

SCIENTIFIC REASONING SCALE IN ITALY: VALIDATION STUDIES

Rossella Caliciuri, Margherita Lanz



UNIVERSITÀ
CATTOLICA
del Sacro Cuore

AIP Sperimentale, 31° Congresso Annuale
11-13 settembre 2025



BACKGROUND & RATIONALE

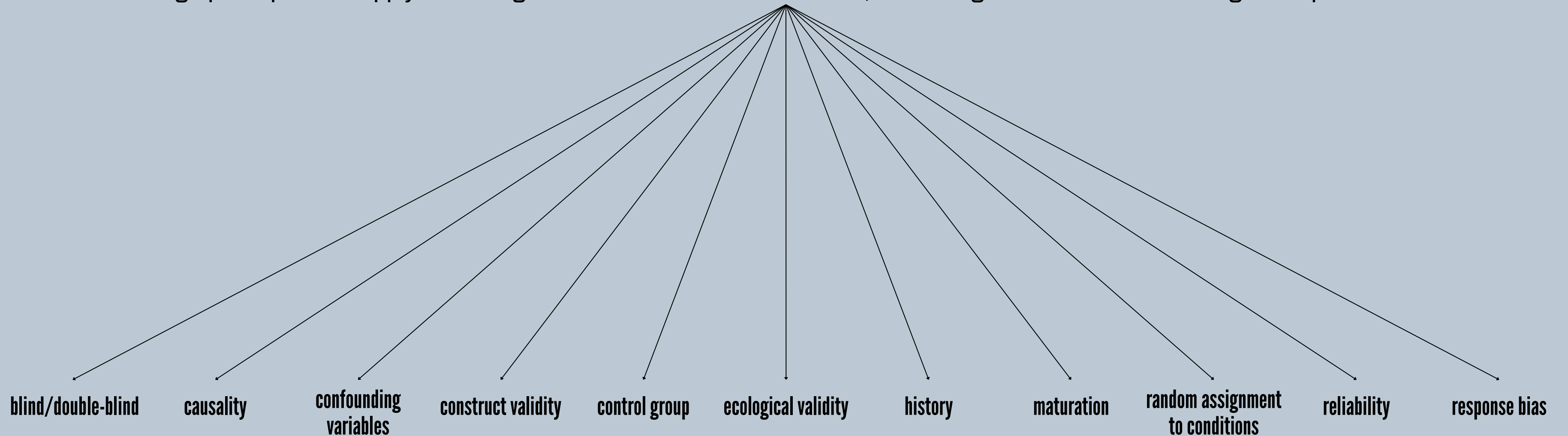
Foundations of the Scientific Reasoning Scale (SRS)

The SRS is a tool created to measure if people can really **think like scientists**, not just know facts.



