Investigating Social Desirability Bias in Student Self-Report Surveys

Angie L. Miller
Indiana University

The frequent use of student self-report surveys in higher education calls into question the possibility of social desirability having an unwanted influence on responses. This research explores the potential presence of social desirability bias with the National Survey of Student Engagement (NSSE), a widely used assessment of student behaviors. Correlations between a short social desirability scale and NSSE benchmarks, subscales, and selected items suggest that the majority of scores have no significant relationship with a measure of social desirability. A series of regression models controlling for demographic variables produce similar results. Effect sizes and estimates of explained variance are also discussed.

Survey research is becoming increasing common in education, in part due to heightened requirements for assessment and accountability at the institutional level (Kuh & Ewell, 2010). Surveys can provide a variety of information relatively easily, especially with the prevalence and availability of online data collection technology. However, this increased reliance on survey data calls into question the quality of the data. Can researchers and practitioners trust that the wealth of information gathered from the surveys is accurate and free from bias?

If there is reason to believe that questions on a survey will prompt respondents to answer untruthfully in an attempt to provide a socially appropriate response, researchers may want to explore the potential presence of social desirability bias (DeVellis, 2003). Many scales have been developed to measure the tendency to respond in a socially desirable manner (Crowne & Marlowe, 1960; Paulhus, 1984).

Generally, these instruments are based on the premise that there are some "ideally good" behaviors that no one actually does, and there are some "ideally bad" behaviors that everyone actually does. For instance, if a person responds "yes" to the statement "I am always a good listener, no matter who I am talking to," this would be considered a socially desirable response. Even if a person is a much better listener than most, there have likely been occasions when he or she has been irritable, or distracted, or tired and was therefore not a good listener. Conversely, if a person responds "no" to the statement "I sometimes feel resentful when I don't get my own way" this would also be considered a socially desirable response. Based on this premise, measures of socially desirability are able to provide a score for one's tendency to respond in a socially desirable way, and this score can then be examined in conjunction with responses to other available instruments. If an instrument is free from social desirability bias, scores on the instrument should not be related to scores on a measure of social desirability.

Investigations of social desirability bias were traditionally only conducted with surveys that contain items of a sensitive nature, such as sexual or drug-related behaviors (Carpenter, 2009). However, some research in higher education suggests that social desirability bias may have an impact on student responses for less sensitive topics as well. Recent studies have found significant relationships between social desirability and perceptions of institutional values, goal orientation, value commitment, major satisfaction, and self-reported gains with samples of undergraduate students (Bowman & Hill, 2011; Ferrari, Bristow, & Cowman, 2005; Ferrari & Cowman, 2004; Ferrari, McCarthy & Milner, 2009; Nauta, 2007). Given this broad array of areas that can be influenced by social desirability, it is important to further explore questions about the validity of using surveys with college students. These findings suggest that in higher

education research, some self-report measures may be subject to social desirability bias.

However, there is also research in higher education that fails to find significant relationships with social desirability. One study did not find a significant difference in social desirability bias between students with high and low imposter tendencies (feelings of intellectual phoniness experienced by high achievers) (Ferrari, 2005), while another did not detect significant differences in social desirability when comparing paper vs. web administered surveys or when comparing anonymous vs. non-anonymous conditions (Hancock & Flowers, 2001). Other research indicates a lack of significant relationships between social desirability and the varied subjects of time use efficiency, self-complexity, and athletic identity (Kelly, 2003; Luo, Watkins, & Lam, 2009; Nasco & Webb, 2006). This wide range of topics, all found to have no relationship to social desirability, suggests that not all student self-report measures are influenced by social desirability.

Given the conflicting results of previous research, it is important to consider whether social desirability is having an impact on student responses to the National Survey of Student Engagement (NSSE). A widely used measure in higher education, NSSE assesses a variety of behaviors related to student engagement with first-year and senior undergraduates (see http://nsse.iub.edu for more information). Items on NSSE require students to reflect on the frequency of various academic and non-academic behaviors as well as on more abstract concepts of growth and development. Several different measures of student engagement are available, including five "Benchmarks of Effective Educational Practice," three "Deep Approaches to Learning" subscales, and three "Gains" subscales. Although student engagement is not what one might consider a "sensitive" topic, students may be aware that answering items

in ways that display higher levels of engagement is desired by their institutions and they want to appear to be "good" students. Therefore, the purpose of the current research is to determine whether students' responses to the NSSE items are influenced by the tendency to respond in a socially desirable manner.

