# A successful test of parallel replication teams in teaching research methods

Lionel G. Standing, Lisa Astrologo, Felecia F. Benbow, Chelsea S. Cyr-Gauthier & Charlotte A. Williams

*This paper describes the novel use of parallel student teams from a research methods course to perform a replication study, and suggests that this approach offers pedagogical benefits for both students and teachers, as well as potentially contributing to a resolution of the replication crisis in psychology today. Four teams, of five undergraduates each, independently attempted exact replications of Study 8 by Gailliot et al. (2007), which reported that participants' self-control is enhanced by consuming a glucose drink. In a 2 × 2 independent groups design, participants (N=306) first consumed a glucose drink or a placebo, and then wrote about death, intended to deplete their self-control, or dental pain as a control condition. Absolute levels of self-control were lower here than in the target article (shown by more items left unsolved in a word puzzle), but its main result was replicated, since self-control overall was raised by the glucose drink. Also, the teams reliably reported similar effects for the experimental treatments (ICC=.928). Two differences from the target study results were noted: the glucose effect occurred only with female participants, and no effect was found from the writing scenario used.*

**Keywords:** *teams, replication, teaching, research methods, glucose, self-control.*

Recently many papers have appeared in the literature that point to a crisis of confidence in psychology and many other disciplines: it is now recognised that published results often cannot be replicated by independent investigators. Although an exact success rate for replication studies cannot be given overall, recent studies commonly suggest that fewer than half of all replication attempts succeed in reproducing the original findings (see, e.g. Ioannidis, 2005; Ioannidis, 2012b; Neuliep, 1990; Pashler & Wagenmakers, 2012; Pashler & Harris, 2012; Ritchie et al., 2012). Although Klein et al. (2014a) were able to replicate 10 of 13 recent studies in cognitive and social psychology (77 per cent), this level of success is unusually high. The PsychFileDrawer.org website, which covers a wider range of content areas for the target studies, currently reports only 25 successes in 84 attempts, a success rate of 30 per cent. Furthermore, the definitive study of reproducibility to date reported success in only 39 out of 100 carefully conducted replication attempts, per-formed by established researchers (Open Science Collaboration, 2015). Although we teach our students that psychology is a discipline based on evidence, on closer inspection the evidence often appears to be unreliable.

This untenable situation is worsened because replication attempts are very seldom published in the standard journals, leading to the pervasive file-drawer problem, in which unsuccessful replication attempts usually remain forgotten in the investigator's files (Rosenthal, 1979); the number of negative results published is actually shrinking (Fanelli, 2011). When mandatory preregistration of National Heart Lung, and Blood Institute medical trials was introduced recently, presumably eliminating the file-drawer problem, the percentage of these cardiovascular studies that reported significant benefits abruptly dropped from 57 per cent to 8 per cent (Kaplan & Irvin, 2015).

In the basic form of a replication study, followed here, a new investigator chooses a target paper that has been published in a

peer-reviewed journal, attempts to perform it again by fulfilling all details of the stated protocol as closely as possible (or exceeding them, e.g. regarding the number of participants), and analyses the resultant data on the basis of the original statistical tests. This comprises a *direct* replication, but alternatively, the investigator may choose to perform a *conceptual* replication in order to test the central idea of the target study while employing, e.g. new materials or testing procedures. A direct replication is generally the appropriate starting point, since a failure to reproduce results in a conceptual replication could be due either to flaws in the theory, or in its generalisability.

An additional reason for performing the team replications reported here was that individual student term projects in an undergraduate research methods class are all too often deficient in their theoretical rationale, methodology, and analysis, as well as having an inadequate sample size (Frank & Saxe, 2012, p.601; Standing et al., 2014). A literature search does not reveal any previous empirical studies which deal directly with parallel student teams and replication.

An earlier paper has reported on a set of four different replication studies performed by teams of student experimenters, only one of which confirmed the findings of the target paper (Standing et al, 2014; Lane et al, 2012). The one success involved an exact replication of Gailliot et al. (2007, Study 8), which was therefore chosen to provide the target study for the present work. This study found that raising the participants' level of glucose will counteract the decrease in self-control (or 'ego-depletion') that is produced by previous efforts at self-control, supporting the authors' view of self-regulation processes as similar to a mental muscle which with exercise becomes tired, but then can be replenished by providing a glucose drink to provide an additional energy source. In a 2 × 2 independent groups design, the self-control of experimental participants was first drained, by requiring them to deal with thoughts of their own mortality (whereas control participants thought about dental pain). All participants were then given either a glucose drink or a placebo drink, and shortly thereafter attempted to solve a word-fragment task: the number of items left uncompleted was taken as a measure of impaired self-control. As predicted, their results showed impairment of self-control only for the mortality-placebo group, which left about 55 per cent of word fragments unsolved, whereas each of the other three groups left approximately 17 per cent unsolved (Gailliot et al., 2007, Figure 3).

