UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Verkoeijen, P. P. J. L., et al. (2018). A Practical Illustration of Methods to Deal with Potential Outliers: A Multiverse Outlier Analysis of Study 3 from Brummelman, Thomaes, Orobio de Castro, Overbeek, and Bushman (2014). *Collabra: Psychology,* 4(1): 30. DOI: https://doi.org/10.1525/collabra.118

ORIGINAL RESEARCH REPORT

A Practical Illustration of Methods to Deal with Potential Outliers: A Multiverse Outlier Analysis of Study 3 from Brummelman, Thomaes, Orobio de Castro, Overbeek, and Bushman (2014)

Peter P. J. L. Verkoeijen, Marike G. Polak and Samantha Bouwmeester

Recently, Brummelman, Thomaes, Orobio de Castro, Overbeek, and Bushman (2014: Study 3) demonstrated that inflated praise benefits challenge seeking of children with high self-esteem, but harms challenge seeking of children with low self-esteem. In the present paper, we examined the original data set on model-fit and prediction outliers according to various reasonable criteria and norms. Subsequently, we carried out a multiverse outlier re-analysis on the data of Brummelman and colleagues' Study 3, employing the same analytical approach as the original authors did but excluding outliers. Out of the twelve re-analyses in the multiverse, six demonstrated that removing only a small number of outliers rendered the originally reported crucial interaction effect between self-esteem and type of praise non-significant and produced a sizeable reduction of the effect size. The present paper illustrates the use of reporting outlier analyses, which lies in allowing a critical evaluation of the empirical evidence and offering a more complete picture that enhances future studies in the field.

Keywords: inflated praise; self-esteem; challenge seeking; outlier analysis; multiverse analysis

In the current Western society, it seems children often receive excessive positive praise for their accomplishments. Brummelman, Thomaes, Orobio de Castro, Overbeek, and Bushman (2014) demonstrated that 25% of the praises adults give is inflated using both an experimental design (Study 1) and a field study with in-home observations (Study 2). Moreover, adults are more likely to direct inflated praise towards children with relatively low self-esteem than to children with relatively high self-esteem. Such behavior appears to be reasonable as common sense dictates that inflated praise will raise low self-esteem. Interestingly however, Brummelman and colleagues (2014) predicted that inflated praise would benefit children with *high* self-esteem, whereas it would actually be harmful to children with *low* self-esteem. Brummelman and colleagues proposed the following psychological mechanism to arrive at their prediction:

> "People with high self-esteem are self-promoting, whereas people with low self-esteem are self-protecting […]. People with high self-esteem are relatively unconcerned with failure, and seek out opportunities to demonstrate their worth and ability. They may interpret inflated praise as an encouragement, and seek challenges to display that they can meet the high standards set for them. In contrast, people with low self-esteem are relatively concerned with failure, and avoid situations that may reveal their worthlessness and low ability. They may cherish inflated praise but avoid challenges because they are afraid that they will be unable to meet the high standards set for them. Thus, paradoxically, inflated praise may backfire in children with low self-esteem and discourage them from taking on challenges." (p. 729)

To investigate the interaction between type of praise and self-esteem on challenge seeking, Brummelman and colleagues conducted a crucial third study. In this study, children made a drawing and afterwards they received no praise, non-inflated praise ("you made a beautiful drawing") or inflated praise ("you made an *incredibly* beautiful drawing"). Subsequently, children took a challenge-seeking test. The findings from this study were in line with Brummelman and colleagues' predictions: compared to non-inflated praise, inflated praise decreased challenge seeking in children with low self-esteem whereas it increased challenge seeking in children with high self-esteem.

Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, NL

Corresponding author: Peter P. J. L. Verkoeijen (p.p.j.l.verkoeijen@essb.eur.nl)

Brummelman and colleagues (2014) used a linear regression in their third study to predict challenge seeking from, self-esteem, type of praise and the interaction between self-esteem and type of praise. It is well known that outliers can have a substantive influence on the outcomes of regression analysis. However, an outlier analysis was not reported by Brummelman and colleagues. In their paper with best-practice recommendations for defining, identifying, and handling outliers, Aguinis, Gottfredson, and Joo (2013) distinguish three types of outliers. Furthermore, these authors describe various outlier diagnostics and show that for each diagnostic various (numerical) cut-offs exist to flag extreme cases. Also, various handling techniques for each outlier type are described. As a result, an outlier analysis is very likely to produce a multiverse of outcomes (cf. Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016).

