Ability Grouping in ESL: Beneficial or Detrimental?

Chris Sheppard (Waseda University, Tokyo, Japan)
Emmanuel Manalo (Waseda University, Tokyo, Japan)
Marcus Henning (The University of Auckland, New Zealand)
Ability Grouping

• Assignment of students to different classes according to outcome measures of their general ability or attainment in particular subject areas

• Aka: “streaming” or “tracking”

• Has been used in educational settings for nearly a century

• But remains controversial: many arguments for and against it

• Reviews of research about the practice consistently fail to produce unequivocal support for its efficacy

• Still, it is widely practiced and often considered necessary, especially in the field of language teaching
Arguments for Ability Grouping

• By utilizing methods and materials appropriate for students at particular levels of achievement, ability grouping has been claimed to facilitate students’ learning development at their own rate – with others who have similar abilities.

• For teachers, the narrower range of abilities to cater for can supposedly lead to enhanced efficiencies in both materials preparation and teaching, and to making class management easier – including encouragement of student participation.
Arguments Against Ability Grouping

• May present problems relating to student attitudes in class (inferiority/superiority depending on group); for students of lesser ability, poses not only the danger of negative stigma, but also deprivation of opportunities to learn with/from students of greater ability

• May render teachers less sensitive to individual differences because of erroneous assumptions about the extent of student similarity along social needs, achievement, learning style, and other dimensions

• Administration also poses many challenges, including high resource demand
Reviews

• Yates and Pidgeon (1959): “it is possible neither to justify the criticisms that have been leveled against it, nor to prove that streaming is a desirable and effective form of organization”

• Esposito (1973): “among studies showing significant effects, the slight gains favoring high ability students is often more than offset by evidence of unfavorable effects on the learning of students of average and below average ability, particularly the latter”

• Slavin (1987): some benefits in grouping plans that involve cross-grade assignment in only one or two subjects; no evidence of benefits in assigning students to self-contained classes

• Hanushek and Wößmann (2006): consistent outcomes of increases in achievement inequality as a consequence of early tracking
More Recent Reports

- Hattie (2009): Synthesis of over 800 meta-analyses revealed that surface indicators of teaching variables (e.g., open instruction, multi-grade classes, ability grouping) were poorer at predicting student achievement compared to indicators of deeper understanding (e.g., mastery learning, formative assessment).

- Robert (2010): Analysis of PISA (Program for International Student Assessment) results from 23 OECD (Organization for Economic Growth and Development) countries showed better performance from more selective schools, but ability grouping does not improve student achievement.
Use in Language Teaching

- Language proficiency continues to be the single most important grouping criterion in most language teaching institutions (Nunan, 1988)

- Believed to be facilitative of the teacher being able to get a clearer idea of the general level of the students in his/her class, and adjust the curriculum and speed of teaching to suit (Hanna, 1968)

- Teachers and students believe that ability grouping enables the provision of appropriate lessons for different levels of students (Kim, 2012)

- Grouping ESL students according to their oral proficiency levels promotes equality in opportunity and responsibility in speaking (Rance-Roney, 2010)
English Language Education in Japan

• Considered very important
  • Included in Japan’s Goals in the 21st Century: “concrete objective of all citizens acquiring a working knowledge of English by the time they take their place in society as adults”
  • In 2003, the Japanese government proposed an action plan to cultivate “Japanese with English Abilities”

• But Japan continues to score lower than most other Asian countries on tests like the TOEFL (Test of English as a Foreign Language) (Silver, Hu, & Iino, 2002)

• Despite the Japanese government’s various initiatives to include English education in the school curriculum, and the overall high uptake of English language courses in the population as a whole, progress is poor
Importance of English Scores

- English entrance examination scores and/or scores in English tests like the TOEFL are used in making acceptance decisions and in placing students in any compulsory English courses
  - English scores are believed to be indicative of students' analytical and logical thinking skills; usually allocated a very high weighting (Butler & Iino, 2005)

- Despite the widespread use of ability grouping in assigning Japanese university students to their English language classes, there appears to be no research to find out if it is in fact useful
The Present Study

• Main question: Is there any evidence that ability grouping is beneficial for ESP students of science and technology in a Japanese university?

