Uncovering Reviewing and Reflecting Behaviors From Paper-based Formal Assessment

Sharon, I-Han Hsiao*
Po-Kai Huang
Hannah Murphy

Arizona State University

2017/03/16
Arizona State University
by 2016 Fall: 20,366 in CS
Data-Driven Engineering to Improve Computing Education
Agenda

- Problems
- Related work
- **WPGA** (Web-based Programming Grading Assistant)
- Behavior Modeling Evaluation Results
Problems

• Challenges in providing feedback for **large scale** and **paper-based** assessments.
  
  o Hand written feedback is time consuming.
  o Graders’ inconsistency.
  o **Limited feedback**.
  o **Advanced learning analytics are not captured**.
  o Delivering logistics is tremendous painful.

![Comic Strips](https://www.phdcomics.com)
Do students focus on their final score or do they put in the effort to review their returned test?

When they do review, how does it associate with their learning?
WPGA (Web-based Programming Grading Assistant)

1. Streamlining the Documentation Process of Paper-based Assessments
2. Augmented Grading & Feedback-giving Interfaces
3. Reflective Feedback Delivery
1. Streamlining Digitization Process

• Past:
Teachers make paper exams → proctor exams in class → collect them back → (Distribute to graders) → grading → (collect them back) → deliver to students → (collect them back) (missing advanced learning analytics)

• Current:
Teachers make paper exams → proctor exams in class → collect them back → feed papers in scanner, upload scanned paper images to WPGA → (Distribute to graders) → grading → deliver to students by publishing graded results on system (harnessed advanced learning analytics)
1. Streamlining Digitization Process

Past:
- Teachers make paper exams
- Proctor exams in class
- Collect them back
- Grading
- Collect them back
- Deliver to students

Current:
- Teachers make paper exams
- Proctor exams in class
- Collect them back
- Grading
- Collect them back
- Deliver to students results on system
- (Harnesses advanced learning analytics)
- Feed papers in scanner, upload scanned paper images to WPGA
- (Harnesses advanced learning analytics)
2. Augmented Grading & Feedback-giving Interfaces

- Draw & Type to give feedback
- Click grading-rubric buttons to deduct points
3. Reviewing & Reflecting UI

Q. Write a Java class called Book. (This class will NOT have a main method.) The class Book will have instance variables “bookId” (int) and “title” (String). Provide a constructor that initializes bookId to 0 and title to “F”, and also provide another constructor that takes two parameters, an integer (for bookId) and a String (for title) and assigns those values to the two variables. Also provide accessor and mutator methods for both variables. Finally, provide toString() method to return a string containing these two variables.

(For instance, toString) method can return a string containing:
Book Id: 10000
Title: Learning Java

public class Book
{
    private int bookId;
    private String title;
    
    public Book()
    {
        bookId = 0;
        title = "F";
    }
    
    public Book(int bookId, String bookTitle)
    {
        bookId = bookId;
        title = bookTitle;
    }

    public void setBookId(int id)
    {
        bookId = id;
    }

    public void setTitle(String title)
    {
        this.title = title;
    }

    public int getBookId()
    {
        return bookId;
    }

    public String getTitle()
    {
        return title;
    }

    @Override
    public String toString()
    {
        return bookId + "", bookTitle + "": 
    }
}
Classroom Study & Data Collection

• In 2016 Fall semester
  o 35 graders (3 instructors, 32 student graders)
  o 6 computing courses across 100~400 levels.
  o 1198 students

• This paper, we looked at 1 course (200 level), 232 students, 6 quizzes, 1 exam.

• 12 distinct operations: exam and question clicks, filter clicks, bookmarks, notes, see the correct answer clicks etc.
Results – (1)

- Students review quizzes regardless of if they count towards academic performances or not.

<table>
<thead>
<tr>
<th>View Avg %</th>
<th>quiz1</th>
<th>quiz2</th>
<th>quiz3</th>
<th>quiz4</th>
<th>quiz5</th>
<th>exam1</th>
<th>quiz6</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.5</td>
<td>36.6</td>
<td>47.6</td>
<td>36.9</td>
<td>24.7</td>
<td>80.9</td>
<td>18.8</td>
<td></td>
</tr>
</tbody>
</table>
Results – (2)

- Hard working reviewers perform better than the others.
  - Reviewer vs. Non-reviewer (91.64±6.68 vs. 83.54±16.46)
Modeling Sequential Review Strategy using Hidden Markov Model

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review</td>
<td>Exam (E)</td>
</tr>
<tr>
<td></td>
<td>Correct Question (C)</td>
</tr>
<tr>
<td></td>
<td>Incorrect Question (I)</td>
</tr>
<tr>
<td></td>
<td>Filter (F): locating question</td>
</tr>
<tr>
<td>Reflect</td>
<td>Reflect (R) : notes, bookmark, checked understanding</td>
</tr>
</tbody>
</table>

Assumptions:
R(E)->R(C)->R(C)->R(I)->Reflect
R(E)->R(I)->Reflect
R(I)->Reflect
Results – (3)

• Both A & B students perform review on **overview first, detail on demand**.
  
  o They look at exam scores first, then review questions.
\( \pi : \) start state  
\( E : \) exam  
\( C : \) correct answer  
\( I : \) incorrect answer  
\( R : \) reflection  
\( F : \) filter
Results – (4)

- A students review & reflect strategically: they strive to get the wrong right.
- B students review persistently, but fail to engage in deeper reflection.

Lack-of-deep-reflection behaviors can be dangerous.
Summary

• Educational technology to capture reviewing & reflecting behaviors and to support grading.
• Connect physical to digital learning analytics.
Ongoing

Future Work

• Strengthen the link between Assessment and Instruction: capture and push for meaningful “reflection”.
  o Social Feedback
  o Recommendation
  o Visual Analytics: visualize what matters! Connect with learning by doing behavior data!
  o Systematic longitudinal study

• Less than a year, 6 professors, ~2300 student users, across all level CS courses.
• Math & Statistics are set to use WPGA in summer.
• Google Sharon Hsiao, WPGA info is on my site.
# What’re differences between WPGA & Gradescope?

<table>
<thead>
<tr>
<th></th>
<th>WPGA</th>
<th>Gradescope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digitalization Process</td>
<td>Scan physical papers and connect to digital world</td>
<td></td>
</tr>
<tr>
<td>Tech</td>
<td>Type in, QR-code</td>
<td>Type in</td>
</tr>
<tr>
<td>Theoretical</td>
<td><strong>Engaging student learning activity</strong>, make grading easy.</td>
<td>Grading makes easy</td>
</tr>
<tr>
<td>Deployment</td>
<td>Free, making it open source; You own the data; Controllable experiment for adaptive technologies*</td>
<td>Free, but not open source</td>
</tr>
<tr>
<td>UI</td>
<td>We are working toward VLA too.</td>
<td>Visual learning analytics</td>
</tr>
</tbody>
</table>

* Semi-supervised hand-writing recognition to facilitate automated grading (Hsiao, 2016);
* Semantic partial credit algorithm (Hsiao, 2016; Hsiao, Huang & Murphy, 2017)
Thank you 😊
Sharon.Hsiao@asu.edu