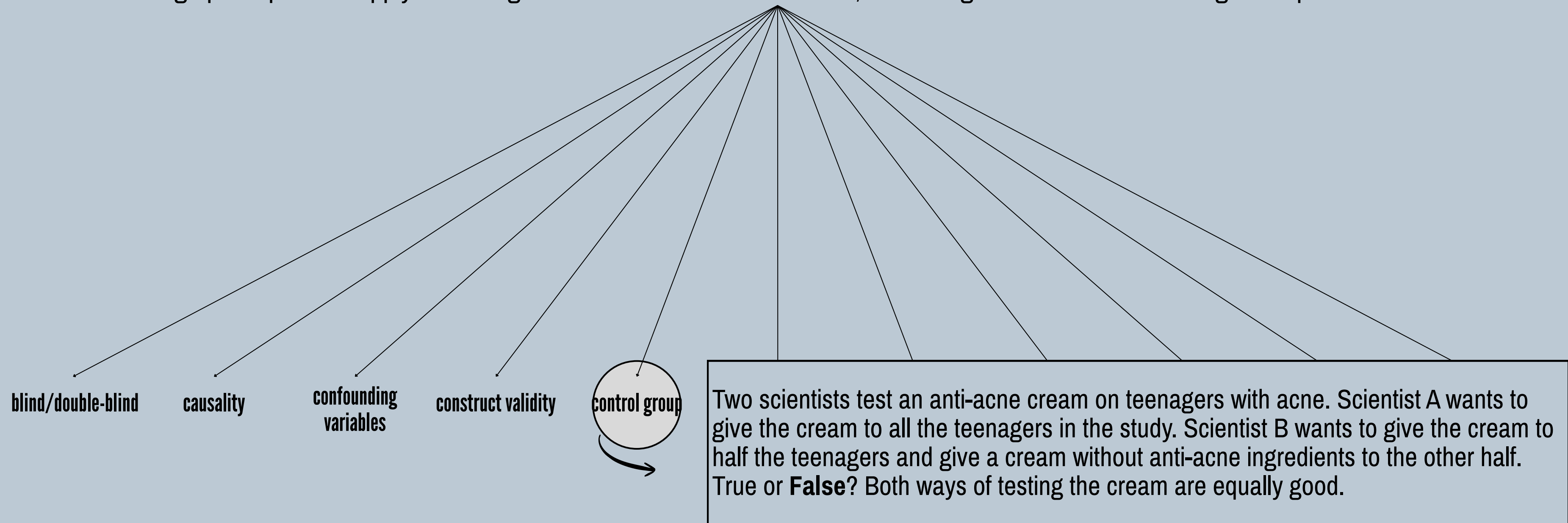
BACKGROUND & RATIONALE

The **Scientific Reasoning Scale** (SRS; Drummond & Fischhoff, 2017) has been originally validated in the US and then in Turkey (Muslu Kaygisiz et al., 2018). The SRS assesses **individuals' ability to evaluate scientific evidence**. The scale consists of **11 true/false items** that challenge participants to apply reasoning skills to brief scientific scenarios, evaluating core scientific reasoning concepts:



BACKGROUND & RATIONALE

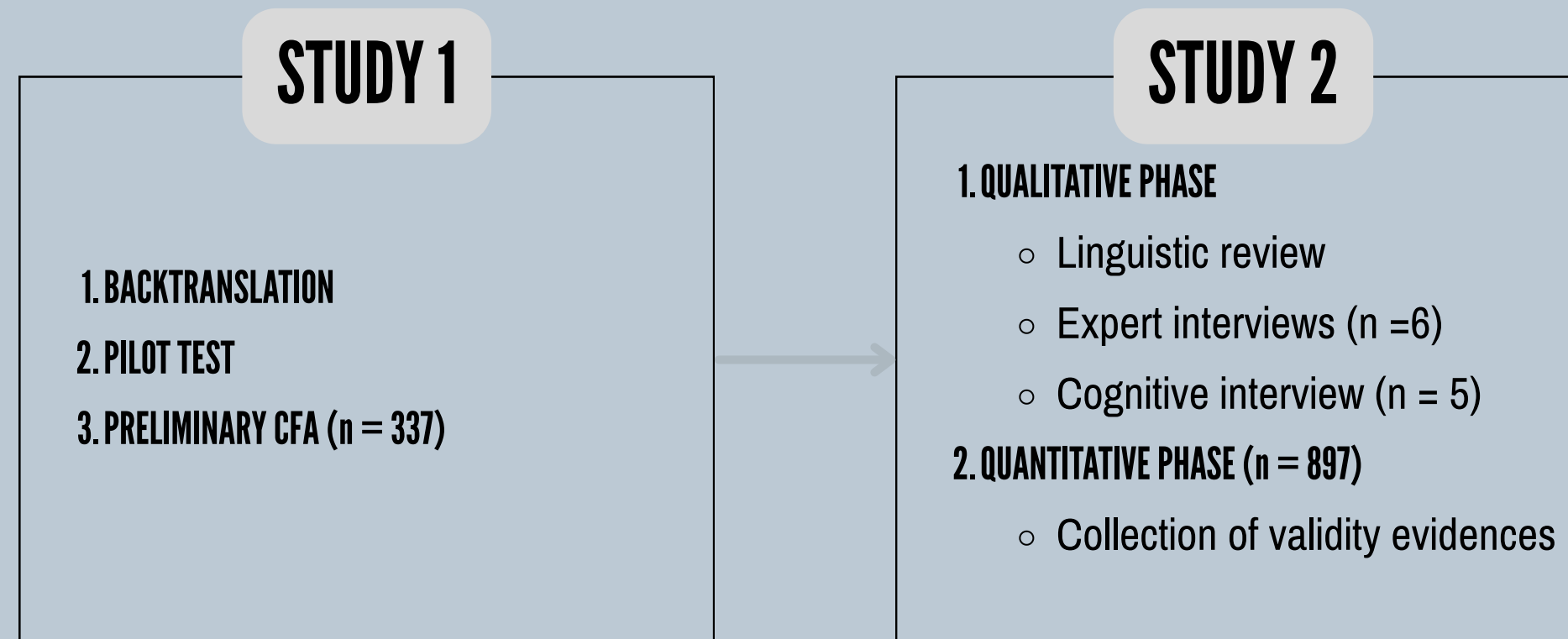
The **Scientific Reasoning Scale** (SRS; Drummond & Fischhoff, 2017) has been originally validated in the US and then in Turkey (Muslu Kaygisiz et al., 2018). The SRS assesses **individuals' ability to evaluate scientific evidence**. The scale consists of **11 true/false items** that challenge participants to apply reasoning skills to brief scientific scenarios, evaluating core scientific reasoning concepts:



AIM & METHODS

The present study aims to **gather validity evidence for the Italian version of the SRS.**

A **multi-study, multi-method** approach was employed:



AIM & METHODS

The present study aims to **gather validity evidence for the Italian version of the SRS**.
A **multi-study, multi-method** approach was employed:

STUDY 2

1. QUALITATIVE PHASE

- Linguistic review
- Expert interviews (n = 6)
- Cognitive interview (n = 5)

2. QUANTITATIVE PHASE (n = 897)

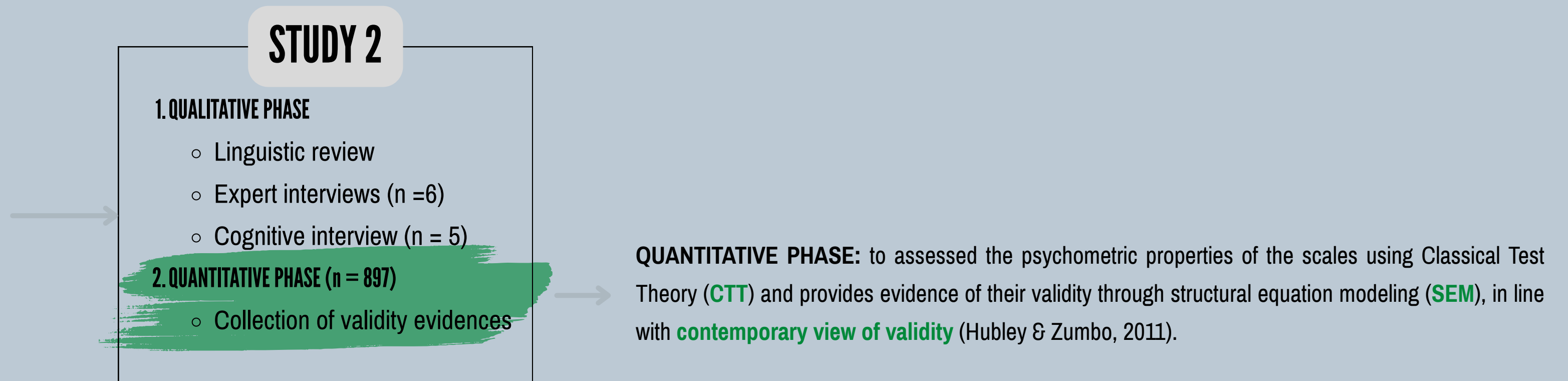
- Collection of validity evidences

NEW VERSION: *Due ricercatori testano una crema per l'acne su alcuni adolescenti con acne. Il ricercatore A vorrebbe dare la crema a tutti gli adolescenti presenti nello studio. Il ricercatore B vorrebbe darla solo a metà di loro, dando all'altra metà una crema neutra senza gli ingredienti per l'acne.*

- **A.** *Somministrare la crema a tutti gli adolescenti presenti nello studio è il modo migliore per verificare se la crema è efficace contro l'acne.*
- **B.** *Somministrare la crema solo a metà degli adolescenti, dando all'altra metà una crema neutra, è il modo migliore per verificare se la crema è efficace contro l'acne.*

AIM & METHODS

The present study aims to **gather validity evidence for the Italian version of the SRS**.
A **multi-study, multi-method** approach was employed:

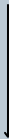


AIM & METHODS - CONTEMPORARY VIEW OF VALIDITY

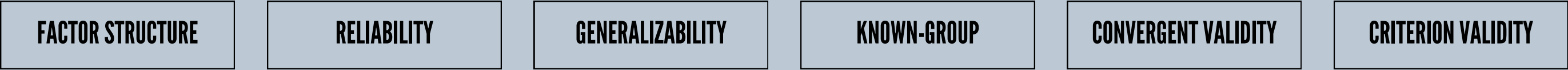
UNIFIED VIEW OF VALIDITY

- validity is contextualized
- unified validity (constrct validity)
- test users in addition to test developers

