



Making WHOOPPEE:

A collaborative approach to creating the modern student peer assessment ecosystem.

Alec Lamon & Dave Comroe



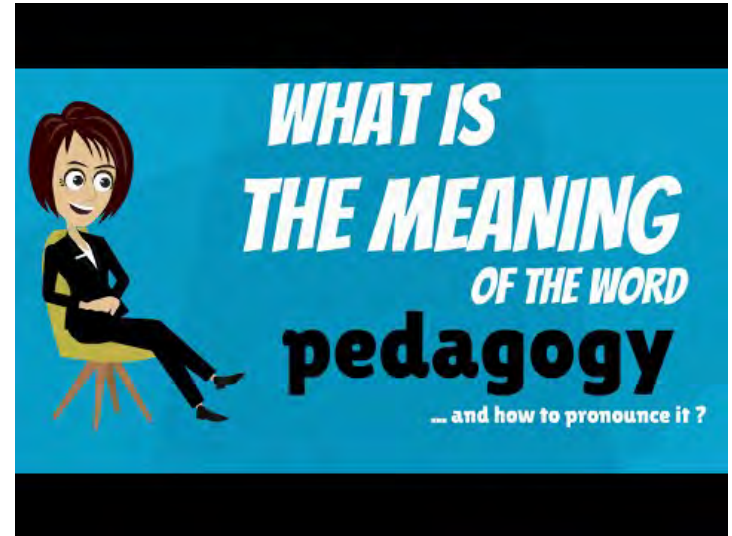
WHOOOPPEE:

Wharton Online Ordinal Peer Performance
Evaluation Engine

WHOOOPPEE Project Goal

Develop the most accurate peer evaluation system.

- Pedagogy
- Technology
- Process
- Analytics



“The Professor”

Dr. Peter Fader

Frances and Pei-Yuan Chia Professor of Marketing

Chair, Wharton Computing Faculty Advisory Committee

Quantitative Marketer with research interests in Predictive Analytics

Coined the name “WHOOOPPEE”



Pete Fader

“The Genius”

Daniel M. McCarthy

Ph.D. Candidate, Expected 2017

Statistics Department, The Wharton School of the
University of Pennsylvania



Dan McCarthy

“Mr. Canvas”

Rob Ditto

Technical Director, Courseware Team

Wharton Computing



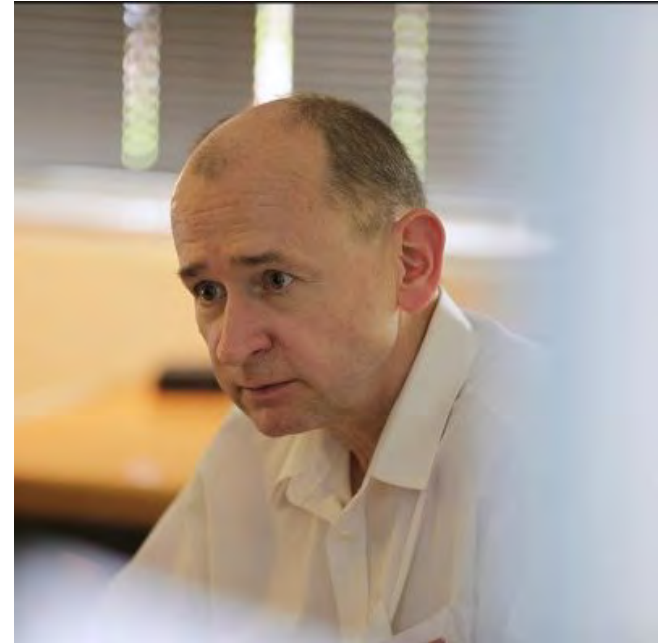
Rob Ditto

“Mr. MOOC”

Don Huesman

Managing Director, Wharton Online Learning

Ed.D., Higher Education Management, University of Pennsylvania



Don Huesman

“Gets to work on all the cool new stuff”

Alec Lamon

Senior IT Director, Research & Analytics

Wharton Computing



Alec Lamon

“Somehow gets to speak first at Educause”

Dave Comroe

Senior IT Director, Client Technology Services

Wharton Computing

Adjunct Professor



David Comroe

Professor Fader

[First clip](#)

http://players.brightcove.net/13421203001/S1ZGqh81I_default/index.html?videoId=5178900960001

The Problem: Grading

Traditional Grading has numerous inherent challenges:

- Bias
- Multiple Graders
- Size of class / multiple sections
- Consistency
- “Extraneous Influences”



The Opportunity: Better Outcomes

Can we develop a better solution for:

- “Better” Grades
- Enhanced Learning Opportunities
- More Student Engagement



Challenges with Peer Evaluation

- Significant Research
- Lots of examples currently in production
- Used heavily in MOOCs with many 1000's of students
- Used ineffectively in MOOCs?
- How do we ensure that the grader is effective?
- Issues with asking students to grade assignments using traditional methods.



