

Analysis of Multiple Years of NSSE Data: Tips and Strategies

NSSE Webinar
April 28, 2009

Allison BrckaLorenz
Bob Gonyea

National Survey
of Student Engagement



Overview

- Introduction
- Five Multi-Year Analysis “Tasks”
 1. Identifying Multi-Year Questions
 2. Methods for Multi-Year Analysis
 3. Data Quality
 4. Changes in NSSE Over Time
 5. Merging Multi-Year Data
- Multi-Year Results



Task 1: Identify and Focus on Specific Questions



Identifying Multi-Year Questions

- Multi-year questions should be
 - Specific*
 - Answerable*
 - Relevant to campus priorities*
- Exploratory studies are possible, but it is best to set limits to focus the effort

Identifying Multi-Year Questions


Possible Multi-Year Questions

- Confirming stability and reliability
 - How stable was our data from one year to the next?*
- Measuring change due to campus initiatives
 - Given the implementation of a specific campus initiative, how much did engagement change before and after?*
- Identifying trends over time
 - What trends in the data are apparent in given engagement measures over time?*

NSSEville State University – Research Question

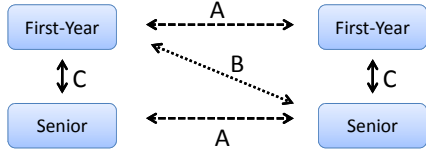
- Previous NSSE results showed a deficit in undergraduate research, especially for women.
- Implemented “Undergraduate Student Research Program” (2006-07) to give resources to students and faculty for research projects outside of class.
- Question: Did research with faculty increase between 2006 and 2008? If so, did changes vary by gender?**

Task 2:
Select and Employ Appropriate Methods of Analysis



Methods for Multi-Year Analysis

NSSE 2005 NSSE 2008



A: Cohort Comparisons B: Longitudinal Comparison C: Cross-sectional Comparisons

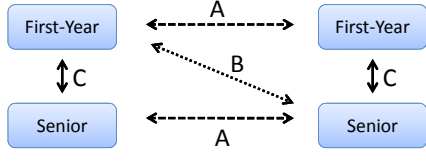
Methods for Multi-Year Analysis

Important Considerations

- Engagement is a process measure, not an achievement measure.
- First and senior years are different educational contexts, with different engagement patterns.
- First-years include those who will leave your institution.
- Seniors include persisters as well as transfers.
- Attrition from survey participation.

Methods for Multi-Year Analysis

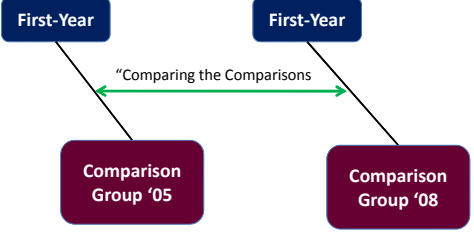
NSSE 2005 NSSE 2008



A: 😊 B: 😐 C: ☹️

Methods for Multi-Year Analysis

NSSE 2005 NSSE 2008




"Comparing the Comparisons"

Methods for Multi-Year Analysis

Statistical Difference

t-tests



Methods for Multi-Year Analysis

- **t-tests** determine whether the means of two groups are statistically different.
- **t-tests** can be quickly calculated by dividing the mean difference by the standard error of the mean.
 - Standard errors of the mean are included in the comparison reports
 - t-scores greater than 2 are $p < .05$, greater than 2.6 are $p < .01$, and greater than 3.3 are $p < .001$.

t-test Calculation

$$\frac{M_{2008} - M_{2006}}{SEM} = t$$

$$\frac{64.8 - 60.4}{.89} = 4.9$$

$(p < .001)$

Methods for Multi-Year Analysis

Statistical Difference

- t-tests
- ANOVA
 - Needs at least *three* years of data
 - Can use statistical controls
- Regression
 - Can use statistical controls



Practical Difference

- Percentage change (frequencies)
- Effect size (means)

Methods for Multi-Year Analysis

- **Effect size:** A measure of the strength of the relationship between two variables; *practical* significance
 - Cohen's *d*: Divide the mean difference by the pooled (average) standard deviation.
 - See [Contextualizing NSSE Effect Sizes](#). For benchmarks, consider:
 - $> |.1|$ = small
 - $> |.3|$ = medium
 - $> |.5|$ = large

