"A systematic review of listening comprehension research designs and measurement tools in SLA research: towards common measurement tools and methodologies?"

Linda Terrier
Domains: English Linguistics and ESP
Scientific Director of *Researching and Teaching Languages for Specific Purposes*
University of Toulouse, France

November 17th, 2017

Fwd: PJ appel à comm Montpellier
Date: Fri, 17 Nov 2017 11:28:20 +0100
Encarnacion Arroyo <earroyo@univ-tlse2.fr>
À: (…) Linda Terrier linda.terrier@univ-tlse2.fr (…)
Bonjour à tous,

Voici un appel à communications pour le 30 novembre (dans 15 jours, je sais...) pour un colloque qui se tiendra afin mai 2018 à Montpellier sur la **méthodologie de la recherche en apprentissage/ didactique des langues étrangères** :

https://amaedmon.wixsite.com/slamethodmontpellier

(…)

En espérant vous rencontrer à Montpellier (aussi bien que dans notre université, bien entendu !),
Bises,
Encarni

[Cogling-L] proposed changes in p-value threshold
> From: Monica Gonzalez-Marquez <
monica.gonzalez-marquez@ifaar.rwth-aachen.de>
> Sent: Tue, 26 Sep 2017 13:19:01 +0200
> To: cogling-l@mailman.ucsd.edu <cogling-l@mailman.ucsd.edu>
> Subject: [Cogling-L] proposed changes in p-value threshold
>
> Hello Everyone,
>
> You may be aware of the **Replication Crisis in science, and of the considerable changes taking place in science practice**. Statistical application, in particular, is being hotly debated largely as a result of widely exposed questionable research practices such as p-hacking, hypothesizing after the fact, etc. If you are not aware of these events, please go here to become better informed,
(…)
> ** Do note that these events affect everyone engaged in science at any level.**

Recherche et pratiques pédagogiques en langues de spécialité
Researching and Teaching Languages for Specific Purposes
Cahiers de l'APLIUT

→ "To what extent have we (researchers in the fields of SLA and SLLT) risen up to the challenge of the Reproducibility crisis?" (focus on LC)

The "reproducibility crisis"

Ioannidis, John P. A. "Why most published research findings are false." *PLoS Med.* Vol. 2, No. 8, e124, 2005.

- Theory that "most published research findings are false" developed thanks to mathematical equations and probabilistic reasoning.

- The complex interactions between
  *overuse of p-values ;
  *different types of bias throughout research procedures
  *competition between research teams when adressing same research questions (rather than collaboration)
  → **make it so that most research findings are false.**

→ **Violation of the falsifiability principle (demarcation criterion of scientific theories) (Popper 1956, p. 9-10)**

The "Reproducibility Project" (2011)

Open Science Collaboration. "Estimating the Reproducibility of Psychological Science." *Science.* Vol. 349, Issue 6251, aac4716, 2015.

- Project: 100 studies published in 2008 in 3 highly-rated Psychology journals to be replicated by 270 contributing authors.

- Key finding: "[…] a large portion of **replications did not reproduce evidence supporting the original results** despite using high-powered designs and original materials when available."

- Two examples of findings:
  *Replication effects were half the magnitude of original effects representing a substantial decline effect;
  *36% of replications had significant results compared to 97% of original studies.

→ Corroborates Ionnadis's hypothesis

→ Violation of the reproducibility criterion (criterion for scientific discoveries) (Popper 1956, p. 23-24)

→ "Science's crisis" (Saltelli & Funtowicz 2017)

A synthesis of the crisis

Munafò, Marcus R. & al. "A manifesto for reproducible science." *Nature Human Behavior*, *Vol. 1, N. 21*, Jan. 10th 2017.

