

# Designing for student facing learning analytics

Kirsty Kitto

Connected Intelligence Centre  
@KirstyKitto • [kirsty.kitto@uts.edu.au](mailto:kirsty.kitto@uts.edu.au)

# what is learning analytics? (LA)

*Learning analytics is 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*

SoLAR definition



# example: course signals at Purdue

PURDUE  
UNIVERSITY

West Lafayette Academic Campus

Build Teach Student View

Fall 2008 - MA-154 - T Delw

Course Tools

- Course Content
- Who's Online
- My Tools
- My Grades

Your location: Home Page

# 154

# Fall 2008

 Caution! You can do better in this CLASS.

 Course Information

 Daily Lessons

 Opt in and Opt out

Questions? Contact Tim Delworth, [delworth@math.purdue.edu](mailto:delworth@math.purdue.edu)

MA 154 Course Web Page

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 267-270). ACM.

# many systems now offering some form of LA

The image displays three separate screenshots of different Learning Management System (LMS) platforms, illustrating various features for tracking student progress and course performance.

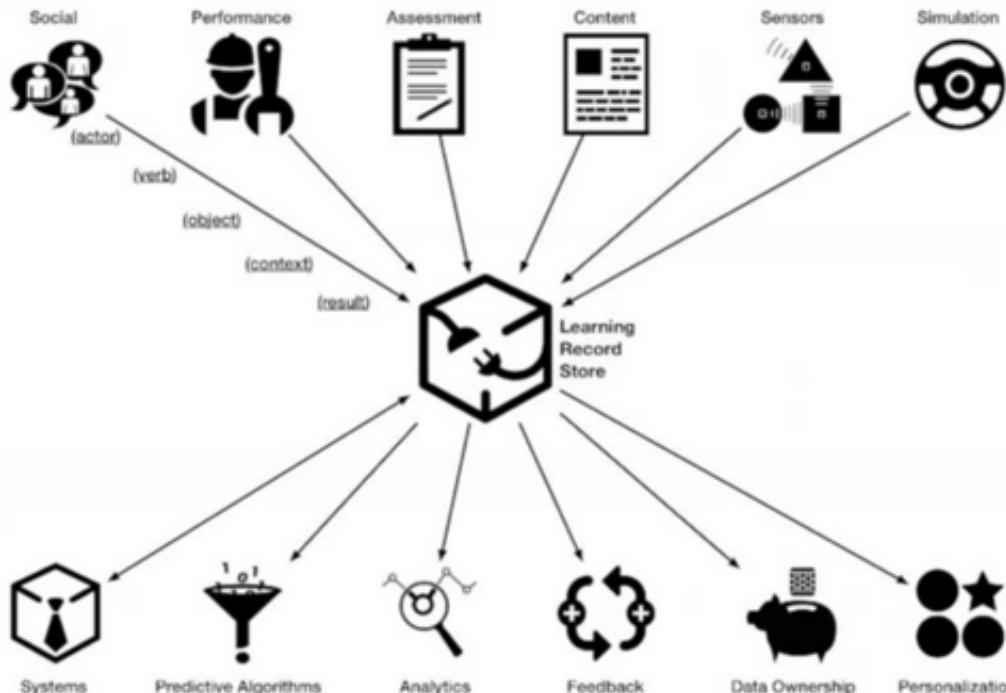
- EdPlace Dashboard:** Shows a green-themed dashboard for a user named "Bobby". It includes a "Worksheets to do" section with a list of assignments (e.g., "Use colors to signal a list", "Write a number of quarters in different ways", "Skeleton and Circulation 2"), a "Badges" section, a "Progress" section showing "Overall" completion at 64%, and a "Next reward" section indicating 41 points needed for a "A day out with your friends".
- Customer Profile View:** Shows a profile for "Kate Spade" with contact information (Email: lam.a.unicorn@gmail.com, Phone: 98125678). It displays "ACTIVE COURSES" (Ballet Intro, Acrobatics 1, Pilates Intro), "REVENUE (\$ THIS MONTH)" (486 DOLLARS), and "CLASSES THIS MONTH" (21 CLASSES). Below this are "RECENT CLASSES" and a bar chart showing "REVENUE (\$ - CLASSES BY MONTH)" for March, April, and May.
- Course Progress Overview:** Shows a grey-themed dashboard for a course titled "Essentials of Pharmacology for Health Professions". It includes sections for "OVERVIEW" (Current Grade: B+, Assignments Done: 12, Assignments Left: 17, Days Remaining: 28), "UPCOMING COURSE MILESTONES" (Unit 7 Practice Quiz - Controlled Narcotics, Final Essay Rough Draft Due, Job Shadow Research Assignment Paper Due), and "RECENT ASSIGNMENT GRADES" (UNIT 4 QUIZ: 88%, UNIT 5 QUIZ: 75%, MID-TERM: 85%). A large circular progress indicator shows "61% Complete".

## BUT is it any good?

# where does learning happen?



but new data standards are emerging...  
new possibilities!





so a lot of data is coming to education!  
but how can we use it effectively?

# but data must be cooked with care!

- *are you capturing all of the relevant data?*
- *is what you are collecting even useful?*
- *or are you just collecting it because you can?*
- *and what metrics are you developing from your data?*

Bowker, G. C. (2005). Memory practices in the sciences (Vol. 205). Cambridge, MA: Mit Press.

# the clicks to constructs problem

low level click steam data  
rarely yields significant  
insights

BUT a careful mapping to  
educational constructs can  
lead to far more useful  
outcomes



Gibson, A., Kitto, K., & Willis, J. (2014). A cognitive processing framework for learning analytics. In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (pp. 212-216). ACM.



student facing LA

# we should give students access to rich LA

*In principle this should help to promote things like:*

- learning to learn
- metacognition and reflection
- interpretation and sensemaking
- data literacy
- lifelong learning

*And ethically... is it reasonable not to give students access to the data that they themselves generate?*

but care is required...

## what would a student do if:

- they were a first in family low SES type student and told in their first year maths class that they were failing?
  - a dashboard showed them at the bottom of a leader board?
  - ... at the top?
  - a social network tool showed them as the only student who was not connected to anyone else in class? ... and they were suffering from anxiety and depression?



