TECHNICAL RESPONSE

WOMEN IN SCIENCE

Response to Comment on "Expectations of brilliance underlie gender distributions across academic disciplines"

Andrei Cimpian[1]* and Sarah-Jane Leslie[2]*

Ginther and Kahn claim that academics' beliefs about the importance of brilliance do not predict gender gaps in Ph.D. attainment beyond mathematics and verbal test scores. However, Ginther and Kahn's analyses are problematic, exhibiting more than 100 times the recommended collinearity thresholds. Multiple analyses that avoid this problem suggest that academics' beliefs are in fact uniquely predictive of gender gaps across academia.

In a nationwide study of academics in 30 disciplines, we found that the distribution of gender gaps in Ph.D. attainment is predicted by the extent to which practitioners of a given discipline believe that success requires raw, unteachable aptitude (1). These field-specific ability beliefs combine with cultural stereotypes linking men, but not women, with such inherent brilliance, thereby decreasing women's participation. Ginther and Kahn claim that these field-specific ability beliefs are irrelevant to female representation at the Ph.D. level (2). According to Ginther and Kahn, what instead accounts for the pattern of gender gaps is the extent to which fields emphasize mathematics-intensive content matter. Ginther and Kahn base this conclusion on regression analyses in which Graduate Record Examination (GRE) scores predict female representation but ability beliefs do not.

Ginther and Kahn's preferred regression models (namely, 3 and 5) are problematic because they include three variables that are highly redundant with one another: the quantitative GRE score, the verbal GRE score, and their ratio. A more appropriate modeling strategy would be, for example, to include just one of the GRE scores and their ratio, as other researchers have done (3). Models with multiple variables that are highly correlated encounter a problem known as multicollinearity, which inflates standard errors and leads to unreliable results (4–6). A typical means of quantifying multicollinearity is to calculate a variance inflation factor (VIF), which is the ratio between a predictor's actual variance in a model and its variance, assuming that it was uncorrelated with the other predictors in the model. The statistics literature typically places an upper bound of 10 on the recommended VIFs (5, 6); caution is

recommended in interpreting models with VIFs above this threshold. By comparison, Ginther and Kahn's models 3 and 5 have average VIFs of more than 1100—more than 110 times the conventional threshold. Moreover, the VIFs for the variables of most interest to Ginther and Kahn (namely, the quantitative GRE scores and the ratio of quantitative to verbal scores) have VIFs ranging from 3258 to 5046. As a sign of the problems caused by this extreme form of collinearity, note the unusual values for the standardized regression coefficients of these variables in Ginther and Kahn's models 3 and 5. For instance, the quantitative:verbal GRE ratio has a standardized coefficient of –17.13 in their model 5. Taken at face value, this coefficient suggests that an increase of 1 standard deviation in the quantitative:verbal ratio would be accompanied by a decrease of 17.13 standard deviations in a field's percentage of female Ph.D.'s (which, to provide an intuitive metric, is equivalent to a decrease that is more than 17 times the difference in female Ph.D.'s between history and mathematics). Such implausible estimates are not uncommon in highly collinear models (5, 6).

Contrary to Ginther and Kahn's claims, we found evidence that academics' ability beliefs predict female representation in a wide range of models that include disaggregated GRE scores, as well as measures of a field's relative emphasis on mathematics versus verbal ability, and that also have acceptable VIFs. We include 22 such models in Table 1. The top half of the table displays unweighted models, and the bottom half displays weighted models (in keeping with Ginther and Kahn's models 4 to 6). As analytic weights, we used the inverse of the variance of the field-specific ability beliefs, which was calculated from the raw, individual-level data. These inverse-variance analytic weights allow fields whose ability beliefs are estimated more precisely (that is, with lower variance) to carry more weight in the regression models (7). In contrast,

