**Goals of Presentation**
- Provide additional help with assessing NSSE benchmark reports
- Emphasize importance of using & contextualizing effect sizes

**Outline of Session**
- Effect Size Review
- NSSE & Benchmark Reports
- Rationale for Analysis
- Methods and Results
- Conclusions
- Q & A

**What is an Effect Size?**
- Any statistic that quantifies the degree to which sample results diverge from the expectations specified in the null hypothesis (Cohen, 1994)
- Provides measure of "practical significance" of a statistical test result (Kirk, 1996), whereas p-values only indicate statistical significance
- Useful with abstract measurement indices

**Effect Size Type**
- Correlations (Explained Variance)
  - Pearson correlations
  - R-squared (OLS regression)
  - Eta-squared (ANOVA)
- Standardized Mean Difference
  - Cohen’s d (used by NSSE)

**Cohen’s d**
- Divides the mean difference between two groups by the pooled standard deviation
- Sample interpretation: “The average student at institution 1 has a .3 standard deviation higher score than the average student at institution 2.”

\[ d = \frac{M_1 - M_2}{SD} \]
Interpreting Cohen’s d

- Context matters

- Rule of thumb (reluctantly provided)
  - .2 = small effect
  - .5 = medium effect
  - .8 = large effect

- Despite caveats, rule widely applied

National Survey of Student Engagement

- NSSE measures...
  1) Student behaviors in & outside classroom
  2) General perceptions about college experience

- Focus on effective educational practices (see works by Pace, Astin, and Kuh)

Benchmarks of Effective Educational Practice

- Level of Academic Challenge
- Active and Collaborative Learning
- Student-Faculty Interaction
- Enriching Educational Experiences
- Supportive Campus Environment

NSSE Benchmark Report

Level of Academic Challenge

<table>
<thead>
<tr>
<th>Class</th>
<th>NSSEvile State</th>
<th>NSSE 2008</th>
<th>Sig</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean a</td>
<td>Mean b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-Year</td>
<td>54.3</td>
<td>52.9</td>
<td>*</td>
<td>.11</td>
</tr>
<tr>
<td>Senior</td>
<td>57.2</td>
<td>56.5</td>
<td></td>
<td>.05</td>
</tr>
</tbody>
</table>

Rationale for Analysis

- Is there a more appropriate rule of thumb for interpreting benchmark comparisons given the context of NSSE data?

- What should be considered a small, medium, and large effect size for NSSE benchmarks?

- What’s a meaningful difference?
Data

Effect sizes used were from:
- NSSE 2007 institutional reports
- \( n = 587 \)
- Comparisons with entire 2007 U.S. NSSE cohort

Frequency of Effect Sizes on NSSE Reports, by Cohen’s Rule

<table>
<thead>
<tr>
<th>Effect Size Range</th>
<th>Trivial</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&lt; .20)</td>
<td>FY</td>
<td>SR</td>
<td>FY</td>
<td>SR</td>
</tr>
<tr>
<td>(.20 to .49)</td>
<td>FY</td>
<td>SR</td>
<td>FY</td>
<td>SR</td>
</tr>
<tr>
<td>(.50 to .79)</td>
<td>FY</td>
<td>SR</td>
<td>FY</td>
<td>SR</td>
</tr>
<tr>
<td>(.80 +)</td>
<td>FY</td>
<td>SR</td>
<td>FY</td>
<td>SR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of Acad. Chal.</th>
<th>FY</th>
<th>SR</th>
<th>FY</th>
<th>SR</th>
<th>FY</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active &amp; Coll. Learn.</td>
<td>54%</td>
<td>56%</td>
<td>37%</td>
<td>36%</td>
<td>7%</td>
<td>7%</td>
</tr>
<tr>
<td>Stu.-Fac. Interaction</td>
<td>60%</td>
<td>48%</td>
<td>34%</td>
<td>38%</td>
<td>6%</td>
<td>11%</td>
</tr>
<tr>
<td>Enriching Educ. Exp.</td>
<td>52%</td>
<td>40%</td>
<td>40%</td>
<td>37%</td>
<td>7%</td>
<td>15%</td>
</tr>
<tr>
<td>Supp. Camp. Envt.</td>
<td>50%</td>
<td>46%</td>
<td>43%</td>
<td>44%</td>
<td>7%</td>
<td>9%</td>
</tr>
</tbody>
</table>

Cohen:
- Small effects are hardly visible.
- Medium effects are noticeable to the eye of the beholder.
- Large effects are plainly evident or obvious.