[Latest Staff](#)[Staff Announcements](#)[Latest Student](#)[Student Announcements](#)[Home](#) > [Announcements](#) > Student focus group: "Learner-facing learner analytics & analysis of student perspectives"

# Staff Announcements

## Student focus group: "Learner-facing learner analytics & analysis of student perspectives"

14 November, 2017 [\[Staff\]](#)

Students are invited to participate next week in a student focus group as part of a national project (Learner-facing learner analytics & analysis of student perspectives GU ref no: 2017/896) that is designed to explore the perceptions of students to learning analytics and the requirements for a potential student facing dashboard.

The aims of the project are too specifically:

- Identify the data / information institutions can provide to students, to support their university studies.
- To determine how can learning analytics can most appropriately be visualized and presented to students.
- Identify the potential concerns students have about the collection and use of data.

Two focus groups will be conducted on Tuesday, 21 November from 1-3.30pm and Thursday, 23 November 10.30am-12pm. Other focus groups can be

### Search announcements

Enter a keyword or a date to search all announcements.

[Search](#)

### Other announcements

- ▶ [Vice Chancellor's Messages](#)
- ▶ [Alerts](#)

### Submit a new announcement or event

Will this communication be for staff or students (or both)?

# things can go very wrong with naïve approaches

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study of intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, 8, 152–161.

Khan, I., & Pardo, A. (2016). Data2U: Scalable real time student feedback in active learning environments. In Proceedings of the international conference on learning analytics and knowledge (pp. 249–253). Edinburgh, Scotland: ACM.

“our combination of leaderboards, badges, and competition mechanics do not improve educational outcomes and at worst can harm motivation, satisfaction, and empowerment”

(Hanus and Fox, 2015)

# ID14-3821: ENABLING CONNECTED LEARNING VIA OPEN SOURCE ANALYTICS IN THE WILD: LEARNING ANALYTICS BEYOND THE LMS

This project is supported by the Australian Government's office for learning and teaching

QUEENSLAND UNIVERSITY OF TECHNOLOGY:

Kirsty Kitto (Lead Investigator), Mandy Lupton, John Banks, Dann Mallet, Peter Bruza

UNIVERSITY OF SOUTH AUSTRALIA

Shane Dawson, Dragan Gašević (Uni of Edinburgh)

UNIVERSITY OF TECHNOLOGY SYDNEY

Simon Buckingham Shum (and now Kirsty Kitto!)

UNIVERSITY OF SYDNEY

Abelardo Pardo

UNIVERSITY OF TEXAS (ARLINGTON)

George Siemens



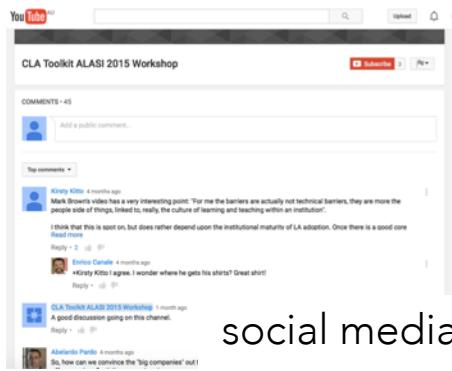
Queensland University  
of Technology



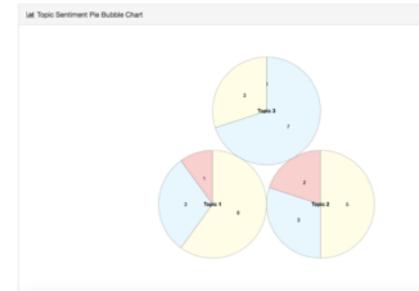
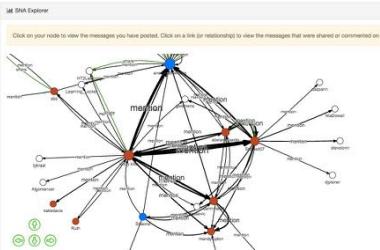
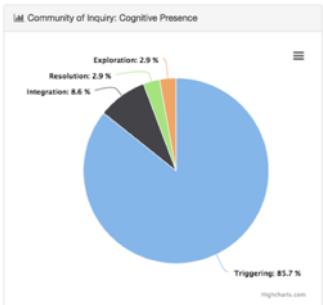
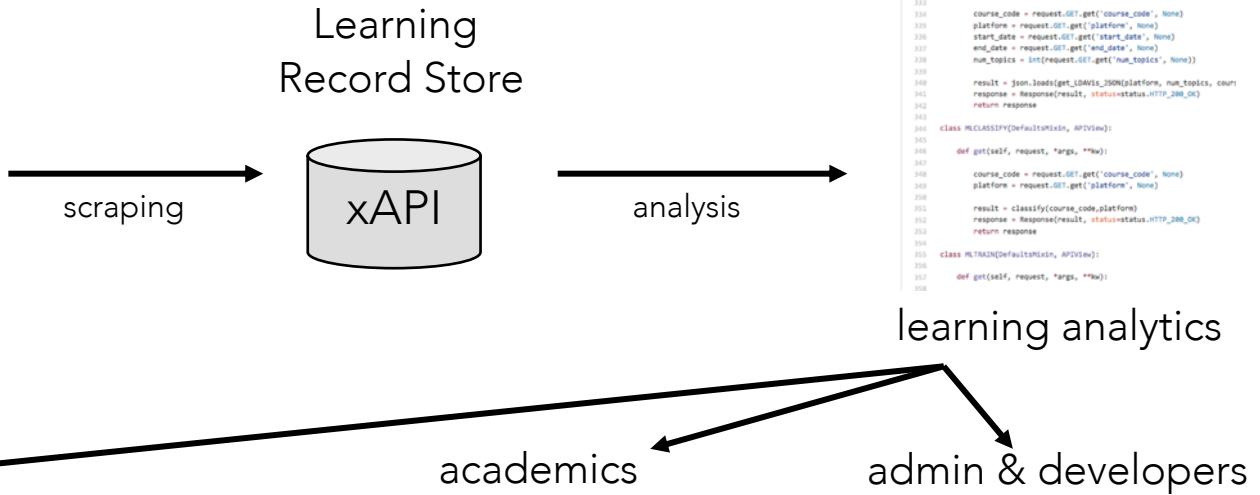
University of  
South Australia



