

Notes on Student Evaluations of Teaching (SET)

P.B. Stark, 17 September 2019.

Last edited 8 September 2020.

Experimental and quasi-experimental results

- weak or negative association with objective measures of learning (Carrell & West, 2010; Braga et al., 2014; Boring et al., 2016)
- substantial bias from gender
 - gender bias can make female instructors rate worse than objectively less effective male instructors (Boring et al., 2016)
 - bias affects ratings of “objective” items like promptness (MacNell et al., 2015; Boring et al., 2016)
 - varies by discipline &c (Boring et al., 2016; Mengl et al., 2018)
- bias from ethnicity & gender (Chisadza et al. 2019)
- strong association with grade *expectations*, but not necessarily with learning (Boring et al., 2016)
- grades–not learning–“rewarded” with high SET (Cho et al., 2015; Carrell & West, 2010; Braga et al., 2014)
- providing cookies during class increases ratings of instructors and of course materials (Hessler et al., 2018)
- the number of points on the rating scale affects gender differences (Rivera & Tilcsik, 2019)

Laboratory studies

- bias in favor of young male instructors (Arbuckle & Williams, 2003)
- ratings predicted by responses to 30 seconds of silent video (Ambady & Rosenthal, 2003)
- race and gender matter (Basow et al., 2013)

Meta-analyses

- weak or negative association with objective measures of learning (Uttl et al., 2016)
- association between author conflicts of interest and conclusions that SET are valid (Uttl et al., 2019)

Observational studies and $n = 1$ experiments

- strong association with student enjoyment (Stark, unpublished)
- data unreliable: substantial fraction of students give demonstrably–apparently deliberately–false answers to objective questions (Stanfel, 1995)
- gendered language in evaluations (Schmidt, 2015, *inter alia*)

- bias against older instructors and female instructors (Bianchini et al., 2013; Wagner et al., 2016)
- bias against non-native English speakers, (Subtirelu, 2015, *inter alia*), URM (Wagner et al., 2016)
- bias in favor of physically attractive instructors (Wolbring & Riordan, 2016; Feeley, 2002; Hamermesh & Parker, 2004)
- biases from physical condition of room, time of day, mathematical level, class size, ... (Bedard & Kuhn, 2005, *inter alia*)
- “halo effect”: students conflate enthusiasm, attractiveness, & other things with effectiveness; enthusiasm not associated with learning (Williams & Ceci, 1997; Feeley, 2002; *inter alia*)
- negative association with learning (Stroebe, 2016)
- association with attractiveness, esp. for female instructors (Babin et al. 2020)

Surveys

- students deliberately falsify ratings (Clayson and Haley, 2011)
- students and faculty use the same adjectives differently (Lauer, 2012)
- comments incommensurable across disciplines (Stark & Freishtat, 2014)
- bias against quantitative classes (Uttl et al., 2013)

Statistical abuses (Stark & Freishtat, 2014)

- averages of categorical data are meaningless/misleading
- response rate matters
- report distributions, not means
- SET are an incomplete census, not a random sample. Nonresponse bias not ignorable.

Litigation/Arbitration

- U. Florida
- Miami U. <https://casetext.com/case/brunarski-v-miami-univ-2>
- Ryerson U. <https://ocufa.on.ca/blog-posts/significant-arbitration-decision-on-use-of-student-questionnaires-for-teaching-evaluation/>
- U. Toronto
- UNLV
- American Sociological Association seeks to crowd fund a class action <https://www.change.org/p/american-sociological-association-end-the-use-of-biased-student-evaluations-of-teaching-in-employment-decisions-e3ff4761-0d64-4b51-9fce-f160d743e690/sign>

Policy changes at other universities

- USC <https://academic senate.usc.edu/teaching-evaluations-update/>
- U. Oregon <https://provost.uoregon.edu/revising-uos-teaching-evaluations>

- Colorado State, University of Kansas, ...