Method

Participants

In the spring 2010 NSSE online administration, 2,352 students at 6 participating institutions completed a short social desirability scale (Ray, 1984) in addition to the NSSE core survey. These institutions represented a variety of NSSE participants in terms of Carnegie classification, size, and geographic region. In this sample, there were 38.4% first-year students and 61.6% senior students. The participants were 43.9% males and 86.6% had full-time enrollment status. In terms of institutional-reported ethnicity, the sample was 63.6% Caucasian/White, 11.1% African American/Black, 7.4% Asian/Pacific Islander, 5.6% foreign, 2.4% Hispanic, 0.3% American Indian/Alaska Native, 0.3% multi-racial, and 9.4% with unknown ethnicity.

Measures

The Benchmarks of Effective Educational Practice were measured using 42 different NSSE items. These items are taken from sets of questions with varying stems, all collecting information about the frequency of their behaviors related to academic and non-academic activities that have taken place during their time at their institution. Responses are used to create 5 benchmark scores, each converted to a 100 point scale with higher scores indicating greater degrees of engagement in the respective areas. The benchmarks are Level of Academic Challenge (11 items), Active and Collaborative Learning (7 items), Student-Faculty Interaction

(6 items), Enriching Educational Experiences (12 items), and Supportive Campus Environment (6 items). With this sample, there were marginally to moderately acceptable levels of internal consistency (McMillan & Schumacher, 2001) for Level of Academic Challenge (Cronbach's α = .718), Active and Collaborative Learning (Cronbach's α = .687), Student-Faculty Interaction (Cronbach's α = .745), Enriching Educational Experiences (Cronbach's α = .634), and Supportive Campus Environment (Cronbach's α = .771).

The Deep Approaches to Learning subscales were measured using 12 different NSSE items. These items are taken from sets of questions with varying stems, all collecting information about the frequency of their behaviors related to academic activities and cognitive behaviors that have taken place during their time at their institution. Responses are used to create 3 subscale scores, each converted to a 100 point scale with higher scores indicating greater degrees of deep approaches to learning. The subscales are Higher Order Learning (4 items), Reflective Learning (3 items), and Integrative Learning (5 items). With this sample, there were acceptably high levels of internal consistency (McMillan & Schumacher, 2001) for Higher Order Learning (Cronbach's α = .843), Reflective Learning (Cronbach's α = .821), and Integrative Learning (Cronbach's α = .723).

The Gains subscales were measured using 16 different NSSE items. These items share a common stem ("To what extent has your experience at this institution contributed to your knowledge, skills, and personal development in the following areas?") with a 4 point response scale ranging from "Very little" to "Very much." Responses are used to create 3 subscales, each converted to a 100 point scale with higher scores indicating greater degrees of gains in each respective area. The subscales are Gains in Practical Competence (5 items), Gains in Personal and Social Development (7 items), and Gains in General Education (4 items). With this sample,

there were acceptably high levels of internal consistency (McMillan & Schumacher, 2001) for Practical Competence (Cronbach's α = .836), Personal and Social Development (Cronbach's α = .869), and General Education (Cronbach's α = .847).

Social desirability bias was measured using a short scale of social desirability (Ray, 1984). This 8-item scale instructs participants to respond with "Yes," "Not sure," or "No" to a set of statements concerning desirable and undesirable social behaviors. Responses are used to create a single measure of social desirability, with higher scores indicating higher levels of social desirability. Ray (1984) reported adequate internal consistency for the shortened 8-item scale (Cronbach's α = .77). With this sample, the internal consistency was slightly lower but still marginally acceptable (Cronbach's α = .696).

Two additional NSSE items were also used for this study. One item measures grades by asking students to report what most of their grades have been at their institution, with response options of "A," "A-," "B+," "B," "B-," "C+," "C," and "C- or lower." Higher scores on this item indicate higher grades. Another item measures overall institutional satisfaction by asking students evaluate their entire educational experience at the institution, with response options on a 4-point scale ranging from "Poor" to "Excellent." Higher scores on this item indicate higher levels of satisfaction.

