JOIBS: July 2023. ISSN 2992-9253

The Value Gap: A School Psychologist's Perspective on Rausch et al. (2023)

Jeffery P. Braden, NC State University, Raleigh, United States. E-mail: jpbraden@ncsu.edu

Funding: The author received no specific funding for this work.

Competing interests: The author has declared no competing interests.

Citation: Braden, J. P. (2023). The value gap: A school psychologist's perspective on Rausch et al. (2023). Journal of Open Inquiry in the Behavioral Sciences. https://doi.org/10.58408/issn.2992-9253.2023.01.01.00000009

Abstract

Rausch et al. (2023) describe an empirical effort to test a number of hypotheses put forward by Lukianoff and Haidt (2018) regarding the values of contemporary (Gen Z) vs. previous generations of undergraduates. My review of their work focuses primarily on the methodologies Rausch et al. employ in executing their research. Strengths of their study include: the empirical replication and extension of prior claims; a priori specification of hypotheses and the methodology to test them; and the insight that prior claims may confound gender and generational status. Weaknesses include: the use of an untested (and unknown) scale to measure their dependent variable; the unfortunate and erroneous classification of majors into "hard" science vs. social science categories; a confound between graduate student status and generational status; and the number and interdependence of the statistical tests they use to test their hypotheses. All of the methodological weaknesses I identify (with the exception of their measurement of academic values) could be strengthened through changes to their procedures and analyses. Overall, their insight that Lukianoff and Haidt's attribution of value differences to generational status may be confounded with gender is worthy of further research.

Keywords: Higher Education, Academic Values, Sex Differences, Generational Differences

In their study, Rausch et al. (2023) attempt to empirically test a number of hypotheses put forward by Lukianoff and Haidt in the duo's 2018 book The Coddling of the American Mind. Rausch et al. do so by recruiting 574 undergraduate and graduate students to complete an online survey of their academic values, a conservatism scale, a personality inventory, and questions about their demographic status (e.g., sex, major, student status). They test their hypotheses using 10 t-test comparisons and 5 multiple regression models (one for each of the five academic values they sought to predict), each of which used 8 independent variables. The total number of statistical contrasts conducted in their study therefore included 50 statistical contrasts (i.e., 10 t-test + 40 (5 x 8) multiple regression contrasts). Based on whether results were statistically significant (rather than clinically meaningful), they conclude that, although some of Lukianoff and Haidt's claims are justified (e.g., that contemporary undergraduates value student well-being more than previous generations), others are not. They also point out that the increase in females on college campuses may confound differences between the current and previous generations of college students.

As suggested by my description of Rausch et al.'s study, I believe my major contribution to a review of their work lies primarily in a critical consideration of their methodological decisions. I am a school psychologist, whose work focuses primarily on the psychoeducational assessment of preK-12 students. Although I teach undergraduates, I do not claim expertise in understanding their values or what motivates them. Instead, I will frame my remarks in terms of what I see as Rausch et al.'s methodological strengths and weaknesses, along with suggestions on how to strengthen their methods to better test their hypotheses.

Study Strengths

The greatest strength of Rausch et al.'s study is the fact that they collected independent data to test claims advanced by Lukianoff and Haidt (2008). Independent tests of claims advanced in the sciences are far too rare in all fields of scientific endeavor (National Academies of Sciences, Engineering, & Medicine, 2019), and particularly in social sciences such as my own discipline of psychology (Open Science Foundation, 2015). I appreciate and respect that Rausch et al. invested the time, energy, and resources to collect, analyze, and discuss their data in the context of directly testing claims from other scholars.

The second major strength of this study is that all of the hypotheses Rausch et al. advance are clearly stated—and, given their unflinching candor in noting where they were and were not supported, clearly done in advance of knowing the results. Far too often, scientists who find the data disagree with their anticipated findings alter their proposed hypotheses to match the outcomes they obtain (i.e., HARKing; Kerr, 1998), the methods used to test their hypotheses (i.e., p-hacking; Head, Holman, Lanfear, Kahn, & Jennions, 2015), or their data/outcomes (i.e., fraud). Although they acknowledge they changed the statistical methods they initially planned to employ, it is clear from their candid acknowledgment of when the data did and did not support their hypotheses that they did not succumb to the temptation to torture their data until it yielded their desired results. In other words, they clearly model one of the most important academic values to scholarly pursuit: Scientific disinterestedness (see Merton, 1942; Macfarlane & Cheng, 2008).