Selected References

- Ambady, N., and R. Rosenthal, 1993. Half a Minute: Predicting Teacher Evaluations from Thin Slices of Nonverbal Behavior and Physical Attractiveness, *J. Personality and Social Psychology*, 64, 431-441.
- Arbuckle, J. and B.D. Williams, 2003. Students' Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations, *Sex Roles*, 49, 507-516. DOI 10.1023/A:1025832707002
- Archibeque, O., 2014. Bias in Student Evaluations of Minority Faculty: A Selected Bibliography of Recent Publications, 2005 to Present. <http://library.auraria.edu/content/bias-student-evaluations-minority-faculty> (last retrieved 30 September 2016)
- Babin, J.J., A. Hussey, A. Nikolsko-Rzhevskyy, and D.A. Taylor, 2020. Beauty Premiums Among Academics, *Economics of Education Review*, 78, 102019. <https://doi.org/10.1016/j.econedurev.2020.102019>
- Basow, S., S. Codos, and J. Martin, 2013. The Effects of Professors' Race and Gender on Student Evaluations and Performance, *College Student Journal*, 47(2), 352-363.
- Basow, S., and Silberg, N.T., 1987. Student Evaluations of College Professors: Are Female and Male Professors Rated Differently?, *Journal of Educational Psychology*, 79, 308-314.
- Bedard, K., and P. Kuhn, 2005. Where Class Size Really Matters: Class Size and Student Ratings of Instructor Effectiveness, Department of Economics, University of California, Santa Barbara. <http://econ.ucsb.edu/~kelly/ucsb4.pdf> (last retrieved 6 October 2016)
- Bianchini, S., F. Lissoni, and M. Pezzoni, 2013. Instructor Characteristics and Students' Evaluation of Teaching Effectiveness: Evidence from an Italian Engineering School. *European Journal of Engineering Education*, 38, 38-57. DOI: 0.1080/03043797.2012.742868
- Boring, A., 2015. Gender Bias in Student Evaluations of Teachers, OFCE-PRESAGE-Sciences-Po Working Paper, <http://www.ofce.sciences-po.fr/pdf/dtravail/WP2015-13.pdf> (last retrieved 30 September 2016)
- Boring, A., K. Ottoboni, and P.B. Stark, 2016. Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness, *ScienceOpen*, DOI 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1
- Braga, M., M. Paccagnella, and M. Pellizzari, 2014. Evaluating Students' Evaluations of Professors, *Economics of Education Review*, 41, 71-88.
- Carrell, S.E., and J.E. West, 2010. Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors, *J. Political Economy*, 118, 409-432.
- Chisadza, C., N. Nicholls, and E. Yitbarek, 2019. Race and Gender biases in Student Evaluations of Teachers, *Economics Letters*, 179, 66-71, DOI 10.1016/j.econlet.2019.03.022.
- Cho, W., W. Baek, and J. Cho, 2015. Why do good performing stu-

- dents highly rate their instructors? Evidence from a natural experiment. *Economics of Education Review*, 49, 172-179.
- Clayson, D.E. and D.A. Haley, 2011 Are Students Telling Us the Truth? A Critical Look at the Student Evaluation of Teaching. *Marketing Education Review*, 21, 101-112.
 - Fan Y., L.J. Shepherd, E. Slavich, D. Waters, M. Stone, R. Abel, and E.L. Johnston, 2019. Gender and cultural bias in student evaluations: Why representation matters. *PLoS ONE*, 14, e0209749. <https://doi.org/10.1371/journal.pone.0209749>
 - Feeley, T.H., 2002. Evidence of Halo Effects in Student Evaluations of Communication Instruction, *Communication Education*, 51:3, 225-236, DOI: 10.1080/03634520216519
 - Hamermesh, D. S., and A. Parker, 2004. Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24(4), 369-376. <https://www.sciencedirect.com/science/article/abs/pii/S0272775704001165>
 - Hessler, M., D.M. Pöpping, H. Hollstein, H. Ohlenburg, P.H. Arnemann, C. Massoth, L.M. Seidel, A. Zarbock & M. Wenk, 2018. Availability of cookies during an academic course session affects evaluation of teaching, *Medical Education*, 52, 1064–1072. doi: 10.1111/medu.13627
 - Hill, M.C., and K.K. Epps, 2010. The Impact of Physical Classroom Environment on Student Satisfaction and Student Evaluation of Teaching in the University Environment, *Academy of Educational Leadership Journal*, 14, 65.
 - Hornstein, H.A., 2017. Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance, *Cogent Education*, 4 1304016 <http://dx.doi.org/10.1080/2331186X.2017.1304016>
 - Johnson, V.E., 2003. *Grade Inflation: A Crisis in College Education*, Springer-Verlag, NY, 262pp.
 - Keng, S.-H., 2017. Tenure system and its impact on grading leniency, teaching effectiveness and student effort, *Empirical Economics*, DOI 10.1007/s00181-017-1313-7.
 - Lake, D.A., 2001. Student Performance and Perceptions of a Lecture-based Course Compared with the Same Course Utilizing Group Discussion. *Physical Therapy*, 81, 896-902.
 - Lauer, C., 2012. A Comparison of Faculty and Student Perspectives on Course Evaluation Terminology, in *To Improve the Academy: Resources for Faculty, Instructional, and Educational Development*, 31, J.E. Groccia and L. Cruz, eds., Jossey-Bass, 195-211.
 - Lee, L.J., M.E. Connolly, M.H. Dancy, C.R. Henderson, and W.M. Christensen, 2018. A comparison of student evaluations of instruction vs. students' conceptual learning gains, *American Journal of Physics*, 86, 531. DOI 10.1119/1.50393300
 - MacNell, L., A. Driscoll, and A.N. Hunt, 2015. What's in a Name: Exposing Gender Bias in Student Ratings of Teaching, *Innovative Higher Education*, 40, 291-303. DOI 10.1007/s10755-014-9313-4
 - Marsh, H.W., and T. Cooper. 1980. Prior Subject Interest, Students

