Effects of Workshop Group Gender Balance on Student Exam Performance

Update on replication studies: 2014-2018

Ross Galloway
School of Physics & Astronomy
University of Edinburgh, U.K.
Context

• Physics 1A class (introductory Newtonian mechanics)
• First-year, first-semester class
• Typical demographics:
  • ~ 270-310 students
  • ~ 75:25 male:female gender balance
  • ~ 50:50 majors:non-majors balance
• All non-majors possess Physics degree entry qualifications
• Taught in ‘Flipped Classroom’ format
Teaching format

• 11-week semester. Each week:
  • 3 × 1-hour Peer Instruction based lecture
  • 1 × 3-hour workshop in teaching studio room
    ▪ Workshops taught in 4 sections: Mon, Tue, Thu, Fri
    ▪ Students randomly allocated to tables of 5 or 6
    ▪ Demographics ⇒ typical group contains 1-2 females
Intervention

- Monday and Thursday sections (‘Balanced’):
  - Where female students are present in group, $\geq \frac{1}{2}$ of students must be female (typically 3 of 6)
  - Remainder of groups are all-male
  - Avoid drawing student attention to what we’re doing
- Tuesday and Friday sections (‘Random’):
  - Purely random allocation as before $\Rightarrow$ ‘control population’
  - Typically 1 of 6 female students in each group
- No other interventions at all
2014-15 results

Controlled Final Exam Score Comparison

Male

Female

Mean Exam Score / %

Balanced

Random

2.3% p=0.250

6.0% p=0.042 (over half a letter grade higher)
2015-16 results

Controlled Final Exam Score Comparison

Male  Female
Mean Exam Score / %
Balanced  Random

Balanced
Tue  Fri
Random
Mon  Thu
2016-17 results

Controlled Final Exam Score Comparison

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>Random</td>
<td>65</td>
<td>60</td>
</tr>
</tbody>
</table>

Balanced: Mon, Fri
Random: Tue, Thu
2017-18 results

Controlled Final Exam Score Comparison

Mean Exam Score / %

- Male: Balanced vs Random
  - Balanced: 58%
  - Random: 51%
  - Difference: 6.6%
  - p=0.16 (not sig.)

- Female: Balanced vs Random
  - Balanced: 54%
  - Random: 51%
  - Difference: 3%
  - p=0.16 (not sig.)
Conclusions

• 2014-15 findings have not been replicated in three subsequent years.

• What’s going on here?
  • Black swan event in 2014-15?
  • Hawthorne Effect?

• If there is a mechanism, it’s very fragile.
  • And we don’t understand it.

• Highlights importance of replication studies.

Acknowledgements: Ross Hunter, James Salmon, Rodger Watt, Sander Roomus