In the present study, we will demonstrate, following guidelines by Aguinis and colleagues how such multiverse arises when an outlier analysis is conducted on the data of Brummelman and colleagues' (2014) third study. By doing so, we aim to underline a more general point, namely that reporting outlier analyses is an important aspect of methodological transparency that allows for a critical evaluation of the empirical evidence for a particular hypothesis. Furthermore, such practice could enhance follow-up studies as it provides a more complete picture of the population(s) of interest to other researchers in the field.

According to Aguinis and colleagues (2013), outliers can be defined as data points that lie at a distance of other data. Furthermore, they distinguish the following three types of outliers: (1) error outliers, which are the result of errors in data coding; (2) interesting outliers, which contain unexpected knowledge because they were sampled from a different population; and (3) influential outliers, which are accurate data points that are not error or interesting outliers but that do have a substantial influence on the conclusions. We will focus on identifying and handling influential outliers from Brummelman and colleagues' (2013) third study. Regarding influential outliers, Aguinis and colleagues make a distinction between model fit outliers, i.e., cases that affect model fit indices, and prediction outliers, i.e., cases that influence parameter estimates. Aguinis and colleagues (**Figure 2**, p. 289) list three general ways of handling outliers, by recommending to report findings with and without either of the following approaches: (1) respecifying the model, (2) removing outliers, and (3) robust approaches. Respecification refers to adding additional terms, such as a term that models nonlinearity, to the regression equation. Here, the main focus will be on the second point, that is, comparing results with and without outliers present.

Identifying outliers in Brummelman and colleagues' Study 3

Model fit outliers

Aguinis and colleagues (2013) recommend to use a two-step procedure for the identification of model-fit outliers. The first step is to identify univariate or multivariate outliers. This is done by combining data visualization with a quantitative approach. The second step involves assessing the change in model fit after excluding the outliers from the first step. For Brummelman and colleagues' (2013) Study 3, we employed this two-step procedure to examine the impact of univariate outliers because they can have a strong impact on the model fit. We examined the distribution of the scores on the dependent variable, i.e., the challenge-seeking score, which was measured on a scale ranging from 0 through 4, and the self-esteem scores, which ranged from 0 through 3.

Challenge seeking scores. The distribution of the challenge seeking scores appeared to be approximately normal. Furthermore, according to conventional criteria (e.g., Tabachnick & Fidell, 2013), that is, z-scores > 3.29 or < −3.29, or scores identified as extremes in a boxplot (with extremely high being defined > third quartile + 1.5* interquartile range and extremely low being defined < first quartile − 1.5* interquartile range), there were no outliers in the challenge seeking scores.

Self-esteem scores. The distribution of the self-esteem scores was skewed to the left with a small number of children having low self-esteem scores. The self-esteem z-scores are presented in **Figure 1**. According to the z-score criterion mentioned above (i.e., z-score > 3.29 or < −3.29), three children qualified as low outliers. These children had raw self-esteem scores of 0.67, 1 and 1. Because the self-esteem scores were not normally distributed, the use of the z-score outlier criterion should be treated with caution. The assumption underlying this criterion is that the data come from a standard normal distribution, where outliers are defined as scores that are unlikely given the assumed probability distribution. The criterion of z > 3.29 or z < −3.29 implies that scores that belong to the most extreme 0.1% of the reference distribution as considered as extremes. However, when the distribution is skewed, which is the case for

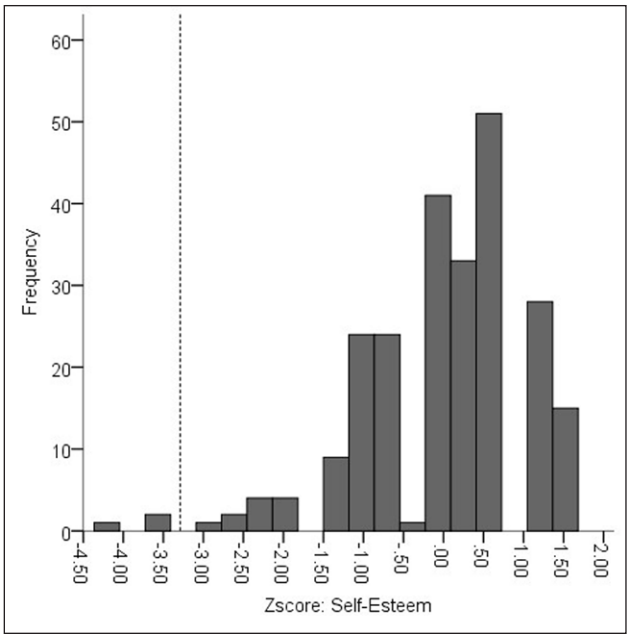

Figure 1: Histogram of standardized self-esteem scores, where the vertical dotted line marks the z < −3.29 cut-off for outliers.

