EAM 2016

UAB

Comparison of the capacity of three nonparametric person -fit indices to detect different aberrant response patterns on real data

Eduardo Doval[1], M. Dolors Riba[1],
Rebeca García-Rueda [2] & Jordi Renom[3]

[1]Departament de Psicobiologia i de Metodologia de les Ciències de la Salut.
Facultat de Psicologia, Universitat Autònoma de Barcelona.

[2]Servei de Llengües-UAB idiomes, Universitat Autònoma de Barcelona.

[1]Departament de Metodologia de les Ciències del Comportament.
Facultat de Psicologia, Universitat de Barcelona.

INTRODUCTION

- In a competence test, the total score is usually used as an indicator of the test-taker's level of competence.

- A total score is only a valid indicator as long as it follows the expected response pattern (*Evidence based on response processes*; AERA, APA & NCME, 2014).

↓

Aberrant response patterns (ARP) + Total Score

INTRODUCTION

- A variety of person-fit indices are available (Meijer and Sitjsma, 2001; Karabatsos, 2003).

- Person-fit indices may be parametric or non-parametric.

- The non parametric group-based are easy to calculate and are based on feasible assumptions: H^T, *U3*, *MCI.*

- Given a fixed Type I error rate of .05, H^T had the highest power to detect ARP, followed by *U3*, and *MCI* (Tendeiro and Meijer (2014) .

- These group-based person-fit indices outperformed parametric statistics like *lz (*Karabatsos, 2003).

AIM

To evaluate, with real data, the sensitivity of the extreme values of the indices to detected different types of ARP:

Harnisch and Linn' Modified Caution Index (*MCI*)

Van der Flier' *U3*

Sijtsma' H^T

RESEARCH METHOD

Data

Test of Basic Language Skills in Catalan.

6th grade primary students in Catalonia (2013-2014).

Participants

65.767 students (91%).