Analyses

To investigate the presence of social desirability bias, bivariate correlations were conducted using the social desirability score and the Benchmarks, Deep Approaches to Learning, and Gains subscales, for first-year students and then again for senior students. Correlations were also conducted between social desirability score and two individual items of interest:

self-reported grades and overall institutional experience. Due to the large number of analyses, a Bonferroni correction (Field, 2009) was used and alpha was set to .002 (.05/26). As is recommended with all analyses using NSSE data, the analyses were weighted by gender, enrollment status, and institution size.

Several sets of regression analyses were also conducted. These analyses explored whether social desirability was a significant predictor of the Benchmarks, Deep Approaches to Learning, and Gains subscales when controlling for gender, enrollment status, first generation status, transfer status, athlete status, living on campus, Greek status, international status, ethnicity, self-reported grades, and overall institutional satisfaction (variables all known to be related to engagement scores). Regression analyses were also performed to determine if social desirability was a significant predictor of self-reported grades and overall institutional satisfaction, but removing these variables from the first step of each respective analysis. A Bonferroni correction was used here as well, and alpha was set to .002. Separate regressions were conducted for first-year and senior students. The regressions were also weighted by gender, enrollment status, and institution size.

Results

Correlations

The number of respondents, the Pearson's *r* value, and the R^2 value, a measure of explained variance that provides an estimate of effect size for bivariate correlations, are presented in Table 1. The number of respondents varies for each correlation because students need to respond to approximately three-fifths of the items within a benchmark or subscale to receive a score. The results of the correlations suggest that for most of the benchmarks and subscales, there is no significant relationship with social desirability.

Table 1: Social Desirability Correlations and Effect Size by Class[1]

	First-Year			Senior		
	r	R^2	N	*r*	R^2	N
Level of Academic Challenge	**.113***	.012	867	.033	.001	1632
Active & Collaborative Learning	.054	.003	860	.059	.003	1627
Student-Faculty Interaction	.057	.003	865	-.018	.000	1629
Enriching Educational Experiences	-.029	.001	866	-.067	.004	1629
Supportive Campus Environment	.028	.001	862	**.129***	.017	1630
Higher Order Learning	.051	.003	867	.030	.001	1629
Reflective Learning	**.140***	.019	867	**.101***	.010	1635
Integrative Learning	.097	.009	867	.074	.005	1630
Gains in Practical Competence	.061	.004	862	.062	.004	1630
Gains in Personal & Social Development	.079	.006	860	**.141***	.020	1624
Gains in General Education	.083	.007	862	**.089***	.008	1631
Self-reported grades	-.080	.006	866	-.045	.002	1635
Overall institutional experience	.050	.003	869	.008	.000	1627

Analyses weighted by gender, enrollment status, and institution size *p<.002

Furthermore, there is no significant relationship between social desirability and the individual items of self-reported grades and overall institutional evaluation. For first-year students, there were significant positive relationships between Level of Academic Challenge and Reflective Learning. For senior students, there were significant positive relationships between social desirability and Supportive Campus Environment, Reflective Learning, Gains in Personal and Social Development, and Gains in General Education.

When significant relationships are detected, the magnitude of the relationships should also be considered as an indicator of effect size. With large samples such as this one, it is important to distinguish statistical significance, which is more easily detected, from practical significance, which is informed by effect size. The magnitude of the significant relationships is very small, according to Cohen's (1992) guidelines, with the largest correlation at .141. Interpreting a correlation coefficient to provide an estimate of explained variance, squaring the r values suggests that social desirability explains only 0.8 to 2.0% of the variance in scores on the statistically significant subscales.

