



GOVERNO DE
PORTUGAL

MINISTÉRIO DA EDUCAÇÃO
E CIÊNCIA

IAVE INSTITUTO
DE AVALIAÇÃO
EDUCATIVA, I.P.

Exames Nacionais: da conceção ao resultado

Jornadas de Reflexão “Metas, Programas
e Exames Nacionais de Biologia e Geologia
– Tendências e Perspetivas”

Helder Diniz de Sousa
11 Outubro 2014

Equipas de autoria de provas

- Constituição e funções
- Seleção/recrutamento

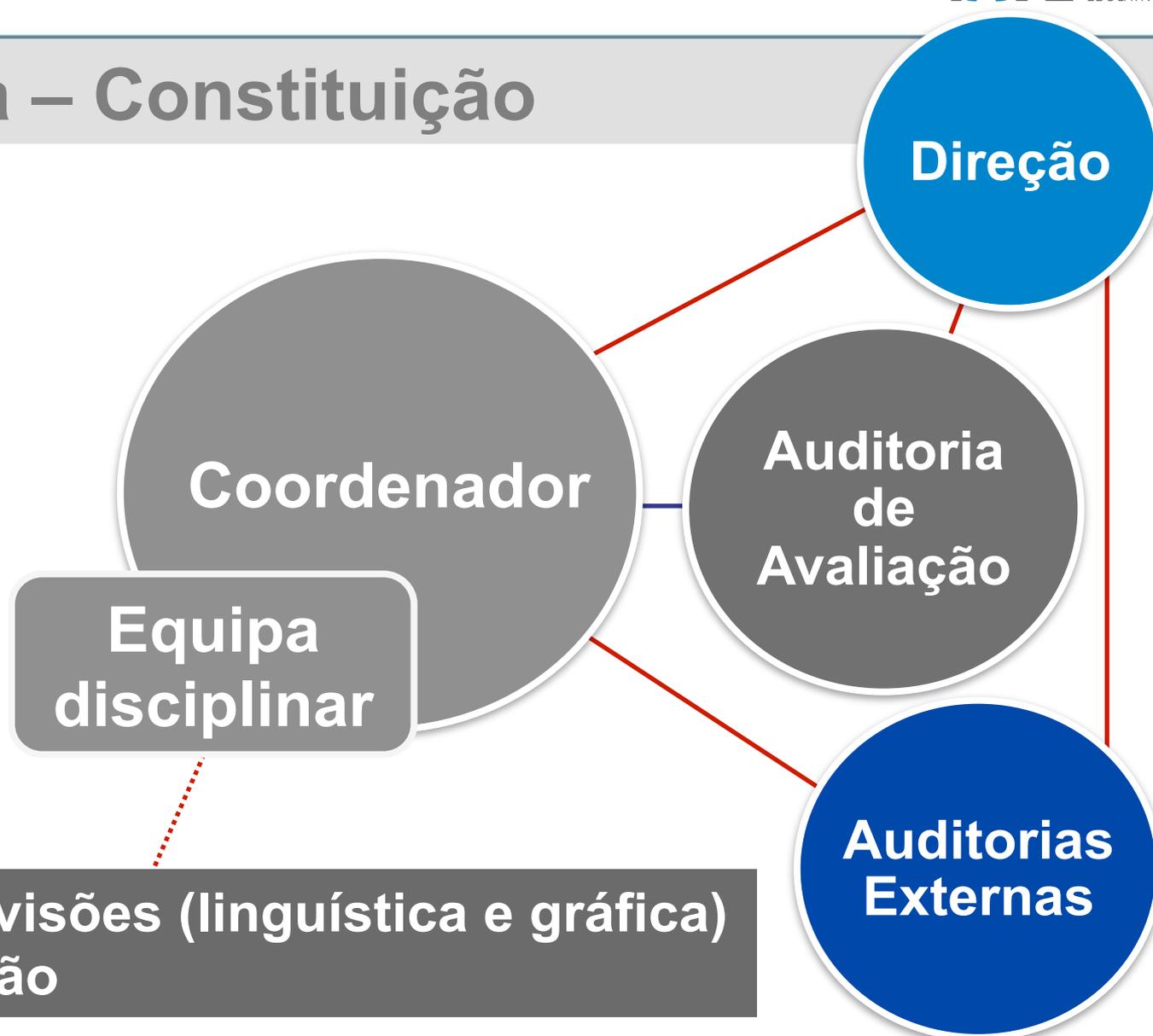
Conceção das provas

- Concetualização
- Workflow* do processo de conceção das provas

Resultados

- Validade, leitura, interpretação e uso
-

Equipa – Constituição



Outras revisões (linguística e gráfica)
Formatação

Equipa disciplinar

Coordenador

Autores

Consultores

Audidores de especialidade

Equipa – Recrutamento

Critérios de seleção (formais e informais)

- Experiência profissional
- Competência técnica e científica
- Ética profissional
- Espírito crítico e humildade intelectual
- Capacidade para trabalhar em equipa
- Capacidade de liderança (coordenador)
- Domínio dos conceitos-chave em avaliação