self-esteem, one would have different expectations for the pattern of scores, and hence the z-score method may not be accurate. One way to deal with this issue is to compare identified extremes according to the z-criterion with extremes identified on the basis of an alternative reference distribution that more closely resembles a skewed distribution. Many alternative distributions could be considered here, for instance the *F*-distribution is a commonly known probability distribution with a skewed shape. We therefore transformed raw self-esteem scores, x_i, so that we could determine the extremity of scores under the *F*-distribution (i.e., $x_i = t_i^2$, where *t*-scores are the standardized differences from the sample mean, following the well-known principle $t^2 = F$; see appendix for the histogram of transformed scores). In the $F(1, 239)$ distribution, the critical value at 0.1% equals 11.10. Given this cut-off, the same three scores as earlier mentioned would be marked as extremes.

An alternative (nonparametric) method to identify univariate outliers, which does not strictly rely on the normality assumption, is the "boxplot criterion". According to the cut-offs associated with the "boxplot criterion", i.e., raw self-esteem score > 3 or raw self-esteem score < 1.50, there were no high outliers, but there were six low outliers.

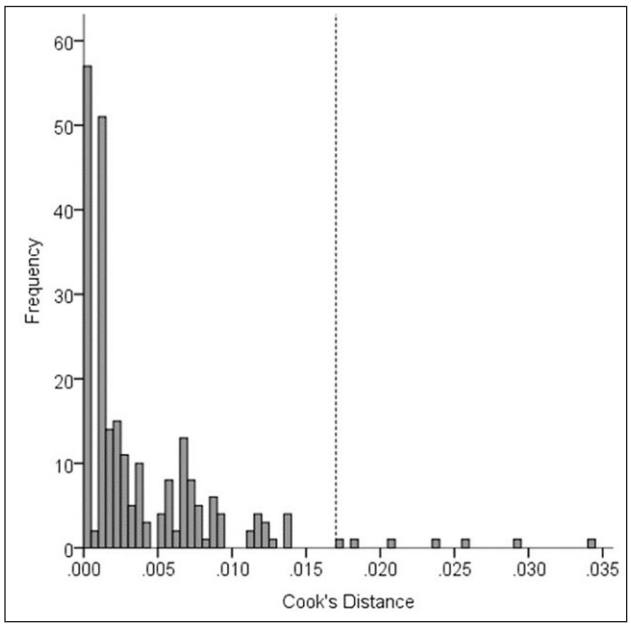

Figure 2: Histogram of Cook's *D* scores for a re-analysis of the original data of Brummelman and colleagues' (2013) third study, where the vertical dotted line marks the $D > 4/(n - k - 1) = 0.0171$ cut-off for outliers.

Prediction outliers

Aguinis and colleagues (2013) propose that the following techniques should be used to identify prediction outliers: calculate for each case (1) $DFFITS_i$ (DIFFerence in FIT, Standardized); (2) Cook's D_i, and (3) $DFBETAS_{ij}$ (DIFFerence in BETA, Standardized). We applied these techniques to the data from Brummelman and colleagues' (2014) Study 3.

$DFFITS_i$ (henceforth DFFITS). DFFITS is a diagnostic that indicates the influence of a single data point in a least squares regression. DFFITS can be understood as the standardized change in the predicted value for a data point when that point is left out of the regression analysis. For a data point to be considered a prediction outlier, Parke (2013) suggests a DFFITS cut-off score of 2. According to this cut-off score, there are no prediction outliers in Brummelman and colleagues' (2014) Study 3. Aguinis and colleagues (2013), however, recommend the following cut-off scores for DFFITS: +– 2*sqrt[$(k + 1)/n$], where *k* represents the number of predictors in the model, and *n* represents the number of observations. The regression model in Brummelman and colleagues' (2013) study had five predictors: one for self-esteem, two for the dummy variables associated with type of praise and two for the interaction terms. Hence, the DFFITS cut-off scores were +– 2*sqrt[(5 + 1)/240] = +– 0.3162. According to these cut-off scores, seven cases in Brummelman and colleagues' Study 3 were identified as prediction outliers.

Cook's distance. Cook's distance (henceforth *D*) quantifies the influence of a single data point on the parameter estimates in a least-squares regression analysis based on the residual and/or the leverage of that data point. With respect to *D*, several cut-off rules exist to identify prediction outliers (e.g., Bollen & Jackman, 1990; Cook & Weisberg, 1982; Cook & Weisberg, 1999; Fox, 1993). When it comes to numerical cut-offs, one guideline is to use $D > 1$, whereas other guidelines suggest using $D > 4/n$ or $D > 4/(n - k - 1)$, with *n* referring to the sample size and *k* to the number of predictors in the model. Cooks' distances for the original regression model applied to the data in Brummelman and colleagues' (2014) third study are presented in **Figure 2**.