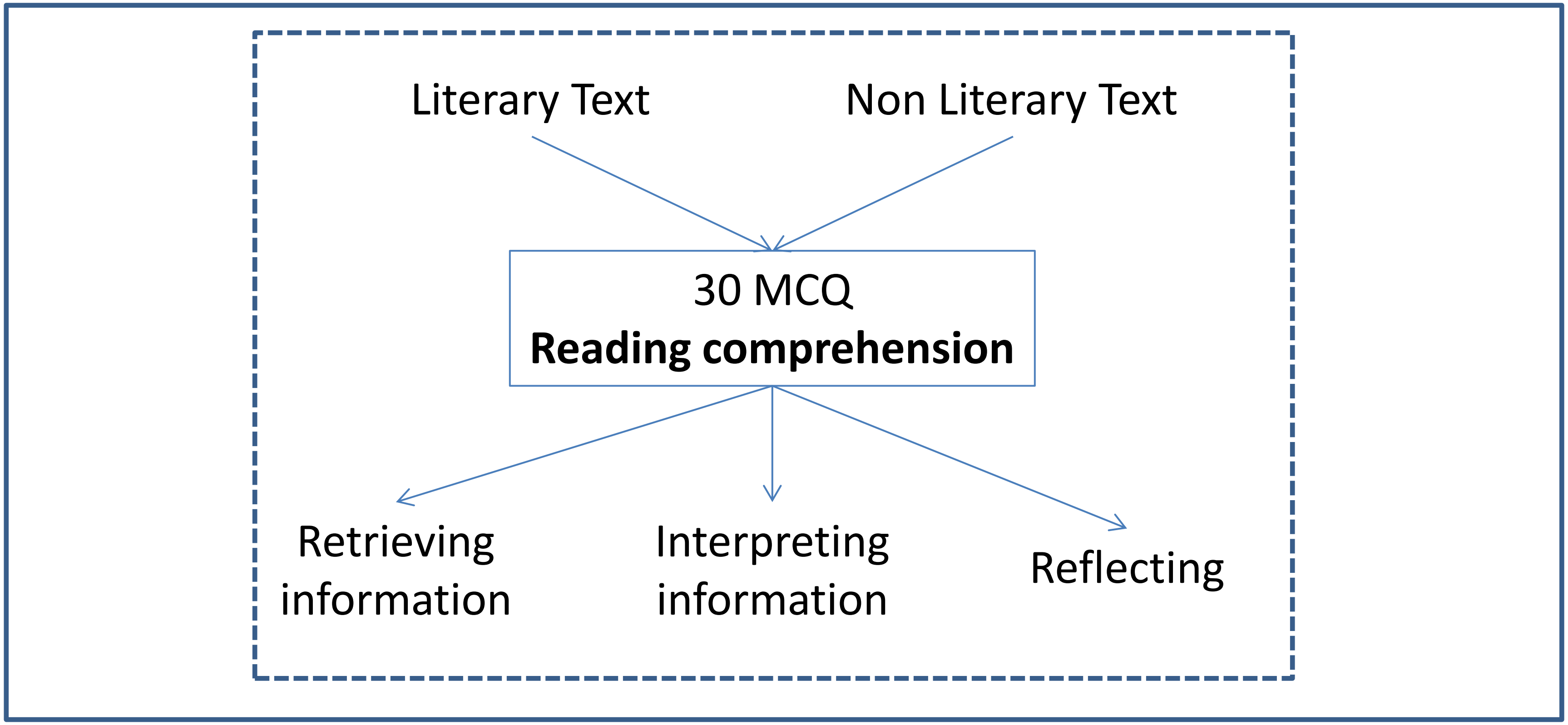
Target population

72.153 students.

RESEARCH METHOD

Test

Basic Language Skills in Catalan.



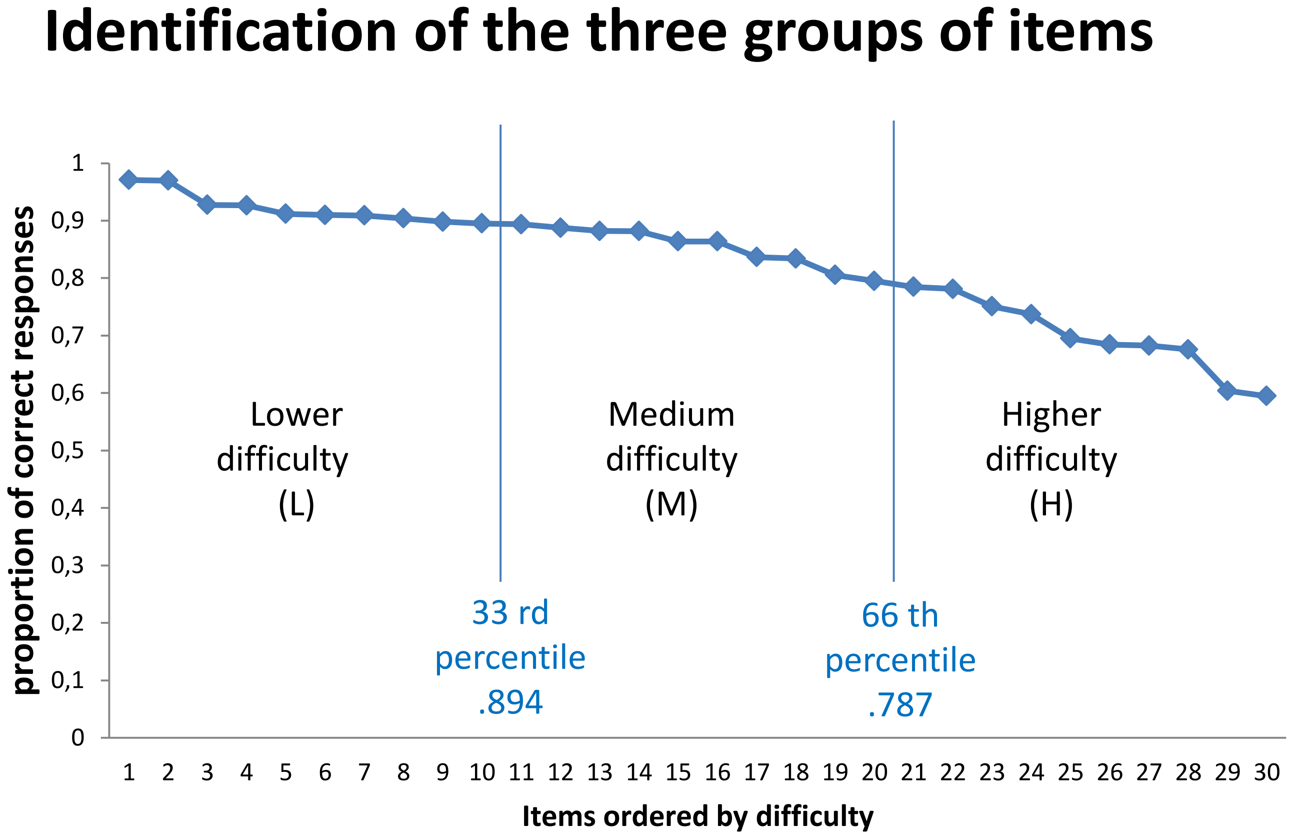
RESEARCH METHOD

Overview of the analysis

1. *MCI*, *U3* and H^T (*PerFit* R-package; Tendeiro, 2015).
2. Selection of 5% extreme cases.
3. Identification of the ARP types.
4. Comparision of % the ARP types detected by each index.