# the connected learning analytics toolkit



social media



Learning Locker

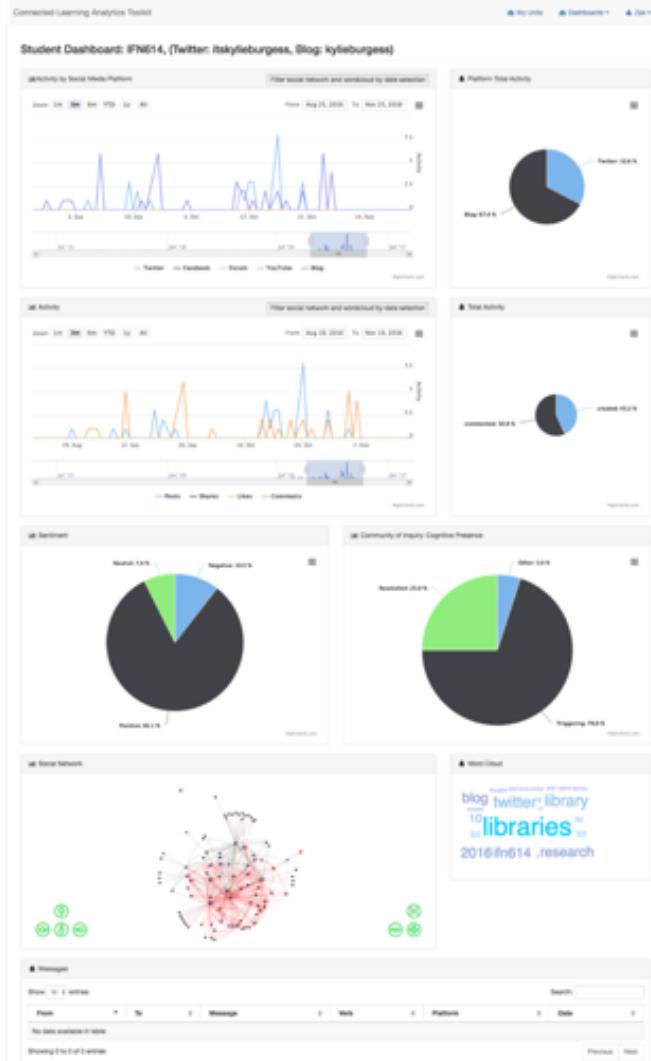
LRS List

Title	Description	User #	Created	Action
Kerry's LRS		1	2015-05-09 11:00:00	
Mandy's LRS		1	2015-05-12 23:02:18	
James's LRS		1	2015-05-12 23:02:36	
Alexander's LRS		1	2015-05-11 23:03:00	
Shane's LRS		1	2015-05-11 23:03:49	
Greer's LRS		1	2015-05-11 23:04:23	
Zain's LRS		1	2015-05-11 23:04:49	
Sebi's LRS		1	2015-05-11 23:05:00	
Simon's LRS		1	2015-05-11 23:05:36	
Annebeth's LRS		1	2015-05-11 23:32:10	

```
130 class TOPICINDEXView(DefaultMixin, APIView):  
131     def get(self, request, *args, **kwargs):  
132         course_code = request.GET.get('course_code', None)  
133         platform = request.GET.get('platform', None)  
134         start_date = request.GET.get('start_date', None)  
135         end_date = request.GET.get('end_date', None)  
136         num_topics = int(request.GET.get('num_topics', None))  
137  
138         result = json.loads(get_LDAvis_JSON(platform, num_topics, count))  
139         response = Response(result, status=status.HTTP_200_OK)  
140         return response  
141  
142 class MCCLASSIFY(DefaultMixin, APIView):  
143     def get(self, request, *args, **kwargs):  
144         course_code = request.GET.get('course_code', None)  
145         platform = request.GET.get('platform', None)  
146         result = classify(course_code, platform)  
147         response = Response(result, status=status.HTTP_200_OK)  
148         return response  
149  
150 class MLTRAINING(DefaultMixin, APIView):  
151     def get(self, request, *args, **kwargs):  
152  
153         pass
```

# some details (CLA toolkit)

1. Has a philosophy of going to the students where they are actually learning (rather than expecting them to come to us)
2. Can currently access data from: wordpress blogs, twitter, youtube, facebook, trello, github, slack
3. Stores data in xAPI format (to ensure future interoperability)
4. Only retrieves data for specific learning activities and only if students sign up
5. And gives students access to their own analytics



Question: How can we give students access to rich LA that encourages metacognition and reflection?

# CAUTION

a “go look at it” approach tends to fail

- students don’t apply knowledge
- limited reflection
- often blindly believe LA instead of questioning it and reinterpreting
- and it can be **hard to use** without scaffolding

# Learning designs for student facing LA

- authentic integration with assessment is necessary
- 3 learning design patterns are being used right now
  - do-analyse-change-reflect
  - active learning squared
  - Groupwork
- More will come in time! (Especially if you come to my workshop)

Kitto, K., Lupton, M., Davis, K., Waters, Z. (2017). Designing for Student Facing Learning Analytics, Australasian Journal of Educational Technology, 33(5), 152-168.

Kitto, K., Lupton, M., Davis, K., Waters, Z. (2016). Incorporating student-facing learning analytics into pedagogical practice. In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), Show Me The Learning. Proceedings ASCILITE 2016 Adelaide, pp. 338-347.

# do-analyse-change-reflect

**Do:** Students are instructed to participate in some sort of activity.

**Analyse:** Students are encouraged to consider LA dashboards that have data collected during the *do* phase.

**Change:** Students encouraged to consider *changing* their behaviour as a result of the analytics that they see in the *analyse* phase.

**Reflect:** Students participate in a reflective process where they explain how they used the LA to make sense of their behaviour, and whether they decided to change as a result (and how).

