Sampling and generalization practices in second language acquisition research

Andringa, S.J.; Thieme, A.M.M.; Smits, Tim

Publication date
2025
Document Version
Final published version

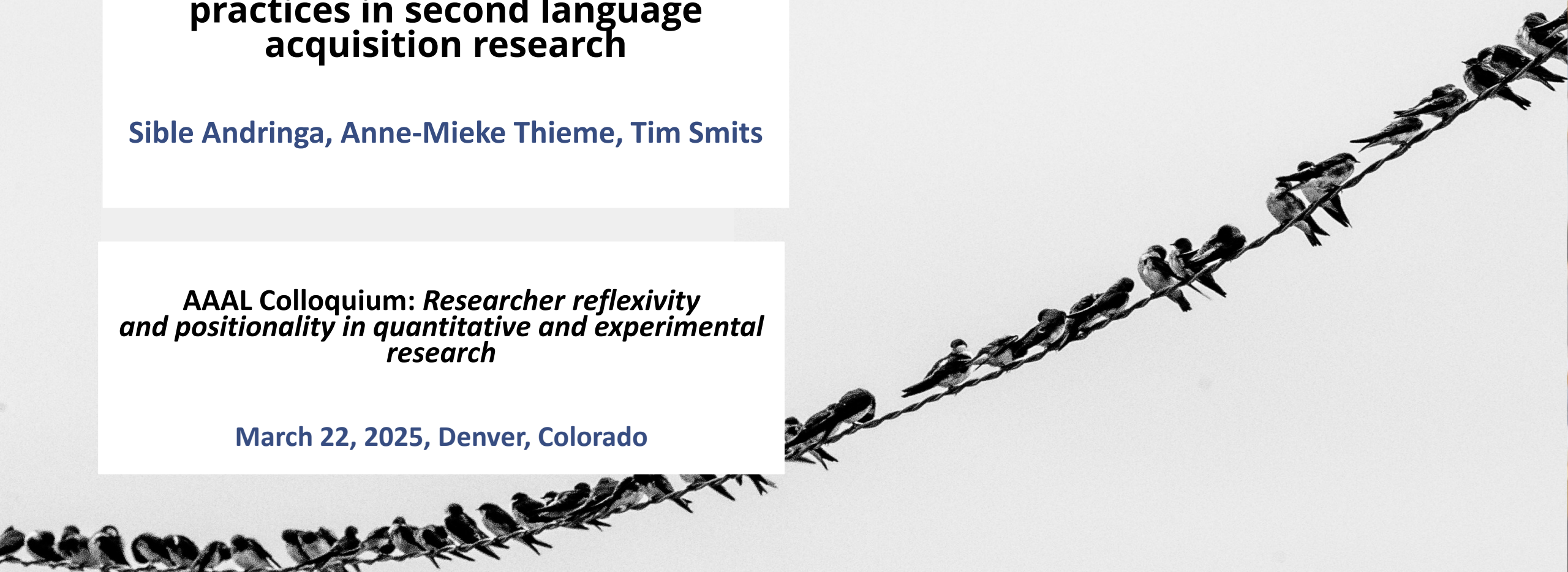
University of Amsterdam
Amsterdam Center for Language and Communication

Sampling and generalization practices in second language acquisition research

Sible Andringa, Anne-Mieke Thieme, Tim Smits

AAAL Colloquium: *Researcher reflexivity and positionality in quantitative and experimental research*

March 22, 2025, Denver, Colorado



Generalization

Quantitative research is governed by positivist en empiricist research philosophies.

- Generalization arises from random sampling from the population, deductive logic
- Science should be value-free
- The researcher is independent, detached, objective

Qualitative research is governed by interpretivist and constructivist research philosophies.

- Generalization arises from exploration of critical, theoretically sampled cases, inductive logic
- Science is socially constructed
- The researcher affects outcomes, co-creates knowledge, reflexive

(e.g., Borgstede & Scholz, 2021; Duff, 2006; Freeman et al. 2007)

Representative samples

- How well a sample represents the entire population (Campbell & Stanley, 1963)
- High levels of representativeness lead to high external validity (Plonsky, 2023; 2024)
- Full representativeness is unattainable, but necessary for populations characteristics that are relevant to the goals of the study (Babbie 2021)

Sampling bias and generalizability

- Sampling practices
  - Research samples are biased with regards to: Target Language (English), First langauge (English), Language pairings (Spanish-English); Educational context (academic); Location (US) (Andringa & Godfroid, 2020; Bylund et al. 2024; Collins & Munoz, 2016; Guo et al., 2021; Plonsky, 2023)

- Generalization practices
  - Concerns haven been raised about the generalizability due to sampling biases (Bigelow & Tarone, 2004; Ortega, 2005; 2019)
  - Systematic reviews of generalization practices in our field are scarce / absent?

The current study

An empirical evaluation of sampling and generalization practices:

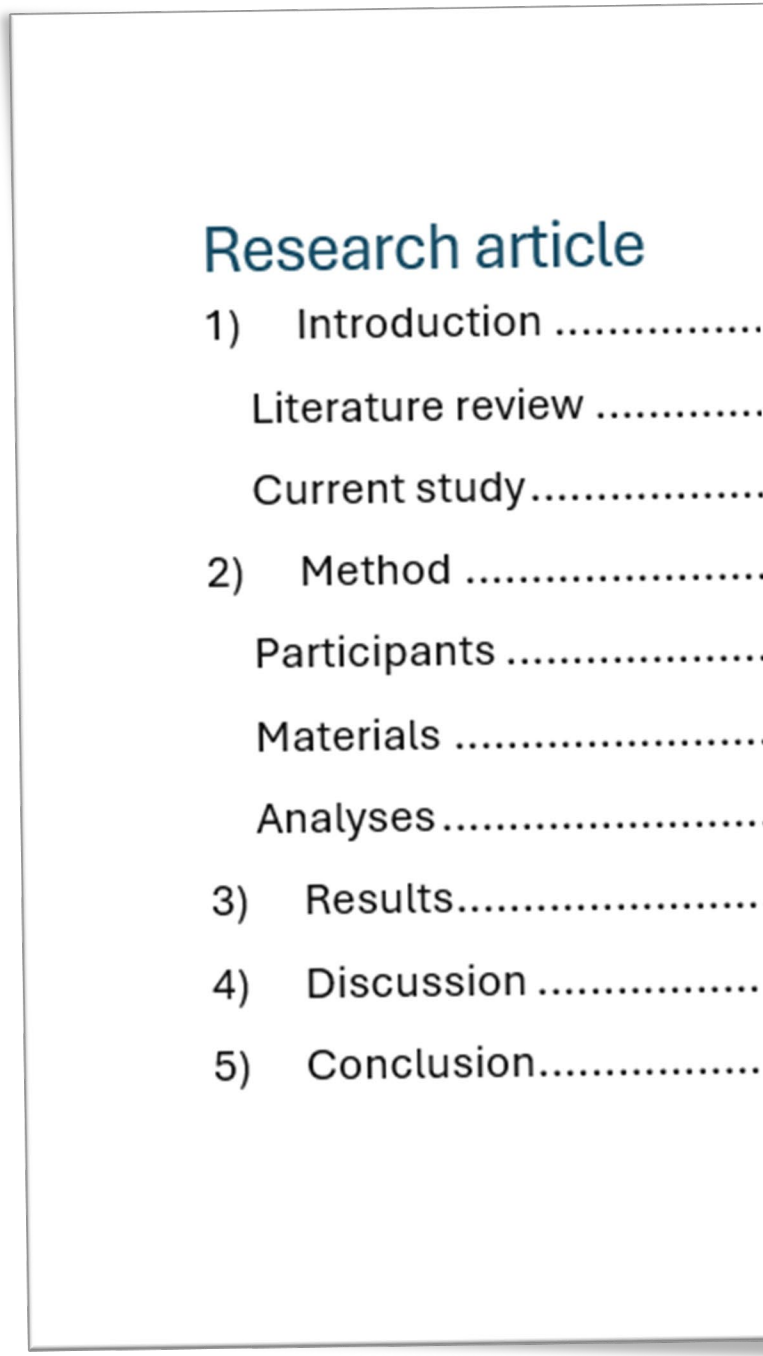
RQ1: What are the targeted populations and actual samples in high impact, general scope SLA journal articles?

RQ2: To what extent are samples statistically representative of the targeted population?

RQ3: Do articles reflect on possible limitations relating to the sample composition and generalizability?

Method

- All articles from 8 high impact journals in the field of SLA, from one year/volume: 2019.
- Only those presenting original research with L2 learners -> **212 articles included** (of 304 in total) .
- Qualitative and quantitative articles, coded for:
  - **Metadata**: authors, title, journal, year, design
  - **Targeted population** by analyzing goals, research questions, hypotheses and neighboring text
  - **Sample composition**: by analyzing method sections
  - **Sample reflection and generalization**: by analyzing dicussion sections.