As can been seen in **Figure 2**, there are no values for *D* larger than 1. The other two guidelines yielded nearly identical cut-offs for outliers, with 4/240 = 0.0167 and 4/(240–5–1) = 0.0171. Both cut-offs resulted in the identification of the same seven cases as prediction outliers, which were the same cases as were identified based on the DFFITS: +– 2*sqrt[$(k + 1)/n$] cutoffs. In **Figure 2**, these seven cases are located to the right of the vertical dotted line. Also based on visual inspection, these data points are clearly deviant from the other scores, that is, they have relatively large Cook's *D* scores (cf. Cook & Weisberg, 1999).

$DFBETAS_{ij}$ (Henceforth DFBETAS). In a least squares linear regression model, DFBETAS can be calculated for each parameter estimate and indicate the difference in a parameter estimate with and without a particular data point. Cut-off recommendations differ for DFBETA. For example, Field (2013) suggests to use a cut-off score of 1, whereas Stevens (2002) suggests a cut-off score of 2. According to these cut-off scores, there are no prediction outliers in Brummelman and colleagues' (2014) third study. Aguinis and colleagues (2013) recommend the following cut-off scores to identify prediction outliers for DFBETAS: –+ 2/sqrt(n). For Brummelman and colleagues' data this means that the cut-offs are: –+ 2/sqrt(240) = –+ 0.1291. Cases were labelled prediction outliers if at least one of the DFBETAS associated with the two interaction effect parameters (i.e., two dummy variables for praise type times × self-esteem) surpassed the cut-off score. Based on this criterion, we identified 17 prediction outliers.

Handling outliers from Brummelman and colleagues' Study 3

The above examination suggests that Brummelman and colleagues' (2014) data from their third study might contain model fit and prediction outliers. Here we will focus on reporting the outcomes of the statistical analysis with and without the outliers (see also, Simons, Nelson, & Simonsohn, 2011; Tabachnick & Fidell, 2013) to provide transparency about the potential effect of these outliers on the results of the main analysis. However, Table 3 in Aguinis and colleagues (2013) paper makes clear that many other ways of handling outliers exist that fall outside the scope of the current paper. For instance, researchers may use robust approaches, such as Bayesian statistics or M-estimation to name a few. This latter approach was adopted by the original authors in a review of a previous version of the current paper. The original authors indicated that they performed a robust regression analysis using Iteratively Reweighted Least Squares with the majorization–minimization (MM) approach (e.g., Fox, 2015): "Self-esteem (centered), two dummy variables to index the three praise conditions (with inflated praise as comparison group), and the interactions between self-esteem and these dummy variables were entered as predictors. The praise × self-esteem interaction was significant, $\chi^2(2) = 8.52$, $p = .014$" (personal communication February 3rd, 2018).

In **Table 1**, we report eleven re-analyses of Brummelman and colleagues' data by removing any outliers identified according to the afore-described criteria and subsequently using the same analytical approach as Brummelman and colleagues. Additionally, we report the outcome of the robust regression analyses conducted by the original authors. The outcomes of these analyses are presented in **Table 1**. **Figures 3** and **4** visually illustrate the consequences of removing the outliers in one of the re-analyses, namely re-analysis 2. **Figure 3** represents the regression lines of challenge seeking scores in Brummelman and colleagues' (2013) third study as a function of self-esteem, type of praise and the interaction between self-esteem and type of praise based on the complete original data. Data points left of the vertical line correspond to self-esteem z-scores lower than –3.29 and are clearly at a distance from the remaining data points in the scatterplot. **Figure 4** shows the results of the same regression analysis but with the three outliers removed. The difference between **Figures 3** and **4** in terms of the slopes of the regression lines may seem subtle, however within the observed range of self-esteem scores (i.e., from 1 to 3) it is clearly visible that the regression lines in **Figure 4** are more turned towards each other, which illustrates the non-significant interaction in this re-analysis.

In the present multiverse outlier analysis, six out of twelve re-analyses (i.e., 50%) revealed an outcome that was different from the originally reported one. In each of these six analyses, excluding a small proportion of cases rendered the crucial interaction between type of praise and self-esteem non-significant and it led to a considerable reduction of the effect sizes as compared to the original study. Also, in each of these six re-analyses, outlier removal rendered the correlation between self-esteem and

Table 1: Re-analyses of Brummelman and colleagues' (2014) Study 3 (original N = 240) after excluding model-fit and prediction outliers and after a robust regression analysis.

