

### *Do the results of analyses with NSSE data differ depending on the treatment of measurement level as interval or ordinal?*

#### **Purpose**

In the social and behavioral sciences, many self-report survey instruments result in data that is difficult to determine the exact scale of measurement. The treatment of responses on Likert-type scales or other sets of vague quantifiers is somewhat controversial, as these types of variables are considered ordinal by some and interval by others (Wuensch, 2005). The responses provide more information than simply rank, as true ordinal scale data would, yet true equal intervals are not able to be determined (Spatz, 2008). Is the interval between “Very often” and “Often” the same as the interval between “Often” and “Sometimes”? When researchers use inferential parametric tests such as *t*-tests, as are included in the NSSE Institutional Reports, the responses are treated as interval data. However, there are less common nonparametric inferential tests that are designed specifically for ordinal data, and some argue that these are more appropriate although the results generally do not differ greatly (Zumbo & Zimmerman, 1993). It is important to consider whether the NSSE means comparisons, which use independent-samples *t*-tests, are robust to potential violations of measurement scale assumptions.

#### **Data**

Using the NSSE 2010 data set, three institutions of varying enrollment sizes were randomly selected for use as sample institutions. The small sample institution, with an enrollment of less than 1,000 students, had data for 76 first-year students and 55 seniors. The medium-sized sample institution, with an enrollment of 5,000 to 9,999 students, had data for 329 first-year students and 255 seniors. The large sample institution, with an enrollment of over 20,000 students, had data for 2,209 first-year students and 2,261 seniors. All other eligible students at U.S. institutions participating in NSSE were used as the referent group, which had data for 165,998 first-year students and 196,231 seniors.

#### **Methods**

To investigate whether the results would differ depending on the analysis used with the NSSE data, a series of independent samples *t*-tests and Mann-Whitney U tests were conducted with the data. Comparisons were made between students at the selected sample institutions and all other U.S. institutions. Therefore, first-year students at the small sample institution were compared to all other first-year students, first-year students at the medium-sized sample institution were compared to all other first-year students, first-year students at the large sample institution were compared to all other first-year students, senior students at the small sample institutions were compared to all other senior

students, senior students at the medium-sized sample institution were compared to all other senior students, and senior students at the large sample institution were compared to all other senior students. These comparisons were done treating the data as both interval level (using the independent samples *t*-test) and ordinal level (Mann-Whitney U test). Comparisons were done for all variables included in the Mean Comparisons of the 2010 NSSE Institutional Reports. The analyses were weighted by gender and enrollment status.

## Results

Results indicated that for the majority of comparisons, the independent samples *t*-test and the Mann-Whitney U test do not yield different conclusions (see Table 1 for detailed results). For the comparisons made on the 85 variables included in the NSSE Institutional Reports, the highest number of mismatches for any group was seven (first-year small sample institution), while the lowest number was two (first-year medium-sized sample institution). Overall, for the 510 different comparisons (85 variables with 6 different groups) there were 25 total mismatches, which is only 4.9% of the total comparisons. Therefore, the conclusions made from the *t*-test and the Mann-Whitney U test are the same over 95% of the time.

For those mismatches, there was not a strong pattern for those where the Mann-Whitney U was significant but the *t*-test was not significant (17), or for those where the *t*-test was significant but the Mann-Whitney U was not significant (8). There may have been a slightly greater chance for Type I error with the Mann-Whitney U test, or Type II error with the independent samples *t*-test, but these results were not consistent for any variable, either by class or size of comparison group.

These findings suggest that treating NSSE response sets as interval data does not violate the assumptions of the parametric independent samples *t*-test, which is also the conclusion of previous studies exploring similar research questions (Zumbo & Zimmerman, 1993). Generally, parametric analyses are more well-known, making their results easier to interpret, understand, and explain to others (Suskie, 1996). Given this information, in addition to the findings that suggest results do not differ dramatically for comparisons using NSSE data, there is evidence to support the independent samples *t*-test as an appropriate analysis for inclusion in NSSE Institutional Reports.

## References

- Spatz, C. (2008). *Tales of distributions: Basic statistics* (9<sup>th</sup> ed.). Belmont, CA: Thomson Wadsworth.
- Suskie, L.A. (1996). *Questionnaire survey research: What works* (2<sup>nd</sup> ed.). Tallahassee, FL: Association for Institutional Research.
- Wuensch, K.L. (2005). Scales of measurement. In B.S. Everitt & D.C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1785-1789). West Sussex, UK: Wiley.
- Zumbo, B.D., & Zimmerman, D.W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology*, 34, 390-400.

**Table 1: Mann-Whitney U and *t*-test Mismatches by Class<sup>a</sup>**

	<i>First-Year</i>			<i>Senior</i>		
	Small sample mismatch	Medium sample mismatch	Large sample mismatch	Small sample mismatch	Medium sample mismatch	Large sample mismatch
clquest						
clpresen						
rewropap						
integrat						
divclass						
clunprep						
classgrp					X <sup>b</sup>	
occgrp						X <sup>c</sup>
intideas						
tutor						
commproj						
itacadem					X <sup>b</sup>	
email						
facgrade						
facplans						
facideas	X <sup>b</sup>					
facfeed						
workhard						
facother		X <sup>c</sup>				
oocideas						
divrstud						
diffstu2						
memorize						
analyze				X <sup>b</sup>		
synthesz						
evaluate						
applying						
readasgn						
readown						
writemor	X <sup>b</sup>		X <sup>b</sup>			
writemid						
writesml						
probseta						
probsetb						
exams						
atdart07			X <sup>b</sup>	X <sup>b</sup>		

	<i>First-Year</i>			<i>Senior</i>		
	Small sample mismatch	Medium sample mismatch	Large sample mismatch	Small sample mismatch	Medium sample mismatch	Large sample mismatch
excrcse05						
worshp05	X <sup>c</sup>					
ownview						X <sup>b</sup>
othrview						
chngrview					X <sup>b</sup>	
intern2						
voluntr2						
lerncom2						
resear2						
forlang2						
studyab2						
indstud2	X <sup>b</sup>					
seniorx2	X <sup>c</sup>					
envstu		X <sup>c</sup>		X <sup>b</sup>		
envfac						
envadm						
acadpr01						
workon01						
workof01						
cocurr01	X <sup>c</sup>					
social05						
carede01						
commute						X <sup>b</sup>
envschol						
envsuprt						
envdivrs						
envnacad			X <sup>b</sup>			
envsocial						
envevent						
envcompt						
gngenled						
gnwork						
gnwrite						
gnspeak						
gnanaly						
gnquant						
gncmpts						
gnothers	X <sup>c</sup>				X <sup>b</sup>	
gncitizn						

	<i>First-Year</i>			<i>Senior</i>		
	Small sample mismatch	Medium sample mismatch	Large sample mismatch	Small sample mismatch	Medium sample mismatch	Large sample mismatch
gning						
gnself						
gndivers						
gnprobsv				X <sup>c</sup>	X <sup>b</sup>	
gnethics			X <sup>b</sup>			
gncommun						
gnspirit						
advise						
entirexp						
samecoll						
<b>Total mismatches</b>	7	2	4	4	5	3

<sup>a</sup> Analyses weighted by gender and enrollment status

<sup>b</sup> Mann-Whitney U test is significant; *t*-test is not significant

<sup>c</sup> Mann-Whitney U test is not significant; *t*-test is significant