Method

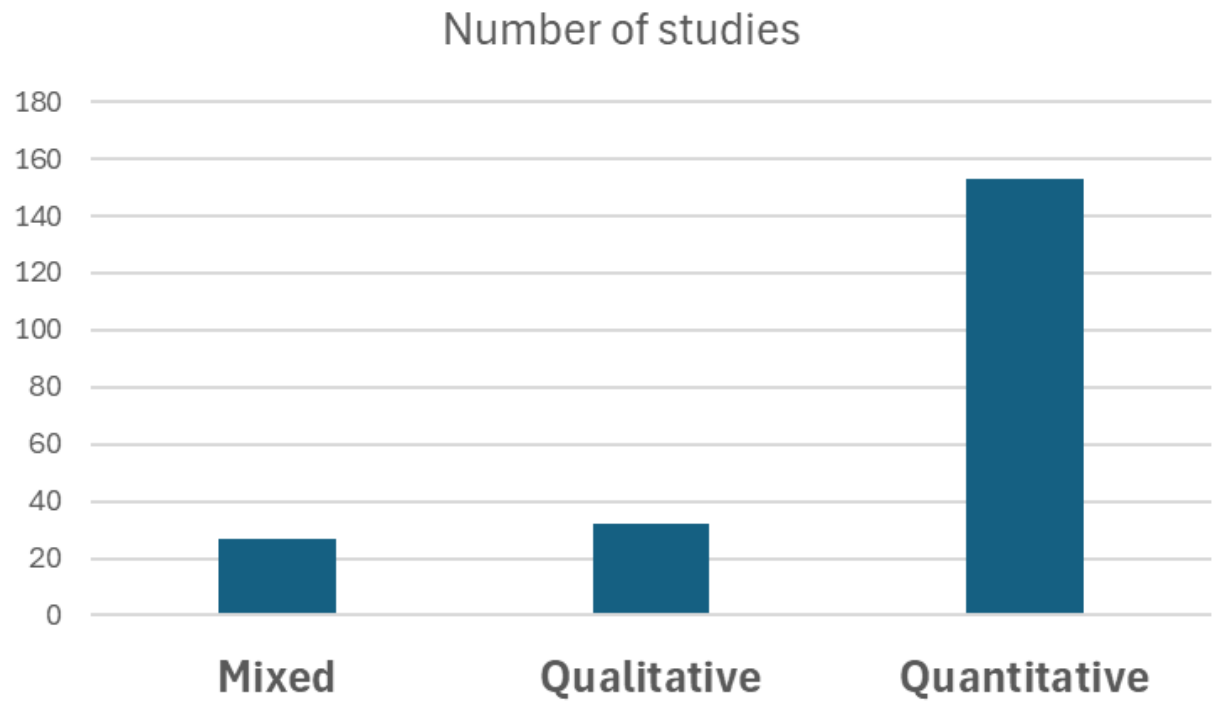
Targeted population	Sample composition	Reflection and generalization
Gender	Gender	Gender
Educational context	Educational context	Educational context
L1	L1	L1
LX	LX	LX
Age	Age	Age
Proficiency	Proficiency	Proficiency
FLA/SLA	FLA/SLA	FLA/SLA
Research context	Research context	Research context
Location	Location	Location
		A-specific sample refl.

Research article

1) Introduction .....
   Literature review .....
   Current study.....
2) Method .....
   Participants .....
   Materials .....
   Analyses.....
3) Results.....
4) Discussion .....
5) Conclusion.....

Results: Descriptives

- Most studies were quantitative in nature: 153 of 212
- Appr. 3988 pps surveyed in 212 studies.
- Quantitative, Qualitative and Mixed studies were similar in behavior.



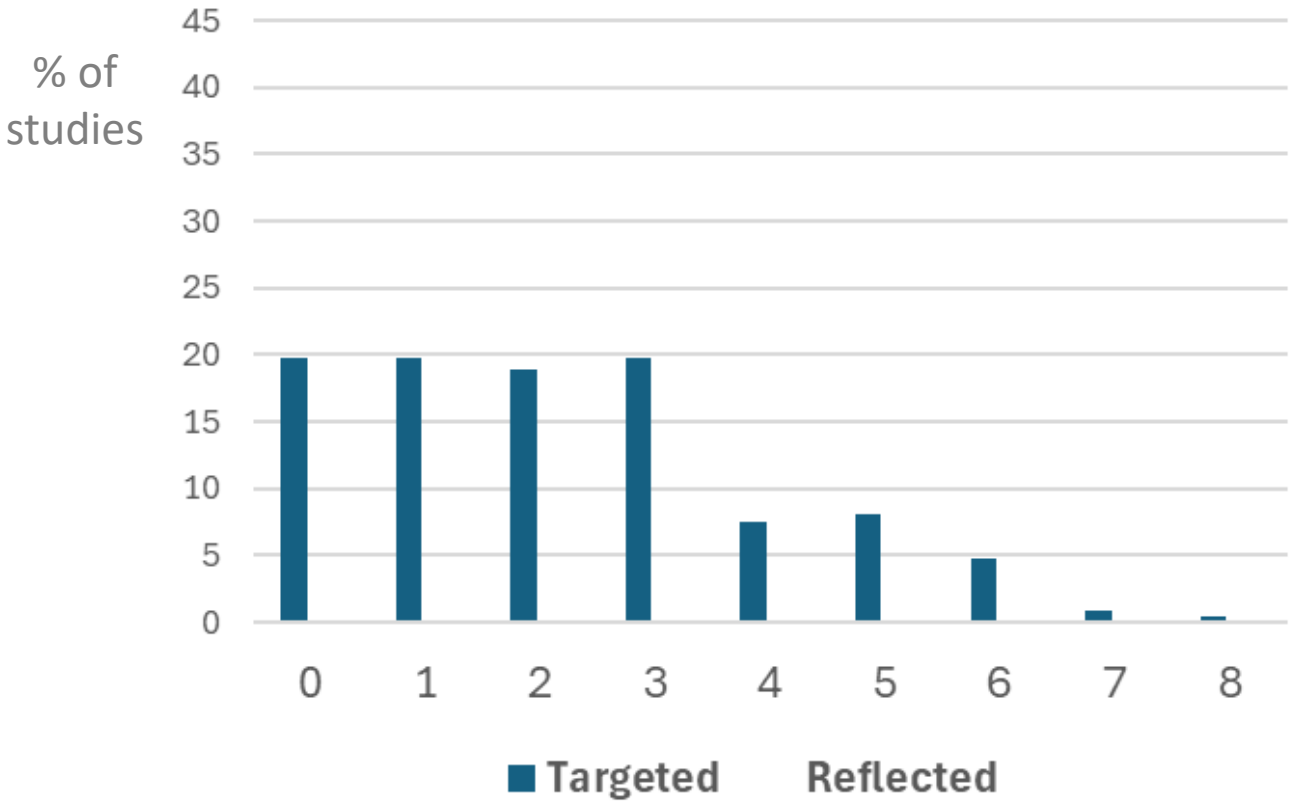
Results

How many population characteristics were targeted?

- 20% does not narrow down the targeted population

"Is there an effect of X on Y?"

"Is there a difference in behaviour for [L1] [LoE] learners of [L2] in [context], given [prof] and [gender]"

