Data, Analytics, GenAI, Standards

FACULTY OF ARTS & SOCIETY

Associate Professor Jon Mason

jon.mason@cdu.edu.au

July 2024
Charles Darwin University acknowledges all First Nations people across the lands on which we live and work, and we pay our respects to Elders both past and present.
Overview – the Request

• Research involving Learning Analytics
• Learning Analytics at my university, region or country
• Generative AI and Learning Analytics
• Standardization of Learning Analytics

... in 15 minutes
Assumptions

• Viewers of this webinar already know a lot about the theory & practice of Learning Analytics
  ➔ what perspective can I offer?

• Worldwide, academic communities are confronting the challenges & opportunities provided by GenAI
  ➔ a paradigm shift in education is underway.
Questions as data: illuminating the potential of learning analytics through questioning an emergent field

Jon Mason, Weiqin Chen & Tore Hoel


Engagement?

Student Activity - Grade Matrix

inactive & higher grade  |  active & higher grade
inactive & lower grade  |  active & lower grade

Grade

Activity

INTERACTIONS  GRADE  MATRIX

1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16
Learner Data in LearnLine

Analytics for Learn – student demographics, access, activity, performance and more

Assessment analytics – student assessment & submission, start to finish time, date last accessed

Class Performance – Compares student unit activity related to grades

Discussion Board Analysis – information about student use, engagement, & text analysis

Feedback Fruits – peer feedback about student progress, comments, feedback, and reviews

Kaltura video analytics – video plays, minutes viewed, average completion rate and drop off rate, most viewed, etc

Test Question analysis – statistics on overall performance, assessment quality, individual questions

Turnitin Analytics – insights into how students engaged with assignments, submission timelines, class similarity scores, number of submissions etc
A Contemporary Technology Enhanced Learning (TEL) Ecology

- Academic Integrity & Writing Tools
- Social Media Platforms
- Data
- Learning Management System
- Productivity & Communication Tools
- Web Apps & Educational Tools
- Data
- Virtual Conferencing / Classroom Tools
- Content Management Systems
- Data
- ePortfolio & Evidencing Rendering Tools
- Lecture Capture & Media Streaming
- Data
- Professional Identity Platforms
- 3rd Party Content & Streaming Services
- Open Education Resources (OER)

@michael_sankey
Overview – my Research Agenda

- Data / Data Literacy
- Analytics
- GenAI
- Standards
- Engagement
- SoTL
- Digital Frontiers
- Sense-Making

Questioning with & within the Digital Environment
Overview – my Research Agenda

- Data / Data Literacy
- Analytics
- GenAI
- Standards
- Engagement
- SoTL
- Digital Frontiers

Pre-Literate, Multi-Literate, Post-Literate: Expanding Horizons for LLMs

Jon Mason¹, Carla C. Eisenberg¹, Cat Kutay¹, Sarah Sutcliffe¹, Joanne Forrest¹, Bev Babbage¹

https://aied2024.cesar.school/
IEEE approved standard (IEEE 9274.1.1-2023)

https://adl.net.gov/projects/xapi/
A customizable personal digital record for all learning experiences.

https://www.1edtech.org/initiatives/digital-credentials/clr
terminology
[tur-muh-nol-uh-jee]
noun, plural ‘terminologies’
1. the system of terms belonging or peculiar to a science, art, or specialized subject; nomenclature.
2. the science of terms, as in particular sciences or arts.

Word Origin and History for ‘terminology’
1. from German Terminologie (1786), a hybrid coined by C.G. Schütz of Jena, from Medieval Latin terminologia, from Latin terminus (stem termin-), terminus (q.v.) + -ologia (see -logy).
Recent Terminology ...

