Opening up the science: Methodological transparency and reproducibility in second language acquisition research

Emma Marsden
University of York, UK
Towards Methodological Transparency & Reproducibility in SLA research

1. Open science & the IRIS digital repository of materials
2. Relationship between transparency and quality
3. Replication & the importance of transparency
4. Challenges and Recommendations
Publicly funded research **should** be made freely available

**OPEN PUBLICATION**
- cited more

**OPEN DATA**
- Corpora of oral production data, L1 and L2: FLLOC, SPLLOC, CHILDES
- Trønso
- CLARIN
- increased citations

- stronger evidence and better statistical reporting


Piwowar & Vision (2013)

OPEN METHODS

Making materials available

Materials = Data collection tools, instruments, stimuli, scoring and coding procedures, analysis protocols
• Some problems:

  Researchers create and keep own instruments
  -> re-invention of wheel, poor systematicity of research

  Maintenance and access to instruments is *ad hoc*

  Publications just have brief descriptions with short samples

  What a contrast with full instrument!
How much of the methods do we see in journals?

(17) Experimental: *[-past] shì ... de (4 tokens)

*Xiaowang shi mingtiān chī-wán nà gè dāngāo de, bù shì jīntiān.
Xiaowang COP tomorrow cat-up that CL cake DE not COP today
Intended: It is tomorrow that Xiaowang (is going to) eat up that cake, not today.

(18) Control A: [+past] shì ... de (4 tokens)

Xiaowang shi zuótiān chī-wán nà gè dāngāo de, bù shì jīntiān.
Xiaowang COP yesterday cat-up that CL cake DE not COP today
It was yesterday that Xiaowang ate up that cake, not today.

(19) Control B: [-past] canonical (4 tokens)

Xiaowang dāsuàn mingtiān chī-wán nà gè dāngāo.
Xiaowang intend tomorrow cat-up that CL cake
Xiaowang intends to eat up that cake tomorrow.

What is IRIS?

Instruments for Research Into Second Languages

- A sustainable digital repository
- 2200+ materials used to collect data:
  - e.g. questionnaires, grammaticality judgment tests, observation & interview schedules, word lists, sound & video files, language tests, pictures, teaching materials, software scripts,
  - And ANALYSIS PROTOCOLS and DATA

- Downloadable
- Uploadable
- Searchable

Scope of content

as wide as the field of L2 research...

Motivation, identity, & strategies

Language learning

... with diverse research aims

... in diverse contexts

... with diverse types of data

L2 perception & processing

Bilingualism

Phonology
Grammar
Vocabulary
Pragmatics

Language teaching, & policy

Part 1: IRIS: Transparency of MET
Marsden, Mackey & Pl...
Rationale behind IRIS

- Materials can be adapted to suit different contexts, learners, languages
- Promotes transparency
- Easier to evaluate quality and reliability
- Quality assurance: only peer-reviewed publications
- Stimulates & facilitates replication
• 2400 files -> **over 1100 data collection instruments**
  • new materials contributed almost daily
• From 1300 researchers
• Searchable across 330+ parameters,
  ✓ research area
  ✓ type of instrument
  ✓ language feature
  ✓ L1, L2
  ✓ participant characteristics: age, proficiency
  ✓ author
• Materials qualify for IRIS if used for:
  *peer-reviewed publications*
  *or PhDs*
• Submission to IRIS supported by 37 journals
  • In acceptance letters from editors to authors, and in author guidelines

• Upload to IRIS in publication guidelines of the American Association of AL

• Formally endorsed by British Association of AL

• Only venue in AL that qualifies articles for Center for Open Science badges

• 16,500+ downloads of research materials

• Cited in handbooks, methods books, position pieces & syntheses

• Used in many research methods training courses
Materials used for ‘year abroad’ research?
Part 2: Open methods to improve rigour and replicability...

The “methodological turn” in applied linguistics

“methodological issues ...demand a kind of professional scrutiny that goes directly to the core of what we do and what we know...”