Cronbach & Meehl, 1955; Hubley & Zumbo, 2011; Messick, 1989; Sorgente & Zumbo, 2025; Zumbo, 2005.
Standards for Educational and Psychological Testing' (AERA, APA, & NCME, 2014)



In this validation process, Structural Equation Modeling (**SEM**) is pivotal as they integrate regression, path analysis, and latent variable models.
This validation process aims to collect **different sources of validity evidence** for the Italian SRS:



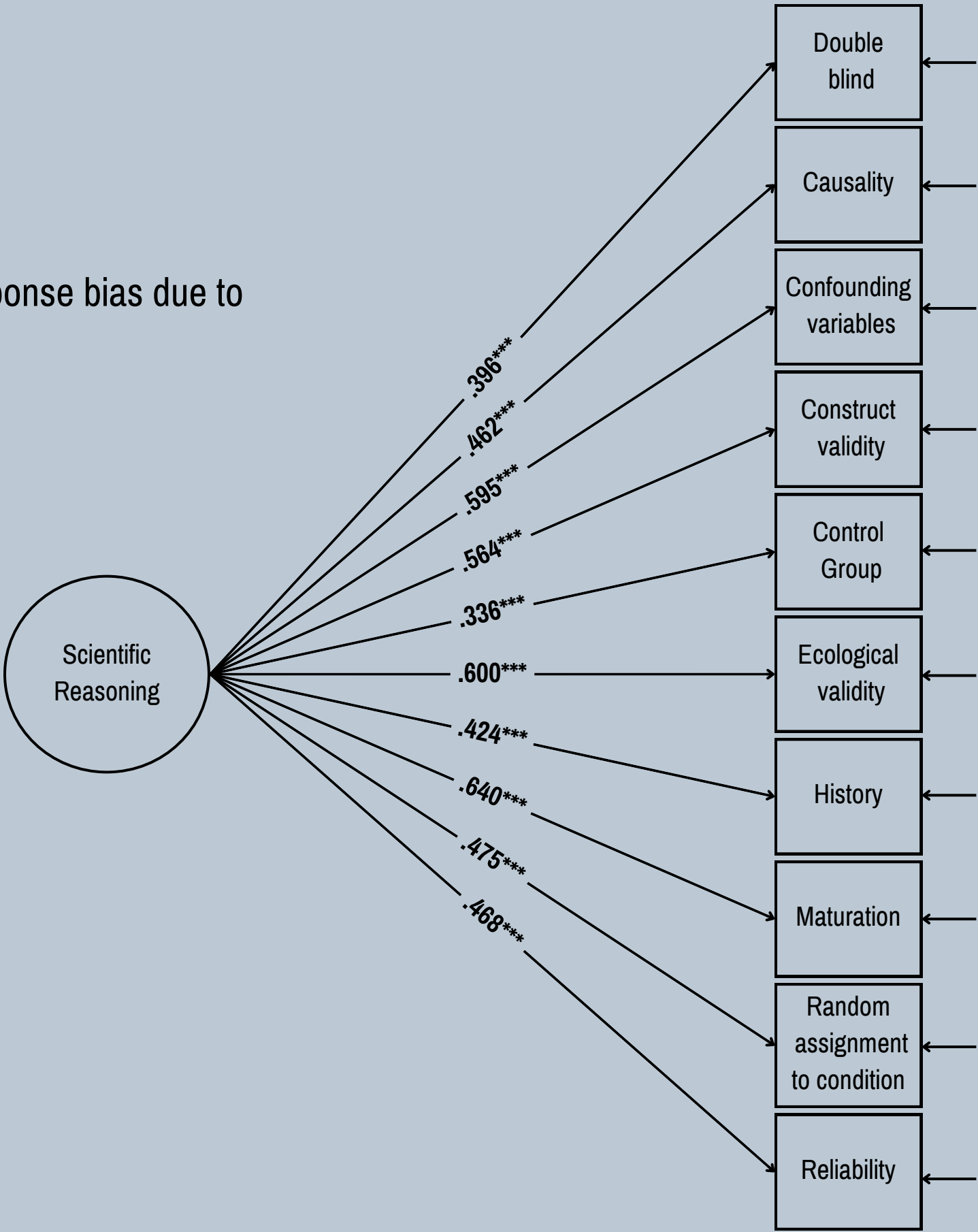
RESULTS & DISCUSSION

DIFFERENT KINDS OF VALIDITY EVIDENCES through SEM:

1

Via CFA **factor structure evidence** was confirmed, excluding the item on response bias due to insufficient factor saturation.

