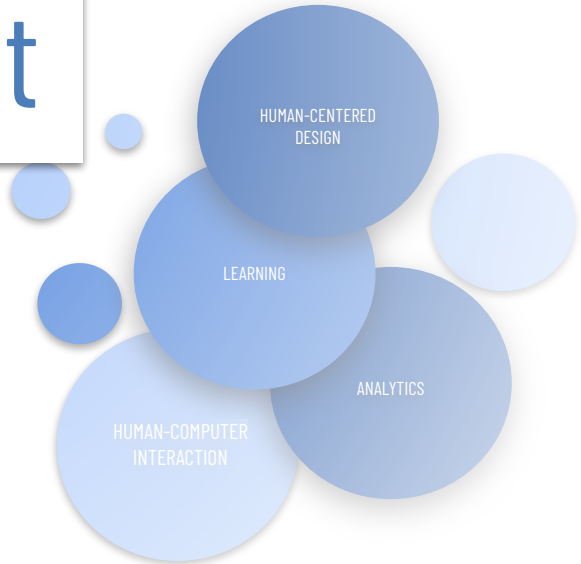




Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE* review, 46(5), 30.

# Learning Analytics

"The measurement, **collection**, **analysis** and reporting of data about learners and their contexts, for purposes of **understanding** and optimising learning and the **environments** in which it occurs" (Siemens, 2011)


Environment

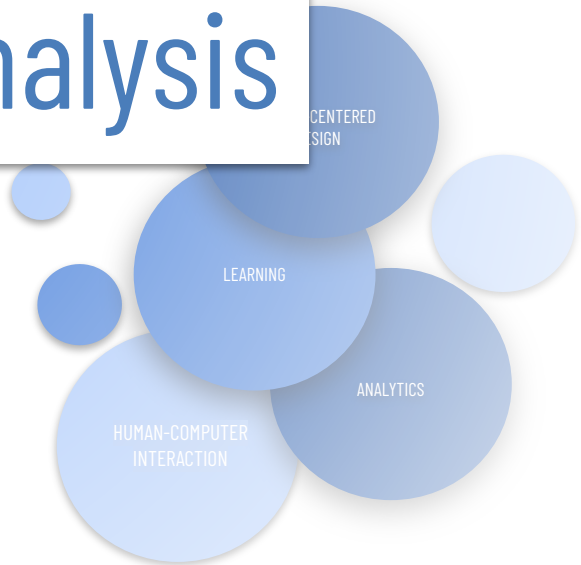



Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE* review, 46(5), 30.

# Learning Analytics

"The measurement, **collection**, **analysis** and reporting of data about learners and their contexts, for purposes of **understanding** and optimising learning and the **environments** in which it occurs" (Siemens, 2011)

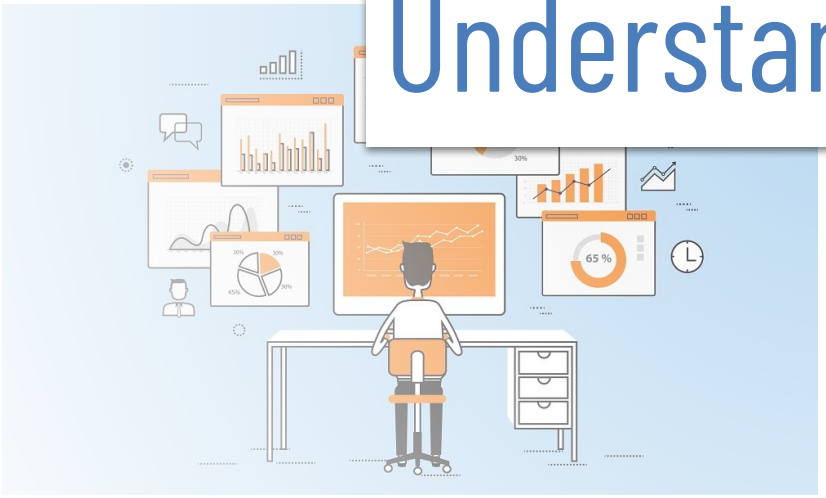
## Collection and Analysis

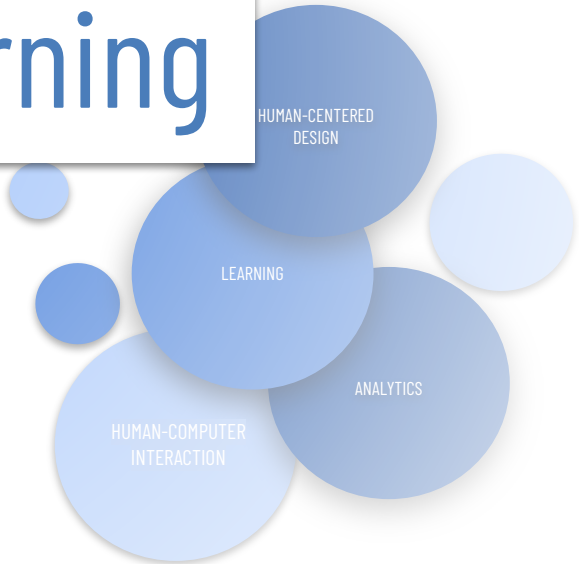

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE* review, 46(5), 30.

# Learning Analytics

"The measurement, **collection**, **analysis** and reporting of data about learners and their contexts, for purposes of **understanding** and optimising learning and the **environments** in which it occurs" (Siemens, 2011)



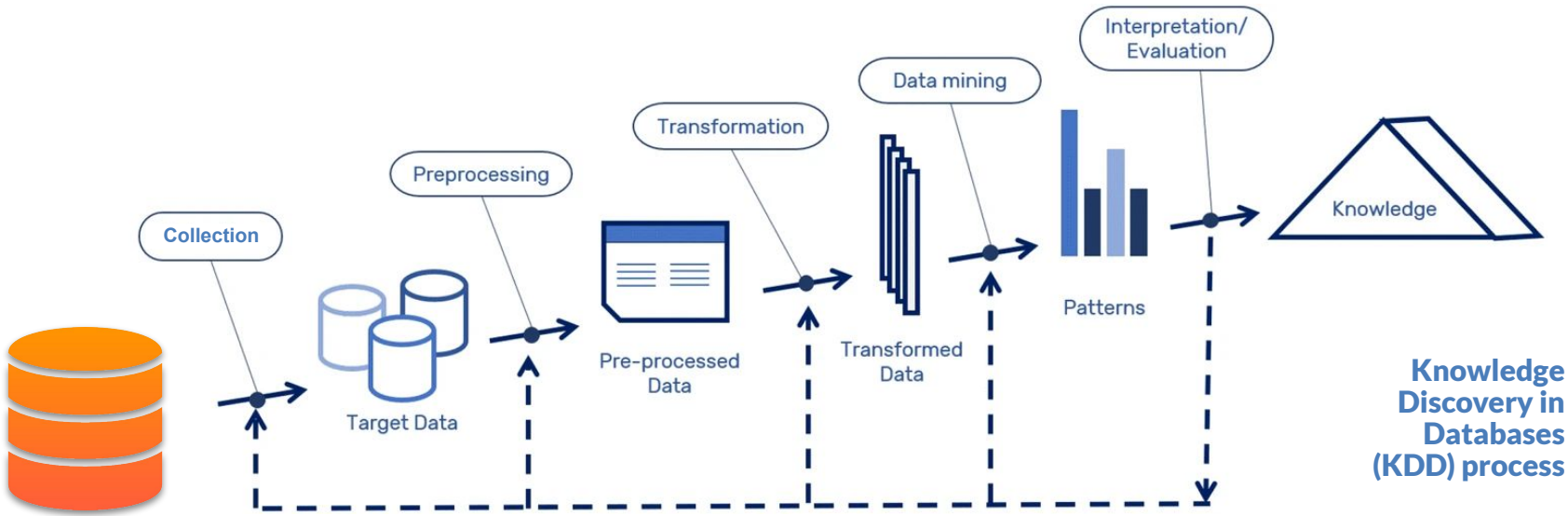

## Understand learning




Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE* review, 46(5), 30.

# Learning Analytics

"The measurement, **collection**, **analysis** and reporting of data about learners and their contexts, for purposes of **understanding** and optimising learning and the **environments** in which it occurs" (Siemens, 2011)

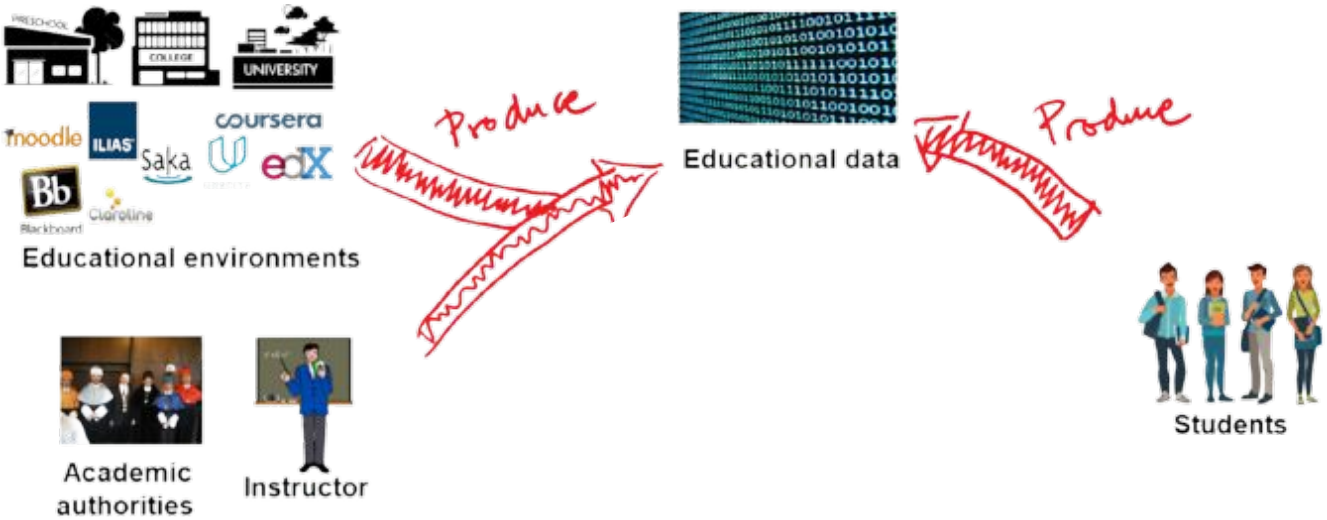



**Knowledge Discovery in Databases (KDD) process**

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE* review, 46(5), 30.

