Research and Development on MOOC Assessment

MOOC assessment interest group
Hoi Suen, Mindy Kornhaber, Hongli Li, Steve Shaffer, Yao Xiong, Deborah Goins, Wik-hung Pun, Xiaojiao Zang, Kyung Sun Chung, Mehmet Baydu, Michael Cook, Youngsun Lyu
Pennsylvania State University
Key problems for online student assessments within MOOC

• For online formative evaluation (assignments, progress check): Scalability and management.

• For online summative evaluation (certification, badges): Test security
Solution for MOOC summative student assessment

In spite of advances in technology (keystroke authentication, biometrics, remote proctoring, remote control of student computers, etc.), the only relatively secured method of summative assessment to date is to administer tests in controlled, secured, and proctored test centers.
Scalable formative assessments within MOOC

- Multiple-choice quizzes/tests
  - Machine grading

- Written assignment
  - Peer review
Peer Assessment within MOOC

• “Crowd-sourced commentary”
  e.g., Post/responses in course forums

• “Calibrated peer review (CPR)”
Calibrated Peer Assessment (CPA)

• Use known-quality assignments for calibration
• Evaluate rating accuracy of individual rater
• Peer ratings are weighted by rater accuracy
Typical CPR Procedures: 1. Student A submits writing assignment

Retrieved from:
http://cpr.molsci.ucla.edu/Images/CPRFlowchart.jpg
Typical CPR Procedures: 2. Calibration of accuracy of Student A as a rater
Typical CPR Procedures: 3. Peer Reviews and weighted scores

1. Calibrations due date

2. Review the instructor's assessment of the calibrations, with explanations for assessment

3. Use the same rating scale to review and evaluate three texts submitted by your peers

4. Review your own text using the same criteria

5. View your score and read feedback from the students who reviewed your text and the texts that you reviewed

End of assignment
Problems of Existing CPR

• Anecdotal evidence: Not accurate nor useful.
• Some potential reasons:
  – No evidence of reliability or validity
  – Poor rubrics
  – Halo effect
  – Nuisance chore and random or systematic grading
  – Within-rater, between-assignment transferability of accuracy
  – Within-rater consistency/certainty
Credibility Index for Peer Assessment
A Proposed Credibility Index

**Accuracy**: similarity between the peer rater’s score and that of the instructor.

**Consistency**: stability of rating scores within the same peer rater for the same assignment.

**Transferability**: maintenance of the same level of accuracy from assignment to assignment

**Credibility**
A Proposed Credibility Index: Refine the last step of CPR

2 Fake assignments
- Student overall rating \((S)\)
- Student maximum/minimum rating \((S_{\text{max}}/S_{\text{min}})\)
- Instructor rating \((I)\)

4 Real assignments
- Student overall rating \((S)\)
- Student maximum/minimum rating \((S_{\text{max}}/S_{\text{min}})\)
Benefits of the Credibility Index

• More accurate/useful feedback
• Potential unobtrusive way of evaluating students’ mastery based on their credibility index value as a supplemental summative evaluation piece.
• Use of credibility to rank comments in discussion forums – instead of popularity – to improve crowd-source commentaries
Research Questions

• Feasibility, proof of concept
• Effectiveness of CI
• Relative reliability/validity compared against CPR
• What combination of the three components (accuracy, consistency, and transferability) produce the optimal solution for rater credibility?
• How to create a cumulative and trustworthy credibility index in the long run?
Language and Quality of Peer Assessment
2.8 million learners have registered for Coursera MOOCs as of March 2013 with the following geographical distribution:

- U.S.: 28%
- India: 9%
- Brazil: 5%
- U.K.: 4%
- Spain: 4%
- Canada: 4%
- Australia: 2%
- Russia: 2%
- Other: 42%

Retrieved from:
http://www.scientificamerican.com/article.cfm?id=massive-open-online-courses-transform-higher-education-and-science
First language of registrants

• While over half of the registrants were from English-speaking countries, a very large minority were from non-English-speaking countries.

• Most MOOCs to date are in English and writing assignments are to be submitted in English.

• Peer assessments are conducted in English.
Language and Peer Assessment

Raters
- ESL students
- English-speaking students

Writers
- ESL students
- English-speaking students
Research Questions

• What are the effects of English proficiency and cultural background on the quality of peer assessment outcomes (impact of language on credibility index values)?

• How to achieve fair peer assessments when assignments are rated by raters with different language and cultural backgrounds?
Factors Related to MOOC Completion
MOOC Completion Issues

Completion is typically <10%. Therefore:

• Cost of MOOC development is very high, esp. relative to potential knowledge and skill gains

• Large loss of any potential revenue stream associated with e.g., completion certificates, etc.
MOOC Completion Rate

Retrieved from: http://www.katyjordan.com/MOOCproject
What Accounts for Attrition on MOOC

Macro-level

Infrastructure level
• MOOC platform, Internet, etc.

Instruction level
• Course contents, instructor, assessment method, etc.

Learner level
• Motivation, SES, language, time, learning behaviors, etc.

Micro-level
Research Questions

• Develop an effective taxonomy of patterns/reasons of non-completion.
• Profiles of different types of learners?
• What variables predict the probability of a learner’s completion on a MOOC?
Auto-Generated High-Order Exam Question
Issues

• One additional tool toward an ultimate secured online summative assessment is to generate exam questions only at the moment of assessment (just-in-time exam).

• Computer-generated exams to date have primarily led to items that tap low-level cognition (e.g., recall items, declarative knowledge, CLOZE tests, MAZE tests).
A Proposed System to Automate Higher-level Question Generation

Relational statements in instructional text or concept maps as bases

(1) natural language processing (NLP)
(2) coherence theory
(3) evolution-based algorithms.

Just-in-time high-level exam questions
Research Questions

• Feasibility of the system?
• Reliability and validity of results from exams generated via the system.
Thank you