

S3 Appendix: Additional reproducibility domain for quality assessment

- (1) Was the time interval between the repeated tests appropriate? [Yes/No/Unclear]
- (2) Were the test conditions similar for the repeated tests (type of administration, environment, instructions)? [Yes/No/Unclear]
- (3) Was a Kappa score calculated? [Yes/No/Unclear].

Like the risk of bias questions in QUADAS-2, the answer to the question “Could the reproducibility data be biased?” [Low risk/High risk/Unclear risk] was based on the answers to the three reproducibility criteria questions.

Based on the work of Mookink et al. (2018) (28).