Embracing Imperfection in Learning Analytics

Kirsty Kitto, Simon Buckingham Shum & Andrew Gibson
UTS Connected Intelligence Centre
@KirstyKitto • kirsty.kitto@uts.edu.au
@sbucketshum • @AndrewResearch
we are recruiting students!

https://utscic.edu.au/research/phd/
the overture
how do you evaluate LA?
how do you evaluate LA?

<table>
<thead>
<tr>
<th>LEARNERS</th>
<th>TEACHERS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATA</strong></td>
<td><strong>DATA</strong></td>
</tr>
</tbody>
</table>
| - For this LA tool it is clear what data is being collected  
- For this LA tool it is clear why the data is being collected |
| **AWARENESS & REFLECTION** | **AWARENESS & REFLECTION** |
| - This LA tool makes me aware of my current learning situation  
- This LA tool makes me forecast my possible future learning situation given my (un)changed behaviour  
- This LA tool stimulates me to reflect on my past learning behaviour  
- This LA tool stimulates me to adapt my learning behaviour if necessary |
| - This LA tool makes me aware of my students' current learning situation  
- This LA tool makes me forecast my students’ possible future learning situation given their (un)changed behaviour  
- This LA tool stimulates me to reflect on my past teaching behaviour  
- This LA tool stimulates me to adapt my teaching behaviour if necessary |
| **IMPACT** | **IMPACT** |
| - This LA tool stimulates me to study more efficiently  
- This LA tool stimulates me to study more effectively |
| - This LA tool stimulates me to teach more efficiently  
- This LA tool stimulates me to teach more effectively |


does LA help learning?

While EDM aims to improve learning outcomes, its “emphasis on the ‘educational’ aspect of educational data mining has been scarce. . . One reason for this is the inclination of researchers to evaluate EDM research primarily for model fits and predictive accuracy rather than for plausibility, interpretability, and generalizable insights.”
does LA help learning?

While EDM aims to improve learning outcomes, its “emphasis on the ‘educational’ aspect of educational data mining has been scarce. . . One reason for this is the inclination of researchers to evaluate EDM research primarily for model fits and predictive accuracy rather than for plausibility, interpretability, and generalizable insights.”

a question: is it easier to evaluate LA in some paradigms than others?

if so, what are the implications for the field?
but what are we evaluating here?

student facing learning analytics
but what type of student facing LA are we talking about?

Are students acquiring: content and skills? or learning to learn?
We should not make nanobots for multiple reasons. As you probably know, in the wrong hands they can be dangerous. So to fined out the rest you are going to have to read the rest of this existing article.

For one, a nanobot could have a bug and start eating anything cardin based or just not work at all. Another thing is that they may also eat the wrong substances, which would only be bad in some cases. What is rile bad if one has a bug it could make mor with the same problem. Now I know that you are wondering what I am talking about, I mean how could it make more of its problem unless it could rewrite other nanobots programs. Well some scientists are trying to figure out how to make it possible for them to copy themselves. So one might be able to become 100.

Also they are planning to make them abule to cile bakteryia, and there they might eat away at the intestens instead. But don't be worryd they might make it so that they will go throw the body with the rest of th fool. Also they might program them to tern of after a certain amount of time.

They are also planning to make small tracking devices so kids won't get lost. I just hope they are hacker safe and they aren't over used. I don't want the goverment to know to much. I also don't want some sikeco thraking me.

So as you can see there are lots of problems. There is bugs, hackers, goverment overuyuos, and falling into the rong hands. There is good noos I think we are stile alitaule fare frome geting a lot of nanobots just yet.
Should the distributed intelligence of the whole system’s performance (humans + technology) be the output measure?

Or, should we also be concerned with the effects on human performance when stripped of the technology?

Learning to learn

“equipping students with knowledge, skills, and dispositions that prepare them for lifelong learning, in a complex and uncertain world”

“Creativity, critical thinking, agency, curiosity, and an ability to tolerate uncertainty…”

Arguably the purpose of analytics-powered pedagogy in such contexts is to provoke productive reflection on one’s strengths and weaknesses — these are higher order competencies, into which a machine can have limited insight.

authentic learning: vital but challenging for LA

wicked problems: how do we provide LA when there is no correct answer?

transformed perspective: the sense that a learner makes of their experience, or a shift in worldview, which by definition is not accessible to the machine, but to which a machine might have partial access

socially and psychologically complex performance: scenarios where the outcome is emergent in nature, a function of many drivers that result in unpredictable and/or unique outcomes, often because social interaction is central to the process

analytics in such contexts will in principle have a high degree of imperfection!
Act 1: perfection is not possible
a cautionary tale from information retrieval

are we measuring what we value?

or merely valuing what we can measure?

a thought experiment: is a perfect classifier desirable in education?
cognitive presence

“extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication.”


https://plus.google.com/u/0/+StefanPSchmid/posts/4wrUbFzFwpJ
we can use machine learning to classify discussion forum text using this construct.

should we use it with students yet?
how accurate does it have to be?

data was unbalanced (solved using boosting)
- is it overfitted for one “type” of learning scenario?
- how accurate will it be if used in another context?
- how different does a situation have to be before we retrain?

how are we going to use it?
- who sees the classifications?
- what happens if the classifier is wrong?
well we have already... and it wasn’t even the state of the art classifier...
well we have already... and it wasn’t even the state of the art classifier...