The third major strength of this study is the contribution it could make (but, as I will argue later, does not yet achieve due to data analysis decisions) to understanding the distinct influences of gender and generation on students' academic values. The fact that they find these outcomes are

statistically significant despite some of the limitations in their analyses and measures suggests that gender may, indeed, influence students' academic values. As I will note later, I believe they could employ methods to make an even stronger test of this claim—but the fact that they find support for it in the analyses they conducted is an intriguing contribution to understanding the characteristics that influence the values of contemporary undergraduates. Given the attention claims about undergraduate values receive in the media, such a contribution could help advance social discourse and public understanding.

Study Weaknesses

I identified four weakness in Rausch et al.'s study. Three of them could be addressed by changes to their data analyses, and one (their measurement of students' academic values) could be partially addressed with changes to the analyses and discussion. I describe the four weaknesses in the following sections.

The Academic Budget Allocation Model. The greatest weakness of Rausch et al.'s study is their decision to measure academic values using a measure apparently lacking any evidence of reliability or validity. In fact, the citation for their measure (which they call the "Academic Budget Allocation Model") is from an oral presentation delivered at the "Heterodox Psychology Workshop: A Training Program for Graduate Students and Young Scholars." Rausch et al. give no other reference to the instrument, and my search of Social Science Citations Index found no subsequent references to the paper (and by extension, the instrument) in the social science literature. Although Rausch et al. acknowledge that the instrument needs more research (p. 18), I am surprised that they offered no data regarding the integrity of the instrument. Instead, they take the instrument at face value without supporting evidence—which is something that even undergraduates in the "soft" science of psychology are taught not to do. I also wonder whether the original study reporting the measure was ever subjected to peer review; when I searched for the term "Heterodox Psychology Workshop," I found a reference to the second annual conference (apparently conducted in 2020) on Chapman University's website (Chapman University, April 15, 2019). I followed the link to the conference, and saw that the second author of both the Rausch et al. and Planke et al. studies (Glenn Geher) was listed as one of the conference organizers. I do not know if the conference employed peer review for conference submissions, but if it did, that process was not mentioned on the site.

I also note that the "budgeting" model employed by the instrument creates forced-choice responses, which means the five dependent variables (which are treated as if they were independent of one another in the analyses) are, by definition, statistically interdependent. That is, once a respondent assigns a number to one value, the numbers available to be assigned to other values are constrained, thus making them statistically interdependent. Although Rausch et al. view the forced-choice or "budgeting" aspect of their measure as a feature, not a bug, they appear to be unaware that forced choice selection causes their five dependent variables to be statistically interdependent. Therefore, their treatment of students' academic values as if they constitute five distinct, independent variables is statistically unwarranted, and undermines both their analyses and the conclusions drawn from them.

I also note that the "budgeting" model employed by the instrument creates forced-choice responses, which means the five dependent variables (which are treated as if they were independent of one another in the analyses) are, by definition, statistically interdependent. That

is, once a respondent assigns a number to one value, the numbers available to be assigned to other values are constrained, thus making them statistically interdependent. Although Rausch et al. view the forced-choice or "budgeting" aspect of their measure as a feature, not a bug, they appear to be unaware that forced choice selection causes their five dependent variables to be statistically interdependent. Therefore, their treatment of students' academic values as if they constitute five distinct, independent variables is statistically unwarranted, and undermines both their analyses and the conclusions drawn from them.

