



IAVE INSTITUTO DE AVALIAÇÃO EDUCATIVA, I.P.

**Avaliar para aprender:
construção de instrumentos de
avaliação**

Tarefas 6. e 7.

AprofGeo

Outubro/Novembro de 2022

1

Tarefas das sessões síncronas e assíncronas

Tarefas 6. e 7.

Análise e interpretação de resultados:

- Noções de psicometria: índices de dificuldade e de discriminação dos itens
- Validade e fiabilidade
- O papel do erro na análise e interpretação dos resultados

Demonstração da construção de itens em formato digital

2

Na interpretação dos resultados de um teste, não se pode ignorar

- o que, de facto, o teste avalia
- a finalidade com que o teste foi construído (aferir as aprendizagens ou certificar?)
- o tipo e a qualidade da informação que se pretende devolver com base nos resultados

3

Uso dos resultados

- A divulgação dos resultados de um teste seria mais útil se fosse além da mera publicação de classificações numéricas.
- A classificação numérica oculta os aspetos onde o aluno falhou e onde teve sucesso.

Cronbach (2002 [1963])

- Com critérios construídos especificamente para o efeito e um sistema de registo adequado, a análise das respostas dos alunos permite a identificação de fragilidades na aprendizagem.

*Intervenção
atempada face
à identificação
de fragilidades*

4

A forma e o conteúdo dos critérios e o sistema de registo são determinantes para a quantidade e qualidade da informação a recolher e, consequentemente, para a qualidade do *feedback* que a informação proporciona.

5

Características de um <i>feedback</i> de qualidade (a utilizar na formulação dos RIPA ou nos relatórios)	
Foco	O trabalho, o desempenho ou o processo, nunca o aluno.
Conteúdo	Específico, relacionado com os objetivos de aprendizagem.
Modo	Claro e sem ambiguidades, utilizando uma linguagem descritiva e recorrendo às qualidades [descritivas] do desempenho.
Comparação	Os critérios ou o progresso.
Função	Descrever (sem julgar) o desempenho por comparação com a descrição dos critérios ou objetivos.
Valência	Positiva. A descrição dos aspetos negativos do trabalho deve ser acompanhada por sugestões positivas de melhoria.

6

Análise dos resultados

Noções básicas de psicometria

7

Dificuldade de um item

Índice de dificuldade

- É uma medida da taxa de sucesso num item
- Traduz a proporção de acerto no item
- Assume valores numa escala de 0 a 1
- Quanto menor é o índice, maior é a dificuldade do item



8

Escala de graus de dificuldade

Índice	Grau de dificuldade
menor ou igual a 0,20	Muito difícil
]0,20 ; 0,40]	Difícil
]0,40 ; 0,60]	Médio
]0,60 ; 0,80]	Fácil
maior do que 0,80	Muito fácil

A “percentagem de acerto” não é mais do que o índice de dificuldade em forma de percentagem

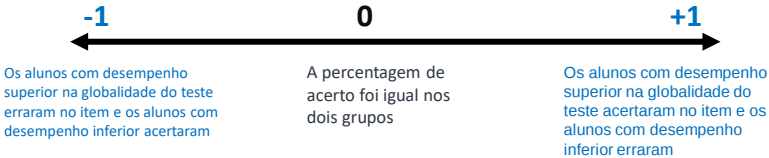
IAVE, 2015

9

Discriminação de um item

Índice de discriminação

- Mede a diferença entre a proporção de acerto no item por parte do grupo de alunos com desempenho superior e a proporção de acerto no item por parte do grupo de alunos com desempenho inferior no teste
- Traduz o grau de consistência entre o sucesso no item e o sucesso no teste como um todo
- Assume valores numa escala de -1 a +1
- Quanto maior é o índice, maior é o poder de discriminação do item



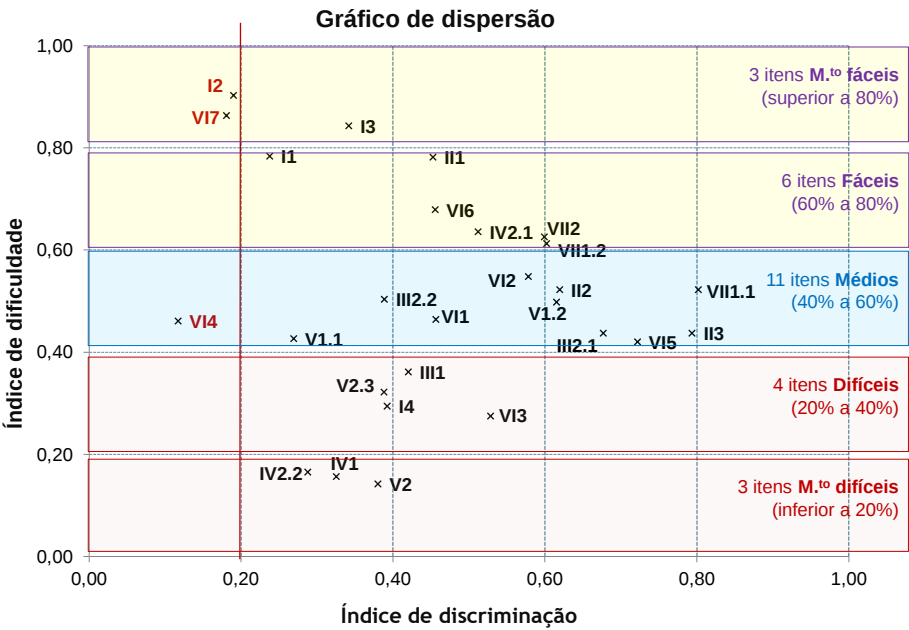
10

Escala de graus de discriminação

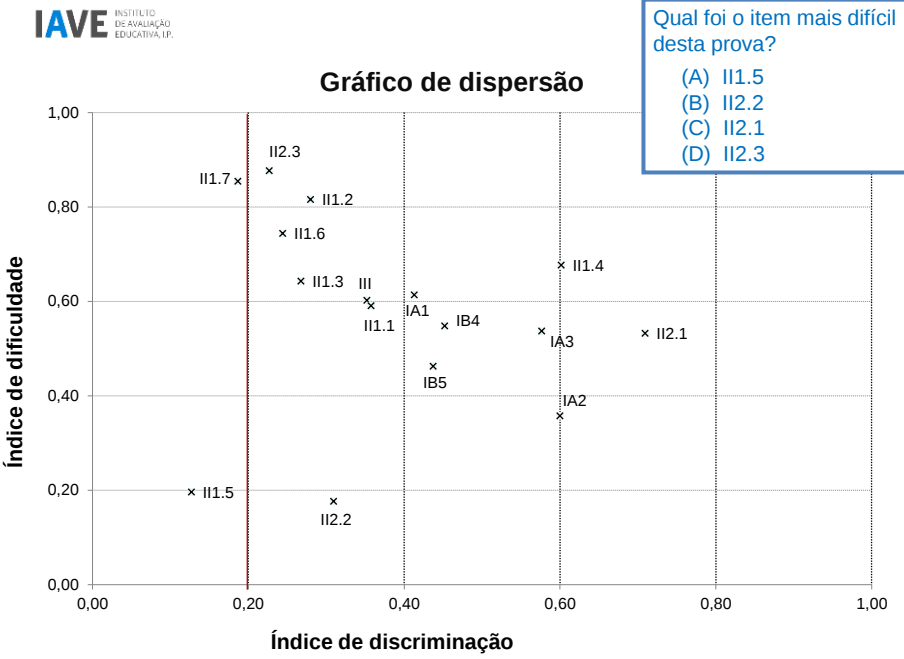
Índice	Grau de discriminação
Superior a 0,40	Elevado
0,20 a 0,40	Moderado
Inferior a 0,20	Baixo (item a rever ou eliminar)

Kelley, Ebel e Linacre, 2002

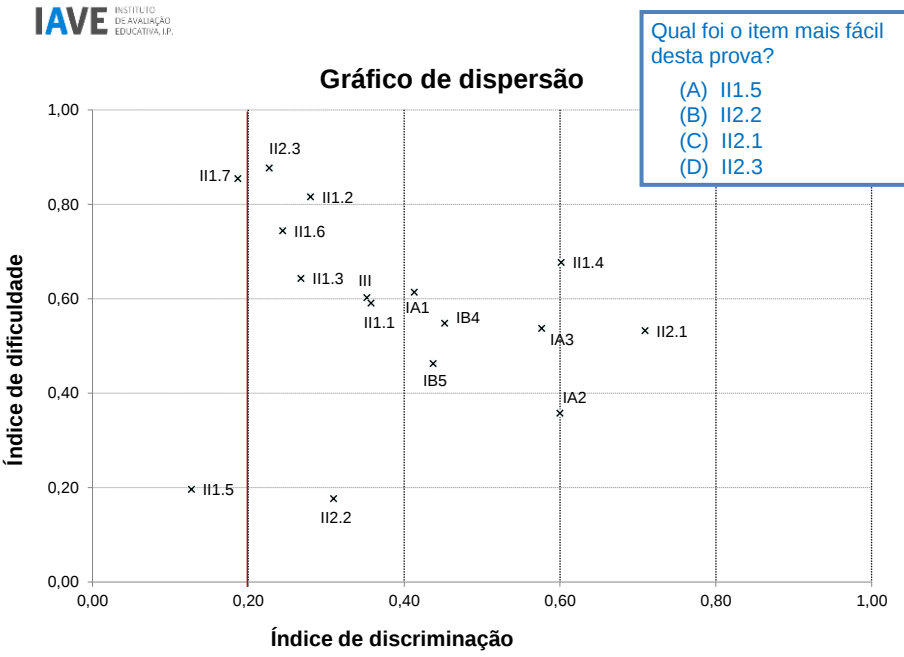
11



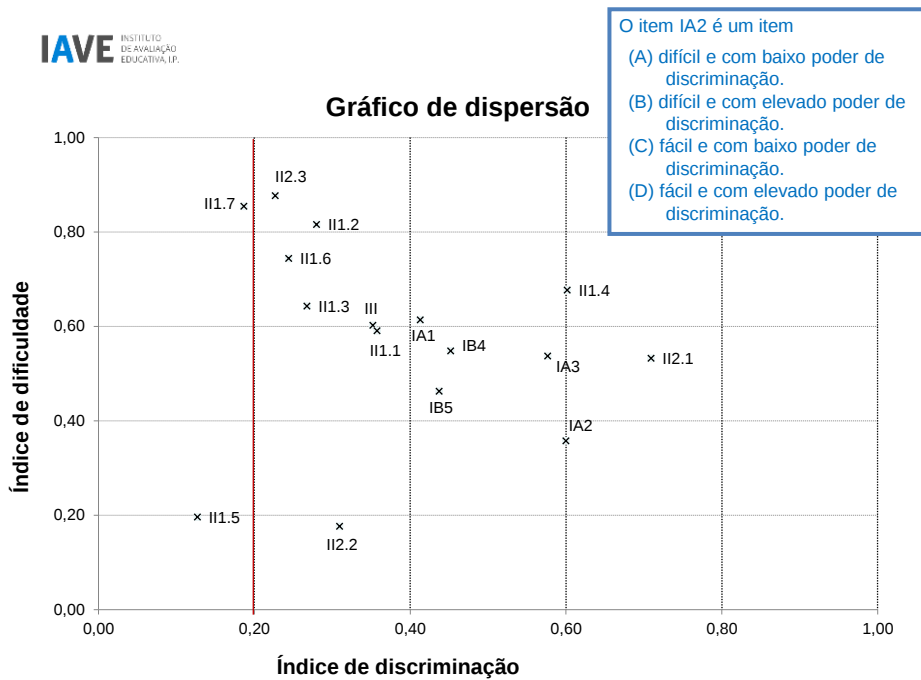
12



13



14



15

Análise dos parâmetros psicométricos

Itens com comportamentos não esperados

- demasiado fáceis
- demasiado difíceis
- que não discriminam
- com discriminação negativa

Distribuição dos itens pelos graus de dificuldade

- ...

Itens com problemas de construção?

16

Análise dos parâmetros psicométricos

Um item pode ser **fácil** porque

- as opções incorretas são pouco plausíveis
- a maior parte dos alunos desenvolveu a aprendizagem
- ...

Itens fáceis

Um item pode ser **difícil** porque

- as opções incorretas são plausíveis
- há mais do que uma resposta correta
- o conteúdo é raro ou trivial
- a tarefa não está claramente definida
- ...

Itens difíceis

17

Análise dos parâmetros psicométricos

Um item pode não **discriminar** porque

- é muito fácil
- é muito difícil
- a opção 'correta' tem falhas
- a opção correta é demasiado óbvia
- a tarefa é muito difícil e os alunos responderam ao acaso
- requer competências que não são consistentes com as avaliadas pelos restantes itens do teste
- ...

*Itens que
não
discriminam*

18

Complexidade vs Dificuldade

Não esquecer que:

A **complexidade** de um item depende dos processos cognitivos requeridos pela sua resolução

A **dificuldade** de um item é uma medida estatística que depende dos resultados obtidos

Withers, 2005

19

Análise psicométrica dos resultados

20

- Assessment Reform Group [ARG] (2002a). *Assessment for Learning: 10 principles*. Cambridge: University of Cambridge Faculty of Education. <http://www.aaia.org.uk/afl/assessment-reform-group/>
- Black, P., & William, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Brookhart, S. (2017) How to use grading to improve learning . Alexandria (USA): ASCD
- Brookhart, S. (2013). *How to create and use rubrics*. Alexandria (USA): ASCD
- Brookhart, S. (2008) How to give Effective Feedback to your students Alexandria (USA): ASCD
- Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational assessment*. London: The Falmer Press.
- Harlen, W. (2006). On the Relationship Between Assessment for Formative and Summative Purposes. In J. Gardner (Ed.) *Assessment and Learning*. (pp. 103-117) Londres: SAGE Publications.
- Stufflebeam, D., & Shinkfield, A. (2007). *Evaluation Theory, Models & Applications*. S. Francisco: Jossey-Bass.
- Izard, J. (2005). *Overview of test construction*, Module 6. Paris: UNESCO. <http://www.unesco.org/iiep>
- Kelley, T., Ebel, R. & Linacre, J. (2002). Item discrimination indices. *Rasch Measurement Transactions*, 16(3), 883-4. <http://www.rasch.org/rmt/rmt163.pdf>
- Leong, S. C. (2006). *On varying the difficulty of test items*. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore.
- Miller, M., Linn, R. & Gronlund, N. (1995, 2009). *Measurement and assessment in teaching* (10th edition). New Jersey: Pearson
- Wiggins, G. (2014). *Assessment grading and rigor*. <https://grantwiggins.wordpress.com/2014/08/07/assessment-grading-and-rigor-toward-common-sense-and-predictable-outcomes-on-tests/>
- Wiggins, G. (1998)). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. EUA: Jossey-Bass.