

Embracing Imperfection in Learning Analytics

Kirsty Kitto, Simon Buckingham Shum & Andrew Gibson
UTS Connected Intelligence Centre
@KirstyKitto • kirsty.kitto@uts.edu.au
@sbuckshum • @AndrewResearch

we are
recruiting
students!

Navigation

Research

Publications:

- PhD in Learning Analytics

动力 Resources

Wellcome LUTS Context Tools Skills & Dispositions Scholarships & Applications

Welcome to the UTS:CIC Doctoral Program

We are delighted to announce that CIC's doctoral program in Learning Analytics is offering 2 new UTS Scholarships to start Spring 2018 semester (August latest) for domestic students (i.e. who do not require a visa).

CIC's primary mission is to maximise the benefits of analytics for UT's teaching and learning. The Learning Analytics Doctoral Program is part of our strategy to cultivate transdisciplinary innovation to tackle challenges at UTs, through rigorous methodologies, arguments and evidence. A core focus is the personalisation of the learning experience, especially through improved feedback to learners and educators.

As you will see from our work, and the PhD topics advertised, we have a particular interest in analytics techniques to nurture in learners the creative, critical, sensemaking qualities needed for lifelong learning, employment and citizenship in a complex, data-saturated society.

We invite you to apply for a place if you are committed to working in a transdisciplinary team to invent user-centered analytics tools in close partnership with the LIPN staff and students who are our "clients".

Please explore this website so you understand the context in which we work, and the research topics we are supervising. We look forward to hearing your ideas which fit into CIC, and how your background, skills, and aspirations can advance this research.

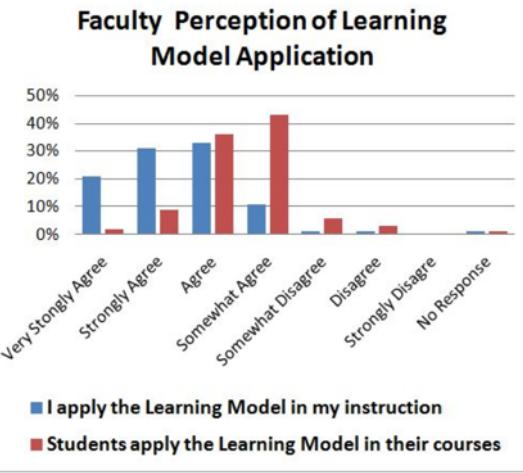
Ph.D.
LEARNING ANALYTICS

<https://utscic.edu.au/research/phd/>

the overture

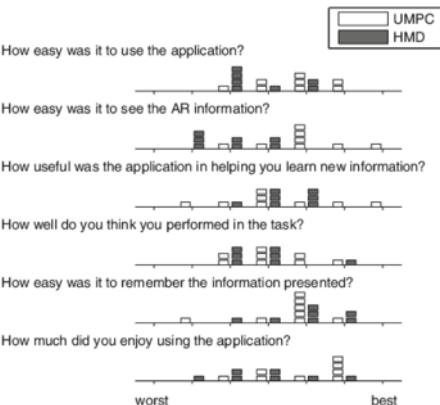
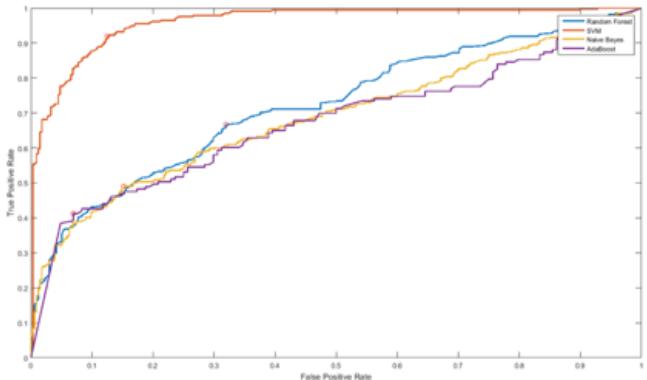


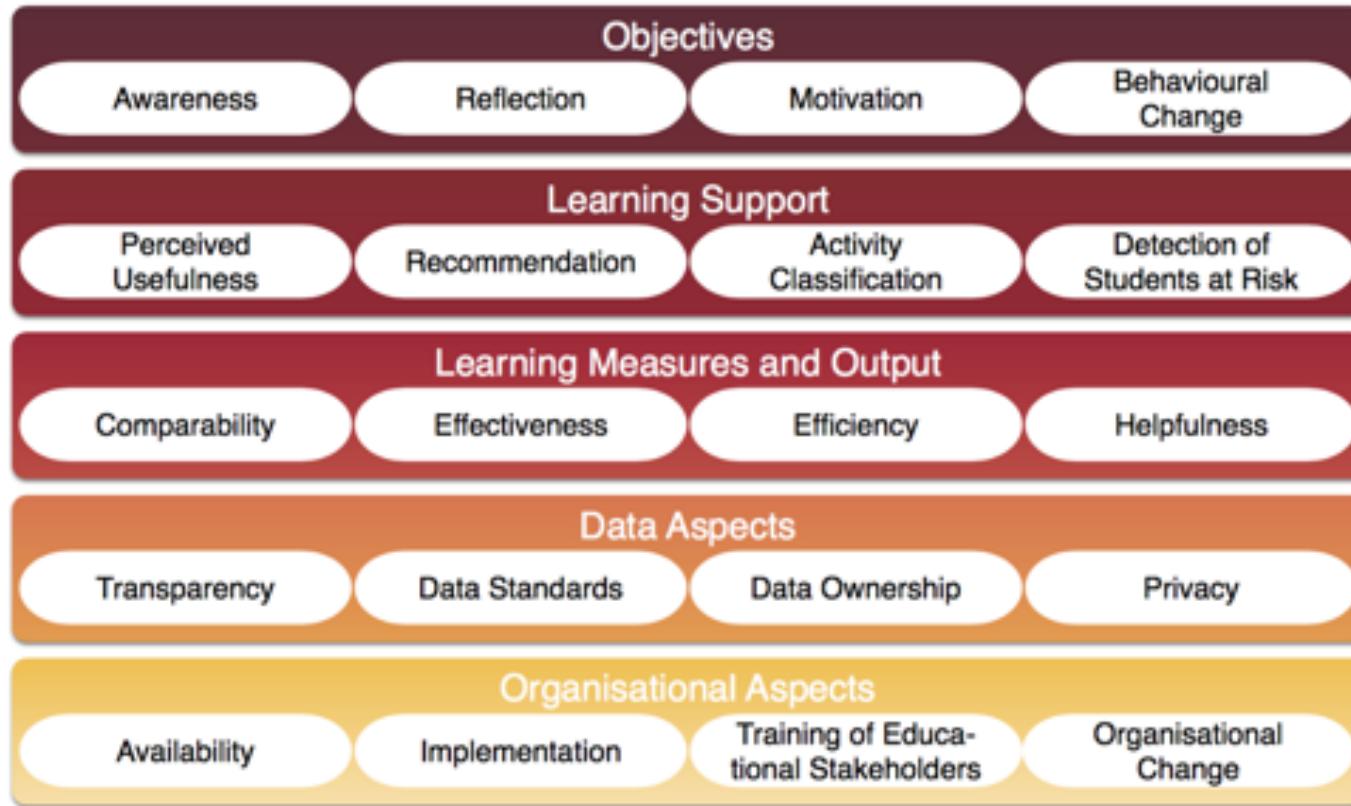
how do you evaluate LA?