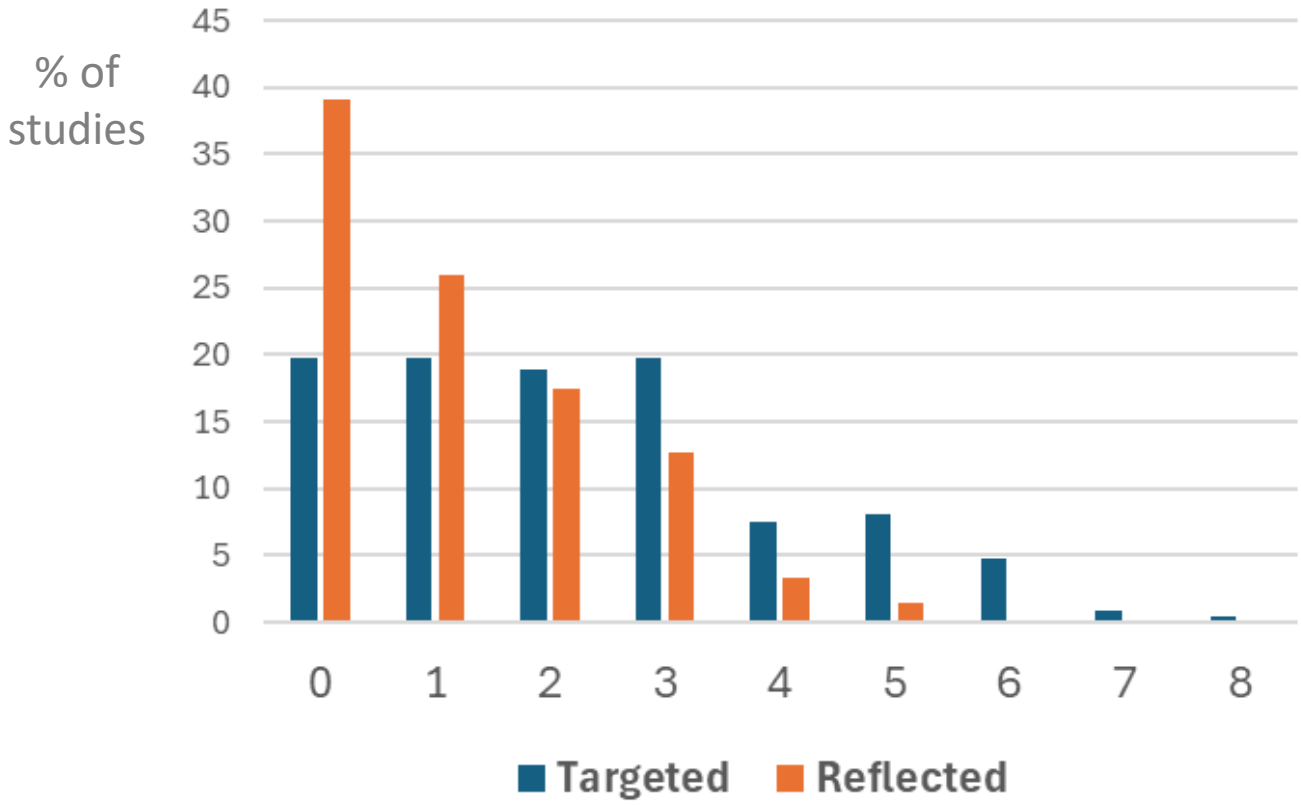


Results

How many sample characteristics were reflected on?

- 39 % did not reflect on any aspect of the sample.
- This includes aspecific sample reflections:

"These results are limited by the small number of participants tested"

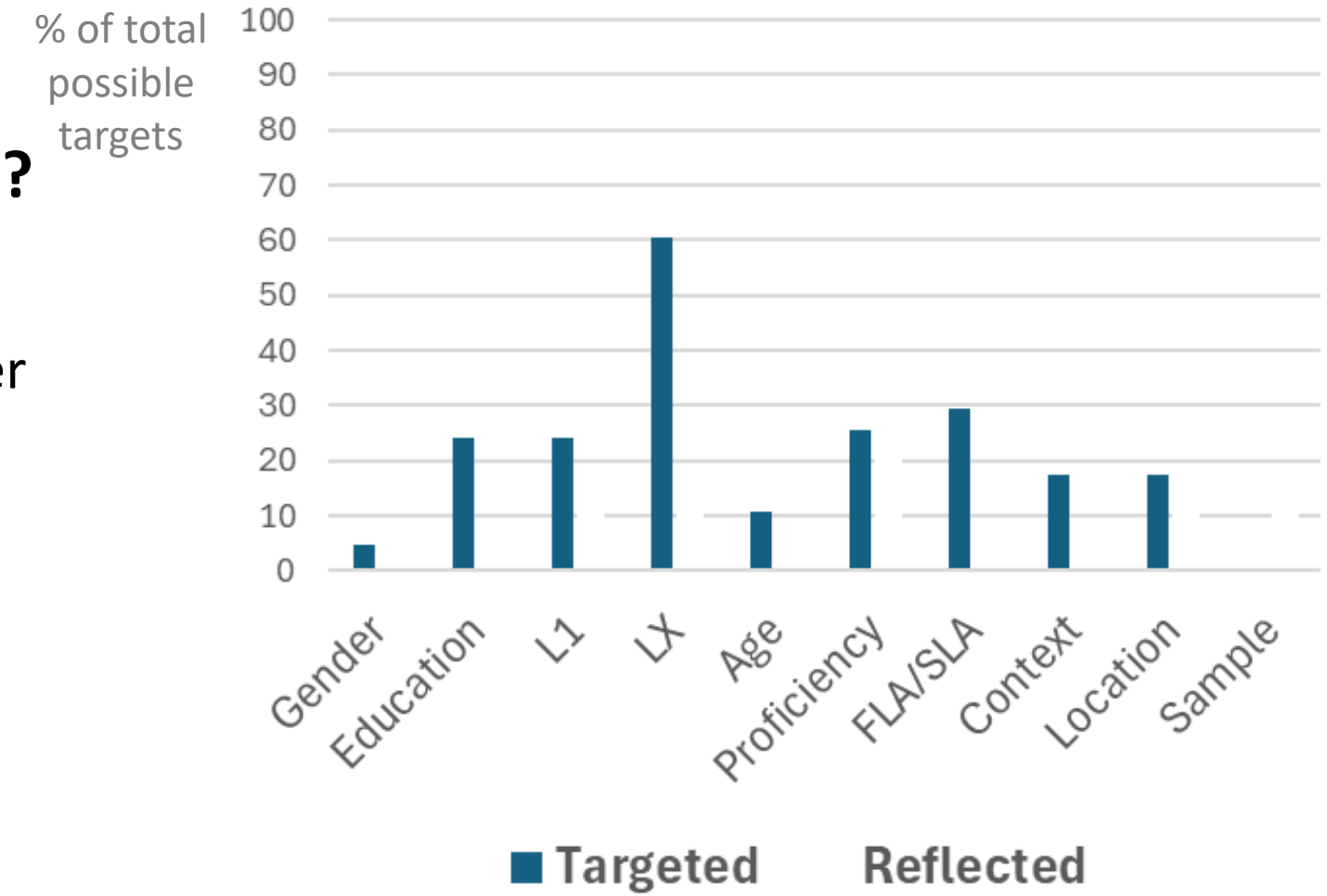


Results

Which population characteristics were targeted?

LX was most often targeted; gender least often.

Many characteristics are often not incorporated in study goals/RQs.