Byrnes, 2013, p. 825

“Methodological practices and study quality need to be measured, not assumed”

Plonsky, 2013
Part 2: Open methods to improve rigour and replicability

Benefits of transparent instrument design

Part 2a: Grammaticality judgment tests

Indicate how acceptable the following sentences are

1) I don’t know what is he eating for dinner tonight.
2) I don’t know what he is eating for dinner tonight.
Frequent use of judgment tests because *assumed* to be...

Easy to...
- develop
- administer
- score

Common conventions
→ greater comparability across studies
JTs are controversial

Tests of *explicit* or *implicit* knowledge?
(e.g., Ellis, 2005; Gutierrez, 2013; Vafaee et al., 2016)

Theoretical perspectives on linguistic knowledge-> methodological decisions

- Do instructions say ‘how acceptable...’ or ‘how correct...’ ?
- Scaled vs. dichotomous response
- Oral vs. written
- ...
Methodological synthesis of JTs

How do we design, administer and report on JTs?

$K = 299$ studies

382 JTs

Total sample size in our synthesis = 24,679!
Transparency?

88% =
306 instruments unavailable for open scrutiny.
*****
252 not available for replication, even if you have a journal subscription

A ‘Special Collection of JT’s’ on IRIS is improving this!
Consequences of lack of availability

- Median $N$ whole study sample:
  - L2 learners = 47
  - NS controls = 20
  - Overall study = 60
  - Field median = 60 (Plonsky, 2013)

- L1s
  - 45% = English
  - 15% = Chinese
  - 9% = Japanese
  - 8% = French

- TLs
  - English = 59%
  - Spanish = 17%
  - French = 10%

- 77% one-shot design -> non-developmental, not within-subject over time

Imagine the ‘n’ if JTs available across the globe?

Think of the L1 – L2 combinations!

Existing JTs used to investigate change over time, e.g. after teaching or year abroad
Part 2a: Instrument transparency: JTs. Plonsky et al. (in progress)

Transparent design -> clearer operationalisation of knowledge type.

-> the reader has to infer what kind of knowledge the researchers elicited ... without full reporting of design features, without seeing the instrument!
Consequences of lack of full transparency: **Construct validity**

*time pressure matters.*

52% we don’t know whether timed or not

*Modality matters.*

18% we don’t know whether written or aural

*Breadth of construct - types and tokens of linguistic feature tested?*

51% we don’t know if there was only one ‘version’

If one version, narrows breadth of construct being elicited
Part 2a: Instrument transparency: JTs.  *Plonsky, Marsden, Gass, Spinner, Crowther (in progress)*

Consequences of lack of transparency:

**Reliability**

- Reliability coefficients can be given alongside instruments on IRIS.
- New coefficients posted every time that JT used?
- Mean reliability = 0.80 (=field average, Plonsky & Derrick, 2016)
- But reliability calculated on whole instrument

**Part 2b: Instrument transparency:**

**Self-paced reading**

**Similar aim to JTs**

Eliciting sensitivity to grammatical structure and norms

But, SPRs identify **precise point** of difficulty *during* parsing,

without an explicit judgment
Self-paced reading: an example

I

Part 2b: Instrument transparency:

Try to understand this sentence. Click for each word to appear. You will be asked a question at the end!

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) Can we train processing routines for abstract syntax
Self-paced reading: an example

don’t

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) *Can we train processing routines for abstract syntax*
Self-paced reading: an example

know

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) Can we train processing routines for abstract syntax.
Self-paced reading: an example

what

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) Can we train processing routines for abstract syntax
Self-paced reading: an example

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) *Can we train processing routines for abstract syntax?*
Self-paced reading: an example

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) *Can we train processing routines for abstract syntax?*
Self-paced reading: an example

having

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) Can we train processing routines for abstract syntax
Self-paced reading: an example

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) Can we train processing routines for abstract syntax?
Self-paced reading: an example

dinner

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) Can we train processing routines for abstract syntax?
Self-paced reading: an example

Part 2b: Instrument transparency:

Stimuli taken from Marsden, Cruickshank, Roberts (in progress) *Can we train processing routines for abstract syntax*
Which picture best matches the sentence:

The word ‘dinner’ - after the critical region

A

B
Transparency of instrument design: SPR

Methodological synthesis

63 studies
reporting a total of 71 SPR tests
### Transparency and reproducibility of SPR tests in SLA research

#### Part 2b: Instrument transparency: SPRs.

*Marsden, Thompson, Plonsky (submitted)*

<table>
<thead>
<tr>
<th>Availability</th>
<th>No. of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full stimulus on IRIS</td>
<td>3</td>
</tr>
<tr>
<td>Another place (author’s website)</td>
<td>3</td>
</tr>
<tr>
<td>Example in article and on IRIS</td>
<td>4</td>
</tr>
<tr>
<td>Full - all items - in article</td>
<td>16</td>
</tr>
<tr>
<td>Just examples in article</td>
<td>28</td>
</tr>
<tr>
<td>Not available at all</td>
<td>16</td>
</tr>
</tbody>
</table>

*96% of full materials not yet available on permanent, open repository*

*A ‘Special Collection of SPRs’ on IRIS is improving this!*
### Consequences of poor availability: *Agenda limiting - L2s?*

<table>
<thead>
<tr>
<th>Language</th>
<th>No. studies: L1</th>
<th>No. studies: target language</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>19</td>
<td>43</td>
</tr>
<tr>
<td>Chinese</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Spanish</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Greek</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>German</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Korean</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Dutch</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Japanese</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>French</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Russian</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Arabic</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Multiple languages</strong></td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

Consequence of poor availability:
agenda limiting: Cross-linguistic influence?

We know the L1 matters for online processing

52 / 71 with learners with one L1

14 compared across different L1s
Consequence of poor availability:
Agenda limiting. Who are the participants?

<table>
<thead>
<tr>
<th>Beginner</th>
<th>Interim</th>
<th>Advanced</th>
<th>Near Native</th>
<th>Bilingual</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>18</td>
<td>50</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>

56 / 71 = university students
- Higher meta-linguistic knowledge could affect reading times
  
  (Keating & Jergusker, 2015)

Part 2b: Instrument transparency: **SPRs.** Marsden, Thompson, Plonsky (submitted)
Consequence of poor availability:
Agenda limiting. Why ‘successful’ comprehension only?

To date, research has focused on advanced learners or bilinguals only analysing ‘correctly comprehended’ sentences. But learners process language even when comprehension is poor! (though see Keating, Jegerski & VanPatten, 2016; Xu, 2014)

Existing SPRs could be used with other learner populations

Part 2b: Instrument transparency: SPRs. Marsden, Thompson, Plonsky (submitted)
Consequence of poor availability:
Agenda limiting. Investigating learning over time?

Growing interest in the role of processing (Altmann & Kamide, Dussias, Gruter, Lew-Williams & Fernald, Kaan, O'Grady, Pickering & Garod, Chang et al., 2006; Elman, 1990; Seidenberg & MacDonald, 1999; Special issue of LinguisFc Approaches to Bilingualism, 5, 4).

Yet, SPR not yet used to investigate development of processing:
76% one shot studies
24% cross-sectional studies, with proficiency as a variable
One a within-subject, longitudinal design

Consequences of lack of transparency:

**Construct validity - Comprehension of what?**

Central tenant of SPR: processing *during reading comprehension*

Comprehension question used at end of trials

Question should not focus on "critical region"

56 / 71 used CQs

BUT what do these CQs focus on?

17 – no example

34 - *one* example

Only 5 studies provided multiple examples of comprehension questions

Part 2b: Instrument transparency: **SPRs.** Marsden, Thompson, Plonsky (submitted)

One example of a CQ does not tell us where attention is repeatedly focussed during reading
Consequences of lack of full transparency: Construct validity

Stimuli design ($k$ items, word frequency, length)

Length matters: affects computational ease (Pienemann & Kessler, 2011).

Yet...