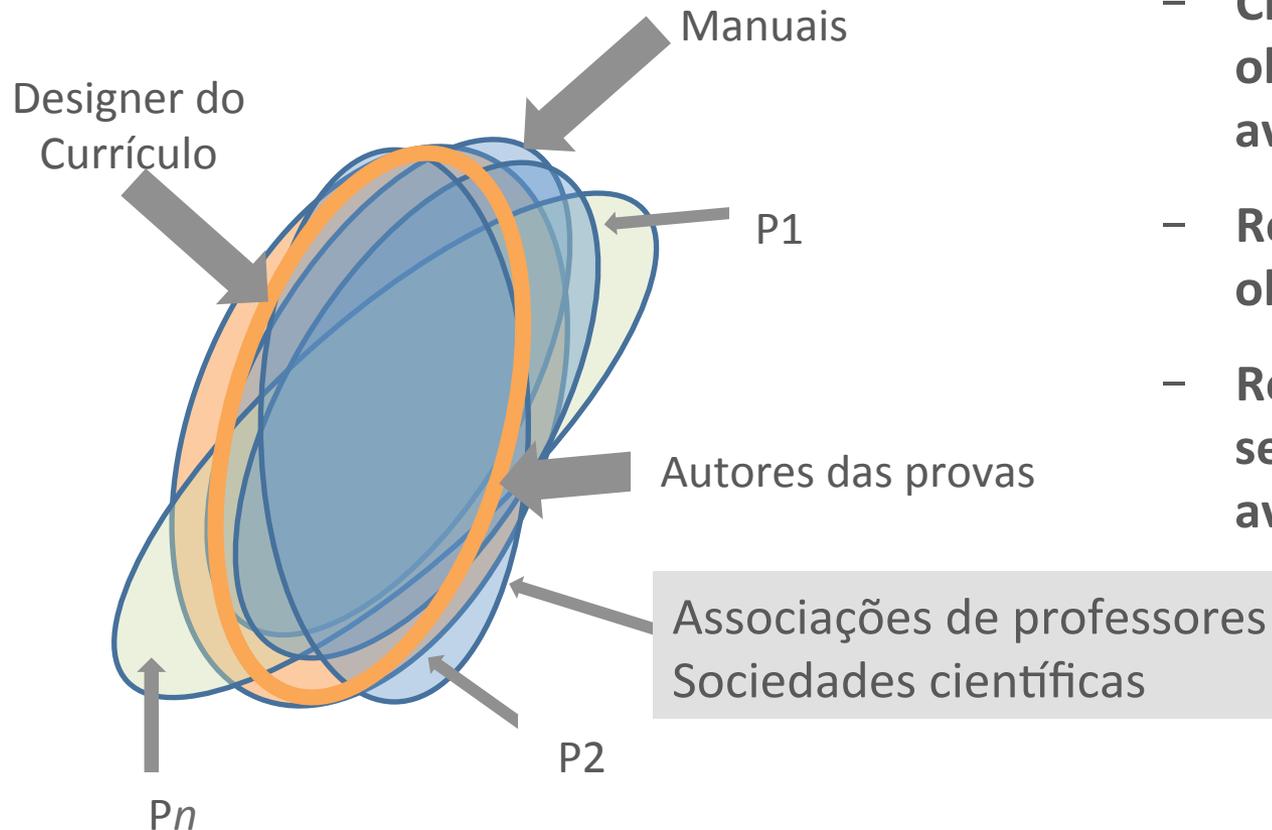
Seleção de colaboradores

Processo de formação interna

Importância da manutenção e estabilidade das equipas

Concetualização de uma prova

A(s) leitura(s) do programa e a definição do objeto