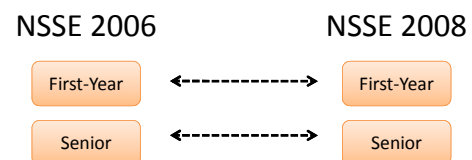
Effect Size Calculation

$$\frac{M_{2008} - M_{2006}}{SD_{pooled}} = d$$

$$\frac{64.8 - 60.4}{13.9} = .32$$

(medium)

NSSEville State University - Methods



- t-tests to determine statistical significance
- Effect size to determine practical significance

Task 3:**Attend to Data
Quality for Each
Year in the Analysis****Data Quality****Data quality measures to consider**

- Response rate
- Sampling Error (frequencies)
(depends on sample and population sizes)
- Standard Error (means)
(depends on variance and sample size)
- Proportional representation
- Missing data

Data Quality: Response Rate

- **Response rate:** Percentage of a sample that completes the questionnaire
 - “Complete” → viewed the demographics page
 - NSSE 2008 average response rate was 37%, with a large majority of institutions ranging from 20% to 60%
- Low response rates do not necessarily produce nonresponse bias
 - Nonresponse bias is the extent to which responders and nonresponders differ on key variables
 - Nonresponse bias is minimal in overall NSSE results

Data Quality: Sampling Error

- **Sampling error:** How much respondents could differ from the population
 - Based on number of respondents (n) relative to total population (N)
 - Preferred sampling errors around $\pm 3\%$ or 5%
 - Sampling errors greater than $\pm 10\%$ need not be dismissed entirely, but should be interpreted with caution

**Data Quality:
Proportional Representation**

- Determine the extent to which respondent demographics match those of your population
 - Weighting may help counter bias
 - NSSE weights data by gender and enrollment status (because women and full-time students respond at higher rates)

Data Quality

- Start with the *Respondent Characteristics* in your **Institutional Report**
 - Response rates
 - Sampling errors
 - Student characteristics
- Check the *NSSE Multi-Year Reporting Logic Over Time* [Handout](#)

NSSE Multi-Year Reporting Logic Over Time

2001-2003		Mean-Freq		Benchmark Report		Annual Report Percentiles	
Weighting	School	None	None	None	None	None	None
Sample	School	1, 2, 3	1, 2, 3	1	1	1	1, 2
Sample11	Reference	1	1	1	1	1	1, 2

Sample11 values of 1 and 2 are random; 3 is targeted
1: Distribution-level comparison (student-level benchmark scores not calculated; average item approach used)
2: Distribution-level comparison

2004		Mean-Freq		Benchmark Report		Annual Report Percentiles	
Weighting	School	None	None	Weight1	Weight2	Weight1	Weight2
Sample	School	1, 2, 3	1, 2, 3	1, 2	1, 2	1, 2	1, 2
Sample11	Reference	1	1	1, 2	1, 2	1, 2	1, 2

Sample11 values of 1 and 2 are random; 3 is targeted
1: Distribution-level comparison (although it is now used in benchmark calculations; begin, peer groups still use average of average approach)
2: Distribution-level comparison

2005		Mean-Freq		Benchmark Report		Annual Report Percentiles	
Weighting	School	None	None	Weight1	Weight2	Weight1	Weight2
Sample	School	1, 2, 3, 4	1, 2, 3, 4	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
Sample11	Reference	1	1	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3

Sample11 values of 1, 2 and 3 are random; 4 is targeted
1: Student-level comparison (peer group averages calculated with all student-level scores); Engagement Index and Devils Report discontinued
2: Student-level comparison


2006		Mean-Freq		Benchmark Report		Annual Report Percentiles	
Weighting	School	None	None	Weight1	Weight2	Weight1	Weight2
Sample	School	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
Sample11	Reference	1	1	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3

Sample11 values of 1, 2 and 3 are random; 4 is targeted
1: Student-level comparison (peer group averages calculated with all student-level scores); Engagement Index and Devils Report discontinued
2: Student-level comparison

