Psychometric Properties of the Scientific Reasoning Scale

ROSSELLA CALICIURI*, MARGHERITA LANZ*
*UNIVERSITÀ CATTOLICA DEL SACRO CUORE, MILANO

scientific reasoning

a set of specialized **groups of cognitive processes** within the realm of thinking (Díaz et al., 2023).

a **process consisting of** (a) problem identification, (b) question identification, (c) hypothesis generation, (d) artifact construction, (e) evidence generation, (f) evidence evaluation, (g) drawing conclusions, and (h) communicating results (Fischer et al., 2014).

SR scale

Scientific Reasoning Scale (SRS – Drummond & Fischhoff, 2017), validated in the US and Turkey (Muslu Kaygisiz et al., 2018) measures **individual's ability to evaluate scientific evidence**. Using an interdisciplinary approach building on behavioral decision research, cognitive developmental psychology, and public understanding of science, the authors define SR skills and measure them with a **11 true/false** item that requires participants to apply their reasoning skills to **brief scientific scenarios**.

item's concepts

1. Blind / Double Blind (item 1)
2. Causality (item 2)
3. Confounding Variables (item 3)
4. Construct Validity (Item 4)
5. Control Group (Item 5)
6. Ecological Validity (Item 6)
7. History (Item 7)
8. Maturation (Item 8)
9. Random Assisgnment to Condition (Item 9)
10. Reliability (Item 10)
11. Response Bias (Item 11)

item's example

8. Subjects in an experiment must press a button whenever a blue dot flashes on their computer screen. At first, the task is easy for subjects. But as they continue to perform the task, they make more and more errors.
True or False? *The blue dot must flash more quickly as the task progresses.*

aim and method

We employed a psychometric methodology that integrates **Item Response Theory** (IRT; Birnbaum, 1986) and **Classical Test Theory** (CTT; Novick, 1966; Spearman, 1904), as proposed by Bean and Bowen (2021).

pre-registered on **OSF**

measures

Participants were asked to complete an anonymous online survey that includes:

- Demographic variables;
- Scientific Reasoning Scale (SRS);
- Some convergent measures;
- Some criterion measures.

sample

The sample comprised **337** adult Italian participants (**61,7% female**; 36,5% male; 1.8% other), aged **20-77** years (M=37, SD=13.64). Regarding education level, the majority of the sample had attained a **master's degree** (**50.6%**), 30.7% had completed a high school diploma, 12% had a bachelor's degree, 4.3% had a Ph.D., and 2.4% had an education level below a high school diploma.

On average, participants **answered 6.9 of the 11 SRS item correctly** (SD=2.2). The percentage of correct responses for each item:

item 1	**57.00%**
item 2	**69.10%**
item 3	**71.20%**
item 4	**56.10%**
item 5	**66.20%**
item 6	**70.90%**
item 7	**66.80%**
item 8	**75.40%**
item 9	**65.30%**
item 10	**40.70%**
item 11	**49.00%**

IRT results

After confirming the unidimensionality of the scale, IRT analyses were conducted with 6 items because items 1, 3, 5, 7, and 11 did not sufficiently saturate the latent factor (<.3).

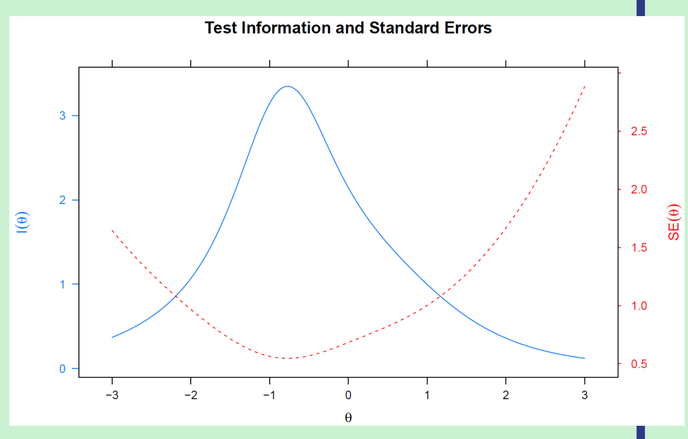
The fit statistics of the 2PL model indicated adequate fit **($M^2(9) = 5.148$, $p = 0.821$; RMSEA = 0, 95% CI [0 – 0], TLI = 1.027, CFI = 1)**.