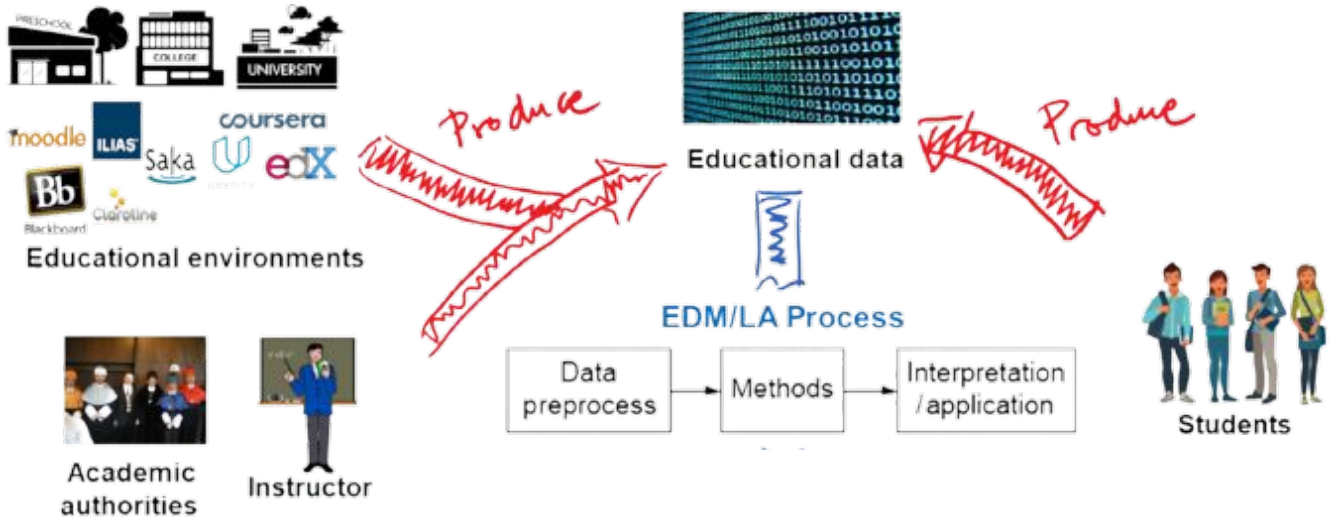
# Educational Knowledge Discovery

# Educational Knowledge Discovery Process



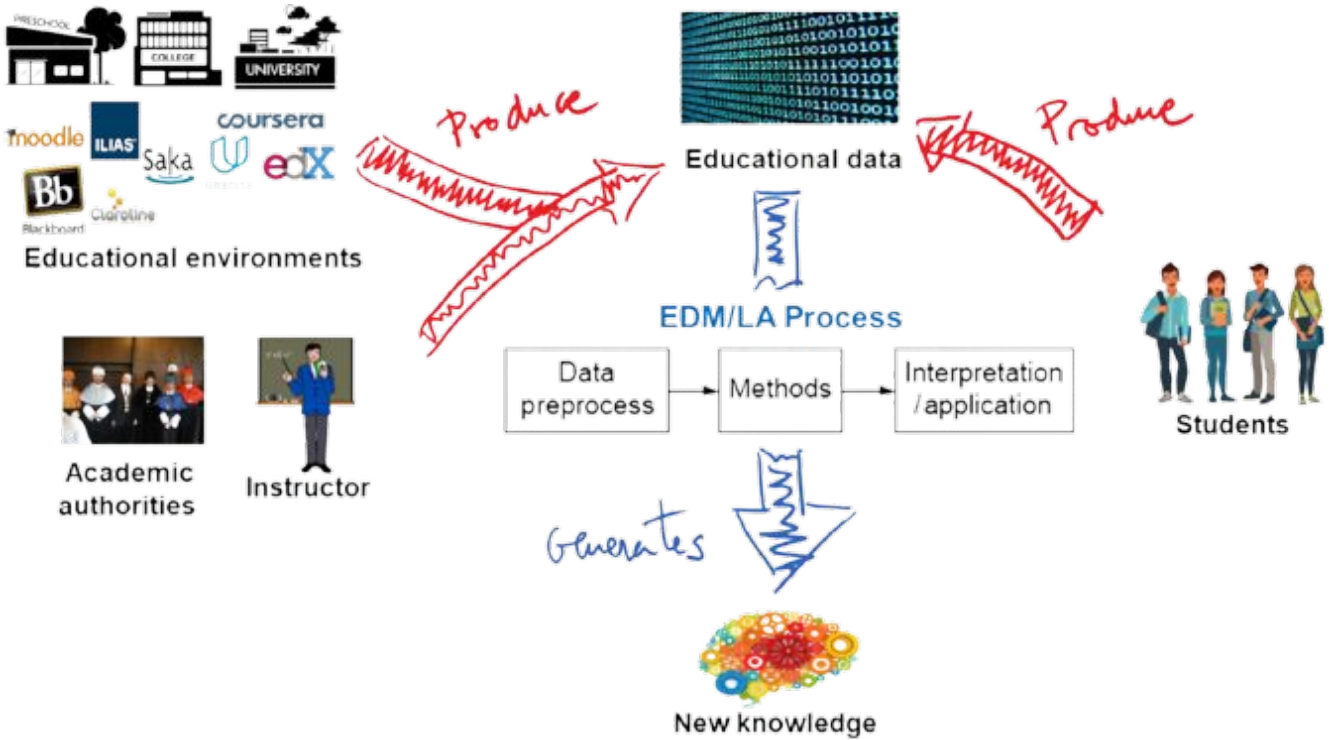

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

# Educational Knowledge Discovery Process



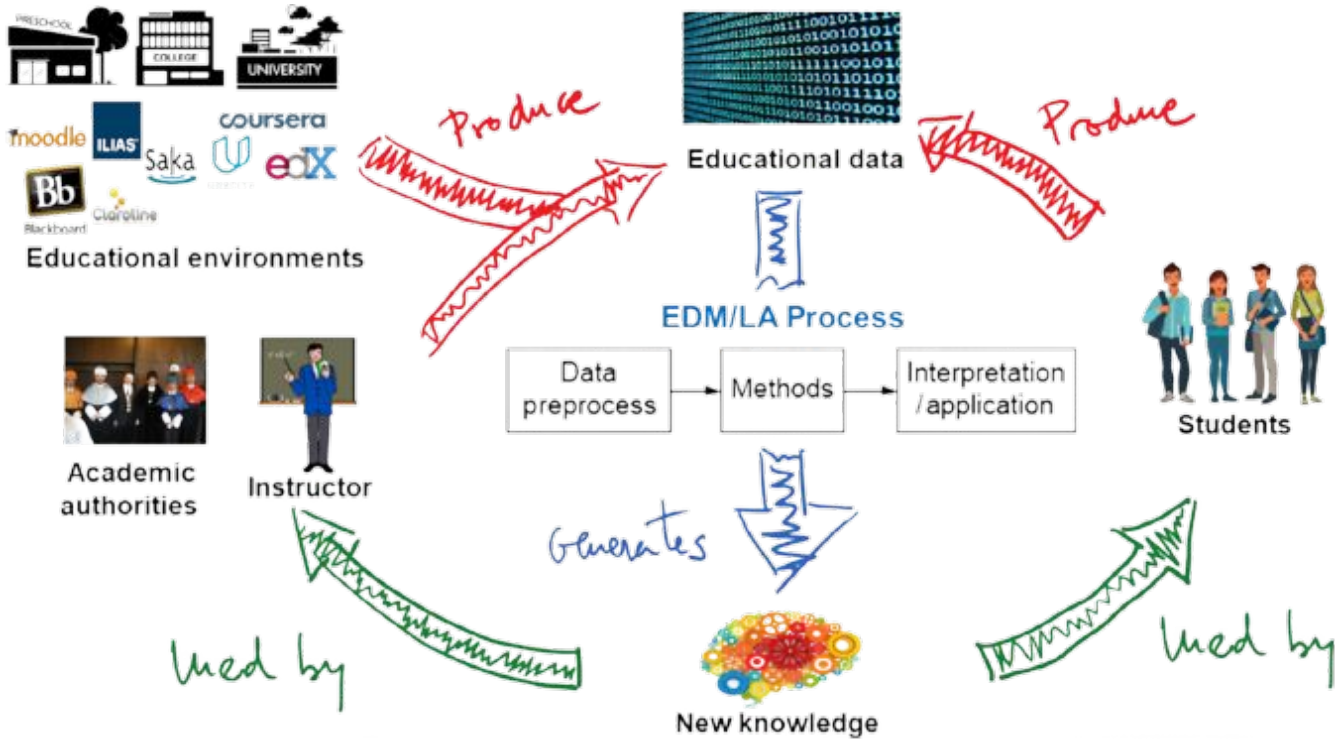

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

# Educational Knowledge Discovery Process




Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

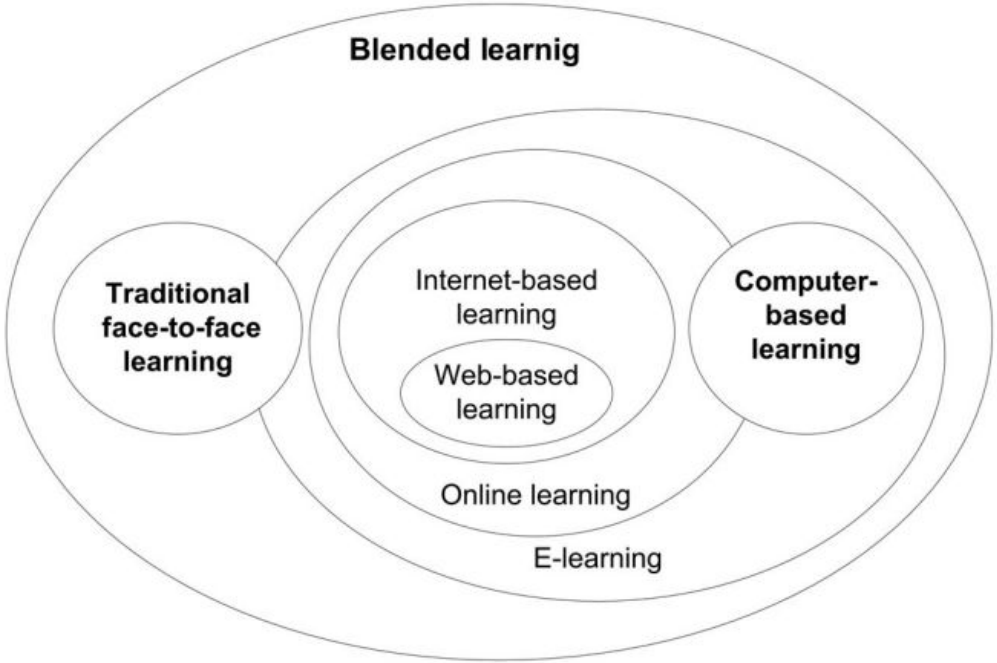
# Educational Knowledge Discovery Process




Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

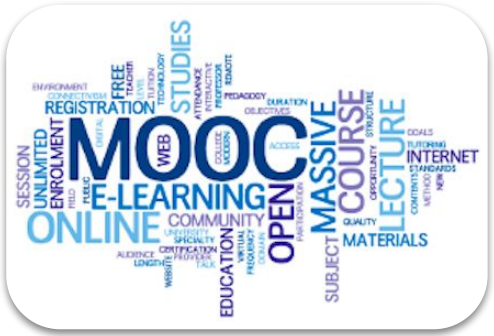
# Environment

# Types of educational environments




Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

# Online Learning



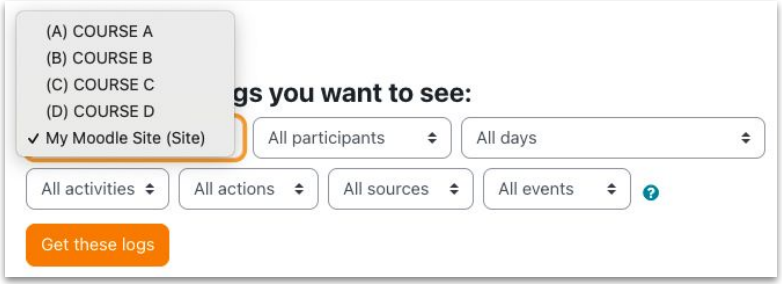



| Sites | Courses | Users | Enrolments |
|---|---|---|---|
| 168,239 | 46,433,872 | 403,776,033 | 2,346,375,317 |
| **Forum posts** | **Resources** | **Quiz questions** | **Countries** |
| 785,968,811 | 409,077,564 | 8,298,312,407 | 241 |

Source: *stats.moodle.org* - December 2023

# Data Collection

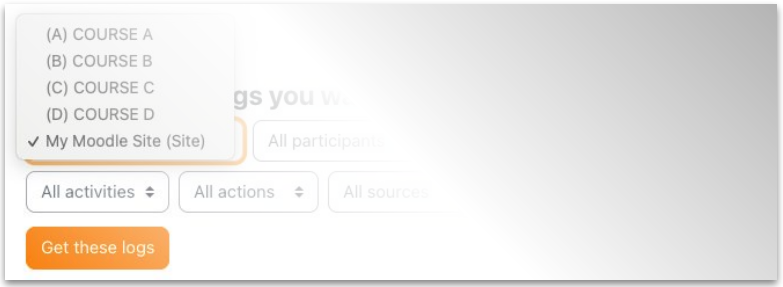
Database




Log Generation Interface


Experience API

Database




Log Generation Interface


Experience API

# Moodle database table: *mdl_logstore_standard_log*

| id | eventname | component | action | target | objecttable | objectid |
|---|---|---|---|---|---|---|
| 8470 | \mod_forum\event \subscription_created | mod_forum | created | subscription | forum_subscriptions | 2 |
| **crud** c | **edulevel** 2 | **contextid** 147 | **contextlevelid** 70 | **contextinstanceid** 66 | **userid** 2 | **courseid** 2 |
| **relateduserid** 2 | **anonymous** 0 | **other** a:1:{s:7:"forumid";s:1:"5";} | **timecreated** 1681205420 | **origin** web | **ip** 0:0:0:0:0:0:0:1 | **relauserid** NULL |

Who

What

Where

When

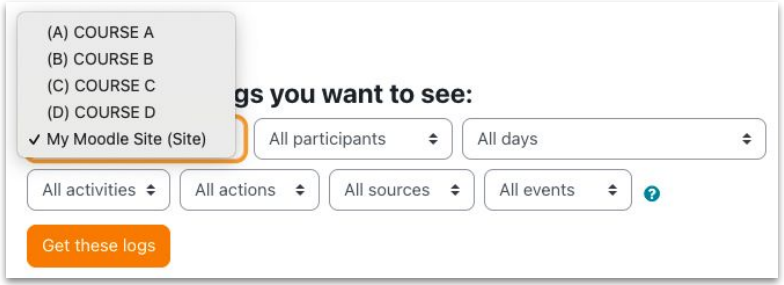
# Moodle database table: *mdl_logstore_standard_log*

| id | eventname | component | action | target | objecttable | objectid |
|---|---|---|---|---|---|---|
| 8470 | \mod_forum\event \subscription_created | mod_forum | created | subscription | forum_subscriptions | 2 |
| **crud** c | **edulevel** 2 | **contextid** 147 | **contextlevelid** 70 | **contextinstanceid** 66 | **userid** 2 | **courseid** 2 |
| **relateduserid** 2 | **anonymous** 0 | **other** a:1:{s:7:"forumid";s:1:"5";} | **timecreated** 1681205420 | **origin** web | **ip** 0:0:0:0:0:0:0:1 | **relauserid** NULL |

BHT Berliner Hochschule für Technik

| Advantages | Drawbacks |
|---|---|
| A basic SQL query to access data. | **Unintelligible** to humans; need of **multiple table joins** to retrieve information (*user roles*, *deleted activities* and *users*) and be human readable; need of **in-depth knowledge** of Moodle database (**416 tables** in the standard configuration + plugins); **admin privileges**. |

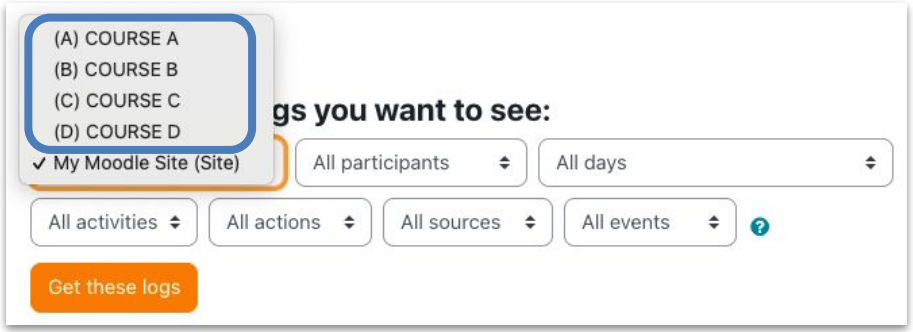
Database

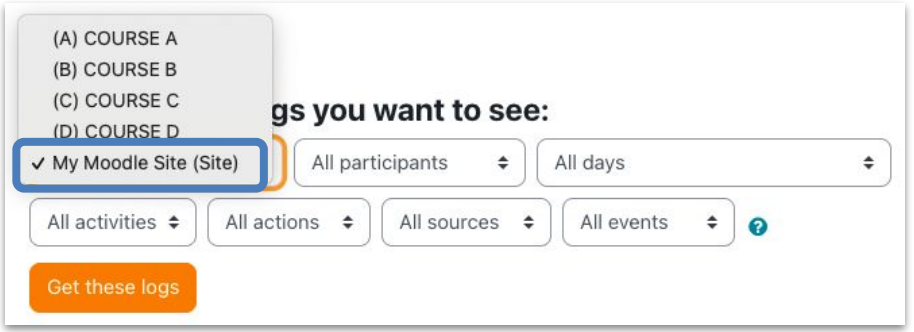



**Log Generation Interface**


Experience API

# *Course* logs VS *Site* logs




- Course Logs
  - a specific course (A, or B, or C,...)

- Site Logs
  - all courses (A, and B, and C, and D,...)
  - all actions performed on the platform

**Rotelli**, D., & Monreale, A. (2022, March). Time-on-task estimation by data-driven outlier detection based on learning activities. In *LAK22: 12th International Learning Analytics and Knowledge Conference*.
**Rotelli**, D., & Monreale, A. (2023, August). Processing and Understanding Moodle Log Data and their Temporal Dimension. *Journal of Learning Analytics*.

# Standard Moodle Logs

| Time | User full name | Affected user | Event context | Componen | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Course module viewed | The user with id '59' viewed the 'lesson' activity with course module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Question answered | The user with id '59' has answered the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:34 | User 59 | - | System | System | Group message sent | The user with id '59' sent a message with id '374' to the conversation with id '92'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | User 62 | User 59 | System | Message viewed | The user with id '59' read a message from the user with id '62'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | - | Lesson: SQL | Lesson | Question viewed | The user with id '59' has viewed the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |

*Example of log data generated in Moodle log data generation interface*
***for both site and course logs***

Who | What | Where | When

# Standard Moodle Logs

| Time | User full name | Affected user | Event context | Componen | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Course module viewed | The user with id '59' viewed the 'lesson' activity with course module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Question answered | The user with id '59' has answered the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:34 | User 59 | - | System | System | Group message sent | The user with id '59' sent a message with id '374' to the conversation with id '92'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | User 62 | User 59 | System | Message viewed | The user with id '59' read a message from the user with id '62'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | - | Lesson: SQL | Lesson | Question viewed | The user with id '59' has viewed the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |

Course ?

*Example of log data generated in Moodle log data generation interface*
***for both site and course logs***

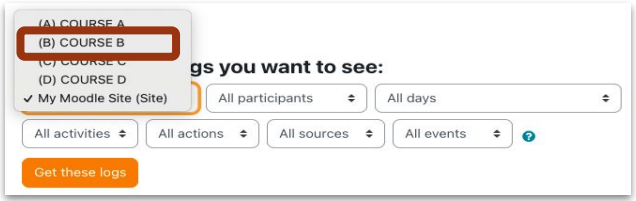
Who

What

Where

When

# Course Logs




5:33 to 6:05 = 32 minutes

| Time | User full name | Affected user | Event context | Component | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 5:33 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 5:33 PM | USER 59 | USER 59 | Attendance: OpenLab | Attendance | Session report viewed | User with id 59 viewed attendance sessions for student with id 59 | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | File: Intro 01 | File | Course module viewed | The user with id '59' viewed the 'resource' activity with course module id '245'. | web | 87.15.222.193 |

Yellow: course B logs

# Course Logs

| Time | User full name | Affected user | Event context | Component | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 5:33 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 5:33 PM | USER 59 | USER 59 | Attendance: OpenLab | Attendance | Session report viewed | User with id 59 viewed attendance sessions for student with id 59 | web | 87.15.222.193 |

???

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 6:05 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | File: Intro 01 | File | Course module viewed | The user with id '59' viewed the 'resource' activity with course module id '245'. | web | 87.15.222.193 |

Yellow: course B logs

# Course Logs

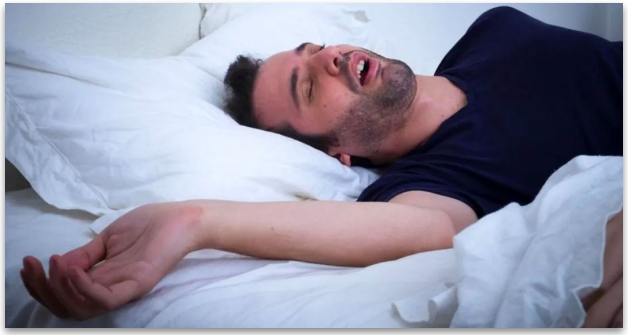
| Time | User full name | Affected user | Event context | Component | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 5:33 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 5:33 PM | USER 59 | USER 59 | Attendance: OpenLab | Attendance | Session report viewed | User with id 59 viewed attendance sessions for student with id 59 | web | 87.15.222.193 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 6:05 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | File: Intro 01 | File | Course module viewed | The user with id '59' viewed the 'resource' activity with course module id '245'. | web | 87.15.222.193 |

Yellow: course B logs

# Course Logs

| Time | User full name | Affected user | Event context | Component | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 5:33 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 5:33 PM | USER 59 | USER 59 | Attendance: OpenLab | Attendance | Session report viewed | User with id 59 viewed attendance sessions for student with id 59 | web | 87.15.222.193 |



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 6:05 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | File: Intro 01 | File | Course module viewed | The user with id '59' viewed the 'resource' activity with course module id '245'. | web | 87.15.222.193 |

Yellow: course B logs

# Course Logs

| Time | User full name | Affected user | Event context | Component | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 5:33 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 5:33 PM | USER 59 | USER 59 | Attendance: OpenLab | Attendance | Session report viewed | User with id 59 viewed attendance sessions for student with id 59 | web | 87.15.222.193 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 6:05 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | File: Intro 01 | File | Course module viewed | The user with id '59' viewed the 'resource' activity with course module id '245'. | web | 87.15.222.193 |

Yellow: course B logs

# Course Logs

| Time | User full name | Affected user | Event context | Component | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 5:33 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 5:33 PM | USER 59 | USER 59 | Attendance: OpenLab | Attendance | Session report viewed | User with id 59 viewed attendance sessions for student with id 59 | web | 87.15.222.193 |

???

