Does a high score from a judge mean that the Technovation student has learned more?

As we scale and grow, we want to make sure our programs continue to have deep impact on participants. A step in that direction is to look at the data we collect, and identify patterns. One such area that we analyzed this year was comparing the scores that judges provide to Technovation students' work to the students' self-reported learning gains.

Technovation is the world's largest technology entrepreneurship competition for girls aged 10-18. Through Technovation, girls identify a problem in their community, develop a mobile app and launch a startup that addresses that problem. They are supported through the whole process by mentors (who could be educators, industry professionals or parents). At the end of the 100-hour program, girls submit mobile app prototypes, business plans, app code, a pitch video, and an app demo video. Judges review these materials online and provide a score.

We were interested to determine if there was any connection between the judges' scores and the girls’
self-reported gains. Our hypothesis was that girls who made significant learning gains would also be scored highly by the judges. However, what we found is that so far, there is no correlation between the two (except for a weak relationship in 2013 indicating that in this year the judges' scores were high and that girls also reported high learning gains).

![Correlation Table](image)

<table>
<thead>
<tr>
<th>Year</th>
<th>n</th>
<th>Correlation Coefficient (r)</th>
<th>Median Judges' Score</th>
<th>Skewness Judges' Distribution</th>
<th>Median Students' Score</th>
<th>Skewness Students' Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>62</td>
<td>0.4310</td>
<td>72</td>
<td>-0.252</td>
<td>77</td>
<td>-0.884</td>
</tr>
<tr>
<td>2014</td>
<td>740</td>
<td>0.1633</td>
<td>72</td>
<td>-0.625</td>
<td>77</td>
<td>-0.467</td>
</tr>
<tr>
<td>2015</td>
<td>615</td>
<td>0.0984</td>
<td>71</td>
<td>-0.539</td>
<td>81</td>
<td>-2.637</td>
</tr>
<tr>
<td>2016</td>
<td>584</td>
<td>0.0443</td>
<td>65</td>
<td>-0.406</td>
<td>78</td>
<td>-1.807</td>
</tr>
<tr>
<td>2017</td>
<td>496</td>
<td>0.0002</td>
<td>69</td>
<td>-0.168</td>
<td>79</td>
<td>-0.640</td>
</tr>
</tbody>
</table>

Judges' Scores were moderately skewed (towards higher values with tails at lower values).

Students' Scores were moderately skewed except for 2015 and 2016, when they were highly skewed.

Our explanation for this is that the students' perceptions of the program's impact is based on the qualitative changes in their self-efficacy; that may be disconnected from the real-world value of their app and business plan.

Over the past 4 years we have tried to make the judging rubric more student-centered (as can be seen from the 2013 and 2017 judging criterion).

### 2013 Judging Criterion

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>App Demo</strong></td>
<td>Does it have adequate functionality?</td>
</tr>
<tr>
<td></td>
<td>Is the app visually appealing?</td>
</tr>
<tr>
<td></td>
<td>Is the app’s user-interface intuitive and easy to use?</td>
</tr>
<tr>
<td><strong>Critical Thinking</strong></td>
<td>Is the app a good solution to a problem in their local community?</td>
</tr>
<tr>
<td></td>
<td>Do they understand the size of their market?</td>
</tr>
<tr>
<td></td>
<td>Do they understand their competition and how they are differentiated?</td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td>Does the team convey their understanding of the computer programming?</td>
</tr>
<tr>
<td></td>
<td>Does the Pitch explain their business plan?</td>
</tr>
<tr>
<td></td>
<td>Is the Pitch clear and concise?</td>
</tr>
<tr>
<td><strong>Strategy &amp;</strong></td>
<td>Do they leverage the capabilities of the platform they are using?</td>
</tr>
<tr>
<td><strong>implementation</strong></td>
<td>Is their app a good representation of their vision?</td>
</tr>
</tbody>
</table>
Do they have a practical vision for extending the capabilities of their apps beyond the prototype?