The present study followed a complementary route to the earlier work of Standing et al. (2014), by forming an undergraduate research methods class into student teams, each of which independently attempted an exact replication of the same target study. Our objective was to replicate the results of Gailliot et al. (2007), and to explore how well the different teams of students would agree in their experimental results. Although Gailliot et al. mention that they studied 51 female and 22 male undergraduates, they reported no data concerning possible gender differences in the observed behaviour, so we added gender as an independent variable, which could be done without changing the basic design.

The total number of participants here was large enough to easily meet the criterion derived by Simonsohn (2015) for a new study to have an adequate chance of replicating an earlier one, which is that it should involve at least 2.5 times the number of participants. The target study had 73 participants, so the minimum acceptable number of participants for our replication was 183.

## Method
### Participants and testers
The participants were 306 unpaid volunteers recruited from various undergraduate psychology classes (with a mean age of approximately 20 years); 68 per cent were female. All were fluent in English, but about 20 per cent were bilingual francophones. Participants were treated in accordance with a protocol

approved by the campus Research Ethics Board.

At an initial class meeting, all 20 members of an undergraduate advanced research methods class were first given the option to create their own individual term project, although none accepted the offer. They were then randomly assigned by the experimenter to form four teams of testers of five members each; one member was designated as the team Coordinator, with responsibilities for day to day issues of subject recruitment, lab booking, supplies, etc. Regular liaison between the instructor and the Coordinators was emphasised, although the teams worked independently. Each team was first trained by the experimenter in the experimental protocol and given detailed written instructions for the specific steps needed to conduct the study. This training session also involved some coaching as to the appropriate general conduct to maintain during testing (no talking with the participants during testing, for example), and included practice runs where the teacher played the role of the subject and dealt with student questions about procedure. Each student had previously written a detailed research proposal in APA format as a graded assignment, to outline the intended study, ensuring that they were thoroughly familiar with both the theory and the practical details involved.

## Materials[1]

A standard drink was used consisting of 410ml of lemonade, made from a sliced lemon and water. This was sweetened with either 36g of glucose powder for the experimental condition, or a packet of Splenda for the placebo condition (a sucralose-based artificial sweetener with zero calories). The drinks were served cool, in plastic cups. Written measures of taste and liking for the drink were obtained on three 5-point scales (How pleasant was it for you while drinking the beverage? How much would you like to drink it again? How appealing is the appearance of the drink?)

A sheet was provided on which participants were asked to write down their thoughts either about what would happen to their body after death (mortality salience scenario), or about dental pain (control scenario); it was assumed, on the basis of terror management theory, that performing the mortality salience task will drain a subject's self-control (Rosenblatt et al., 1989; Parry, 2015). A sheet of 20 simple word fragments (e.g. _ _ATULA) to be completed by the participant was employed; the number of fragments that were left unsolved provided a measure of the depletion of self-control. An easy crossword puzzle was also used as a filler task.

## Procedure

Each participant, having signed a consent form, was first randomly assigned to one of the four cells of the $2 \times 2$ independent groups design: Drink Type (glucose or placebo) $\times$ Scenario (dental pain or mortality writing task). They then consumed their assigned drink, and completed the appropriate scenario in writing. Next they worked on a crossword puzzle and completed a filler questionnaire for six minutes. This delay was designed to allow adequate time for the stress of writing about death to drain the participants' self-control. Participants then were given the sheet of incomplete word fragments and asked to work at solving them. No time limit was imposed. (The Marlowe-Crowne Social Desirability Scale, given at this point in the target study, was not used, as it revealed no effect there). Finally, participants estimated how many calories they thought their drink had contained, to check for any perceived difference between the drinks which might influence their persistence in the word-fragment task. Testing was performed under double-blind conditions, in that neither the testers nor the participants knew whether a given subject

1 The written materials are available on request from the first author

consumed the glucose drink or the placebo. However, it was not possible to blind testers concerning the writing scenario that participants were assigned to.

Following testing, the Coordinators and instructor collated the data and distributed the total data set to the whole class, who then analysed the scores with SPSS and wrote a graded final report of the study in APA format, with the option to collaborate with fellow team members as joint authors.

## Results
### Overall effects of mortality salience and glucose on self-control (2 × 2 ANOVA)

Using all participants, a 2 × 2 independent groups ANOVA was performed on the number of word fragments that were left unsolved, as a function of the type of drink consumed (glucose vs. placebo) and the writing scenario (mortality vs. dental pain). This showed that the mean proportion of fragments left unsolved was lower with the glucose drink overall, $F(1, 302)=5.01$, $p=.026$, indicating that self-control was enhanced. The effects of the scenario and the Drink Type × Scenario interaction were both non-significant, $F(1, 302)=0.014$, $p= .905$, and $F(1, 302)=.747$, $p=.388$, respectively. The mean proportions of word fragments left unsolved are shown in Figure 1, where they are compared with the results obtained by Gailliot et al. (2007). This overall result for the drink type variable replicates the central result of the target study, in that the glucose drink again increased the participants' self-control, the key comparison being that the placebo-mortality group left more fragments unsolved than did the glucose-mortality group, $t(156)=2.33$, $p=.011$. The lack of a scenario effect does not contradict the mental muscle theory, but it differs from the pattern seen by Gailliot et al., since the number of unsolved word fragments left, averaged over glucose and placebo conditions, was slightly (non-significantly) *higher* under dental pain than mortality salience.

Thus there was no sign of mortality threat here.

### Male-female differences in self-control

The effect of the participants' gender was examined as some check on the generality of the results. A 2 × 2 ANOVA (Drink Type × Gender), ignoring scenario since it had been found nonsignificant, again replicated the results of the target paper: glucose as compared to the placebo drink produced more self-control (fewer fragments left unsolved), even though the trend showed only marginal significance at the overall level, $F(1, 290)=3.20$, $p=.07$. However, the female participants clearly showed this enhancement of self-control under glucose conditions, $t(198)=2.75$, $p< .005$, whereas for the males no effect of the drink type was found, $t(92)=2.84$, $p=.39$. This gender difference is illustrated in Figure 2.

### Comparison of the results obtained by the four teams

The mean level of self-control shown for drink and scenario conditions was examined as a function of the team which had tested the subject. This three-way ANOVA (Drink × Scenario × Team) confirmed that more word fragments were left unsolved under the placebo than the glucose condition, $F(1, 290)=5.61$, $p=.018$, although it was not affected by the scenario, $F(1, 290)=.001$, $p=.976$. This glucose-placebo difference was seen with all four teams, as shown in Figure 3, although the trend reached only marginal statistical significance within two teams ($p=.071$ and $.067$) and was not significant for the other two ($p=.76$ and $.31$). No interactions were significant (all $p > .4$). The ANOVA also showed that significant differences existed between the four teams regarding overall proportions of fragments left unsolved, pooled over drink and scenario conditions, $F(1, 290)$ 2.99, $p=.031$, and the LSD procedure revealed that the scores for Team 4 were lower than for Teams 1 and 2, $p= .034$ and $.007$, respectively. However, these differences were not large, as the
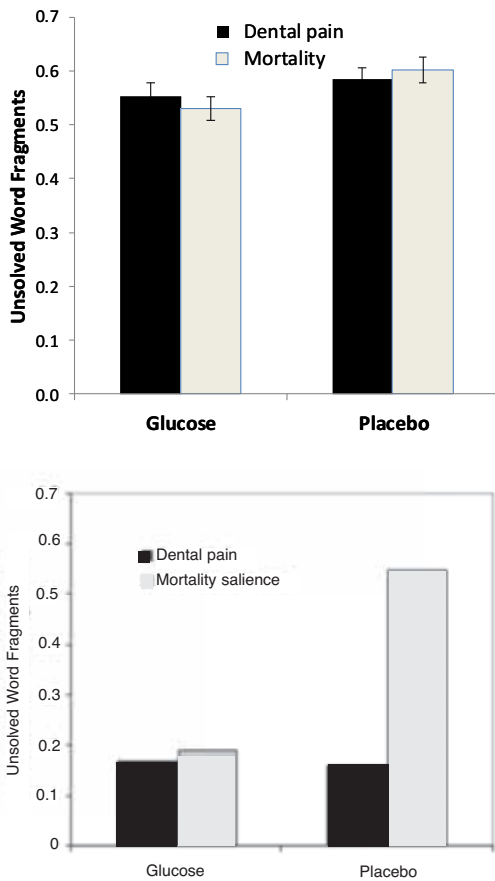
Figure 1. Self-control (proportion of word fragments left unsolved) for all teams as a function of writing scenario (dental pain vs. mortality salience) and drink type (glucose vs. placebo), compared below with the results of Gailliot et al., Study 8. Fewer word fragments left unsolved indicate more self-control. Error bars show standard error of the mean.

overall proportions of fragments left unsolved, pooled over drink and scenario conditions, for teams 1 to 4 were .58, .60, .56, and .52, respectively.

The responses to the four treatments (Drink × Scenario) obtained by each of the four teams is illustrated in Figure 4, based on all participants, showing fairly similar although not identical patterns. These measures show good reliability, with an intraclass coefficient of correlation, or ICC, of .928, as

shown by online computation (Chinese University of Hong Kong, 2014; Model 3, meaned). An alternative index of inter-team agreement was obtained by ranking the four means (Drink × Scenario) obtained by each team in order of relative magnitude. These rankings yielded a Krippendorff's alpha of .906, as computed online (Freelon, 2015), again indicating satisfactory reliability.

However, it is noteworthy that the absolute level of these means for the proportion of unsolved word fragments is about three times higher than is shown for three out of four groups in the target article (as illustrated in Figure 1), i.e. values of about .6 rather than .2 were now seen, indicating that for unknown reasons our participants showed much less self-control overall than did the participants of Gailliot et al. The effect size *(d)* observed here for the glucose drink compared to the placebo is 0.26, as compared to a value of 0.65 that was reported in the target paper, or 0.30 as found by Standing et al. (2014). This change represents a decline from a medium to a small effect size (Cohen, 1992).

## Discussion

At a basic level, the results of this study represent a successful replication of the key idea in the target experiment, since more self-control was shown by participants after they had ingested glucose rather than a placebo drink, although the absolute level of self-control shown here was lower than before (with more word fragments left uncompleted), and the treatment differences were smaller. Accordingly, the present study was posted as a successful replication on the PsychFile-Drawer website (Astrologo, Benbow, Cyr-Gautier, Williams, & Standing, 2015). However it should be noted that the enhancement of self-control by glucose, although commonly observed (as argued in a review by Gailliot, 2015), may still fail to occur (Clohecy, Standing, & McKelvie, 2015).

Concerning the question of major interest here, the four teams of testers were found to
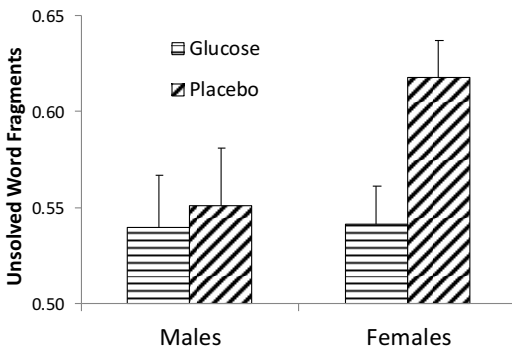
Figure 2. Word fragments left unsolved as a function of participants' gender and the drink type (glucose vs. placebo), summed over scenarios. Error bars show SEMs.

produce a similar pattern of results across the four Drink × Scenario combinations, for the overall data. However, the participants unexpectedly showed no more ego-depletion when tested with the mortality scenario as compared to the dental pain scenario, suggesting that death for them held no more terrors than the dentist's chair. While it was noted that team 4 produced significantly higher self-control scores than the others (pooled over drink and scenario conditions), the difference was not great, and the relative pattern of response across the treatment conditions was fairly similar within each team's results, indicating adequate reliability.

Gender was not analysed in the target study, which mentions only that 70 per cent of the sample were female (Gailliot et al., 2007, p.331), nor in the other eight experiments within that paper, so this potential moderator variable was left unexplored. The present data show the predicted glucose effect only for females, for reasons which are unclear but merit further study, while the failure of both males and females to respond to mortality threat was unexpected in terms of terror management theory (Rosenblatt et al., 1989).

We may conclude from this initial study that the use of parallel student teams to conduct a replication study was quite easy to

arrange, and gave consistent results. We feel that we can recommend this approach in teaching research methods at an intermediate level, although not for an introductory course, and the students seemed to respond positively to the experience of working in small groups. Not only does this approach acquaint students with some recent research ideas, but it provides many potentially valuable teaching moments related to methodological issues such as the conduct of double-blind testing, statistical power, and effect sizes, as well as the importance of being alert for potential moderator variables, such as the participants' gender, which can affect the data. The results obtained can often be reported on the PsychFileDrawer website, and thus should also benefit the discipline of psychology as a whole, as well as encouraging students to take public responsibility for their published work. Finally, an instructor may be glad to see that a given trend was obtained consistently by all the teams and cannot be dismissed as due to ineptitude or a simple fluke.

The problems encountered here appear to include the inherently low reliability of psychological findings in general, and the failure of the previous experimenters to provide their data broken down according to major demographic variables such as gender. We would see the present approach as complementary to the Many Labs Project (for example), where a number of different laboratories were all organised to attempt replications of the same papers, a project which found good cross-tester reliability (Klein et al., 2014a). Another valuable approach involves the use of preregistered replication reports, in which the details of the analysis are specified in advance, and the results must always be published whether or not they are positive (Wagenmakers et al., 2015; Kaplan & Irvin, 2015).

These various approaches are not mutually exclusive and may all have merit, although the present one appears to us to be the easiest and fastest to implement, as well as providing pedagogical benefits to the
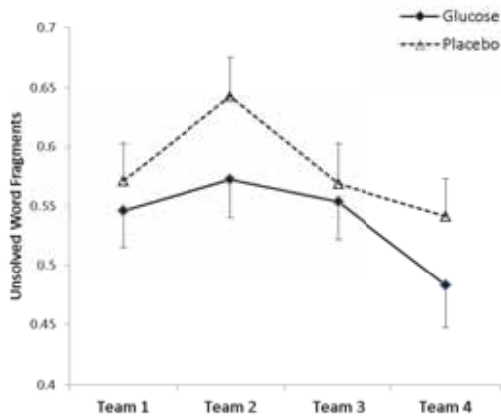
Figure 3. Mean proportion of word fragments left unsolved as a function of glucose versus placebo drink, pooled over mortality and dental pain scenarios, according to each of the four teams of testers. (All participants). Error bars show SEMs.

students in terms of experience, critical thinking about experimental control and statistical power, the analysis of data, and scientific writing. Our impression was that the students became involved in the study, because it was 'theirs', and seemed interested in the controversy over replication today. The flood of published papers in the journals today would require massive amounts of testing in order to validate them, and classes of undergraduate or graduate students potentially represent a vast and almost untapped resource, as argued by Grahe et al. (2012). Even though students are relatively inexperienced in terms of testing skills, Standing et al. (2014) provide data which suggest that their success rate in conducting replications is at least as high as that seen with other groups of testers. Altogether, we feel that there is now some empirical support for the enthusiastic arguments advanced in favour of student replications by Frank and Saxe (2012), although the present study represents no more than a start.

While an instructor might wish to enrich the variety of the material covered by using

teams that tackle several different target papers, as was done by Standing et al. (2014), the use of parallel teams has the overriding advantage that the combined *N* will be much larger, with an associated increase in statistical power. Power is a crucial consideration here, particularly as it may be weakened due to encountering an effect size that is lower than the value reported in the target paper, as was seen in the present study. The replication attempts that are posted on PsychFile-Drawer.org report the use of about the same number of participants overall as were used in the target studies (with median values of 99 and 90, respectively), but this is less than half the number that is needed to provide enough power for an adequate test of replicability, according to the calculations of Simonsohn (2010).

It must be recognised that there may be subtle methodological issues with replication studies, so that the possible goals of investigators may be more diverse than simply to obtain a statistically significant result, e.g. they may wish to establish precisely an effect size (Anderson & Maxwell, 2015). Again, a combined series of studies does not necessarily give a more precise estimate of effect size (Nuijten, van Assen, Veldkamp, & Wicherts, 2015). Nor do the results of replication study always fall into a simple yes/no paradigm, but may require quite complex levels of analysis (e.g. Rohrer et al., 2015). There is also the recently-recognised problem that a reported treatment effect, e.g. the response of patients to cognitive-behavioural therapy or a heart drug, may not be constant but rather decline over time (Johnsen & Friborg, 2015; Lehrer, 2010), so that a replication study potentially may be aiming at a moving target; issues such as these are discussed by Klein et al. (2014b). It should also be noted that replications may carry only spurious credibility if they are performed by the original investigators (Ioannidis, 2012a), and that an exact rather than a conceptual replication is normally the best starting point.

Despite these issues, surely the main

*Lionel G. Standing, Lisa Astrologo, Felecia F. Benbow, Chelsea S. Cyr-Gauthier & Charlotte A. Williams*
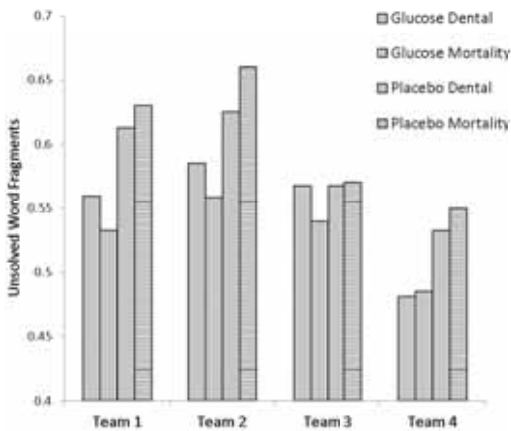
Figure 4. Mean proportion of unsolved word fragments left as a function of drink (glucose vs. placebo) and writing scenario (dental pain vs. mortality), as measured by each of the four teams of testers. (All participants).

problem responsible for the crisis today is simply that so few replication attempts are performed and published. Since there are many thousands of research methods classes active worldwide each year, we believe that the approach described here may substantially benefit both the skills and critical thinking of psychology students, and the discipline of psy-

chology in general. Some practical suggestions in this endeavour are given by Standing (2016). Any new approach to teaching which may encourage researchers to perform more replication studies would also be an important part of the influential 'New Statistics' methodological reform movement that has been proposed by Cumming (2014).

## Author note

Correspondence may be sent to:
**L. G. Standing,**
Department of Psychology,
Bishop's University,
2600 College St., Sherbrooke QC,
Canada J1M 1Z7, or lstandin@ubishops.ca

## References

Anderson, S.F. & Maxwell, S.E. (2015, July). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods, 20.* doi: 10.1037/met0000051

Astrologo, L., Benbow, F., Cyr-Gautier, C., Williams, C., & Standing L.G. (2015). Glucose promotes self-control following depletion (#223). (Online publication in PsychFileDrawer archive). Retrieved from http://www.psychfiledrawer.org/replication.php?attempt=MjIz

Chinese University of Hong Kong, Department of Obstetrics and Gynaecology (2014). Intraclass correlation. (Online calculator). Retrieved from http://department.obg.cuhk.edu.hk/ research-support/IntraClass_correlation.asp

Clohecy, E.D., Standing, L.G. & McKelvie, S.J. (2015). What enables self-control? A test of glucose, fructose, and vagus nerve activation as possible factors. *Escritos de Psicología, 8*(1), 38–43.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. doi:10.1037/0033-2909.112.1.155

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1) 7–29. doi: 10.1177/0956797613504966

Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90,* 891. doi:10.1007/s11192-011-0494-7

Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science, 7,* 600–604. doi:10.1177/1745691612460686

Freelon, D. (2015). ReCal for ordinal, interval, and ratio data (OIR). (Online calculator). Retrieved from http://dfreelon.org/utils/recalfront/recal-oir/

Gailliot, M. T. (2015). Energy and psychology. *Global Journal for Research Analysis, 4*(7), 253–255.

Gailliot, M.T., Baumeister, R.F., DeWall, C.N., Maner, J.K., Plant, E.A., Tice, D.M., Brewer, L.E., Schmeichel, B.J. (2007). Self-control relies on glucose as a limited energy source: willpower is more than a metaphor. *Journal of Personality and Social Psychology, 92*(2), 325–336. doi:10.1037/0022-3514.92.2.325

Grahe, J.E., Reifman, A., Hermann, A.D., Walker, M., Oleson, K.C., Nario-Redmond, M. & Wiebe, R.P. (2012). Harnessing the undiscovered resource of student research projects. *Perspectives on Psychological Science, 7*(6), 605–607. doi:10.1177/1745691612459057

Ioannidis, J.P. (2005). Why most published research findings are false. *PLoS Med, 2*(8): e124. doi:10.1371/journal.pmed.0020124

Ioannidis, J.P. (2012a). Scientific inbreeding and same-team replication: Type D personality as an example. *Journal of Psychosomatic research, 73*(6), 408-410. doi:10.1016/j.jpsychores. 2012.09.014

Ioannidis, J.P. (2012b). Why science is not necessarily self-correcting. *Perspectives on Psychological Science, 7*(6), 645–654. doi:10.1177/ 1745691612464056

Johnsen, T.J. & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin, 141*(4), 747–768. doi:10.1037/bul0000015

Kaplan, R.M. & Irvin, V.L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS One, 10*(8): e0132382. doi:10.1371/journal.pone.0132382

Klein, R.A., Ratliff, K.A., Vianello, M., Adams, R.B. Jr., Bahník, S., Bernstein, M.J., Nosek, B.A. et al. (2014a). Investigating variation in replicability: A "Many Labs" replication project. *Social Psychology, 45*(3), 142–152. doi:http://dx.doi.org/10.1027/1864–9335/a000178

Klein, R.A., Ratliff, K.A., Vianello, M., Adams, R.B. Jr., Bahník, S., Bernstein, M.J., Nosek, B.A. et al. (2014b). Theory building through replication: Response to commentaries on the ''Many Labs'' replication project. *Social Psychology, 45*(4), 307–310.

Lane, E.A., Beadman, K.S., Bélisle, A.M., Campbell, R.M., Cournoyer Lemaire, E., Stewart, D. & Standing, L.G. (2012). Glucose promotes self-control processes (#145). (Online publication in PsychFile-Drawer archive). Retrieved from http://www.psych-filedrawer. org/replication.php?attempt=MTQl

Lehrer, J. (2010, Dec 13). The truth wears off: Is there something wrong with the scientific method? *New Yorker*. Retrieved from www.newyorker.com/magazine/2010/12/13/the-truth-wears-off

Neuliep, J.W. (Ed.). (1990). *Handbook of replication research in the behavioral and social sciences.* Corte Madera, CA: Select Press.

Nuijten, M.B., van Assen, M.A., Veldkamp, C.L. & Wicherts, J.M. (2015). The replication paradox: Combining studies can decrease accuracy of effect size estimates. *Review of General Psychology, 19*(2), 172-182. doi:10.1037/gpr0000034

Open Science Collaboration (2015, Aug 28). Estimating the reproducibility of psychological science. *Science, 349*(6251), 943-951. doi:10.1126/science.aac4716

Parry, M. (2015, May 22). Death denial: Does our terror of dying drive almost everything we do? *Chronicle of Higher Education: The Chronicle Review.* Retrieved from http://chronicle. com/article/Mortal-Motivation/230303

Pashler, H. & Wagenmakers, E. (2012). Editors' introduction to the Special Section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528–530. doi:10.1177/1745691612465253

Pashler, H. & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*(6), 531–536. doi:10.1177/ 1745691612463401

Ritchie, S.J., Wiseman, R. & French, C.C. (2012). Replication, replication, replication. *The Psychologist, 25*(5), 346–357.

Rohrer, D., Pashler, H. & Harris, C.R. (2015, August). Do subtle reminders of money change people's political views? *Journal of Experimental Psychology: General, 144*(4), e73–e85. doi:10.1037/xge0000058

Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T. & Lyon, D. (1989). Evidence for terror management theory I: The effects of mortality salience on the reactions to those who violate or uphold cultural values. *Journal of Personality and Social Psychology, 57*, 681-690. doi:10.1037/0022-3514.57.4.681

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641. doi:10.1037/0033-2909.86.3.638

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559–569. doi:10.1177/0956797614567341

Standing, L.G. (2016). How to use replication team projects in a research methods course. *Essays from Xcellence in Teaching, XV* (Society for the Teaching of Psychology), 26-31. Retrieved from http://teachpsych.com/ebooks/eit2015/index.php

Standing, L.G., Grenier, M., Lane, E.A., Roberts, M.S. & Sykes, S. J. (2014). Using replication projects in teaching research methods. *Psychology Teaching Review, 20*(1), 94–102.

Wagenmakers, E., Beek, T. F., Rotteveel, M., Gierholz, A., Matzke, D., Steingroever, H., et al. (2015). Turning the hands of time again: a purely confirmatory replication study and a Bayesian analysis. *Frontiers of Psychology, 6*, 1–6. doi:10.3389/fpsyg.2015.00494