# does it work? ... maybe

Unit	Semester	Aim/pattern	Linked to assessment	N=
<b>IFN614 Information Programs</b>	S2, 2015	Piquing students curiosity Examine, relabel classifier	No	S:12 AL:6
<b>IAB260 Social Technologies</b>	S1, 2016	Do-analyse-change-reflect	Yes	S:23 B:17
<b>IFN614 Information Programs</b>	S2, 2016	Do-analyse-change-reflect (predict, compare)	Yes	S:21 B:11



IFN614 Information Programs

UNIT SITE

Need help? 

Ask your peers: tweet with the hashtag #ifn614

Important question? Tweet the teaching team: @ifn614

Personal question? Email the teaching team: [out.ifn614@gmail.com](mailto:out.ifn614@gmail.com)

Latest posts from the teaching team blog

How to export your blog's content and put it somewhere else

How to export your blog's content and import it into your personal blog or you

Author: Kate

Posted: November 23, 2016, 8:36 am

Good luck and a reminder

Just a quick note to remind you that your assignment all needs to be finalised by 11.59pm Sunday

Author: Kate

Posted: November 4, 2016, 9:21 am

Get questions about the CLA Toolkit?

A few of you have mentioned you are having difficulties with the CLA Toolkit. If that's you:

psia

Author: Kate

Posted: October 30, 2016, 10:01 am

That's that (quite) all folks!

It's Week 13 already and that means it's time for a host of lasts our last Twitter chat, tonight

IFN614 Information Programs

UNIT SITE

Latest tweets from the teaching team

Latest #ifn614 tweets

#ifn614

Hmm, an empty timeline. That's weird.

Check for Tweets

What is a Twitter chat?

Here's everything you need to know to participate in our Twitter chats this semester.

On this page:

- What is a Twitter chat?
- Prepare for your first Twitter chat
- Use a Twitter chat monitoring tool
- Things to do in advance of the chat
- Things to do right before the Twitter chat
- Things to do during the chat
- Things to do before you log off
- Things to do after the Twitter chat

Twitter chat tips

Recent Posts

- How to export your blog's content and put it somewhere else
- Good luck and a reminder
- Got questions about the CLA Toolkit?
- That's that (quite) all folks!
- Return of Assignment 2

Recent Comments

- Service review: QUT Library Reference Service - 'Ask a Librarian' feature shown on Week 3 Reference
- Kate on Assignment 2 & 3: EDI and grant application
- All on Assignment 2 & 3: EDI and grant application

# for example (Trial 3)

**Do:** blogging assignment was introduced in the first week of semester

**Analyse:** In week 2 students were introduced to the Col model (Garrison et al., 2001) and were encouraged to sign up for the CLA toolkit (optional)

- a class provided an overview of the Col model and the CLA toolkit
- 23/40 signed up (eventually)
- Students blogged about role and activity they were aiming for

**Change:** Students encouraged to think about how they were contributing to the community using data in the CLA toolkit dashboard and to change

**Reflect:** In week 14 students were required to critically evaluate their engagement with respect to their aims in week 2 (assessed!)

# final blog post prompt for Trial 3

- What role did you want to play in the community this semester? Did you achieve that?
- How many comments did you make on your peers' posts
- Why did you comment as much as you did; what factors influenced the volume of your contributions?
- Did you need to modify your instinctive behaviour to engage the way you wanted to, or felt you should, engage?

Score	Level of analysis	N = 11
1	Included some /all graphs with no reference or analysis	1
2	Included some/all graphs, quantitative analysis relating activity to personality &/or interest	2
3	Included some/all graphs, quantitative analysis relating activity to personality &/or interest, basic analysis on activity in relation to week 2 aim	5
4	Included some/all graphs, referred back to week 2 aim, compared & contrasted, mentioned qualitative aspects	3

Out of 21 who signed up, 40 total!

# 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]



Home   The Team   Scenarios   Case Studies   Tool   Institutional Adoption   Dissemination   News   Contact

Providing personalised,  
timely support actions to  
large student cohorts.

LEARN MORE



FOLLOW US





**PERSONALISATION  
IS NOT ABOUT  
FIRST/LAST NAME.  
IT'S ABOUT  
RELEVANT  
CONTENT.**

Dan Jak



**CATHERINEMONGINA**



catherinemongina

# weekly personalised feedback to 800+ students

(Acknowledgement: Jurgen Schulte, UTS Science)



Lecture  
Response

Attendance  
Engagement  
Understanding  
Workload



Assignments  
Exams

Participation  
Engagement  
Solving Skills  
Time  
Management



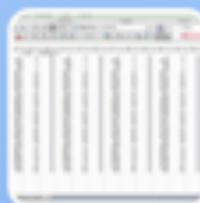
Laboratory

Participation  
Time  
Management  
Solving Skills  
Reporting  
Skills



Exams

Selected  
Material  
Comprehension  
Solving Skills  
Engagement



Dashboard

Engagement  
Participation  
Working  
Pattern

# rapid, personalised feedback at scale to students

(developed by Jurgen Schulte, UTS Science)

End of **week 3** feedback **case 3**

cond 1      Dear Osiri,

cond 2      Quite a few students had to move lab classes the past two weeks. This is just to confirm that I have you on record that your are now in lab **Group 18** and that your online lab report should be submitted at our **Group 18** pages.

