AEA Papers and Proceedings 2018, 108: 1–6
https://doi.org/10.1257/pandp.20181055

A New Look at Lake Wobegon: Who's in Your Canoe?[†]

By Chris Bollinger, Tisha L. N. Emerson, Linda English, and Gail M. Hoyt*

> …*Lake Wobegon, where all the women are strong, all the men are good looking, and all the children are above average.*
> —Garrison Keillor, *A Prairie Home Companion* (Lee 1991)

Over 20 years ago, Maxwell and Lopus (1994) identified what has since been referred to as the "Lake Wobegon" effect in economic education research. They found that students generally tend to overstate their academic accomplishments and that low-achieving students are unlikely to report their achievement at all, with the combined effect being an upward bias in scores. In research from the broader health literature, misreporting of certain behaviors is often tied to survey conditions, Brener, Billy, and Grady (2003). For instance, Hoyt and Chaloupka (1994) find that adolescents overreport illicit substance use when surveyed in the presence of a friend and they systematically underreport use when a family member is present.

Given that student self-reported data is a ubiquitous feature of empirically based economic education research, the potential for systematic misreporting raises significant concerns about biased estimates and misinterpretation of results when nonrandom reporting occurs. In fact, Becker and Powers (2001) argue against using student-provided data for aptitude measures due to their unreliability. Systematic self-reporting is especially troubling if the economic research evaluates the impact of educational inputs, pedagogical techniques, or course policies, and potentially biased results lead educators to implement misguided policy.

Using data from 509 students surveyed at two universities, we too find the Lake Wobegon effect. We consider factors that influence the degree and direction of the potential misreporting of current cumulative GPA. Consistent with past studies, we examine differences in means and use OLS estimation. Going beyond earlier work, as our survey was administered in the classroom, we incorporate information about situational factors relevant to the student while completing the survey. We also evaluate density, deciles, and use quantile analysis to more completely characterize the sources of the Lake Wobegon effect. This additional analysis shows that what might at first glance appear to be a difference in mean values might actually be driven by heteroscedasticity associated with the GPA, which differs across gender. Also, survey conditions have a systematic impact on self-reported values of academic performance. A better understanding of these influences might more aptly inform economic education research in both study techniques and conclusions drawn from data.

I. Data and Methodology

The 509 study participants were enrolled in one of five sections of principles of microeconomics taught at Baylor University and at the University of Kentucky in the fall semester of 2012.[1] At the beginning of the course, participants completed surveys in the classroom soliciting information about gender, ethnicity, age, high school economics background, state in which the student attended high school, intended major, and motivation behind enrolling

*Bollinger: Department of Economics, University of Kentucky, Lexington, KY 40506 (email: crboll01@uky.edu); Emerson: Department of Economics, Baylor University, Waco, TX 76798 (email: Tisha_Nakao@baylor.edu); English: Department of Economics, Baylor University, Waco, TX 76798 (email: Linda_English@baylor.edu); Hoyt: Department of Economics, University of Kentucky, Lexington, KY 40506 (email: ghoyt@uky.edu). We thank Daniel Tannenbaum for his helpful comments.

[†] Go to https://doi.org/10.1257/pandp.20181055 to visit the article page for additional materials and author disclosure statement(s).

[1] The microeconomic principles course serves as a required prerequisite for taking macroeconomic principles at both Baylor University and the University of Kentucky.

in the principles course. Students' SAT (and/or ACT) scores and GPAs were collected directly from students on the surveys and acquired directly from university student records.[2] To account for survey conditions that might potentially influence self-reported data, students were asked if they were sitting next to a good friend, a casual acquaintance, someone they did not know, or some combination. Additionally, we administered the third edition of the microeconomics version of the Test of Understanding in College Economics (TUCE) prior to providing any course instruction.[3]

Table 1 presents the means of the main variables used in the analysis for both men and women. The record, or official, GPA is quite similar, but the difference in average reported GPA is 0.21, nearly a quarter of a grade, and is statistically significantly different. We also note that men are statistically significantly more likely to report sitting next to a friend than women, while women are statistically significantly more likely to report sitting next to a stranger.

