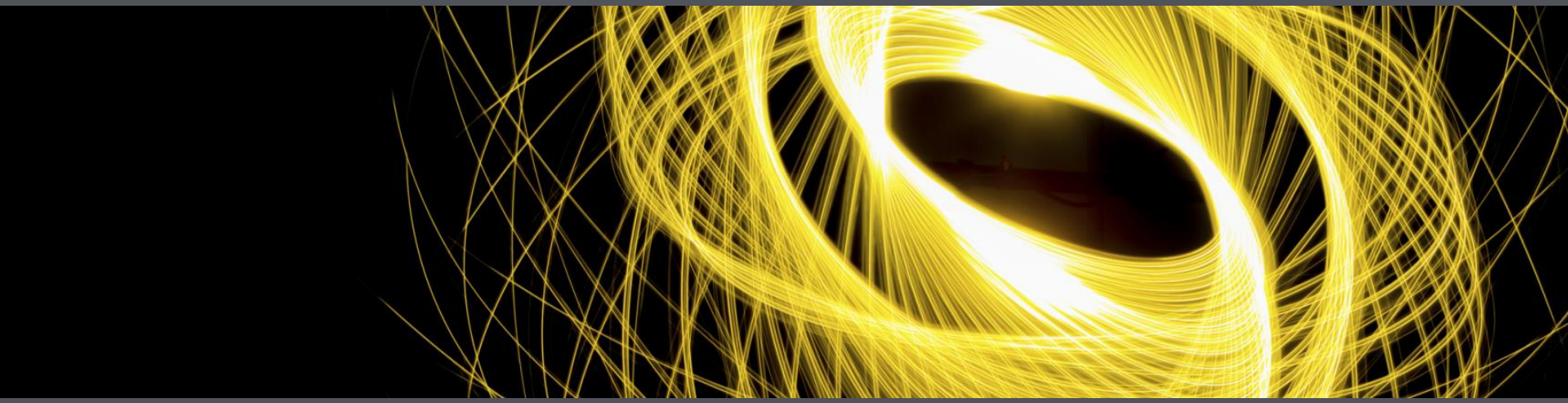
ISLI Assessment Team

University of Reading

Analysing tests and assessments using item analysis



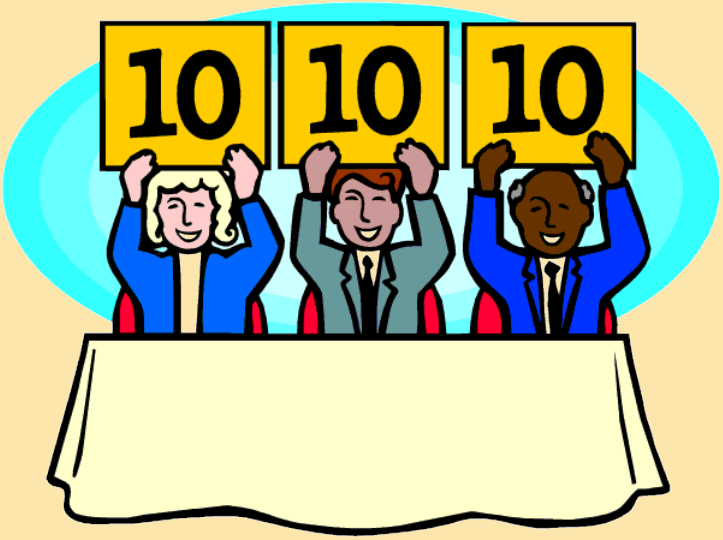
Svetlana Mazhurnaya

Phil Smyth

International Study and Language Institute (ISLI)

Judging the test

University of Reading



- When designing a test or assessment task, how do we know if the task or items on the test are working effectively?
- How do we know how difficult the items are?
- How do we know that test takers are getting the items right based on language knowledge and not something else?

Session aims

University of Reading

Key concepts in item analysis

Excel demonstration

Results interpretation

Key Concepts in Item Analysis (Classical Test Theory)

University of Reading

- Objectively marked items (no judgement)
- Reliability (consistency)
- Facility values
- Discrimination indices

Sample Items

1. Please take this MCQ test individually: https://forms.office.com/e/YfXubjMH9H
2. Identify some potentially problematic items in your breakout rooms (10 min.)


3. One of the _____ Black poets in America, Lucy Terry, was a slave in Deerfield, Massachusetts.
(A) first
(B) first was
(C) first to
(D) first has the

4. _____ engineering projects use mathematics extensively.
(A) All
(B) An
(C) Every
(D) Any

Preparing your data file

University of Reading

1. Record MCQ answers as **A-D**
2. Record dichotomous items as **1** = right, **0** = wrong, **-** missing data
3. Convert A-Ds into 1, 0, manually or using a converter here: Excel Spreadsheets for Classical Test Analysis (languagetesting.info)
4. Calculate the total scores

Reliability (Cronbach's Alpha)

- Split-half reliability
- Looking for homogeneity or internal consistency
- Reported between 0.0 and 1.0
- Above 0.7 likely acceptable (Pallant, 2007, p. 98)

Estimating overall test reliability

University of Reading

- Use the Alpha calculator here:
- Excel Spreadsheets for Classical Test Analysis (languagetesting.info)

University of Reading

Classical test theory (CTT) analysis

Facility value (FV)

Discrimination index (DI)

University of Reading

Classical test theory (CTT) analysis

Facility value (FV)

- The percentage of students who answer the item correctly (reported between 0.0 (difficult) and 1.0 (easy)
- Can also calculate the proportion of test takers who chose different distractors (distractor analysis)
- Good FV is 0.5 (widest scope for variation) (Popham, 2000).
- Acceptable range is 0.3 - 0.7 (Bachman, 2004)

University of Reading

Classical test theory (CTT) analysis

Discrimination index (DI)

- How well the item discriminates between high-scoring and low-scoring students
- 0.4 and above = very good
- 0.3 - 0.39 = reasonably good
- 0.2 - 0.29 = marginal items*
- 0.19 and below = poor items (Popham, 2000)

Task

What will you do with the analysed test items – keep/tweak/drop?

How did your initial hunches about the test items compare with the statistical analysis?

Item analysis in SPSS

Reliability Statistics

Cronbach's Alpha	N of Items
.736	13

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q1	8.70	5.959	.406	.714
Q3	8.88	6.317	.247	.736
Q4	8.63	6.907	.023	.762
Q5	8.50	6.308	.353	.721
Q8	8.30	7.190	.000	.741
Q18	8.65	5.977	.415	.713
Q19	8.60	5.733	.558	.693
Q21	8.40	6.349	.489	.711
Q26	8.38	6.292	.618	.704
Q28	8.65	6.028	.391	.716
Q31	8.43	6.456	.365	.720
Q33	8.60	6.195	.337	.723
Q34	8.90	5.682	.533	.696

Questions?

s.mazhurnaya@reading.ac.uk

phil.smyth@reading.ac.uk

References

This Photo by Unknown Author is licensed under CC BY

Alderson, J. C., Clapham, C., Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.

Bachman, L. F. (2004) *Statistical Analysis for Language Assessment.* Cambridge: Cambridge University Press.

Fulcher, G. (2023, January, 10). *Excel Spreadsheets for Classical Test Analysis*. http://languagetesting.info/statistics/excel.html

Green, R. (2013). *Statistical Analyses for Language Testers*. Palgrave Macmillan UK.

Hughes, A. (2002;1998;2010;). *Testing for Language Teachers*. Cambridge University Press.

Pallant, J. (2007). *SPSS Survival Manual*. (3rd ed.) Maidenhead. Open University Press.

Popham, W. J. (2000) Modern Educational Measurement. (3rd ed.) . Boston., Allyn & Bacon.

University of Reading

Thank you for joining the session

TAFSIG Statistical Workshop
Feedback Form