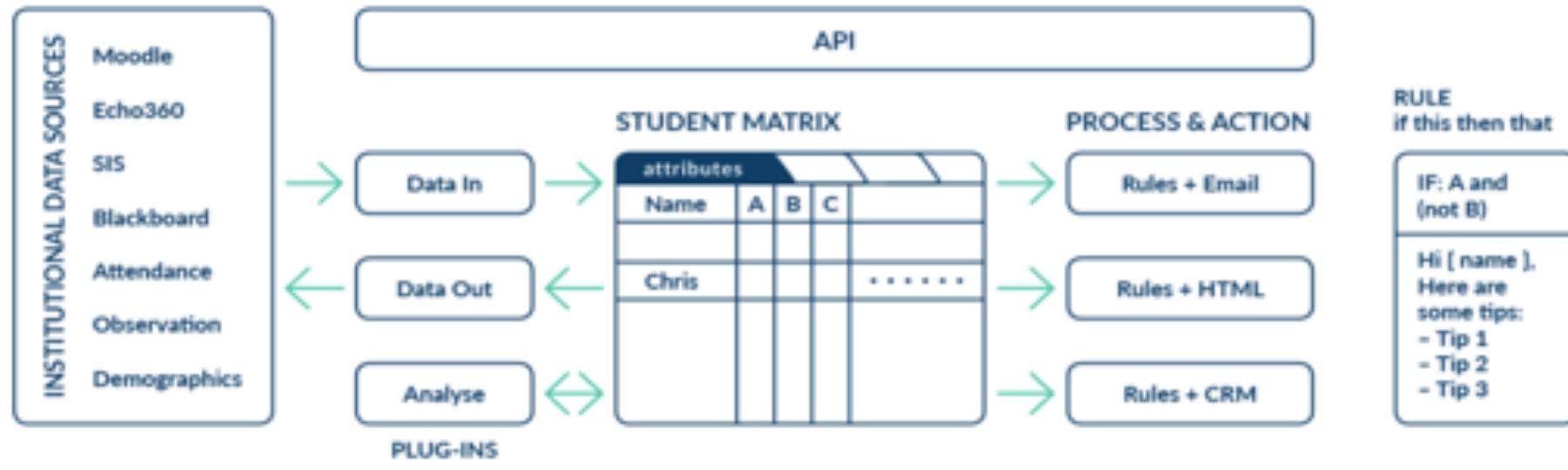
cond 6      You had a good start with Physical Modelling and seem to be well on track.  
cond 7      You managed to achieve 9 out of 10 marks in your WileyPLUS assignments.  
cond 8      Your lab reports came back with 7 out of 7 marks.  
cond 9      I noticed you are a keen participant of our lecture exercises. Did you know that they can be accessed before as well as after the lecture, not just during lecture?  
cond 5      You seem to have had problems with one of the forces questions. Please have a look at HRW Chapter 3.2.2 where this case is discussed in more detail.

Please don't forget that the our **third** homework assignment has been released already. This assignment will be due 11.00 pm Friday next week.

Kind regards,  
Jurgen Schulte

# to national funding...

Office for Learning & Teaching: <http://OnTaskLearning.org>



Support



Educational  
Designers



Academics



Students



The 8th International

# Learning Analytics & Knowledge Conference

SMC Conference & Function Centre, Sydney, NSW, Australia

March 5-9, 2018 #LAK18



HOME

ABOUT

SCHEDULE

CALL FOR PAPERS

KEYNOTES

PROCEEDINGS

SPONSOR LIST

REGISTRATION



## WELCOME TO LAK'18 – SYDNEY, AUSTRALIA

Welcome to the official web site for the 8th International Learning Analytics and Knowledge (LAK) Conference. This year's conference will be held at the SMC Conference & Function Centre in Sydney, Australia on March 5-9, 2018. The preparations for LAK'18 are currently ongoing and we will be releasing information in the coming months.



Tweets by @lak18syd

LAK18 @lak18syd

Notifications for #lak18syd are now out. Register for early bird rates at [lak18syd.com/conference-reg...](http://lak18syd.com/conference-reg...) before 9th January



Nov 23, 2017

LAK18 Retweeted

Rebecca Ferguson @RDevecaF

If you would like to join the @lak18syd doctoral consortium, today is the deadline for applications



Nov 10, 2017

don't forget!

# Panel – solving a common problem

Kirsty Kitto & Andrew Gibson  
Connected Intelligence Centre  
 [@KirstyKitto](mailto:@KirstyKitto) • [kirsty.kitto@uts.edu.au](mailto:kirsty.kitto@uts.edu.au)

# how can we improve discussion forums?

student discussion forums often lapse into naïve arguments or banal commentary...

how can we encourage students to develop their communication skills in this online format?

Welcome to the Instructional Designers group! This is a group for teachers and instructional designers to help each other, collaborate, and discuss elements of course design in Canvas.

If learners think it looks bad, you may have lost a good percentage of the battle in getting them to pay attention.  
- Paul Shanks (@PaulShanks)

ASK A QUESTION START A DISCUSSION VIEW ALL RESOURCES VIEW ALL MEMBERS

Tools Anyone here Interested in learning analytics iaitinstructionaldesign

I'm considering a job switch from instructional design to learning analytics. Is learning analytics all that different from standard data analytics?

17 comments share save hide report

WANT TO ADD TO THE DISCUSSION? Post a comment!

CREATE AN ACCOUNT

I am trying to write a Shiny app graphing the density of a variable VAL, by categories of age (AGE) or sex (SEX). The user selects "SEX" or "AGE" in the dropdown menu, and I have been trying to use `fill = input$var` or `group = input$var` in `ggplot` and `geom`.

In `ggplot`:

```
output$plot <- renderPlot(ggplot(data=ggfR, aes(x=VAL, y=group, fill=group)) + geom_density(aes(group=group), position="stack"))
```

EDIT: as docento pointed out, this can be fixed for `ggplot` using `aes_string`:

```
output$plot <- renderPlot(ggplot(data=ggfR, aes(x=VAL, y=group, fill=group), aes_string(fill=fill)))
```

In `geom`:

```
gvis <- reactive(function(gpar, GDF=ggfR$group, select) { ggvis(x = ~VAL, y = ~group, group_by = ~group, select = ~VAL, layer_densities=adjust(x=ggfR$group, select = ~VAL, add_axis="y"), title = "Density", ticks="none") %>% vis_label("Density", "group", "VAL") %>% gvis %> bind_shiny("ggvis", "ggvis_id") })
```

SOLUTION: using `as.name(input$group, select)` will render the graph properly!

This is (was) what is rendered: `ggvis link to shiny output`. Interestingly, it seems that `group_by` correctly interprets `input$group`, whereas `input$group, select` is treated as a constant in `fill=input$group, select`.

Any ideas on how I could get the plots to render correctly?