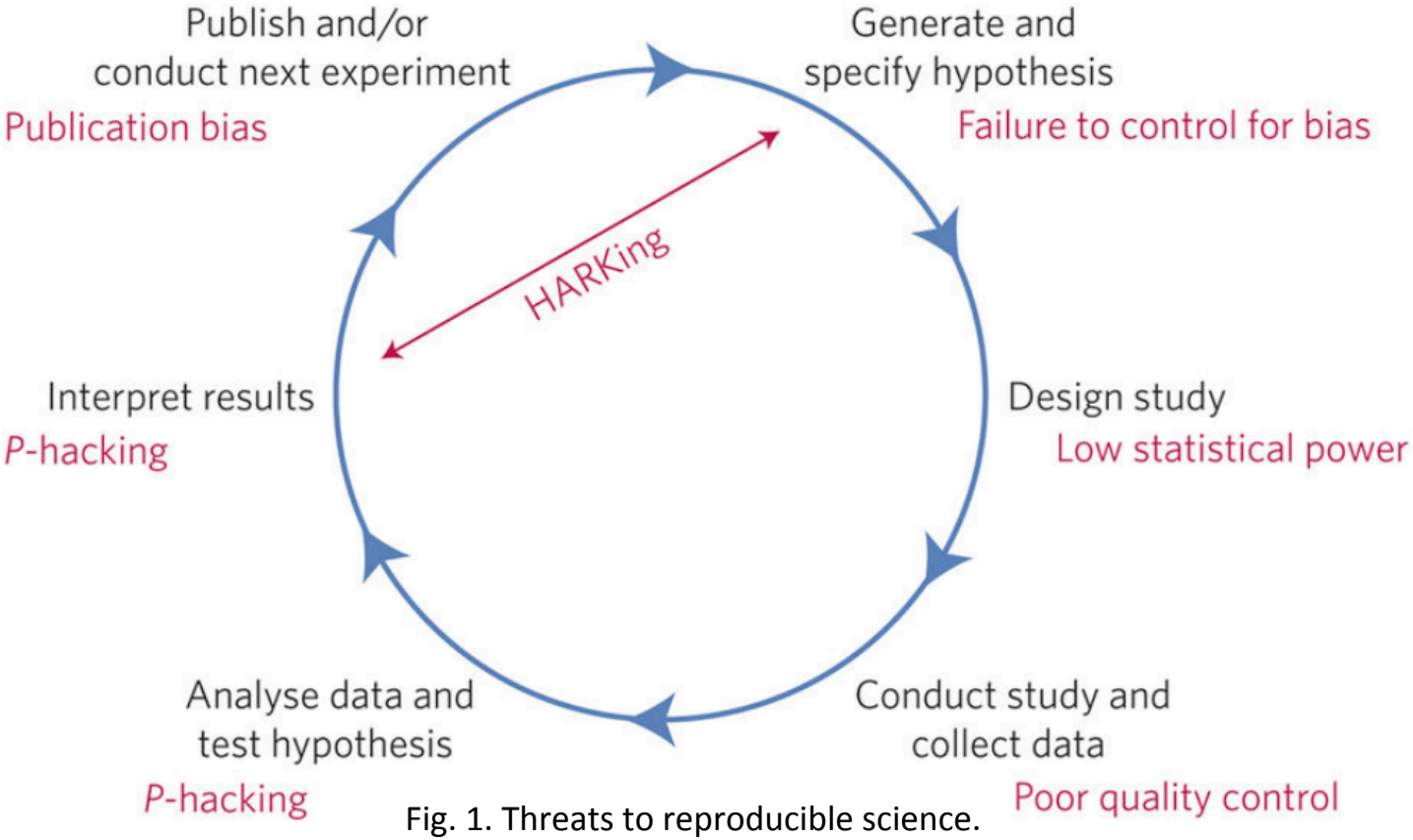


Fig. 1. Threats to reproducible science.
(HARKing=Hypothesing After the Results are Known)

→ Questionable research practices (QRPs) every step of the way

→ General « publish or perish » research environment

→ Insufficient expertise in statistics and data analysis (Plonsky 2013)

Research methodology in the field of second language acquisition and learning

Problem for SLA and SLLT!

→ Systematic reviewing to check for implementation of recommendations suggested to counter the crisis

Main research question:

"To what extent have we (researchers in the field of SLA and SLLT) risen up to the challenge of the Reproducibility crisis?" (focus on listening comprehension)

Recommendations to counter the crisis

Munafò, Marcus R. & al. "A manifesto for reproducible science." *Nature Human Behavior*, *Vol. 1, N. 21*, Jan. 10th 2017.