- Artificial Communication / Cognition
- Artificial Emotional Intelligence
- Digital Ecosystem
- Conversational Agents
- Prompt Engineering
- Learning Fabric
- Learning Experience Design
- Cybersecurity Mesh
- GenAI
- Data Observability
- Hybrid Learning Spaces
- Post-Literate
- Neuroadaptive Learning
- Metaverse
- Web 3.0
- Decentralised Identity
International Standards

ISO/IEC TR
20748-1
Information technology for learning, education and training -- Learning analytics interoperability
— Part 1: Reference model
Published (2016)

ISO/IEC TR
20748-2
Information technology for learning, education and training -- Learning analytics interoperability
— Part 2: System requirements
Published (2017)

ISO/IEC TS
20748-3
Information technology for learning, education and training — Learning analytics interoperability
— Part 3: Guidelines for data interoperability
Published (2020)

ISO/IEC TS
20748-4
Information technology for learning, education and training — Learning analytics interoperability
— Part 4: Privacy and data protection policies
Published (2019)
6.4 Data collection

Data collection is the process of measuring and gathering information on matters of interest from learning and teaching activities. Tracking data from learners emanate from a wide variety of platforms, e.g., when accessing learning material, using desktop computers and mobile devices, including wearable technologies and the Internet of Things. In this process, the requirements related to the data authority, control of data source, interoperability of data, and efficient flow and exchange of data are addressed.

6.4.1 Accessibility

<table>
<thead>
<tr>
<th>Requirement ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R4.1.1</td>
<td>Accessibility requirements <em>should</em> be registered prior to data collection.</td>
</tr>
</tbody>
</table>

6.4.2 Aggregation/Integration of data

<table>
<thead>
<tr>
<th>Requirement ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R4.2</td>
<td>To access heterogeneous learning systems or tools, aggregated profiles for the user <em>should</em> be supported.</td>
</tr>
</tbody>
</table>

6.4.3 Data interoperability

<table>
<thead>
<tr>
<th>Requirement ID</th>
<th>Description</th>
</tr>
</thead>
</table>
| R4.3.1         | To improve accuracy of collected data, standardized information model and controlled vocabularies *should* be applied to data collection API.  
*EXAMPLE 1:* recipe for xAPI specification is an example to define activity stream type and vocabularies.  
*EXAMPLE 2:* IMS Caliper Metric Profile is an example to define learning activity types and vocabularies. |
| R4.3.2         | Data information model *should* cover a wide range of data types.  
*EXAMPLE:* xAPI is an example to define statement and IMS Caliper is an example for metric profile in terms of information model. |
| R4.3.3         | Data information model *should not* be dependent on dominant products or services.  
Note: xAPI and IMS Caliper are good example for open specification about information model. Institution can use both specifications to develop its profile in terms of specific |
## Use cases on advanced learning analytics services using emerging technologies

### Use-Case Template

<table>
<thead>
<tr>
<th>ID:</th>
<th>Unique ID of this use case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Enter the goal of the use case - preferably as a short, active verb phrase</td>
</tr>
<tr>
<td>Description:</td>
<td>Describe the goal and context of this use case. This is usually an expanded version of what you entered in the &quot;Title&quot; field</td>
</tr>
<tr>
<td>Primary Actor:</td>
<td>A person or a software/hardware system that interacts with your system to achieve the goal of this use case</td>
</tr>
<tr>
<td>Preconditions:</td>
<td>Describe the state the system is in before the first event in this use case</td>
</tr>
<tr>
<td>Postconditions:</td>
<td>Describe the state the system is in after all the events in this use case have taken place</td>
</tr>
<tr>
<td>Main Success Scenario:</td>
<td>Describe the flow of events from preconditions to postconditions, when nothing goes wrong. This is the meat of the use case</td>
</tr>
<tr>
<td>Extensions:</td>
<td>Describe all the other scenarios for this use case - including exceptions and error cases</td>
</tr>
<tr>
<td>Frequency of Use:</td>
<td>How often will this use case be used</td>
</tr>
<tr>
<td>Status:</td>
<td>Development status</td>
</tr>
<tr>
<td>Owner:</td>
<td>Who owns this use case, in your project team</td>
</tr>
<tr>
<td>Priority:</td>
<td>Priority of this use case</td>
</tr>
</tbody>
</table>
relationship among elementary, middle, and high school, by mapping to achievement statements and each level of topics/units.

