FSSE faculty survey of student engagement

Item Nonresponse Bias

Other Data Quality Indicators

The quality and usefulness of the survey data depends on the validity of the responses respondents give on each survey question. In survey research, missing data occurs for many reasons and may come from a variety of sources such as non-coverage, total or unit response, item non response and partial response (Brick & Kalton, 1996). Because the sources of missing data may result in a non-representative sample of a larger population, the results or the inferences made based on the sample might be inaccurate. Because most surveys are self-reported and have some level of item nonresponse, this might in turn bias the estimates that represent particular educational outcomes. While assessing the quality of survey data, the major concern is the extent to which respondents provide high-quality data by providing accurate and complete responses to all survey items (Schaeffer & Presser, 2003) and whether the item nonresponse patterns have an impact on the outcome measures or depend on certain characteristics of respondents.

The purpose of this study was to investigate the prevalence of item nonresponse bias among participants in the FSSE survey and its impact on the estimates of ten FSSE scale scores, by comparing item nonresponse patterns across faculty-level characteristics such as gender identity, racial or ethnic identification, citizenship, employment status, academic rank, and the number of undergraduate or graduate courses taught.

Data

The data from this study come from the 2014 administration of the Faculty Survey of Student Engagement (FSSE) from 18,860 faculty members at 143 bachelor's-granting colleges and universities. To be included in this study, faculty had to have provided their demographic information at the end of the FSSE survey. The item nonresponse was examined among fifty survey items on the FSSE 2014 survey. These fifty items were selected because they are included in ten FSSE scales, the main measures of faculty involvement in student engagement on the FSSE survey. These scales are 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, and Supportive Environment.

Methods

To answer the first research question about whether or not item nonresponse depends on faculty-level characteristics, item responses to fifty survey items were dichotomously recoded to indicate whether or not the item was responded to or missing. Next, chi-square tests of independence were used to examine whether the item nonresponse for each item differed by the faculty characteristics gender identity, racial or ethnic identification, citizenship, employment status, academic rank, and the number of undergraduate or graduate courses taught. Significance was noted for $p < .001$ due to the large sample size. Additionally, descriptives of individual items were examined to ensure that responses were well distributed and that there was no evidence that items were difficult to answer (such as unusually high item nonresponse). To answer the second research question about the magnitude of the item nonresponse for different groups, the effect size for the chi-square test of independence was calculated using the phi coefficient ($\phi = \sqrt{\frac{\chi^2}{N}}$) for 2×2 table or Cramer's V coefficient ($V = \sqrt{\frac{\chi^2}{N(L-1)}}$, where L refers to the smaller number, row or column, in the contingency table) for a contingency table larger than 2×2, to describe the magnitude of association between the dichotomous or categorical variables. The range of phi or Cramer's V coefficient is between 0 and 1, which corresponds to the complete independence to complete dependence of the variables. To measure the effect size of a chi-square test of independence, Cohen (1988) has suggested that phi values .10, .30 and .50 correspond to small,

Suggested citation: Chiang, Y.-C. & BrckaLorenz, A. (2015). Nonresponse bias. FSSE Psychometric Portfolio. Retrieved from fsse.indiana.edu.

medium and large effects, respectively. To interpret the magnitude for Cramer's V, a Cramer's V value needs to be converted to *w*, which is equivalent to a phi coefficient and is done by multiplying its value by $\sqrt{(L-1)}$. For example, a Cramer's V value of .10 obtained from a 2×4 table is equivalent to a phi or w value of .10 and so the effect is small.

Results

As shown in Table 1, the items within each scale do not have high proportions of item nonresponse. The highest proportion of item nonresponse (i.e., 14-15%) was found in the Discussions with Diverse Others scale and the lowest proportion of item nonresponse (i.e., 2-3%) was in the Supportive Environment scale. Overall, the item nonresponse rate for items within FSSE scales is between 2 to 15%. Using chi-square tests of independence to examine the item nonresponse by faculty characteristics, we find that most of the item nonresponse does not vary by faculty characteristics. As shown in Table 3, there were some significant differences found for item nonresponse by faculty characteristics, but the effect size (i.e., the magnitude of the relationship) was very small (less than .1). For example, there are significant differences in item nonresponse by gender identify and the number of undergraduate courses being taught among items in the Discussions with Diverse Others scale, but the magnitude of the differences are less than or equal to .05.