• Are there any differences based on the level of the students?
Participants

• Participants
  • Drawn from our cohort of 8500 Japanese EFL students
  • 1 year university Science and Engineering Students
  • Learning English for Specific (Scientific) Purposes

• Ability Grouping
  • Currently Students are placed into ability groups based on their English proficiency (As measured by their TOEIC scores) and their slot in the timetable which they are available to take the English classes.
  • There is no adjustment of the course goals. Individual teachers are encouraged to adapt the materials and the aims of the teaching
Measuring Ability Grouping

Placement
- All students in our program are placed by their TOEIC scores. We cannot compare placed and unplaced students.
- Instead we compared students in high variation classes and low variation classes.
  - Different time slots have different numbers of students assigned to them.
    - Minimum of 56 students → Maximum of 304 students.
    - Resulting in a difference in the variability of TOEIC scores depending on the class.
- High degree of placement.
  - Students placed in classes with a lower variability (standard deviation) in TOEIC scores.
- Low degree of placement.
  - Students placed in classes with a higher variability (standard deviation) in TOEIC scores.
Measuring Improvement

• Improvement was measured by:
  • Improvement in TOEIC scores
    • The difference between each participants placement TOEIC (IP) score and their score at the end of the first semester course
      (Due to a change in policy, the improvement of students who enrolled after 2011 was calculated from the difference between their placement and end of year TOEIC score)
  • Course grades
    • The grade assigned to each student by their course teacher
    • Coded based on the University GPA standard
      • A+ → 4, A → 3.5, B → 3, C → 2, F → 0
Procedure

- TOEIC IP placement test
- Students placed based on their TOEIC scores
- 14 weeks of 90 min class held once per week
- Graded by course teachers
- TOEIC IP end of semester (end of year) test
Analysis

• Created 8 groups of differing proficiency levels
  • <295 and then every 100 points to 990

• Select “high” and “low” placed participants
  • Bottom 33% class standard deviation = “high” placement
  • Top 33% class standard deviation = “low” placement

• High students and low students were then matched and paired based on their placement TOEIC scores.
Analysis

- Create pairs with the same TOEIC placement scores

<table>
<thead>
<tr>
<th>TOEIC score</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>170-295</td>
<td>164</td>
</tr>
<tr>
<td>300-395</td>
<td>646</td>
</tr>
<tr>
<td>400-495</td>
<td>322</td>
</tr>
<tr>
<td>500-595</td>
<td>324</td>
</tr>
<tr>
<td>600-695</td>
<td>326</td>
</tr>
<tr>
<td>700-795</td>
<td>158</td>
</tr>
<tr>
<td>800-895</td>
<td>56</td>
</tr>
<tr>
<td>900-990</td>
<td>58</td>
</tr>
</tbody>
</table>
Analysis

- Pairwise t-tests were used to determine the degree of difference between low and high placement.
- In order to compare the degree of difference, Cohen’s D was calculated for each proficiency band.
### Results

#### Change in TOEIC Scores – Descriptive Statistics

<table>
<thead>
<tr>
<th>Mean TOEIC Score</th>
<th>High Placement</th>
<th>Low Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>260.55</td>
<td>74.88</td>
<td>81.23</td>
</tr>
<tr>
<td>349.54</td>
<td>46.19</td>
<td>73.59</td>
</tr>
<tr>
<td>452.43</td>
<td>18.23</td>
<td>75.65</td>
</tr>
<tr>
<td>542.58</td>
<td>5.38</td>
<td>87.92</td>
</tr>
<tr>
<td>633.06</td>
<td>7.02</td>
<td>75.65</td>
</tr>
<tr>
<td>736.65</td>
<td>-8.10</td>
<td>72.10</td>
</tr>
<tr>
<td>835.00</td>
<td>1.96</td>
<td>39.85</td>
</tr>
<tr>
<td>941.72</td>
<td>-8.10</td>
<td>27.20</td>
</tr>
</tbody>
</table>
• **Change in TOEIC Scores – Comparisons between high and low placement**

<table>
<thead>
<tr>
<th>Mean TOEIC Score</th>
<th>t-value</th>
<th>df</th>
<th>p-value</th>
<th>Cohen's D</th>
<th>Cohen's D (standard error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>260.55</td>
<td>0.24</td>
<td>81</td>
<td>0.81</td>
<td>0.03</td>
<td>0.11</td>
</tr>
<tr>
<td>349.54</td>
<td>0.84</td>
<td>322</td>
<td>0.40</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>452.43</td>
<td>-1.24</td>
<td>180</td>
<td>0.21</td>
<td>-0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>542.58</td>
<td>-3.44</td>
<td>181</td>
<td>0.00</td>
<td>-0.25</td>
<td>0.07</td>
</tr>
<tr>
<td>633.06</td>
<td>-2.63</td>
<td>182</td>
<td>0.01</td>
<td>-0.19</td>
<td>0.07</td>
</tr>
<tr>
<td>736.65</td>
<td>-3.05</td>
<td>78</td>
<td>0.00</td>
<td>-0.34</td>
<td>0.11</td>
</tr>
<tr>
<td>835.00</td>
<td>-1.54</td>
<td>27</td>
<td>0.13</td>
<td>-0.29</td>
<td>0.19</td>
</tr>
<tr>
<td>941.72</td>
<td>-0.10</td>
<td>28</td>
<td>0.92</td>
<td>-0.02</td>
<td>0.19</td>
</tr>
</tbody>
</table>
## Results