Regression analyses
The results summarizing Step 2 of the 26 separate regression models can be found in Table 2. Analyses were completed for first-year and senior students, using each of the Benchmarks, Deep Approaches to Learning subscales, and Gains subscales as the dependent variables. The independent variables for Step 1 included gender, enrollment status, first generation status, transfer status, athlete status, living on campus, Greek status, international status, ethnicity, self-reported grades, and overall institutional experience. The independent variables for Step 2 included all of the control variables from Step 1, and social desirability score. Regressions using grades and institutional experience as the

Table 2: *Summary of Step 2 Regression Results by Class*[1]

Dependent Variable	First-Year				Senior			
	b	*SE b*	β	ΔR^2	*b*	*SE b*	β	ΔR^2
Level of Academic Challenge	.447	.151	.107	.010	.082	.105	.020	.000
Active & Collaborative Learning	.427	.207	.076	.005	.248	.119	.051	.002
Student-Faculty Interaction	.242	.238	.038	.001	.057	.149	.010	.000
Enriching Educational Experiences	.099	.158	.023	.000	.023	.117	.005	.000
Supportive Campus Environment	.092	.207	.014	.000	.623	.126	**.113***	.011
Higher-Order Learning	.395	.264	.056	.003	.024	.161	.004	.000
Reflective Learning	.786	.297	.099	.009	.479	.177	.070	.004
Integrative Learning	.466	.227	.073	.005	.290	.141	.051	.002
Gains in Practical Competence	.298	.255	.038	.001	.399	.147	.062	.003
Gains in Personal & Social Development	.180	.277	.022	.000	.672	.181	**.089***	.007
Gains in General Education	.174	.255	.022	.000	.345	.156	.051	.002
Self-reported grades	-.304	.021	-.057	.003	-.008	.011	-.018	.000
Overall institutional experience	.021	.010	.079	.006	.014	.006	.064	.004

Analyses weighted by gender, enrollment status, and institution size *p<.002

dependent variables were also completed. For the analyses using grades as the dependent variable, the independent variables for Step 1 included gender, enrollment status, first generation status, transfer status, athlete status, living on campus, Greek status, international status, ethnicity, and overall institutional experience; the independent variables for Step 2 included all of the control variables from Step 1, and social desirability score. For the analyses using institutional experience as the dependent variable, the independent variables for Step 1 included gender, enrollment status, first generation status, transfer status, athlete status, living on campus, Greek status, international status, ethnicity, and self-reported grades; the independent variables for Step 2 included all of the control variables from Step 1, and social desirability score.

These results of the regression models indicate that for first-year students, social desirability does not significantly predict scores on any of the Benchmarks, Deep Approaches to Learning subscales, Gains subscales, self-reported grades, or institutional experience even after controlling for gender, enrollment status, first generation status, transfer status, athlete status, living on campus, Greek status, international status, ethnicity, self-reported grades, and overall institutional experience. For senior students, social desirability bias does not significantly predict scores on any of the Deep Approaches to Learning subscales, Level of Academic Challenge, Active and Collaborative Learning, Student-Faculty Interaction, Enriching Educational Experiences, Gains in General Education, Gains in Practical Competence, self-reported grades, or institutional experience even after controlling for gender, enrollment status, first generation status, transfer status, athlete status, living on campus, Greek status, international status, ethnicity, self-reported grades, and overall institutional experience. The only models for which social desirability was a significant predictor for senior

students were those with Supportive Campus Environment and Gains in Personal and Social Development as the dependent variables. These findings were consistent with the results from the bivariate correlations for these variables. The standardized beta coefficients for these significant relationship were relatively small (β = .113 for Supportive Campus Environment; β = .089 for Personal and Social Development), and there was a change in R^2 of only .011 and .007, respectively, after social desirability score was entered in the model for Step 2. This very small change in the percent of explained variance from the models after social desirability score is entered indicates that a tendency to respond in a socially desirable manner is having very little, if any, impact on responses to the NSSE items. The other significant bivariate correlations with social desirability (Level of Academic Challenge and Reflective Learning for first-years; Reflective Learning and Gains in General Education for seniors) did not show a significant relationship in the regression model after controlling for the other independent variables.