Our Approach

- Effects of various sizes are modeled using actual distribution of NSSE institutional means, done in two stages:
  1. institution-level distribution to establish percentile groupings
  2. student-level effect size calculations

Our Approach

- Small effect
  - Difference between students… 2\textsuperscript{nd} and 3\textsuperscript{rd} quartiles
  - \( \text{Small ES} = (\bar{X}_{A} - \bar{X}_{B})/\text{SD} \)

- Medium effect
  - Difference between students… lower and upper halves of distrib.
  - \( \text{Medium ES} = (\bar{X}_{C} - \bar{X}_{D})/\text{SD} \)
Our Approach

- **Large effect**

\[ \text{Difference between students...} \]

\[ 1^{st} \text{ and } 4^{th} \text{ quartiles} \]

\[ \text{Group E} \]

\[ 25\% \]

\[ \text{Group F} \]

\[ 25\% \]

Large ES \( \approx \frac{(X_f - X_e)}{SD} \)

Our Approach

- **Very large effect**

\[ \text{Difference between students...} \]

\[ \text{lowest and highest } 10\% \text{ of distrib.} \]

\[ \text{Group G} \]

\[ 10\% \]

\[ \text{Group H} \]

\[ 10\% \]

Very large ES \( \approx \frac{(X_h - X_g)}{SD} \)

Effect Sizes from Percentile Group Comparisons

<table>
<thead>
<tr>
<th>First-year Level of Acad. Chal.</th>
<th>S</th>
<th>M</th>
<th>L</th>
<th>VL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active &amp; Coll. Learn.</td>
<td>.18</td>
<td>.42</td>
<td>.60</td>
<td>.75</td>
</tr>
<tr>
<td>Stu.-Fac. Interaction</td>
<td>.13</td>
<td>.34</td>
<td>.49</td>
<td>.63</td>
</tr>
<tr>
<td>Enriching Educ. Exp.</td>
<td>.16</td>
<td>.38</td>
<td>.56</td>
<td>.77</td>
</tr>
<tr>
<td>Supp. Camp. Envt.</td>
<td>.18</td>
<td>.41</td>
<td>.57</td>
<td>.72</td>
</tr>
<tr>
<td>Minimum</td>
<td>.13</td>
<td>.34</td>
<td>.49</td>
<td>.63</td>
</tr>
<tr>
<td>Maximum</td>
<td>.18</td>
<td>.42</td>
<td>.60</td>
<td>.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Seniors Level of Acad. Chal.</th>
<th>S</th>
<th>M</th>
<th>L</th>
<th>VL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active &amp; Coll. Learn.</td>
<td>.12</td>
<td>.32</td>
<td>.51</td>
<td>.74</td>
</tr>
<tr>
<td>Stu.-Fac. Interaction</td>
<td>.11</td>
<td>.35</td>
<td>.52</td>
<td>.69</td>
</tr>
<tr>
<td>Enriching Educ. Exp.</td>
<td>.24</td>
<td>.54</td>
<td>.87</td>
<td>.99</td>
</tr>
<tr>
<td>Supp. Camp. Envt.</td>
<td>.16</td>
<td>.45</td>
<td>.61</td>
<td>.76</td>
</tr>
<tr>
<td>Minimum</td>
<td>.11</td>
<td>.32</td>
<td>.51</td>
<td>.69</td>
</tr>
<tr>
<td>Maximum</td>
<td>.24</td>
<td>.54</td>
<td>.87</td>
<td>.99</td>
</tr>
</tbody>
</table>

**Proposed Reference Values for Effect Size Interpretation** (NSSE Benchmark Comparisons)

<table>
<thead>
<tr>
<th>EFFECT SIZE</th>
<th>SMALL</th>
<th>MEDIUM</th>
<th>LARGE</th>
<th>VERY LARGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.1</td>
<td>.3</td>
<td>.5</td>
<td>.7</td>
</tr>
</tbody>
</table>

Frequency of NSSE Effect Sizes – Proposed Reference Values

<table>
<thead>
<tr>
<th>Effect Size Range</th>
<th>Trivial</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
<th>V. Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY SR</td>
<td>FY SR</td>
<td>FY SR</td>
<td>FY SR</td>
<td>FY SR</td>
<td></td>
</tr>
</tbody>
</table>

| Level of Acad. Chal. | 27% | 34% | 43% | 44% |
| Active & Coll. Learn. | 29% | 29% | 44% | 46% |
| Stu.-Fac. Interaction | 34% | 25% | 45% | 40% |
| Enriching Educ. Exp. | 25% | 21% | 46% | 32% |
| Supp. Camp. Envt.    | 27% | 23% | 41% | 42% |

Sample University Benchmark Comparisons (Seniors)