WHOOOPPEE

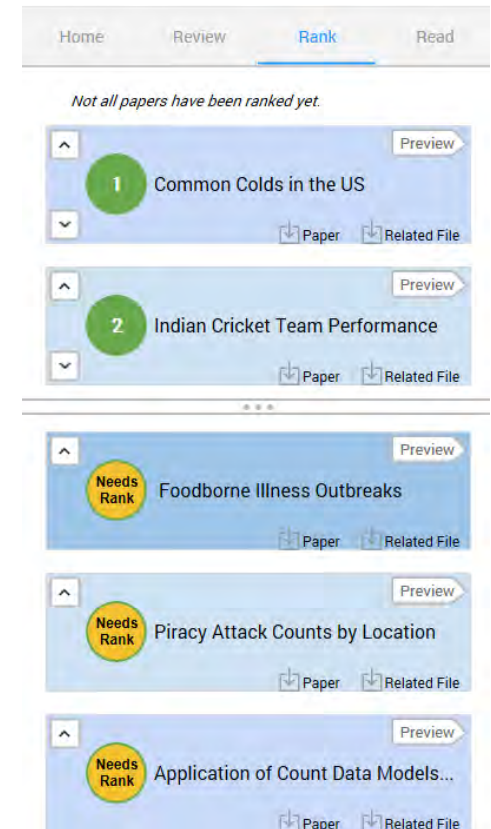
What makes WHOOOPPEE different?

- Ordinal Ranking Assessment
- Secret Sauce – WHOOOPPEE Algorithm
- Gold Standard Reviews
- Pedagogical Improvements



Ordinal Ranking

- Class given clear grading rubric in advance
- Students randomly assigned batches of 5 anonymous assignments
- Assignments uniquely *ranked* from Best to Worst
- “Easy” for the student



WHOOOPPEE Algorithm

- Translates Ordinal ranks to True Scores
- Accounts for variability in batch difficulty
- Strongly correlates a student's ability to create a quality assignment with ability to assess quality
- Identifies and corrects for a student's ineffectiveness in assessing quality or attempts to 'game the system'



Gold Standard

- Professor and TA rank batches
- Ensure all assignments are ranked by teaching team
- Gold Standard Reviews have weights equal to that of the best overall student for each assignment
- Helped reduce student uneasiness with peer assessment in early runs
- Pete loves doing it, optional in other classes.



Pedagogical Improvements

- Students more engaged
- Students learn 5 other perspectives
- Students can inherently judge their own performance
- General belief that WHOOPPEE grades were more valid than using traditional grading
- Outliers can be explained through data



Student (dis)satisfaction

- Initial uneasiness with peer evaluation
- I don't trust other students!
- It's more work for the student!
- It's less work for the teacher!
- Messaging is very important.
- “Black Box” Grading



To Build Or Not To Build?

Does the experiment have legs?

Does it match goals?

Is the upside > downside?

Will you have the right people in the room?

Will they buy in?



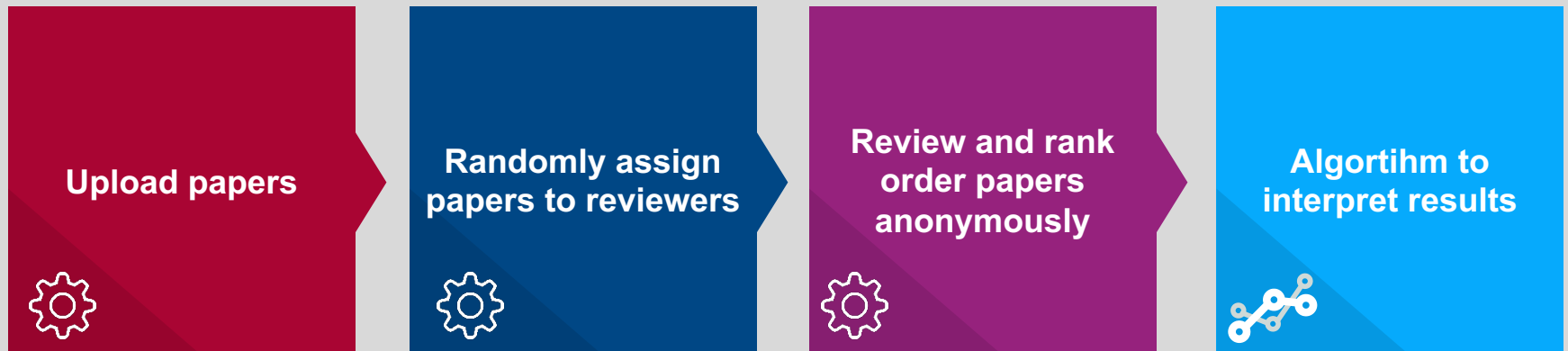
Experimental Tech in a Real-World Setting

- Strong short-term focus through a long-term lens
 - Full inventory of functionality
 - Identify the most critical pieces of functionality
 - Use manual effort for the remainder until it's no longer an experiment
 - All the while noting the pieces necessary for later scaling. Iterative learning as the experiment proceeds.



