META-ANALYSIS STUDY PROTOCOL

The Reliability of the Serial Reaction Time task: Meta-Analysis of Test–Retest Correlations

**Contributors**:

Cátia M. Oliveira*[1]

Marianna E. Hayiou-Thomas[1]

Lisa. M. Henderson[1]

[1]Department of Psychology, University of York

*address correspondence to cmfo500@york.ac.uk

## BACKGROUND

A dominant causal theory of developmental language disorder and dyslexia proposes an impairment in the procedural memory system (Ullman & Pierpont, 2005; Ullman et al., 2020). A large body of research over the last decade has used the Serial Reaction Time (SRT) task to test this "procedural deficit hypothesis"; however, findings have been inconsistent, and the reliability of this task has been questioned. Whilst the SRT task shows robust procedural learning effects (i.e., faster and more accurate responses to probable versus improbable trials), the stability of this effect across sessions has been shown to be suboptimal ($r < .70$; Burlingame et al., 1995).

Here, we present a meta-analysis that examines the reliability of the SRT task with the aim of determining the strength of the test-retest reliability for this task and potential participant-related (e.g. age) and methodological factors (e.g. interval between test and retest) that may affect its reliability.

## STUDY OBJECTIVES

Herein, we describe a protocol for a meta-analysis of studies investigating the test-retest reliability of the SRT task. This investigation aims to assess the stability of the SRT task, whilst considering possible moderating variables (e.g., age, length of interval between sessions).

Based on previous literature, we address the following research questions and hypotheses:

1. Suboptimal test-retest reliability will be expected ($r < .70$)
    a. Test-retest reliability will be lower for children than adults (West et al., 2020);
    b. Test-retest reliability is expected to be higher for studies with shorter testing intervals (Calamia et al., 2013);
    c. Measures that take into account individuals' speed (e.g. ratio, regression slopes) are expected to have higher test-retest reliability than those which do not (e.g. difference scores) (West et al., 2020).

Exploratory:

1. Does the number of trials moderate the test-retest reliability of the SRT task?
2. Does learning the same or different sequences moderate the test-retest reliability?

3. Does the SRT task variant (e.g. deterministic vs probabilistic sequences) moderate the test-retest reliability (Stark-Inbar et al., 2017)?

## ELIGIBILITY CRITERIA

### INCLUSION CRITERIA

- Used a strictly visual deterministic, probabilistic or alternating SRT task with procedural learning computed as the difference between sequenced/probable and random/improbable trials;
- Reported Pearson's correlation (or equivalent) coefficients between procedural learning at least at two-time points;
- If the same results are published in multiple articles, these will only be reported once in the meta-analysis;
- Language of publication: English;

### EXCLUSION CRITERIA

- Studies that used adaptations that considerably deviate from the task proposed by (Nissen & Bullemer, 1987)
- Dual task paradigms;
- Studies with active interventions or studying populations whose performance is expected to change over time (e.g. stroke patients);

## SEARCH STRATEGY

We will perform our search in Medline, PsycINFO and Embase, as well as the BASE - Bielefeld Academic Search Engine for grey literature. Citation searching will also be conducted to ensure that all relevant papers are identified.

The following search strategy will be used:

1. Procedural learning
2. Procedural memory

3. Sequence learning
4. Implicit learning
5. Statistical learning
6. Procedural knowledge.sh.
7. Serial Reaction Time task (.tw for PsycINFO)
8. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7
9. Reliab*
10. Consistency
11. Stab*
12. Individual differences
13. Valid*
14. Psychometr*
15. Measurement
16. 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16
17. 7 AND 16

## SELECTION OF STUDIES

One reviewer will independently screen all articles and identify those who are relevant to include in the review. This screening will be done at the title and abstract level. At the full-article level the list of papers will be screened by all reviewers to determine whether they fit the criteria. Discrepancies in article inclusion will be resolved by discussion between the three reviewers. Once the list of full articles is agreed upon, the first reviewer will code the data using the table presented in the appendix.

The prism flow diagram (Page et al., 2021) in figure 1 will be used to track the number of records identified, included and excluded; as well as the reasons for exclusions.