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Sample Univ.</th>
<th>Comparison group</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of Acad. Chal.</td>
<td>65.8</td>
<td>55.6</td>
<td>.72</td>
</tr>
<tr>
<td>Active &amp; Coll. Learn.</td>
<td>57.7</td>
<td>50.1</td>
<td>.44</td>
</tr>
<tr>
<td>Stu.-Fac. Interaction</td>
<td>42.8</td>
<td>41.2</td>
<td>.08</td>
</tr>
<tr>
<td>Enriching Educ. Exp.</td>
<td>44.0</td>
<td>39.8</td>
<td>.23</td>
</tr>
<tr>
<td>Supp. Camp. Envt.</td>
<td>62.7</td>
<td>56.9</td>
<td>.30</td>
</tr>
</tbody>
</table>
Examine Individual Items for ES Interpretation

- Make benchmark scores and effect sizes more tangible and observable
- More actionable observations exist at the item level
- ES can result from many small differences or few large differences

Level of Academic Challenge (ES=.72)

<table>
<thead>
<tr>
<th>Percent of students who...</th>
<th>Sample Univ.</th>
<th>Comp. Group</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Said the institution emphasizes studying and academic work</td>
<td>82%</td>
<td>78%</td>
<td>4%</td>
</tr>
<tr>
<td>Worked harder than expected to meet an instructor’s expectations</td>
<td>66%</td>
<td>57%</td>
<td>10%</td>
</tr>
<tr>
<td>Said courses emphasized analyzing ideas, experiences, or theories</td>
<td>95%</td>
<td>84%</td>
<td>11%</td>
</tr>
<tr>
<td>Said courses emphasized synthesizing</td>
<td>90%</td>
<td>74%</td>
<td>16%</td>
</tr>
<tr>
<td>Said courses emphasized making judgments about the value of info</td>
<td>86%</td>
<td>71%</td>
<td>15%</td>
</tr>
<tr>
<td>Said courses emphasized applying theories to new situations</td>
<td>95%</td>
<td>79%</td>
<td>16%</td>
</tr>
<tr>
<td>Read more than 10 assigned books or book-length packs of readings</td>
<td>68%</td>
<td>52%</td>
<td>16%</td>
</tr>
<tr>
<td>Wrote at least one paper or report of 20 pages or more</td>
<td>62%</td>
<td>49%</td>
<td>12%</td>
</tr>
<tr>
<td>Wrote more than 4 papers or reports between 5 and 19 pages</td>
<td>72%</td>
<td>46%</td>
<td>27%</td>
</tr>
<tr>
<td>Wrote more than 10 papers or reports of fewer than 5 pages</td>
<td>37%</td>
<td>31%</td>
<td>6%</td>
</tr>
<tr>
<td>Spent more than 10 hours/week preparing for class (studying, etc.)</td>
<td>68%</td>
<td>55%</td>
<td>13%</td>
</tr>
</tbody>
</table>

Student-Faculty Interaction (ES=.08)

<table>
<thead>
<tr>
<th>Percent of students who...</th>
<th>Sample Univ.</th>
<th>Comp. Group</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussed grades or assignments with an instructor</td>
<td>57%</td>
<td>58%</td>
<td>1%</td>
</tr>
<tr>
<td>Talked about career plans with a faculty member or advisor</td>
<td>47%</td>
<td>40%</td>
<td>7%</td>
</tr>
<tr>
<td>Discussed ideas from classes with faculty outside of class</td>
<td>27%</td>
<td>27%</td>
<td>0%</td>
</tr>
<tr>
<td>Received prompt written or oral feedback from faculty</td>
<td>64%</td>
<td>62%</td>
<td>2%</td>
</tr>
<tr>
<td>Worked with faculty members on activities other than coursework</td>
<td>22%</td>
<td>21%</td>
<td>1%</td>
</tr>
<tr>
<td>Worked on research project with a faculty member outside of class</td>
<td>19%</td>
<td>19%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Conclusions

1. Cohen’s rule of thumb does not adequately classify effect sizes for NSSE benchmark comparisons.

2. The proposed thresholds of .1, .3, .5, and .7 are grounded in actual NSSE findings and allow for refined interpretations of NSSE results.

3. Examining individual item frequencies confirms and provides a richer explanation for the labels applied to effect size results.

4. The approach in this study to contextualize effect sizes in normative data appears to be useful and appropriate, and may apply to other institutional research data analyses.

Contact Information:

Bob Gonyea
rgonyea@indiana.edu

Shimon Sarraf
ssarraf@indiana.edu

Paper available on NSSE web site
http://nsse.iub.edu/pdf/effect_size_guide.pdf