- Clareza dos objetivos da avaliação?
- Relevância do objeto a avaliar?
- Relação entre o que se ensina e o que se avalia?

A produção de uma Informação-Exame

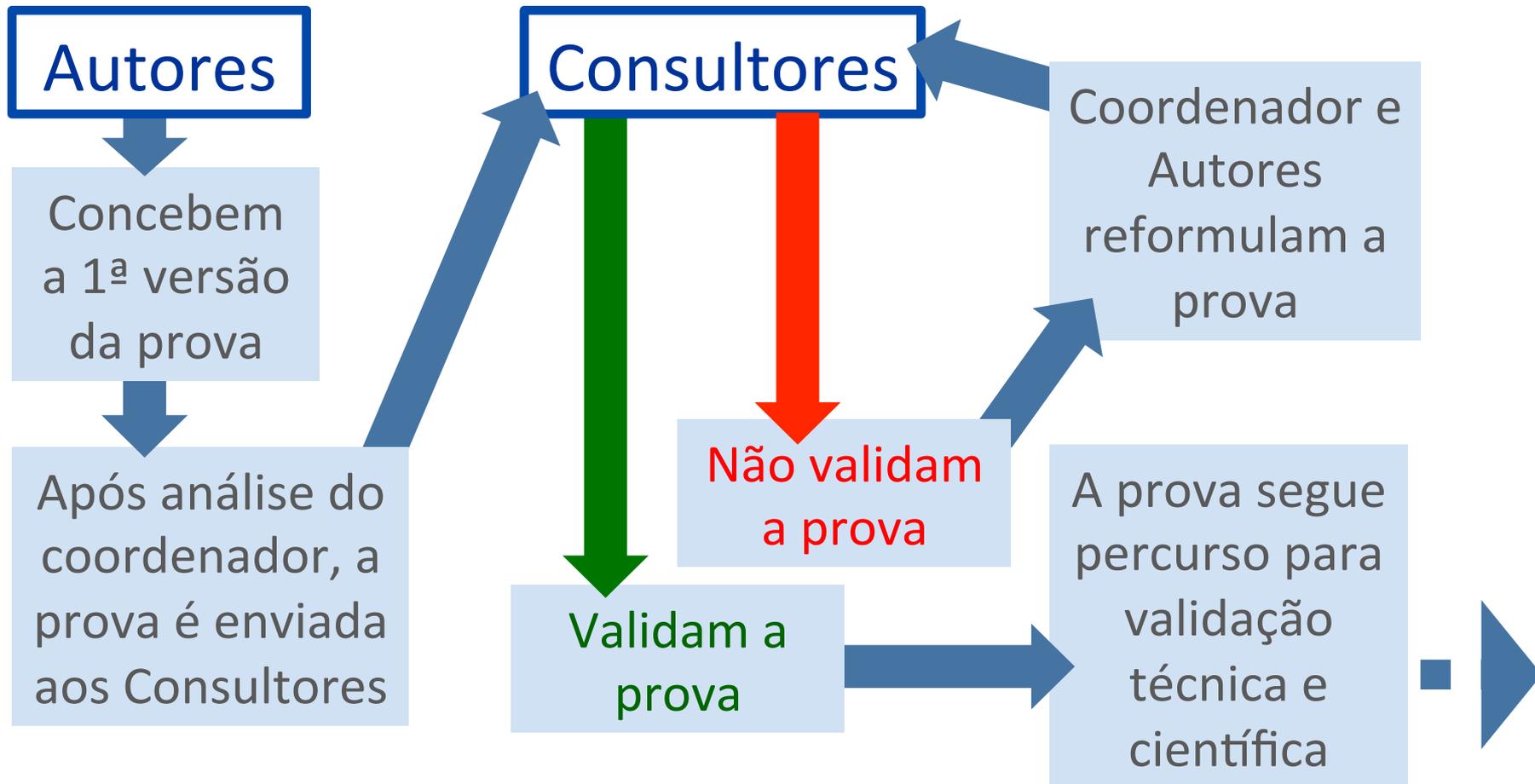
A construção de uma matriz geral e específica