NSSEville State University – Data Quality

<p>2006</p> <ul style="list-style-type: none"> 📌 Response Rate: 32% (30% FY/34% SR) 📌 Sampling Error: 4.1% FY/4.4% SR 📌 81% FY Female, 74% SR Female 	<p>2008</p> <ul style="list-style-type: none"> 📌 Response Rate: 35% (36% FY/34% SR) 📌 Sampling Error: 3.5% FY/3.5% SR 📌 69% FY Female, 68% SR Female
--	--

Task 4: Take into Account Changes in NSSE Items and Reports Across Years



Changes in NSSE Over Time

- 📌 Big changes in 2004!
- 📌 *NSSE Multi-Year Variable Tracking Sheet*
 - 📌 Question changes
 - 📌 Response value changes
- 📌 Benchmark considerations
 - 📌 See our new *NSSE Multi-Year Data Analysis Guide* and our new *Multi-Year Benchmark Report!*

Changes in NSSE Over Time

- 📌 Use the *NSSE Multi-Year Variable Tracking Sheet Codebook Handout* to interpret the Excel spreadsheet ([.xls](#))
- 📌 Important variables to *always* consider:
 - 📌 **classran**: Institution reported class rank (1 = First-Year, 4 = Senior)
 - 📌 **inelig**: Identifies eligibility (1 = Eligible)
 - 📌 Sample type
 - 📌 **Smp101** (2001-2003): 1 & 2 are random
 - 📌 **Smp105** (2004+): 1-3 are random
 - 📌 Weight (**stuwt2** for 2001-2003, **weight1** for 2004+)
 - 📌 Any other controlling, independent, or dependent variables

NSSEville State University – Data Quality

SORI_07	N_07	Var_07	Q_07	07to06 Q comp	07to08 Q comp	RV_07	07to06 RV comp	07to08 RV comp	BM_07	BM comp_07 to06	BM comp_07 to08
45	74	RESRCH04	Work on a research project with a faculty member outside of course or program requirements	1	1	Have Not decided. Do Not plan to do; Plan to do; Done	1	1	SFI	1	1

- 📌 For my analyses, I want to “keep” the variables classran, ineligible, smp105, weight1, gender, and RESRCH04

Task 5: Merge Multiple Years of Data



Merging Multi-Year Data

- Account for any changes in variable names
- Don't forget to create a variable to account for the data's year!
- Merge options
 - SPSS pull-down menus
 - Data -> Merge Files -> Add Cases
 - Write your own syntax

Merging Multi-Year Data with SPSS

- Save a copy of each year of your data with only the variables you want to "keep"
- In each year of your data create the same variable **Year** with a different value representing each year
 - 0 = 2001, 1 = 2004, 3 = 2007, 4 = 2008, etc.
- Open your "base year" and add the next year
 - Data -> Merge Files -> Add Cases
 - Repeat for each additional year

Merging Multi-Year NSSEville State University Data

See *Using SPSS to Merge Multiple Years of Data* [handout](#)

```

GET FILE='C:\temp\NSSEville 2006.sav' /KEEP classran inelig smpl05 weight1 gender RESRCH04.
COMPUTE Year = 0 .
VARIABLE LABELS Year "The year the data was collected".
VALUE LABELS Year
  0 '2006' .
EXECUTE .
SAVE OUTFILE='C:\temp\2006.sav'.

GET FILE='C:\temp\NSSEville 2008.sav' /KEEP classran inelig smpl05 weight1 gender RESRCH04.
COMPUTE Year = 1 .
VARIABLE LABELS Year "The year the data was collected".
VALUE LABELS Year
  1 '2008' .
EXECUTE .
SAVE OUTFILE='C:\temp\2008.sav'.

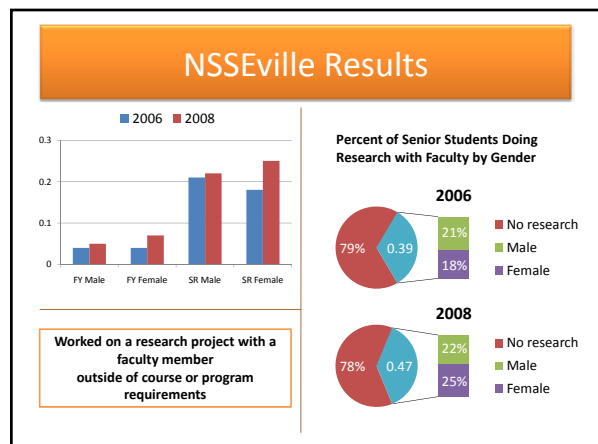
GET FILE='C:\temp\2006.sav'.
ADD FILES /FILE=*
/FILE='C:\temp\2008.sav'.
EXECUTE .
SAVE OUTFILE='C:\temp\NSSEville 2006 2008.sav'.
    
```