Assignment of majors into the "hard" and social sciences. It is difficult to put aside the tired trope of classifying STEM disciplines as "hard" sciences, not least because two of the major categories—Technology and Engineering—are not sciences. However, I will look past the implication that other, nonSTEM disciplines are "soft," and instead address a problem of classification: Including humanities disciplines (they specifically mention History) in the social sciences. The humanities (including History) are disciplines that embrace the narrative and criticism as their primary tools of inquiry. In contrast, social sciences share with all sciences their primary tools of inquiry: observation and experimentation. Rausch et al.'s classification of History as a "social science" suggests they may have also subsumed other humanities majors (e.g., philosophy, English, languages, literature, rhetoric, religious studies) in their "social sciences" majors. Therefore, their contrasts built on academic major (i.e., the dichotomous categorization of majors into "hard" sciences and social sciences) is clearly mistaken.

Confounding generational and graduate status. Rausch et al. divide their sample into Gen Z (n = 448) and Millennials Plus (n = 126) categories. Subsequent contrasts presume that the only meaningful influence on academic values represented in the two subsamples is generational. However, 353, or 77%, of the Gen Z sample are undergraduates; in contrast, just 23, or 18%, of the Millennials Plus sample are undergraduates. Thus, the samples confound generational difference with graduate student status. It is likely that graduate students embrace different academic values than undergraduates independent of generational status—in part because of the academic values that led them to choose graduate education, and in part because of the effects of graduate education on their values. For example, the finding of a "generational" difference with respect to the value of academic rigor could be due to self selection (i.e., those who value academic rigor are more likely to pursue graduate education), or it could be an outcome of their graduate education (which should teach the value of academic rigor for scholarly inquiry). Regardless, it is implausible to assume that the two samples represent (only) a difference in generations. In much the same way as Rausch et al. argue Lukianoff and Haidt (2018) confound generational status and gender, I would argue that Rausch et al. confound generational status with graduate student status.

Experiment-wise error rate and false positive findings. Rausch et al. use the conventional $p < .05$ cutoff to control for false positive findings (i.e., Type I error) for each of the contrasts they test. They apparently assume that each of their statistical tests functions independently, rather than recognizing that there is a clear relationship between the number of contrasts one conducts and the likelihood of false positive findings (i.e., $1 - (1 - \alpha)^n$, where $\alpha = .05$ and n = the number of contrasts). Putting aside their questionable characterization of some contrasts as "approaching significance" (dichotomous decision rules do not allow for near misses), the probability of at least one false positive occurring in their results is 92% (i.e., $1 - (1 - .05)^{50}$). In other words, the odds are better than 9 to 1 that at least one of the statistically significant contrasts to which they

ascribe meaning is simply the result of running a large number of contrasts on the same sample. Which contrast(s) might be erroneous is not possible to know, but they fail to recognize the overwhelming odds that at least one of the results they identify as statistically significant is, in fact, a false positive. I would also point out that, like most social scientists, they emphasize the statistical significance of their findings without regard to the practical significance (usually defined as the proportion of total variance explained by one or more model parameters). For example, the entire model (i.e., all independent variables in the analysis) predicted only 1.5% of the variance in the academic freedom value, meaning the single variable of gender (which they identify as statistically significant) would predict even less. Most social scientists hold that individual model parameters should account for 5% or more of the total variance to be considered non-trivial. The proportion of variance accounted for by individual model parameters is not given, which does not allow for readers to determine the degree to which a contrast that meets the criterion for statistical significance actually accounts for a meaningful proportion of variance.

Suggested Responses/Improvements

Although I see a number of weaknesses in the version of the MS I reviewed, I believe Rausch et al. could address most of these weaknesses by providing additional data or making different analytic choices. I offer my suggestions in the spirit of helping to advance their inquiry into this interesting area.

The Academic Budget Allocation Model. My concerns regarding Rausch et al.'s approach to measuring academic values, which is the backbone of their study, could all be addressed by providing additional data and justification for the measure they employ. I note that reliability and validity data are available for the other two instruments they used in their study; similar data should be included for their Academic Budget Allocation Model. Although I recognize Rausch et al. are not test publishers (and do not have the resources available to test publishing companies), I would highly recommend they consult chapters 1 and 2 in the current edition of standards for educational and psychological tests (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education, 2014) to get a sense of the type of information needed to support their use of the Academic Budget Allocation Model. I would also encourage them to justify why a forced-choice decision model is better than one in which respondents could rank each academic value independently, especially given they treat each academic value as if it was independent of others in their academic analyses. To put it another way, if a student strongly values academic rigor and academic freedom, but also thinks other values are important, the model used in this study forces the student to put less value on some items to assign importance to others. It may be that this approach is justified due to social desirability or other factors (e.g., if each value is rated independently from 1-10, students could respond with all 1s or all 10s); but if it is, supporting data and clear elaboration for how statistical interdependence among value rankings are subsequently analyzed should be included in their study.