Task 1	Task 2	Task 3	Task 4	Task 5
C	D	C	C	C
HD	HD	HD	HD	HD
C	HD	C	D	HD
HD	D	D	HD	HD
C	D	C	HD	HD
C	HD	C	C	C
C	D	D	C	D
HD	HD	HD	HD	HD
D	D	D	HD	HD
HD	HD	HD	HD	HD
HD	D	C	D	HD
D	D	C	C	D
C	HD	HD	D	D
D	HD	C	D	D
D	HD	HD	D	C

how do you evaluate LA?





Scheffel, M., Drachsler, H., Stoyanov, S., Specht, M.: Quality Indicators for Learning Analytics. *Educational Technology & Society* 17(4), 117–132 (2014)

LEARNERS

TEACHERS

DATA

- 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

- 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

- 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

- 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

Scheffel, M., Drachsler, H., Stoyanov, S., Specht, M.: Quality Indicators for Learning Analytics. Educational Technology & Society 17(4), 117–132 (2014)

Scheffel, M. (2017). *The Evaluation Framework for Learning Analytics*.

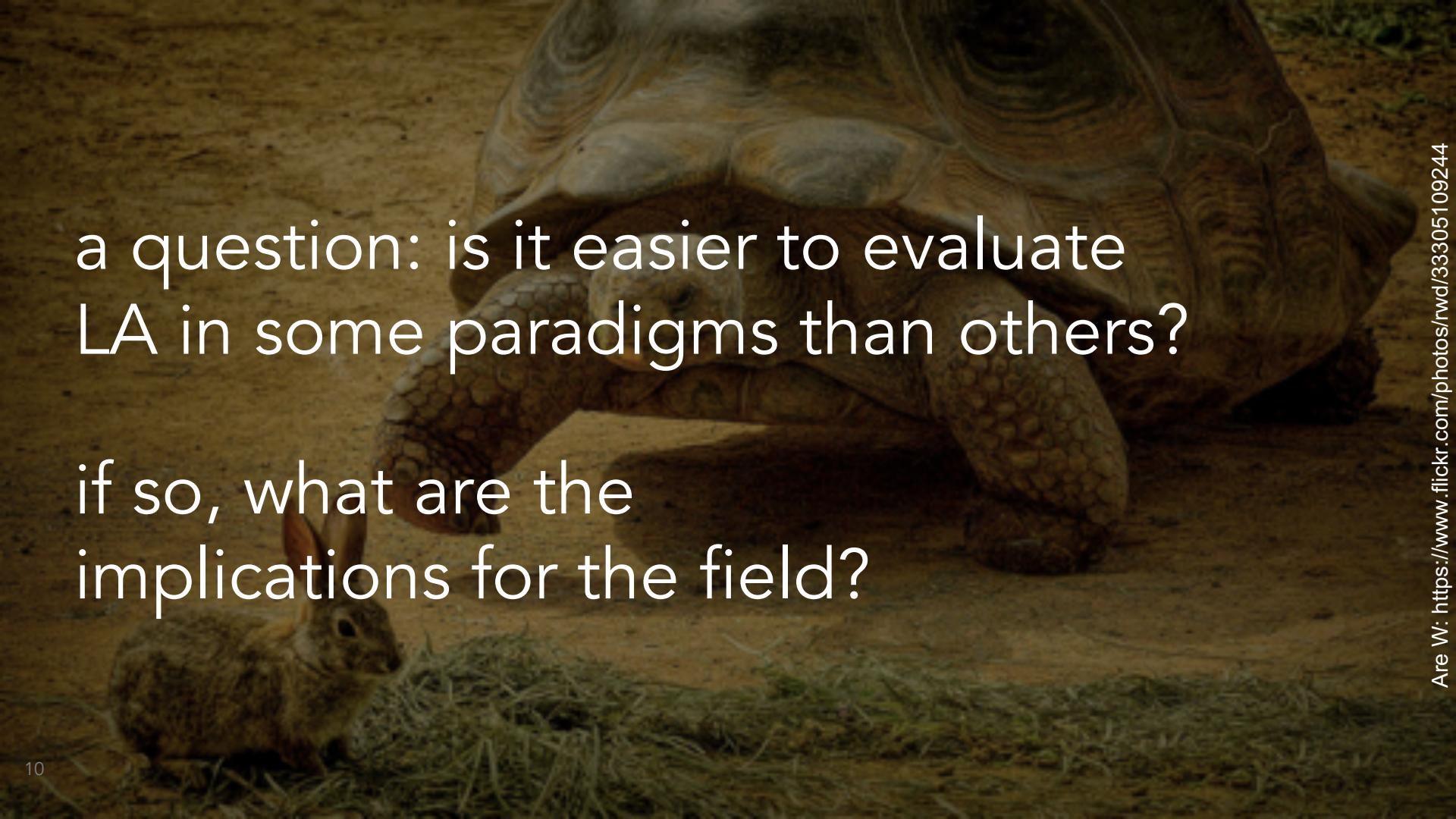
does LA help learning?

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.”

