

# PERSON-FIT STATISTICS: A USEFUL TOOL FOR FORMATIVE ASSESSMENT

**Marta Fuentes<sup>1</sup> - Eduardo Doval<sup>2</sup> - M. Dolors Riba<sup>2</sup> - Jordi Renom<sup>3</sup>**

1 Departament de Psicologia Bàsica, Evolutiva i de l'Educació. *Universitat Autònoma de Barcelona*

2 Departament de Psicobiologia i Metodologia de les Ciències de la Salut. *Universitat Autònoma de Barcelona*

3 Departament de Metodologia de les Ciències del Comportament. *Universitat de Barcelona*

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.



# Introduction

Generally, scores on a test reflect the number of correct answers. However, since not all test questions have the same characteristics, answer one or the other should not have the same evaluative value. An obvious example of these features is the difficulty.

The test score should identify the type of items that the evaluated person is able to overcome. If a person exceeds half of the test items, it is logical that these are the easiest. A pattern of response of this type would be expected.

If, however, either answers the most difficult items and fails the easiest, its pattern of response would be unexpected or unusual, and therefore would have a reasonable doubt about whether your score correctly reflects their level of knowledge or competence. This is a *atypical response pattern (ARP)*.

The International Test Commission (2012) recommends identifying patterns of response that can lead to scores of dubious validity. This type of identification allows to use formative assessment multiple choice tests, since the information to be drawn from the pattern of response can be used to identify learning styles that are not optimal.

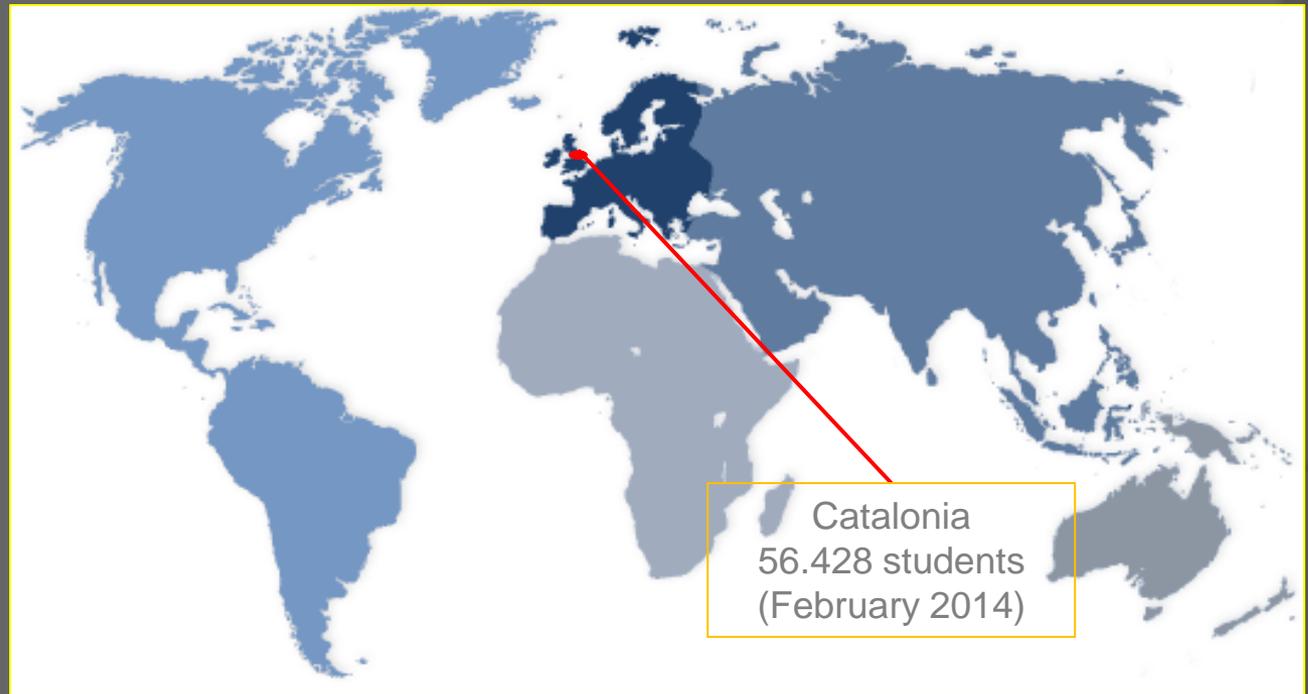
# Objectives

1. To quantify the atypical patterns of response observed in a language proficiency test.
2. To identify some of these response patterns to conclude about the kind of information that, educational level, can be drawn from them.