- The initial fit indices were satisfactory; however, item 11 did not sufficiently saturate the latent factor (loading < .3; Merenda, 1997; Peterson, 2000). Consequently, this item was removed, and a new CFA was conducted using 10 items instead of 11.
- The fit indices for the revised model were good: [$\chi^2 (35) = 61.690$, $p = .004$; RMSEA = .029 (.017 .041), $p = .999$; CFI = .966; WRMR = .953]. All factor loadings were high (greater than .3) and significant ($p < .001$), indicating a robust factor structure.



RESULTS & DISCUSSION

DIFFERENT KINDS OF VALIDITY EVIDENCES through SEM:

2

Reliability evidence ($\omega = .612$) confirmed the scale's reliability in the Italian sample.

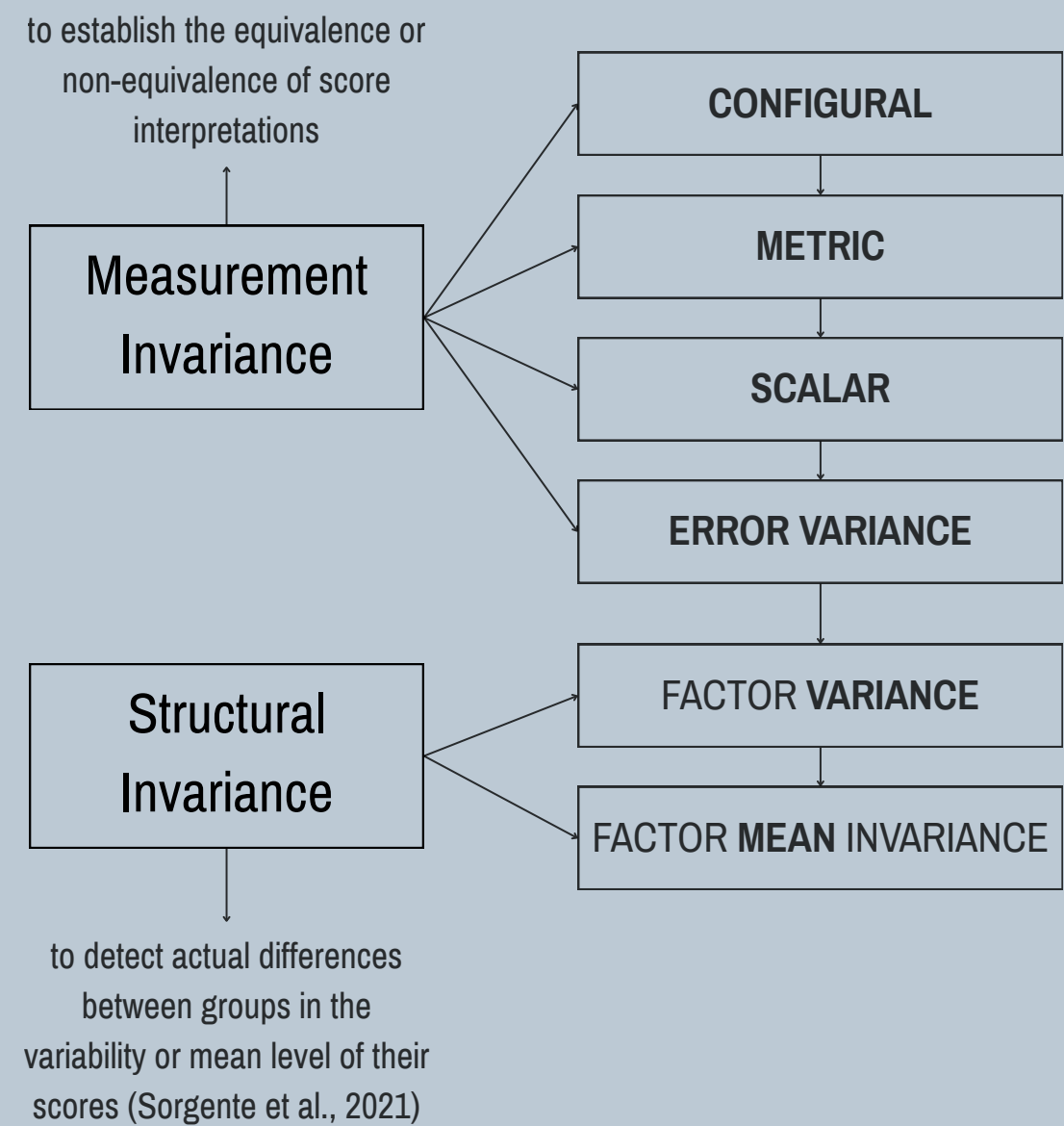
This value reflects the broad nature of the scientific reasoning construct (Little et al., 1999), which includes 10 distinct content domains (facets).