### Grades – Descriptive Statistics

<table>
<thead>
<tr>
<th>Mean TOEIC Score</th>
<th>High Placement Mean</th>
<th>Std. Dev</th>
<th>Low Placement Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>261</td>
<td>2.17</td>
<td>1.13</td>
<td>2.53</td>
<td>0.84</td>
</tr>
<tr>
<td>350</td>
<td>2.76</td>
<td>0.82</td>
<td>2.79</td>
<td>0.69</td>
</tr>
<tr>
<td>452</td>
<td>2.96</td>
<td>0.67</td>
<td>3.02</td>
<td>0.71</td>
</tr>
<tr>
<td>543</td>
<td>3.08</td>
<td>0.43</td>
<td>3.38</td>
<td>0.56</td>
</tr>
<tr>
<td>633</td>
<td>3.19</td>
<td>0.60</td>
<td>3.49</td>
<td>0.64</td>
</tr>
<tr>
<td>737</td>
<td>3.35</td>
<td>0.59</td>
<td>3.78</td>
<td>0.27</td>
</tr>
<tr>
<td>835</td>
<td>3.32</td>
<td>0.49</td>
<td>3.77</td>
<td>0.76</td>
</tr>
<tr>
<td>942</td>
<td>3.88</td>
<td>0.22</td>
<td>3.97</td>
<td>0.13</td>
</tr>
</tbody>
</table>
## Results

- **Grades – Comparisons**

<table>
<thead>
<tr>
<th>Mean TOEIC Score</th>
<th>t-value</th>
<th>df</th>
<th>p-value</th>
<th>Cohen's D</th>
<th>Cohen's D (Standard Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>261</td>
<td>-2.34</td>
<td>81</td>
<td>0.02</td>
<td>-0.26</td>
<td>0.11</td>
</tr>
<tr>
<td>350</td>
<td>-0.64</td>
<td>322</td>
<td>0.52</td>
<td>-0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>452</td>
<td>-0.85</td>
<td>180</td>
<td>0.40</td>
<td>-0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>543</td>
<td>-6.18</td>
<td>181</td>
<td>0.00</td>
<td>-0.46</td>
<td>0.07</td>
</tr>
<tr>
<td>633</td>
<td>-4.39</td>
<td>182</td>
<td>0.00</td>
<td>-0.32</td>
<td>0.07</td>
</tr>
<tr>
<td>737</td>
<td>-5.85</td>
<td>78</td>
<td>0.00</td>
<td>-0.66</td>
<td>0.11</td>
</tr>
<tr>
<td>835</td>
<td>-2.50</td>
<td>27</td>
<td>0.02</td>
<td>-0.47</td>
<td>0.19</td>
</tr>
<tr>
<td>942</td>
<td>-1.72</td>
<td>28</td>
<td>0.10</td>
<td>-0.32</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Results

- Cohen's D and mean TOEIC score for Change in TOEIC score and Grade

![Graph showing the relationship between Average TOEIC score at placement and Effect Size (Cohen's D). The graph includes data points for Grade and TOEIC Change.]
Results

- Correlation between Average TOEIC placement score and Cohen’s D (the degree of difference between low and high placement)
  - TOEIC Improvement
    - \( r = -0.92, p = 0.003 \)
  - Grade
    - \( r = -0.73, p = 0.062 \)
Discussion

• Degree of placement has an effect on both final course grade and the improvement in TOEIC score

• However, the extent of this effect depends on the proficiency level of the student when they begin the course
  • This is negatively linearly related
  • Low level students (less that 328 on their TOEIC placement test) are more likely to benefit from a higher degree of placement
  • Average to High level students (greater than 328) are more likely to benefit from a lower degree of placement
Discussion

• Previous claims about placement effects (i.e., more beneficial for higher level students, less beneficial for lower level students) are not supported
  • Opposite effect is suggested by the results: possibly beneficial for lower level students, but detrimental for higher level students

• Possible reasons and explanations need to be considered

• Future directions for research
References