Table 1 | A manifesto for reproducible science.

Theme	Proposal	Examples of initiatives/potential solutions (extent of current adoption)	Stakeholder(s)
Methods	Protecting against cognitive biases	All of the initiatives listed below (* to ****) Blinding (**)	J, F
	Improving methodological training	Rigorous training in statistics and research methods for future researchers (*) Rigorous continuing education in statistics and methods for researchers (*)	I, F
	Independent methodological support	Involvement of methodologists in research (**) Independent oversight (*)	F
	Collaboration and team science	Multi-site studies/distributed data collection (*) Team-science consortia (*)	I, F
Reporting and dissemination	Promoting study pre-registration	Registered Reports (*) Open Science Framework (*)	J, F
	Improving the quality of reporting	Use of reporting checklists (**) Protocol checklists (*)	J
	Protecting against conflicts of interest	Disclosure of conflicts of interest (***) Exclusion/containment of financial and non-financial conflicts of interest (*)	J
Reproducibility	Encouraging transparency and open science	Open data, materials, software and so on (* to **) Pre-registration (**** for clinical trials, * for other studies)	J, F, R
Evaluation	Diversifying peer review	Preprints (* in biomedical/behavioural sciences, **** in physical sciences) Pre- and post-publication peer review, for example, Publons, PubMed Commons (*)	J
Incentives	Rewarding open and reproducible practices	Badges (*) Registered Reports (*) Transparency and Openness Promotion guidelines (*) Funding replication studies (*) Open science practices in hiring and promotion (*)	J, I, F

Estimated extent of current adoption: *, <5%; **, 5–30%; ***, 30–60%; ****, >60%. Abbreviations for key stakeholders: J, journals/publishers; F, funders; I, institutions; R, regulators.

→ **A manifesto for Open Science**

→ **A set of recommendations for training in statistics and research methods**

Recommendations to counter the crisis

Andrew Gelman is a professor of statistics and political science and director of the Applied Statistics Center at Columbia University.

Recommendations to counter the crisis

"Fixing the reproducibility crisis: Openness, Increasing sample size, and Preregistration ARE NOT ENUF!!!!"
(Gelman, Title of an article from his blog, April 15th 2017)

"There should be **a stronger focus on collecting accurate and relevant data**, and an openness to designs that allow within-person comparisons, even if this represents more effort in data collection." (Gelman 2017 : 20)

Collecting accurate and relevant data when researching LC:

*Using common measurement tools

*Designing protocols that will provide relevant measurements to answer the RQ

*Promoting within-person comparisons (vs. between-person comparisons)

*Blinding instructors to the experimental condition the participants are assigned to

Research methodology in the field of second language acquisition and learning

Sub-question 1:

"Are we implementing Open Science practices?"

Sub-question 2:

"Are we working towards common measurement tools and research methodologies?"

Systematic coding scheme

Table 1 | A manifesto for reproducible science.

Theme	Proposal	Examples of initiatives/potential solutions (extent of current adoption)	Stakeholder(s)
Methods	Protecting against cognitive biases	All of the initiatives listed below (* to ****) Blinding (**)	J, F
	Improving methodological training	Rigorous training in statistics and research methods for future researchers (*) Rigorous continuing education in statistics and methods for researchers (*)	I, F
	Independent methodological support	Involvement of methodologists in research (**) Independent oversight (*)	F
	Collaboration and team science	Multi-site studies/distributed data collection (*) Team-science consortia (*)	I, F
Reporting and dissemination	Promoting study pre-registration	Registered Reports (*) Open Science Framework (*)	J,
	Improving the quality of reporting	Use of reporting checklists (**) Protocol checklists (*)	J
	Protecting against conflicts of interest	Disclosure of conflicts of interest (***) Exclusion/containment of financial and non-financial conflicts of interest (*)	J
Reproducibility	Encouraging transparency and open science	Open data, materials, software and so on (* to **) Pre-registration (**** for clinical trials, * for other studies)	J, F, R
Evaluation	Diversifying peer review	Preprints (* in biomedical/behavioural sciences, **** in physical sciences) Pre- and post-publication peer review, for example, Publons, PubMed Commons (*)	J
Incentives	Rewarding open and reproducible practices	Badges (*) Registered Reports (*) Transparency and Openness Promotion guidelines (*) Funding replication studies (*) Open science practices in hiring and promotion (*)	J, I, F

Estimated extent of current adoption: *, <5%; **, 5–30%; ***, 30–60%; ****, >60%. Abbreviations for key stakeholders: J, journals/publishers; F, funders; I, institutions; R, regulators.