Here is the full Shiny code:

```
ui.R
```

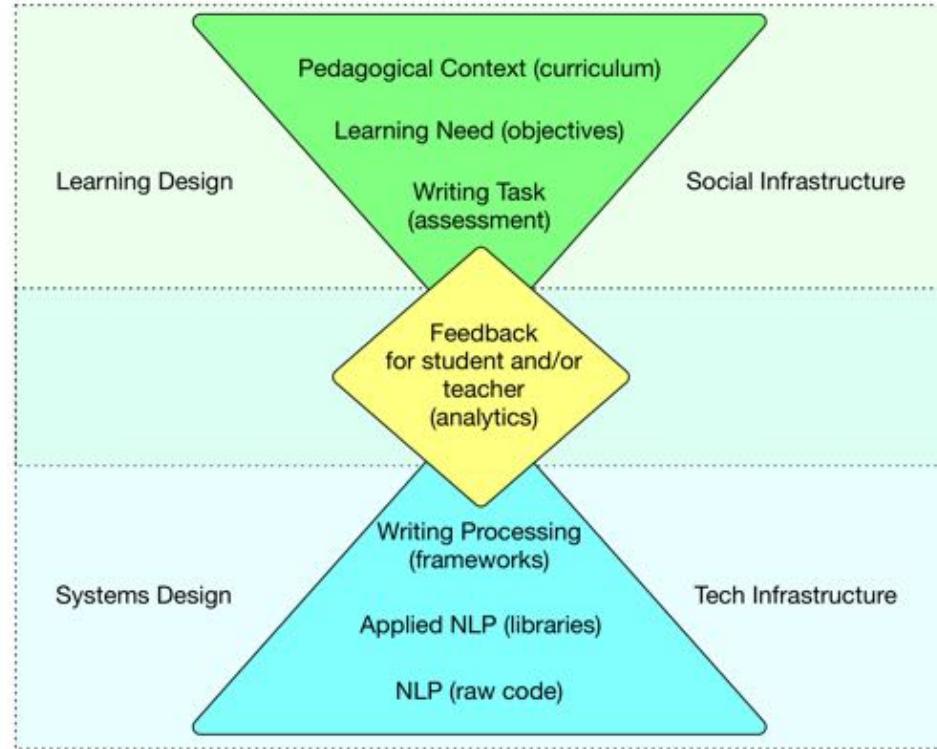
```
library(shiny)
```

Linked

Insert reactive input variable to aes of ggplot

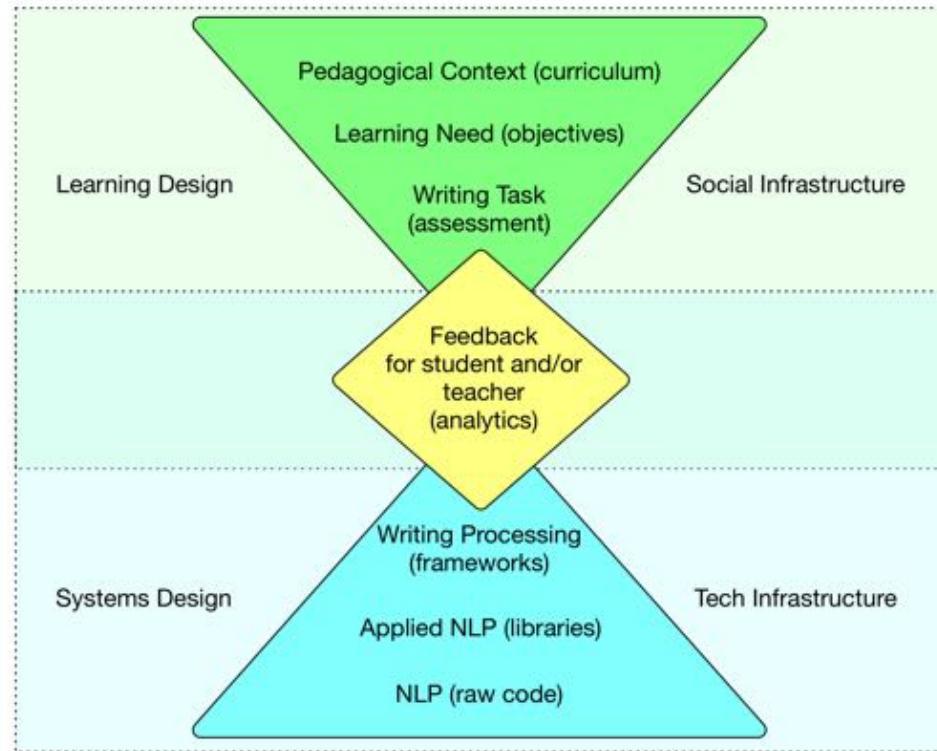
# The Writing Analytics Approach

- View the issue not as solely an analytics problem
- Joint pedagogic and computational approach
- Socio-technical solution



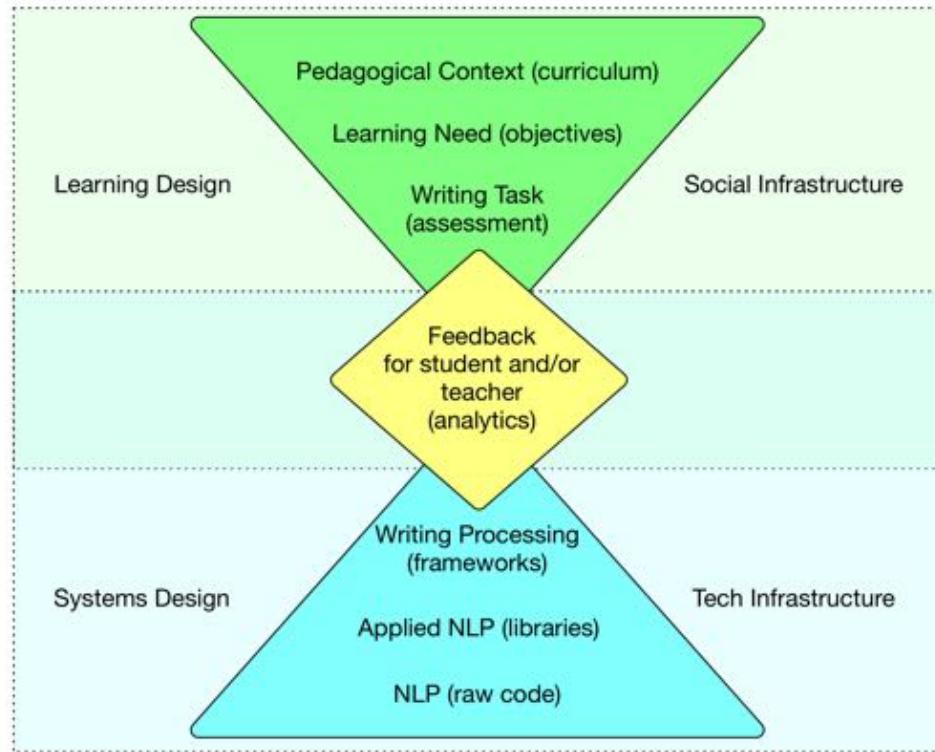
# Formative feedback – learning value

- Providing individual feedback to each contributor on their discourse
- The extent to which desirable language is used
- The extent to which undesirable language is not used