# Method

The data are from an institutional assessment conducted by the Department of Education of the Government of Catalonia.

The test was answered by all students of fourth course of Secondary Education of Catalonia schools.



# Method

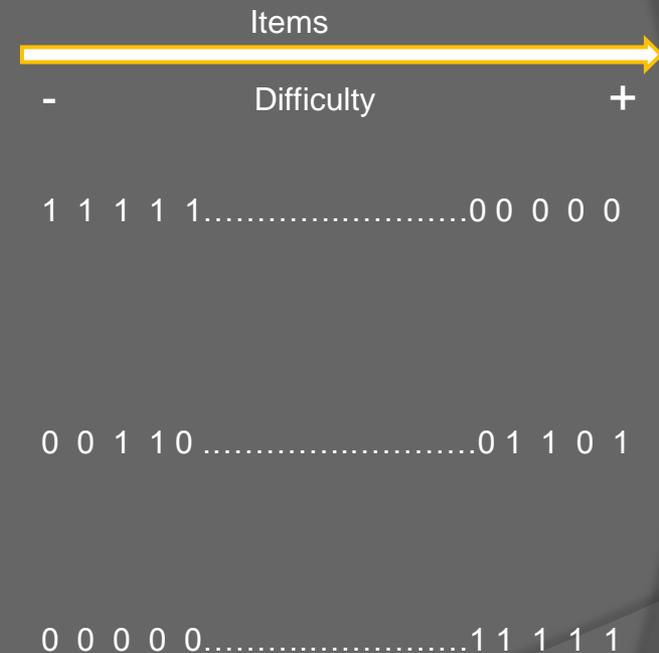
Modified Caution Index (MCI) was calculated to quantify the presence of Atypical Response Patterns (ARP). Analyses were carried out with PerFit R package (Tendeiro, 2015).

To analyze the patterns of response is necessary to order the items by difficulty.

A perfect answer pattern corresponding to a  $k$  score is one in which they respond correctly to  $k$  easiest items. A perfect pattern corresponds to a  $MCI = 0$  value.

The ARP are those that deviate from a pattern of correct answer.

A totally atypical pattern is one in which the  $k$  most difficult items are answered correctly and incorrectly answered other items. In this case the value of  $MCI = 1$ .



# Instruments

Test of communicative competence in English (more details are available at “quaderns d’avaluació 28” [http://csda.gencat.cat/ca/arees\\_d\\_actuacio/publicacions/quaderns\\_avaluacio/](http://csda.gencat.cat/ca/arees_d_actuacio/publicacions/quaderns_avaluacio/))

40 multiple choice item (3 responses categories) to assess Oral Comprehension and Reading Comprehension (Direct meaning comprehension and Inferred meaning comprehension).