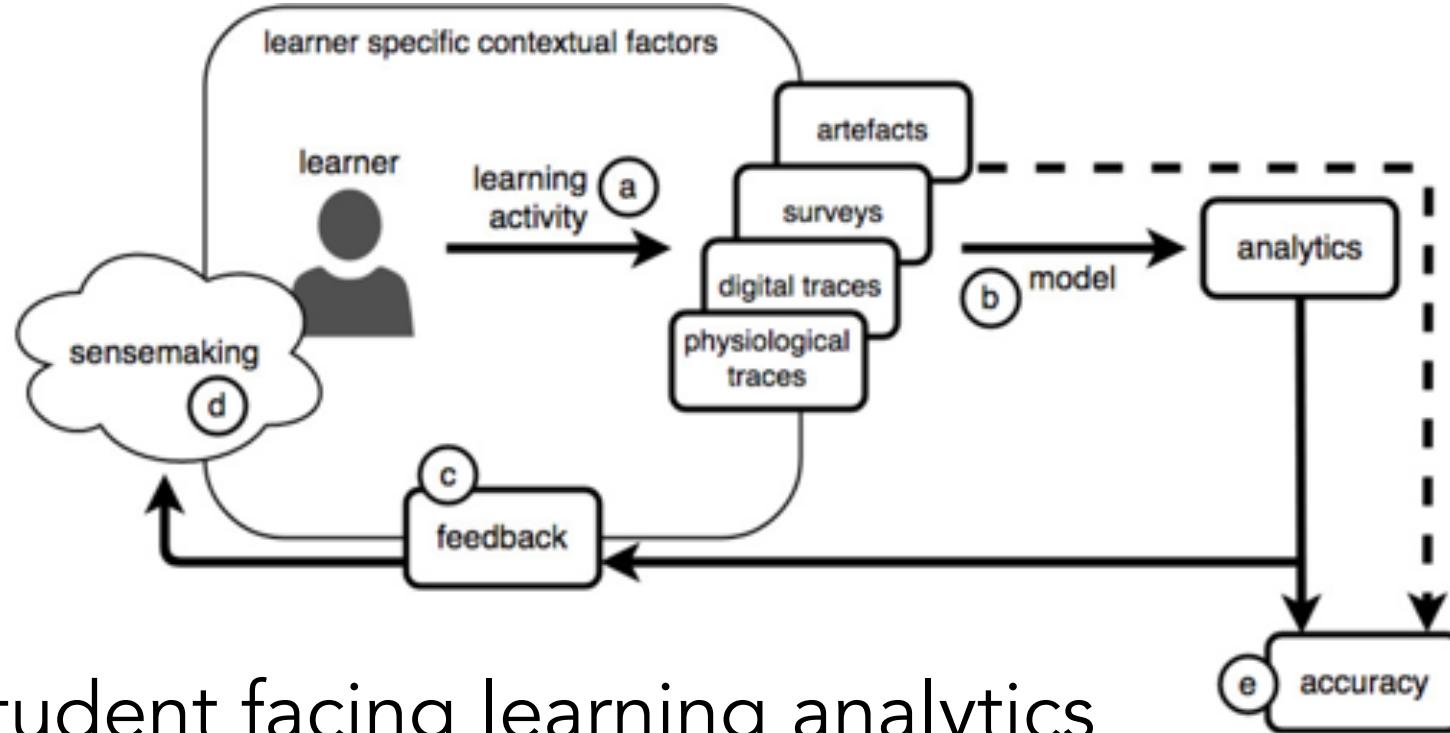
Ran Liu and Kenneth R Koedinger. 2017. Closing the loop: Automated data-driven cognitive model discoveries lead to improved instruction and learning gains. JEDM-Journal of Educational Data Mining 9, 1 (2017), 25–41.

A photograph of a large tortoise and a small rabbit in a field. The tortoise is in the background, its head and front legs visible as it walks across dry, brown grass. In the foreground, the back of a small, light-colored rabbit is visible, also walking through the same dry grass.

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?



should b for pleas reason a wrong
We shold not make nanobots fore multipul reesuns. As you probibly know in the rong hands thay can be dangerous. So to fined out the rest you are going to have to read the rest of this exsithing artikul. cle an

For one a nanobot could have a bug and start eeting enything cardin basted or just not work at all. Another thing is that thay may also eat the rong substins, wich wold onle be bad in some cases. Wat is rile bad if one has a bug it cold make mor with the same problem. Now I know that you are wondering wat I am tolking abot, I mean how could it make mor of its problem inles it colud rerite uther nanobots programs. Well some sientintists are tring to figyer out how to mak it posibul for them to copy themselfs. So one might be able to bekum 100.

Also thay are planing to make them abule to cile bakterya, and there thay might eat away at the intestens insted. But don't be worryd thay mite make it so that thay will go throw the body with the rest of th foot. Also thay might program them to term of after a serten amout of time.

Thay are also planing to make smal traking divises so kids wont get lost. I just hope thay are haker safe and thay 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, hakers, goverment overyuos, and faling into the rong hands. There is good noos I think we are stile alitaule fare frome geting a lot of nanobots just yet. news

We should not make nanobots for multiple reasons. As you probably know in the wrong hands they can be dangerous. So to find 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 carbon based or just not work at all.

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?

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 government to know too much. I also don't want some sicker tracking me.

So as you can see there are lots of problems. There are bugs, hackers, government overzealous, and failing in a lot of nanotech.

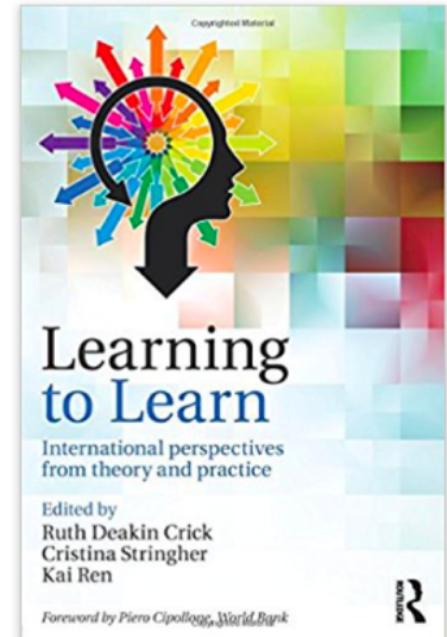
Gavriel Salomon, David N Perkins, and Tamar Globerson. 1991. Partners in cognition: Extending human intelligence with intelligent technologies. Educational researcher 20, 3 (1991), 2–9.

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?