Ginther and Kahn used as their analytic weights the number of respondents within each field, which here does not track the precision with which ability beliefs were measured at the field level: There was no correlation between the number of respondents within a field and the variance of the ability belief estimate for that field, $r(28) = -0.01$, $P = 0.929$. There is thus no reason to penalize (i.e., downweight) fields that are small but whose ability beliefs were nevertheless measured with as much precision as those of larger fields. Within each half of Table 1, the models are sorted in increasing order of average VIF; thus, as one moves down the table, the models become more collinear. All of the models in Table 1 include as predictors the ability beliefs of each field's practitioners and the three competing variables tested in (1) (namely, on-campus hours worked, selectivity, and systemizing versus empathizing; these coefficients are not displayed in the table). In addition, the models include various permutations of the following GRE-based variables: the quantitative score, the verbal score, the analytical writing score, the quantitative:verbal ratio score, and the quantitative–verbal difference score. Adjusting for quantitative GRE scores provides a conservative test of our hypothesis, as academics and nonacademics alike believe that success in mathematics depends largely on raw ability (1, 8). Thus, young men and women's quantitative GRE scores may already reflect the influence of mathematics-specific ability beliefs, so adjusting for these scores in our analyses may underestimate the true impact of ability beliefs on gender gaps in representation.

The results displayed in Table 1 make it clear that academics' ability beliefs are a significant predictor of female representation above and beyond whether a discipline (i) requires mathematical ability (as indicated by the quantitative GRE score) and (ii) privileges this ability relative to verbal ability (as indicated by the quantitative: verbal ratio or the quantitative–verbal difference) (see, e.g., models 7, 8, 9, 20, and 21). More generally, academics' beliefs are a statistically reliable predictor of gender gaps in all 22 models with acceptable average VIFs. Additional models in which we weighted the observations by the inverse of the standard error of the field-specific ability beliefs, rather than the inverse of their variance, revealed the same results [$Ps \leq 0.063$ for ability beliefs in parallel versions of models 13 to 23 (not shown)]. Academics' ability beliefs were also a significant predictor of gender gaps when we used the National Science Foundation's definition of science, technology, engineering, and mathematics (STEM) disciplines, as supplied by Ginther and Kahn [$Ps \leq 0.029$ for parallel versions of models 1 to 11 and 13 to 23 (not shown)]. It is only in models whose average VIFs are greater than 1000 (models 12 and 24 in Table 1, which parallel Ginther and Kahn's models 3 and 5) that ability beliefs no longer predict female representation.

In light of these analyses, the claims we made in Leslie, Cimpian *et al.* (1) remain valid as originally stated: Fields whose practitioners idolize brilliance and genius have fewer women.

[1]Department of Psychology, University of Illinois, Champaign, IL 61821, USA. [2]Department of Philosophy, Princeton University, Princeton, NJ 08544, USA.
*Corresponding author. E-mail: acimpian@illinois.edu (A.C.); sjleslie@princeton.edu (S.-J.L.)

Table 1. Multiple regression models predicting women's Ph.D. representation. The table displays standardized coefficients (and *P* values) from models predicting female representation at the Ph.D. level (*10*) on the basis of field-specific ability beliefs; the three control variables in (*1*) (on-campus hours worked, selectivity, and systemizing versus empathizing; coefficients not shown); a STEM indicator variable; and various combinations of quantitative, verbal, and analytical writing GRE scores. Dashes indicate that a predictor was not included in the model. All models included the 28 disciplines for which GRE scores were available, as supplied in (*2*). The *P* values were calculated using standard errors that were robust to heteroskedasticity (*9*). We used the inverse of the variance of the field-specific ability beliefs as analytic weights.