Table 1—Sample Means by Gender

Variable	Men	Women
Student report GPA	3.054	3.264
Record GPA	2.998	3.048
SAT	1,151	1,118
Pre-TUCE score	9.944	9.441
Hispanic	0.070	0.068
Black	0.052	0.059
Asian	0.045	0.059
Other race	0.021	0.050
Sitting near acquaintance	0.150	0.185
Sitting near friend	0.575	0.446
Sitting near stranger	0.317	0.455
Sample size	287	222

II. OLS Estimation

We begin by reporting OLS results in Table 2. We choose a specification in which reported GPA is the dependent variable and the official record GPA is a regressor. We select this specification over using the difference between reported and official scores as it represents a cognitive reporting relationship: individuals are reporting conditional upon the actual rather than making decisions on how they will make errors. While the male and female coefficients are not statistically significantly different at conventional levels, there are notable differences in the estimates which are economically significant.[4] The two intercepts reflect the difference seen overall in the means in Table 1 (the p-value here is 0.057). Women report a higher GPA, all else equal, including their actual recorded GPA. Women are also more responsive to differences in their recorded GPA as reflected in the coefficient on the record GPA.

Table 2—OLS Estimates of GPA Regressions **AQ 1**

	Male	Female
Record GPA	0.300	0.354
	(5.06)	(4.84)
SAT	−0.0001	−0.0005
	(0.26)	(1.86)
Pre-TUCE score	0.024	0.011
	(2.78)	(1.20)
Hispanic	−0.015	−0.040
	(0.17)	(0.31)
Black	−0.238	−0.202
	(1.94)	(1.71)
Asian	−0.010	0.199
	(0.07)	(1.79)
Other race	0.209	0.004
	(0.86)	(0.02)
Near acquaintance	0.101	0.008
	(1.27)	(0.10)
Near friend	0.154	0.069
	(2.44)	(0.84)
Near stranger	0.095	0.028
	(1.36)	(0.36)
Intercept	1.854	2.577
	(7.23)	(8.77)
R^2	0.18	0.17
Observations	287	222

The most interesting aspect, however, is the difference between men and women in how they

[2] ACT measures were converted into SAT equivalents using http://www.act.org/aap/concordance/.

[3] The TUCE is a standardized test of economic knowledge aimed at principles level college students. See Saunders (1991) for additional information about the TUCE. To incentivize effort on the TUCE, students were given bonus points based on the number of correct answers.

[4] While many focus on statistical significance, the best estimate of the parameter is still provided by the estimated difference. A difference of nearly a quarter point is important in understanding average grade and, in particular, in the contest of measurement error models. See Ziliak and McCloskey (2008) and Wasserstein and Lazar (2016).

respond to the proximity of others. The three variables are not mutually exclusive (one can be sitting "near" at least two people). The omitted category of "sitting near no one" provides the reference. On average, men report higher GPAs when sitting next to other individuals. This is most pronounced when men are sitting next to a friend. While the female coefficient on sitting next to a friend is the highest for women as well, it is less than half the coefficient for men. As we saw in Table 1, men are also more likely to be sitting next to a friend. The Lake Wobegon phenomenon appears to manifest itself differently. Women overreport generally, while men overreport when they are in proximity to others.

In Figure 1, we plot both the male and female difference between self-reported GPA and record GPA. Here we see the shift up for women near the mean, but we also see that this difference depends on the point in the distribution. Men have a slightly longer right tail, suggesting that while on average women overreport, many of the most egregious overreports are from men. We also note that part of the reason men's average report is closer to the record is that while many men overreport, a large portion also underreport and this underreporting is more pronounced than for women. Although not reported in Table 1, the standard deviation of the difference for men is slightly larger than for women.

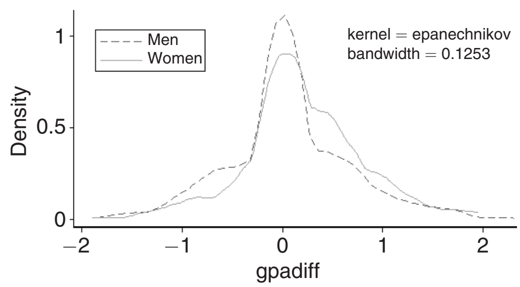

Figure 1. Density of GPA Difference

AQ 2

III. Quantile Regression Estimation

To examine the role of survey conditions in the response profile, we estimate quantile regressions by gender. Quantile regression measures the quantile or percentile of the distribution conditional on a set of variables. By estimating these regressions at various quantiles we can understand how the entire distribution is changing. As we shall see, in some cases, the mean difference observed in Table 2, is dominated by only one part of the distribution. One can think of these as lines parallel (or potentially parallel) to the OLS regression line which passes through the mean of the conditional distribution. Low quantiles pass through the lower end of the "residual" distributions, below the mean line, while high quantiles pass through the high end of the "residual" distributions above the regression line. If the residuals in the OLS regression were perfectly homoscedastic and symmetric, all quantile regression lines would be parallel to the OLS line.