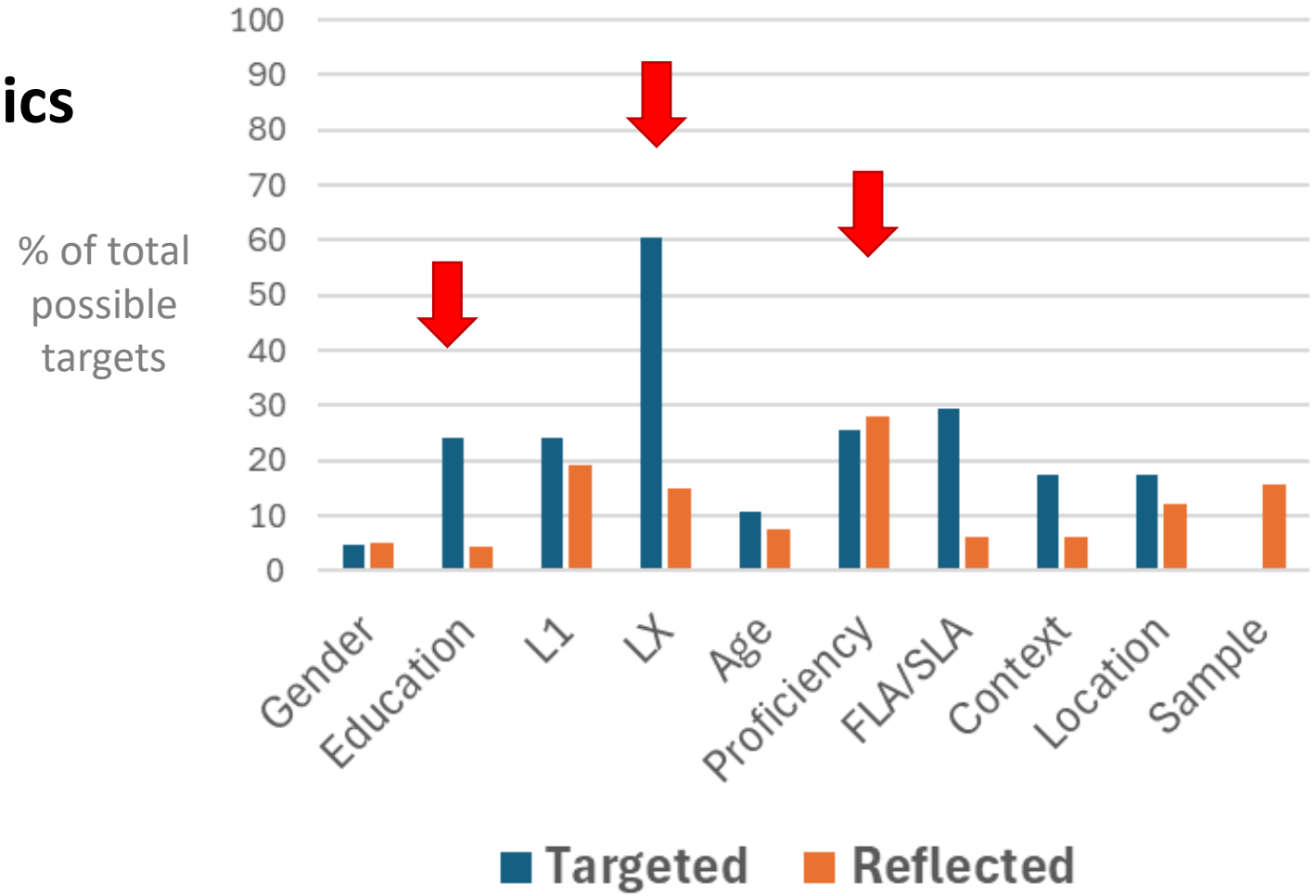


Results

What population characteristics are reflected on?

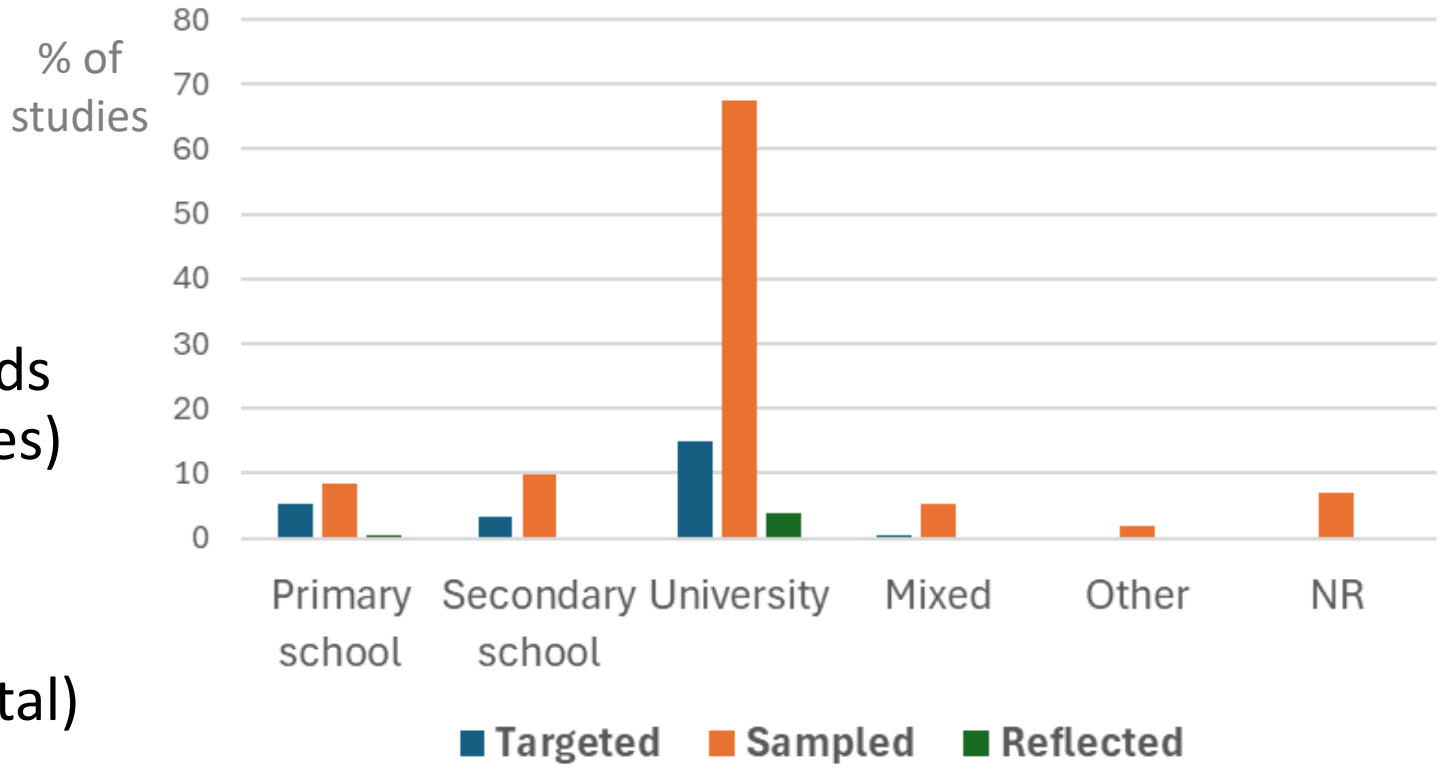
Proficiency is reflected on most frequently;

Educational context the least



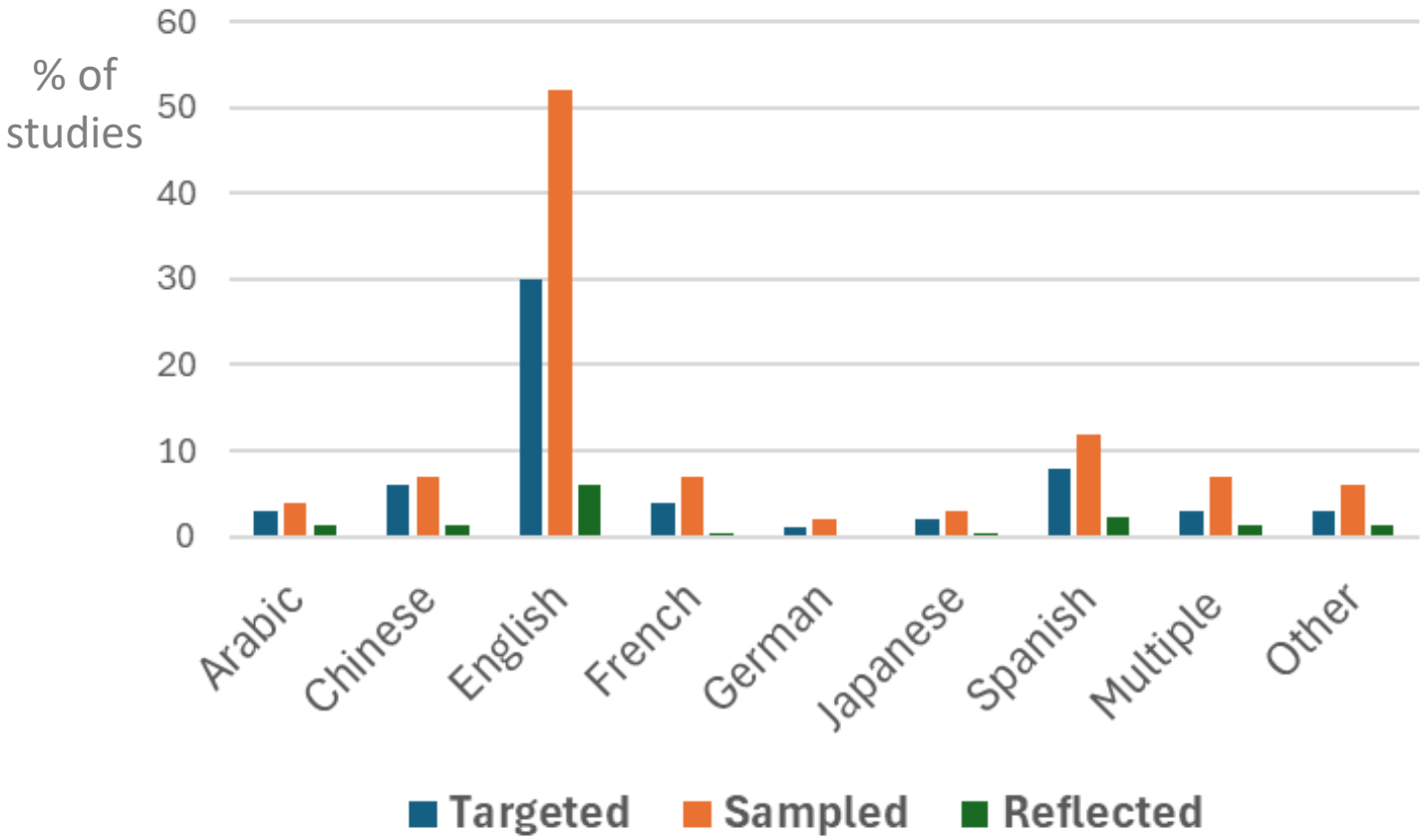
Results: Educational context

- 24% of the studies targeted an educational context
- Samples were heavily biased towards university contexts (67%; 113 studies)
- There is hardly reflection on educational context (9 studies in total)