Weighting?	Model no.	Average VIF*	Field-specific ability beliefs	Quantitative GRE	Verbal GRE	Analytical writing GRE	Quantitative/verbal GRE ratio	Quantitative/verbal GRE difference
No	1	3.38	−0.53 (*P* = .001)	—	—	—	−0.49 (*P* = .011)	—
No	2	3.38	−0.53 (*P* = .001)	—	—	—	—	−0.49 (*P* = .014)
No	3	3.52	−0.37 (*P* = .015)	−0.63 (*P* = .004)	—	—	—	—
No	4	3.53	−0.38 (*P* = .037)	−0.63 (*P* = .004)	+0.01 (*P* = .935)	—	—	—
No	5	3.95	−0.38 (*P* = .037)	—	−0.33 (*P* = .092)	—	—	−0.76 (*P* = .004)
No	6	3.96	−0.37 (*P* = .043)	—	−0.34 (*P* = .076)	—	−0.77 (*P* = .003)	—
No	7	5.59	−0.39 (*P* = .031)	−0.58 (*P* = .104)	—	—	−0.06 (*P* = .870)	—
No	8	5.67	−0.38 (*P* = .037)	−0.61 (*P* = .092)	[Dropped out]†	—	—	−0.03 (*P* = .935)
No	9	5.67	−0.38 (*P* = .037)	−0.61 (*P* = .092)	—	—	—	−0.03 (*P* = .935)
No	10	9.26	−0.39 (*P* = .030)	−0.48 (*P* = .231)	−0.39 (*P* = .533)	+0.48 (*P* = .545)	—	—
No	11	25.27	−0.39 (*P* = .030)	−1.20 (*P* = .160)	[Dropped out]†	+0.48 (*P* = .545)	—	+0.87 (*P* = .533)
No	12‡	1120.82	−0.17 (*P* = .329)	+11.90 (*P* = .024)	−6.88 (*P* = .018)	—	−15.06 (*P* = .017)	—
Yes	13	3.46	−0.56 (*P* = .001)	—	—	—	−0.47 (*P* = .014)	—
Yes	14	3.47	−0.57 (*P* = .001)	—	—	—	—	−0.46 (*P* = .017)
Yes	15	3.58	−0.40 (*P* = .012)	−0.60 (*P* = .004)	—	—	—	—
Yes	16	3.66	−0.40 (*P* = .036)	−0.60 (*P* = .005)	+0.01 (*P* = .999)	—	—	—
Yes	17	4.13	−0.40 (*P* = .036)	[Dropped out]†	−0.35 (*P* = .084)	—	—	−0.73 (*P* = .005)
Yes	18	4.13	−0.40 (*P* = .036)	—	−0.35 (*P* = .084)	—	—	−0.73 (*P* = .005)
Yes	19	4.15	−0.38 (*P* = .042)	—	−0.36 (*P* = .070)	—	−0.75 (*P* = .004)	—
Yes	20	5.77	−0.41 (*P* = .030)	−0.58 (*P* = .094)	—	—	−0.03 (*P* = .936)	—
Yes	21	5.86	−0.40 (*P* = .036)	−0.60 (*P* = .084)	—	—	—	−0.01 (*P* = .999)
Yes	22	9.71	−0.41 (*P* = .034)	[Dropped out]†	−0.59 (*P* = .234)	+0.36 (*P* = .668)	—	−0.60 (*P* = .214)
Yes	23	10.13	−0.41 (*P* = .034)	−0.50 (*P* = .214)	−0.31 (*P* = .652)	+0.36 (*P* = .668)	—	—
Yes	24§	1205.87	−0.16 (*P* = .403)	+11.55 (*P* = .027)	−7.15 (*P* = .020)	—	−14.81 (*P* = .019)	—

*VIF, Variance inflation factor. The models are sorted in increasing order of the average VIF of their predictors (where higher VIFs indicate higher collinearity in the model).   †The variable dropped out of the model because of collinearity.   ‡Parallel to Ginther and Kahn's model 3. The results are not an exact match because we used more than two-decimal precision for the variables from (*1*) and robust standard errors.   §Parallel to Ginther and Kahn's model 5. The results are not an exact match because we used more than two-decimal precision for the variables from (*1*) and robust standard errors, and weighted the observations by the inverse of the variance of a field's ability beliefs (rather than its sample size).

REFERENCES AND NOTES

1. S. J. Leslie, A. Cimpian, M. Meyer, E. Freeland, *Science* **347**, 262–265 (2015).
2. D. K. Ginther, S. Kahn, *Science* **349**, 391 (2015).
3. C. Riegle-Crumb, B. King, E. Grodsky, C. Muller, *Am. Educ. Res. J.* **49**, 1048–1073 (2012).
4. J. Fox, *Applied Regression Analysis, Linear Models, and Related Methods* (SAGE, Thousand Oaks, CA, 1997).
5. S. Chatterjee, A. S. Hadi, B. Price, *Regression Analysis by Example* (Wiley, New York, ed. 3, 2000).
6. M. H. Kutner, C. J. Nachtsheim, J. Neter, *Applied Linear Regression Models* (McGraw-Hill, New York, ed. 4, 2004).
7. J. M. Wooldridge, *Introductory Econometrics: A Modern Approach* (Cengage, Mason, OH, ed. 5, 2012).
8. M. Meyer, A. Cimpian, S. J. Leslie, *Front. Psychol.* **6**, 235 (2015).
9. H. White, *Econometrica* **48**, 817–830 (1980).
10. National Science Foundation, Survey of Earned Doctorates (2011); www.nsf.gov/statistics/srvydoctorates.