Figure 3 Knowledge space tracing part on the AI model for math Ed

Technologies / Technical Approaches and Applications

Prediction and recommendation technologies

Prediction technology on maths education is mainly used to recommend the next assessment item by calculating the probability that a learner will answer the correct answer to unsolved questions. The prediction AI model usually uses an RNN (recurrent neural network) based deep learning model. Recently, many research are underway to improve prediction accuracy by utilizing as known as the transformer model. This use case shows the accuracy is little bit over 85%, and more than 90% accuracy is recommended for reliable service.

In detail, four types of datasets – (a) assessment item annotation (IRT values regarding assessment item), (b) learner level annotation (IRT values regarding assessment), (c) node on the knowledge space for each assessment item, (d) correct and incorrect answer table of learners - are used for AI training. To improve accuracy for the prediction probability through the data processing module both skill index sequence by learner and correct answer sequence for each learner are generated from the input datasets (a) and (b). The training process of the AI predictive model consists of three steps. First, predict the probability of correct answers to assessment items the learner has not yet solved through the AI prediction model for correct answers for each item. Then, the loss is calculated by comparing the predicted value with the actual result of the correct and incorrect answer table, and then the AI training process is performed in the direction of minimizing the loss function.

Figure 4 AI training process of prediction function regarding correct answer for math

In this case, the process of tracing the learner's weak knowledge consists of an iterative process. If an assessment item with a low probability of correct answer via the AI prediction model is actually found as an true result, the AI diagnostic model
(some) Questions Arising

- The data might tell a story – but is storytelling a science?
- What data might be missing?
- Can we trust the data?
- How should consent be managed?
- Who ‘owns’ the data?
- Are educational institutions acting ethically in their collection & use of student data?
- Who determines the boundaries of personal & institutional data?
- What are the ethical considerations?
- What kind of data governance model best suits learning analytics?
- How can we balance compliance within a local jurisdiction & an international context?
- Self-directed learning involves a blend of formal & informal learning – how can we meaningfully capture data from both domains?
THE BASIC STEPS OF A PIA

GATHER
Gather all the information you need to do the PIA.

RISK
Identify any real privacy risks and how to mitigate them.

TAKE ACTION
Take action.

CHECK
Check against the privacy principles.

PRODUCE
Produce a PIA (use our report template to help).

REVIEW
Review and adjust the PIA as necessary as the project develops.

What are the disciplinary trajectories of Data Science & Artificial Intelligence?

- Natural sciences
- Social sciences
- Life sciences
- Applied Sciences
  - Technology
  - Engineering
  - Industrial

- Physics
- Chemistry
- Information & Computing Sciences
- Engineering
- Social Sciences
Data Governance

- The origins and destinations of data flows are complex
- Humans are also agents (‘tools’) of data production – data that is monetized by Big Tech
- ‘Ownership’ of personal data is contestable & this is why privacy laws like GDPR exist
- A data governance framework requires specification of rules, roles, & responsibilities ... and policies to ensure privacy, security, & data quality are managed safely, ethically & responsibly.
https://thesociologicalreview.org/magazine/june-2023/artificial-intelligence/the-search-continues/
Wrapping up ...

• Digital ‘ecosystems’ increasingly complex
• The digital environment is more than a collection of tools
• Data does not ‘speak for itself’
  • instruments that produce it require scrutiny
• Text production is now trivial
  • & so is data production – raising validity concerns around existing LA systems
• A paradigm shift is underway in Education
  • Dialogic Web – moving beyond ‘search’
  • Post-literate
  • Human-Computer Collaboration
  • Co-production of knowledge
  • New conceptual frameworks needed

jon.mason@cdu.edu.au