### 2017 Judging Criterion

<table>
<thead>
<tr>
<th>Ideation (15 points)</th>
<th>Score</th>
</tr>
</thead>
</table>
| 5 pts | The team clearly demonstrates how their app idea aligns with at least one of the 6 Technovation themes of the UN Sustainable Development Goals:  
- Poverty  
- Environment  
- Peace  
- Gender Equality  
- Health  
- Education  
Learn more about these goals at [sustainabledevelopment.un.org](http://sustainabledevelopment.un.org) |
| 5 pts | The team has clearly outlined and provided evidence through facts and statistics of the problem they are addressing. |
| 5 pts | The app effectively addresses the problem that was outlined. |

**Ideation Total Score:**

<table>
<thead>
<tr>
<th>Technical (20 points)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 pts</td>
<td>The app is fully functional and has no noticeable bugs.</td>
</tr>
</tbody>
</table>
| 10 pts | The team has completed and you have verified the [Technical Checklist](#).  
*Please note this score should range from 0 to 10.* |
| 5 pts | The demo video demonstrates the functionality of the app well. |

**Technical Total Score:**

<table>
<thead>
<tr>
<th>Pitch (20 points)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 pts</td>
<td>The team clearly communicates the problem they are addressing.</td>
</tr>
<tr>
<td>5 pts</td>
<td>The team presents a compelling argument to support the solution they are proposing.</td>
</tr>
<tr>
<td>5 pts</td>
<td>The team’s passion and energy for solving the problem is evident in their pitch delivery.</td>
</tr>
<tr>
<td>5 pts</td>
<td>The pitch is specific and to the point.</td>
</tr>
</tbody>
</table>

**Pitch Total Score:**
We have also worked to improve the training provided to the judges, so they have a better understanding of the program the girls are going through. Despite these changes, it is hard for the
judges to not compare the Technovation student products with what they are used to seeing every day in professional environments.

A similar example of this disconnect is from our Curiosity Machine program that is implemented with younger children. Students build hands-on engineering-based designs out of easily available materials such as cardboard and popsicle sticks. An external judge may review their design and assign a low score based on what they think is good design executed with sophisticated materials. However, the student may have made huge gains in understanding how a particular concept is applied, and/or in her own ability to create, troubleshoot and innovate.

This disconnect is interesting to find and explore further as we improve our curriculum, training and judging processes.

**Method of analysis:**

**Survey Scores**
- Responses were collected from individual student post-survey questions related to attitude, expectations, and satisfaction about the Technovation experience.
- A completely positive experience resulted in maximum Likert scores and were normalized to a value of 100.

**Judges’ Scores**
- Team scores from App Judging were assigned to each team member, and were normalized to a value of 100.
- The Judges’ team score was compared to each team members’ post-survey normalized Likert score to give scatter plots and distributions. Skewness was calculated:
  - Between -0.5 to +0.5: approximately symmetric
  - Between -1 to -0.5 or +0.5 to +1: moderately skewed
  - Less than -1 or Greater than +1: highly skewed
2013

**Judges' Scores**
- Skewness = -0.2523
- Median = 72

**Student Self-Assessed Scores**
- Skewness = -0.8843
- Median = 77

Judge Scores compared to Student Self-Assessed Scores
2014

Judges' Scores
Skewness = 0.625
Median = 72

Student Self-Assessed Scores
Skewness = -0.467
Median = 77

Judge Scores compared to
Student Self-Assessed Scores
2015

Judges' Scores
Normalized = 100/55
Skewness = -0.539
Median = 72

Student Self-Assessed Scores
Skewness = -2.637
Median = 77

Judge Scores compared to Student Self-Assessed Scores
2016

**Judges' Scores**
Normalized = 100/57  
Skewness = -0.406  
Median = 72

**Student Self-Assessed Scores**
Skewness = -1.807  
Median = 78

Judge Scores compared to Student Self-Assessed Scores
2017

**Judges' Scores**
Skewness = -0.1685  
Median = 72

**Student Self-Assessed Scores**
Skewness = -0.1685  
Median = 79

**Judge Scores compared to Student Self-Assessed Scores**
**Further Exploration**

We need to further explore the skewness in student scores in 2015 and 2016, compared to other years and identify any particular reasons for it. **We would also like to find ways to build a higher correlation between student self-assessed scores and their paired judge scores.** A high correlation would indicate that students can accurately estimate how they did by understanding the rubric and using a checklist.

In addition, we would like to dig a bit deeper and see if there are any hidden biases in the scoring—exploring connections between scores to language, country, region, culture, etc. Our goal is to make sure the judging process is as fair as possible, and studying the existing data is one way to ensure that.