# What to look for?

- simplistic or throw-away statements
- me-too statements
- Conflict
- Agreement
- statements describing problems
- statements positing solutions
- identifications of self change



# An example...

UTS: CIC | xerox  Home t3 Reflection

■ AWA Feedback ■ TAP Output

Depth	Intention						What change is likely to lead to future benefits?
	Integration			What impact on my goals/inspirations?	What other ideas could I use to change myself?	How do others address these challenges?	How can I learn from other perspectives?
	Internalisation		What do these feelings say about me?	How is this a problem that challenges me?	Why do I need to change?	How can I change?	
	Interpretation	What does it mean for me?	Why do I feel this way?				
	Impression	What do I notice about my situation?					
		Thoughts	Feelings	Challenge	Self critique	Potential solution	Learning opportunity
CONTEXT				CHALLENGE		CHANGE	
Narrative							

## General feedback on your writing:

- Your document does not appear to have a good balance of the key sentence level features that AWA expects to find in reflective writing. These are represented as coloured icons located at the beginning of sentences. In reflective writing, AWA does not expect to have very few of one type of sentence feature together with a large quantity of another. Check this against your assessment rubric or with one of your subject tutors as you may be missing a key element of good reflective writing.
- Your document appears to have a number of words that AWA does not recognise. Try using word processing software to locate the specific mistakes, correct your document, and try submitting it again to AWA.

## JIM's Business communications reflection

Effective dialogue is essential not only in an organisational environment, but within my day-to-day life. Communication starts with the self. **● By becoming aware of my learned cultural viewpoints and limitations, I can ascertain how to overcome them.** **▲ I must understand my own biases, why I have them, and how this impacts my environment and the way I collectively work.** **● The reflective essay has illustrated disconnects between the communication theories I learned and wrote about in assignment one and the group dynamics, communication culture and my own personal limitations and dialogue flaws within the group essay development.** **● Not only have the theories informed my educational development, but also the issues discovered during the group case study have informed my professional environment.** There are many obstacles to becoming an effective and authentic communicator in the 21st century including time and physical locations, yet through self-reflection & personal knowledge of my limitations & bias, I am able to navigate obstacles in order to pursue authentic dialogue.

The group assignment was an important project to not only collectively formulate a communication case study

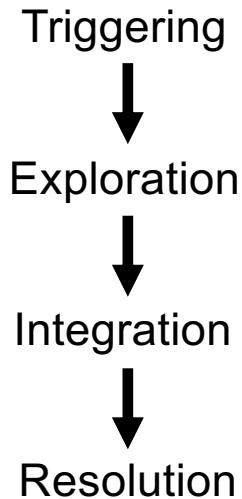
Feedback



active learning squared

# cognitive presence

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



The screenshot shows a Google+ post from Stefan P. Schmid. The post includes several annotations:

- Problem presentation, Triggering phase:** Points to the first comment by Stephan F. Schmid.
- Relation to previous post, Triggering phase:** Points to the second comment by Stephan F. Schmid, which refers back to the first.
- Tagging participants:** Points to the 'Tagged' section where Stephan F. Schmid is mentioned.
- Shared resource, Triggering phase:** Points to a blue link labeled "The secret of creating super-industry business message!"
- Insight, Opinion, Trigger phase Exploration phase:** Points to Stephan F. Schmid's response with a detailed explanation.
- Related video, Exploration phase:** Points to a video thumbnail.
- Reference to previous post, Integration phase:** Points to Stephan F. Schmid's response referring back to his own previous post.
- Related original content, Integration phase:** Points to a link in Stephan F. Schmid's response.
- Insight, opinion, Resolution phase:** Points to Stephan F. Schmid's final response summarizing the discussion.

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

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

# current state of the art uses 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

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

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

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 70.5% 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, beyond those granted by a license under copyright law or terms of service, require prior permission 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