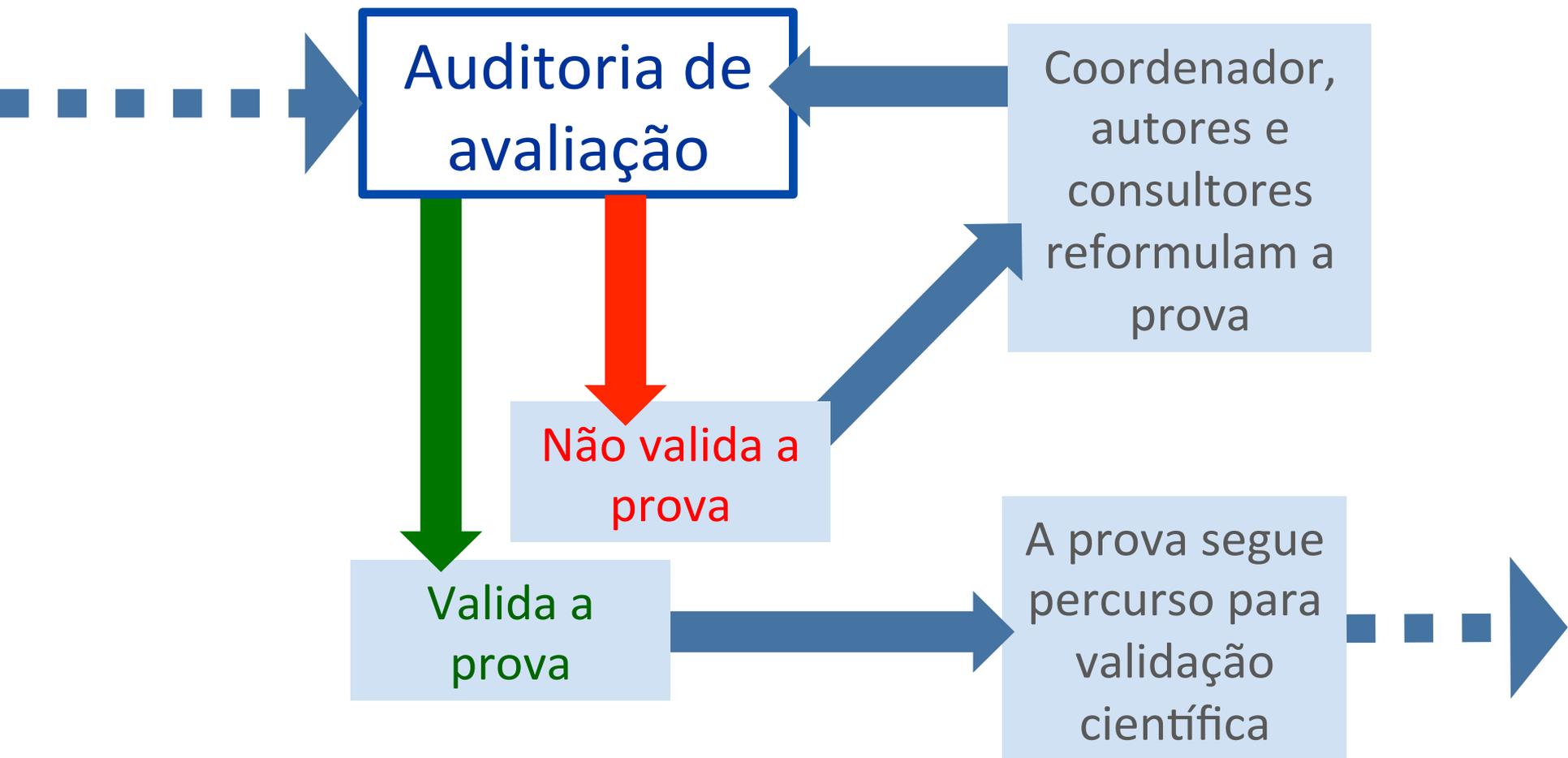
**A análise do histórico de resultados por item e a
previsão de resultados**