- 25 / 71 length of sentence not specified
- 5 / 71 reported number of syllables per word
- 19 / 29 using segments: number of segments not specified

Part 2b: Instrument transparency: SPRs. Marsden, Thompson, Plonsky (submitted)

A bank of stimuli helps instrument design and research training
Part 2b: Instrument transparency: SPRs. Marsden, Thompson, Plonsky (submitted)

Consequence of lack of transparency:

Data cleaning and analysis protocols

Outlier removal?
- 12 did not report any outlier removal.
- 44 used participant RTs.
- 15 used both participant RTs and item RTs.

Outlier cut-off?
- 14 used 2.5 SD.
- 5 used 200ms-2000ms (as recommended by Jegerski & Keang, but for an advanced L2 reader).
- Wide variety of ranges 100ms – 25000ms.

Analyse all data or just correct responses?
- 7/56 studies analysed all
- 28/56 analysed only correct responses
- 21/56 didn’t report

Having analysis protocols available would make analyses more systematic and reproducible.
Summary of Part 2:
Greater transparency and availability should help:

- Agenda setting and research aims (cross-sectional, L1-L2 influence)
- Scope and power of studies (different proficiencies, n)
- Operationalise our constructs (design: items, coming, stimuli)
- Comparability across studies (data cleaning & analysis)

‘Better reporting’ unlikely to fully address all these problems
( across all journals & techniques...)

The actual materials necessary

More transparency and systematicity of methods ->
Benefits for all types of validity & reliability
Part 3: Replication research in SLA

“Conducting a research study again, in a way that is either identical to the original procedure or with small changes (e.g., different participants), to test the original findings” (Mackey & Gass, 2005: 364).

“essential ... support for theory” (Porte, 2012)

Strong replication movement in Psychology

• “Many Labs” & Reproducibility projects
• Pre-registration of materials & analyses
Commentaries and calls for replication

46 published commentaries & calls for replication in L2 research

Santos 1989... Polio & Gass 1997... Porte 2012... Vandergirft & Cross 2017

23 + from other disciplines: Psychology, Education, Sociology, Business, Marketing, Organisation Science

Replications... lacking prestige, originality, or excitement

(Makel et al 2012, citing Lyndsay & Ehrenberg, 1993; Neuliep et al 2014)

Other syntheses of replications:
for Psychology: Makel, Plucker & Hegarty (2012)
for Education research: Makel & Plucker (2014)
How long does it take for a study to be replicated?

- mean 6.64 years (sd 6.16)
- range 0-37 years
Proportion of replication in journals that published most replications across the 26 journals that have published replications = 0.26% 

<table>
<thead>
<tr>
<th></th>
<th>SSLA</th>
<th>The MLJ</th>
<th>FLA</th>
<th>LL</th>
<th>AP</th>
<th>Total articles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TOTAL articles</strong></td>
<td>562</td>
<td>1009</td>
<td>1528</td>
<td>855</td>
<td>1030</td>
<td>1528</td>
</tr>
<tr>
<td><strong>k replications</strong></td>
<td>13</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>% of total</strong></td>
<td>2.31%</td>
<td>0.79%</td>
<td>0.33%</td>
<td>0.58%</td>
<td>0.39%</td>
<td>0.70%</td>
</tr>
</tbody>
</table>

1973 – 2015, last complete year before synthesis

Compared to other disciplines?

- 0.07% Psychology, but NB: pre-replication-boom & top 100 journals
- 0% - 3% Business, Marketing, Communication
- 0.13% Education (top 100, 1938-2014)
What do we replicate? Study design and findings

Participants? The WEIRDest Western Educated Industrialized Rich Democratic (Mishra et al. 2012) … speaking (1/4) or learning (1/2) English

‘ages’ and ‘proficiency’ not reported (67%, 37% studies) (Thomas 1994) -> replicability??

Of replications whose initial findings were…

<table>
<thead>
<tr>
<th>Null</th>
<th>Null but trend</th>
<th>Stat sig differences</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>83</td>
<td>7</td>
</tr>
</tbody>
</table>

need to replicate null findings too

When power (n) is low, “null findings” ≠ “no effect”

How much do we change when we replicate?

