Contextualizing NSSE Effect Sizes: Empirical Analysis and Interpretation of Benchmark Comparisons

Bob Gonyea
Shimon Sarraf
IU Center for Postsecondary Research

2009 INAIR Annual Conference
March 26, 2009
Indianapolis, Indiana

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

	NSSEville State	NSSE 2008		
Class	*Mean* [a]	*Mean* [a]	*Sig* [b]	*Effect Size* [c]
First-Year	54.3	52.9	*	.11
Senior	57.2	56.5		.05

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?

Methods and Results

Data

Effect sizes used were from:

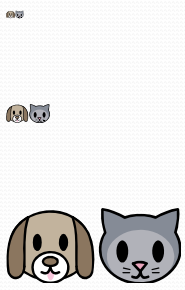
- NSSE 2007 institutional reports
- n=587
- Comparisons with entire 2007 U.S. NSSE cohort

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

	Effect Size Range							
	Trivial (< .20)		Small (.20 to .49)		Medium (.50 to .79)		Large (.80 +)	
	FY	SR	FY	SR	FY	SR	FY	SR
Level of Acad. Chal.	50%	62%	42%	30%	7%	7%	1%	1%
Active & Coll. Learn.	54%	56%	37%	36%	7%	7%	2%	1%
Stu.-Fac. Interaction	60%	48%	34%	38%	6%	11%	1%	3%
Enriching Educ. Exp.	52%	40%	40%	37%	7%	15%	1%	8%
Supp. Camp. Envt.	50%	46%	43%	44%	7%	9%	1%	1%

Cohen:

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



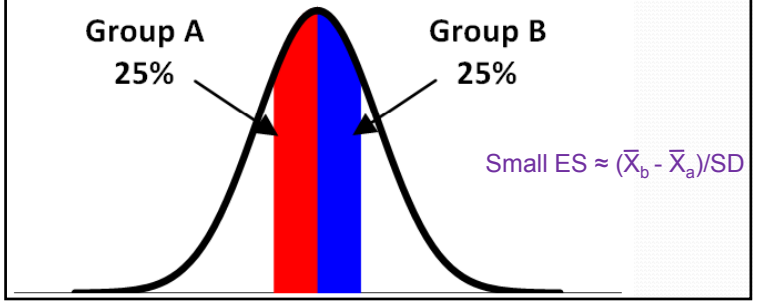
Our Approach

- 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

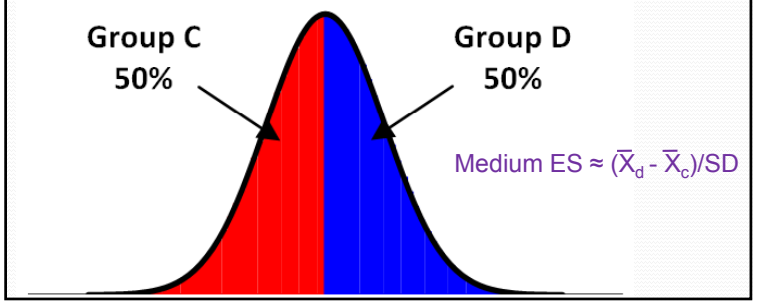
Our Approach

- *Small* effect
  - *Difference between students… 2nd and 3rd quartiles*



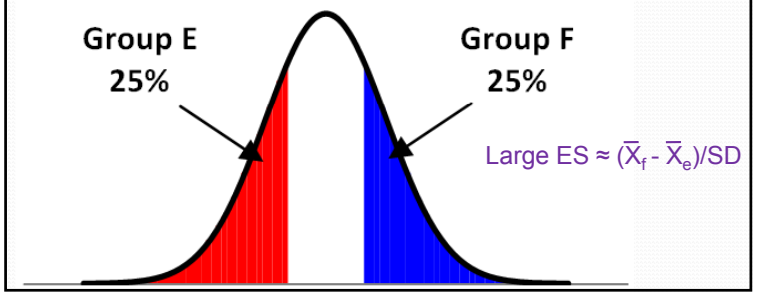
Our Approach

- *Medium* effect
  - *Difference between students… lower and upper halves of distrib.*



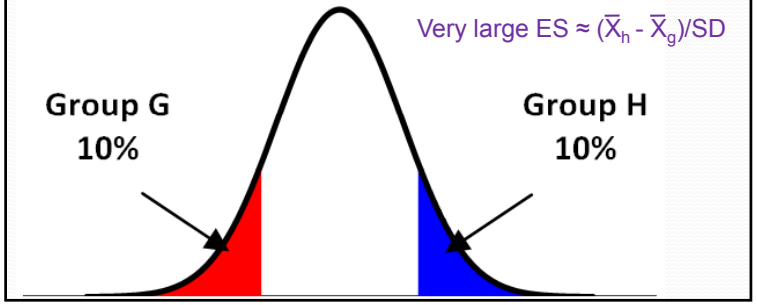
Our Approach

- *Large* effect
  - *Difference between students… 1st and 4th quartiles*



Our Approach

- *Very large* effect
  - *Difference between students… lowest and highest 10% of distrib.*