Analysis	Outlier exclusion criterion	#Excl. (%)	*F*-test interaction term	η2 interaction	η2 Reduction
Original study	NA	0 (0%)	$F(2, 234) = 4.491$, $p = .012$	0.037	NA
Re-analysis 1	–3.29 > Challenge seeking z-score > 3.29	0 (0%)	$F(2, 234) = 4.491$, $p = .012$	0.037	0%
Re-analysis 2	Self-esteem z-score < –3.29	3 (1.3%)	$F(2, 231) = 2.436$, $p = .090$	0.021	43%
Re-analysis 3	Self-esteem raw score < first quartile – 1.5*IQR < 1.50	6 (2.5%)	$F(2, 228) = 0.996$, $p = .371$	0.009	76%
Re-analysis 4	*DFFITS* > 2	0 (0%)	$F(2, 234) = 4.491$, $p = .012$	0.037	0%
Re-analysis 5	*DFFITS* > –2*sqrt[(5 + 1)/240] = 0.3162 or < –.3162	7 (2.9%)	$F(2, 227) = 2.022$, $p = .135$	0.018	51%
Re-analysis 6	*D* > 1	0 (0%)	$F(2, 234) = 4.491$, $p = .012$	0.037	0%
Re-analysis 7	*D* > 4/240 > 0.0167	7 (2.9%)	$F(2, 227) = 2.022$, $p = .135$	0.018	51%
Re-analysis 8	*D* > 4/(240–5–1) > 0.0171	7 (2.9%)	$F(2, 227) = 2.022$, $p = .135$	0.018	51%
Re-analysis 9	*DFBETA* > 1	0 (0%)	$F(2, 234) = 4.491$, $p = .012$	0.037	0%
Re-analysis 10	*DFBETA* > 2	0 (0%)	$F(2, 234) = 4.491$, $p = .012$	0.037	0%
Re-analysis 11	*DFBETAS* interaction term > –2/sqrt(n) = 0.1291 or < –.1291	17 (7.1%)	$F(2, 217) = 2.194$, $p = .114$	0.020	46%
Re-analysis 12	NA: robust regression analysis	0 (0%)	$\chi^2(2) = 8.52$, $p = .014$	NA	NA

Note. IQR = interquartile range; NA = not applicable.