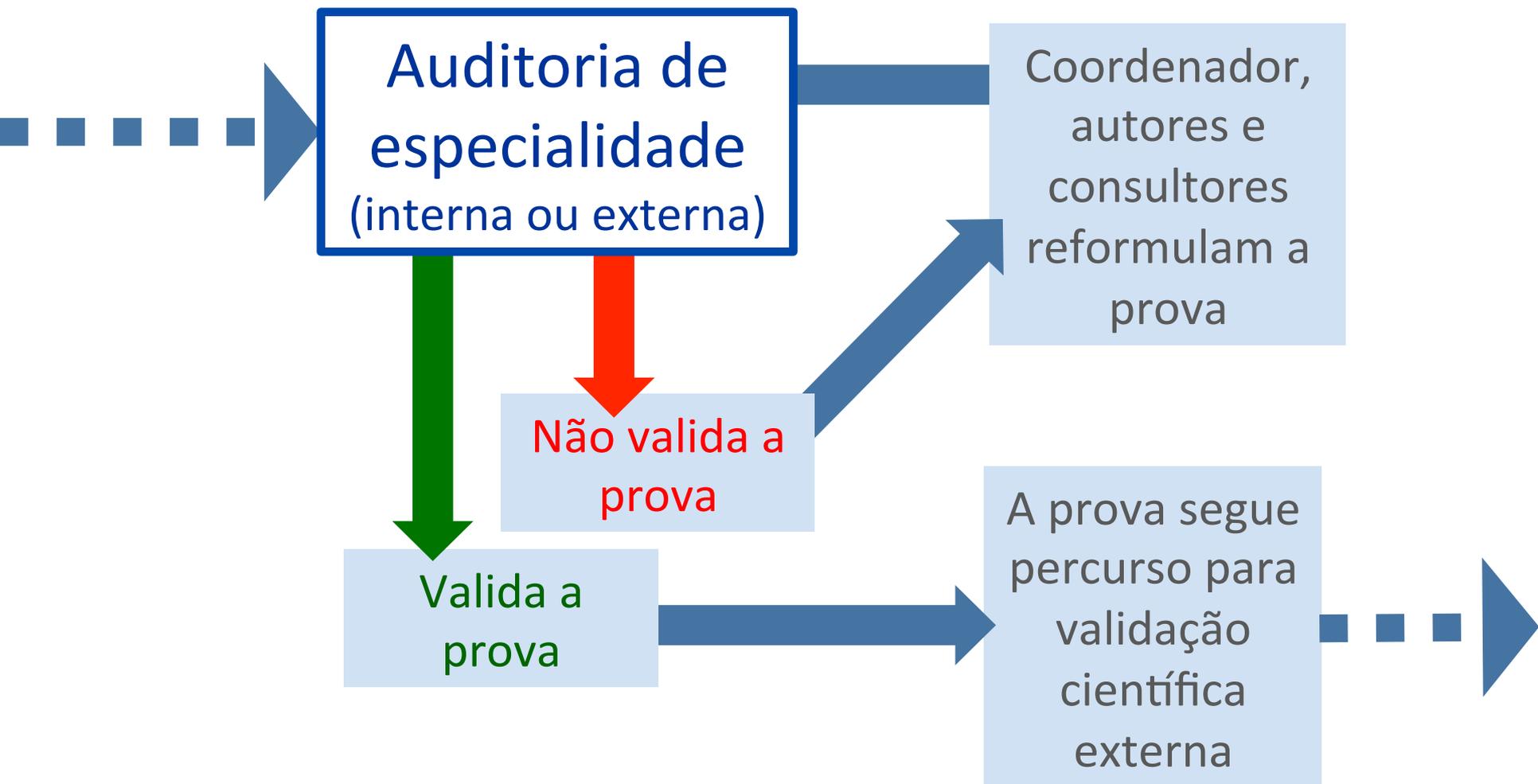
A seleção de materiais de suporte