Evaluations, and Instructional Effectiveness. Paper presented at the annual meeting of the American Educational Research Association.

- Marsh, H.W., and L.A. Roche. 1997. Making Students' Evaluations of Teaching Effectiveness Effective. *American Psychologist*, 52, 1187-1197
- Mengl, F., J. Sauermann, and U. Zölitz, 2018. Gender Bias in Teaching Evaluations. *Journal of the European Economic Association*, jvx057, DOI 10.1093/jeaa/jvx057
- Nilson, L.B., 2012. Time to Raise Questions about Student Ratings, in *To Improve the Academy: Resources for Faculty, Instructional, and Educational Development*, 31, J.E. Groccia and L. Cruz, eds., Jossey-Bass, 213-227.
- Rivera, L. and A. Tilcsik. 2019. Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation, *American Sociological Review*, in press.
- Schmidt, B., 2015. Gendered Language in Teacher Reviews, <http://benschmidt.org/profGender> (last retrieved 30 September 2016)
- Stanfel, L.E., 1995. Measuring the Accuracy of Student Evaluations of Teaching. *Journal of Instructional Psychology*, 22(2), 117-125.
- Stark, P.B., and R. Freishtat, 2014. An Evaluation of Course Evaluations, *ScienceOpen*, DOI 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1
- Stroebe, W., 2016. Why Good Teaching Evaluations May Reward Bad Teaching: On Grade Inflation and Other Unintended Consequences of Student Evaluations, *Perspectives on Psychological Science*, 11(6), 800–816. DOI 10.1177/1745691616650284
- Subtirelu, N.C., 2015. “She does have an accent but...”: Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com, *Language in Society*, 44, 35-62. DOI 10.1017/S0047404514000736
- Uttl, B., C.A. White, and A. Morin, 2013. The Numbers Tell it All: Students Don't Like Numbers!, *PLoS ONE*, 8(12): e83443, DOI 10.1371/journal.pone.0083443
- Uttl, B., C.A. White, and D.W. Gonzalez, 2016. Meta-analysis of Faculty's Teaching Effectiveness: Student Evaluation of Teaching Ratings and Student Learning Are Not Related, *Studies in Educational Evaluation*, DOI: 0.1016/j.stueduc.2016.08.007
- Uttl B., K. Cnudde, and C.A. White, 2019. Conflict of interest explains the size of student evaluation of teaching and learning correlations in multisection studies: a meta-analysis. *PeerJ* 7:e7225 <https://doi.org/10.7717/peerj.7225>
- Vasta, R., and R.F. Sarmiento, 1979. Liberal Grading Improves Evaluations but not Performance, *J. Educational Psychology*, 71, 207-211.
- Wagner, N., M. Rieger, and K. Voorvelt, 2016. Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams, Institute for Social Studies Research, Working paper 617. *Economics of Education Review*, 54, 79-94. DOI 10.1016/j.econedurev.2016.06.004.

- Williams, W.M., and Ceci, S.J., 1997. "How'm I doing?": Problems with Student Ratings of Instructors and Courses, *Change: The Magazine of Higher Learning*, 29, 12-23. DOI: 10.1080/00091389709602331
- Wolbring, T., and P. Riordan, 2016. How Beauty Works. Theoretical Mechanisms and Two Empirical Applications on Students' Evaluations of Teaching, *Social Science Research*, 57, 253-272. DOI: 10.1016/j.ssresearch.2015.12.009
- Worthington, A.C., 2002. The Impact of Student Perceptions and Characteristics on Teaching Evaluations: A Case Study in Finance Education. *Assessment and Evaluation in Higher Education*, 27, 49-64.