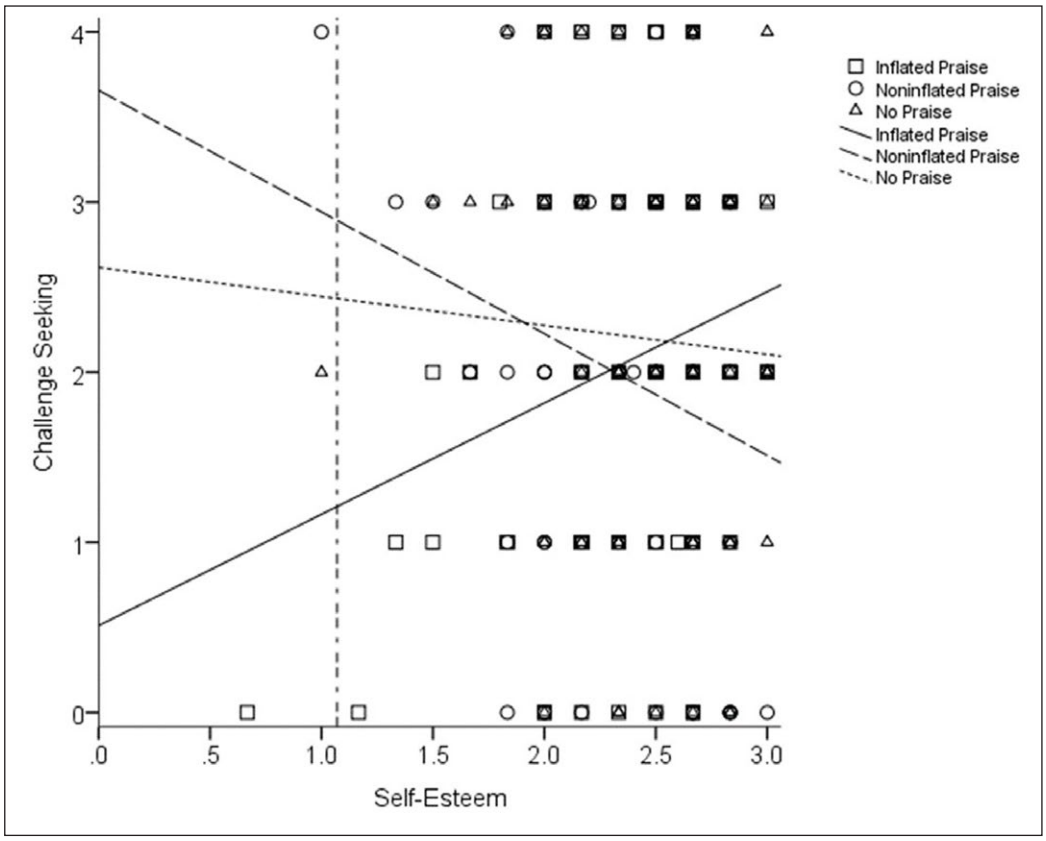

Figure 3: Re-analysis of original data: Scatterplot and regression lines of challenge seeking scores in Brummelman and colleagues' (2013) third study as a function of self-esteem and type of praise. Data points left of the vertical line correspond to self-esteem z-scores lower than –3.29.

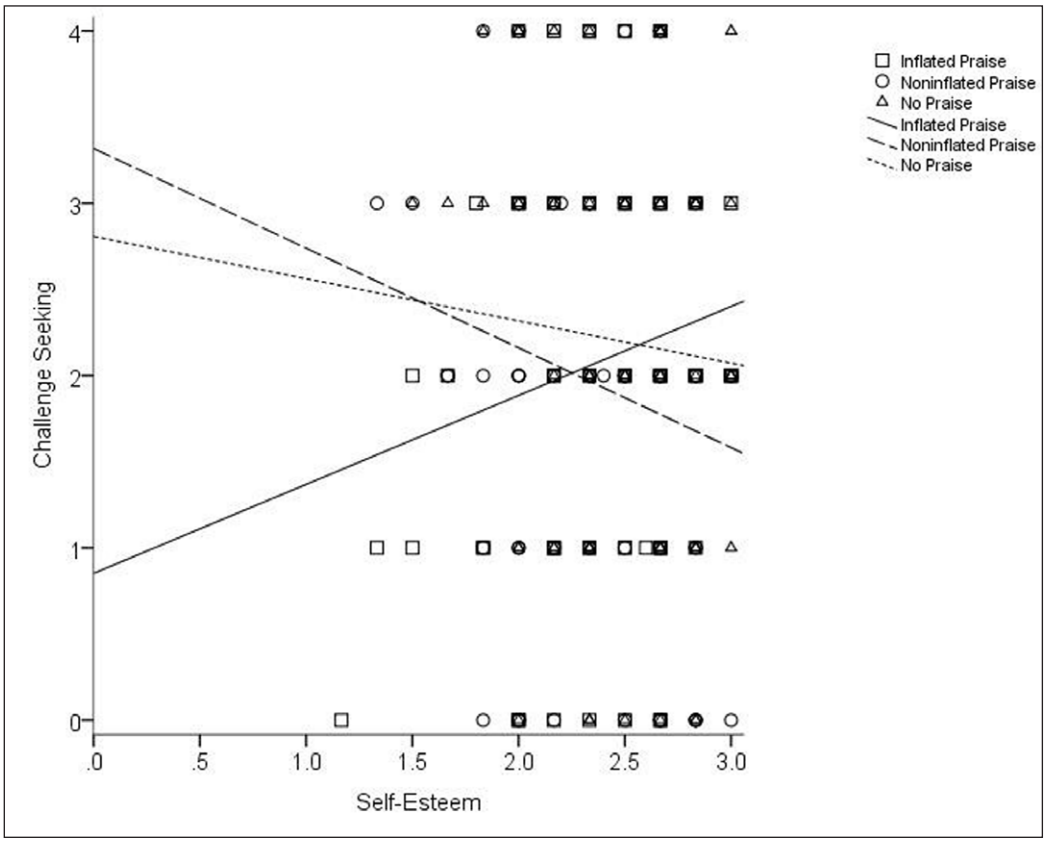

Figure 4: Re-analysis 2: Scatterplot and regression lines of challenge seeking scores in Brummelman and colleagues' third study as a function of self-esteem and type of praise after removal of three extreme cases.

challenge seeking in each of the praise type conditions (cf. simple slopes) non-significant. It could be argued that these non-significant findings were due to a reduced power to pick up the interaction effect from the original study, although this explanation seems implausible given the small number of excluded observations (the number of excluded cases varied between 3 (1.3%) and 17 (7.1%)) and given that the effect sizes in the re-analyses were much smaller than in the original study.