should we worry?
In Week 2 I was very aspirational about the role I wanted to play; ‘I would like my profile to be professional, respectful, organised, connected and visible. I aim to be an active participant within “reflection and critical discourse that is the core dynamic of a community of inquiry”’. I achieved my aim of being an active participant as I made over 75 comments on my peers’ posts, averaging over 5 per week. However I feel I did not participate fully in all 4 phases of the cognitive presence in the Practical [sic] Inquiry Model; triggering event, exploration, integration and resolution – despite having sentence openers taped next to my computer! Triggering events and some exploration were met by sharing an interesting article relevant to a post I had read and also asking some questions, but I felt a lot of my posts were agreeing with and complimenting upon the erudite musings of my peers. I was definitely wary of confronting differing ideas and promoting a critical discourse. This participation in all cognitive phases needs improving so the sentence openers will remain up! [score=4]
Act 2:

perfection is not desirable
the Navajo rug

In a Navajo rug there is always an imperfection woven into the corner. And interestingly enough, it's where “the Spirit moves in and out of the rug.” The pattern is perfect and then there's one part of it that clearly looks like a mistake …

Perfection is not the elimination of imperfection. That's our Western either/or, need-to-control thinking. Perfection, rather, is the ability to incorporate imperfection!

_Breathing Under Water: Spirituality and the 12 Steps_, by Richard Rohr
the Navajo rug

In a Navajo rug there is always an imperfection woven into the corner. And interestingly enough, it's where “the Spirit moves in and out of the rug.” The pattern is perfect and then there's one part of it that clearly looks like a mistake ...

Perfection is not the elimination of imperfection. That's our Western either/or, need-to-control thinking. Perfection, rather, is the ability to incorporate imperfection!

Breathing Under Water: Spirituality and the 12 Steps, by Richard Rohr
active learning squared (AL$^2$)

the student trains the classifier...

...while it is training the student...

Community of Inquiry Classification

Want to learn about your participation within your learning community?

When you start this activity, you will see one of your posts. We have used machine learning to categorise your cognitive presence according the Community of Inquiry model.

However, our machine learning tool is still learning and it could be wrong. We would like you to:

1. Think about how your post was classified
2. Choose what category you believe your post belongs to
3. If you like, you may highlight text from your post that you used in making your decision, or add remarks to the text-box about what helped you come to your conclusion
4. You can view your history below

What is Cognitive Presence?

Cognitive presence has four phases: Triggering, Exploration, Integration, and Resolution.

**Triggering Phase** initiates discussion about a particular issue/topic for inquiry.
**Exploration Phase** posts explore the issue at hand by exchanging knowledge between members of the community.
**Integration Phase** interactions build upon the ideas shared and explored in the Exploration phase and begin to construct understanding or a solution about a topic or issue.
**Resolution Phase** are messages in a discussion that test the solutions or understanding developed in the Integration phase.

Begin
Community of Inquiry Classification

Want to learn about your participation within your learning community?

When you start this activity, you will see one of your posts. We have used machine learning to categorise your cognitive presence according the Community of Inquiry model.

However, our machine learning tool is still learning and it could be wrong. We would like you to:

1. Think about how your post was classified
2. Choose what category you believe your post belongs to
3. If you like, you may highlight text from your post that you used in making your decision, or add remarks to the text-box about what helped you come to your conclusion
4. You can view your history below

What is Cognitive Presence?

Cognitive presence has four phases: Triggering, Exploration, Integration, and Resolution.

Triggering Phase initiates discussion about a particular issue/topic for inquiry.
Exploration Phase posts explore the issue at hand by exchanging knowledge between members of the community.
Integration Phase interactions build upon the ideas shared and explored in the Exploration phase and begin to construct understanding or a solution about a topic or issue.
Resolution Phase are messages in a discussion that test the solutions or understanding developed in the Integration phase.

Begin
Community of Inquiry Classification

Was classified as: Triggering

Here's a free definition for your buzzword bingo card

Conspicuous: an approach to defining the levels at which an institution collects in a given content area. It's about the depth of collecting and there are standard indicators, which you can read about in this IFLA guide to collection development policies. Conspicuous is also an approach that can be taken to collection development policy writing, where the policy sets the target level of depth in particular areas of collecting. It's not used much in Australian libraries any more, and is a bit out of fashion internationally (though used by some research libraries still).
see the paper for second example of imperfection: automated formative feedback on reflective writing
cognitive dissonance provides a teachable moment... and other closely related fields have used similar methods already!

embracing imperfection

so imperfection in our LA tools opens up new opportunities

▪ teachable moments
▪ intelligence augmentation
▪ mindful engagement with automated feedback
▪ learning to challenge computational decisions
▪ accelerates presence of more advanced LA in education

but to get to this point we need to ensure that mature LA tools are evaluated holistically!
as machine intelligence reduces, we can increase human agency (and learning) through good LD

“nonautomatic, effortful and thus metacognitively guided processes”

but then how can we evaluate success?
towards comprehensive evaluation for LA

mature student facing LA (that aims to help students learn how to learn) needs to be evaluated across a range of criteria in the paper we explore

1. Learning design
2. Model
3. Feedback
4. Sensemaking/gain
5. Accuracy
applying this to AL$^2$

**Learning design:** this learning design aims to teach (i) data literacy (i.e. that ML can be wrong) and (ii) a basic educational construct

**Model:** dual process model of cognition

**Feedback:** automatic classifications are appended to student comments and presented in a new display

**Sensemaking/gain:** The interface allows the student to (i) change the classification of their post, (ii), highlight components of the post that they feel are indicative of the classification they have chosen, (iii) leave a comment about why they chose that classification.

**Accuracy:** to date - very low in pilot trials (30.2%!)
conclusions

- perfect accuracy in LA is **unlikely to be possible** in a wide range of authentic learning scenarios…
- … nor is it always desirable — embracing imperfection opens up new possibilities for **teachable moments**!
- imperfection is sometimes **a feature not a bug**
Thanks!
Debate!

@kirstykitto
@suckshum
@andrewresearch

Open Access reprint: