Supplementary Online Material

Additional Analyses

In the literature, the primary argument for using difference scores is that they are intuitive, straightforward, and face-valid (e.g., see Griffin, Murray, & Gonzalez, 1999, for an excellent discussion on why researchers tend to use difference scores). However, as we note in the main text, the practice of using difference scores has been criticized widely by methodologists and psychometricians (e.g., Cohen, Cohen, West, & Aiken, 1983; Cronbach & Furby, 1970; Edwards, 1994a; Griffin et al., 1999; Johns, 1981; Zimmerman, 1997). There seems to be at least three problems with difference scores:

1. The first problem is *the interpretational ambiguity problem*. When a researcher correlates a difference score with an outcome measure, the meaning of that correlation can sometimes be hard to interpret. In fact, it is impossible to know what that correlation means without looking at the relations between the components of the difference score and the outcome measure. This problem is commonly known as the *confounding of difference scores with their constituents* (Edwards, 1994b; Johns, 1981). Moreover, difference scores are hard to interpret because the components that make up the difference score (e.g., self-perceived performance and actual performance) are often correlated, and therefore, non-orthogonal (Griffin et al., 1999). These circumstances can only be diagnosed by looking at the individual correlations that together make up the difference score, and is one of the primary arguments for using the joint testing approach as advocated by Edwards (1995).
2. The second problem with difference scores is that they are *vulnerable to ceiling effects* (Cohen et al., 1983; Cronbach & Furby, 1970; Griffin et al., 1999; John & Robins, 1994; Johns, 1981; Lord, 1956; McNemar, 1958; Schultheiss et al., 2005; Tucker, Damarin, & Messick, 1966; see Cheng, Kornienko, & Granger, 2018). For example, when data are skewed and there is a ceiling effect, difference scores may not reflect the true amount of change (http://methods.sagepub.com/reference/encyclopedia-of-measurement-and-statistics/n135.xml). Furthermore, as Ehrlinger, Mitchum, and Dweck (2016, p. 96) have noted, "percentile estimates have a ceiling of 100. Thus, an individual scoring in the 95th percentile can be, at most, be overconfident by 5 percentile points while lower-scoring individuals are able to display much more overconfidence."
3. The third problem with difference scores is that they sometimes exhibit low reliability (Lord, 1958; Rogosa & Willett, 1983). For example, when component measures are positively correlated, the reliability of the difference score is usually less than the reliability of either component measure (Edwards, 2001).

To examine whether these problems apply to our studies, we conducted additional analyses in Study 1, which is the study that yielded the most inconsistent support for our hypothesis. We did three things here. First, we computed basic descriptive statistics to examine how well the participants in this study performed in the Flashcard Game. Here, we found that the mean for actual performance was 15.39 ($SD = 4.36$), which means that, on average, individuals correctly answered approximately 75% of the overall trials. Second, we examined the distribution of performance and found that over 63% of our participants ($n = 96,013$) were high-performing, scoring at least 15 out

of 20 (75% correct), and that over 1/5 of our sample ($n = 31,495$) had a perfect score on the exercise (i.e., they scored 20 out 20).

These numbers clearly indicate that in terms of performance, most participants in Study 1 were on the higher end of the scale. This leads us to suspect that the difference score approach might not be the most appropriate to use for this particular study. Specifically, because a large number of participants in this study did so well on the Flashcard Game, using the difference score approach in this study could potentially fail to capture the true amount of self-enhancement for a large majority of participants. For comparison, we ran these same analyses for Studies 3 and 4 (which are the studies that yielded the clearest support for our hypothesis). In these studies, the overall average performance on our 15-item test wasn't nearly so high ($M_{Study3} = 8.14$, $SD_{Study3} = 1.89$; $M_{Study4} = 7.18$, $SD_{Study4} = 1.93$). Only 1 participant (from Study 3) got a perfect score (none in Study 4), and in both studies, very few scored at least 11 out of 15 (Study 3: 10.20%; Study 4: 3.23%).

Second, we inspected the correlation between actual performance and self-perceived performance across our studies. In Studies 1 ($r = .22$, $p < .001$) and 2 ($r = .37$, $p < .001$), these two components were correlated, indicating that, in general, those who thought that they did better than others *did in fact perform better than others.* In Studies 3 ($r = .05$, $p = .11$) and 4 ($r = .08$, $p = .20$), there was no such correlation. Therefore, it is plausible that the difference scores in Studies 1 and 2 are somewhat problematic, in that those indices confound both performance and perception, unlike in Studies 3 and 4, in which self-perceived performance and actual rank seem orthogonal.

Finally, we note that in Studies 1 and 2, social class was positively associated with actual performance. This means that if we use the difference score approach, higher-class participants will be more limited in displaying overconfidence, while lower-class participants would be able to display much more overconfidence (e.g., Ehrlinger et al., 2016). This might give the false impression that higher-class individuals are *less* overconfident compared to their lower-class counterparts.

Together, these explanations might give clues as to why the difference score approach produced results that are inconsistent with the residual score approach, and potentially suggests that the (negative) correlation between objective social class and overplacement in Study 1 could be a statistical artifact (see Griffin et al., 1999; for another empirical example in in which the residual score approach and the difference score approach produce divergent conclusions). Therefore, we strongly urge readers to treat the results generated by the difference score approach with caution.

Supplemental References

Cheng, J. T., Kornienko, O., & Granger, D. A. (2018). Prestige in a large-scale social group predicts longitudinal changes in testosterone. *Journal of Personality and Social Psychology, 114*(6), 924.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change" ¬ or should we? *Psychological Bulletin, 74*, 68¬80.

Edwards, J. R. (1994a). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58*, 51-100.

Edwards, J. R. (1994b). Regression analysis as an alternative to difference scores. *Journal of Management, 20*(3), 683-689.

Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes, 64,* 307-324.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods, 4*(3), 265-287.

Ehrlinger, J., Mitchum, A. L., & Dweck, C. S. (2016). Understanding overconfidence: Theories of intelligence, preferential attention, and distorted self-assessment. *Journal of Experimental Social Psychology, 63*, 94-100.

Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships, 6*(4), 505-518.

John, O. P. & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206–219.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance, 27*(3), 443-463.

Lord, F. M. (1956). The measurement of growth. *ETS Research Bulletin Series, 1*, 1-22.

Lord, F. M. (1958). The utilization of unreliable difference scores. *Journal of Educational Psychology, 49*(3), 150.

McNemar, Q. (1958). On growth measurement. *Educational and Psychological Measurement*, *18*(1), 47-55.

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability the difference score in the measurement of change. *Journal of Educational Measurement, 20*(4), 335-343.

Schultheiss, O. C., Wirth, M. M., Torges, C. M., Pang, J. S., Villacorta, M. A., & Welsh, K. M. (2005). Effects of implicit power motivation on men's and women's implicit learning and testosterone changes after social victory or defeat. *Journal of Personality and Social Psychology*, *88*(1), 174-188.

Tucker, L. R., Damarin, F., & Messick, S. (1966). A base-free measure of change. *Psychometrika*, *31*(4), 457–473.

Supplementary Online Material

**General Mental Ability Test in Study 2**

Today, you'll take a test that is designed to measure your general cognitive and mental abilities. This test has twenty questions. You'll have a maximum of 20 seconds to work on each question. The page will automatically advance to the next question after 20 seconds; however, you may continue to the next question before the time limit expires. Good luck! [1]

Which of the following is the earliest date?
○ Jan. 16, 1898
○ Feb. 21, 1889
○ Feb. 2, 1898
○ Jan. 7, 1898
○ Jan. 30, 1889*

LOW is to HIGH as EASY is to

○ SUCCESSFUL
○ PURE
○ TALL
○ INTERESTING
○ DIFFICULT*
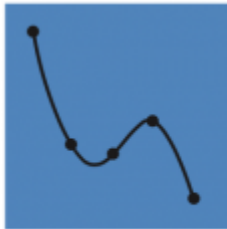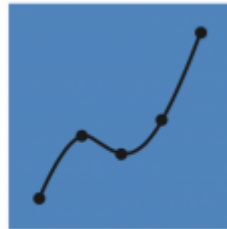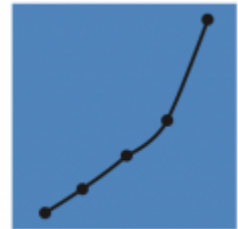

A featured product from an Internet retailer generated 27, 99, 80, 115 and 213 orders over a 5-hour period. Which graph above best represents this trend?
○ A
○ B
○ C
○ D*
○ E

---

[1] Responses designated with an * are the correct answer

What is the next number in this series: 29   41   53   65   77   ?
O 75
O 88
O 89*
O 98
O 99

One word below appears in red. What is the OPPOSITE of that word?
She gave a complex answer to the question and we all agreed with her.
O Long
O Better
O Simple*
O Wrong
O Kind

Jose's monthly parking fee for April was $150; for May it was $10 more than April; and for June $40 more than May. What was his average monthly parking fee for these 3 months?
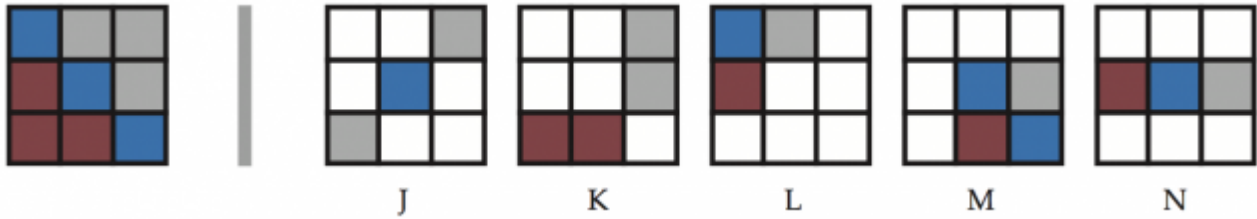O $66
O $160
O $166
O $170*
O $200

1. Sandra is responsible for ordering all office supplies.
2. Notebooks are office supplies.
3. Sandra is responsible for ordering notebooks.

If the first two statements are true, is the final statement true?
O Yes*
O No
O Uncertain

Which THREE choices are needed to create the figure on the left? Only pieces of the same color may overlap.

- ❑ J
- ❑ K*
- ❑ L*
- ❑ M*
- ❑ N

Which THREE of the following words have similar meanings?

- ❑ Observable
- ❑ Manifest
- ❑ Hypothetical*
- ❑ Speculative*
- ❑ Theoretical*

Last year, 12 out of 600 employees at a service organization were rewarded for their excellence in customer service, which was _____ of the employees?

- ❍ 1%
- ❍ 2%*
- ❍ 3%
- ❍ 4%
- ❍ 6%

The boy plays baseball.
All baseball players wear hats.
The boy wears a hat.

Assume the first two statements are true. Is the final one:

- ❍ True*
- ❍ False
- ❍ Not certain

Paper sells for 21 cents per pad. What will four pads cost?

○ 80 cents
○ 82 cents
○ 84 cents*
○ 88 cents

How many of the five pairs of items below are exact duplicates?

Nieman K.M.   Neiman, K.M.
Thomas, G.K.   Thomas, C.K.
Hoff, J.P.   Hoff, J.P.
Pino, L.R.   Pina, L.R.
Warner, T.S.   Wanner, T.S.

○ 0
○ 1*
○ 2
○ 3
○ 4

PRESENT and RESERVE have:
○ Similar meanings
○ Contradictory meanings
○ They mean neither the same nor opposite*

A train travels 20 feet in 1/5 second. At this same speed, how many feet will it travel in three seconds?
○ 60 feet
○ 100 feet
○ 200 feet
○ 300 feet*
○ 400 feet

When rope is selling at 10 cents a foot, how many feet can you buy for 60 cents?
○ 6 feet*
○ 10 feet
○ 600 feet

The ninth month of the year is...
○ October
○ January
○ June
○ September*
○ May

Which number below represents the smallest amount?
○ 7
○ .8
○ 31
○ .33*
○ 2

Tom greeted Beth.
Beth greeted Dawn.
Tom did not greet Dawn.

Assume that the first two statements are true. Is the final one:
○ True
○ False
○ Not certain*

A boy is 17 years old and his sister is twice as old. When the boy is 23 years old, what will be the age of his sister?
○ 34
○ 40*
○ 43
○ 46

Trivia Test in Study 3

Thank you for participating in this study. Today, you will take a fun trivia game! This game has 15 questions, but each question has a time limit: You'll have a maximum of 5 seconds to answer each. The game moves quickly, so please make sure that you give this survey your undivided attention for the next few minutes. Even though the game is NOT diagnostic of your intellectual ability, please do your best and try as hard as you can to perform well on the game! When you are ready, please click the next button to start the game! Good luck!"

What does a conchologist study?
◯ Shells*
◯ Balls

What is the oldest swimming stroke?
◯ Backstroke
◯ Breaststroke*

Which of these two deserts has a larger area?
◯ Gobi
◯ Sahara*

What is the correct name for a female badger?
◯ Doe
◯ Sow*

Of which planet is Callisto a moon?
◯ Neptune
◯ Jupiter *

Which of these is not a method of cutting vegetables?
◯ Brunoise
◯ Julienne
◯ Brochette*

Is 147 a prime number?
◯ Yes
◯ No*

Where is the world's largest gold depository?
○ New York *
○ Switzerland

Ascorbic acid is better known as
○ Vitamin C*
○ Ethanol

Dishabiliophobia is the fear of
○ Untidiness
○ Getting undressed in front of others*

Which of these countries produces more wine?
○ Chile
○ Spain*

What is a baby rabbit called?
○ Kitten*
○ Hamlin

Alexander Dumas is the author of
○ Man in the Iron Mask*
○ Madame Butterfly

Which country set up the world's first chemistry lab in 1650?
○ Netherlands*
○ France

What martial arts name means "gentle way"?
○ Judo*
○ Taichi

Note: The same 15-item trivia test was used in Study 4. In that study, the prompt was as follows:
Today, we're going to ask you to answer a general knowledge test. It consists of fifteen questions, and you'll have four seconds to answer each before the page automatically advances to the next question. When you are ready to take the test, you may click next. Good luck!

Study 2 Measures

Time 1 Survey Measures

To what extent do you agree or disagree with the following statements?

(*1 = strongly disagree, 7 = strongly agree***)**

I see myself as someone who

1. …is reserved
2. …is generally trusting
3. …tends to be lazy
4. …is relaxed, handles stress well
5. …has only a few artistic interests
6. …is outgoing, sociable
7. …tends to find fault with others
8. …does a thorough job
9. …gets nervous easily
10. …has an active imagination

To what extent do you agree or disagree with the following statements?[2]

(*1 = strongly disagree, 7 = strongly agree***)**

1. In uncertain times, I usually expect the best.
2. I'm always optimistic about my future.
3. Overall, I expect more good things to happen to me than bad
4. I hardly ever expect things to go my way.
5. If something can go wrong for me, it will.
6. I rarely count on good things happening to me.

To what extent do you agree or disagree with the following statements?

(*1 = strongly disagree, 7 = strongly agree***)**

1. Having some groups on top really benefits everybody.
2. It's probably a good thing that certain groups are at the top and other groups are at the bottom.
3. An ideal society requires some groups to be on top and others to be on the bottom.
4. Some groups of people are simply inferior to other groups.
5. Groups at the bottom are just as deserving as groups at the top.
6. No one group should dominate in society.
7. Groups at the bottom should not have to stay in their place.
8. Group dominance is a poor principle.

Time 2 Survey Measures

Please answer the following questions.

(1 = *not at all*, 7 = *extremely*)

- How much do you desire having higher social status compared to others?
- How much do you desire being more well-regarded compared to others?
- How much do you desire having more prestige compared to others?

---

[2] In Studies 2 to 4, we administered the full optimism scale. Upon conducting reliability analysis, we found that some items did not hang with the scale for unknown reasons in Studies 2 and 4 particularly (Study 2: items 3 & 6, Study 4: items 3 to 6). Thus, in our final analysis, we computed optimism by excluding responses to those items, leaving a four-item scale in Study 2 ($\alpha = .92$) and a two-item scale in Study 4 ($\alpha = .86$). Importantly, using the full-scale in our analysis in those studies yielded virtually identical results.

- How much do you desire to be more respected compared to others?

To what extent do you agree or disagree with the following statements?
(*1 = strongly disagree, 7 = strongly agree*)
- I would like to be in a powerful position in an organization.
- I aspire to be in a high-ranking position in an organization.
- I seek opportunities to advance to a higher rank in the work place.
- I want to be in a high-ranking position in an organization where I have a lot of power.
- I would like to be a powerful figure at work.
- At work, I want to be the one with the highest position of power.
- At work, I want to be the one who is most "in command".
- In an organizational setting, I want to have the most authority.
- In an organizational setting, I want to be in a position with the most power.
- At work, I want to have a great deal of power.

To what extent do you agree or disagree with the following statements?
(*1 = strongly disagree, 7 = strongly agree*)
- I think I would enjoy having authority over other people.
- If given the chance, I would make a good leader of people.
- I think I am usually a leader in my group.
- I enjoy planning things and deciding what other people should do.
- I like to give orders and get things going.
- People take notice of what I say.
- When a group I belong to plans an activity, I would rather direct it myself than just help out and have someone else organize it.

To what extent do you agree or disagree with the following statements?
(*1 = strongly disagree, 7 = strongly agree*)
- I would like an important job where people looked up to me.
- I like talking to people who are important.
- I want to be an important person in the community.
- I like to be admired for my achievements.
- I like being the center of attention.
- I like to have people come to me for advice.
- I find satisfaction in having influence over others because of my position in the community.