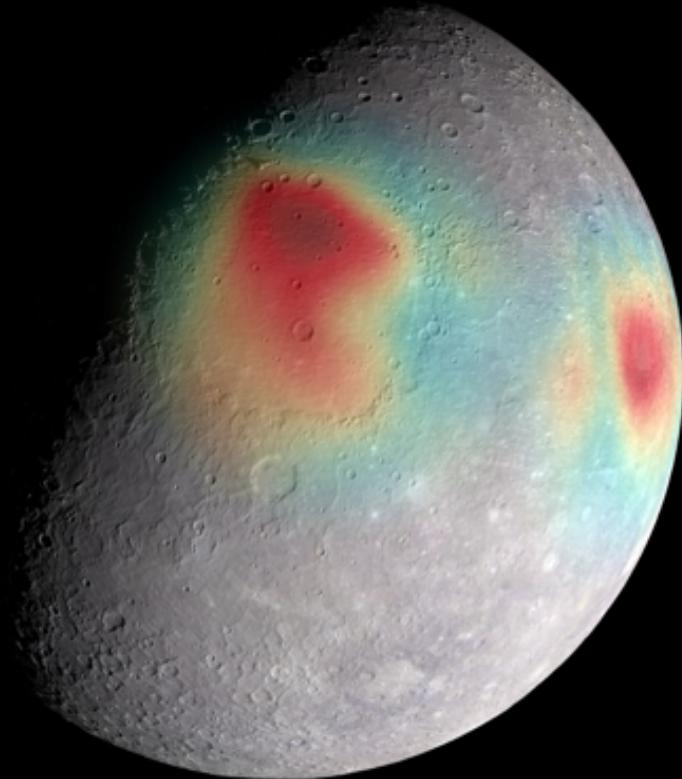
transformed perspective: the sense that a learner makes of their experience, or a shift in worldview, which by definition is not accessible to the machine, but to which a machine might have partial access

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

Turpin, Scholer (2006). User performance versus precision measures for simple search tasks. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 11-18). ACM.

User Performance versus Precision Measures for Simple Search Tasks

Andrew Turpin^{*} Falk Scholer
School of Computer Science & IT
RMIT University, GPO Box 2476V
Melbourne, Australia, 3001.
(aht,scholer)@cs.rmit.edu.au

ABSTRACT

Several recent studies have demonstrated that the type of improvements in information retrieval system effectiveness reported in forums such as SIGIR and TREC do not translate into a benefit for users. Two of the studies used an instance recall task, and a third used a question answering task, so perhaps it is surprising that the precision-based measures of IR system effectiveness on one-shot query evaluation do correlate with user performance on these tasks. In this study, we evaluate two different information retrieval tasks – TREC Web track data: a precision-based user task, measured by the length of time that a user needs to find a single document that is relevant to a TREC topic; and, a simple recall-based task, represented by the total number of relevant documents that users can identify within five minutes. Users employ search engines with controlled mean average precision (MAP) of between 55% and 56%. Our results show that there is no significant relationship between system effectiveness measured by MAP and the precision-based task. A significant, but weak relationship is present for the precision at one document returned metric. A weak relationship is present between MAP and the simple recall-based task.

Categories and Subject Descriptors

I.4 Information Storage and Retrieval; I.3.6 User Interfaces; D.2.8 Software Engineering; Metrics—complexity measures, performance measures

General Terms

Performance, Design, Experimentation, Human Factors

^{*}Supported in part by a Discovery Project Grant from the Australian Research Council.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without prior permission if copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SIGIR'06, August 6–11, 2006, Seattle, Washington, USA.
Copyright 2006 ACM 1-59593-369-7/06/08...\$5.00.

Keywords

Search engines, information retrieval evaluation, user study

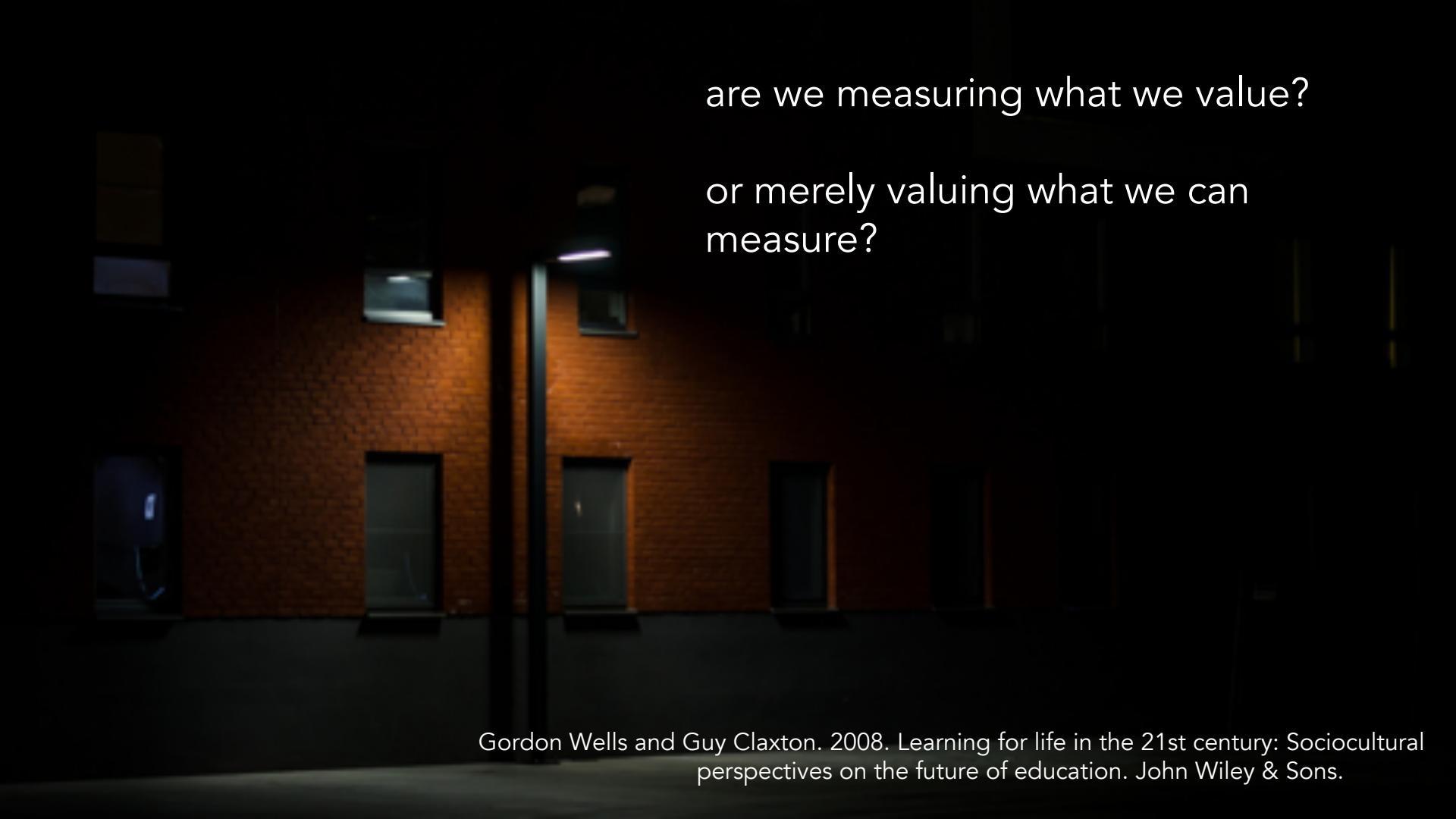
1. INTRODUCTION

The field of information retrieval has a well-established tradition of experimental evaluation, dating back to Cleverdon's "Craufield" experiments [7], and continuing through the ongoing series of Text REtrieval Conference (TREC)¹. The general approach for evaluating ad-hoc retrieval, where a static collection is searched for documents that are relevant to previously unknown topics, requires a collection of documents that is to be searched; a set of queries that represent user information needs and are run against the collection; and a set of relevance judgments that indicate, for each query, which documents satisfy the current information need and which do not. Evaluations are typically run as a batch process, where the retrieval system fetches a pre-specified number of answers document for each query, with no user interaction. Performance is quantified using a variety of metrics derived from the number of relevant answers that have been found. Commonly reported measures include mean average precision (MAP), precision at 10 documents retrieved (P@10), and bpref (these metrics are defined in Section 2). Indeed, much IR research focuses on demonstrating improvements in these metrics.

However, recent studies have demonstrated that improvements in these metrics do not translate into a direct benefit for users. A study by Henih et al. [13] shows that instance recall – where users try to identify different aspects of a question within a limited timeframe – does not improve with small increases in mean average precision of the underlying search system on the scale that is commonly reported in IR results. Alata et al. [1] confirm this result (using bpref), but also show that larger, specific increases in bpref, users do benefit on the instance recall task. Turpin and Henih [17] demonstrate a lack of improvement when users are engaged in a question answering task for a small number of questions.

A possible reason for the lack of correlation between underlying system effectiveness and user performance could be the nature of the search tasks that have been examined. Instance recall – as its name implies – is inherently recall-oriented [1, 13]. However, mean average precision, while including a recall component, evaluates systems predom-

¹<http://trec.nist.gov>

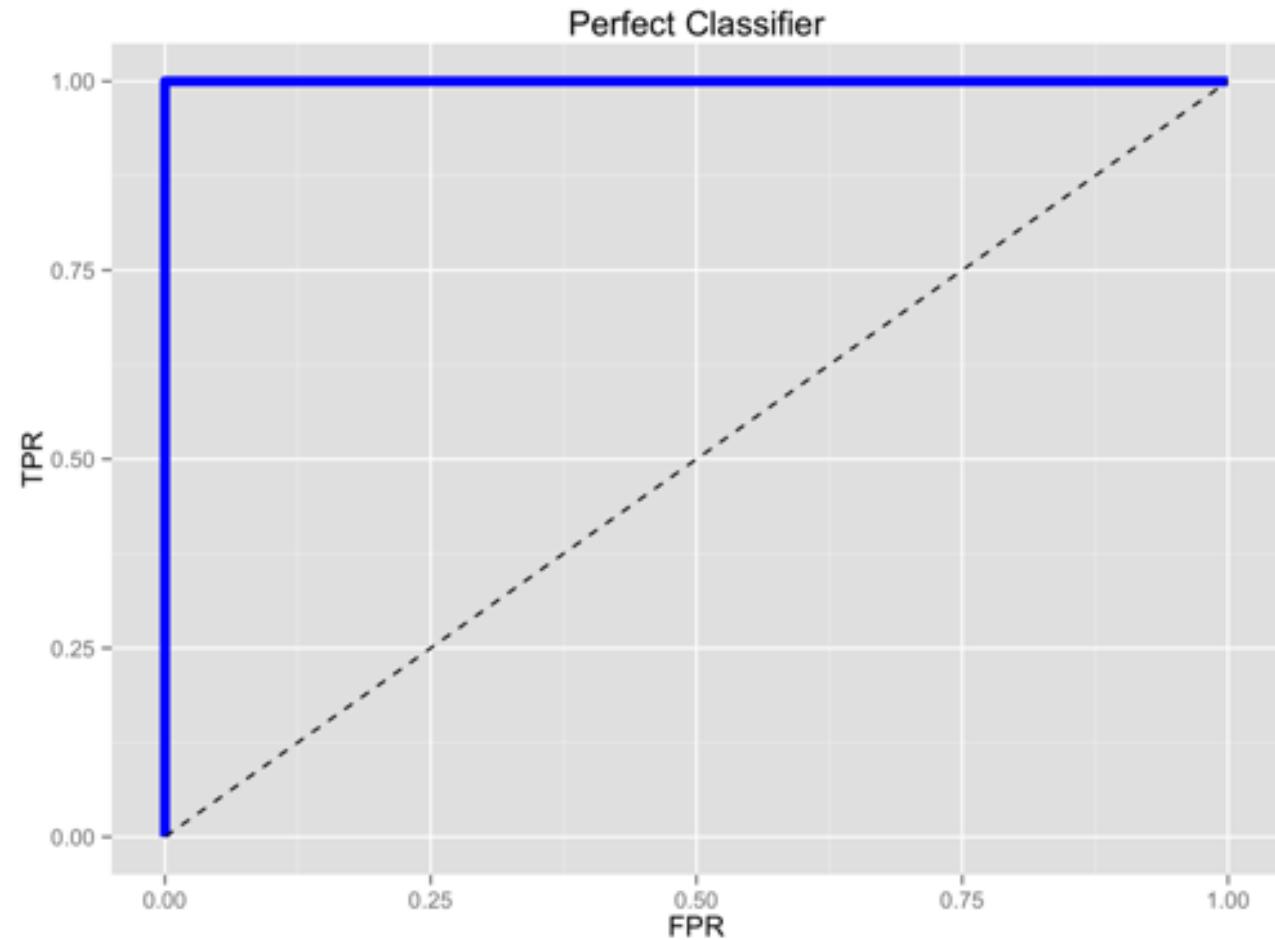


are we measuring what we value?

or merely valuing what we can
measure?

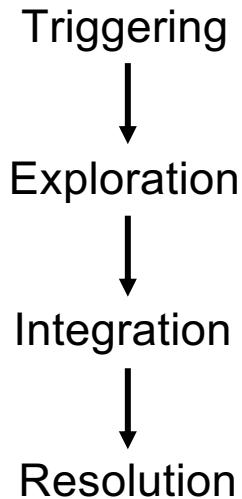
Gordon Wells and Guy Claxton. 2008. Learning for life in the 21st century: Sociocultural perspectives on the future of education. John Wiley & Sons.

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."