Discussion

In the present study, we performed a multiverse outlier analysis on the data from Brummelman and colleagues' (2014) third study. Because there are different criteria to identify influential outliers, different cut-off scores within a specific criterion, and different ways of handling outliers, our analysis produced a small multiverse of twelve re-analyses. The robust regression analysis performed by the original authors yielded the same outcome as the original study. Furthermore, according to some often reported absolute cut-off scores (i.e., *DFFITS* > 2, Cook's *D* > 1, *DFBETA* > 1, and *DFBETA* > 2), there are no influential cases in the dataset. However, relative cut-off scores that take into account sample size and model complexity recommended by Aguinis and colleagues (2013), in combination with a visual inspection of the data pattern, indicate that Brummelman and colleagues' dataset contains both a small number of model-fit and prediction outliers. Following the guidelines by Aguinis and colleagues (2013), we compared the outcomes of the statistical analysis with and without the identified outliers. These re-analyses, which comprised 50% of the multiverse, demonstrated that removing those cases had a substantial effect on the crucial interaction between self-esteem and type of praise on challenge seeking. Specifically, the interaction was no longer statistically significant in any of these re-analyses and the interaction effect size was considerably reduced relative to the originally reported effect size.

The present study underlines the importance of being transparent about the multiverse of outcomes likely to result from outlier analyses. Consistent with the example we presented here based on Brummelman and colleagues' (2014) data, we think good practice would require that researchers indicate which type of outliers they are going to examine and that they report all outcomes that emerge from applying various reasonable identification criteria, various cut-off scores and various handling techniques. This enables other researchers to critically evaluate the strength of the empirical evidence for a particular hypothesis based on as much information as possible rather than on a possibly biased subset of this information. With a reported multiverse outlier analysis, researchers can (and should) still argue why one outcome (or a subset of outcomes) from the multiverse should be prioritized. This point may be important in the context of removing outliers from an analysis. Without substantive – as opposed to statistical – concerns researchers may have good arguments to refrain from dropping influential cases (cf. Simmons et al., 2011).

The re-analyses that were part of the present multiverse outlier analysis may provide some interesting avenues for future research. Both the scatterplots of Brummelman and colleagues' (2014) third study (see **Figures 3** and **4**) and a substantial subset of re-analyses suggest that the interaction between self-esteem and type of praise on challenge seeking might depend on a few low self-esteem scores. When these cases were removed, the interaction was no longer statistically significant and the same applied to the relationship between self-esteem and challenge seeking in each of the praise type conditions. This might suggest that Brummelman and colleagues' (2014) effect does not generalize to children across a wide range of self-esteem scores and might be confined to children with (very) low levels of self-esteem. It may be interesting to investigate this possibility in upcoming research. When doing so, it should be noted that researchers (e.g., Egberink & Meijer, 2011; Van den Bergh & Van Ranst, 1998) have identified substantial problems with the reliability and scalability of the Dutch Self-worth subscale for boys aging 8 through 12. Consequently, when targeting the same Dutch age group as Brummelman and colleagues (2014) did, it would be good to restrict the sample to girls or to use an alternative measure of self-worth with better psychometric properties.

As a final point, we emphasize that the approach in the present study was limited to an illustration of Aguinis and colleagues' (2013) recommendation to identify model-fit and prediction outliers within the context of a regression analysis through various reasonable diagnostics and to report the outcomes of the regression analysis with and without the identified extreme cases. In our view, this approach has a considerable practical utility as a screening tool for influential outliers in regression analysis because it aligns well with the statistical knowledge and skills of the vast majority of researchers in the field of psychology. However, there are other – more advanced – approaches that could be used to identify and handle outliers. Aguinis and colleagues (2013) provide a rich overview of options in table 1 through 3 of their review paper, for example, robust techniques, such as M-estimation (e.g., Fox, 2015, Chapter 19), might be employed. Furthermore, researchers could consider univariate outlier detection methods that do not assume normality, but instead are adjusted to the distribution of the majority of data points (e.g., van der Loo, 2010). Additionally, in case of multivariate response data, person fit statistics could be calculated based on an Item Response Theory Framework to identify cases with deviant response pattern (e.g., Felt, Castaneda, Tiemensma, & Depaoli, 2017; Meijer, 2002).

Conclusion

In closing, the multiverse outlier analysis in the present study contains various outcomes with respect to outlier identification and outlier handling techniques (i.e., our re-analyses and the robust regression analysis). Reporting these results is important because it enhances transparency. In addition, and quoting Aguinis and

colleagues (2013), it serves to … "(a) place the burden of determination for the most "accurate conclusions" on the reader and (b) ensure complete transparency so that the handling technique does not appear to have been chosen because it supported one's hypotheses" (p. 291). Furthermore, we think that the outcomes of our re-analyses nuance Brummelman and colleagues' (2014) central claim because in half of these re-analyses the interaction between self-esteem and type of praise on challenging seeking seems to hinge on children with low levels of self-esteem. However, we invite the reader of this paper to formulate her of his own conclusion based on our multiverse outlier analysis.