Innovation often requires a shallow depth of field at first

Key Functionality



Mapped to Current Skills/Tools

Upload papers



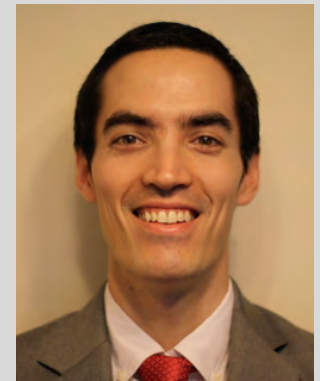
Random assignment
of papers to
reviewers



Review and rank
order papers
anonymously



Algorithm to
interpret results



Best Short-Term Solution

- Canvas for paper submission and general assignment process
- TurnItIn Peermark for anonymous rankings and assignments
- Algorithm completely separate
- Results imported back into Canvas



Even Ella is giving some side-eye to our initial tradeoffs

Minding the Gaps

Documentation!

In-class face time

Behind the scenes effort

Joke contest as practice:



Algorithm

- Variation of the Bradley-Terry-Luce model
- Runs thousand of random comparisons to find the likelihood of finding true scores based on the data, given the parameters of the model
- R with C++ libraries for speed
- 130 students takes ~4 hours to compute
 - Can be sped up with fewer iterations
 - Running the stochastic optimization iterations in parallel would increase speed even further

<http://www.cs.cmu.edu/~jkbradle/papers/shahetal.pdf>

```
300 r.log.new=function(w,assign,a,b,gold=0){
301   #calculate the (log) likelihood for the given rank in a
302   # When w is J x 1 matrix, a,b are scalars, output is th
303   nr=nrow(assign)
304   n.g=ncol(assign)
305   J=length(w)

   assign=as.matrix(assign,nrow(assign),n.g)
   a=matrix(a,1,1)
   b=matrix(b,1,1)
   if(gold==0){g=log(1+w)**a+rep(1,J)**b}
   if(gold==1){g=log(1+rep(max(w),nr)*1.0)**a+rep(1,nr)}
   highest_val = as.numeric(a)*log(2)+as.numeric(b)
   g[g > highest_val] = highest_val
   g=t(g)
   r=matrix(w,J,1)
   grade=matrix(w[assign,nr,n.g] #no byrow=TRUE becau
   gamma=exp(c(g)*w.grade)
   rowSums(log(gamma[, -n.g]))-rowSums(log(t(apply(gam
   n(re)

a,b) when there are bad reviewers.
new=function(w,assign,a,b,bads=NULL){
  the (log) likelihood for the given rank in a
  J x 1 matrix, a,b are scalars, output is th
  numbers in 'assign' associated with reviewe
  this function will not be used for 'assign'
  gn)
  gn)

  ix(assign,nrow(assign),n.g)

  rep(1,J)**b

  numeric(a)*log(2)+as.numeric(b)
  = highest_val

  ign],nr,n.g) #no byrow=TRUE becau
  e)
  -n.g]))-rowSums(log(t(apply(gam
  26
```

Initial Spring 2015 Run

- ~150 students across two sections
- No grades were changed from the algorithm
 - All outliers were explainable
 - Detailed information on algorithm and specific scores provided to students
 - General belief that WHOOPPEE grades were more valid than in previous years using traditional grading
- Student Survey (N=87):
 - 50.5% were confident their work was accurately assessed
 - 93% felt peer review improved their understanding of concepts
 - 46% significantly improved

2015A (Spring 2015)

Home

Pages

Announcements

Assignments

Discussions

People

Class Recordings

Quizzes

Grades

References

Files

Resources

View All Pages

Project #1

Project #1 Guidelines and A

Project #1 Observations (fro

WHOOPEE overview and t

Project 1 grading process

WHOOPEE Instructions for

Optional reading (details abo

Shah Wainwright et al - A C

Where we are now

Second clip:

http://players.brightcove.net/13421203001/S1ZGqh81I_default/index.html?videoid=5178888069001



QUESTIONS?

A large blue sign for the Wharton School of the University of Pennsylvania. The sign features the Wharton crest on the left, which is a shield with a red top section containing a white building and a blue bottom section containing three white circles. To the right of the crest, the word "Wharton" is written in a large, white, serif font. Below "Wharton", the words "UNIVERSITY of PENNSYLVANIA" are written in a smaller, white, serif font, with "of" in a script font.

Wharton
UNIVERSITY of PENNSYLVANIA



Wharton
UNIVERSITY *of* PENNSYLVANIA