Garrison, Anderson, Archer (2001) Critical thinking, cognitive presence, and computer conferencing in distance education. American journal of distance education, 15(1):7-23

The diagram illustrates the four phases of cognitive presence using a screenshot of a Google+ post. The post is titled "The terror of tweeting" and includes a link to a blog post. It has several comments:

- Triggering phase:** The original post and its first comment, which links to a blog post, represent the triggering phase.
- Exploration phase:** Subsequent comments expressing opinions and sharing resources, such as a video, fall under the exploration phase.
- Integration phase:** References to previous posts and related content, such as a link to a Twitter thread, are categorized as integration.
- Resolution phase:** The final comment, which provides an insight or opinion, marks the resolution phase.

<https://plus.google.com/u/0/+StefanPSchmid/posts/4wrUbFzFwpJ>

we can use machine learning to classify discussion forum text using this construct

Kovanović, Joksimović, Waters, Gašević, **Kitto**, Hatala, Siemens (2016). Towards automated content analysis of discussion transcripts: a cognitive presence case. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16). ACM, New York, NY, USA, 15-24.

Towards Automated Content Analysis of Discussion Transcripts: A Cognitive Presence Case

Vitomir Kovanović
School of Informatics
The University of Edinburgh
Edinburgh, UK
v.kovanovic@ed.ac.uk

Dragan Gašević
Moray House School of
Education and School of
Informatics
The University of Edinburgh
Edinburgh, UK
dgasevic@acm.org

George Siemens
LINK Research Lab
University of Texas at Arlington
Arlington, USA
gsiemens@uta.edu

Srećko Joksimović
Moray House School of
Education
The University of Edinburgh
Edinburgh, UK
s.joksimovic@ed.ac.uk

Kirsty Kitto
Queensland University of
Technology
Brisbane, Australia
kirsty.kitto@qut.edu.au

Zak Waters
Queensland University of
Technology
Brisbane, Australia
z.waters@qut.edu.au

Marek Hatala
School of Interactive Arts and
Technology
Simon Fraser University
Burnaby, Canada
mhatala@sfu.ca

ABSTRACT

In this paper, we present the results of an exploratory study that examined the problem of automating content analysis of student online discussion transcripts. We looked at the problem of encoding discussion transcripts for the levels of cognitive presence, one of the three main constructs in the Community of Inquiry (CoI) model of distance education. Using Cosine Matrix and L2WC features, together with a set of custom features developed to capture discussion contexts, we developed a random forest classification system that achieved 90.5% classification accuracy and 0.63 Cohen's kappa coefficient. This is higher than reported in previous studies. Besides improvement in classification accuracy, the developed system is also less sensitive to overfitting as it uses only 205 classification features, which is around 100 times less features than in similar systems based on bag-of-words features. We also provide an overview of the classification features most indicative of the different phases of cognitive presence that gives additional insights into the nature of cognitive presence learning cycle. Overall, our results show great potential of the proposed approach, with an added benefit of providing further characterisation of the cognitive presence coding scheme.

Keywords

Community of Inquiry (CoI) model, content analysis, content analytics, online discussions, text classification

Permissions to make digital or hard copies of all or part of this work for personal or classroom use in general settings, without prior permission or written license, are granted by copyright owner for those registered participants in the conference or workshop that published the first work. Copyright for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or to republish, to post on servers, to redistribute to lists, to use other than for the purpose of this conference or workshop, requires prior specific permission and/or a fee. Request permission from permissions@acm.org.

LAK '16, April 25–26, 2016, Edinburgh, United Kingdom
© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ISBN 978-1-4503-4790-5/16/04...\$15.00
DOI: <http://dx.doi.org/10.1145/2869851.2869950>

I. INTRODUCTION

Online discussions are commonly used in modern higher education, both for blended and fully online learning [42]. In distance education, given the absence of face to face interactions, online discussions represent an important component of the whole educational experience. This is particularly true for the social-constructivist pedagogies which emphasize the value of social construction of knowledge through interactions and discussions among a group of learners [27]. In this regard, the Community of Inquiry (CoI) model [22, 23] represents perhaps one of the best researched and validated models of online and distance education, focused on explaining important dimensions – also known as presences – that shape students' online learning experience.

The most commonly used approaches to the analysis of online discussion transcripts are based on the quantitative content analysis (QCA) [12, 54, 31, 35]. According to Kruegerhoff [37] content analysis is "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" [p18]. In the case of the study presented in this paper, contexts is online learning environments. QCA is a well defined research technique commonly used in social science research, and it makes use of specifically designed coding schemes to analyze text artifacts with respect to the defined research goals and objectives. For instance, the CoI model defines a set of coding schemes which are used by the educational researchers to assess the levels of three CoI presences.

In the domain of educational research, QCA of student discussion data have been mainly used for the retrospective and research after the courses are over without an impact on the courses' learning outcomes [53]. In the field of content analysis [38], which focuses on building analytical models based on the learning content including student produced content such as online discussion messages – there have been some attempts to automate some of these coding schemes. Most notable are the efforts of McKinn [44] and Corch et al. [11] on automation of the CoI coding schemes, which served

A classroom scene showing several students at wooden desks, each equipped with a computer monitor, keyboard, and mouse. The students are focused on their work. In the foreground, a student in a blue hoodie is looking down at their keyboard. Another student in a white shirt is visible in the background. The room has light-colored walls.

should we use it with students yet?

how accurate does it have to be?