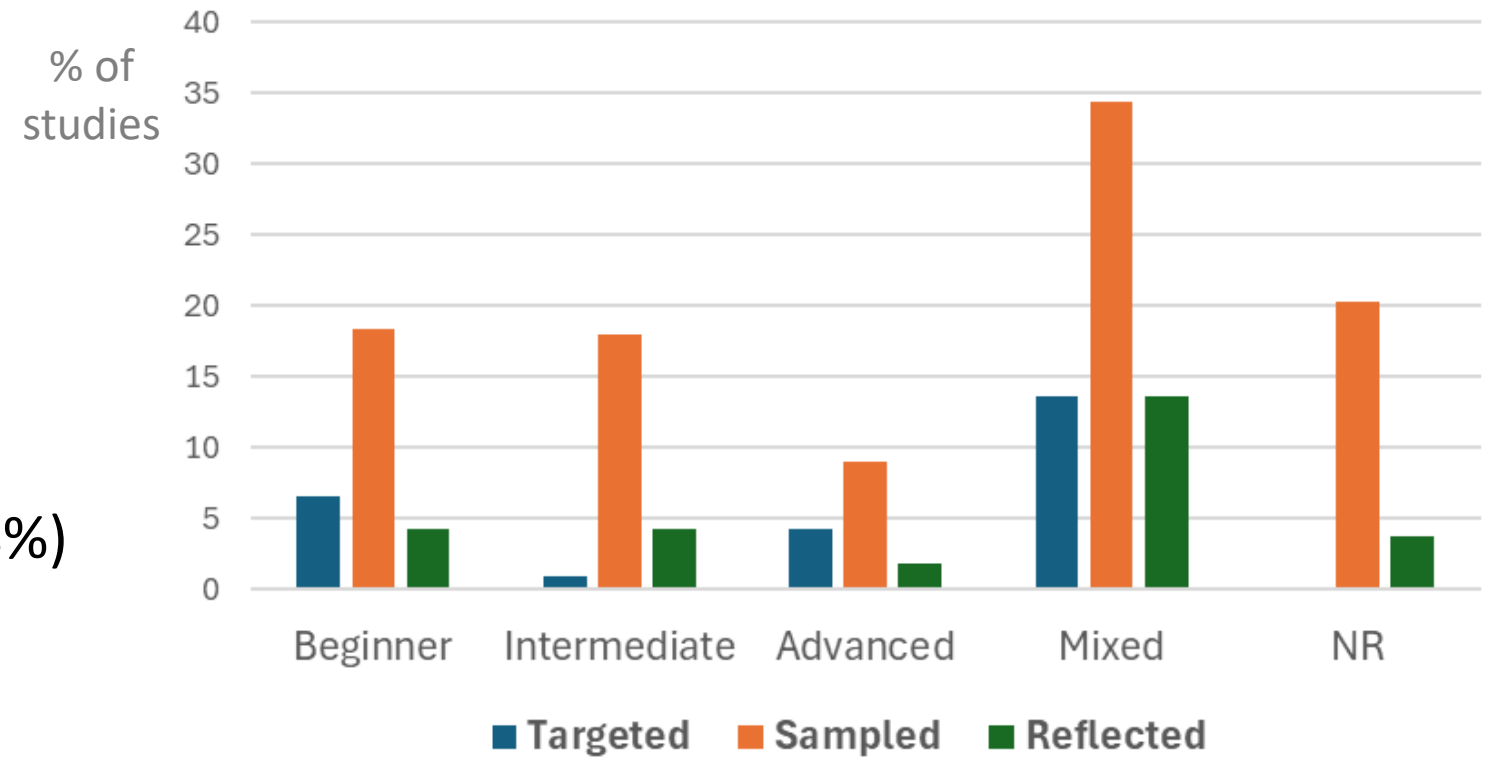
Results: LX

- 60% of the studies targeted an LX, 30% targeted English.
- 52% of the samples were English language learners.
- 15% reflected on the LX



Results: Proficiency

- 26% of the studies targeted some level of proficiency
- No clear proficiency biases in the samples
  - 20% does not report proficiency
- Profiency is most reflected on (28%)



Discussion

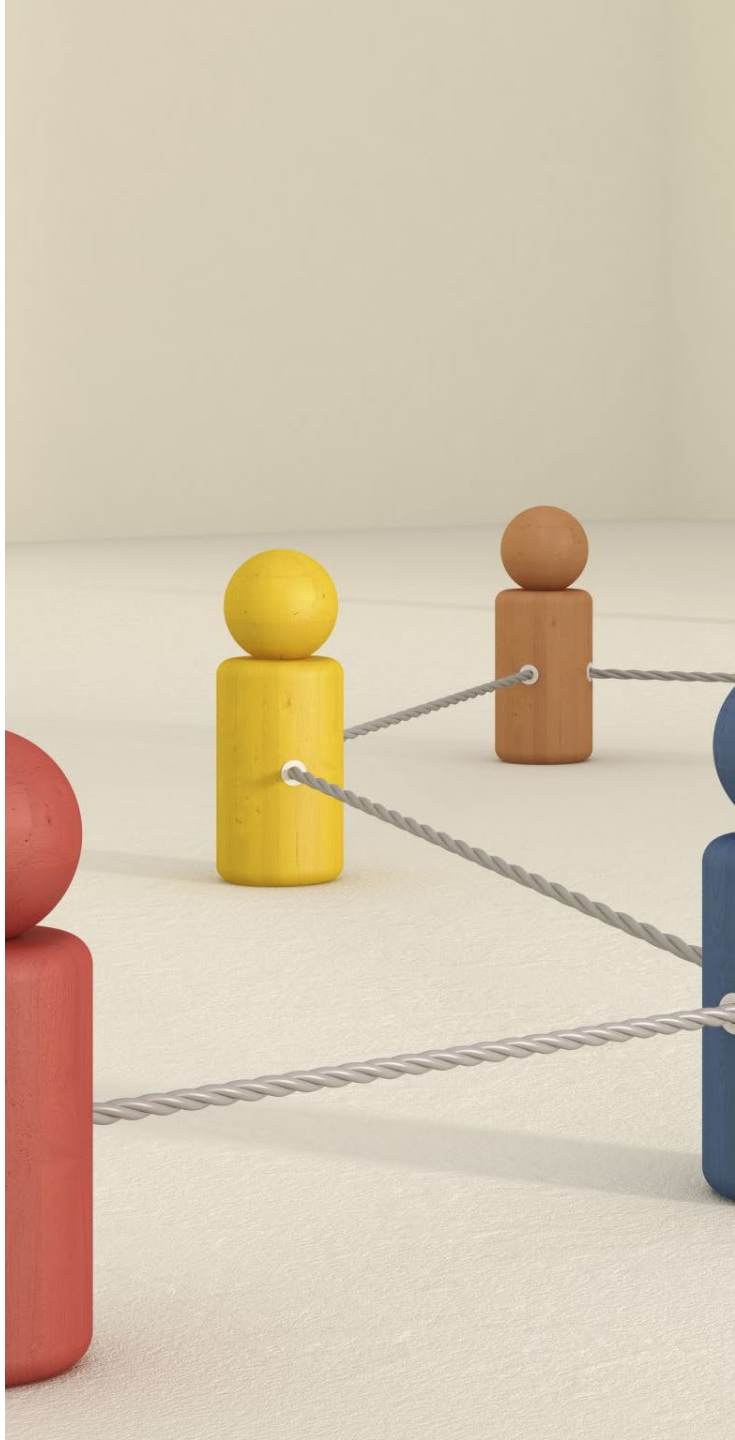
(Sampling) limitations of the present study

High impact journals:

- Articles intended to have high impact & wide generalizability

- Bias towards publishers, research institutions, and scholars situated in western, English-speaking countries.

- Representative for qualitative research approaches?