% studies with changes between I and R...

<table>
<thead>
<tr>
<th>Number of changes</th>
<th>changes = reasons for replication</th>
<th>% studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Mean per study (st dev.)</td>
<td>1.34 (1.31)</td>
<td></td>
</tr>
</tbody>
</table>
How much do we change when we replicate?

% studies with changes between I and R…

<table>
<thead>
<tr>
<th>Number of changes</th>
<th>changes = reasons for replication</th>
<th>changes acknowledged, but not reason for replication % studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>33</td>
<td>45</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Mean per study (st dev.)</td>
<td>1.34 (1.31)</td>
<td>0.91 (1.04)</td>
</tr>
</tbody>
</table>
How much do we change when we replicate?

% studies with changes between I and R:

<table>
<thead>
<tr>
<th>Number of changes</th>
<th>s = reasons for replication</th>
<th>s acknowledged, but not reason for replication</th>
<th>changes not acknowledged by authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mean per study (sd)</td>
<td>1.34 (1.31)</td>
<td>0.91 (1.04)</td>
<td>1 (1.18)</td>
</tr>
</tbody>
</table>

no relationship to what the replication calls itself (e.g., ‘partial replication,’ just ‘replication’, extension, conceptual)
Do “authorship overlaps” relate to whether findings are supportive of initial study?

% “supportive of Initial”, as function of authorship independence

<table>
<thead>
<tr>
<th>Author overlap? (% total R studies)</th>
<th>Findings in relation to initial study (% of replication studies)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% not or partially not supportive</td>
<td>% partially or very supportive</td>
</tr>
<tr>
<td>Overall</td>
<td>28</td>
<td>68</td>
</tr>
</tbody>
</table>

None (69%) 37 58
Some/all (31%) 10 91
How do relations between authorship teams relate to the nature of the conclusions?

**Author overlap?**

<table>
<thead>
<tr>
<th>% not supportive</th>
<th>% partially supportive</th>
<th>% very supportive</th>
<th>% not reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K studies total)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No overlap</td>
<td>20</td>
<td>17</td>
<td>30</td>
</tr>
<tr>
<td>Some overlap</td>
<td>5</td>
<td>5</td>
<td>43</td>
</tr>
<tr>
<td>Overall</td>
<td>15</td>
<td>13</td>
<td>34</td>
</tr>
</tbody>
</table>

**Findings in relation to initial study (% of studies):**

**Psychology:**
- 2% supportive if no overlap
- 5% supportive if some overlap

**Education:**
- 9% supportive with overlap, in same publication
- 1% with overlap, in new publication
- 4% when no author overlap

---

Questionable Research Practices (bias, p-hacking)

Or

Materials availability and fidelity to initial study??
How do studies use the *findings* and the *data* of the Initial study?

Only 6% of Replications provided Initial study’s effect size.

Only 6% of Replications used Initial’s raw data in a new analysis

---

**Data Sharing!**

Open data is associated with strength of evidence and quality of reporting

---

How do Replicators compare their findings to Initial Study?

<table>
<thead>
<tr>
<th>% of Replication Studies that compared their findings to Initial Study using...</th>
</tr>
</thead>
<tbody>
<tr>
<td>% narrative Comparisons</td>
</tr>
<tr>
<td>93</td>
</tr>
</tbody>
</table>
How do Replicators compare their findings to Initial Study?

Learning from the Reproducibility Project in Psychology:

- **mass coverage:**
  - “in only 36% of the studies were the original results replicated” [because $p > 0.05$]

BUT... 77% of replication effect sizes were within a 95% prediction interval of original effect size

(Patil et al. 2016)
Transparency: How do replicators get hold of materials?

<table>
<thead>
<tr>
<th>Availability in Initial Study</th>
<th>How passed to replicator</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (k = 12)</td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td>parCal examples (k = 29)</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>one full instrument, but not all (k = 26)</td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>Only 4% of Initial studies had materials in IRIS or some other open access</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Only 4% of Initial studies had materials in IRIS or some other open access.
1) Replication: cultural changes in academia

Replication *not* an easy route; it’s an *essential* route

Facilitate student apprenticeship model

“a lot is hidden behind the final [published article].”
Roxana, a replicator, SLA grad student/trainee teacher from Vasquez & Harvey 2010 p.436

"Ralph is doing a preliminary study of re-inventing the wheel."
Effects of publication bias on replication effort

Only 4/67 tried to replicate a study that had null findings.

Fear that replication won’t ‘replicate findings of original’

How to make ‘null findings’ more publishable?

Transparent replication e.g. via pre-registration should help

Methods fully reviewed and approved -> Principle Acceptance IPA

Researchers cannot ‘argue out’ on basis of methodological flaw

Perspectives in Psychological Science, Cortex, Journal of Child Language

Part 3: Replication
Part 3: Example of a multi-site replication

Morgan-Short, Marsden, Heil, et al. (manuscript)

Effects of attention to form on second language comprehension: A multi-site replication study

Contributors: Kara Morgan-Short, Jeanne Heil, Alaidde Berenice Villanueva Aguilera, Emma Marsden, Zerbrina Valdespino-Hayden, Charlotte Oliver

Date created: 2014-05-14 10:48 AM | Last Updated: 2016-09-12 10:39 AM

Description:
This project has two main aims: (1) To investigate the feasibility of carrying out multi-site replication studies, with a view to promoting and facilitating more such projects in the future. (2) To deepen our understanding about the extent to which learners' attending to language form interferes with comprehension, whilst reading and listening. Aim 1: There have been many calls for increased replication research, including in the field of second language (L2) acquisition. Of particular interest for L2 research are exact replications accompanied by conceptual replications manipulating factors such as first language (L1) or modality. The proposed project will provide a clear protocol and materials for 5 replication studies, some of which will be conducted by researchers who respond to an open call. The challenges of this undertaking will be evaluated in terms of several factors, including buy-in, communication, data collection and analysis, and financial and reporting arrangements. Aim 2: The learning and teaching of grammar and the allocation of attention to form are ongoing concerns to researchers and teachers. The proposed project addresses the issue of simultaneous attention to form and meaning through replications of Leow et al. (2008) and Morgan-Short et al. (2010), which were conceptual replications of the highly cited VanPatten (1990) study. The proposed replications include: (a) an exact replication (in the UK) to further confirm that focusing attention on written forms does not interfere with L2 reading comprehension; (b) quasi-replications with learners with different L1s (Italian, Chinese), which may inform theories of learned attention (Ellis, 2008); and (c) quasi-replications in the aural modality (US, Italy) that will be informative to interpreting differences in results among previous studies.

License: No license

Citation:
http://osf.io/ed9ws

Components

Meeting 12/9/2016 - Emma & Kara
1. Kara ask Jeanne whether would like to talk at SLRF
2. Kara to ask Jeanne to make slides about participant bio details, levels, etc
3. Slide of descriptives with proficiency and check marks
4. Slide descriptives on comprehension, then laid over with statistical significance on comp*condition
5. Kara to play around with bar charts – maybe one slide with 16 bars (4
Part 3: Example of multi-site replication in SLA. Morgan-Short, Marsden, Heil et al.

The study: Is comprehension affected by attending to grammatical or lexical forms?

**EITHER** read a written text (and spot forms):

El pueblo azteca, como pueblo primitivo, podía encontrar una solución a los problemas presentados por las fuerzas de la (La1) naturaleza. Daba mucha importancia a su religión. En ella su Dios principal y todopoderoso (Sol1) era TonaKuh (el sol) (Sol1). Se lo admiró mucho. Tonatiuh tenía las bondades y los defectos humanos, pero con un gran poder sobrenatural. Según la (La2) religión azteca, el sol (Sol2) Tonatiuh necesitaba que lo alimentaran (- n1) con una sustancia...