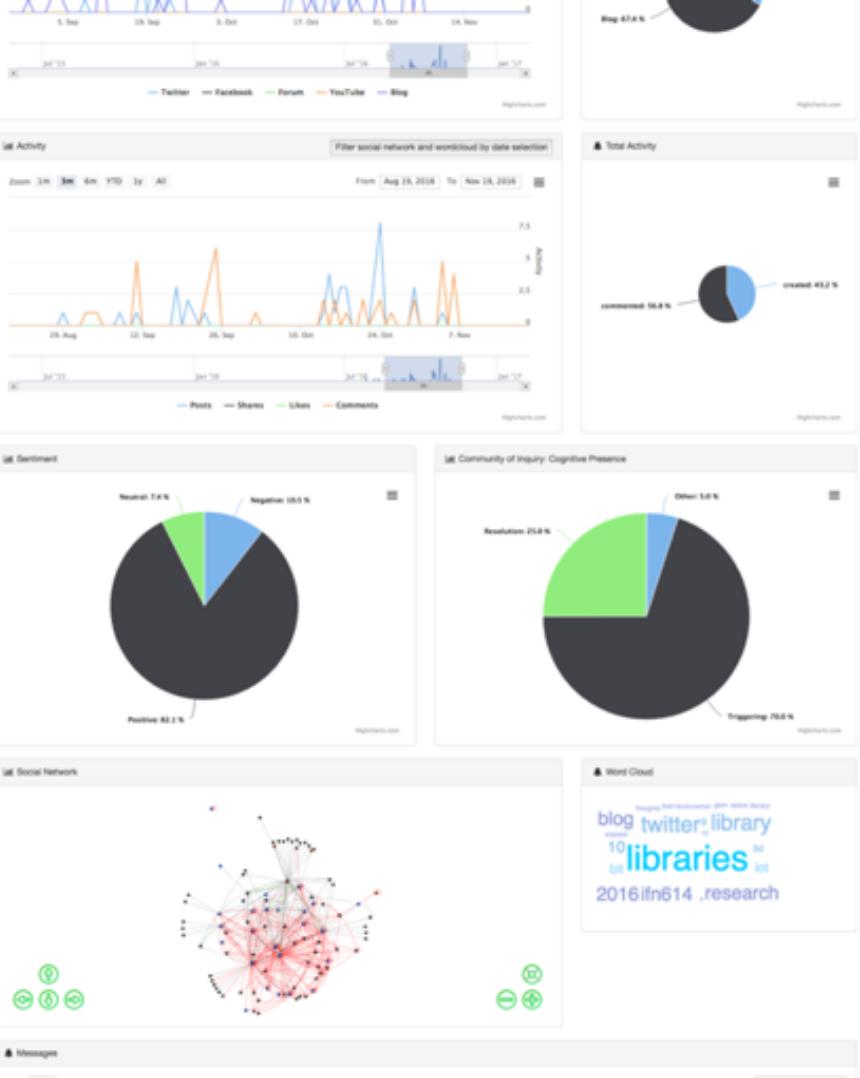
data was unbalanced (solved using boosting)

- 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?



A very strong reflection from most recent trial?

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!

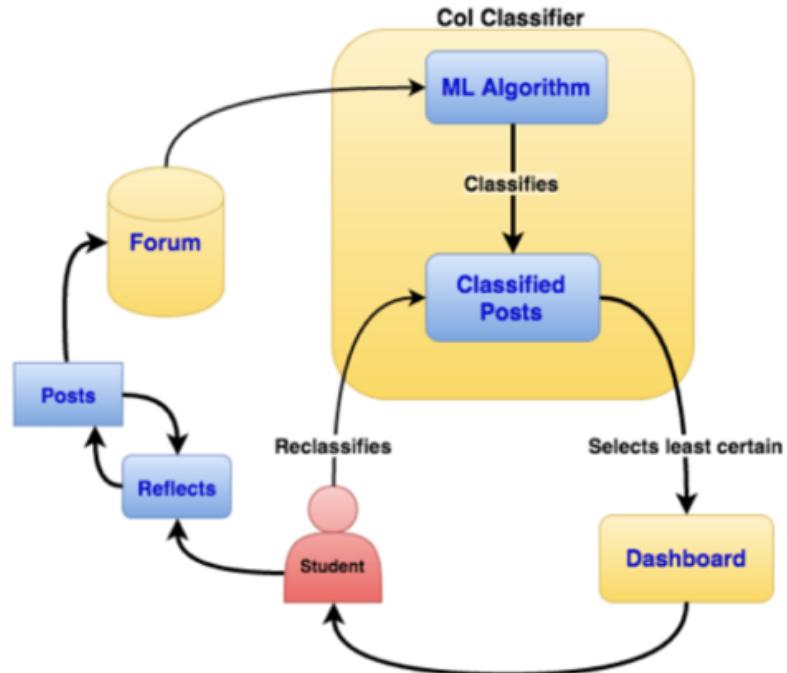
Breathing Under Water: Spirituality and the 12 Steps,
by Richard Rohr



active learning squared (AL²)

the student trains the classifier...

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



Kirsty Kitto, Mandy Lupton, Kate Davis, and Zak Waters. 2017. Designing for student-facing learning analytics. *Australian Journal of Educational Technology*, 33, 5 (2017), 152–168.

Community of Inquiry Classification

View Community of Inquiry Classifications

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

Community of Inquiry Classifications

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

View Community of Inquiry Classifications

What is this?

Was classified as: Triggering

Here's a free definition for your buzzword bingo card:

Conspectus: 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. Conspectus is also an approach that can be taken to collection development policy writing, where the policy sets out 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).

Sharing information/outside links

Triggering

Exploration

Integration

Resolution

Other

Preview:

Author	Posts
Kate Davis	1462

July 27, 2015 at 8:52 pm



Kate Davis
Keymaster

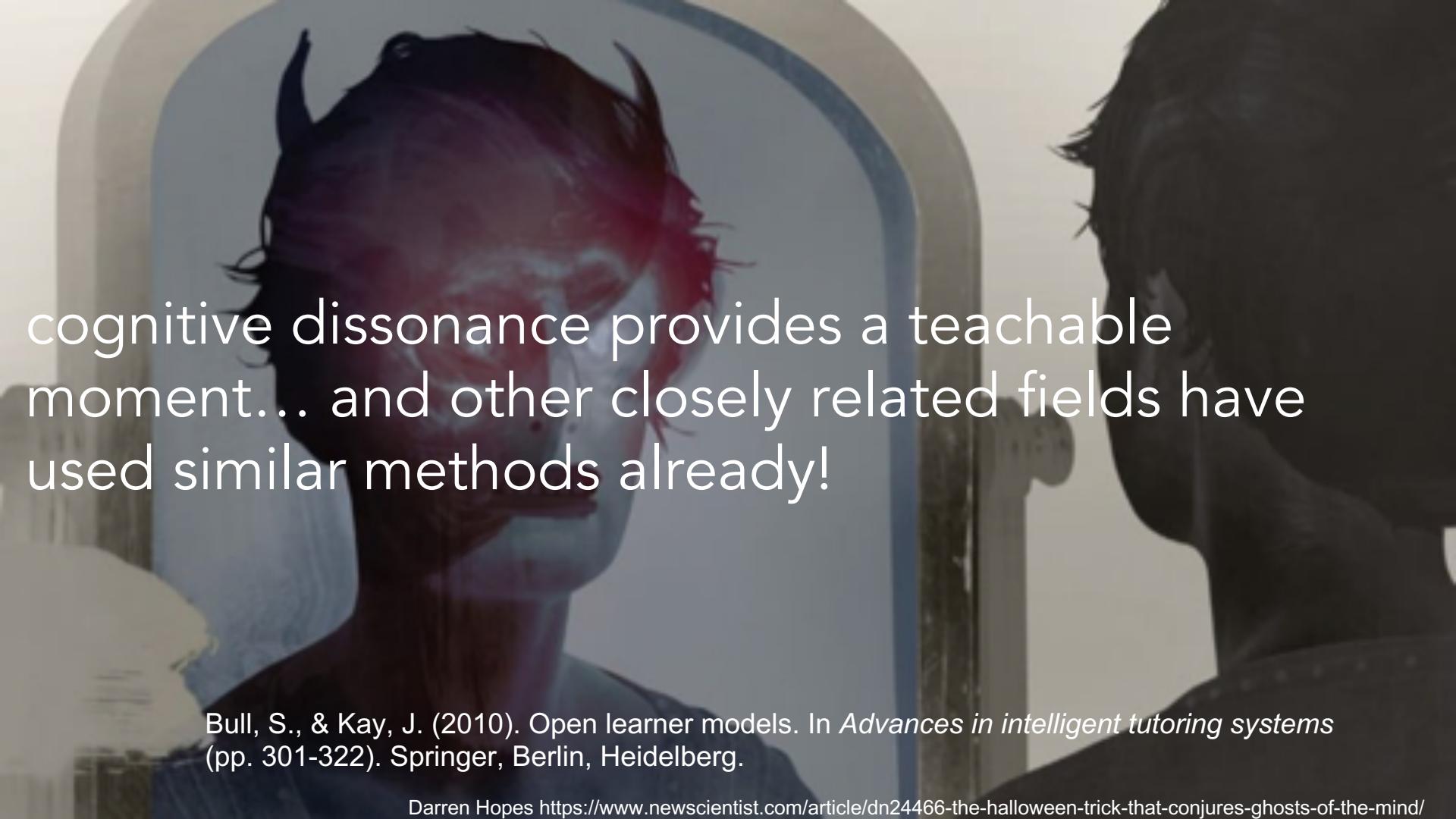
Here's a free definition for your buzzword bingo card...

