Assessing Online Computer Coaches for Problem Solving: Educational Impact

Presented by: Evan Frodermann\(^1\)

Qing (Xu) Ryan\(^{1,3}\), Ken Heller\(^1\), Leon Hsu\(^1\), Jia-Ling Lin\(^1\), Judy Hill\(^1\), Bijaya Aryal\(^2\), Koblar Alan Jackson\(^4\)

1. University of Minnesota–Twin Cities
2. University of Minnesota–Rochester
3. University of Colorado–Boulder
4. Central Michigan University

AAPT Summer 2014 Meeting
Minneapolis, MN

Supported by NSF DUE #0715615 and DUE-1226197, and by the University of Minnesota
Research Question

• Do the computer coaches improve students’ problem-solving?

Version 1 Summary

• 3 sections of a university calculus-based introductory mechanics course
  • Fall 2011: One section of 221 students
    • Students could complete their homework using WebAssign or the coaches.
  • Spring 2013: Two sections of 148/103 students.
    • Although coaches were available to help with some problems, students were required to complete their homework using WebAssign.
Final exam scores Sp. 2013

<table>
<thead>
<tr>
<th>Final Exam</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (0-20%) Coach Use (N=72)</td>
<td>70%±3%</td>
<td>71% ± 6%</td>
<td>70% ± 3%</td>
</tr>
<tr>
<td>Medium (40-60%) Coach Use (N=38)</td>
<td>65% ± 4%</td>
<td>68% ± 4%</td>
<td>66% ± 3%</td>
</tr>
<tr>
<td>Heavy (80-100%) Coach Use (N=49)</td>
<td>73% ± 2%</td>
<td>68% ± 4%</td>
<td>71% ± 2%</td>
</tr>
</tbody>
</table>

- Heavy users ≈ Light users in final exam
  - However, the confidence and FCI score of the students started lower as shown in previous talk.
  - Need to examine this closer.
Matching Historical Data

• 4 to 1 match of historical students to Sp2013 students.
  – Matched on pre-class FCI, gender, expected grade, and expected study time.
    • ~85% student perfect match. 3145 students from Spring 2008 to Fall 2011.

• Normalize the exam scores.
  – Exam differences absorbed by equating the exam averages.
  – Retained the exam distributions.

• Avoided unintended selection bias or specific class bias into the results by using varied selection rules.
- Low-use characterized students in historical match predict there should be a significant difference (11%±2%) as compared to the high-user characterized students.
- Coached group shows no significant difference (0%±4%) between heavy and low user groups in coached class from Sp2013.
How to assess the impact of the coaches?

- Other tools to assess student problem-solving?

- TA grading vs. rubric assessment
  - Rubric from Jennifer Docktor (2009)
    - J Docktor et al, PERC proceedings (2009) AIP
    - Built upon previous work: P. Heller et al, Vol 60 #7, AJP (1992)
  - 5 categories assessed on scale 0-5
    - Useful Description, Physics Approach, Specific Application, Mathematical Procedures, Logical Progression
The TA grades are highly correlated with independently assessed rubric scores.

Note: Our TAs have gone through TA training in the Minnesota model but not the rubric assessment.
Test differences controlled through a match of low users from Spring 2013 to the Fall 2011 control class low user characterization.

No perceived difference in low user characterization students.

The difference in heavy use characterization still present in Fall 2011.
Summary

• Coaches improved problem solving scores
  • Near full letter grade difference
  • Less prepared, less self confident student characteristics
  • More female than the rest of the class (about 1/3 of the class)
  • $70\% \pm 3\%$ Coached heavy users; $61\% \pm 2\%$ Non-coached matched

• For TAs with our TA support, grading is a good indication of problem solving skills.
  • High correlation between TA grading and Rubric assessment

• For more information see:
  • Talks: GC09/10, FD08/09 (Wednesday)
  • Posters: PST2C13/14/15 (Tuesday)