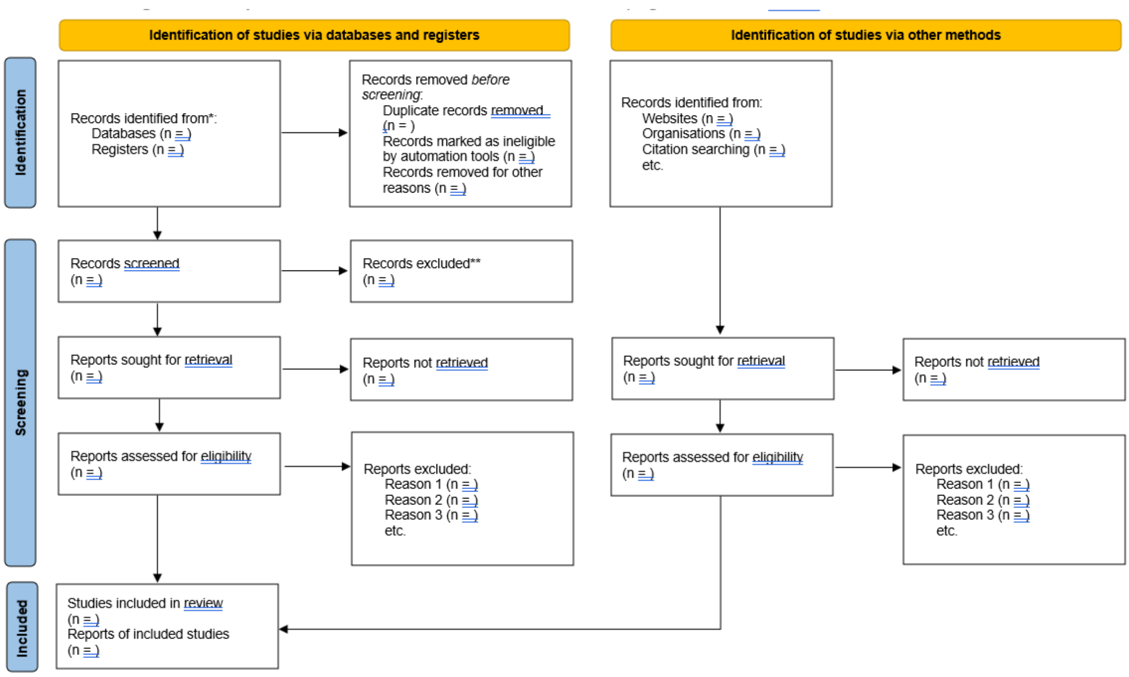


Figure 1. PRISMA flow diagram for study inclusion (Page et al., 2021)

## DATA EXTRACTION

Meta-analysis

Articles included in the meta-analysis will be coded for the items in appendix 1. The following information from each study will be recorded:

a) Authors;
b) Year;
c) Study ID;
d) Number of participants
e) Age of participants;
f) Correlation;
g) Measure of procedural learning;
h) Group;
i) SRT type;
j) Sequence complexity;
k) Number of trials;
l) Interval between test and retest sessions;

## ANALYSIS

All continuous moderators will be centred on their mean. Centering moderators does not change the regression coefficients. All correlation coefficients will be converted from Pearson's r to Fisher's z scale as Pearson's r is not normally distributed (Moeyaert et al., 2017; Quintana, 2015).

The *metafor* package (Viechtbauer, 2010) will be used for model fitting. Appropriate syntax will be used following the guidelines from Pustejovsky and Tipton (2021) for model specification to reflect the levels of dependency in the dataset. If teams contribute multiple correlation coefficients for the meta-analysis, to deal with the lack of independence across effect sizes and avoid reducing power by calculating the average effect sizes for these studies, multilevel random/mixed effects models will be estimated using the function *rma.mv()* from the *metafor* package (Viechtbauer, 2010). Alongside, if necessary, RVE standard errors, hypothesis tests, and confidence intervals for the meta-regression coefficient estimates will be computed using the functions *coef_test()* or *conf_int()* from the *clubSandwich* package (Pustejovsky, 2021) to avoid model misspecification (Pustejovsky & Tipton, 2021).

### Modelling

A random-effects model will be initially run to estimate the overall test-retest reliability (i.e., the average correlation coefficient) of the SRT task. Following the intercept-only model, separate meta-regression models will be performed for each mediator variable (e.g., age, interval between sessions) to determine which factors influence the test-retest reliability of the SRT task. After performing the meta-analytic calculations, Fisher's *z* will be converted back to Pearson's *r* for reporting the average correlation and 95% CI for each model.

### Bias analysis

Cook's distances and standardised residuals will be used to identify potential outliers and influential cases(Viechtbauer, 2010; 2011). To detect evidence of bias, funnel plots, contour-enhanced funnel plots, Egger's regression test will be analysed (Viechtbauer, 2021).