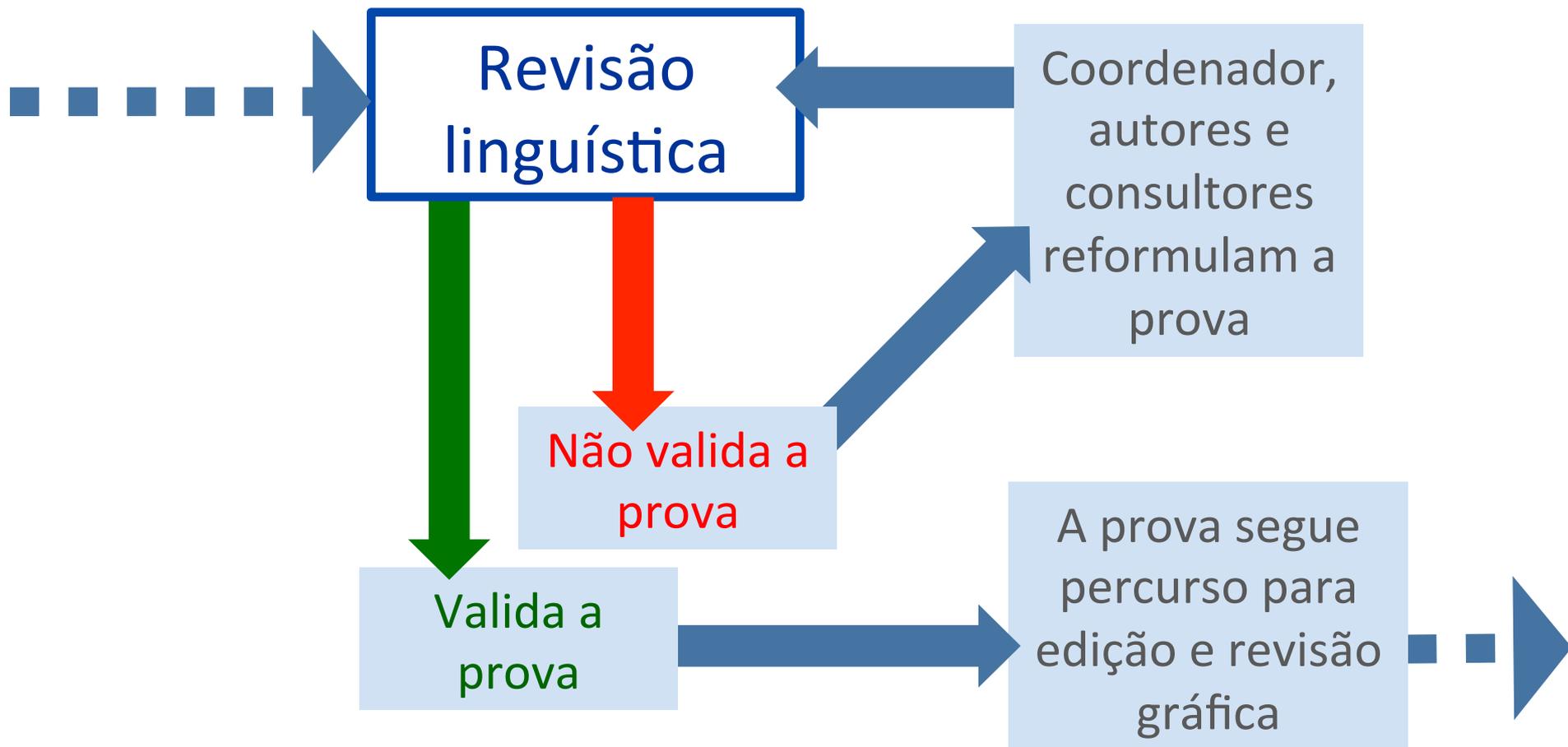
A originalidade vs previsibilidade da prova