There are no notable item nonresponse bias indications among the items in the ten FSSE scales. The proportion of missing responses to these items does not vary by the faculty characteristics gender identity, racial or ethnic identification, citizenship, employment status, academic rank, and the number of undergraduate or graduate courses taught. Institutions and researchers using these items can feel confident that item nonresponse does not bias their results.

Although the item nonresponse for items within the Discussions with Divers Others scale is not extremely high, it is higher than the items in other scales. This may be an indication that these items are more difficult to answer than other survey items or it may be that these items are towards the end of the survey and respondents are dropping off. Cognitive interviews or additional testing could help with understanding why these items have more missing responses. One limitation of this study is that it did not account for when participants stop responding to the survey. It is possible that patterns of bias could emerge after respondents that did not finish the survey are removed from the analysis. Additionally, this study does not account for response bias. It is unknown whether or not there exists bias in who responds to the FSSE in general.

Although some statistically significant differences in item nonresponse were found for some of the tested items, the magnitude of these differences was trivial. Institutions and researchers using these items can feel confident that item nonresponse does not bias the results of these items. Future faculty characteristics, such as faculty's disciplinary appointment, age, and earned doctorate, could additionally be examined to further strengthen these results.

References

Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical methods in medical research*, *5*(3), 215-238.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, 65-88.

Table 1. Item Nonresponse and Statistical and Practical Differences by Groups Among Select FSSE Items

		Item Nonresponse		Statistical and Practical Differences						
		N	%	Gender	Employment Status	Academic Rank	Race/ Ethnicity	Citizenship	Number of Undergrad. Courses	Number of Graduate Courses
Supportive Environment	fSEacademic	513	3							
	fSElearnsup	456	2							
	fSEdiverse	385	2							
	fSEsocial	389	2							
	fSEwellness	424	2							
	fSEnonacad	425	2							
	fSEactivities	560	3							
	fSEevents	561	3							
Quality of Interactions	fQIstudent	545	3							
	fQIadvisor	730	4		ϕ=.039*	V=.038*				
	fQIfaculty	664	4							
	fQIstaff	1244	7						V=.048*	
	fQIadmin	1104	6				V=.039*		V=.046*	
Student-Faculty Interaction	fSFcareer	976	6		ϕ=.030*					
	fSFotherwork	1113	6		ϕ=.035*					
	fSFdiscuss	1032	6		ϕ=.041*					
	FSFperform	1089	6							
Effective Teaching Practices	fETgoals	986	6							
	fETorganize	1091	6							
	fETexample	1115	6							
	fETvariety	1021	6							
	fETreview	1035	6							
	fETstandards	1045	6							
	fETdraftfb	1115	6							
	fETfeedback	1195	7							
Quantitative Reasoning	fQRconclude	1973	11		ϕ=.035*					
	fQRproblem	2044	12							
	fQRevaluate	2108	11		ϕ=.031*					
Reflective & Integrative Learning	fRIintegrate	1997	11							
	fRIsocietal	2124	12							
	fRIdiverse	2096	12				V=.049*			
	fRIownview	2121	12				V=.039*			
	fRIperspect	2160	12							
	fRInewview	2131	12				V=.039*			
	fRIconnect	2155	12				V=.053*			
Collaborative Learning	fCLaskhelp	2202	13						V=.048*	
	fCLexplain	2336	13							
	fCLstudy	2342	13							
	fCLproject	2315	13						V=.048*	
Learning Strategies	fLSreading	2322	13				V=.043*			
	fLSnotes	2397	14		ϕ=.029*				V=.063*	
	fLSsummary	2378	14						V=.049*	
Discussions with Diverse Others	fDDrace	2448	14				V=.045*		V=.048*	
	fDDeconomic	2563	15				V=.045*		V=.045*	
	fDDreligion	2665	15	V=.039*			V=.041*		V=.050*	
	fDDpolitical	2710	15	V=.037*			V=.043*		V=.051*	
Higher-Order Learning	fHOapply	2369	14		ϕ=.031*					
	fHOanalyze	2360	13							
	fHOevaluate	2392	14				V=.041*			
	fHOform	2462	14							

Note: *p < .001. For dichotomous variables, employment status and citizenship, values are computed using Continuity Correction (2*2 table) instead of a Chi-Square test.