Discussion

These results suggest that while a few NSSE benchmarks and subscales have significant relationships with a measure of social desirability, the overall influence on survey response is minimal. First-year students' scores on Level of Academic Challenge and Reflective Learning were significantly correlated with social desirability, and senior students' scores on Supportive Campus Environment, Reflective Learning, Gains in Personal and Social Development, and Gains in General Education were significantly correlated with social desirability. Of the 26 separate correlations that were conducted, only 6 were significant. Of these six significant correlations, the effect size was trivial, explaining at most two percent of the variance. When social desirability scores were

included in regression models that controlled for a variety of demographic characteristics that were known to be related to benchmark and subscale scores, social desirability bias had even less influence.   After controlling for demographic characteristics, social desirability was not a significant predictor of any of the benchmarks or subscales for first-year students.   For senior students, social desirability was a significant predictor for Supportive Campus Environment and Gains in Personal and Social Development, but contributed only minimally to changes in the explained variance of the model.

Pike (1999) found that first-year students were more likely than seniors to exhibit halo effects, which occur when perceptions of positive qualities about one thing results in the perceptions of positive qualities about all other related things. However, this study did not find strong evidence to extend these findings concerning halo effects to social desirability bias in first-year students. Instead, there appears to be more of an influence of social desirability for senior students, as the seniors in the sample had more significant correlations and regression coefficients. While it may be true that the first-year students in this sample perceived their engagement, learning, and gains as high due to a positive overall experience at their institution, the results suggest that they generally were not engaging in this behavior due to a desire to respond with socially appropriate answers. It may be that as first-year students, there is less of a social obligation to appear to have high levels of engagement, learning, or gains in development. There was a significant correlation for social desirability and academic challenge for first-years, which may be reflective of a social obligation that college is supposed to be academically difficult, compared with their high school experiences. They hear this message from high school teachers, admissions counselors, and various staff during orientation, and perhaps

this communicates a social obligation of what their experience "should" be like.

For senior students, the pattern of results was very similar, with a few exceptions. It may be that seniors, after spending several years in higher education at their institutions, feel more of a social obligation to respond that their institution has contributed to engagement, learning, and gains in these areas. Also, seniors may be more cognizant that their institutions want them to show a positive attitude about their institutional experiences. After all, if their campus environment is not supportive, how should they justify spending so much time there? Furthermore, if they have not made gains in their skills and competencies, did they just waste all of the money spent on tuition?

Potential reasons for the significant correlations for both first-year and senior students between social desirability and reflective learning should be contemplated as well. Although these correlations were very weak, and were no longer significant when controlling for other demographic variables, a brief consideration may be beneficial. The NSSE reflective learning items (i.e. "How often have you… Examined the strengths and weaknesses of your own views on a topic or issue") rely on reporting the frequency of more cognitive activities, as opposed to the reporting of outwardly observable behaviors (i.e. "How often have you… Made a class presentation"). It may be the case that different types of behaviors, some of which are more difficult to recall and report than others, are more influenced by social desirability.

Limitations
There are some limitations of this research that should be considered when interpreting the results. While there were a variety of institutions that were included in this sample, they should not be considered as representative of all undergraduate institutions in the entire country. In addition,

this study includes only first-year and senior students, so generalizations to sophomore and junior students should be made with caution. Data was collected from six institutions, so while the results are not institution-specific, only having six institutions in the sample did not allow for examinations of the influences of institutional characteristics, such as Carnegie classification, geographic region, or institutional control (public versus private). It could be that social desirability has a disparate impact depending on the type of institution. Finally, some of the analyses may have been affected by the lower than desirable Cronbach's alphas for a few of the measures. Better evidence for the internal consistency of the benchmarks and subscales could strengthen the conclusions drawn from the results of the analyses.

Future Research
Further research on the topic of social desirability bias should be done in order to gain a clearer picture about the precise function of social desirability bias in self-report surveys. It might be important to identify differences in social desirability bias among various student characteristics. Do some students have a stronger tendency to respond in a socially desirable manner? Exploring potential differences for gender, enrollment status, place of residence, transfer status, and ethnicity, among others may reveal critical pieces of information that contribute to greater understanding of bias in student surveys. It may also be interesting to investigate the potential influence of social desirability in other types of student self-report data. After all, students complete many types of self-report instruments for their institutions. Does social desirability bias influence responses to course evaluations? Are surveys concerning social activities more or less influenced by social desirability bias than surveys on academic activities? Does one's interpretation of social

desirability bias depend on the type of question that is being asked, not only concerning the sensitivity of the topic, as is suggested in previous literature, but also on the difficulty of responding?