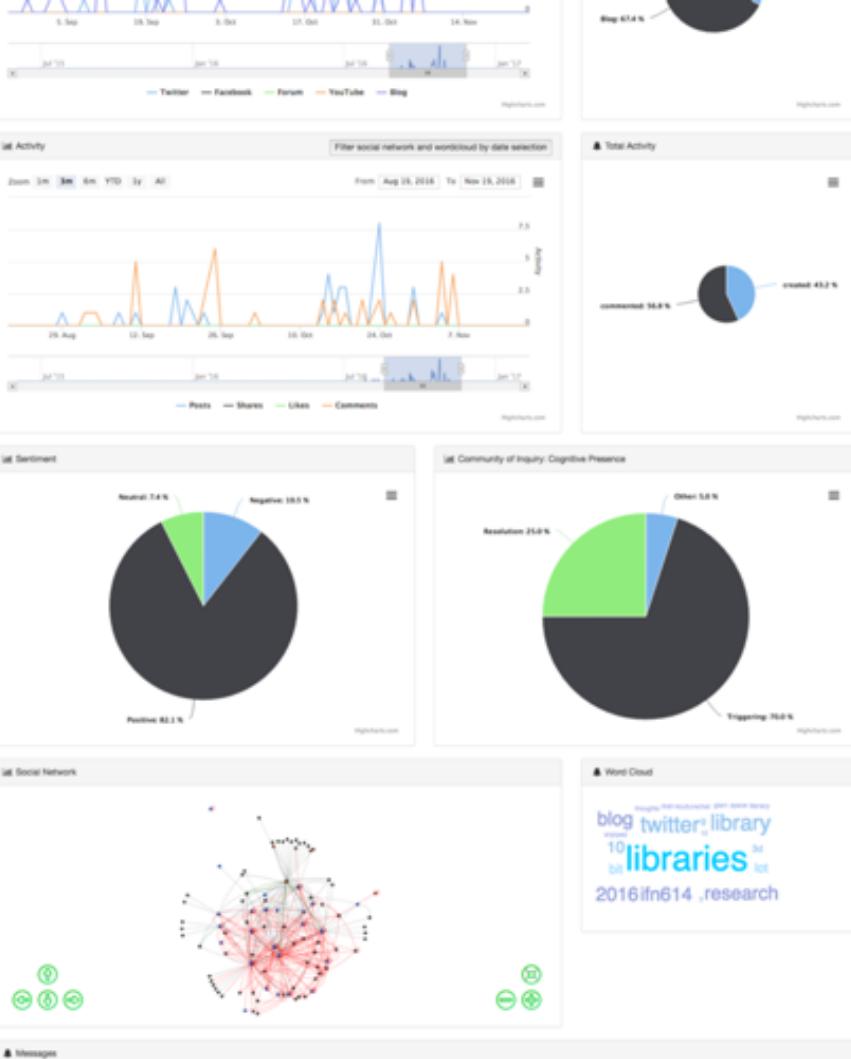
Online discussions are commonly used in modern higher education, both for blended and fully online learning [42]. In distance education, given the absence of face to face interactions, online discussions represent an important component of the whole educational experience. This is particularly true for the social-constructivist pedagogies which emphasize the value of social construction of knowledge through interactions and discussions among a group of learners [2]. 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 are 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 Corlett et al. [11] on automation of the CoI coding schemes, which served

# but contextuality...

- training data sets are rarely shared in education
- and cohorts change – a lot!
- the Col report uses (not very) accurate Machine Learning
- need to be able to rapidly train classifiers for new cohorts
- does this provide a new teachable moment?



# Active Learning Squared

the student trains the classifier while it is training the student...

Connected Learning Analytics Toolkit

## 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.

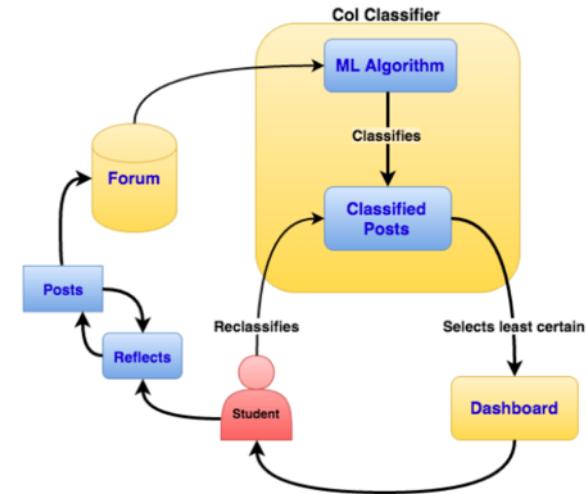
Triggering Phase initiates discussion about a particular issue/topic for inquiry.

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**



# Active Learning Squared

Connected Learning Analytics Toolkit

## Community of Inquiry Classification

Let Community of Inquiry Classifications What is this?

We're 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 PLA guide to collection development policies. Conspicuous 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 [redacted]

Triggering	Exploration	Integration	Resolution	Other
------------	-------------	-------------	------------	-------

Preview:

Author	Posts
July 27, 2015 at 9:52 am	4402
 Jane Smith	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 PLA guide to collection development policies. Conspicuous 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).

# Does it work?

	posts	class	agree	ToTut	ToClas	%ToM
A	8	8	0.125	1:56	3:18	0:28
B	10	10	0.333	0:58	0:55	0:06
C	7	5	0.200	2:06	2:07	0:32
D	19	19	0.181	1:47	4:06	0:12
E	4	4	0	1:22	0:49	0:16
F	18	18	0.050	5:12	4:42	0:17
Av	11	10.67	0.143	2:13	2:40	0:19

**Table 3:** Key performance indicators for each IS student attempting the  $AL^2$  task.

	IRR ( $\kappa$ )	IRR (%)	EC-UnSM	EC-SM	NB
IS dataset	0.09	43.0	0.473	0.305	0.302
ALASI15	0.3	47.4	0.342	0.368	0.078

**Table 1:** Accuracy of the three different classifiers investigated in this work for the IS and ALASI15 datasets. IRR between the two expert coders is also given, both as a kappa value ( $\kappa$ ) and as an percentage of agreement (%) for the two datasets.

	class	agree	ToTut	ToClas	%ToM
A	13	0.153	0:20	3:56	0:18
B	10	0.400	2:31	2:10	0:13
C	13	0.428	3:44	8:35	0:40
D	10	0.500	0:45	2:07	0:13
E	8	0.375	3:47	2:04	0:16
F	3	0.333	0:29	0:19	0:17
G	35	0.114	3:04	5:02	0:09
H	2	0	1:26	0:45	0:45
I	12	0.333	5:01	5:43	0:29
J	8	0.250	6:36	3:40	0:28
K	19	0.450	3:08	7:02	0:22
L	6	0.167	0:21	2:17	0:46
M	7	0.142	1:55	4:59	0:43
N	27	0.259	1:31	9:38	0:21
O	35	0.228	1:51	2:58	0:05
P	15	0.400	0:20	5:12	0:21
Q	6	0.333	4:22	5:22	0:54
R	27	0.222	0:35	11:38	0:39
S	1	0	5:10	0:00	0:00
T	7	0	3:47	5:58	0:51
Av	12.61	0.254	2:32s	4:28	0:25

**Table 4:** Key performance indicators for each ALASI15 participant attempting the  $AL^2$  task.

only Trial 1?  
why did it not run  
with Trial 3?

- it did
- no students used it
- no link to assessment  
(made the go look at it mistake again)

watch this space ;)





Questions?



[utscic.edu.au](http://utscic.edu.au)