Conspectus: 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. Conspectus is also an approach that can be taken to collection development policy writing, where the policy sets out 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

Key	Feedback (Reflective)
<ul style="list-style-type: none"><input type="checkbox"/> Words associated with strong feelings<input checked="" type="checkbox"/> Expressions indicating belief, learning, or knowledge.<input checked="" type="checkbox"/> Expressions indicating self critique<input checked="" type="checkbox"/> One or more keywords missing<input checked="" type="checkbox"/> Sentence too long, might disengage the reader. Try breaking it into smaller sentences<input checked="" type="checkbox"/> Initial thoughts and feelings about a significant experience.<input checked="" type="checkbox"/> The challenge of new surprising or unfamiliar ideas, problems or learning experiences.<input checked="" type="checkbox"/> Deeper reflection, personally applied.<input checked="" type="checkbox"/> How new knowledge can lead to a change	<p>Auto feedback: <input type="checkbox"/> Get Feedback <input type="checkbox"/> Save <input type="checkbox"/> Export to PDF <input type="checkbox"/> Key</p> <p>Feedback (Reflective)</p> <p>Prior to starting my clinical placement, I honestly had no idea what sort of challenges <i>I would</i> have to face in a Community Pharmacy setting. It has essentially provided me with a perspective of the expectations of a pharmacist as a health care professional. I personally saw it as a journey which exposed my strengths and weaknesses. I saw my preceptor as someone who guided me to help address my weaknesses. However, <i>I began to realise</i> that this was only to a certain extent. The most important thing <i>I learnt</i> from these experiences is that I can only develop my skills if I actively contribute to the pharmacy by demonstrating initiative. This initiative was a product of my inner passion and motivation to practise as a pharmacist in future. Various encounters along my journey proved to me that every day presents with a new challenge. I initially could not comprehend just how diverse the members of the community were, particularly in regards to their health issues and understanding of their condition. I found that my clinical placement allowed me to see things from a perspective that <i>I would never have imagined</i>. In order to illustrate these notions, I have decided to reflect upon two major ideas.</p> <p>Effective patient communication was a skill I had significantly developed during my clinical placement. A specific example was when I dispensed rosuvastatin for a patient. It was one of the first weeks of clinical placement and by this time I</p>



cognitive dissonance provides a teachable moment... and other closely related fields have used similar methods already!

Bull, S., & Kay, J. (2010). Open learner models. In *Advances in intelligent tutoring systems* (pp. 301-322). Springer, Berlin, Heidelberg.

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”

Gavriel Salomon, David N Perkins, and Tamar Globerson.
1991. Partners in cognition: Extending human intelligence with intelligent technologies. *Educational researcher* 20, 3 (1991), 2–9.



but then how can
we evaluate
success?

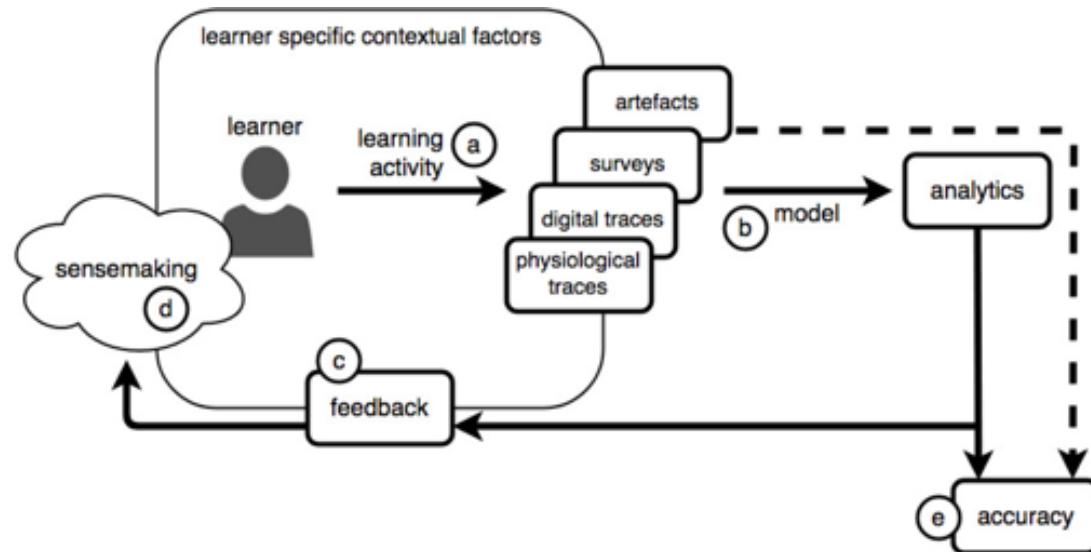


towards comprehensive evaluation for LA

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 AL²

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

@sbuckshum

@andrewresearch

Open Access reprint:

<http://bit.ly/NavajoRugLAK18>