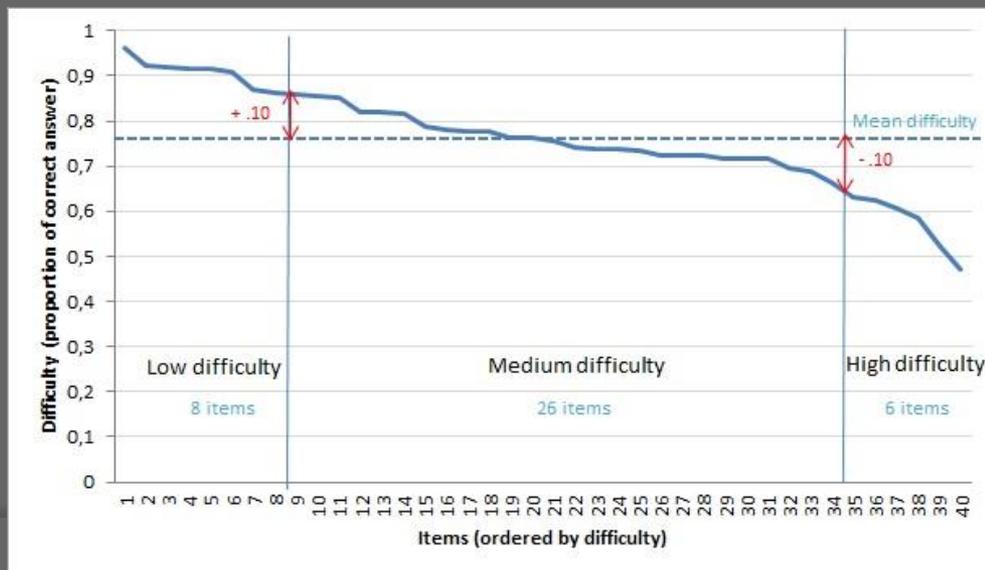
The item mean difficulty was 0.76.

To facilitate interpretation of the results, the items have been grouped into three difficulty levels :

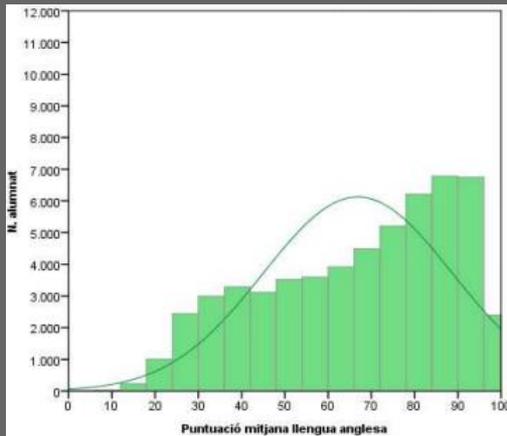
**Medium difficulty** (26 items with a difficulty index between mean difficulty  $\pm 0.10$ . Percentage of correct answers between 66% and 86%)

**Low difficulty** (8 items with a difficulty index greater than mean difficulty + 0.10. More than 86% correct)

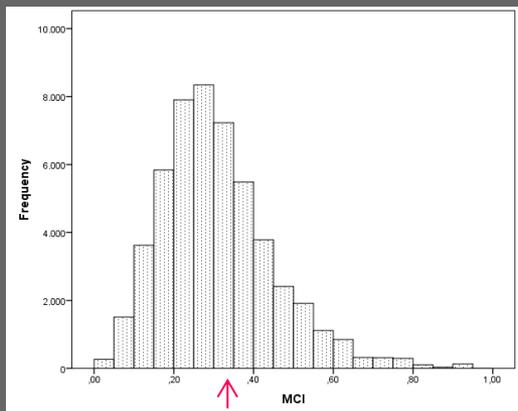
**High difficulty** (6 items with a difficulty index lower than mean difficulty - 0.10. Less than 66% correct)



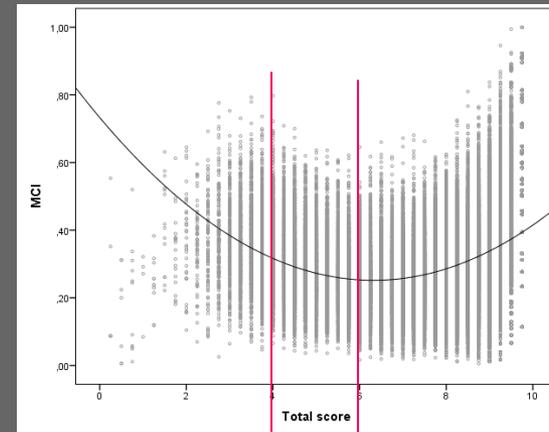
# Results



Distribution of total scores.  
17.9% with scores between 4 and 6.



