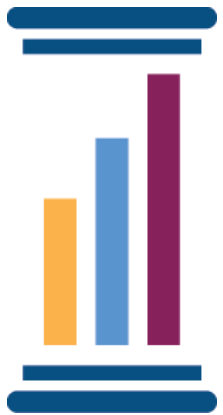


MOVING BEYOND STATISTICAL SIGNIFICANCE: USING EFFECT SIZES IN NSSE



Louis Rocconi & Bob Gonyea
National Survey of Student Engagement
Indiana University Bloomington

June 23, 2015



Outline

- **What is an effect size?**
- **Interpreting effect sizes**
- **Brief overview of NSSE EI and HIPs**
- **Methods**
- **Results**
- **Conclusions**

What is an effect size?

- A statistic that quantifies the degree to which sample results diverge from the expectations specified in the null hypothesis (Cohen, 1994)
- Provides a measure of “practical significance” of a statistical result, whereas p -values indicate statistical significance
- Useful with abstract measurement indices (such as NSSE’s Engagement Indicators)

We need more than statistical significance

- A lot of researchers do not know what the p -value actually means: probability of these data (or more extreme data) given that H_0 is true: $P(D | H_0)$
 - $P(D | H_0) \neq P(H_0 | D)$
 - A small $P(D | H_0)$ does not imply $P(H_0 | D)$ is also small
- The $p < .001$ error
- Statistical significance evaluates sample size
- APA & AERA require providing an effect size estimate when reporting a p -value.

Types of effect sizes

Kirk (1996), Rosnow & Rosenthal (2003) list three types of effect size measures:

1. Measures of Differences
Ex.) Cohen's d , h ; Hedge's g ; Glass's g
2. Strength of association (correlations)
Ex.) r , r^2 , μ^2 , ω^2
3. Other measures (ratios)
E.g., Odds ratio, Relative risk, Risk difference

**NSSE
Institutional
Reports use
Cohen's d & h**

Cohen's d

- Divides the mean difference between two groups by the pooled standard deviation

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Cohen's h

- **Difference between arcsine transformed proportions**

$$h = (2 \sin^{-1} \sqrt{P_1}) - (2 \sin^{-1} \sqrt{P_2})$$

- **Why an arcsine transformation? To make the proportions comparable in the sense of having variances independent of the parameter**

$$VAR(p) = \frac{(p)(1-p)}{n}$$

Interpreting effect sizes

- **Context matters:** Reporting and interpreting effect sizes in the context of previously reported effects is essential (APA, 1999).
- **Cohen's rules of thumb (reluctantly provided)**
 - .2 = small effect
 - .5 = medium effect
 - .8 = large effect
- **Despite caveats, Cohen's rules of thumb are widely used**

Context matters

- **McCartney & Rosenthal (2000)** note that in research involving hard to change outcomes, such as the incidence of heart attacks, the largest effect size found was below .20. However, those “small” effects correspond to reducing the incidence heart attacks by about half, an enormous practical significance.
- **Ellis (2010)** found that around two-thirds of effect sizes reported in international business were small ($r < .30$).
- **Hill, et al. (2008)** summarized average effect sizes for educational intervention research and found an average effect size of .33 in elementary school, .51 for middle school, and .27 for high school.
- **Lipsey et al. (2012)** found effect sizes investigating academic performance on standardized reading and mathematics achievement tests to rarely be as large as .30.

Engagement Indicators

- **Higher-Order Learning**
- **Reflective & Integrative Learning**
- **Learning Strategies**
- **Quantitative Reasoning**
- **Collaborative Learning**
- **Discussions with Diverse Others**
- **Student-Faculty Interaction**
- **Effective Teaching Practices**
- **Quality of Interactions**
- **Supportive Environment**

High-Impact Practices

- **Learning Community**
- **Service-Learning**
- **Research with Faculty**
- **Internship or Field Experience**
- **Study Abroad**
- **Culminating Senior Experience**

Purpose & Research Questions

The purpose of this study is to examine the distribution of statistical comparisons and their effects between institutions and their comparison groups using measures from NSSE, and to make recommendations for the interpretation of effect sizes from engagement results.

1. How do the effect sizes from NSSE institutional comparisons distribute within Cohen's small, medium, and large ranges?
2. Is it possible to derive more useful effect size cut points that fit the context of institutional engagement results?