**OR** hear the same in the oral modality (and spot forms)

**THEN** comprehension measured by 10 multiple choice questions

from Leow et al. (2008)
Participants: Sites by Modality

704 participants across 7 research sites

• 4 sites ran **listening version**
  • University of Illinois at Chicago*
  • University of Oregon
  • Southampton University
  • Kazimierz Wielki University & Adam Mickiewicz University

• 3 sites ran **reading version**
  • University of York*
  • Georgetown University
  • University of South Carolina
Demonstration of the need for replication:

- Clearly replicated findings in 4 sites
- Not clear in 2 sites
- Not replicated in 1 site
Challenges for multi-site replication

1. Seeking collaborators
2. Parity in proficiency at different sites (Thomas, 1994)
3. Compatibility of software (E Prime / superlab scripts)
4. Responsibilities for data entry & analysis
   we provided detailed protocols
Benefits of multi-site replication!

1. N=704

2. Protocols ready to be used again
   Four further replications ongoing:
     1 in China, 1 Native speakers, 1 Heritage learners, 1 after year abroad

3. Having different sites tempered our claims:
   One individual study could have concluded on basis of site with difference.
   Similar findings across 6 sites suggest something special in one ‘odd’ site
   = More reliable reason for generating new hypothesis
CONCLUSIONS:
Challenges for open methods and
Recommendations
Some challenges for open materials

Concern 1: Open materials will encourage bad use of materials

...still need critical thinking about...

- purpose of materials
- use or adaptation of materials
- analysis and interpretation


“I thought I was on to something, but I can’t figure out how to move it.”
Concern 2: Reluctance to share

Others might misuse my materials”
  • Put notes about use on IRIS
  • We gatekeep to reduce bad science: peer review

My next study might be ‘scooped’”
  • the existing study should be open to full scrutiny
  • Our plan won’t be same as others’

I might be proved wrong”
  • Good!
  • but... unlikely to be so clear-cut
  • >citations!

“I don’t have time”
  • 15 minutes of your time vs 3 years of PhD student’s!
  • Sharing magnifies impact of time & public money

“I can’t find my materials or data”
  • All the more reason for IRIS to exist!
Concluding remark (1)

researchers:
• Ethics: sharing with all

• Etic: one study individualism vs synthetic, across contexts and over time
  (Plonsky 2012, Norris & Ortega 2005)

• a collective methodological memory
Concluding remark (2)

to you, future reviewers and editors of journals and books:

Without seeing FULL materials, can reviewers properly evaluate?

Without open access to materials and data, can researchers have properly:

- built on previous methods systematically?
- compared their data to previous data?
- reduced potential bias? (given that independence helps)
Ask your journal editors...

*incentivise and recognise open science*

**Open Science Badges** (from the Center for Open Science)

*Language Learning* adopted the scheme in 2015.

*Studies in Second Language Acquisition*, available from now

*Applied Linguistics*, board approved

*The Modern Language Journal*, board approved

*Linguistic Approaches to Bilingualism*, pending

*Language, Interaction and Acquisition*, pending

- Blohowiak, B., Cohoon, J., de-Wit, L., Farach, F., Hasselman, F., & DeHaven, A. (2016). Badges to acknowledge open practices. osf.io/tyyvx
Effectiveness of journals recognising open materials with badges

Heh guys, have you seen these on www.iris-database.org?

Thank you for listening.
With thanks to:
my collaborators
Susan Gass, Patti Spinner, Dustin Crowther (JT)
Alison Mackey (IRIS)
Ros Mitchell & Florence Myles (Learner language corpora projects)
Kara Morgan-Short, Jeanne Heil, David Abugaber (Replication)
Luke Plonsky (IRIS, JT, SPR, Replication)
Sophie Thompson (SPR, Replication)

Phew! I’m glad I went to www.iris-database.org
Selected references


UK Data Archive MANAGING AND SHARING DATA a best practice guide for researchers 2nd ed
http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/localsites/researchdatamanagement/documents/managingsharing.pdf


Wicherts, J., Bakker, M., Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. PLoS One , 6, 11, e26828