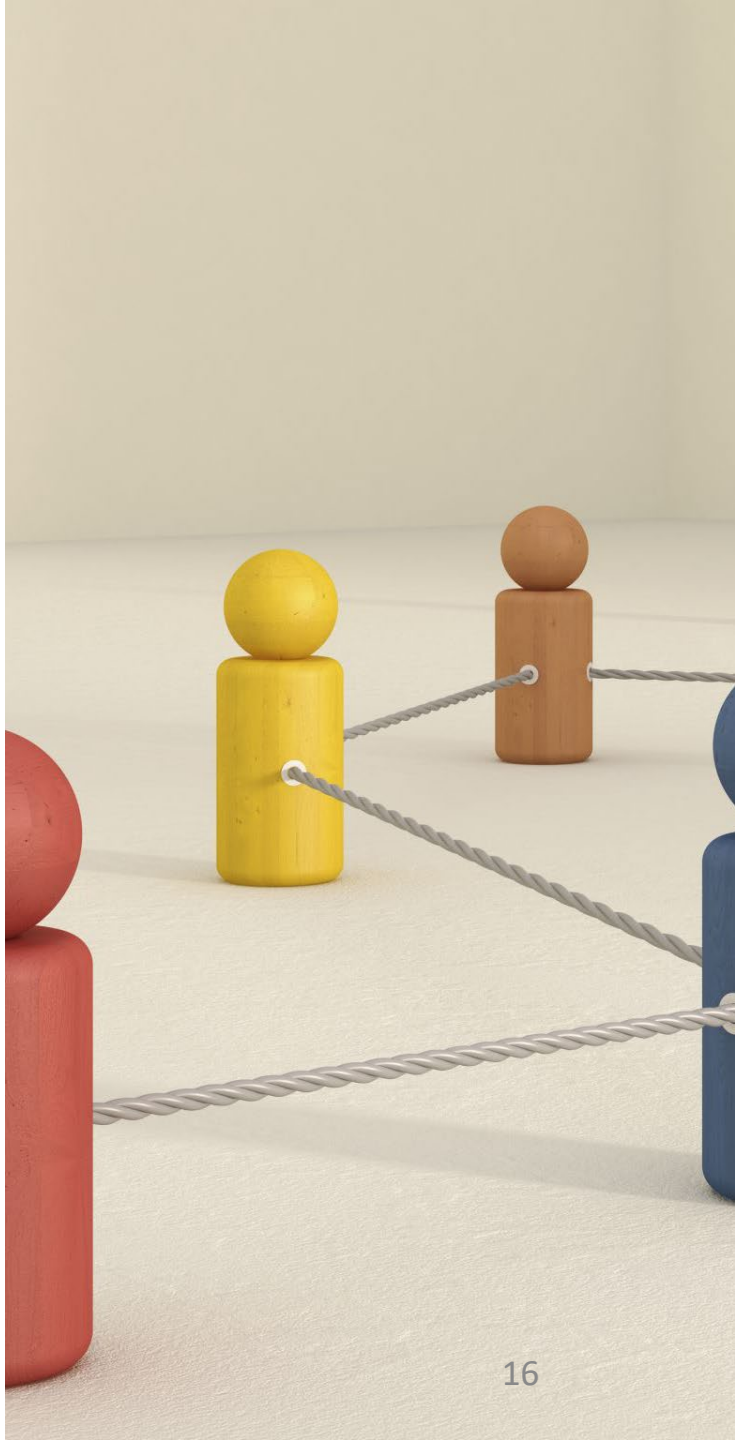


Why?

- Sampling choices are a prime site where the reseacher affects study outcomes: the generalization process needs to be transparent.

- Why is there so little transparency?
  - Word limits?
  - Lack of awareness?
  - Academic publishing culture / publish or perish, incentives us to:
    - Aggrandize?
    - Downplay limitations?
    - Conceal imperfections (QRPs? Isbel et al. 2022)

- **Paradox: We need more reflexivity to guarantee objective knowledge generation in quantitative approaches.**



Conclusion

Having better standards and increased transparency would…

a) increase awareness of our own role in shaping study outcomes
b) lead to better design choices, because they need to be justified.
c) result in more impactful research

"Overall, it seems that we have convinced ourselves that our neglect of certain L2 populations is motivated exclusively by innocent logistic and methodological reasons, in other words, that it is simply a technical neglect with neither ideological roots nor ethical consequences" (Ortega, 2005, p. 433)

Thank your listening!

Questions or comments, come find me!

Or email at s.j.andringa@uva.nl

References

Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics, 40*, 134–142.

Bigelow, M., & Tarone, E. (2004). The role of literacy level in second language acquisition: Doesn't who we study determine what we know? *TESOL Quarterly, 38*(4), 689–700.

Borgstede, M., & Scholz, M. (2021). Quantitative and qualitative approaches to generalization and Replication–A representationalist view. *Frontiers in Psychology, 12*

Bylund, E., Khafif, Z., & Berghoff, R. (2024). Linguistic and geographic diversity in research on second language acquisition and multilingualism: An analysis of selected journals. *Applied Linguistics, 45*(2), 308–329.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), Handbook of research on teaching (pp. 1-84). Rand McNally & Company.

Collins, L., and Muñoz, M.. 2016. The foreign language classroom: Current perspectives and future considerations. Modern Language Journal 100: 133–47.

Duff, P. A. (2006). Beyond generalizability. *Inference and Generalizability in Applied Linguistics: Multiple Perspectives*, 12, 65.

Godfroid, A., & Andringa, S. (2023). Uncovering sampling biases, advancing inclusivity, and rethinking theoretical accounts in second language acquisition: Introduction to the special issue SLA for all? *Language Learning, 73*(4), 981–1002.

Guo, Q., Zhou, X. E., & Gao, X. A. (2021). Research on learning and teaching of languages other than english in system. *System, 100*, 102541.

Isbell, D. R., Brown, D., Chen, M., Derrick, D. J., Ghanem, R., Arvizu, M. N. G., Schnur, E., Zhang, M., & Plonsky, L. (2022). Misconduct and questionable research practices: The ethics of quantitative data handling and reporting in applied linguistics. The Modern Language Journal, 106(1), 172–195.

Ortega, L. (2005). For what and for whom is our research? the ethical as transformative lens in instructed SLA. *The Modern Language Journal, 89*(3), 427–443.

Ortega, L. (2019). SLA and the study of equitable multilingualism. *The Modern Language Journal, 103*, 23–38.

Plonsky, L. (2023). Sampling and generalizability in lx research: A second-order synthesis. *Languages*, *8*(1), 75.

Plonsky, L. (2024). Study quality as an intellectual and ethical imperative: A proposed framework. Annual Review of Applied Linguistics, , 1–15.

Method

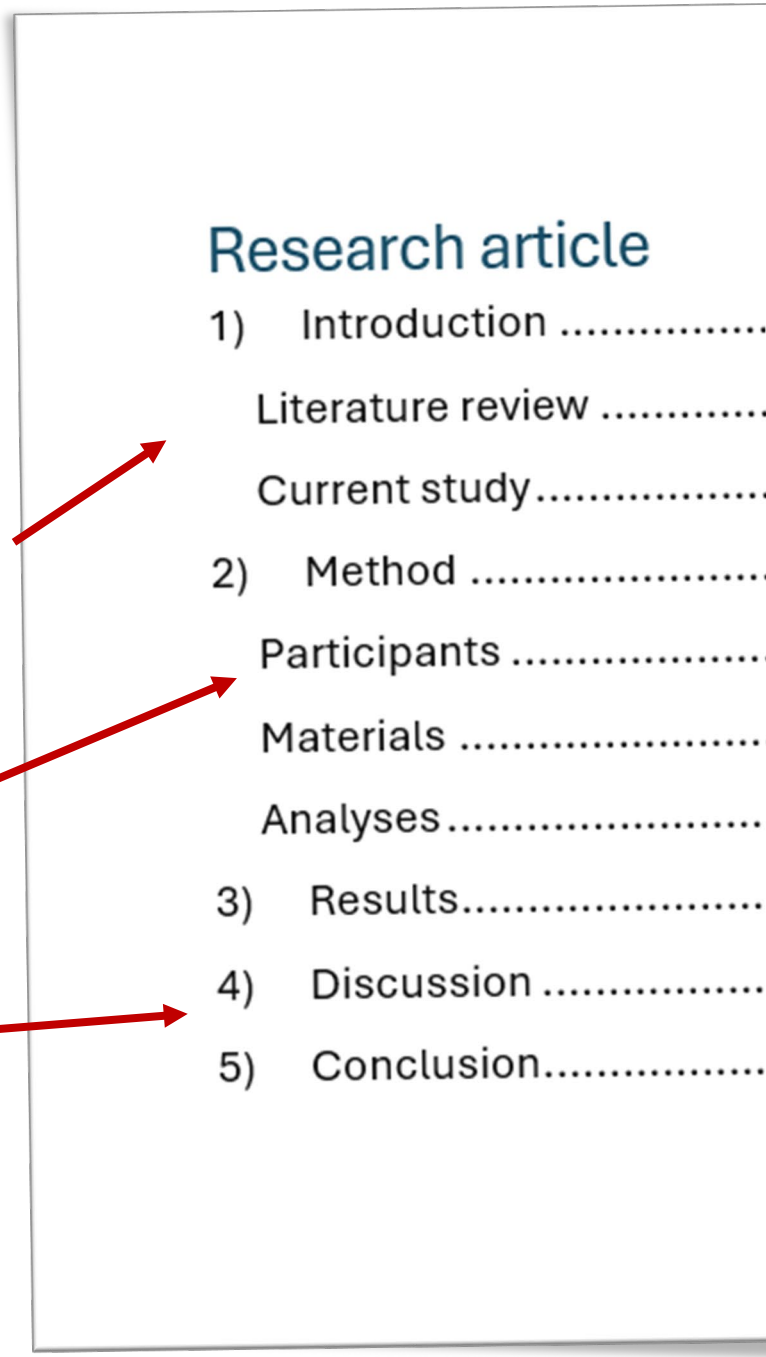
- Reflection and generelization coded for:

Reflection/limitations	
Sample size	"Number of participants was small"
Future research	"… should include a broader range of proficiencies"
Generalizability	"Because of [sample], we caution that the findings of this study many not be true for…"
Sample	"The sample consisted mainly of [characteristic]. This means that …"
Design	"This study was conducted with [characteristic]; it would be valuable to…"

Recommendations for reporting

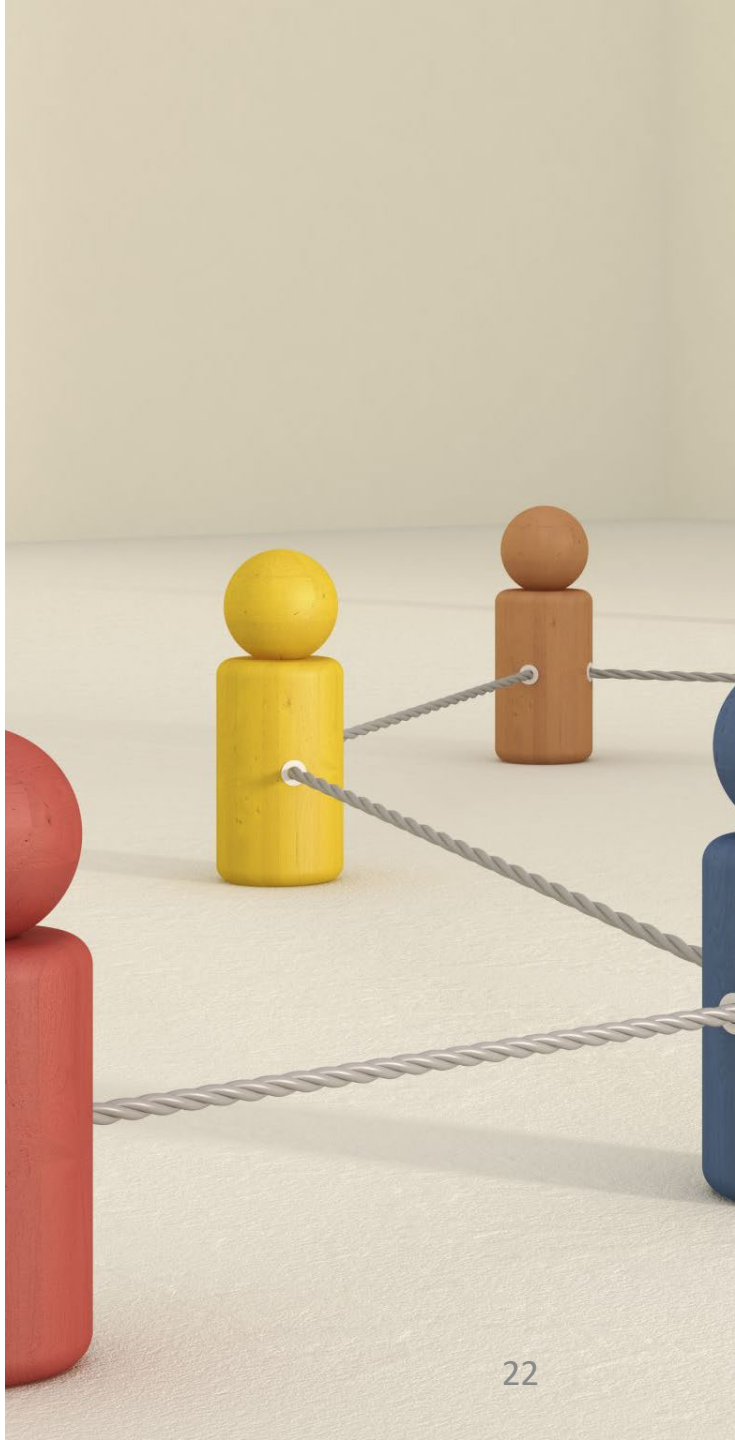
Make the generalization process explicit

- Narrow down the target population when needed, even if you cannot justify the target population theoretically.
- Be explicit about how your sampling choices were constrained by practical considerations, and what they mean for generalizabiltiy
- Discuss (limits to) generalizability due to sampling choices



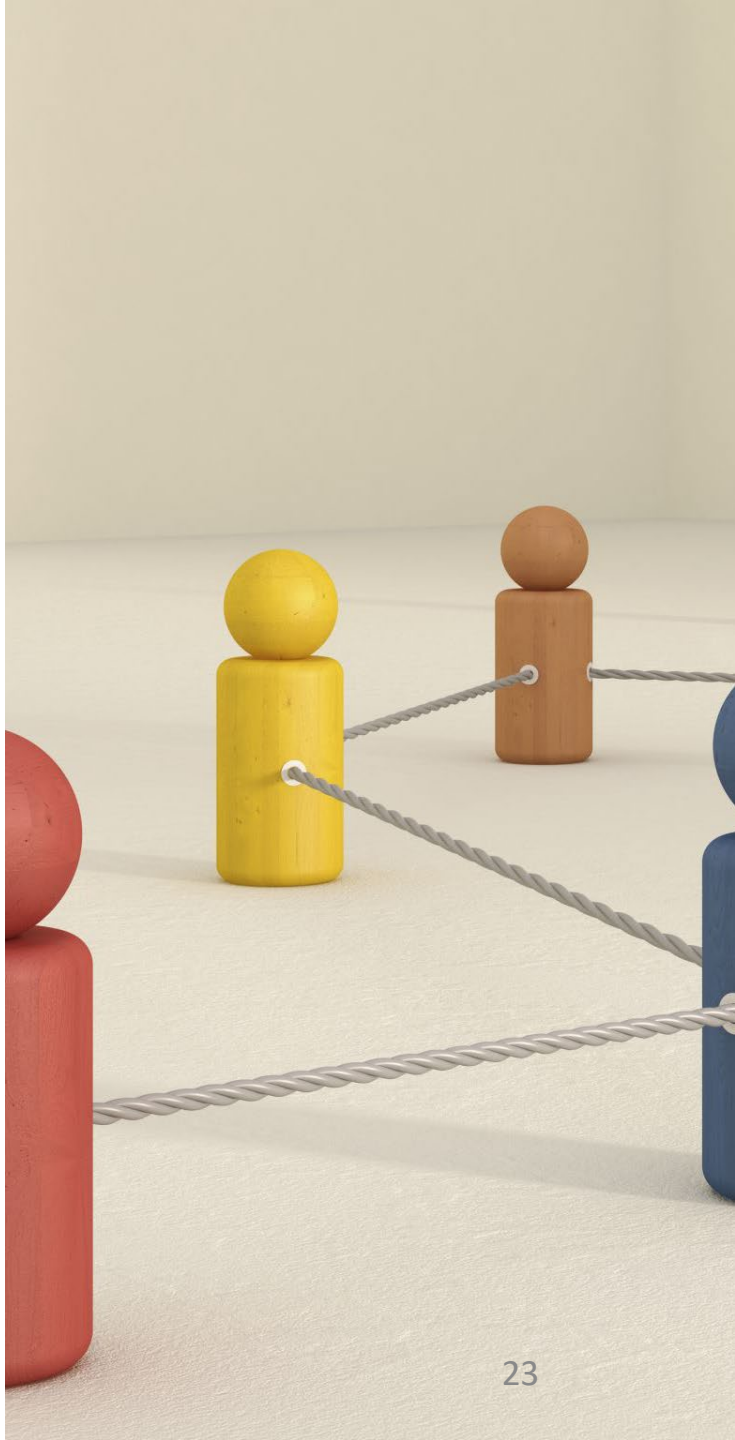
Discussion

- When specific population characteristics *are* targeted, then samples *do* match.
  - We did not assess if targeted characteristics were theoretically motivated.
  - Theoretical justifications can be very difficult to provide (Godfroid & Andringa 2023)
  - This, however, should not mean you don't have to narrow down the target population.



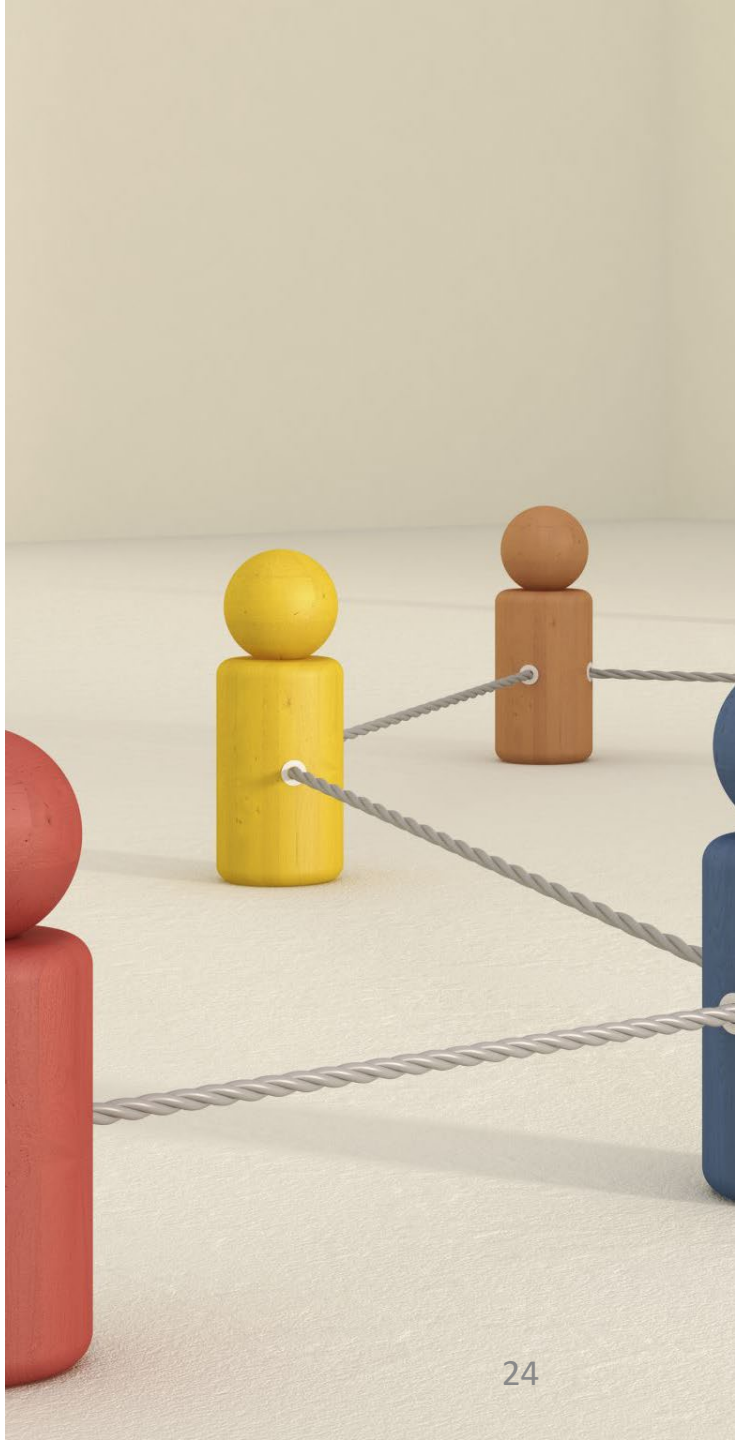
Discussion

- Few studies clearly narrow down the population they aim to generalize to, which implies they aim for wide generalizability.
    - No assessment of the appropriateness of claims made in conclusions, but given the biases present, unlimited generalization is hardly ever justified.
    - Generalization intentions should be explicated more



Discussion

- Not many studies clearly specify or narrow down the population they aim to generalize to, which implies they aim for wide generalizability.
- When specific population characteristics *are* targeted, then samples *do* match.
- Sampling choices imposed limits on generalizability (Andringa & Godfroid, 2020; Bylund et al. 2024; Collins & Munoz, 2016; Guo et al., 2021; Plonsky, 2023)
- Few studies discuss the implications of the sampling choices they made. Generalizability is not addressed.

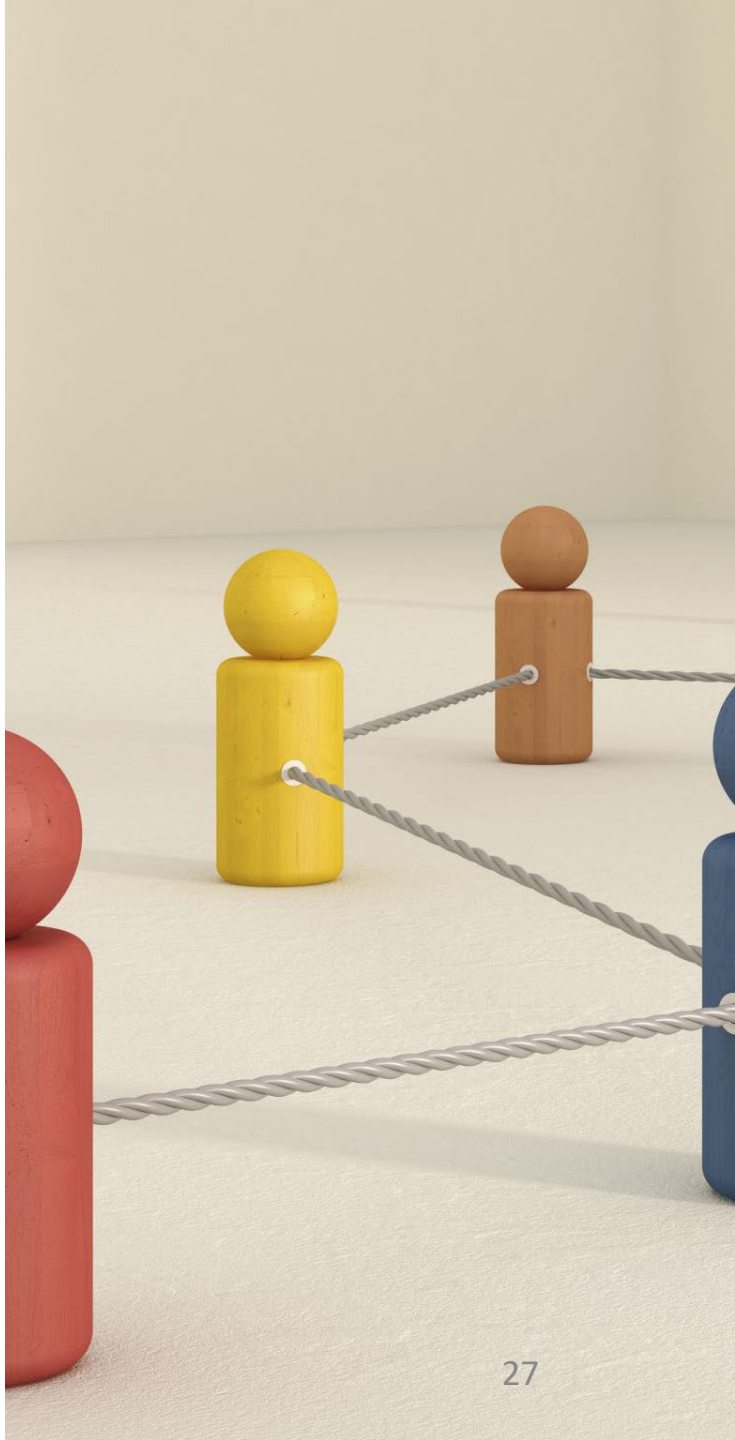


Why?

- Sampling choices are a prime site where the reseacher affects study outcomes -> transparency needed.

- Why is there so little transparency?
  - Word limits?
  - Lack of awareness?
  - Academic publishing culture / publish or perish, incentives us to:
    - Agrrandize?
    - Downplay limitations?
    - Conceal imperfections (QRPs? Isbel et al. 2022)

"Overall, it seems that we have convinced ourselves that our neglect of certain L2 populations is motivated exclusively by innocent logistic and methodological reasons, in other words, that it is simply a technical neglect with neither ideological roots nor ethical consequences" (Ortega, 2005, p. 433)



Conclusion

There don't seem to be reporting standards towards generalization.

Having standards and increased transparency would…

a) increase awareness of our own role in shaping study outcomes

b) probably lead to different design choices, because they need to be justified.

c) result in more impactful research

Thank your listening!

Questions or comments, come find me!

Or email at s.j.andringa@uva.nl