Data

- **2013 & 2014 administrations of NSSE**
- **984 U.S. institutions**
- **More than 580,000 students**
- **Comparisons with the entire 2013 & 2014 cohort**

Methods

Effects of various sizes were modeled using actual distribution of NSSE institutional means, done in two stages:

- 1. Institution level:**

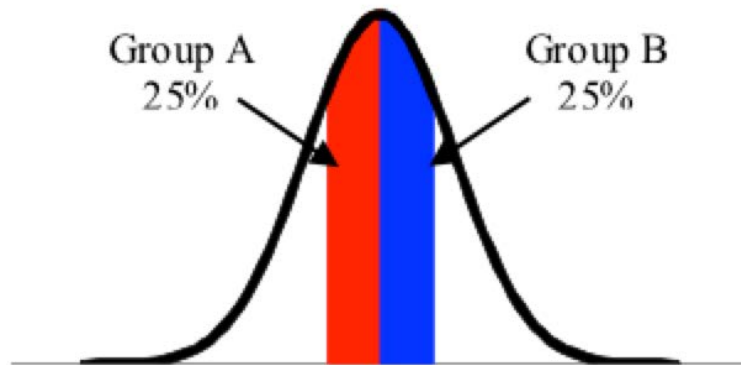
We assigned percentile ranks to the institutional scores for each EI and HIP, separately by class. Then we grouped the institutions into four quartiles, with the lowest 25% of scores in the first quartile, the next 25% of scores in the second quartile, etc.

- 2. Student level:**

Using the students from these groups of institutions, we computed effect sizes for contrived small, medium, and large comparisons.

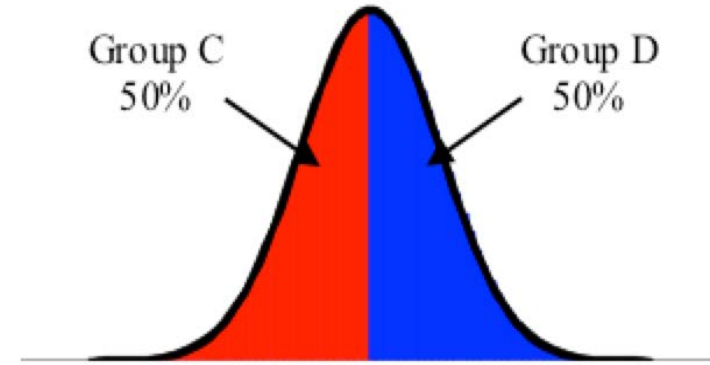
Methods

Small



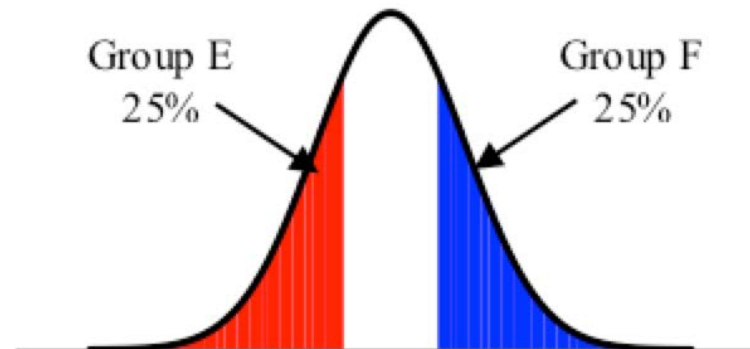
Distribution of Institutional Scores

Medium



Distribution of Institutional Scores

Large



Distribution of Institutional Scores

Frequency of NSSE Effect Sizes by Cohen's Ranges

	Trivial		Small		Medium		Large	
	ES < .2		.2 ≤ ES < .5		.5 ≤ ES < .8		ES ≥ .8	
	FY	SR	FY	SR	FY	SR	FY	SR
HO	72%	75%	26%	23%	1%	1%	<1%	<1%
RI	71%	68%	26%	28%	2%	3%	<1%	1%
LS	75%	66%	22%	33%	2%	1%	<1%	<1%
QR	76%	79%	20%	18%	2%	2%	1%	<1%
CL	64%	58%	30%	35%	4%	5%	2%	2%
DD	61%	63%	34%	33%	4%	3%	<1%	1%
SF	60%	41%	33%	39%	6%	16%	1%	4%
ET	68%	71%	30%	27%	1%	2%	<1%	<1%
QI	59%	59%	37%	37%	2%	4%	<1%	0%
SE	61%	55%	34%	38%	4%	6%	<1%	<1%
Lrn Com	57%	69%	38%	26%	3%	3%	1%	1%
Serv-Lrn	47%	46%	36%	36%	11%	13%	6%	5%
Res w/Fac	84%	55%	15%	32%	1%	11%	0%	2%
Intern	--	43%	--	38%	--	15%	--	4%
Study Ab	--	40%	--	43%	--	10%	--	7%
SR Cap	--	36%	--	36%	--	17%	--	10%