Data Accessibility Statement

The present paper reports re-analyses of the Brummelman and colleagues' (2014) data from Study 3. Because we did not collect these data, we are not allowed to make them accessible. Please contact Dr. Eddie Brummelman for access to the original dataset.

Additional File

The additional file for this article can be found as follows:

- **Appendix S1.** Histogram of transformed self-esteem scores. Docx. DOI: https://doi.org/10.1525/collabra.118.s1

Acknowledgements

We are grateful to Eddie Brummelman for sharing the original data.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

- Contributed to conception: PV, MP, SB
- Contributed to acquisition of data: Not applicable
- Contributed to analysis and interpretation of data: PV, MP, SB
- Drafted and/or revised the article: PV, MP, SB
- Approved the submitted version for publication: PV, MP, SB

References

Aguinis, H., Gottfredson, R. K., & **Joo, H.** (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods, 16*, 270–301. DOI: https://doi.org/10.1177/1094428112470848

Bollen, K. A., & **Jackman, R. W.** (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In: Fox, J., & Long, J. S. (Eds.), *Modern methods of data analysis*. Newbury Park, CA: Sage.

Brummelman, E., Thomaes, S., Orobio de Castro, B., Overbeek, G., & **Bushman, B. J.** (2014). "That's not just beautiful-that's incredibly beautiful!": The adverse impact of inflated praise on children with low self-esteem. *Psychological Science, 25*, 728–735. DOI: https://doi.org/10.1177/0956797613514251

Cook, R. D., & **Weisberg, S.** (1982). *Residuals and influence in regression*. New York, NY: Chapman & Hall.

Cook, R. D., & **Weisberg, S.** (1999). *Applied regression including computing and graphics*. New York, NY: Wiley. DOI: https://doi.org/10.1002/9780470316948

Egberink, I. J. L., & **Meijer, R. R.** (2011). An item response theory analysis of Harter's Self Perception Profile for Children or why strong clinical scales should be distrusted. *Assessment, 18*(2), 201–212. DOI: https://doi.org/10.1177/1073191110367778

Felt, J. M., Castaneda, R., Tiemensma, J., & **Depaoli, S.** (2017). Using person fit statistics to detect outliers in survey research. *Frontiers in Psychology, 8*, 863. DOI: https://doi.org/10.3389/fpsyg.2017.00863

Field, A. (2013). *Discovering statistics using SPSS* (4th ed.). Thousand Oaks, CA: Sage.

Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*. Thousand Oaks, CA: Sage.

Meijer, R. R. (2002). Outlier detection in high-stakes certification testing. *Journal of Educational Measurement, 39*, 219–233. DOI: https://doi.org/10.1111/j.1745-3984.2002.tb01175.x

Parke, C. (2013). *Essential first steps to data analysis: Scenario based examples using SPSS*. London, England: Sage. DOI: https://doi.org/10.4135/9781506335148

Simmons, J. R., Nelson, L. D., & **Simonsohn, U.** (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366. DOI: https://doi.org/10.1177/0956797611417632

Steegen, S., Tuerlinckx, F., Gelman, A., & **Vanpaemel, W.** (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*, 702–712. DOI: https://doi.org/10.1177/1745691616658637

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Erlbaum.

Tabachnick, B. G., & **Fidell, L. S.** (2013). *Using multivariate statistics*. Boston: Pearson/Allyn & Bacon.

Van den Bergh, B. R. H., & **Van Ranst, N.** (1998). Self-concept in children: Equivalence of measurement and structure across gender and grade of Harter's Self-Perception Profile for Children. *Journal of Personality Assessment, 70*(3), 564–582. DOI: https://doi.org/10.1207/s15327752jpa7003_13

van der Loo, M. P. J. (2010). Distribution based outlier detection for univariate data. Discussion paper 10003. The Hague: Statistics Netherlands.

Peer review comments
The author(s) of this paper chose the Open Review option, and the peer review comments are available at: http://doi.org/10.1525/collabra.118.pr

How to cite this article: Verkoeijen, P. P. J. L., Polak, M. G., & Bouwmeester, S. (2018). A Practical Illustration of Methods to Deal with Potential Outliers: A Multiverse Outlier Analysis of Study 3 from Brummelman, Thomaes, Orobio de Castro, Overbeek, and Bushman (2014). *Collabra: Psychology,* 4(1): 30. DOI: https://doi.org/10.1525/collabra.118

Senior Editors: Simine Vazire, Victoria Savalei

Submitted: 23 October 2017 **Accepted:** 30 July 2018 **Published:** 23 August 2018

UNIVERSITY of CALIFORNIA PRESS Collabra: Psychology

Collabra: Psychology is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS