### Status as a replication

Replication studies can provide important insights into phenomena as they assess whether procedures, analyses, results, and claims can be reproduced and generalized (REF). There is no single definition of what does and does not constitute a replication, and there is ongoing debate about what nomenclature and taxonomy is useful in distinguishing between subtypes such as conceptual versus direct replications (e.g., REFs). However, a common theme in attempts to describe the defining features of replication is a strong overlap between studies in their hypothesis, design, measures, and/or analyses. The “success” or “failure” of a replication study is typically defined in terms of the congruence in its results and support for the hypothesis between the replication study and the original study. Such overlaps serve to constrain Researcher Degrees of Freedom, that is, flexibility in conduction of a study which may (inadvertently or not) serve to weaken the evidence supporting a study’s claims and give rise to non-replicable results. Researcher Degrees of Freedom have been argued to play a central role in the replication crisis in psychology (Nosek et al., 2022; Simmons et al., 2011).

Foody et al. (2013) state that “the research was a replication of a previous study by Luciano et al. (2011)”. Their article therefore gains additional importance by identifying as a replication, as should therefore represent a higher bar of evidence for the underlying claim. In order to assess whether it is appropriate to label Foody et al. (2013) as a replication study – a distinction which influences the evidentiary value we should place on its results – it is therefore useful to examine the overlap between Foody et al. (2013) and Luciano et al. (2011). Foody et al. (2013) do acknowledge one important difference between their work and Luciano et al. (2011) when self-identifying the work as a replication: “the research was a replication … except that we were able to use less intensive interventions with our non-clinical sample” (p. 384). However, many other unacknowledged differences exist (see Table 1).

Table 1. Comparisons between Foody et al. (2013) and Luciano et al. (2011)

|  |  |  |
| --- | --- | --- |
|  | Luciano et al. (2011) | Foody et al. (2013) |
| Sample | Clinical sample of adolescents | Non-clinical sample of adults |
| Design | Quasi-experiment (non-random assignment to defusion intervention conditions based on severity) | Laboratory analogue study (random assignment to defusion intervention condition) |
| Distress induction procedure | No | Yes |
| Interventions | Value clarification (all participants)  Defusion (distinction vs. hierarchical conditions; non-random assignment) | Defusion (distinction vs. hierarchical conditions) |
| Timepoints | 7 therapy sessions over 4 months | Single experimental session with 3 measurements (baseline, post distress induction, post defusion intervention) |
| Outcome variables | Multi-item self-report Likert scales.  *Established measures:*  Spanish Avoidance and Fusion Questionnaire (AFQ-S)  Accepting without Judgment Scale of the Kentucky Inventory of Mindfulness Skills (KIMS)  *Ad hoc measures:*  Impulsive Behavior Inventory (IBI)  Emotional Behavior Inventory (EBI) | Single item Visual Analogue Scales  *Ad hoc measures:*  Discomfort  Anxiety  Stress  e.g., “How much discomfort do you feel right now?”  n.b., these are collectively described as measures of “distress” and all conclusions related to distress rather than |
| Analytic method | Mann-Whitney *U* tests between conditions | Not tested direct. RM-ANOVAs fitted to all three time points between conditions |
|  |  |  |
|  |  |  |
|  |  |  |

However the overlap between the two studies is very limited and Foody et al. (2013) does not meet most definitions of a replication. Foody et al. (2013) argue there is conceptual overlap between the two intervention conditions, although they use different terminology and content. The designs differ substantially: Foody et al. (2013) is a lab based analogue study on adults that uses inferential statistics only, whereas Luciano et al. (2011)