Distribution values of Modified Caution Index (MCI). According to Karabatsos (2001), A MCI  $\geq 0.26$  indicates the presence of some type of ARP (59.2% at this study)



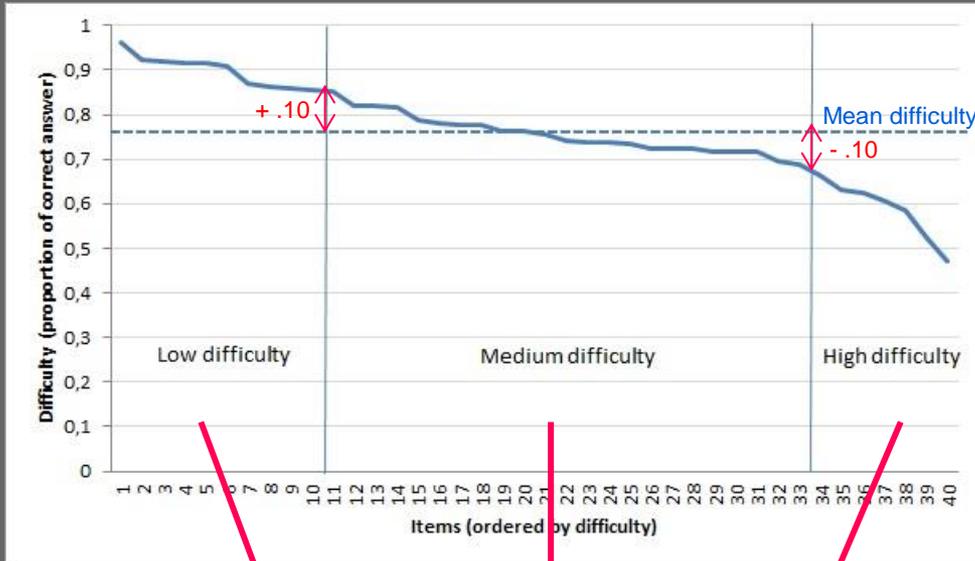
Distribution values of MCI according to the total score.

The consequences of the presence of atypical patterns in the score range between 4 and 6 are important (This segment is marked in red): These scores may underestimate or overestimate the real skill of the student.

56% of scores between 4 and 6 have some type of atypical response pattern .

# RESULTS

## Type of atypical response patterns



	Low difficulty (8 items)	Difficulty Media (26 items)	High difficulty (6 items)	MCI
Perfect pattern	100%	50%	0%	0
Case 1	87.5%	34.6%	66.7%	.386
Case 2	50%	53.9%	33.3%	.389
Case 3	37.5%	61.5%	16.7%	.480

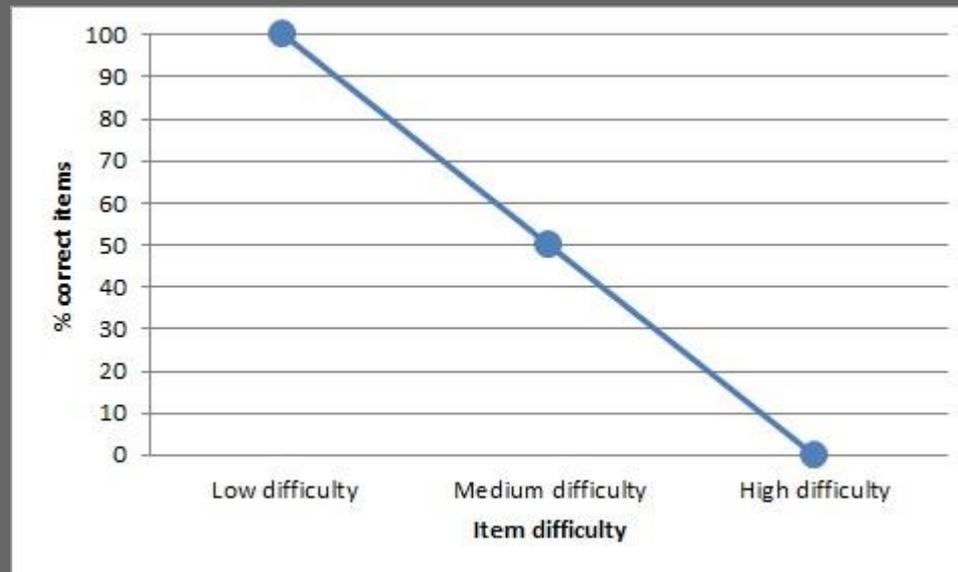
Different response patterns for people with a score of 5 out of 10 (20 out of 40)



# RESULTS

## Type of atypical response patterns

### Perfect case

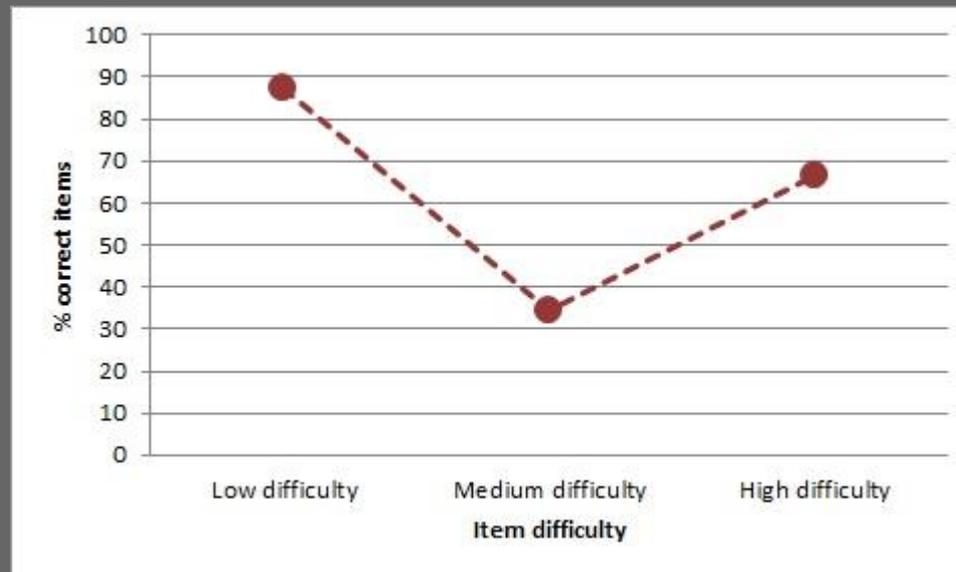


Perfect case for a pattern corresponding to 5 points over 10: correctly answered all items of low difficulty, half of the items of medium difficulty and incorrectly answered all items of high difficulty..

# RESULTS

## Type of atypical response patterns

### Case 1

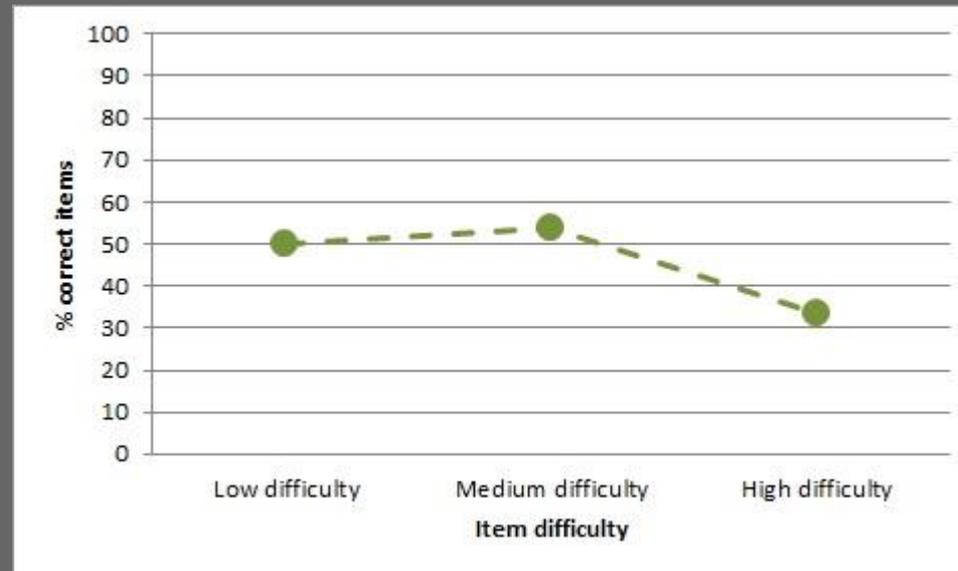


Atypical Response Pattern for a pattern corresponding to 5 points over 10: Case 1. The percentage of correct answers at the high difficulty items are higher than the expected by the perfect pattern.

# RESULTS

## Type of atypical response patterns

### Case 2

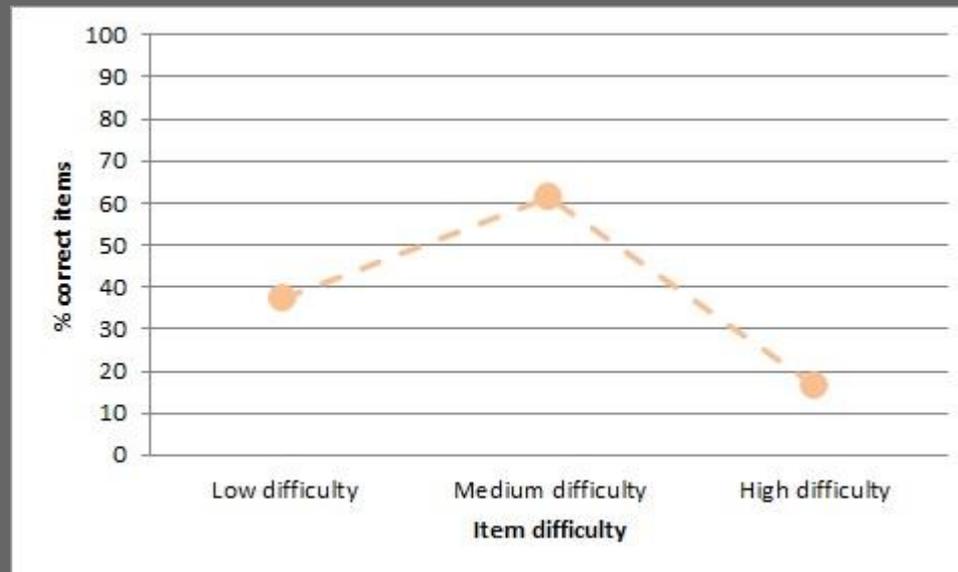


Atypical Response Pattern for a pattern corresponding to 5 points over 10: Case 2. The percentage of correct answers is quite similar in all three segments of difficulty..

# RESULTS

## Type of atypical response patterns

### Case 3



Atypical Response Pattern for a pattern corresponding to 5 points over 10: Case 3. The percentage of correct answers at items with medium difficulty is higher than items more easy, and the percentage of correct answers at items with high difficulty is considerable.

# CONCLUSIONS

In the range of scores between 4 and 6, the percentage with some kind of atypical response pattern is high (56.1%).

According to the literature (Meijer & Sitjsma, 2001), it has been possible to identify three types of ARP.

- The profile of a student responds using the random corresponds to the case 2.
- The cheater student profile is identified with the case 1.
- Case 3 is showing more interest in educational level. It corresponds to a student who seems quite capable of overcoming the items of medium difficulty and instead responds incorrectly to most items of low difficulty. This pattern of response could be due to an inadequate strategy study, and possibly have paid more attention to the central contents of the subject neglecting basic concepts. With an educational attention from the teacher aimed at promoting a change of strategy in the study of matter, this student could reach a score significantly higher and possibly more appropriate to the level that has been shown

# CONCLUSIONS

The same test score can be obtained in many different ways. Factors such as motivation toward the subject, learning style or level of focus or concentration during test execution may be possible explanations for the variability patterns.

Throughout the academic year, teachers tend to accumulate sufficient evidence of their students. The contrast of this evidence in the presence of atypical patterns in multiple choice tests can be helpful for the design of personalized learning strategies.

While multiple choice tests are used exclusively within the scope of the accreditation assessments, adding this type of analysis that complement count of correct answers is a useful tool for these tests; also can be considered as formative assessment instruments.

# REFERENCES

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

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

Tendeiro, J.N. (2015). Package 'PerFit' [Programa informático]. University of Groningen. Available in <http://cran.r-project.org/web/packages/PerFit>

**Thanks for your interest**

Contact:  
Eduardo.Doval@uab.cat  
Universitat Autònoma de Barcelona

**Edulearn 2015**