Our Approach

- *Small* effect ➢ *Difference between students… 2nd and 3rd quartiles*
- *Medium* effect ➢ *Difference between students… lower and upper halves of distrib.*
- *Large* effect ➢ *Difference between students… 1st and 4th quartiles*
- *Very large* effect ➢ *Difference between students… lowest and highest 10% of distrib.*

Effect Sizes from Percentile Group Comparisons

	First-year				Seniors			
	S	M	L	VL	S	M	L	VL
Level of Acad. Chal.	.18	.42	.60	.75	.12	.32	.51	.74
Active & Coll. Learn.	.13	.37	.56	.70	.11	.35	.52	.69
Stu.-Fac. Interaction	.13	.34	.49	.63	.14	.39	.63	.87
Enriching Educ. Exp.	.16	.38	.56	.77	.24	.54	.87	.99
Supp. Camp. Envt.	.18	.41	.57	.72	.16	.45	.61	.76
Minimum	*.13*	*.34*	*.49*	*.63*	*.11*	*.32*	*.51*	*.69*
Maximum	*.18*	*.42*	*.60*	*.77*	*.24*	*.54*	*.87*	*.99*

Proposed Reference Values for Effect Size Interpretation

(NSSE Benchmark Comparisons)

	EFFECT SIZE
SMALL	.1
MEDIUM	.3
LARGE	.5
VERY LARGE	.7

Frequency of NSSE Effect Sizes – Proposed Reference Values

	Effect Size Range									
	Trivial (0 to .09)		Small (.10 to .29)		Medium (.30 to .49)		Large (.50 to .69)		V. Large (.70+)	
	FY	SR	FY	SR	FY	SR	FY	SR	FY	SR
Level of Acad. Chal.	27%	34%	43%	44%	22%	14%	6%	5%	2%	3%
Active & Coll. Learn.	29%	29%	44%	46%	18%	17%	6%	6%	3%	2%
Stu.-Fac. Interaction	34%	25%	45%	40%	15%	20%	5%	9%	2%	5%
Enriching Educ. Exp.	25%	21%	46%	32%	22%	24%	5%	11%	2%	11%
Supp. Camp. Envt.	27%	23%	41%	42%	24%	25%	7%	7%	1%	3%

Sample University Benchmark Comparisons (Seniors)

Benchmark	Sample Univ.	Comparison group	*Effect size*	
Level of Acad. Chal.	65.8	55.6	*.72*	***Very large***
Active & Coll. Learn.	57.7	50.1	*.44*	***Medium***
Stu-Fac. Interaction	42.8	41.2	*.08*	***Trivial***
Enriching Educ. Exp.	44.0	39.8	*.23*	***Small***
Supp. Camp. Envt.	62.7	56.9	*.30*	***Medium***

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
- See Table 6 of handout.

Level of Academic Challenge (ES=.72)

Percent of students who...	Sample Univ.	Comp. Group	Differ-ence
Said the institution emphasizes studying and academic work	82%	78%	4%
Worked harder than expected to meet an instructor's expectations	66%	57%	10%
Said courses emphasized analyzing ideas, experiences, or theories	95%	84%	11%
Said courses emphasized synthesizing	90%	74%	16%
Said courses emphasized making judgments about the value of info	86%	71%	15%
Said courses emphasized applying theories to new situations	95%	79%	16%
Read more than 10 assigned books or book-length packs of readings	68%	32%	36%
Wrote at least one paper or report of 20 pages or more	62%	49%	12%
Wrote more than 4 papers or reports between 5 and 19 pages	72%	46%	27%
Wrote more than 10 papers or reports of fewer than 5 pages	37%	31%	6%
Spent more than 10 hours/week preparing for class (studying, etc.)	69%	55%	14%

Student-Faculty Interaction (ES=.08)

Percent of students who...	Sample Univ.	Comp. Group	Differ-ence
Discussed grades or assignments with an instructor	57%	58%	-1%
Talked about career plans with a faculty member or advisor	47%	40%	6%
Discussed ideas from classes with faculty outside of class	27%	27%	0%
Received prompt written or oral feedback from faculty	64%	62%	1%
Worked with faculty members on activities other than coursework	22%	21%	1%
Worked on research project with a faculty member outside of class	19%	19%	0%

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.

Conclusions

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