Each item showed a non-significant S-χ^2 value, indicating that all 6 items fit under the 2PL unidimensional model.

	a	b
2 - Causality	**1.23**	**-0.84**
4 - Confounding Variables	**0.64**	**-0.41**
6 - Control Group	**1.20**	**-0.95**
8 - Maturation	**2.78**	**-0.81**
9 - Random Assignment to Condition	**1.03**	**-0.75**
10 - Reliability	**1.60**	**0.34**

Item Information Functions (IIF): items 2, 4, 6, 7, 8, and 9 are informative for medium-low trait levels, while items 10 is informative for medium-high.

Test Information Function (TIF), it emerges that, overall, the SRS is suitable for detecting medium-low levels of scientific reasoning (information is most informative in the trait range between -.1 logits and -0.5 logits).



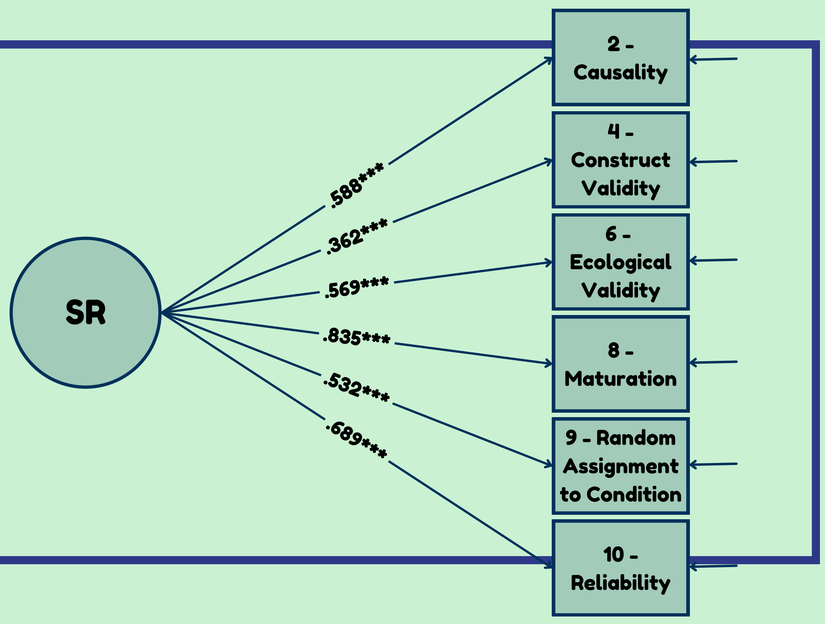
CTT results

Items 1, 3, 5, 7, 11 did not sufficiently saturate the latent factor (<.3).

Fit indices of 6 item's **CFA** were good: **[χ^2 (9) = 4.734, p = .856); RMSEA = .000 (.000 - .033), p = .987; CFI = 1.000; WRMR = .402].**

The 6 factor loadings were all >.362 and significant ($p \leq .001$).

ω = **.615**;
percentage of correct responses: **3.8 (SD=1.6).**



bibliography

- Bean, G. J., & Bowen, N. K. (2021). Item response theory and confirmatory factor analysis: complementary approaches for scale development. Journal of Evidence-Based Social Work, 18(6), 597-618.
- Díaz, C., Dorner, B., Hussmann, H., & Strijbos, J. W. (2023). Conceptual review on scientific reasoning and scientific thinking. Current Psychology, 42(6), 4313-4325.
- Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. Proceedings of the National Academy of Sciences, 114(36), 9587-9592.
- Drummond Otten, C., & Fischhoff, B. (2023). Calibration of scientific reasoning ability. Journal of Behavioral Decision Making, 36(3), e2306.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., ... & Eberle, J. (2014). Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. Frontline Learning Research, 2(3), 28-45.

discussion

A model that integrates CTT and IRT offers a comprehensive assessment: some information derived from both approaches providing a **triangulated evaluation of item quality**, some aspects are **unique** to each method (Bean & Bowen, 2021):

- IRT allowed us to exclude 5 items from our sample and identified the difficulty and discrimination parameters for each: most of them discriminate for a low-to-moderate level of scientific reasoning; only one item discriminates for a high level.
- CFA indicated that the original scale, developed in the USA, does not adequately fit the Italian data. Therefore, we considered the scale as consisting of 6 out of 11 items.

Overall, this scale seems to work in Italy by using **6 items aimed at measuring a low-to-moderate level of scientific reasoning**.

Future developments should focus on understanding how to develop items that effectively measure concepts related to scientific reasoning, which were excluded from this scale. These concepts include double-blind procedures, confounding variables, control groups, history effects, and response bias.

scan me to the poster

ISSBD 2024
LISBON | PORTUGAL
37TH BIENNIAL MEETING OF THE INTERNATIONAL SOCIETY FOR THE STUDY OF BEHAVIOURAL DEVELOPMENT

PLEASE, FOR ANY QUESTION, WRITE TO
rossella.caliciuri@unicatt.it