Assignment of majors into the "hard" and social sciences. The fixes here are straightforward and easy: Avoid the pejorative labeling of sciences into "hard" and social, and specify all of the majors included in each condition. As a former dean of humanities and social sciences at a STEM-intensive university, I might suggest labeling the two groups "STEM" and "non-STEM" or "Other" majors. (I confess to having flirted with other distinctions, such as the "natural" and "unnatural" sciences, or the "social" and "asocial" sciences, but such terms exacerbate the bias I wish to

avoid.) I would also suggest deciding whether there are a sufficient number of humanities majors to constitute a third category for analysis, or whether humanities majors (such as History) should simply be excluded from the analyses altogether (as was done for art, education, and business majors).

Confounding generational and graduate status. The response to this concern is straightforward: Add a variable to the regression model for graduate status. Doing so may reduce some of the variance captured by generational status, but failing to do so ignores the substantial confound in undergraduate representation between the generational samples (77% vs. 18%). Lukianoff and Haidt's primary argument was that contemporary undergraduates differ from previous generations of undergraduates; failing to tease apart the unique influences of generational difference from graduate student status undermines the discussion of generational influence at the core of their study.

Experiment-wise error rate and false positive findings. Again, this is a weakness where a different statistical approach would address the concern. The most conservative approach would be to treat all 50 statistical contrasts as correlated, and then choose a Type I error rate for each contrast that keeps the overall likelihood of a false positive for the entire study at 0.05. Using $\alpha = 1 - (1 - .05)^{1/n}$, where the experiment-wise error rate is .05 and the number of contrasts is 50, the α for each contrast would be 0.0102. Alternatively, analyses could be clustered into "families" (i.e., groupings where it is reasonable to assume statistical independence between families of contrasts, such as regression analyses being one family and contrasts of means constituting a second family) with appropriate justification. I would also strongly encourage the inclusion of the proportion of variance accounted for by each model parameter, and the effect size for each contrast between means, to help readers judge the degree to which independent variables relate to variation in academic values.

Closing Comments

I hope that these comments are helpful to Rausch et al. in considering ways to enhance their arguments, and to readers for understanding the strengths and limitations of the reported findings. As a consumer, rather than producer, of research on undergraduates' values, I enjoyed reading this study and hope I will have the opportunity to review a revised manuscript—and see Rausch et al.'s responses to my criticisms.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for educational and psychological testing (3rd ed.)*. Washington, DC: American Educational Research Association.

Chapman University (April 25, 2019). 2nd Biennial Heterodoxy in Psychology Conference. *Research and Creative Activity at Chapman.* https://blogs.chapman.edu/research/2019/04/25/%CF%88-2ndbiennial-heterodoxy-in-psychology-conference/.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology, 13*, e1002106. http://dx.doi.org/10.1371/journal.pbio.1002106.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217. https://doi-org.prox.lib.ncsu.edu/10.1207/s15327957pspr0203_4.

Lukianoff, G. & Haidt, J. (2018). *The coddling of the American mind*. New York: Penguin Press.

Macfarlane, B. & Cheng, M. (2008). Communism, universalism and disinterestedness: Re-examining contemporary support among academics for Merton's scientific norms. *Journal of Academic Ethics, 6*, 67–78. https://doi.org/10.1007/s10805-008-9055-y.

Merton, R. K. (1942). The normative structure of science. In N. Storer (Ed.) *The sociology of science: Theoretical and empirical investigations* (pp. 267–278). Chicago: The University of Chicago Press.

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25303.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716. https://doi.org/10.1126/science.aac4716.