Recently, there has been some criticism concerning the use of student self-report surveys in higher education (Porter, Rumann, & Pontius, 2011). While social desirability bias should continue to be a concern for any researchers employing a survey methodology, this study did not find strong support for the influence of social desirability on the National Survey of Student Engagement. There was little evidence for social desirability bias for first-year students, especially after controlling for demographic characteristics. The significant relationships for social desirability and select benchmarks and subscales for senior students had very small effect sizes, and explained very little of the variance. The lack of significant relationships between social desirability and a majority of the benchmarks and subscales in this study, along with minimal effect sizes, suggest that social desirability bias is not a primary threat to the quality of student self-report survey responses concerning engagement, learning, and gains. These findings are important to consider, due to the widespread use of NSSE and many other self-report surveys in higher education. The trend for survey research is very likely to continue in the field, given the move toward increased accountability and assessment, so it is essential that we also continue to examine the quality of the data that is collected.

References

Bowman, N.A., & Hill, P.I. (2011). Measuring how college affects students: Social desirability and other potential biases in college student self-reported gains. In S. Herzog & N.A. Bowman (Eds.), *Validity and Limitations of College Student Self-Report Data* (pp. 73-85).

New Directions for Institutional Research. doi:10.1002/ir.390

Carpenter, C.M. (2009). Development of a drug use resistance self-efficacy (DURSE) scale. *American Journal of Health Behavior, 33(2),* 147-157.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychotherapy. *Journal of Counseling Psychology, 24,* 349-354.

DeVellis, R.F. (2003). *Scale development: Theory and applications* (2nd ed). Thousand Oaks, CA: Sage Publications.

Ferrari, J.R. (2005). Impostor tendencies and academic dishonesty: Do they cheat their way to success? *Social Behavior and Personality, 33,* 11-18.

Ferrari, J.R., Bristow, M., & Cowman, S.E. (2005). Looking good or being good? The role of social desirability tendencies in student perceptions of institutional mission and values. *College Student Journal, 39,* 7-14.

Ferrari, J.R. & Cowman, S.E. (2004). Toward a reliable and valid measure of institutional mission and values perception: the DePaul Values Inventory. *Journal of Beliefs and Values, 25,* 43-54.

Ferrari, J.R., McCarthy, B.J., & Milner, L.A. (2009). Involved and focused? Students' perceptions of institutional identity, personal goal orientation and levels of campus engagement. *College Student Journal, 43,* 886-896.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications.

Hancock, D.R. & Flowers, C.P. (2001). Comparing social desirability responding on world wide web and paper-administered surveys. *Educational Technology Research & Development, 49,* 5-13.

Kelly, W.E. (2003). As achievement sails the river of time: The role of time use efficiency in grade point average. *Educational Research Quarterly, 27(4)*, 3-8.

Kuh, G. D. & Ewell, P. T. (2010). The state of learning outcomes assessment in the United States. *Higher Education Management and Policy, 22*(1), 1-20.

Luo, W., Watkins, D., & Lam, R.Y.H. (2009). Validating a new measure of self-complexity. *Journal of Personality Assessment, 91,* 381-386.

McMillan, J.H., & Schumacher, S. (2001). *Research in education: A conceptual introduction.* New York: Longman.

Nauta, M.M. (2007). Assessing college students' satisfaction with their academic majors. *Journal of Career Assessment, 15,* 446-462.

Nasco, S.A. & Webb, W.M. (2006). Toward an expanded measure of athletic identity: The inclusion of public and private dimensions. *Journal of Sport & Exercise Psychology, 28,* 434-453.

Paulhus, D.L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598-609.

Pike, G.R. (1999). The constant error of the halo in educational outcomes research. *Research in Higher Education, 40,* 61-86.

Porter, S.R., Rumann, C. & Pontius, J. (2011). The validity of student engagement survey questions: Can we accurately measure academic challenge? In S. Herzog & N.A. Bowman (Eds.), *Validity and Limitations of College Student Self-Report Data* (pp. 87-98). New Directions for Institutional Research. doi:10.1002/ir.391

Ray, J.J. (1984). The reliability of short social desirability scales. *The Journal of Social Psychology, 123,* 133-134.