Table 3 presents the unconditional quantiles of the difference between the reported and recorded GPA. We note (and the general observation is true for each group) that the median for women is slightly positive, reflecting the general shift up in the entire distribution, while the median for men is zero. Quantiles of 30 percent and lower reflect an underreport of GPA by the respondent while quantiles of 60 percent and above reflect an overreport. This aids in interpreting the slope coefficients below.

We look first at the GPA quantiles for men in Table 4. The coefficient on GPA record is important. The coefficient at the low end of the distribution is much larger than at the high end. This indicates, as one might expect, that there is heteroscedasticity around the OLS line, with more spread around low recorded GPA and a tighter spread at the higher reported GPA. Thus those who are underreporting or overreporting their GPA the most have low record GPA. Some of this result likely derives from the upper bound of GPA (see Haley, Johnson, and McGee 2010).

Examining the coefficient on sitting next to a friend, we see higher values for the lower quantiles. This implies that those who would report below their actual GPA but are sitting next to a friend, underreport that GPA by less than those who underreport that GPA and are not sitting next to anyone. The smaller coefficient on this variable at the high quantile implies that sitting next to a friend induces higher reporting there as well, which exacerbates their already higher report.

In all cases, the median (50th percentile) has the smallest coefficient. Those who are near the middle of the reporting distribution are least

Table 3—Reporting Difference Centiles by Gender

Centile	All	Men	Women
10	−0.640	−0.697	−0.451
20	−0.250	−0.364	−0.176
30	−0.047	−0.091	−0.020
40	0.000	0.000	0.000
50	0.030	0.000	0.103
60	0.145	0.050	0.315
70	0.370	0.238	0.481
80	0.587	0.535	0.624
90	0.909	0.839	0.969

Table 4—Male GPA Regression Quantile Estimates

	10	25	50	75	90
Record GPA	0.405	0.384	0.323	0.335	0.262
	(4.87)	(4.59)	(5.45)	(4.78)	(3.30)
SAT	−0.000	−0.000	0.000	−0.000	−0.000
	(1.21)	(0.20)	(0.54)	(1.10)	(1.10)
TUCE score	0.055	0.045	0.018	0.014	0.005
	(3.80)	(3.08)	(1.71)	(1.11)	(0.39)
Hispanic	0.180	0.116	−0.035	−0.101	−0.134
	(1.01)	(0.64)	(0.28)	(0.67)	(0.79)
Black	−0.328	−0.301	−0.177	−0.171	−0.294
	(1.60)	(1.45)	(1.21)	(0.99)	(1.50)
Asian	−0.016	−0.137	−0.055	−0.218	−0.178
	(0.07)	(0.63)	(0.36)	(1.20)	(0.87)
Other race	−0.024	0.154	0.081	−0.035	0.548
	(0.08)	(0.49)	(0.37)	(0.13)	(1.84)
Acquaintance	0.193	0.124	0.078	0.050	0.119
	(1.41)	(0.90)	(0.80)	(0.43)	(0.91)
Friend	0.181	0.178	0.104	0.188	0.087
	(1.59)	(1.55)	(1.28)	(1.96)	(0.80)
Stranger	0.113	0.101	0.078	0.108	0.242
	(0.95)	(0.84)	(0.91)	(1.08)	(2.13)
Intercept	1.053	1.132	1.692	2.510	3.102
	(2.46)	(2.63)	(5.55)	(6.96)	(7.60)
Observations	287	287	287	287	287

affected by the presence of people around them. It is those making extreme reports already who are changing their behavior the most in the presence of other individuals.

In Table 5, we examine the same regressions for women. We see an even more pronounced pattern on the recorded GPA variable for women, indicating even larger heteroscedasticity associated with the level of GPA for women. In contrast to the men, the coefficients on the variables measuring proximity display different patterns. For the proximity of a friend, the coefficient is highest at the low quantiles, indicating less likelihood to underreport when near a friend (similar to men, but with less impact),

Table 5—Female GPA Regression Quantile Estimates

	10	25	50	75	90
Record GPA	0.409	0.562	0.501	0.330	0.114
	(1.98)	(6.23)	(6.82)	(3.82)	(1.70)
SAT	−0.000	−0.000	−0.001	−0.000	−0.000
	(0.14)	(1.03)	(1.87)	(1.21)	(1.17)
Pre-TUCE score	0.049	0.017	0.011	0.011	−0.003
	(1.48)	(1.17)	(0.91)	(0.79)	(0.32)
Hispanic	−0.192	−0.258	−0.087	0.008	−0.102
	(0.47)	(1.43)	(0.59)	(0.05)	(0.77)
Black	−0.304	−0.057	−0.097	−0.291	−0.318
	(0.68)	(0.29)	(0.61)	(1.57)	(2.21)
Asian	0.014	0.300	0.149	0.135	0.055
	(0.03)	(1.56)	(0.95)	(0.73)	(0.39)
Other race	0.130	0.016	−0.081	0.082	0.001
	(0.27)	(0.08)	(0.48)	(0.41)	(0.01)
Acquaintance	0.013	0.046	−0.080	−0.002	−0.065
	(0.05)	(0.38)	(0.81)	(0.02)	(0.71)
Friend	0.080	−0.010	0.004	0.067	-0.021
	(0.33)	(0.09)	(0.05)	(0.66)	(0.27)
Stranger	0.074	−0.038	−0.001	0.072	0.068
	(0.31)	(0.36)	(0.01)	(0.72)	(0.87)
Intercept	1.055	1.582	2.357	2.931	3.917
	(1.08)	(3.71)	(6.78)	(7.19)	(12.42)
Observations	222	222	222	222	222

and at the high end it is negative, indicating less likelihood to overreport when sitting next to a friend. Women in general appear to report more accurately when sitting near friends. Caution should be exercised as these are both small in magnitude and statistical significance. However, sitting near a stranger makes an underreporting women (low quantile) less likely to underreport, but makes an overreporting woman (high quantiles) more likely to overreport. **AQ 3**

IV. Concluding Remarks

Haley, Johnson, and McGee (2010) notes that mismeasurement of GPA can be detrimental to estimation when GPA is used as a regressor. This is even more important when mismeasurement differs among individuals and drawing conclusions about educational interventions will be suspect. We note, in particular, that many interventions are targeted at the lower end of the GPA distribution, where response error is largest. We also note that the variation of response error can depend upon other factors. The weak association between self-reported and actual GPA, coupled with heteroscedasticity point to **AQ 4** nonclassical measurement error. Using classical measurement error, and constant error variance to correct for this would be suspect. Examining the impact of response error throughout the reported GPA distribution allows us to understand how those conclusions may be impacted. We note that response error is most prevalent in the tails of the distribution. We also note that how that error manifests itself differs between men and women. With these issues in mind, we recommend that researchers use administrative data, as opposed to self-reported, when possible. We also suggest considering fielding surveys in a different manner. Typically, student surveys are passed out during the first day of class. It appears that sitting next to friends may impact survey reliability. While further investigation is warranted, an online survey may reduce this problem. Finally, while our data only speak to response error in GPA reporting, there is reason to believe that response error exists with other measures as well. As such, we caution researchers to adopt procedures to minimize (possibly through random seating assignments or employing empty alternative seats) and consider possible effects of remaining response error on study outcomes.

REFERENCES

Becker, William E., and John R. Powers. 2001. "Student Performance, Attrition, and Class Size Given Missing Student Data." *Economics of Education Review* 20 (4): 377–88.

Brener, Nancy D., John O. Billy, and William R. Grady. 2003. "Assessment of Factors Affecting the Validity of Self-Reported Health-Risk Behavior Among Adolescents: Evidence from the Scientific Literature." *Journal of Adolescent Health* 33 (6): 436–57.

Haley, M. Ryan, Marianne F. Johnson, and M. Kevin McGee. 2010. "A Framework for Reconsidering the Lake Wobegon Effect." *Journal of Economic Education* 41 (2): 95–109.

Hoyt, Gail Mitchell, and Frank J. Chaloupka. 1994. "Effect of Survey Conditions on Self-reported Substance Use." *Contemporary Economic Policy* 12 (3): 109–21.

Lee, Judith Yaross. 1991. *Garrison Keillor: A Voice of America*. Jackson, MS: University Press of Mississippi.

Maxwell, Nan L., and Jane S. Lopus. 1994. "The Lake Wobegon Effect in Student Self-Reported Data." *American Economic Review* 84 (2): 201–05.

Saunders, Philip. 1991. "The Third Edition of the Test of Understanding in College Economics." *Journal of Economic Education* 22 (3): 255–72.

Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose." *American Statistician* 70 (2): 129–33.

Ziliak, Stephen T., and Deirdre N. McCloskey. 2008. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor: University of Michigan Press.

AUTHOR, PLEASE ANSWER ALL QUERIES (numbered with "AQ" in the margin of the page).

AQ#	Question	Response
1.	If you added "self-reported" in front of GPA in the title it would make it clear that you are regressing reported on record.	
2.	If you specify that difference means reported—record, it would be clearer.	
3.	Should this be an "under-reporting man" rather than "women"?	
4.	Is this what you really mean to say? [Mismeasurement} is even more important when .... drawing conclusions about educational interventions will be suspect?	