Reliability of a Domain-General Adaptation of the Downs-Black (1998) Tool

Anthony Cruz

12/8/2021

# Methods

## Pre-Registration

The design and pre-analytic plan for this study were pre-specified in an open-ended registration uploaded to Open Science Framework in October 2021 (<https://osf.io/8ycj9>).

## Sample

The sample size for this study was chosen using an a priori power analysis in G\*Power 3.1.9.7 (Faul et al., 2007). With power=.95, false positive rate α=.05, and effect size |ρ|=.77 (from Downs & Black, 1998), the projected sample size needed for this study was N=12. No dropouts or incomplete data were anticipated due to the nature of this study, so we maintained a final sample size of N=12. 12 empirical research articles (published between 2006 and 2018) were chosen arbitrarily from a set of articles that advanced past title & abstract screening for a related systematic review. The goal of this related review is to catalogue the extant research examining neural evidence of changes in stimulus representation that result from category learning. Each paper was not fully screened and may not be included in the final review, but it can be assumed that their contents are related to the topic of the review.

## Survey Design

Our modification of the Downs-Black tool is homologous to the original Downs-Black tool in terms of scoring and subscales. Each checklist item receives a score and the total score can fall between 0 and 31. The total score will be referred to as the Quality Index. This checklist includes the following subscales: Reporting (items 1-10), External Validity (items 11-13), Internal Validity - Bias (Items 14-20), Internal Validity - Confounding (Items 21-26), Validity (items 11-26), and Power (item 27). At the end of the checklist, each Rater is asked to provide a subjective rating, ranging from 0 to 10, of the overall quality of the study. This measure will be referred to as the Global Score. Several items were reworded to support use outside of the clinical domain. Of the 27 survey items, 19 were modified. The final survey was uploaded to Qualtrics. Responses were required for all items, barring an optional question asking if Raters have any questions or comments related to the assessment. Raters completed the assessment independently and remotely for each study in no particular order.

## Test Procedure

Two reviewers assessed each study. These reviewers were familiar with the review protocol and assisted in title & abstract screening in the related systematic review. However, they were not involved in the development of this checklist or the development of the review protocol. They were also not explicitly instructed that they were piloting this tool for a study on reliability. Rather, they were instructed that we were assessing the feasibility of conducting risk of bias assessment on Qualtrics. Data collection ran from the 25th of June 2021 to the 5th of November 2021.

# Results

## Inter-Rater Reliability

### Level 3

#### Level 4.

##### Level 5.

## Internal Consistency

# Discussion