Section 1

Section 2

Section 3

Section 4

Section 0: profile of articles and journals

Section 5: "Designs" (Gelman 2017)
+Accurate measurement tools
+Relevant protocols
+Within-person comparison
+Blinding instructor

→ **Data coding scheme available on demand (could not upload it on Iris depository)**

Selection of the corpus

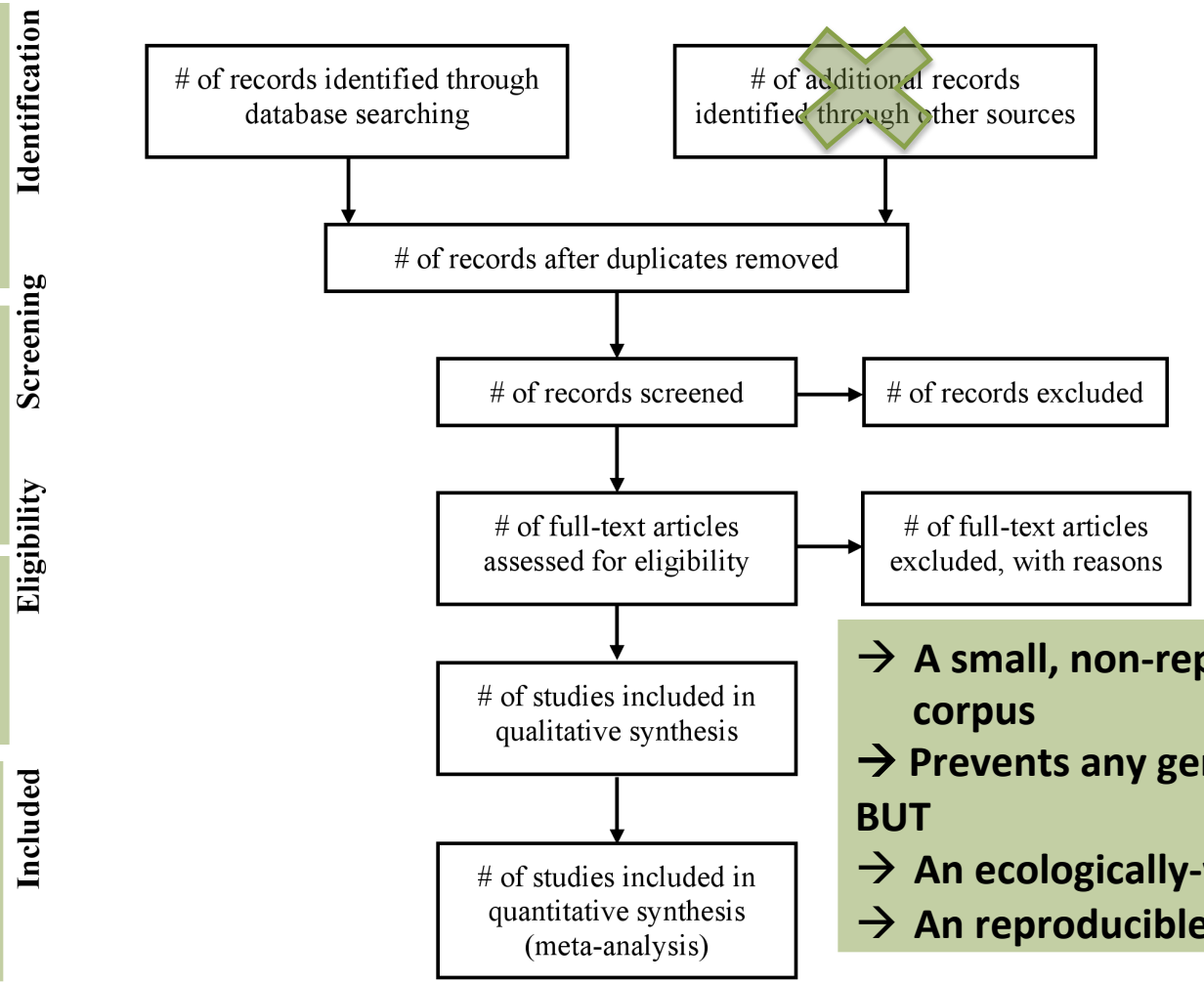
Moher, David & al. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. " *PLoS Med*, Vol. 6, No. 7, e1000097, July 21st 2009.

University database Summon 2.0 ((Title:(listening comprehension)) May 26 2016 to April 26 2018 = 65 records

26 records excluded
Predatory journals (18)
Duplicates (3)
Wrong ref. (2)
Not in English (3)

26 eligible
8 excluded
Library subscription stopped

19 records included in qualitative synthesis
(meta-analysis not relevant for my study)



→ A small, non-representative corpus
→ Prevents any generalization
BUT
→ An ecologically-valid study
→ An reproducible study

Figure 1. Flow of information through the different phases of a systematic review.

→ Raw coded data available on demand.

Final corpus

- **19 articles from 5 continents** - Africa (1), North-America (3), Eastern and Western Asia (8+2), Australia (1), Europe (4)

- **Published in 15 different journals** (3 full Open Access journals; 3 OA journals operating on Article Processing Charge; 8 with subscription which include a few articles in OA/APC)

Advances in Language and Literary Studies (**OA**-3articles); *Applied Psycholinguistics* (**Sub**.); *British Journal of Educational Technology* (**Sub**.); *Cogent Education* (**OA+APC**); *e-FLT: Electronic Journal of Foreign Language Teaching* (**OA**); *Journal of Educational and Social Research* (**OA+APC**); *Journal of English for Academic Purposes* (**Sub**.); *Journal of the National Council of Less Commonly Taught Languages* (**OA**); *Language Teaching* (**Sub**.); *Language Teaching Research* (**Sub**.); *Language Testing* (**Sub**.); *ReCALL : the Journal of EUROCALL* (**Sub**.-2articles); *Revista de Lingüística y Lenguas Aplicadas* (**OA**); *Romanian Journal of English Studies* (**OA+APC**); *TESOL Quarterly* (**Sub**.-2articles)

- **Different types of study:** 13 quantitative studies, 1 qualitative study, 3 mixed-methods studies, 2 research reports with no experiments.

- **Classroom research** (**SLLT** ≠ SLA).

Results – To what extent have recommendations when **performing research ("Methods")** been implemented?

PERFORMING RESEARCH	**Yes**	No	N/A	Total
1.1- Study design based on commonly accepted designs (ie. ref to other existing designs)	**0**	17	2	19
1.3- Blinding data collectors to the experimental condition the participants are assigned to	**1**	16	2	19
1.4- Blinding data analyst to parts of the data	**0**	17	2	19
1.5- Pre-registration of the study	**0**	17	2	19
1.6- Involvement of methodologists in the research	**0**	17	2	19
1.7- Independent oversight of the research	**0**	17	2	19
1.8- Multi-site studies/distributed data collection	**3**	14	2	19
1.9- Team consortium	**0**	19	0	19

= WHAT IS *NOT* IMPLEMENTED (OR VERY LITTLE)

PERFORMING RESEARCH	**Yes**	No	N/A	Total
1.2- Blinding participants to the experimental condition they are assigned to	**13***	0	6	19

= WHAT *IS* IMPLEMENTED… BUT NOT DONE SO WELL

→ 8 recommendations for "methods" out of 9 are <u>not</u> implemented or very little;

→ *"Unethical" blinding of participants in 7 out of the 13 cases

Results – To what extent have recommendations when **communicating research ("Reporting")** been implemented?

COMMUNICATING RESEARCH	**Yes**	No	N/A	Total
2.1- Pre-registration of the study	**0**	17	2	19
2.2- Pre-specification of the study's design, primary outcome and analysis plan	**0**	17	2	19
2.3- Use of reporting checklist in the study	**0**	16	3	19
2.4- Use of a protocol checklist in the study	**0**	16	3	19

= WHAT IS *NOT* IMPLEMENTED (OR VERY LITTLE)

COMMUNICATING RESEARCH	**Yes**	No	N/A	Total
2.5- Disclosure statement	**9***	8	2	19

= WHAT *IS* IMPLEMENTED

→ "Reporting": 4 recommendations out of 5 <u>not</u> implemented

→ *Disclosure statement for <u>financial</u> conficts of interest quite well adopted

Results – To what extent have recommendations to **allow for verification of the research ("Reproducibility")** been implemented?

ALLOWING FOR VERIFICATION OF RESEARCH	**Yes**	No	N/A	Total
3.1- Raw coding available	**0**	17	2	19
3.2- Open data	**1**	16	2	19
3.3- Full disclosure	**0***	17	2	19

= WHAT IS *NOT* IMPLEMENTED (OR VERY LITTLE)

→ "Reproducibility": None of the 3 recommendations are implemented

→ *Disclosure statement for financial conflicts of interest is well adopted (see preceding slide)

BUT

→ *FULL disclosure is NOT implemented

→ Emerging paradigm of Open Science not adopted

→ No systemic change

Results – To what extent have recommendations when **publishing research ("Evaluation" and "Incentives")** been implemented?

PUBLISHING RESEARCH	**Yes**	No	N/A	Total
4.1- Post article on a "pre-print services"	**0**	19	0	19
4.2- Post article on a "pre-print services" with public comments/peer-review	**0**	19	0	19
4.4- Post article on a "post-print services" with public comments/peer-review	**0**	19	0	19
4.3- Publish article in Registered Reports format	**0**	19	0	19

= WHAT IS NOT IMPLEMENTED (OR VERY LITTLE)

PUBLISHING RESEARCH	**Yes**	No	N/A	Total
4.5-Publish article in a journal that encourages authors to follow "Open Science" practices	**11***	8	0	19

= WHAT *IS* IMPLEMENTED … BUT NOT DONE SO WELL

→ "Publishing": 4 recommendations out of 5 not implemented

→ *Paradox

"Authors are encouraged to (...) provide a Data Availability Statement. Cogent Education applies the Taylor & Francis Basic Data Sharing Policy."
"Applied Psycholinguistics encourages authors to consider uploading their data collection materials to a repository such as the IRIS database."

→ Open Science paradigm not implemented either by authors or journal editors and publishing houses (in the corpus)

→ No systemic change

Research methodology in the field of second language acquisition and learning

Sub-question 1:

"Are Open Science practices implemented in the SLA-SLLT articles comprising the corpus?"

Answer from the data:

- No, not yet
- In line with other scientific fields (see "extent of current adoption" in Munafo & al, table 1)

Explanatory hypothesis:

- No paradigm change
- Systemic change comes at high cost

Research methodology in the field of second language acquisition and learning

Sub-question 2:

"Are the SLA-SLLT articles comprising the corpus working towards common measurement tools and research methodologies?"

Results – To what extent have recommendations for **accurate measurement tools in LC** been implemented?

5.1- Accurate tool for measuring LC	**Occurrences**
A test (no description)	**3**
"In-house" test	**5**
Cambridge Preliminary English Test (Listening)	**1**
IELTS listening module	**2**
TOEFL iBT	**2**
TOEIC listening test	**2**
N/A	**4**
Total	**19**

→ "The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true." (Ionnadis 2005)

→ "The system [on which these tests are based] works very well for reading or grammar tests, for which [they were] developed, but not so well for listening tests." (Buck 2001, p. 212)

→ Meaurement accuracy?

Results – To what extent have recommendations for **relevant protocol designs in LC** been implemented?

5.2- Relevant protocol for measuring LC	**Occurrences**
"One-shot" survey	**3**
Pre-Post-test with average of 10-session treatement (of 1 to 2 hours)	**10**
N/A	**6**
Total	**19**

	Yes	No	N/A	Total
5.3- Use of within-person comparison	**0**	13	6	19
5.4- Blinding "instructor(s)	**1**	8	10	19

→ Protocols used are likely to make for low-power studies

→ Recommendations to reduce-bias in data collection not implemented

Research methodology in the field of second language acquisition and learning

Sub-question 2:

"Are the articles comprising the corpus moving towards common measurement tools and research methodologies?"

Answer from the data:

- No, not yet
- Problem: limits to the study not mentionned
- In line with Plonsky's finding on statistical robusteness

Explanatory hypothesis:

- Lack of awareness (Plonsky 2013)
- Lack of formalised tools and recommendations originating from the expert community on which to rely

Concluding remarks

- The articles comprising **the corpus** show **room for improvement** on every recommendation to counter the reproducibility crisis as coded for this study.
- **Threats to reproducibility** in all the studies from the corpus and **likely false results** in the quantitative studies.
- In this study, the situation is explained by the general **cost of systemic change** (here for Open Science), **lack of awareness for rigor, lack of formalized material/ recommendations in our field** on which to rely.

- Non-surprising results: recent studies / recent recommendations + findings in line with other scientific fields and with Plonsky for SLA.
- Limits: Results not to be generalized + non-representative corpus, **but** an ecologically-valid study → researcher faced with these studies in "real-life";

- Some possible perspectives in SLA-SLLT as a community:
  - *Validate common measurement tools and protocols for LC (ex.: Metacognitive Awareness Listening Questionnaire, Vandergrift et al. 2006)
  - *Favor continuous training in statistics AND research designs thanks to such symposium
  - *Favor Open Science thanks to commons tools (ex. Iris depository, OSF) and common stances (//AI community Statement on Nature Machine Intelligence)

Montpellier – May, 31st 2018 – Linda Terrier

"A systematic review of listening comprehension research designs and measuring tool in SLA research: towards a common measuring tool and methodology?"

SELECTED BIBLIOGRAPHY: REFERENCES CITED DURING THIS PRESENTATION

Blom, Elma & Sharon Unsworth . "Introduction." *Experimental Methods in Language Acquistion*, *Coll. Language Learning and Language Teaching*, John Benjamin Publishing, 2017, pp. 1-10.

Gelman, Andrew. "The failure of null hypothesis significance testing when studying incremental changes, and what to do about it." *Personality and Social Psychology Bulletin*, Vol. 44, N. 1, pp. 16-23, Sept. 10th 2017.

Gelman, Andrew. "Fixing the reproducibility crisis: Openness, Increasing sample size, and Preregistration ARE NOT ENUF!!!! " *Statistical Modeling, Causal Inference, and Social Science.* Posted on 15 April 2018, 9:43. Mackey, Alison & Susan M. Glass. *Second Language Research*. Routledge, 2010.

Dornyei, Zoltan. *Research Methods in Applied LInguistics*. Oxford University Press, 2007.

Ioannidis, John P. A. "Why most published research findings are false." *PLoS Med*. Vol. 2, No. 8, e124, 2005.

Moher, David & al. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. " *PLoS Med*, Vol. 6, No. 7, e1000097, July 21st 2009.

Munafò, Marcus R. & al. "A manifesto for reproducible science." *Nature Human Behavior*, Vol. 1, N. 21, 2017.

Nosek, Brian A. & al. "Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability." *Perspectives on Psychological Science,* Vol. 7, no. 6, pp. 615-631, 2012.

Plonksy, Luke. "STUDY QUALITY IN SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research." *Studies in Second Language Acquisition*, Vol. 35, Issue 4, 2013, pp. 655-687.

Plonsky, Luke (editor). *Advancing quantitative methods in second language research*. Routledge, 2015.

Open Science Collaboration. "Estimating the reproducibility of psychological science." *Science*, Vol. 349, Issue 6251, aac4716, 2015.

Popper, Karl R. *The Logic of Scientific Discovery*. Hutchinson, 1959.

Peng, Roger D. "The reproducibility crisis in science. A statistical counterattack." Significance, June 2015, pp. 30-32, 2015.

Saltelli, Andrea & Silvio Funtowicz. "What is science's crisis really about?." *Futures*. Vol. 91, 2017, pp. 5-11.