RESEARCH METHOD

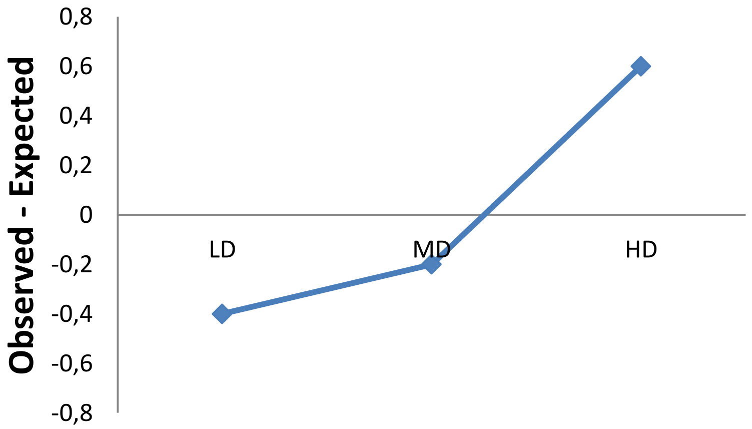
Identification of the ARP type



RESEARCH METHOD

Identification of the ARP type

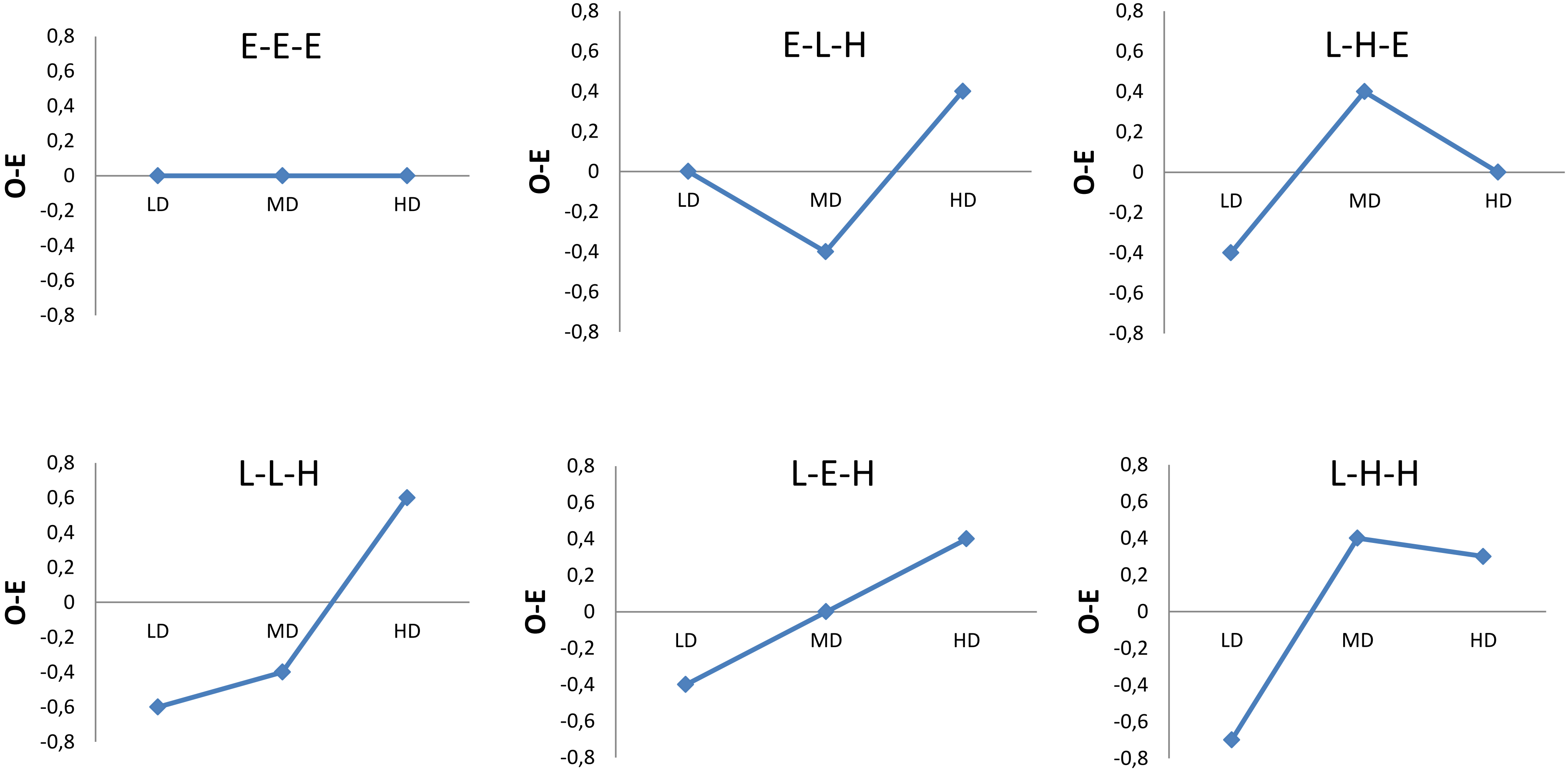
		Lower difficulty										Medium difficulty										Higher difficulty									
	Difficulty	.97	.97	.93	.93	.91	.91	.91	.9	.9	.9	.89	.89	.88	.88	.86	.86	.84	.83	.81	.8	.78	.78	.75	.74	.7	.68	.68	.68	.6	.59
Score=16	Expected	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Observed	1	0	0	1	0	0	1	1	1	1	0	1	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	0	1
Proportion of correct responses	Expected	1										0.6										0									
	Observed	0.6										0.4										0.6									
Differences (observed-expected)		Lower than expected (L)										Lower than expected (L)										Higher than expected (H)									



RESEARCH METHOD

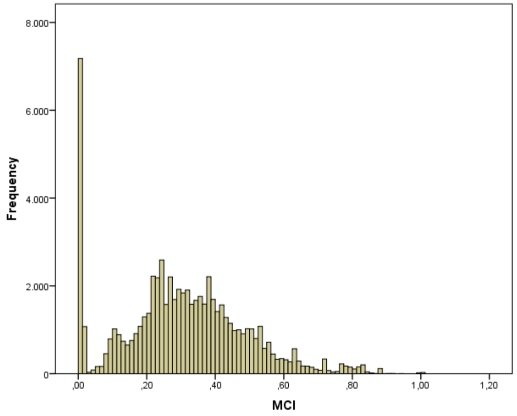
Identification of the ARP type

6 response pattern types

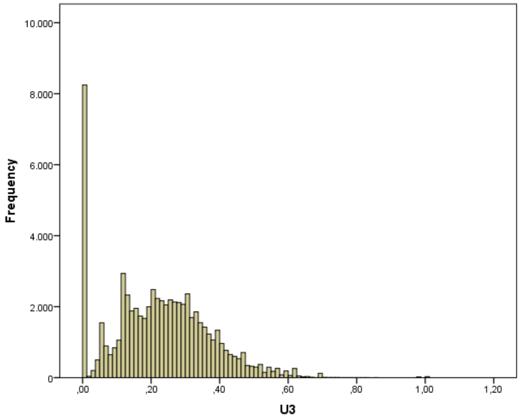


RESULTS

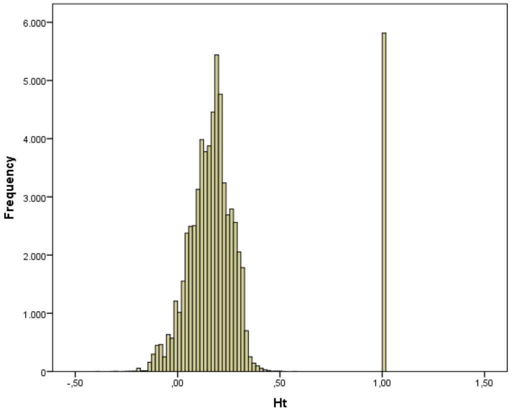
MCI



U3



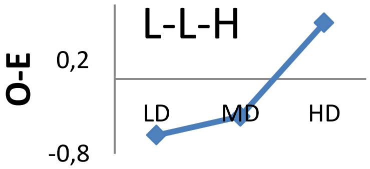
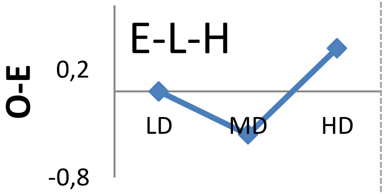
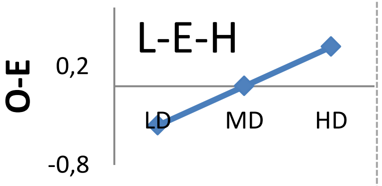
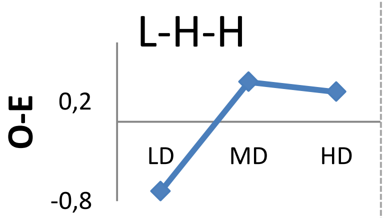
H^T



- 4,475 ARP (6.8% of the total).
- ARP detected:

  Three indices = 47,4% (2120)
  Two indices = 31.7% (1420)
  One index = 20.9% (935)

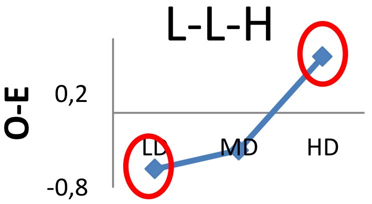
RESULTS

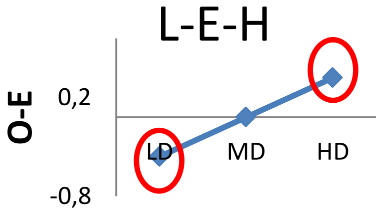
		MCI+U3+H^T	MCI+H^T	MCI+U3	H^T +U3	MCI	H^T	U3
	42.2% (1890)	48%	9.2%	0.2%	9.4%	0.4%	0.5%	32.2%
	29.5% (1319)	32.6%	64.1%	0%	0%	3.1%	0.2%	0%
	23.6% (1054)	73.8%	5.4%	1.1%	5.9%	2.6%	0.3%	13.9%
	4.7% (211)	1,9%	0%	0%	55.9%	0%	6%	36%

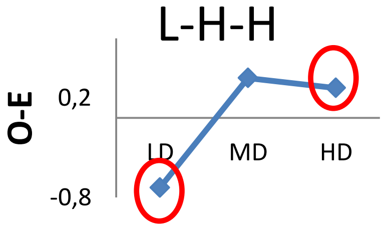
DISCUSSION

- Tendeiro and Meijer (2014) found that H^T outperformed *U3* and *MCI*.

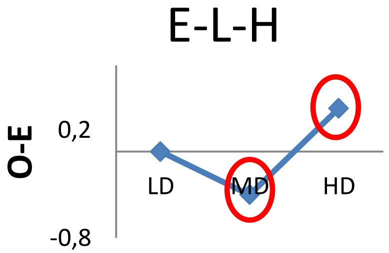
- In our study *U3* is more sensitive to patterns involving extreme difficulty values.







- H^T is more sensitive to patterns involving medium and high difficulty values.



CONCLUSION

- The use of both indices, H^T and *U3*, is recommended to detect the greatest number of ARP cases and a wider variety of ARP types.

REFERENCES

AERA, APA y NCME (2014). *The Standards for Educational and Psychological Testing*. Washington, DC: AERA.

International Test Comission. (2012). Quality Control in Scoring, Test Analysis, and Reporting of Test Scores. [www.intestcom.org]

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education 16*, 277–298.

Meijer, R.R. and Sitjsma, K. (2001). Methodology Review: Evaluating Person Fit. Applied Psychological Measurement,25(2), pp. 107–135.

Tendeiro, J.N. (2015). Package ‘PerFit’ [Sotfware]. University of Groningen. Available at http://cran.r-project.org/web/packages/PerFit

Tendeiro, J. N., & Meijer, R. R. (2014). Detection of invalid test Scores: The usefulness of simple nonparametric statistics. Journal of Educational Measurement, 51, 239-259. doi:10.1111/jedm.12046

Thank you for your attention

rebeca.garcia@uab.cat

eduardo.doval@uab.cat

dolors.riba@uab.cat

jrenompinsach@ub.edu

This work received financial support from the Dirección General de Investigación y Gestión del Plan Nacional de I+D+I, of the Spanish Ministry for Economy and Competitiveness (project EDU2013-41399-P).

We thank the "Consell d'Avaluació de Catalunya " the support given to this study.