RESULTS & DISCUSSION

DIFFERENT KINDS OF VALIDITY EVIDENCES through SEM:

3

Generalizability
& Known-group
evidence



The scale demonstrated **full measurement invariance** across:

- gender: (achieving invariance required freeing item 9);
- age groups: 18-27 (Gen Z), 28-44 (Gen Y), 45-60 (Gen X);
- education level: up to diploma and up to PhD
- employment statuses (students, workers, and ‘others’);

Structural invariance was supported across gender, age groups, and employment status, but not across education level. **Individuals with at least a bachelor's degree had higher factor means than those with at most a high school diploma.**

This full invariance enhances the **scale's applicability** to Italian adults regardless of their background, ensuring that **any observed differences are real** and not due to measurement bias.

RESULTS & DISCUSSION

DIFFERENT KINDS OF VALIDITY EVIDENCES through SEM:

4

Convergent validity evidence was assessed by correlating the SRS with measures of Cognitive Reflection (CRT-Long) and Probabilistic Reasoning (PRS).

The fit of this model was good:
[χ^2 (53) = 88.14, p = .002; RMSEA = .027 (.017 .037), p = 1; CFI = .973; WRMR = .909]

	Cognitive Reflection	Probabilistic Reasoning
SRS	.523***	.578***

This validity is supported by **significant correlations** between the SRS and the CRT and PRS (consistent with Drummond & Fischhoff, 2017). Thus, higher scientific reasoning scores are associated with higher scores in both cognitive reflection (.523) and probabilistic reasoning (.578), and vice versa.

RESULTS & DISCUSSION

DIFFERENT KINDS OF VALIDITY EVIDENCES through SEM:

5

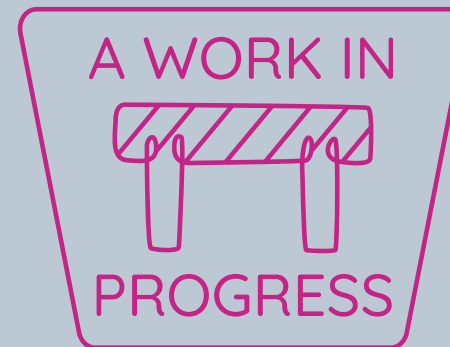
Criterion-related evidence was examined by testing a SEM model in which the score of SRS was related to the five factors of the Paranormal Health Beliefs Scale (parapsychological, superstitious, religious, extraordinary events, and pseudo-scientific).

The fit of the model was good
[χ^2 (98) = 123.85, p = .040); RMSEA = .017 (.004 .026), p = 1; CFI = .989; WRMR = .750]

Results showed **significant negative correlations** with five paranormal health belief factors, indicating that higher scientific reasoning scores are associated with fewer paranormal health beliefs.

Paranormal Health Beliefs	SRS
Parapsychological	-.316***
Superstitious	-.421***
Religious	-.398***
Extraordinary Events	-.406***
Pseudo-scientific	-.394***

OPEN QUESTIONS & FUTURE RESEARCH



VALIDITY IS NOT SOMETHING THAT IS PROVEN ONCE AND FOREVER IN JUST ONE STUDY

In line with contemporary view of validity, **future research** could provide further evidence:

- **ecological validity:** one idea is to create and test a version of the scale for daily life. This could be compared to the current version, which is more focused on lab situations.
- **measurement models:** other studies could also integrate alternative measurement models, such as Item Response Theory; this would allow for triangulated validity evidence regarding the models used to validate the scale;
- **cross-cultural research:** since SRS currently exists in four countries (Italy, US, Turkey, Israel), future research could investigate its cross-cultural invariance, examining whether the construct maintains the same meaning across these diverse cultural contexts.

GRAZIE



Questions, comments, suggestions, existential doubts – all welcome!

rossella.caliciuri@unicatt.it