Unskilled and Unaware—But Why?
A Reply to Krueger and Mueller (2002)

Justin Kruger
University of Illinois

David Dunning
Cornell University

J. Kruger and D. Dunning (1999) argued that the unskilled suffer a dual burden: Not only do they perform poorly, but their incompetence robs them of the metacognitive ability to realize it. J. Krueger and R. A. Mueller (2002) replicated these basic findings but interpreted them differently. They concluded that a combination of the better-than-average (BTA) effect and a regression artifact better explains why the unskilled are unaware. The authors of the present article respectfully disagree with this proposal and suggest that any interpretation of J. Krueger and R. A. Mueller’s results is hampered because those authors used unreliable tests and inappropriate measures of relevant mediating variables. Additionally, a regression–BTA account cannot explain the experimental data reported in J. Krueger and D. Dunning or a reanalysis following the procedure suggested by J. Krueger and R. A. Mueller.

In 1999, we published an article (Kruger & Dunning, 1999) suggesting that the skills that enable one to perform well in a domain are often the same skills necessary to be able to recognize good performance in that domain. As a result, when people are unskilled in a domain (as everyone is in one domain or another), they lack the metacognitive skills necessary to realize it. To test this hypothesis, we conducted a series of studies in which we compared perceived and actual skill in a variety of everyday domains. Our predictions were borne out: Across the various studies, poor performers (i.e., those in the bottom quartile of those tested) overestimated their percentile rank by an average of 50 percentile points.

Along the way, we also discovered that top performers, although they estimated their raw test scores relatively accurately, slightly but reliably underestimated their comparative performance, that is, their percentile rank among their peers. Although not central to our hypothesis, we reasoned that top performers might underestimate themselves relative to others because they have an inflated view of the competence of their peers, as predicted by the well-documented false consensus effect (Ross, Greene, & House, 1977) or, as Krueger and Mueller (2002) termed it, a social-projection error.

Krueger and Mueller (2002) replicated some of our original findings, but not others. As in Kruger and Dunning (1999), they found that poor performers vastly overestimate themselves and show deficient metacognitive skills in comparison with their more skilled counterparts. Krueger and Mueller also replicated our finding that top performers underestimate their comparative ranking. They did not find, however, that metacognitive skills or social projection mediate the link between performance and miscalibration. Additionally, they found that correcting for test unreliability reduces or eliminates the apparent asymmetry in calibration between top and bottom performers. They thus concluded that a regression artifact, coupled with a general better-than-average (BTA) effect, is a more parsimonious account of our original findings than our metacognitive one is.

In the present article we outline some of our disagreements with Krueger and Mueller’s (2002) interpretation of our original findings. We suggest that the reason the authors failed to find meditational evidence was because of their use of unreliable tests and inappropriate measures of our proposed mediators. Additionally, we point out that the regression–BTA account is inconsistent with the experimental data we reported in our original article, as well as with the results of a reanalysis of those data using their own analytical procedure.

Does Regression Explain the Results?

The central point of Krueger and Mueller’s (2002) critique is that a regression artifact, coupled with a general BTA effect, can explain the results of Krueger and Dunning (1999). As they noted, all psychometric tests involve error variance, thus “with repeated testing, high and low test scores regress toward the group average, and the magnitude of these regression effects is proportional to the size of the error variance and the extremity of the initial score” (Krueger & Mueller, 2002, p. 184). They go on to point out that “in the Kruger and Dunning (1999) paradigm, unreliable actual percentiles mean that the poorest performers are not as deficient as they seem and that the highest performers are not as able as they seem” (p. 184).

Although we agree that test unreliability can contribute to the apparent miscalibration of top and bottom performers, it cannot fully explain this miscalibration. If it did, then controlling for test

The writing of this reply was supported financially by University of Illinois Board of Trustees Grant 1-2-69853 to Justin Kruger and by National Institute of Mental Health Grant RO1 56072 to David Dunning.

Correspondence concerning this article should be addressed to Justin Kruger, Department of Psychology, 709 Psychology Building, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820, or to David Dunning, Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14853-7601. E-mail: jkruger@s.psych.uiuc.edu or dad6@cornell.edu
reliability, as Krueger and Mueller (2002) do in Figure 2, should cause the asymmetry to disappear. Although this was the case for the difficult test that Krueger and Mueller used, this was inevitable given that the test was extremely unreliable (Spearman–Brown = .17). On their easy test, which had moderate reliability of .56, low-scoring participants still overestimated themselves—by approximately 30 percentile points—even after controlling for test unreliability, just as the metacognitive account predicts. When even more reliable tests are used, the regression account is even less plausible. For instance, in Study 4 of Kruger and Dunning (1999), in which test reliability was quite high (Spearman–Brown = .93), controlling for test unreliability following the procedure outlined by Krueger and Mueller failed to change the overall picture. As Figure 1 of this article shows, even after controlling for test unreliability, low-scoring participants continued to overestimate their percentile score by nearly 40 points (and high scorers still underestimated themselves). In sum, although we agree with Krueger and Mueller that measurement error can contribute to some of the apparent miscalibration among top and bottom scorers, it does not, as Figure 1 of this article and Figure 2 of theirs clearly show, account for all of it.

Do Metacognition and Social Projection Mediate Miscalibration?

Krueger and Mueller (2002) did more than merely suggest an alternative interpretation of our data, they also called into question our interpretation. Specifically, although they found evidence that poor performers show lesser metacognitive skills than top performers, they failed to find that these deficiencies mediate the link between performance and miscalibration.

The fact that these authors failed to find mediational evidence is hardly surprising, however, in light of the fact that the tests they used to measure performance, as the authors themselves recognized, were either moderately unreliable or extremely so. It is difficult for a mediator to be significantly correlated with a crucial variable, such as performance, when that variable is not measured reliably.

In addition, even if the tests were reliable, we would be surprised if the authors had found evidence of mediation because their measures of metacognitive skills did not adequately capture what that skill is. Metacognitive skill, traditionally defined, is the ability to anticipate or recognize accuracy and error (Metcalfe & Shimamura, 1994). Krueger and Mueller (2002) operationalized this variable by correlating, across items, participants’ confidence in their answers and the accuracy of those answers. The higher the correlation, the better they deemed the individual’s metacognitive skills.

There are several problem with this measure, however. Principal among them is the fact that a high correlation between judgment and reality does not necessarily imply high accuracy, nor does a low correlation imply the opposite. To see why, consider an example inspired by Campbell and Kenny (1999) of two weather forecasters, Rob and Laura. As Table 1 shows, although Rob’s predictions are perfectly correlated with the actual temperatures, Laura’s are more accurate: Whereas Rob’s predictions are off by an average of 48 degrees, Laura’s are off by a mere 7.

How can this be? Correlational measures leave out two important components of accuracy. The first is getting the overall level of the outcome right, and this is something on which Rob is impaired. The second is ensuring that the variance of the predictions is in harmony with the variance of the outcome, depending on how strongly they are correlated (Campbell & Kenny, 1999). Correlational measures miss both these components. However, deviational measures, that is, ones that simply assess on average how much predictions differ from reality, do take these two components into account. We suspect that this fact, coupled with the problem of test unreliability, is the reason the deviational measures of metacognition we used in our studies mediated the link between performance and miscalibration, whereas the correlational measure used by Krueger and Mueller (2002) did not.¹

Note that this point applies equally well to Krueger and Mueller’s (2002) social-projection measure (how well others are doing) as it does to their metacognition measure (how well oneself is

---

¹ Krueger and Mueller’s (2002) operationalization of metacognitive accuracy is problematic on other grounds. As researchers in metacognition have discovered, different correlational measures of skill (e.g., Pearson’s $r$, gamma) often produce very different results when applied to the exact same data (for an excellent discussion, see Schwartz & Metcalfe, 1994).

---

Table 1

<table>
<thead>
<tr>
<th>Day</th>
<th>Actual temperature$^a$</th>
<th>Rob</th>
<th>Laura</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>70</td>
<td>20</td>
<td>65</td>
</tr>
<tr>
<td>Tuesday</td>
<td>80</td>
<td>35</td>
<td>75</td>
</tr>
<tr>
<td>Wednesday</td>
<td>60</td>
<td>5</td>
<td>70</td>
</tr>
<tr>
<td>Thursday</td>
<td>70</td>
<td>20</td>
<td>75</td>
</tr>
<tr>
<td>Friday</td>
<td>90</td>
<td>50</td>
<td>80</td>
</tr>
<tr>
<td>$r$</td>
<td>1.00</td>
<td>.53</td>
<td></td>
</tr>
<tr>
<td>Average deviation</td>
<td>48</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ In degrees Fahrenheit.
percentile points were then within 1 point of their actual test score. Counter to our hypothesis, Krueger and Mueller did not find any evidence for a social-projection problem among high performers. However, their measure of social projection is irrelevant to our original assertion. Of key importance is whether high performers overestimated how well they thought their peers had performed overall. Krueger and Mueller’s measure, instead, focuses on the correlation between participants’ confidence in their own responses across items and their confidence in their peers’ responses. Note that an individual might have a very inflated view of the performances of their peers (or a very deflated one), but that this within-subject correlational measure would capture none of this misestimation, instead measuring only how individual item self-confidence covaries with individual item—other confidence.

What Does Experimental Evidence Suggest?

To be fair, we believe that there is more to Krueger and Mueller’s (2002) regression effect argument than simple measurement error. After all, as we noted in our original article (Kruger and Dunning, 1999, p. 1124) and elsewhere (see Kruger, Savitsky, & Gilovich, 1999), whenever two variables are imperfectly correlated, extreme values on one variable are likely to be matched by less extreme values of the other. In the present context, this means that extremely poor performers are likely to overestimate that performance to some degree.

For this reason, we collected additional data to directly test our interpretation. The crucial test of the metacognitive account does not come from demonstrating that regression effects cannot explain our data. Rather, the crucial test comes from experimentally manipulating metacognitive skills and social projection to see whether this results in improved calibration. This was the approach we took in our original studies, and we believe the data we obtained provide the most conclusive support for our own interpretation and against Krueger and Mueller’s (2002) regression-BTA interpretation. If our results were due merely to a regression artifact, then we should have observed the same regressive correlations regardless of whatever experimental manipulation we used. However, we found in Studies 3b and 4 of our original article that we could make the regression effect evaporate under experimental conditions as exactly predicted by our theoretical analysis.

In Study 4, for instance, we gave 140 participants a test of logical reasoning and compared actual performance on the test with perceived performance. Next, we asked participants to grade their own test (i.e., to indicate which problems they thought they had answered correctly and which they had answered incorrectly) and to estimate their overall performance once more. Half of the participants, however, did something else. Just prior to grading their test, they completed a crash course on logical reasoning adopted from Cheng, Holyoak, Nisbett, and Oliver (1986). What we found was that participants who had received training—but only participants who had received training—became substantially more calibrated with respect to their test performance. Incompetent participants who had, just prior to training, overestimated their test score by 5 points (out of 10) and their percentile score by 36 percentile points were then within 1 point of their actual test score and within 17 points of their percentile score. Mediational analyses revealed that this new-found accuracy was a direct result of the increased metacognitive skill. In passing, Krueger and Mueller (2002) took issue with the overall strategy we pursued in this study, but provided no alternative account of the results of our experimental manipulation.

We took a similar approach in demonstrating the role of false consensus in the slight underestimation we observed with extremely skilled participants. If top performers underestimate themselves because of an inflated view of the comparison group, then they (and only they) should become more calibrated if they are given an accurate picture of the true skills of their peers. In Study 3, that is exactly what we observed: After they were provided with a representative sample of tests that had been completed by their peers, top-scoring (but not bottom-scoring) participants arrived at more accurate self-assessments. This, too, cannot be explained by statistical regression artifact.

Krueger and Mueller (2002) took issue with this finding as well, pointing out that most people increased their percentile estimates after seeing the performances of their peers, and that “the increase [among high performers] . . . was not significantly larger than the increase among the poor performers” (p. 185). Actually, it was, but this is not the issue: The hypothesis derived from the false consensus account of the underestimation of top performers is not that top-scoring participants will increase their self-estimates more than will poor performers, but that top-scoring participants will improve their self-estimates more than will poor performers, who will show no overall improvement (that is, they will not lower their self-estimates).2 This, too, is precisely what we observed.

Final Thoughts

We end on two final thoughts. First, we cannot help but notice the obvious irony presented by this exchange. Krueger and Mueller (2002) dismissed our original account of our data, and we have spent a hefty portion of this reply dismissing theirs. The discerning reader may have noticed that both camps seem to be rather confident in their conclusions, although, given the contradictions, someone must be wrong. Whoever is wrong, they do not seem to know it.

Second, although we strongly believe, for the reasons outlined in this reply, that regression alone cannot explain why the unskilled are unaware, we do not believe Krueger and Mueller’s (2002) alternative interpretation should be dismissed lightly. Regression effects are notoriously difficult to spot but easy to misinterpret—by laypeople and researchers alike (Kahneman & Tversky, 1973; Lawson, 2001; Nisbett & Ross, 1980). Although regression effects cannot explain our original data, the simple fact remains that more work needs to be done. No single study, or even set of studies, can be taken as the final word on the issue, and it remains to be seen which account—ours, theirs, or one yet to come—best explains why the unskilled are unaware.

---

2 The interaction term from the 2 (quartile: top vs. bottom) × 2 (estimate: Time 1 vs. Time 2) analysis on participants’ perceptions of their percentile ability was $F(1, 34) = 4.54, p = .04$, although this was not reported in our original article because it did not pertain to our hypothesis.
References


Received August 13, 2001
Accepted August 15, 2001

Members of Underrepresented Groups:
Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to Demarie Jackson at the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In your letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Demarie Jackson, Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.