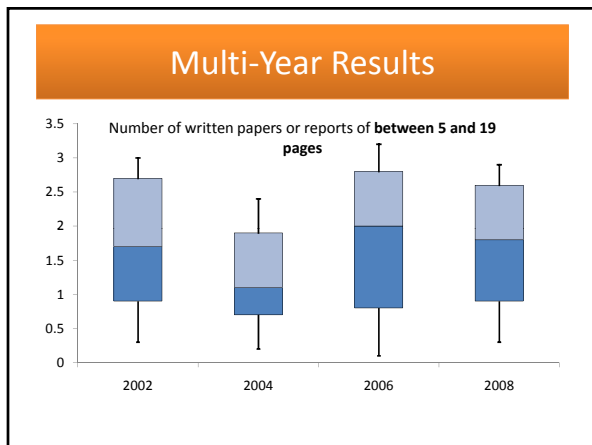
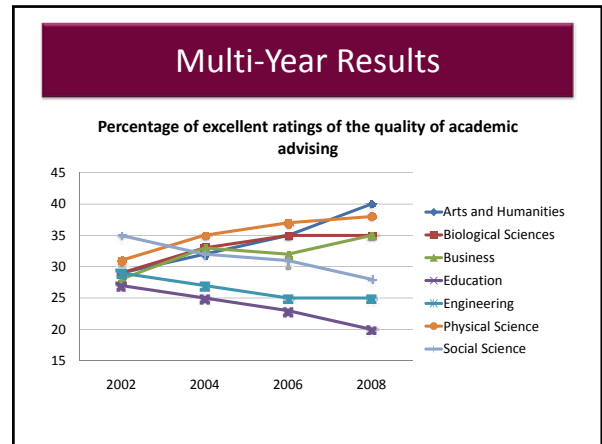
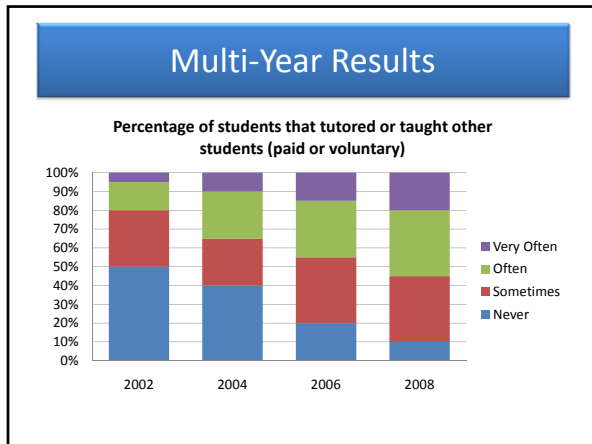
After running this syntax, I now have an SPSS dataset called "NSSEville 2006 2008" with the six variables I wanted to use in my analyses and a new variable that identifies the year of the data.

NSSEville Results

Worked on a research project with a faculty member outside of course or program requirements

	N	Percent "done"	Statistical Difference?	Effect Size
2006 F-Y Male	69	4%	NO	
2008 F-Y Male	133	5%		
2006 F-Y Female	306	4%	YES**	.13 (small)
2008 F-Y Female	309	7%		
2006 SR Male	82	21%	NO	
2008 SR Male	143	22%		
2006 SR Female	238	18%	YES***	.17 (small)
2008 SR Female	325	25%		





Questions?

Allison BrckaLorenz
abrckalo@indiana.edu

Bob Gonyea
rgonyea@indiana.edu

www.nsse.iub.edu

**National Survey
of Student Engagement**