Effect sizes from percentile group comparisons

	First-year			Senior		
	Small	Medium	Large	Small	Medium	Large
HO	0.09	0.22	0.37	0.10	0.25	0.36
RI	0.11	0.26	0.39	0.10	0.27	0.41
LS	0.09	0.23	0.35	0.08	0.20	0.31
QR	0.09	0.24	0.35	0.11	0.30	0.47
CL	0.13	0.36	0.55	0.12	0.38	0.59
DD	0.13	0.33	0.50	0.12	0.32	0.51
SF	0.12	0.34	0.54	0.19	0.49	0.74
ET	0.10	0.28	0.41	0.09	0.25	0.37
QI	0.14	0.32	0.46	0.13	0.36	0.52
SE	0.12	0.31	0.49	0.14	0.34	0.53
Minimum d	0.09	0.22	0.35	0.08	0.20	0.31
Maximum d	0.14	0.36	0.55	0.19	0.49	0.74
Average d	0.11	0.29	0.44	0.12	0.32	0.48

Effect sizes from percentile group comparisons

	First-year			Senior		
	Small	Medium	Large	Small	Medium	Large
Learning Community	0.10	0.35	0.51	0.10	0.29	0.43
Service-Learning	0.18	0.43	0.73	0.17	0.43	0.69
Research with Faculty	0.06	0.17	0.25	0.16	0.41	0.61
Internship	--	--	--	0.20	0.50	0.76
Study Abroad	--	--	--	0.20	0.50	0.78
Senior Capstone	--	--	--	0.25	0.60	0.92
Minimum h	0.06	0.17	0.25	0.10	0.29	0.43
Maximum h	0.18	0.43	0.73	0.25	0.60	0.92
Average h	0.11	0.31	0.50	0.18	0.46	0.70

Proposed reference values

	<i>d</i> range	<i>h</i> range*
Small	$\geq .1$	$\geq .2$
Medium	$\geq .3$	$\geq .5$
Large	$\geq .5$	$\geq .8$

* Particularly for Service-Learning, Internship, Study Abroad, and Capstone

Frequency of effect sizes using proposed references

	Trivial		Small		Medium		Large	
	ES ≤ .1		.1 < ES ≤ .3		.3 < ES ≤ .5		ES > .5	
	FY	SR	FY	SR	FY	SR	FY	SR
HO	45%	46%	44%	45%	9%	8%	1%	1%
RI	40%	40%	47%	44%	11%	12%	2%	4%
LS	44%	38%	46%	46%	8%	15%	2%	1%
QR	47%	49%	42%	41%	8%	7%	3%	3%
CL	34%	30%	46%	48%	14%	14%	5%	7%
DD	33%	35%	47%	47%	15%	14%	4%	4%
SF	33%	23%	43%	34%	17%	23%	6%	20%
ET	38%	41%	48%	46%	12%	11%	1%	2%
QI	34%	30%	46%	48%	16%	18%	3%	4%
SE	36%	30%	45%	46%	15%	18%	4%	6%

EI Examples: Higher-Order Learning

Effect Size:

Small
-.13

Medium
.35

Large
.55

	Small -.13			Medium .35			Large .55		
	Institution A	Comp Group A	Percentage Point Difference	Institution B	Comp Group B	Percentage Point Difference	Institution C	Comp Group C	Percentage Point Difference
Applying facts, theories, or methods to practical problems or new situations	76%	80%	-4%	77%	77%	0%	74%	62%	11%
Analyzing an idea, experience, or line of reasoning in depth by examining its parts	71%	77%	-6%	82%	78%	4%	74%	53%	21%
Evaluating a point of view, decision, or information source	65%	70%	-5%	89%	76%	14%	76%	53%	23%
Forming a new idea or understanding from various pieces of information	65%	72%	-7%	91%	74%	17%	72%	60%	12%
	Average Difference per Item		5%	Average Difference per Item		9%	Average Difference per Item		17%

HIP Examples: Service-Learning

Effect Size:	Small			Medium			Large		
	Institution D	Comp Group D	Percentage Point Difference	Institution E	Comp Group E	Percentage Point Difference	Institution F	Comp Group F	Percentage Point Difference
Participation in a community-based project as part of a course	78%	69%	10%	85%	63%	22%	92%	61%	31%

Conclusions

- **Cohen's rules of thumb do not adequately classify effect sizes for NSSE Engagement Indicator comparisons**
- **The proposed thresholds of .1, .3, .5 for Engagement Indicators and .2, .5, .8 for High-Impact Practices are grounded in actual NSSE data and allow for refined interpretations of NSSE results**
- **Contextualizing effect sizes in normative data appears to be useful and appropriate for interpreting effect sizes in the context of student engagement**

Questions & Contact Info

Louis Rocconi

lrocconi@indiana.edu

Bob Gonyea

rgonyea@indiana.edu