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 6:05 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | File: Intro 01 | File | Course module viewed | The user with id '59' viewed the 'resource' activity with course module id '245'. | web | 87.15.222.193 |

Yellow: course B logs

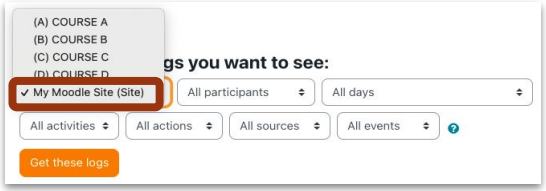
# *Site* Logs

| Time | User full name | Affected user | Event context | Component | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 5:33 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 5:33 PM | USER 59 | USER 59 | Attendance: OpenLab | Attendance | Session report viewed | User with id 59 viewed attendance sessions for student with id 59 | web | 87.15.222.193 |
| 15 January 2021, 5:35 PM | USER 59 | USER 59 | User: USER 59 | System | Dashboard viewed | The user with id '59' has viewed their dashboard | web | 87.15.222.193 |
| 15 January 2021, 5:42 PM | USER 59 | USER 59 | User: USER 59 | System | Dashboard viewed | The user with id '59' has viewed their dashboard | web | 87.15.222.193 |
| 15 January 2021, 6:02 PM | USER 59 | - | Course: A | System | Course viewed | The user with id '59' viewed the course with id '21'. | web | 87.15.222.193 |
| 15 January 2021, 6:02 PM | USER 59 | - | H5P: Answer the questions | H5P | Course module viewed | The user with id '59' viewed the 'h5pactivity' activity with course module id '36'. | web | 87.15.222.193 |
| 15 January 2021, 6:02 PM | USER 59 | - | H5P: Answer the questions | H5P Package | H5P content viewed | The user with id '59' has viewed the H5P with the id '28'. | web | 87.15.222.193 |
| 15 January 2021, 6:04 PM | USER 59 | - | H5P: Answer the questions | H5P | xAPI statement received | The user with the id '59' send a tracking statement for a H5P activity with the course module id '36'. | web | 87.15.222.193 |
| 15 January 2021, 6:04 PM | USER 59 | - | Course: A | System | Course viewed | The user with id '59' viewed the course with id '21'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | Category: Overall Site | System | Category viewed | The user with id '59' viewed the course category with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | File: Intro 01 | File | Course module viewed | The user with id '59' viewed the 'resource' activity with course module id '245'. | web | 87.15.222.193 |

**32** minutes

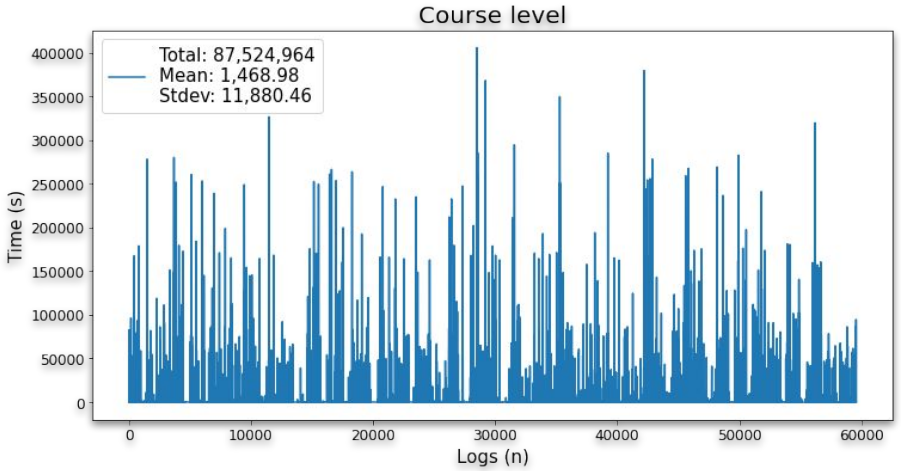
**2** minutes

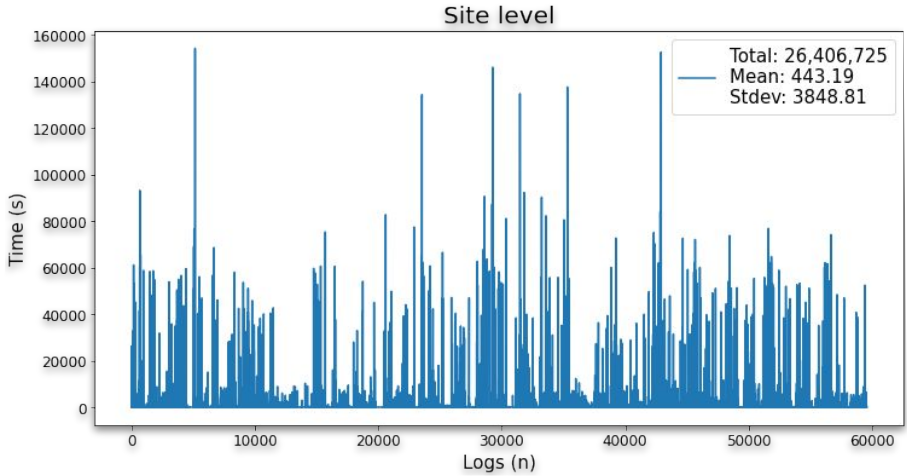
Yellow: **course B logs**; Red: **platform logs**; Green: **course A logs**



# Comparison of *time spent online*

$$( d_i = t_{i+1} - t_i )$$




Procedure:

- extraction of Course **B logs**
- duration calculation

Procedure:

- extraction of **all logs** in the platform
- duration calculation
- filter on Course **B logs**

**Rotelli**, D., & Monreale, A. (2023, August). Processing and Understanding Moodle Log Data and their Temporal Dimension. *Journal of Learning Analytics*.

# How can we understand the course?

| Time | User full name | Affected user | Event context | Component | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Course module viewed | The user with id '59' viewed the 'lesson' activity with course module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Question answered | The user with id '59' has answered the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:34 | User 59 | - | System | System | Group message sent | The user with id '59' sent a message with id '374' to the conversation with id '92'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | User 62 | User 59 | System | Message viewed | The user with id '59' read a message from the user with id '62'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | - | Lesson: SQL | Lesson | Question viewed | The user with id '59' has viewed the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |

Course ?

*Example of log data generated in Moodle log data generation interface*
***for both site and course logs***

# Database logs vs Log Generation Interface logs

| id | eventname | component | action | target | objecttable | objectid |
|---|---|---|---|---|---|---|
| 8470 | \mod_forum\event \subscription_created | mod_forum | created | subscription | forum_subscriptions | 2 |
| **crud** | **edulevel** | **contextid** | **contextlevelid** | **contextinstanceid** | **userid** | **courseid** |
| c | 2 | 147 | 70 | 66 | 2 | 2 |
| **relateduserid** | **anonymous** | **other** | **timecreated** | **origin** | **ip** | **relauserid** |
| 2 | 0 | a:1:{s:7:"forumid";s:1:"5";} | 1681205420 | web | 0:0:0:0:0:0:0:1 | NULL |

| Time | User full name | Affected user | Event context | Componen | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Course module viewed | The user with id '59' viewed the 'lesson' activity with course module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Question answered | The user with id '59' has answered the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:34 | User 59 | - | System | System | Group message sent | The user with id '59' sent a message with id '374' to the conversation with id '92'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | User 62 | User 59 | System | Message viewed | The user with id '59' read a message from the user with id '62'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | - | Lesson: SQL | Lesson | Question viewed | The user with id '59' has viewed the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |

Where

Course ?

Who

What

Where

When

# Database logs vs Log Generation Interface logs

| id | eventname | component | action | target | objecttable | objectid |
|---|---|---|---|---|---|---|
| 8470 | \mod_forum\event \subscription_created | mod_forum | created | subscription | forum_subscriptions | 2 |
| **crud** | **edulevel** | **contextid** | **contextlevelid** | **contextinstanceid** | **userid** | **courseid** |
| c | 2 | 147 | 70 | 66 | 2 | 2 |
| **relateduserid** | **anonymous** | **other** | **timecreated** | **origin** | **ip** | **relateduserid** |
| 2 | 0 | a:1:{s:7:"forumid";s:1:"5";} | 1681205420 | web | 0:0:0:0:0:0:0:1 | NULL |

| Time | User full name | Affected user | Event context | Componen | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Course module viewed | The user w… ed the 'lesson' … se module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:35 | User 59 | - | Lesson: SQL | Lesson | Question answered | The user with id '59' has answered the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |
| 19 Jan 21, 17:34 | User 59 | - | System | System | Group message sent | The user with id '59' sent a message with id '374' to the conversation with id '92'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | User 62 | User 59 | System | Message viewed | The user with id '59' read a message from the user with id '62'. | web | 154.23.39.193 |
| 19 Jan 21, 17:33 | User 59 | - | Lesson: SQL | Lesson | Question viewed | The user with id '59' has viewed the True/false question with id '264' in the lesson activity with course module id '156'. | web | 154.23.39.193 |

Role

Time

Who

What

Where

When

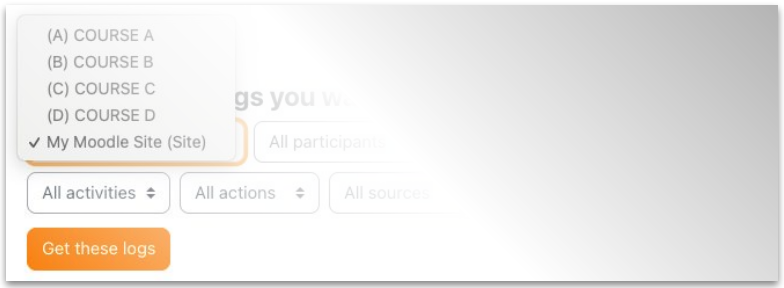
# Our solution

| courseid | timecreated | userid | Time | User full name | Affected user |
|---|---|---|---|---|---|
| 98 | 1615214006 | 37 | 2021-03-08T15:33 | Alex Scyar | - |
| **Event context** | **Component** | **Event name** | **Description** | **Origin** | **IP address** |
| Lesson: Introduction | Lesson | Question viewed | The user with id '37' has viewed the Numerical question with id '484' in the lesson activity with course module id '641' | web | 154.23.39.193 |

*Combined log record*

| Advantages | Drawbacks |
|---|---|
| Directly generated from Moodle. Human readable. | Exceeding of **PHP memory size limit**; need of **table joins** to retrieve roles data and distinguish between site logs and course logs; **admin privileges** for collecting site logs |

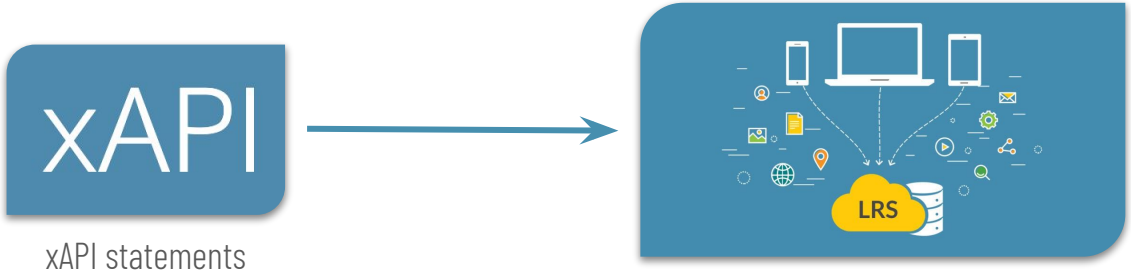
Database




Log Generation Interface


**Experience API**


Source: https://github.com/adlnet/xAPI-Spec/blob/master/xAPI-About.md\#50-xapi-components




xAPI statements

Learning Record Store

# xAPI statement - JSON format

```
{
    "actor": {
        "name": "Alex Scyar",
        "mbox": "mailto:alex@example.com"
    },
    "verb": {
        "id": "http://activitystrea.ms/schema/1.0/read",
        "display": {
            "en": "read"
        }
    },
    "object": {
        "definition": {
            "type": "http://id.tincanapi.com/activitytype/book",
            "name": {
                "en": "book"
            }
        }
    }
}
```

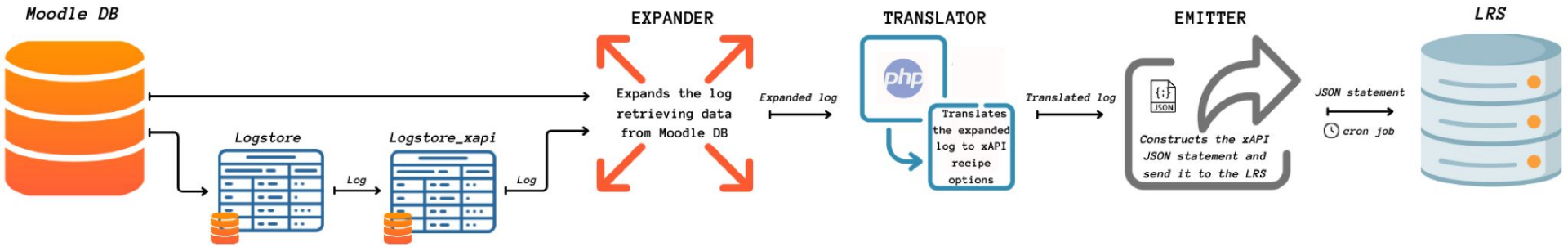
(IRI)

[actor]

[verb]

[object]

# Logstore xAPI plugin




**The plugin architecture.** Source: Daniela Rotelli. Icons @Flaticon.com

**Rotelli**, D., Noël, Y., Lallé, S., Luengo, V., and Pesce, D. (2023, September). A Moodle plugin for rich xAPI data logging. In *Responsive and Sustainable Educational Futures: 18th European Conference on Technology Enhanced Learning, EC-TEL 2023, Aveiro, Portugal, September 4–8, 2023*, Proceedings. Cham: Springer International Publishing.

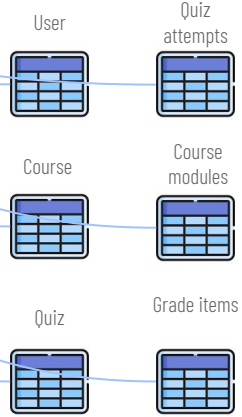
# From PHP to JSON

Logstore

| eventname | target | objecttable | objectid | contextinstanceid | relateduserid | courseid |
|---|---|---|---|---|---|---|
| mod_quiz\attempt_submitted | attempt | quiz_attempts | 4 | 62 | 24 | 38 |

```php
function attempt_submitted(array $config, \stdClass $event) {
    $repo = $config['repo'];
    $user = $repo->read_record_by_id('user', $event->relateduserid);
    $course = $repo->read_record_by_id('course', $event->courseid);
    $attempt = $repo->read_record_by_id('quiz_attempts', $event->objectid);
    $coursemodule = $repo->read_record_by_id('course_modules', $event->contextinstanceid);
    $quiz = $repo->read_record_by_id('quiz', $attempt->quiz);
    $gradeitem = $repo->read_record('grade_items', [
        'itemmodule' => 'quiz',
        'iteminstance' => $quiz->id,
    ]);
    $attemptgrade = $repo->read_record('grade_grades', [
        'itemid' => $gradeitem->id,
        'userid' => $event->relateduserid
    ]);
    $lang = utils\get_course_lang($course);

    return [[
        'actor' => utils\get_user($config, $user),
        'verb' => utils\get_verb('completed', $config, $lang),
        'object' => utils\get_activity\course_quiz($config, $course, $event->contextinstanceid),
        'timestamp' => utils\get_event_timestamp($event),
        'result' => utils\get_attempt_result($config, $attempt, $gradeitem, $attemptgrade),
        'context' => [
            'platform' => $config['source_name'],
            'language' => $lang,
            'extensions' => utils\extensions\base($config, $event, $course),
            'contextActivities' => [
                'other' => [
                    utils\get_activity\quiz_attempt($config, $attempt->id, $coursemodule->id),
                ],
                'grouping' => [
                    utils\get_activity\site($config),
                    utils\get_activity\course($config, $course),
                ],
                'category' => [
                    utils\get_activity\source($config),
                ]
            ],
        ]
    ]];
}
```





```json
{
    "id":"7ccd3322-e1a5-411a-a67d-6a735c76f119",
    "timestamp": "2023-04-23T12:17:00+00:00",
    "actor":{
        "objectType": "Agent",
        "name":"Alex Scyar",
        "mbox":"mailto:alex@example.com"
    },
    "verb":{
        "id":"http://activitystrea.ms/schema/1.0/submit",
        "display":{
            "en":"submitted"
        }
    },
    "object":{
        "id":"yoursite/mod/quiz/attempt.php?attempt=15&cmid=29",
        "definition":{
            "type": "http://adlnet.gov/expapi/activities/attempt"
            "name":{
                "en":"quiz attempt"
            },

        }
    },
    "result":{
        "score":{
            "scaled":0.43
        },
        "success":true,
        "completion":true,
        "duration": "PT385S"
    }
}
```


**Rotelli**, D., Noël, Y., Lallé, S., Luengo, V., and Pesce, D. (2023, September). A Moodle plugin for rich xAPI data logging. In *Responsive and Sustainable Educational Futures: 18th European Conference on Technology Enhanced Learning, EC-TEL 2023, Aveiro, Portugal, September 4-8, 2023*, Proceedings. Cham: Springer International Publishing.

# Logstore xAPI plugin characteristics

| | |
|---|---|
| **Events** | 216 events, related to 74 objects ⟶ ALL **participating** events (student role) of Moodle standard configuration |
| **Historical events** | Management of conversion issues for different **Php versions** and when copying the *logstore* table into the *logstore_xapi* table in case of **deleted items** from the database:<br><br>Course and user not found<br>Deleted modules and activities<br>Missing answers<br>Events implying a deletion |
| **Privacy** | Option to hash data and specify a **secret key** in the plugin configuration |

SCIENCES SORBONNE UNIVERSITÉ

**Test**: 49,993,394 entries of the actions performed by 25,386 users

0 errors ⟶ **no failed statements**

**Rotelli**, D., Noël, Y., Lallé, S., Luengo, V., and Pesce, D. (2023, September). A Moodle plugin for rich xAPI data logging. In *Responsive and Sustainable Educational Futures: 18th European Conference on Technology Enhanced Learning, EC-TEL 2023, Aveiro, Portugal, September 4-8, 2023*, Proceedings. Cham: Springer International Publishing.

# Comparison between Moodle log data extraction methods

| Collection method | Advantages | Drawbacks |
|---|---|---|
| Database | A basic SQL query to access data | **Unintelligible** to humans; need of **multiple table joins** to retrieve information (*user roles*, *deleted activities* and *users*) and be human readable; need of **in-depth knowledge** of Moodle database (**416 tables** in the standard configuration + plugins); **admin privileges** |
| Moodle Log Generation Interface | Directly generated from Moodle. Human readable. | Exceeding of **PHP memory size limit**; need of **table joins** to retrieve roles data and distinguish between site logs and course logs; **admin privileges** for collecting site logs |
| xAPI statements | Comprehensive information, no additional table joins, no admin privileges | Demanding and **laborious process**, dependent on xAPI requirements, creation of **new statements** for each new plugin, installation of a **LRS** |

# Data Preprocessing

# Combined log record

| courseid | timecreated | userid | Time | User full name | Affected user |
|---|---|---|---|---|---|
| 98 | 1615214006 | 37 | 2021-03-08T15:33 | Alex Scyar | - |
| **Event context** | **Component** | **Event name** | **Description** | **Origin** | **IP address** |
| Lesson: Introduction | Lesson | Question viewed | The user with id '37' has viewed the Numerical question with id '484' in the lesson activity with course module id '641' | web | 154.23.39.193 |

# Data integration: role

student (roleid=5), teacher (roleid=3), and non-editing teacher (roleid=4)

```
SELECT cx.instanceid as courseid, u.id as userid
FROM mdl_course c
LEFT OUTER JOIN mdl_context cx ON c.id = cx.instanceid
LEFT OUTER JOIN mdl_role_assignments ra ON cx.id = ra.contextid AND ra.roleid = '???' AND
      cx.instanceid <> 1
LEFT OUTER JOIN mdl_user u ON ra.userid = u.id WHERE cx.contextlevel = '50'
```

manager (roleid=1) and course creator (roleid=2)

```
SELECT distinct userid
FROM mdl_role_assignments
WHERE roleid = '???'
```

admin

```
SELECT id, shortname (or fullname)
FROM mdl_course
```

# Data integration: area categorisation





Icons@Freepik

**Rotelli, D.**, & Monreale, A. (2023). Processing and Understanding Moodle Log Data and their Temporal Dimension. *Journal of Learning Analytics*.

# Data integration: area categorisation


courseid = 0 or 1

| Area | Event name | Annotation |
|---|---|---|
| Authentication | User has logged in, User logged out. | |
| Moodle Site | Blog entries viewed[1], Calendar event created, Calendar event deleted, Calendar event updated, Calendar subscription created, Calendar subscription updated, Category viewed, Course module instance list viewed[2], Course module viewed[2], Course viewed[2], Courses searched, Discussion created[3], Discussion subscription created[3], Discussion viewed[3], Discussion deleted[3] Grade overview report viewed[4] Notification viewed, Subscription created[2], User report viewed[5]. | [1]The Affected user matches the User full name; [2]Courseid=1; [3]Courseid=1 and the component of these events is *Forum*; [4]The *Grade overview report* can be reached from the profile page. [5]The Affected user matches the User full name and the component is *Forum*; |
| Profile | Badge viewed[6], Course user report viewed[7], Dashboard reset, Dashboard viewed, Grade overview report viewed[7], Tag added/removed to/from an item[6], Tag created/deleted[6], User password updated, User updated, User profile viewed[8]. | [6]Courseid=0; [7]Courseid=1; [8]The Affected user matches the User full name; |
| Social interaction | Blog entries viewed[9], Group message sent, Message contact added, Message contact removed, Message deleted, Message sent[10], Message viewed, User profile viewed[11], User report viewed[12]. | [9]Courseid=0; [10]The component is 'System'. The event 'Message sent' is also available for the 'Chat' component and is recorded at the course level; [11]The Affected user does not match the User full name and courseid=0; [12]The Affected user does not match the User full name and the component is *Forum*. |

Icons@Freepik

# Data integration: component redefinition and alignment

| Time | User full name | Affected user | Event context | Component | Event name | Description | Origin | IP address |
|---|---|---|---|---|---|---|---|---|
| 15 January 2021, 5:33 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 5:33 PM | USER 59 | USER 59 | Attendance: OpenLab | Attendance | Session report viewed | User with id 59 viewed attendance sessions for student with id 59 | web | 87.15.222.193 |
| 15 January 2021, 5:35 PM | USER 59 | USER 59 | User: USER 59 | System | Dashboard viewed | The user with id '59' has viewed their dashboard | web | 87.15.222.193 |
| 15 January 2021, 5:42 PM | USER 59 | USER 59 | User: USER 59 | System | Dashboard viewed | The user with id '59' has viewed their dashboard | web | 87.15.222.193 |
| 15 January 2021, 6:02 PM | USER 59 | - | Course: A | System | Course viewed | The user with id '59' viewed the course with id '21'. | web | 87.15.222.193 |
| 15 January 2021, 6:02 PM | USER 59 | - | H5P: Answer the questions | H5P | Course module viewed | The user with id '59' viewed the 'h5pactivity' activity with course module id '36'. | web | 87.15.222.193 |
| 15 January 2021, 6:02 PM | USER 59 | - | H5P: Answer the questions | H5P Package | H5P content viewed | The user with id '59' has viewed the H5P with the id '28'. | web | 87.15.222.193 |
| 15 January 2021, 6:04 PM | USER 59 | - | H5P: Answer the questions | H5P | xAPI statement received | The user with the id '59' send a tracking statement for a H5P activity with the course module id '36'. | web | 87.15.222.193 |
| 15 January 2021, 6:04 PM | USER 59 | - | Course: A | System | Course viewed | The user with id '59' viewed the course with id '21'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | Category: Overall Site | System | Category viewed | The user with id '59' viewed the course category with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | Course: B | System | Course viewed | The user with id '59' viewed the course with id '2'. | web | 87.15.222.193 |
| 15 January 2021, 6:05 PM | USER 59 | - | File: Intro 01 | File | Course module viewed | The user with id '59' viewed the 'resource' activity with course module id '245'. | web | 87.15.222.193 |

Yellow: course B logs; Red: platform logs; Green: course A logs

| New Component | Event name | Annotation |
|---|---|---|
| Badge | Badge viewed, Badge listing viewed | |
| Blog | Blog entries viewed | |
| Calendar | Calendar subscription updated, Calendar subscription created, Calendar event updated, Calendar event deleted, Calendar event created | |
| Course home | Course viewed[1] | [1]The Event Context is "Course: course_name". |
| Courses list | Category viewed, Courses searched | |
| Dashboard | Dashboard viewed, Dashboard reset | |
| Grades | Course user report viewed, Grade overview report viewed | |
| Messaging | Message sent[2], Group message sent, Message deleted, Message viewed, Message contact added | [2]The component is 'System'. |
| Notification | Notification viewed | |
| Participant profile | User list viewed, User profile viewed[3] | [3] The Affected user does not match the User full name. |
| Quiz* | Question viewed[4] | [4]The event 'Question viewed' is also available for the 'Lesson' component. |
| Tag | Tag added/removed to an item, Tag created/deleted | |
| User profile | User password updated, User updated, User profile viewed[5] | [5]The Affected user matches the User full name. |
| Site home | Course viewed[6] | [6]The Event Context is "Front page" and not the name of the course. |

| Component | Old Component |
|---|---|
| Assignment | File submissions, Submission comments, Online text submissions |
| Book | Book printing |
| Grades | User report |
| H5P | H5P Package |

# Data cleaning

- Cron jobs
  - user name (-)
  - event name: '*Prediction process started*', '*Web service token created*'
  - 136,446 out of 318,249 logs
- User login failed
- Deleted activities
- Deleted users
- NO Missing data

# Consolidated Moodle logs

| Unix_Time | Date&Time | Role | Username | userid |
|---|---|---|---|---|
| 1589034114 | 2020-05-09 14:21:54 | Student | Alex Scyar | 37 |
| **courseid** | **Course_Area** | **Context** | **Component** | **Event_name** |
| 24 | History of Arts | Pablo Picasso | Lesson | Question answered |

Example of a consolidated record extracted from the Moodle **log generation interface**.

# Research questions?

# Time management

"Time management is the ability of students to schedule, plan, and manage their personal study time" (Pintrich, 1991)

Academic Performance
(Trueman, & Hartley, 1996)

Well-being
(stress reduction)
(Misra, & McKean, 2000)

Misra, R., & McKean, M. (2000). College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction. American journal of Health studies, 16(1), 41.
Trueman, M., & Hartley, J. (1996). A comparison between the time-management skills and academic performance of mature and traditional-entry university students. Higher education, 32(2), 199-215.
Pintrich, P. R. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).

# What are the challenges?

### Data Accuracy

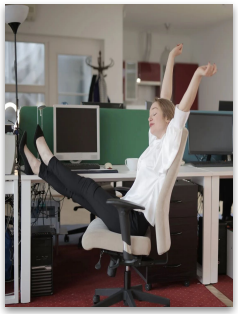
Distinguishing between
**time-on-task** and **time-off-task**

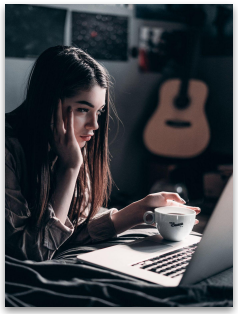
Source of images: https://www.pexels.com/, https://unsplash.com/

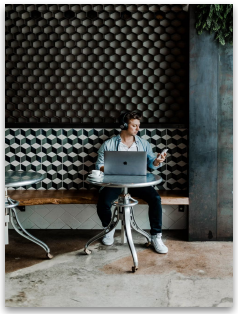
7h59m59s

Source of images: https://unsplash.com/; https://www.pexels.com/

7h59m59s







Source of images: https://unsplash.com/; https://www.pexels.com/

Elapsed time

Time-on-task

Time-off-task

Time-on-task

Source of images: https://unsplash.com/; https://www.pexels.com/

**Time-on-task**: the amount of time students are actively involved in quality learning

**Time-off-task**: the time spent in activities other than learning (Carroll, 1963)

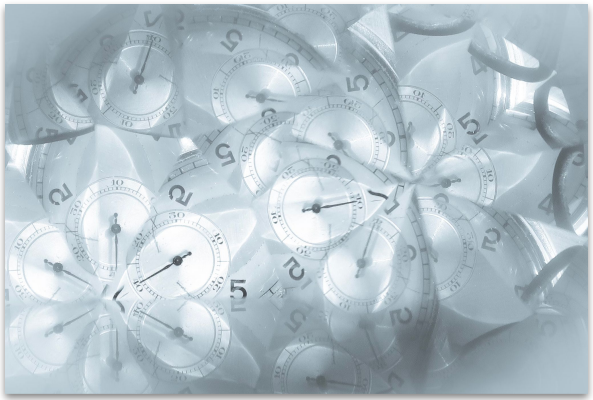
**Elapsed time** = time_**on**_task + time_**off**_task

John Carroll. A model of school learning. *Teachers college record, 64(8):723–723, 1963*

# What are the challenges?

### Data Accuracy

Distinguishing between **time-on-task** and **time-off-task**

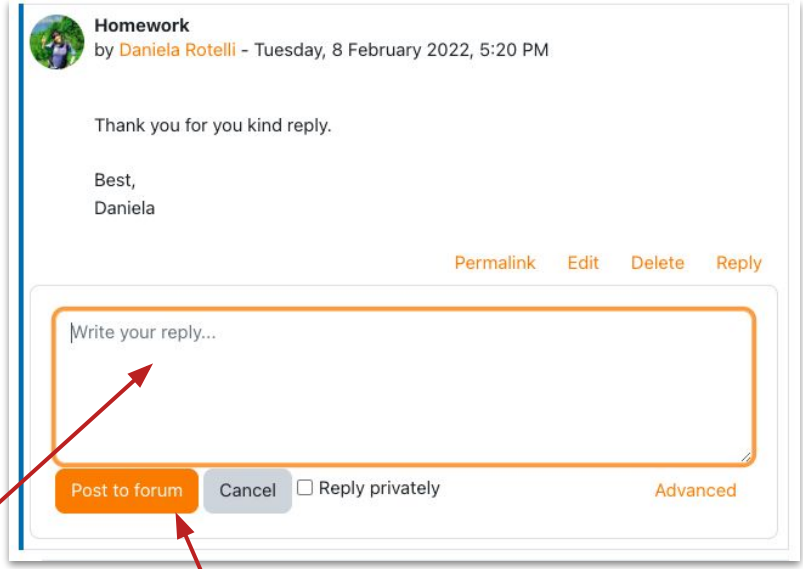


### Interpretation

**Understanding the educational context** is fundamental to avoid erroneous conclusions.

Source of images: https://www.pexels.com/, https://unsplash.com/

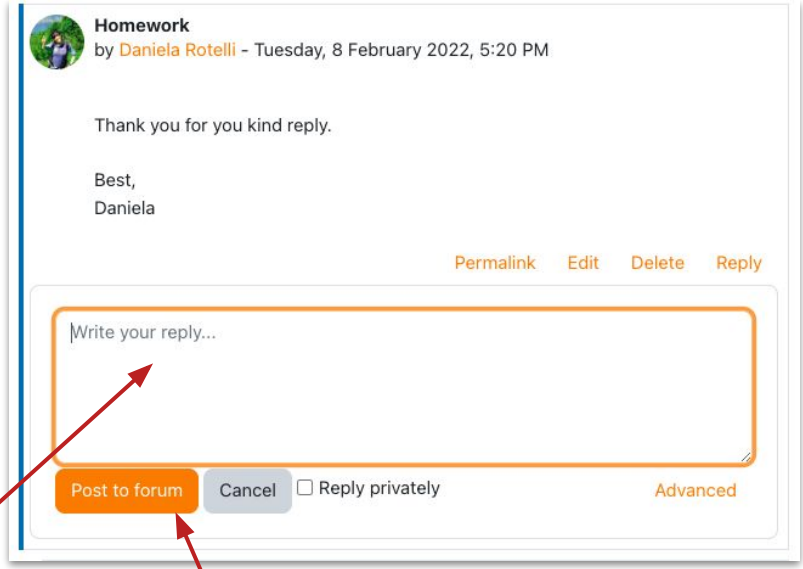
# The temporal dimension of Moodle Log Data




Writing in a forum
(AJAX)

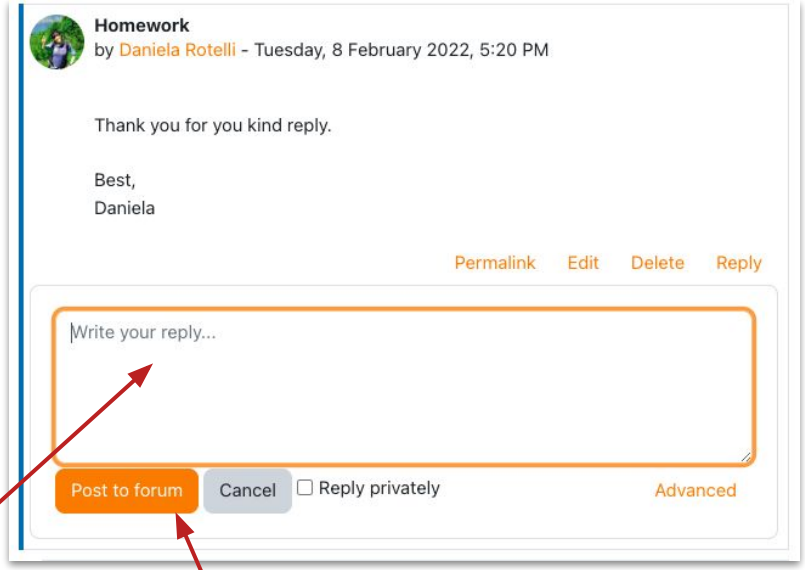
**When does the writing start?**

**End of writing**




Writing in a forum (AJAX)

$$d_i = t_{i+1} - t_i$$




Writing in a forum
(AJAX)

When does the writing start?

End of writing

$d_i = \times_1 - t_i$

# Our solution

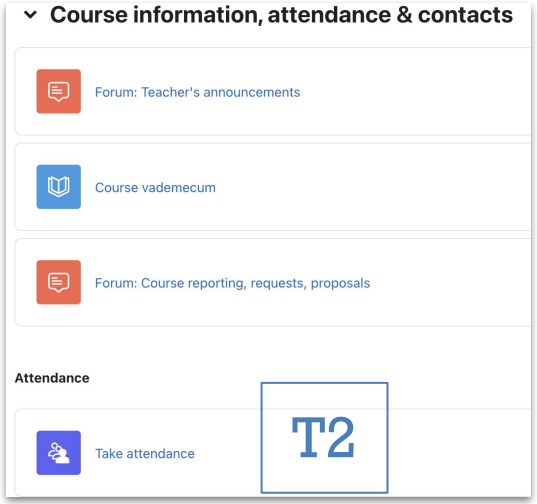
Categories of events

- Starting
- Ending
- Simultaneous
- Opening & closing
- Instantaneous

# Starting events

| T | Date&Time | Difference | Event name | Category | Computation | Duration |
|---|---|---|---|---|---|---|
| 0 | 2021-01-20T17:58:36 | 4 | Course viewed | Starting | $d_i = t_{i+1} - t_i$ | 4 |
| 1 | 2021-01-20T17:58:40 | 12 | Session report viewed | Starting | $d_i = t_{i+1} - t_i$ | 12 |
| 2 | 2021-01-20T17:58:52 | 3 | Course viewed | Starting | $d_i = t_{i+1} - t_i$ | 3 |
| 3 | 2021-01-20T17:59:55 | 37 | Course module viewed | Starting | $d_i = t_{i+1} - t_i$ | 37 |

T1



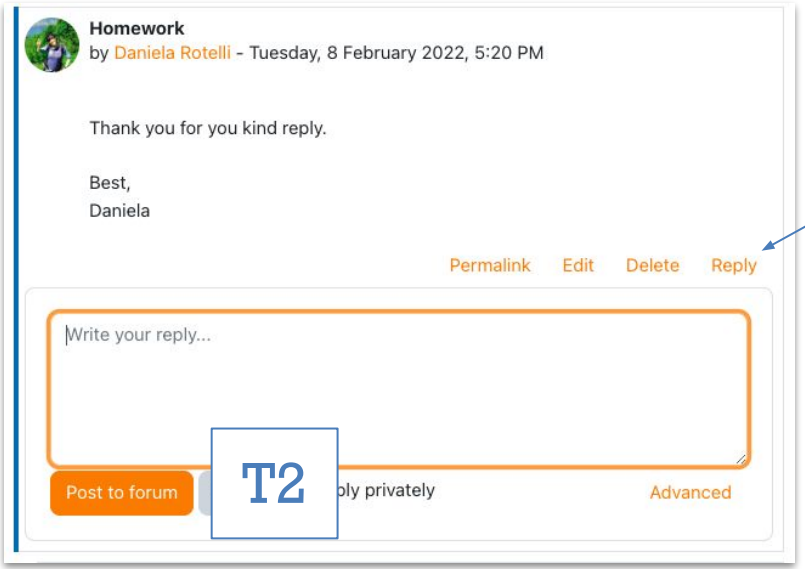

$t_{i+1} - t_i = d_i$

**Rotelli**, D., & Monreale, A. (2023, August). Processing and Understanding Moodle Log Data and their Temporal Dimension. *Journal of Learning Analytics*.

# Ending events

| T | Date&Time | Difference | Event name | Category | Computation | Duration |
|---|---|---|---|---|---|---|
| 0 | 2021-01-13T16:34:52 | 2 | Course module viewed | Starting | $d_i = t_{i+1} - t_i$ | 2 |
| 1 | 2021-01-13T16:34:54 | 97 | Discussion viewed | Starting | - | - |
| 2 | 2021-01-13T16:36:31 | 0 | Some content has been posted. | Ending | $d_i + d_{i+1} = t_{i+1} - t_i$ | 97 |

T1




T2

?

$$t_{i+1} - t_i = d_i + d_{i+1}$$

**Rotelli**, D., & Monreale, A. (2023, August). Processing and Understanding Moodle Log Data and their Temporal Dimension. *Journal of Learning Analytics*.

# Simultaneous events

| T | Date&Time | Difference | Event name | Category | Computation | Duration |
|---|---|---|---|---|---|---|
| 0 | 2021-04-29T18:12:20 | 0 | Course module viewed | Simultaneous | $d_i = 0$ | 0 |
| 1 | 2021-04-29T18:12:20 | 3 | The status of the submission has been viewed. | Starting | $d_i = t_{i+1} - t_i$ | 3 |
| 2 | 2021-04-29T18:12:23 | 0 | Course module viewed | Simultaneous | $d_i = 0$ | 0 |
| 3 | 2021-04-29T18:12:23 | 6 | Submission form viewed. | Starting | $d_i = t_{i+1} - t_i$ | 6 |
| 4 | 2021-04-29T18:12:29 | 0 | Course module viewed | Simultaneous | $d_i = 0$ | 0 |
| 5 | 2021-04-29T18:12:29 | 0 | A file has been uploaded. | Simultaneous | $d_i = 0$ | 0 |
| 6 | 2021-04-29T18:12:29 | 0 | Submission created. | Ending | $d_i = 0$ | 0 |
| 7 | 2021-04-29T18:12:29 | 0 | Course module viewed | Simultaneous | $d_i = 0$ | 0 |
| 8 | 2021-04-29T18:12:29 | 5 | The status of the submission has been viewed. | Starting | $d_i = t_{i+1} - t_i$ | 5 |

$$d_i = 0$$

# Opening & Closing events

| T | Date&Time | Difference | Event name | Category | Computation | Duration |
|---|---|---|---|---|---|---|
| 0 | 2021-01-19T08:23:58 | 0 | Course module viewed | Simultaneous | $d_i = 0$ | 0 |
| 1 | 2021-01-19T08:23:58 | 76 | Content page viewed | Starting | $d_i = t_{i+1} - t_i$ | 76 |
| 2 | 2021-01-19T08:25:14 | 0 | Course module viewed | Simultaneous | $d_i = 0$ | 0 |
| 3 | 2021-01-19T08:25:14 | **21** | **Question viewed** | Opening | - | - |
| 4 | 2021-01-19T08:25:35 | **13** | Question answered | Closing | $d_{i+1} = t_{i+1} - t_i$ | **21** |
| 5 | - | - | **Question reviewed** | Starting | $d_i = t_{i+1} - t_i$ | **13** |
| 6 | 2021-01-19T08:25:48 | 0 | Course module viewed | Simultaneous | $d_i = 0$ | 0 |
| 7 | 2021-01-19T08:25:48 | 367 | Content page viewed | Starting | $d_i = t_{i+1} - t_i$ | 367 |

T3

**What is an Information System?**

(More answers are needed)

- [ ] It supports the organization activities by managing information.
- [ ] It is a component of an organization.
- [ ] Any organization has an Information System.
- [ ] It collects, stores, processes and communicates the information.

T4

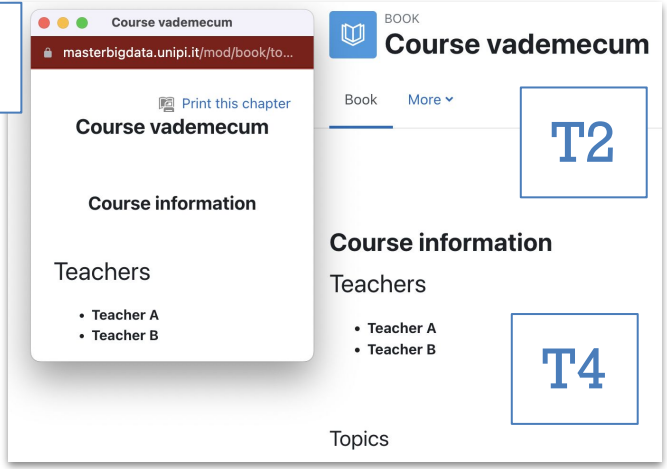
Submit

$$t_{i+1} - t_i = d_{i+1}$$

**Rotelli**, D., & Monreale, A. (2023, August). Processing and Understanding Moodle Log Data and their Temporal Dimension. *Journal of Learning Analytics*.

# Instantaneous events

| T | Date&Time | Difference | Event name | Category | Computation | Duration |
|---|---|---|---|---|---|---|
| 0 | 2021-03-17T12:53:54 | 0 | Course module viewed | Simultaneous | $d_i = 0$ | 0 |
| 1 | 2021-03-17T12:53:54 | 14 | Chapter viewed | Starting | $d_i = t_{i+1} - t_i$ | 14 |
| 2 | 2021-03-17T12:54:08 | 13 | Chapter viewed | Starting | $d_{i-1} = (t_{i+1} - t_i) + (t_i - t_{i-1})$ | 106 |
| 3 | 2021-03-17T12:54:21 | 93 | Chapter printed | Instantaneous | $d_i = 0$ | 0 |
| 4 | 2021-03-17T12:55:44 | 11 | Chapter viewed | Starting | $d_i = t_{i+1} - t_i$ | 11 |




$$d_i = 0$$

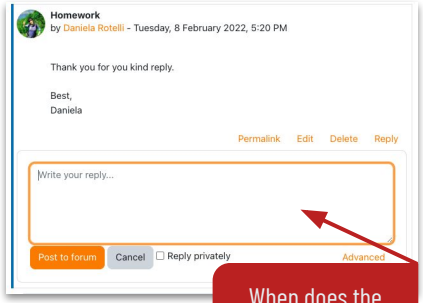
$$(t_{i+1} - t_i) + (t_i - t_{i-1}) = d_{i-1}$$

# Event categorisation

| Component | Event name | Category | Annotation |
|---|---|---|---|
| **Assignment** | A file has been uploaded | Simultaneous | |
| | A submission has been submitted | Simultaneous | The event is *instantaneous* if students are requested to press the "Submit" button.[27] |
| | An online text has been uploaded | Simultaneous | |
| | Comment created | Ending | The time that begins when this event is recorded must be considered "The status of the submission has been viewed" because the user remains on the same page. |
| | Comment deleted | Instantaneous | |
| | Course module instance list viewed | Starting | |
| | Course module viewed | Simultaneous | |
| | Feedback viewed | Simultaneous | |
| | Remove submission confirmation viewed | Starting | |
| | Submission confirmation form viewed | Starting | |
| | Submission created | Ending | This event must be swapped with the "Course module viewed" event to be *ending* (Table 6). |
| | Submission form viewed | Starting | |
| | Submission updated | Ending | This event must be swapped with the "Course module viewed" event to be *ending* (Table 6). |
| | The status of the submission has been updated | Simultaneous | |
| | The status of the submission has been viewed | Starting | |
| | The user duplicated their submission | Ending | This event must be swapped with the "Course module viewed" event to be *ending* (Table 6). |
| **Badge** | Badge listing viewed | Starting | |
| | Badge viewed | Starting | |
| **Blog** | Blog entries viewed | Starting | |
| **Book** | Book printed | Instantaneous | |
| | Chapter printed | Instantaneous | |
| | Chapter viewed | Starting | |
| | Course module viewed | Simultaneous | |
| **Calendar*** | Calendar event created | Ending | Since these events are not accompanied by other concurrent events indicating the type of action (such as in the Forum), the time that begins when they are recorded should be considered "Calendar viewed." Whenever a user with the teacher role creates activities with a due date, these events are *Simultaneous*. |
| | Calendar event deleted | Ending | |
| | Calendar event updated | Ending | |

Some examples of classification

**Rotelli**, D., & Monreale, A. (2023, August). Processing and Understanding Moodle Log Data and their Temporal Dimension. *Journal of Learning Analytics*.

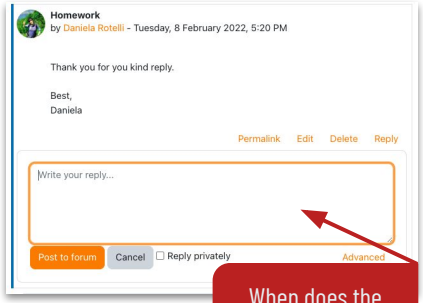



When does the writing start?

Time spent (in seconds) in six forums of Course B

| | Events | | | | | | |
|---|---|---|---|---|---|---|---|
| Forum ID | Course module viewed | Discussion viewed | Discussion created | Post created | Post updated | Some content has been posted | Forum Total |
| 1 | 7,685 | 122,303 | 0 | 0 | 0 | 0 | 129,988 |
| 2 | 5,835 | 197,106 | 0 | 76,559 | 0 | 0 | 279,500 |
| 3 | 63,972 | 132,590 | 0 | 41,223 | 0 | 0 | 237,785 |
| 4 | 321 | 0 | 0 | 0 | 0 | 0 | 321 |
| 5 | 1,165 | 12,643 | 0 | 16,026 | 0 | 0 | 29,834 |
| 6 | 716 | 1,762 | 0 | 1,611 | 0 | 0 | 4,089 |

Starts to **read** not to write




When does the writing start?

Time spent (in seconds) in six forums of Course B

| Forum ID | Events | | | | | | Forum Total |
|---|---|---|---|---|---|---|---|
| | Course module viewed | Discussion viewed | Discussion created | Post created | Post updated | Some content has been posted | |
| 1 | 7,685 | 122,303 | 0 | 0 | 0 | 0 | 129,988 |
| 2 | 5,835 | 197,106 | 0 | 76,559 | 0 | 0 | 279,500 |
| 3 | 63,972 | 132,590 | 0 | 41,223 | 0 | 0 | 237,785 |
| 4 | 321 | 0 | 0 | 0 | 0 | 0 | 321 |
| 5 | 1,165 | 12,643 | 0 | 16,026 | 0 | 0 | 29,834 |
| 6 | 716 | 1,762 | 0 | 1,611 | 0 | 0 | 4,089 |

Starts to **read** not to write

$$d_i = t_{i+1} - t_i$$

| Forum ID | Without event category | | | | | With event category | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Module view | Reading | Writing | Reading & Writing | Forum Total | Module view | Reading | Writing | Reading & Writing | |
| 1 | 7,685 | 122,303 | 0 | 0 | 129,988 | 7,685 | 122,303 | 0 | 0 | |
| 2 | 5,835 | 273,575 | 0 | 0 | 279,500 | 5,771 | 268,734 | 64 | 4,920 | |
| 3 | 63,972 | 173,813 | 0 | 0 | 237,785 | 62,607 | 172,231 | 1,365 | 1,582 | |
| 4 | 321 | 0 | 0 | 0 | 321 | 321 | 0 | 0 | 0 | |
| 5 | 1,165 | 28,669 | 0 | 0 | 29,834 | 663 | 27,558 | 502 | 1,111 | |
| 6 | 716 | 3,373 | 0 | 0 | 4,089 | 469 | 3,150 | 247 | 223 | |

*Starting:* $\hat{d}_i = d_{i+1} - d_i$

*Ending:* $\hat{d}_i + \hat{d}_{i+1} = d_{i+1} - d_i$

*Open&Clos:* $\hat{d}_{i+1} = d_{i+1} - d_i$

*Instantaneous:* $\hat{d}_i = 0$

*Simultaneous:* $\hat{d}_i = 0$

**Rotelli**, D., & Monreale, A. (2023, August). Processing and Understanding Moodle Log Data and their Temporal Dimension. *Journal of Learning Analytics*.

# Research questions?

**Elapsed time** = time_**on**_task + time_**off**_task

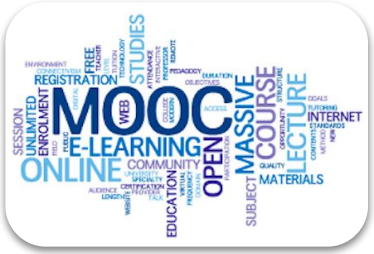
1) How can we estimate the **time-on-task**?
2) How can we identify **outliers** (extremely long durations)?

# Online Learning Environments

docebo

edmodo

CANVAS




Bb
Blackboard
LEARNING MANAGEMENT SYSTEM

moodle


udemy

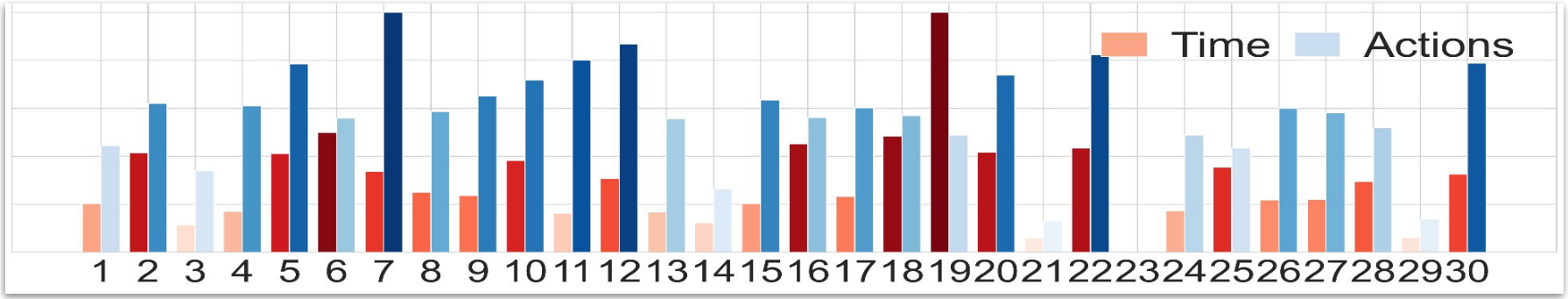
Cp
Adobe Captivate

coursera

OPEN edX

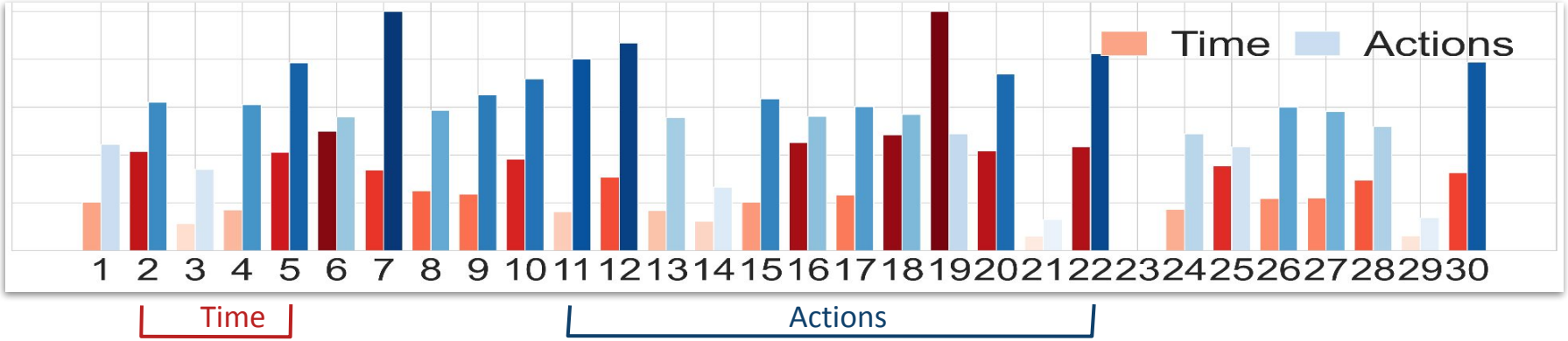
# Data Transformation

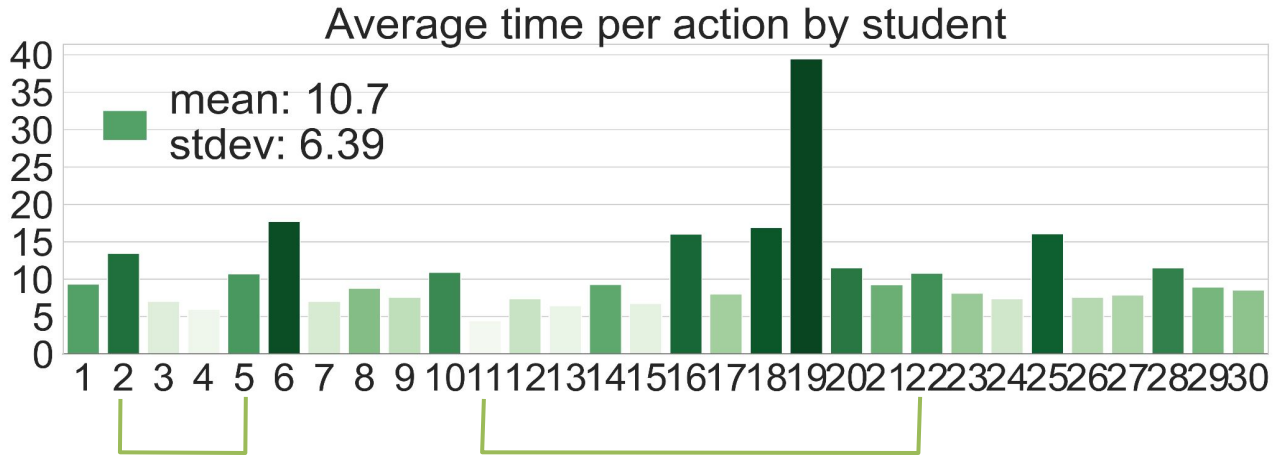
# Total time spent VS Number of actions

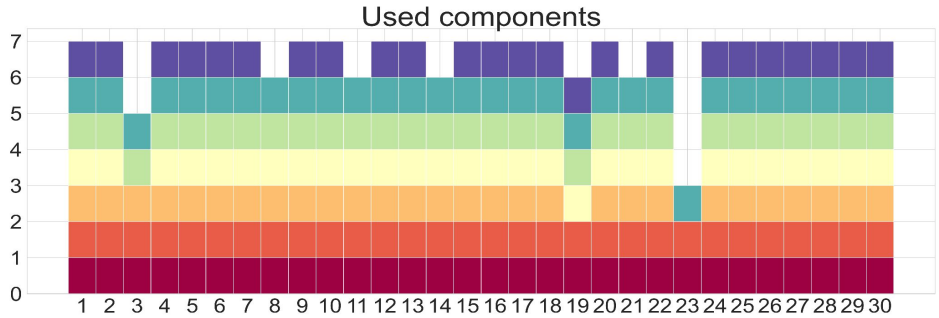



**X-axis**: student ID

# Total time spent VS Number of actions



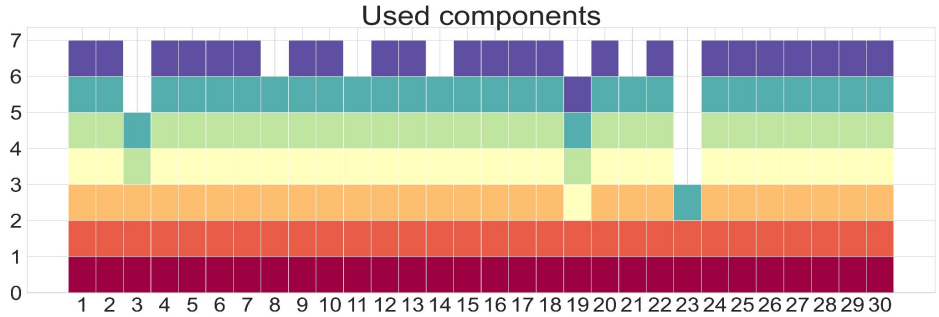





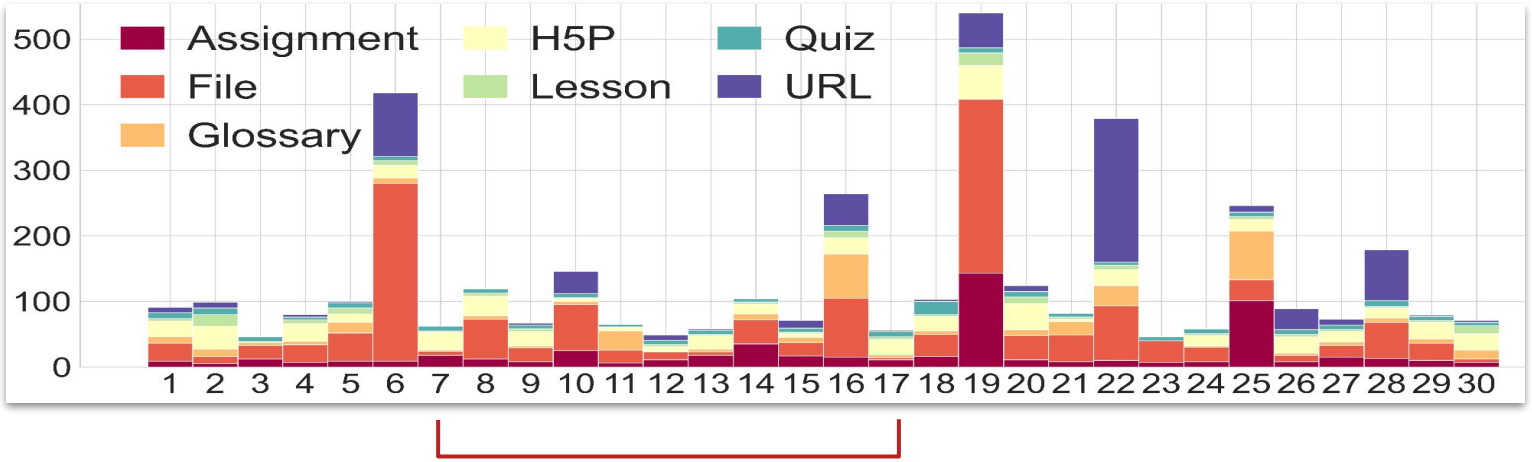


**Rotelli**, D., Monreale, A., & Guidotti, R. (2022, September). Uncovering Student Temporal Learning Patterns. In *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption: 17th European Conference on Technology Enhanced Learning, EC-TEL 2022, Toulouse, France, September 12–16, 2022, Proceedings (pp. 340-353)*. Cham: Springer International Publishing.

# Average time-on-task








**Rotelli**, D., Monreale, A., & Guidotti, R. (2022, September). Uncovering Student Temporal Learning Patterns. In *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption: 17th European Conference on Technology Enhanced Learning, EC-TEL 2022, Toulouse, France, September 12–16, 2022, Proceedings (pp. 340-353)*. Cham: Springer International Publishing.

# At what time students learn?

Student 1

| t time-windows | mon | tue | wed | thu | fri | sat | sun |
|---|---|---|---|---|---|---|---|
| morning | 1 - Book, File<br>2- File, Lesson | 1 - File<br>2- Lesson | 1- Book, Lesson | 1 - Forum, Chat<br>2- File,, Chat | | | |
| afternoon | | | | | | | |
| evening | | | | | | | |

d day-intervals

Student 2

| t time-windows | mon | tue | wed | thu | fri | sat | sun |
|---|---|---|---|---|---|---|---|
| morning | | | | | | 1 - Video<br>2- H5P, Quiz | 1 -, Assignment, Wiki<br>2- Wiki, Forum |
| afternoon | | | | | | | |
| evening | | | | | | | |

d day-intervals

# Which days students learn?

Student 1

| t time-windows | mon | tue | wed | thu | fri | sat | sun |
|---|---|---|---|---|---|---|---|
| morning | 1 - Book, File<br>2- File, Lesson | 1 - File<br>2- Lesson | 1- Book, Lesson | 1 - Forum, Chat<br>2- File,, Chat | | | |
| afternoon | | | | | | | |
| evening | | | | | | | |

d day-intervals

Student 2

| t time-windows | mon | tue | wed | thu | fri | sat | sun |
|---|---|---|---|---|---|---|---|
| morning | | | | | | 1 - Video<br>2- H5P, Quiz | 1 -, Assignment, Wiki<br>2- Wiki, Forum |
| afternoon | | | | | | | |
| evening | | | | | | | |

d day-intervals

# Lectures, Laboratories, Days off?

Student 1

| t time-windows | mon | tue | wed | thu | fri | sat | sun |
|---|---|---|---|---|---|---|---|
| morning | 1 - Book, File<br>2- File, Lesson | 1 - File<br>2- Lesson | 1- Book, Lesson | 1 - Forum, Chat<br>2- File,, Chat | | | |
| afternoon | | | | | | | |
| evening | | | | | | | |

d day-intervals

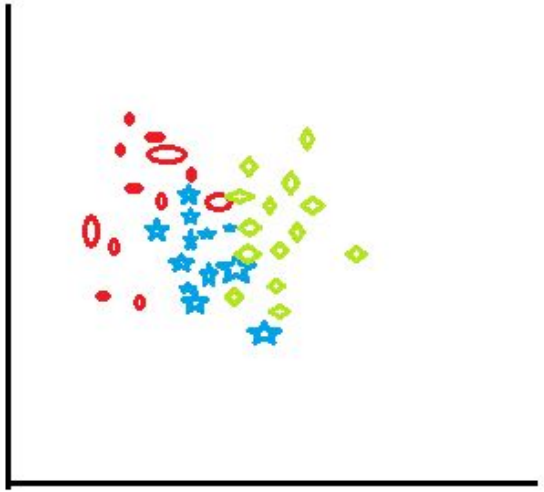
Student 2

| t time-windows | mon | tue | wed | thu | fri | sat | sun |
|---|---|---|---|---|---|---|---|
| morning | | | | | | - Video<br>- H5P, Quiz | 1 -, Assignment, Wiki<br>2- Wiki, Forum |
| afternoon | | | | | | | |
| evening | | | | | | | |

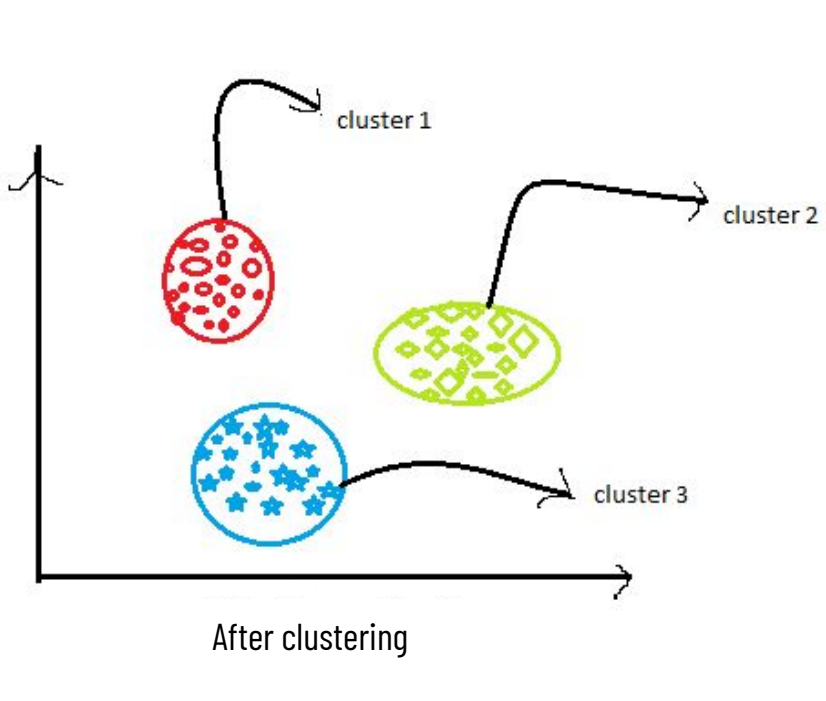
d day-intervals

# Data Mining

# Clustering: K-means

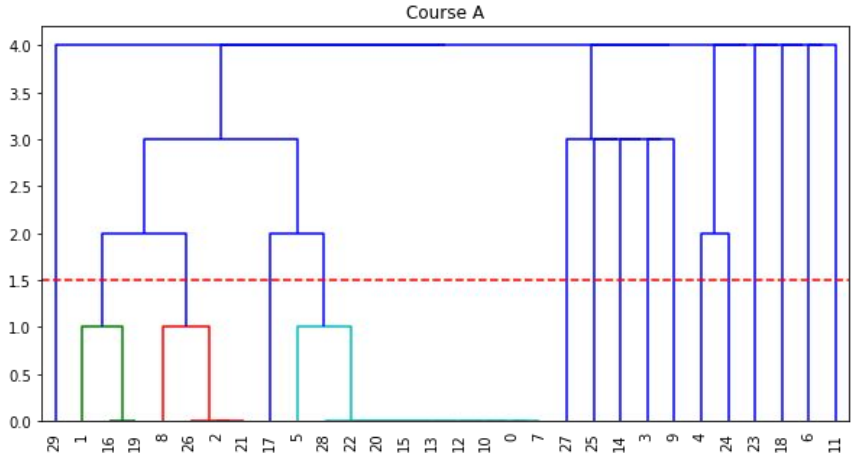


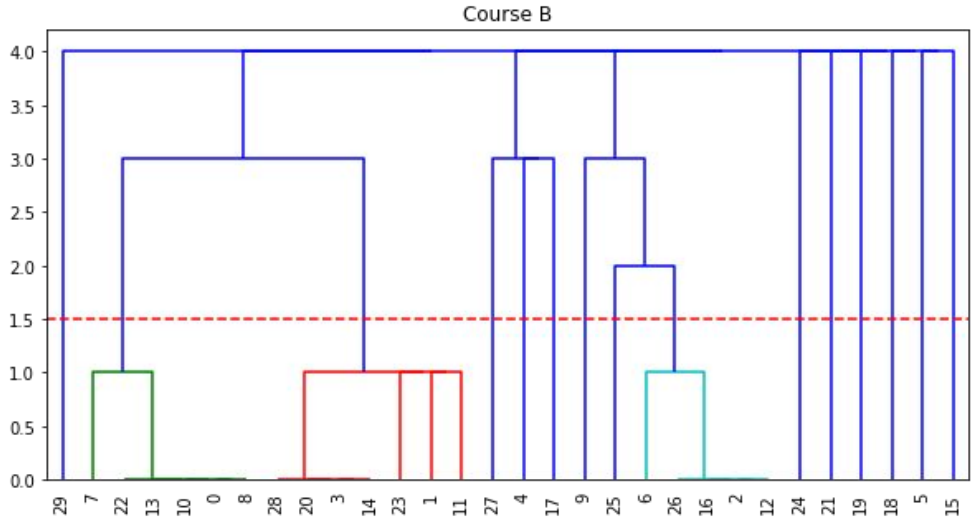
Before clustering

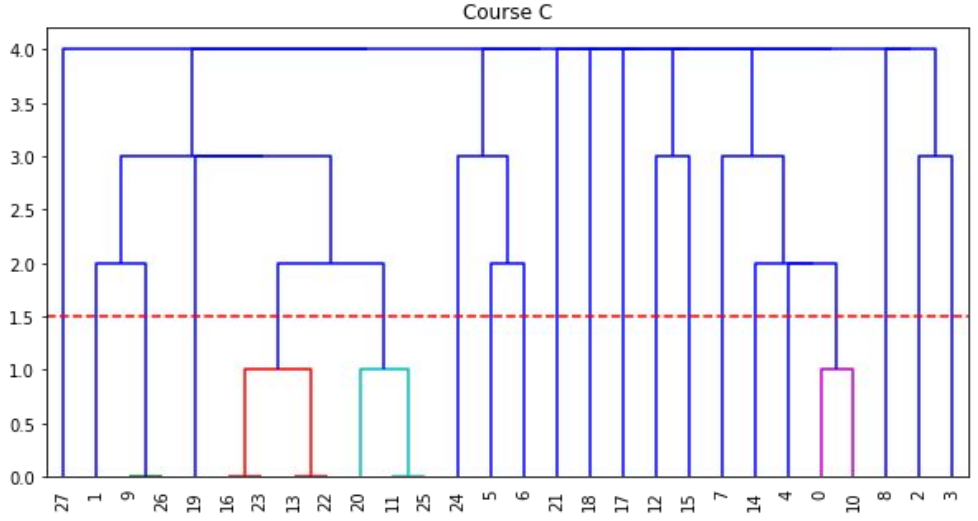



After clustering

# Hierarchical Clustering



# Methods for educational knowledge discovery

| Title | Description | Key applications |
|---|---|---|
| **Causal mining** | To find causal relationship or to identify causal effect in data | Finding what features of student behavior cause learning, failure etc. |
| **Clustering** | To identify groups of similar observations | Grouping similar students based on their interaction patterns |
| **Distillation of data** | To represent data in intelligible ways, e.g. with summarization | Helping instructors to visualize and analyze the ongoing activities |
| **Knowledge tracing** | To estimate student mastery of skills | Monitoring student knowledge over time |
| **Outlier detection** | To point out significantly different individuals | Detection of students with difficulties or irregular learning processes |
| **Prediction** | To infer a target variable from some combination of other variables | Predicting student performance and detecting student behaviors |
| **Process mining** | To obtain knowledge of the process from event logs | Reflecting students' behavior based on traces of their evolution |
| **Recommendation** | To predict the rating or preference a user would give to an item | To make recommendations to students with respect to their activities |
| **Relationship mining** | To study relationships among variables and to encode rules | Identifying relationships in learner behavior and diagnosing difficulties |
| **Social network analysis** | To analyze the social relationships between entities in networks | Interpretation of the structure and relations in collaborative activities |
| **Text mining** | To extract high-quality information from text | Analyzing the contents of forums, chats, web pages, and documents |
| **Visualization** | To show a graphical representations of data | To produce data visualizations that communicate research results |

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

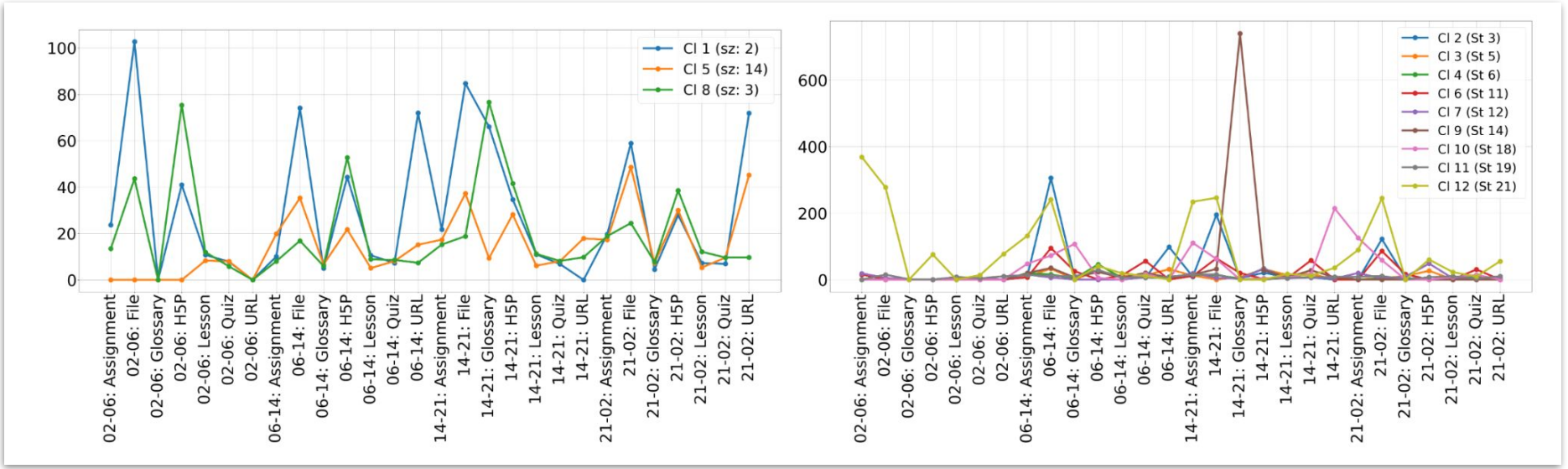
# Applications of educational knowledge discovery

| Topics of interest | Description |
|---|---|
| **Analyzing educational theories** | To analyze how learning theories and learning analytics could be integrated in educational research |
| **Analyzing programming code** | To apply techniques focused on analyzing code from programming courses, programming assignments/submissions |
| **Collaborative learning and teamwork** | To analyze collaborative learning and to predict the team grade in teamwork groups |
| **Curriculum mining and analytics** | To analyze program structure, course grading, and administrative data to improve curriculum development, program quality |
| **Dashboards and visual analytics** | To apply a visualization technique to explore and understand relevant user traces |
| **Early warning systems** | To predicting student's performance and students at risk as soon as possible to intervene early to facilitate student success |
| **Emotional learning analytics** | To study affect during learning and the importance of emotion to learning |
| **Game learning analytics** | To apply data-mining and visualization techniques to player interactions in serious games |
| **Interpretable learner models** | To develop “white box” interpretable, explanatory, usable, and highly comprehensible leaner models |
| **Learning foreign language** | To apply techniques for improving of foreign language learning |
| **Measuring self-regulated learning** | To apply techniques to measure self-regulated learning feature and behaviors in students |
| **Multimodal learning analytics** | To apply machine learning and sensor technologies for providing new insights that happen across multiple contexts |
| **Providing personalized feedback** | To generate personalized feedback automatically or semi-automatically to support the student learning |

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, *10*(3), e1355.

# Data Interpretation

# Time-windows




Course B - Ward linkage

**Cluster 1**
**File, URL**: morning, evening
**Glossary, H5P**: afternoon

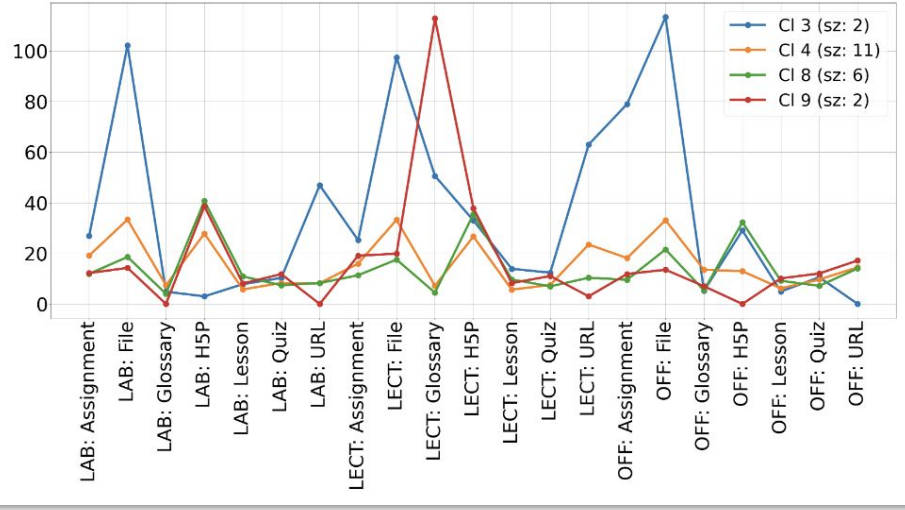
**Cluster 8**
**H5P**: morning, evening
**Glossary**: afternoon

**Student 11**
**File**: morning, afternoon, evening
**Quiz**: morning, afternoon

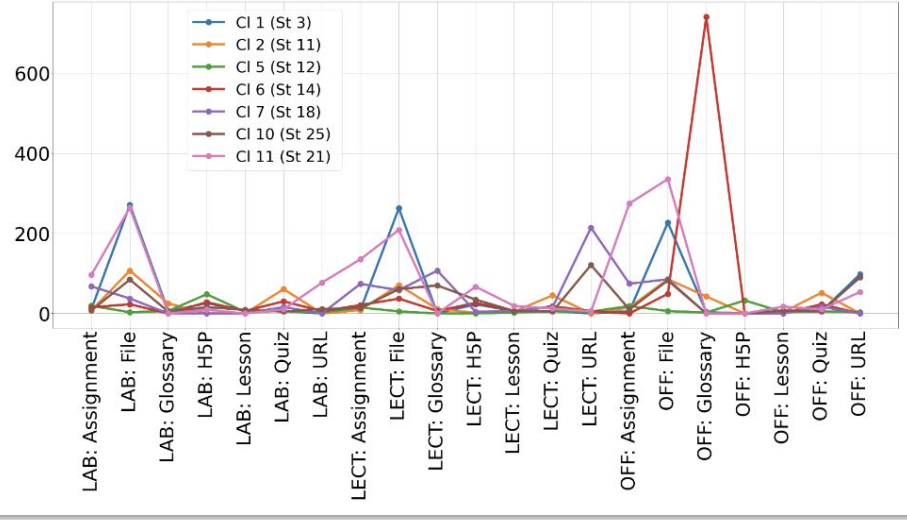
**Student 18**
**Glossary**: morning
**URL, Assignment**: afternoon

# Domain knowledge

## Clusters




## Singletons




Course B - Ward linkage

**Cluster 3**
**File, URL**: laboratory, lecture
**File, Assignment**: day off

**Cluster 8**
**H5P**: laboratory, lecture
**File**: day off

**Student 11**
**File, Quiz**: laboratory, lecture, day off

**Student 18**
**Assignment, Glossary, URL**: lecture

# Clusters depend on variables?

- Selection of variables:
  - subjective process
  - overlooked variables
  - no relevance to the study
  - results: similar or dissimilar behaviours

- Group dynamics: how students migrate across clusters?
  - similarity (or uniqueness) regardless of the profile configuration

**Rotelli**, D., Monreale, A., & Guidotti, R. (2022, September). Uncovering Student Temporal Learning Patterns. *In Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption: 17th European Conference on Technology Enhanced Learning, EC-TEL 2022, Toulouse, France, September 12–16, 2022, Proceedings (pp. 340-353).* Cham: Springer International Publishing.

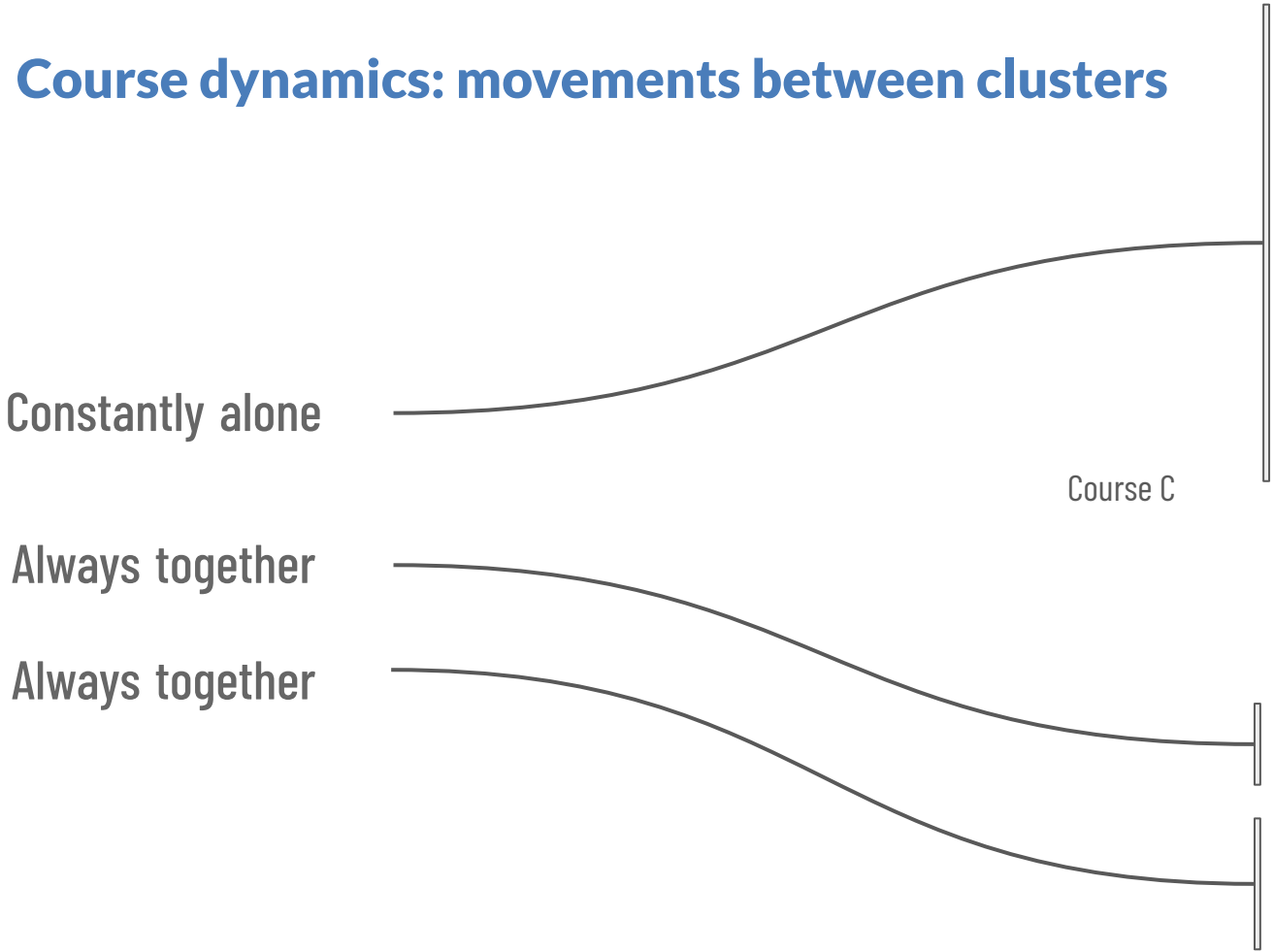
# Clusters depend on variables?

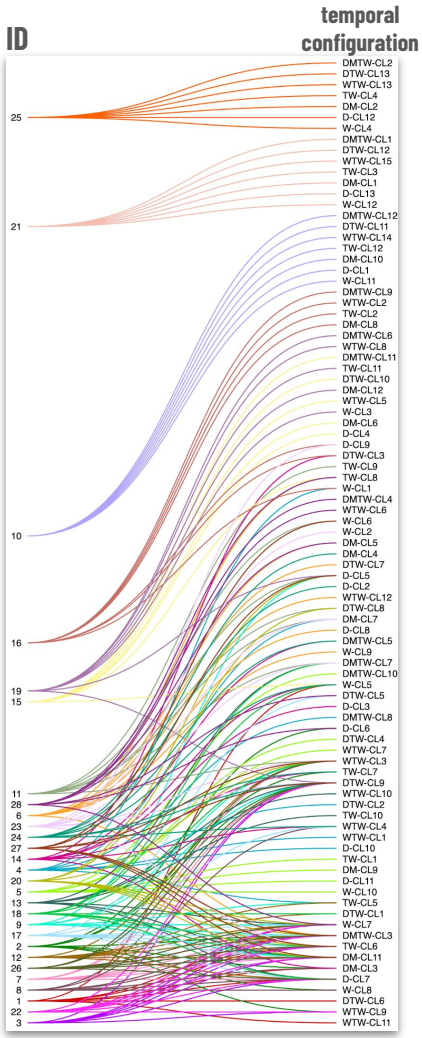
- Selection of variables:
  - subjective process
  - overlooked variables
  - no relevance to the study
  - results: similar or dissimilar behaviours

- Group dynamics: how students migrate across clusters?
  - similarity (or uniqueness) regardless of the profile configuration

**Rotelli**, D., Monreale, A., & Guidotti, R. (2022, September). Uncovering Student Temporal Learning Patterns. *In Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption: 17th European Conference on Technology Enhanced Learning, EC-TEL 2022, Toulouse, France, September 12–16, 2022, Proceedings (pp. 340-353).* Cham: Springer International Publishing.

# Course dynamics: movements between clusters



# Who can benefit from an understanding of student temporal learning patterns?

Source of images: pch.vector on Freepik

# Who can benefit from an understanding of student temporal learning patterns?

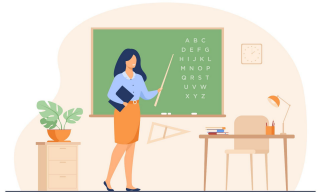
## Students

Dashboard: learning how to **manage and organise their time** to improve their performance avoiding drop-out

Source of images: pch.vector on Freepik

# Who can benefit from an understanding of student temporal learning patterns?





### Students

Dashboard: learning how to **manage and organise their time** to improve their performance avoiding drop-out

### Teachers

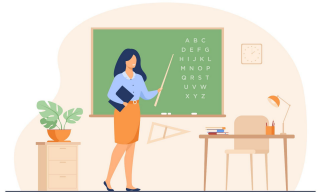
Understanding student regularities to better fit students needs and enable **different teaching strategies** tailored to each learner characteristics.

Source of images: pch.vector on Freepik

# Who can benefit from an understanding of student temporal learning patterns?

### Students

Dashboard: learning how to **manage and organise their time** to improve their performance avoiding drop-out



### Teachers

Understanding student regularities to better fit students needs and enable **different teaching strategies** tailored to each learner characteristics.

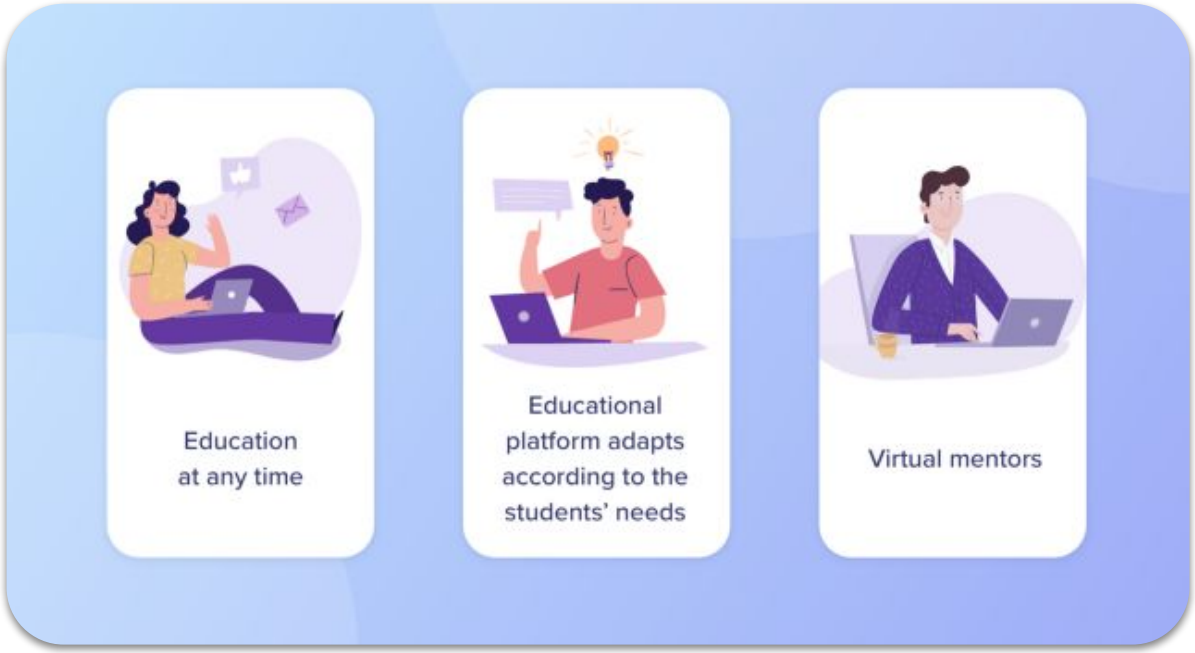
### Researchers

**Analysing learning** to develop and evaluate **insightful** models of student learning strategies

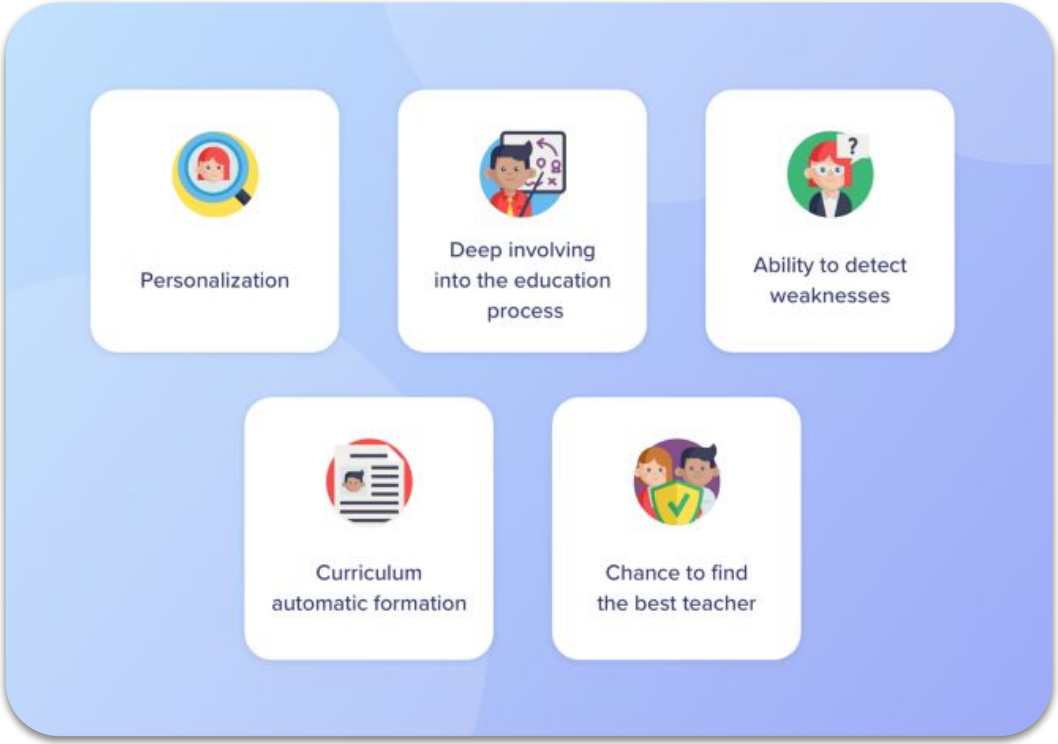
Source of images: pch.vector on Freepik

# AI in Education

# AI in Education: Benefits for students

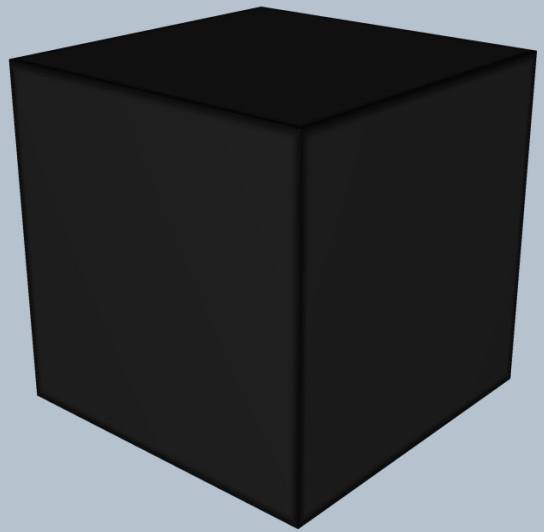


# AI in Education: Benefits for schools and teachers



# AI in Education: Ethics

# AI in Education: Ethics

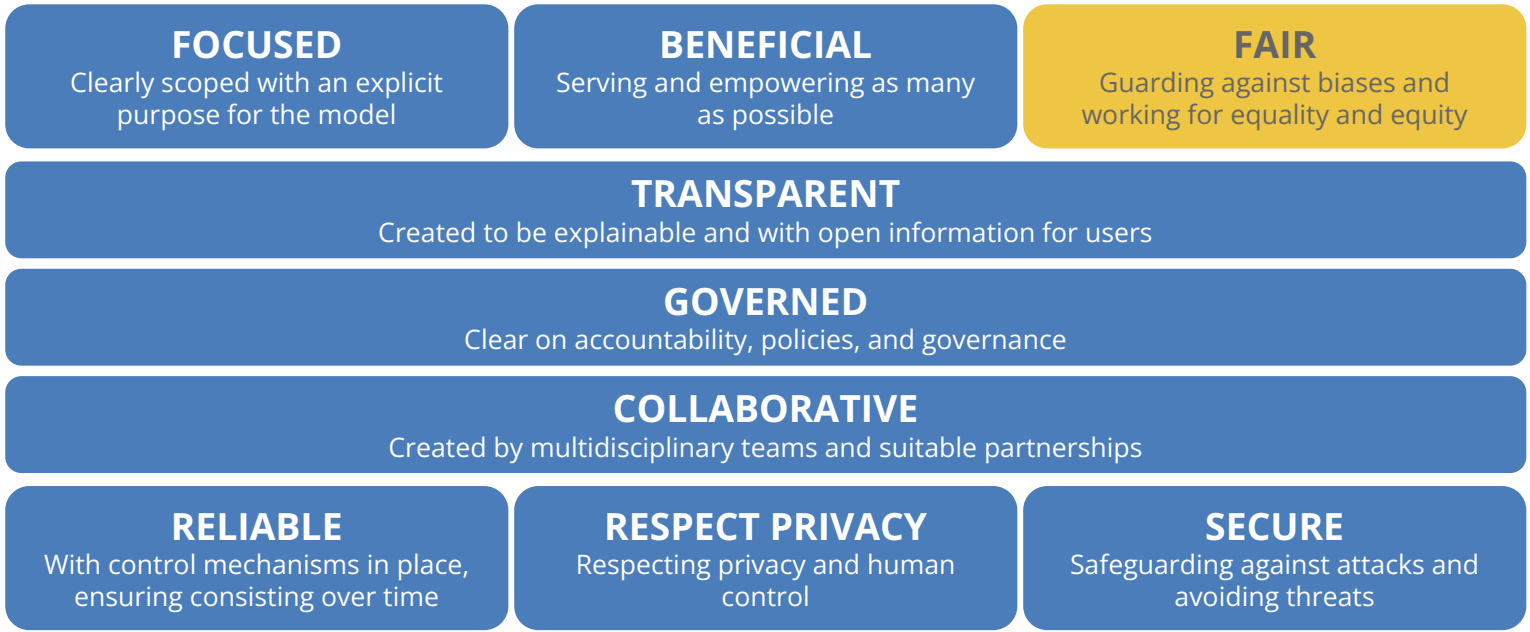


- Opaque algorithms
- Biased datasets
- Incorrectly predicting

Unesco AI in Education [https://en.unesco.org/artificial-intelligence/education]

# Responsible Knowledge Discovery in Education

# Educational knowledge discovery must be trustworthy




*UNICEF https://www.unicef.org/globalinsight/media/2196/file*

# Discrimination is not a general concept

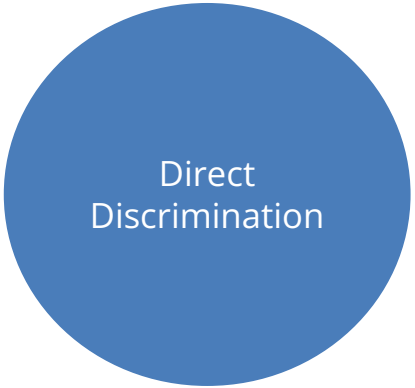
- **It is domain specific**
  - **Credit** (Equal Credit Opportunity Act)
  - **Education** (Civil Rights Act of 1964; Education Amendments of 1972)
  - **Employment** (Civil Rights Act of 1964)
  - **Housing** (Fair Housing Act)
  - **Public Accommodation** (Civil Rights Act of 1964)

- **It is feature specific**

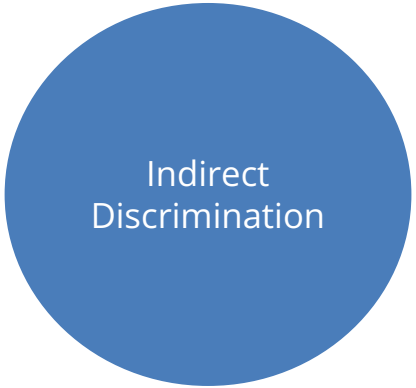
The Civil Rights Act of 1964 prohibits discrimination on the basis of race, color, religion, sex or national origin.

Education Amendments of 1972 (Title IX) prohibits sex (including pregnancy, sexual orientation, and gender identity) discrimination in any education program or activity receiving federal financial assistance.

# Is discrimination direct or indirect?




direct discrimination happens when **protected attributes** of groups or individuals **explicitly** result in **non-favorable outcomes** toward them




individuals **appear to be treated fairly** based on non-protected attributes; however, protected groups get to be treated unjustly as a result of **implicit** effects from protected attributes

Barocas, S., Hardt, M., & Narayanan, A. (2017). Fairness in machine learning. Nips tutorial, 1, 2.

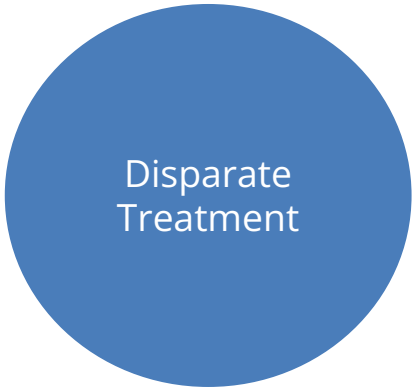
# Granularity of discrimination


when a system gives **unfairly different predictions** to individuals who are considered similar for that task

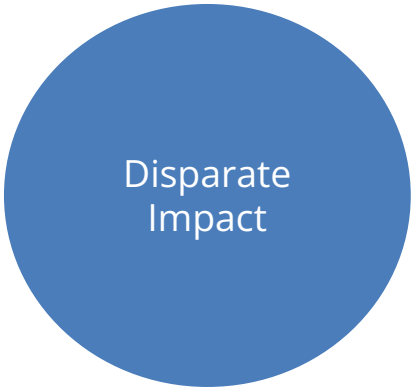

when a system **systematically treats individuals** who belong to different groups **unfairly**

Barocas, S., Hardt, M., & Narayanan, A. (2017). Fairness in machine learning. Nips tutorial, 1, 2.

# Types of disparity in groups




when members of different groups are treated differently




when members of different groups obtain different outcomes


when members of different groups have different error rates

Barocas, S., Hardt, M., & Narayanan, A. (2017). Fairness in machine learning. Nips tutorial, 1, 2.

# Definitions of fairness

### Equalized Odds

an algorithm is fair if the groups have equal rates for true positives and false positives

### Equal Opportunity

an algorithm is fair if the groups have equal true positive rates

### Demographic Parity

an algorithm is fair if the likelihood of a positive outcome should be the same regardless of the group

### Fairness through awareness

an algorithm is fair if it gives similar predictions to similar individuals

### Fairness through unawareness

an algorithm is fair as long as any protected attribute is not explicitly used in the decision-making process

### Equality of Treatment

an algorithm is fair if the ratio of false negatives and false positives is the same for all the groups

Barocas, S., Hardt, M., & Narayanan, A. (2017). Fairness in machine learning. Nips tutorial, 1, 2.

# Techniques for bias treatment

## Pre-processing

before model training

Pre-processing techniques try to **transform the data** so that the bias is mitigated. If the algorithm is allowed to modify the training data, pre-processing can be used

## In-processing

during model training

In-processing techniques try to modify **learning algorithms** to mitigate bias during training process. If it is allowed to change the learning procedure, in-processing can be used

## Post-processing

after model training

Post-processing is performed by **re-ranking items** of the lists obtained after model training. If the algorithm can treat the learned model as a black box, post-processing can be used

# Interesting questions

- Does the study **provide detail on the demographics** of the population involved in the study?
- Does the study describe how **privacy rights** were respected in the data collection process?
- Does the study explicitly **check for algorithmic bias** in some fashion across population(s)?
- Does the study include a discussion about **potential positive and negative impacts**?
- Does the study discuss **ethical concerns** related to the work?
- Does the study discuss the **limitations** of the work?
- Does the study describe the **computational power** needed for training or testing the given models?
- Does the study provide **open-source code and datasets** for the community?

# Challenges #1

1. **Purpose and gain**. It is necessary to make the goals of the educational knowledge discovery initiative *transparent, clarifying* exactly what is going to happen with the information and explicitly what is not.

2. **Representation and actions**. To choose the right environment for the learner's feedback, *the correct visualization technique* to provide recommendations and results to the students.

3. **Data**. A policy needs to be created for educational knowledge discovery that aligns with organization principles. *Transparent communication about where the data are stored*, what is being done to ensure data security and privacy and how the data are evaluated and used.

4. **IT infrastructure**. Efforts should be made from the beginning to search for possible solutions to set up the necessary *internal or external IT infrastructure* and contact and establish connections with relevant people.

5. **Development and operation**. Scalability is maybe one of the most frequently underestimated problems in today's IT industry. A distinction must be made as to whether processes have to be carried out manually, semi-automatically or fully automatically.

6. **Privacy**. All educational knowledge discovery implementations have *to ensure the privacy of the involved parties*. The general lifetime of personal data is a topic that requires further discussion.

7. **Ethics**. Educational knowledge discovery implementers must find a suitable way to *meet high ethical standards* and ensure a beneficial outcome for all stakeholders.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

# Challenges #2

1. **Transferability**: The (learning system) Wall Transfer student model from learning system A to learning system B. Improve an already-good student model in learning system B. Change behavior of learning system B in runnable fashion.

2. **Effectiveness**. Assign students to control or experimental group. Use analytics, only within experimental group, to assign intervention collect longer-term outcome measure. Demonstrate that experimental / analytics-intervention group performs better than experimental / analytics-no-intervention group. But that experimental/analytics-no-intervention group does not perform better.

3. **Interpretability**. Take a complex model of a learning analytics phenomenon. Develop a no-human-in-the-loop method of explaining the model. Present the explanation to five (new) data scientists and users. Ask the participants to explain what decision the model will make, and why, for five case studies. Code the explanations of the model's decisions. Verify if the scientists and users interpret the model the same way.

4. **Applicability**: Knowledge Tracing Beyond the Screen. Take data from at least four students completing learning activity together. Model at least four distinct skills for each student. Predict immediate future performance for these skills.

5. **Generalizability: The General-Purpose Boredom Detector**. Build an automated detector of affect. Demonstrate that the detector works for an entirely new learning system with different interactions and with AUC ROC ≥0.65. Alternatively, build an automated detector for a commonly-seen outcome or measure. *Collect a new population distinct from the original population.* Demonstrate that the detector works for the new population with degradation of quality under 0.1 (AUC ROC, Pearson/Spearman correlation) and remaining better than chance.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, *10*(3), e1355.

# Conferences about educational knowledge discovery

| Title | Acronym | 1° year |
|---|---|---|
| **International Conference on Artificial Intelligence in Education** | AIED (Core A) | 1982 |
| **International Conference on Intelligent Tutoring Systems** | ITS (Core B) | 1988 |
| **IEEE International Conference on Advanced Learning Technologies** | ICALT (Core B) | 2000 |
| **European Conference on Technology-Enhanced Learning** | EC-TEL (Core B) | 2006 |
| **International Conference on Educational Data Mining** | EDM (Core B) | 2008 |
| **ACM Conference on User Modeling, Adaptation, and Personalization** | UMAP (Core B) | 2009 |
| **International Conference on Learning Analytics and Knowledge** | LAK (Core A) | 2011 |
| **Learning at Scale** | L@S (No core) | 2014 |
| **Learning and Students Analytics Conference** | ISAC (No core) | 2017 |

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

# Journals about educational knowledge discovery

| Title | Acronym | Free and open access |
|---|---|---|
| **Journal of Learning Analytics** | JLA | Yes |
| **Computers and Education** | C&S | No |
| **British Journal of Educational Technology** | BJET | No |
| **Journal of Educational Data Mining** | JEDM | Yes |
| **Journal of Artificial Intelligence in Education** | JAIED | No |
| **IEEE Transactions on Learning Technologies** | TLT | No |
| **Journal of Computer Assisted Learning** | JCAL | No |
| **International Journal on Technology Enhanced Learning** | JTEL | No |
| **User Modeling and User-Adapted Interaction** | UMUAI | No |
| **Internet and Higher Education** | I/HE | No |
| **Computer Applications in Engineering Education** | CAEE | No |

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10*(3), e1355.

# Public data sets for educational knowledge discovery

| Title | Description |
|---|---|
| **ASSISTments Competition Dataset** | Competition where data miners can try to predict an important longitudinal outcome using real-world educational data |
| **Canvas Network dataset** | De-identified data from Canvas Network open courses ( January 2014–September 2015), along with related documentation |
| **DataShop** | DataShop provides a central repository to secure and store research ITS data and set of analysis and reporting tools |
| **Educational Process Mining Dataset** | Students' logs during sessions over a simulation environment in digital electronics |
| **HarvardX-MITx dataset** | De-identified data from the first year of MITx and HarvardX MOOC courses on the edX along with related documentation |
| **KDD Cup 2010 Dataset** | Challenge to predict student performance on mathematical problems from logs of student interaction with ITS |
| **Learn Moodle dataset** | Anonymized data from the “Teaching with Moodle August 2016” course from learn. moodle.net |
| **MOOC-Ed Dataset** | Communications taking place between learners in two offerings of the Massively Open Online Course for Educators |
| **NAEP Data Mining Competition 2019** | Competition for measuring students' test activities, and helps develop and test evaluation methods for educational analysis |
| **NUS Multisensor Present. Dataset** | It contains real-world presentations recorded in a multisensor environment. |
| **OU Learning Analytics Dataset** | It contains data about courses, students and their interactions with Moodle for seven selected courses |
| **xAPI-Educational Mining Dataset** | Students' Academic Performance Dataset collected from e-learning system called Kalboard 360 |

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, *10*(3), e1355.

# References


1. Baker, R. S., & Hawn, A. (2021). Algorithmic bias in education. International Journal of Artificial Intelligence in Education, 1-41.
2. Barocas, S., Hardt, M., & Narayanan, A. (2017). Fairness in machine learning. Nips tutorial, 1, 2.
3. Fenu, G., Galici, R., & Marras, M. (2022). Experts' view on challenges and needs for fairness in artificial intelligence for education. In AIED 2022, Springer.
4. Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., ... & Gašević, D. (2022). Explainable artificial intelligence in education. C & E: AI, 3, 100074.
5. Kizilcec, R. F., & Lee, H. (2020). Algorithmic fairness in education. arXiv preprint arXiv:2007.05443.
6. Memarian, B., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI), and higher education: A systematic review. C & E: AI, 100152.
7. Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. Frontiers in Big Data, 2, 13.
8. Rachha, A., & Seyam, M. (2023). Explainable AI In Education: Current Trends, Challenges, And Opportunities. SoutheastCon 2023, 232-239.
9. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley interdisciplinary reviews: Data mining and knowledge discovery, 10(3), e1355.
10. Rotelli, D., & Monreale, A. (2023). Processing and Understanding Moodle Log Data and Their Temporal Dimension. Journal of Learning Analytics, 1-23.
11. Rotelli, D., (2023). A Data Science Perspective on Online Student Temporal Learning Patterns and Dynamics.

# “The Student is Not Like Me”

*Kenneth R. Koedinger*

**Daniela Rotelli**


Department of Computer Science
University of Pisa
Pisa, Italy