Study heterogeneity will be analysed using the Q statistics, I2 and $\tau 2$ (Higgins & Thompson, 2002;  Viechtbauer, 2021).

## QUALITY CONTROL AND QUALITY ASSURANCE

All article screening will be done by one reviewer, yet discrepancies will be resolved amongst the three authors.

Conflicts of Interest

There are no relevant competing interests.

Upload date

This protocol was uploaded to osf.io on the 30th of June, 2021.

## REFERENCES


Burlingame, G. M., Lambert, M. J., Reisinger, C. W., Neff, W. M., & Mosier, J. (1995). Pragmatics of tracking mental health outcomes in a managed care setting. *The Journal of Mental Health Administration*, *22*(3), 226–236. https://doi.org/10.1007/BF02521118

Calamia, M., Markon, K., & Tranel, D. (2013). The robust reliability of neuropsychological measures: Meta-analyses of test-retest correlations. *Clinical Neuropsychologist*, *27*(7), 1077–1105. https://doi.org/10.1080/13854046.2013.809795

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

James Pustejovsky (2021). clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. R package version 0.5.3. https://CRAN.R-project.org/package=clubSandwich

Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, *20*(6), 559–572. https://doi.org/10.1080/13645579.2016.1252189

Nissenandpetekbullemer, M. J. (1987). Nissen, Bullemer - 1987 - Attention Requirements of Learning Evidence from Performance Measures.pdf. *Cognitive Psychology*, *19*. https://ac.els-cdn.com/0010028587900028/1-s2.0-0010028587900028-main.pdf?_tid=bec66475-4a81-467d-852a-1417270cd733&acdnat=1543287406_5f864c76f3d75cceef7b9aa67f51ada2

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *372*. https://doi.org/10.1136/bmj.n71

Pustejovsky, J. E., & Tipton, E. (2021). Meta - analysis with Robust Variance Estimation : Expanding the Range of Working Models. *Prevention Science*, *0123456789*. https://doi.org/10.1007/s11121-021-01246-3

Quintana, D. S. (2015). From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in Psychology*, *6*(OCT), 1–9. https://doi.org/10.3389/fpsyg.2015.01549

Stark-Inbar, A., Raza, M., Taylor, J. A., & Ivry, R. B. (2017). Individual differences in implicit motor learning: Task specificity in sensorimotor adaptation and sequence learning. *Journal of Neurophysiology*, *117*(1), 412–428. https://doi.org/10.1152/jn.01141.2015

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The Neurocognition of Developmental Disorders of Language. *Annual Review of Psychology*, *71*(1), 389–417. https://doi.org/10.1146/annurev-psych-122216-011555

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. In *Cortex*. https://doi.org/10.1016/S0010-9452(08)70276-4

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor. *Journal of Statistical Software*, *36*(3), 1–48.

Viechtbauer, W. (2021). Model checking in meta-analysis. In C. H. Schmid, T. Stijnen, & I. R. White (Eds.), *Handbook of meta-analysis* (pp. 219-254). Boca Raton, FL: CRC Press.

West, G., Shanks, D. R., & Hulme, C. (2020). Sustained Attention, Not Procedural Learning, is a Predictor of Reading, Language and Arithmetic Skills in Children. *Scientific Studies of Reading*, *00*(00), 1–17. https://doi.org/10.1080/10888438.2020.1750618

**APPENDIX**

| Authors | Study ID | Year | N | Age | Cor | Measure | Group | SRT_type | Seq_complexity | Trial_no | Interval |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

**Authors** represents the citation of the study – authors;

**Year** represents the citation of the study – year;

**Study ID** refers to a random code given to each study;

**N** indicates the number of participants for each correlation;

**Age** refers to the age of participants;

**Cor** provides the Pearson correlation between language/literacy measure and sequence learning;

**Measure** refers to the measure of procedural learning (0 – does not account for individual speed (e.g. difference scores / 1 – accounts for individual speed e.g. regression slopes, ratio scores);

**Group** indicates which sample is being analysed – TD, DD or DLD groups;

**SRT_type** includes which SRT task was used – deterministic, probabilistic or alternating;

**Seq_complexity** refers to the type of sequence used – first (FOC) or second (SOC) order conditional;

**Trial_no** indicates the number of trials the participants was exposed to by the time of the correlational analysis;

**Interval** indicates the interval between test and retest sessions;