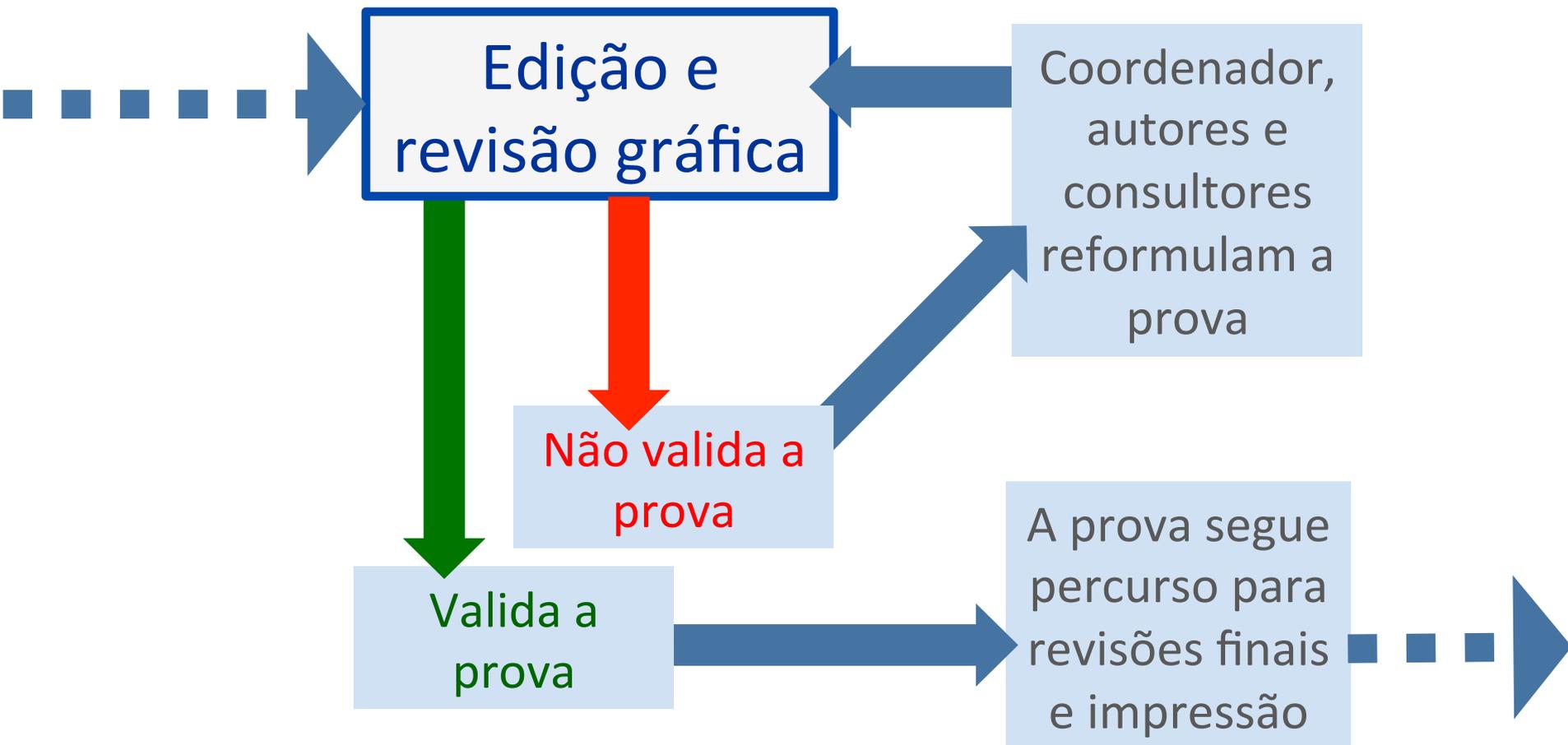
Workflow da conceção de uma prova



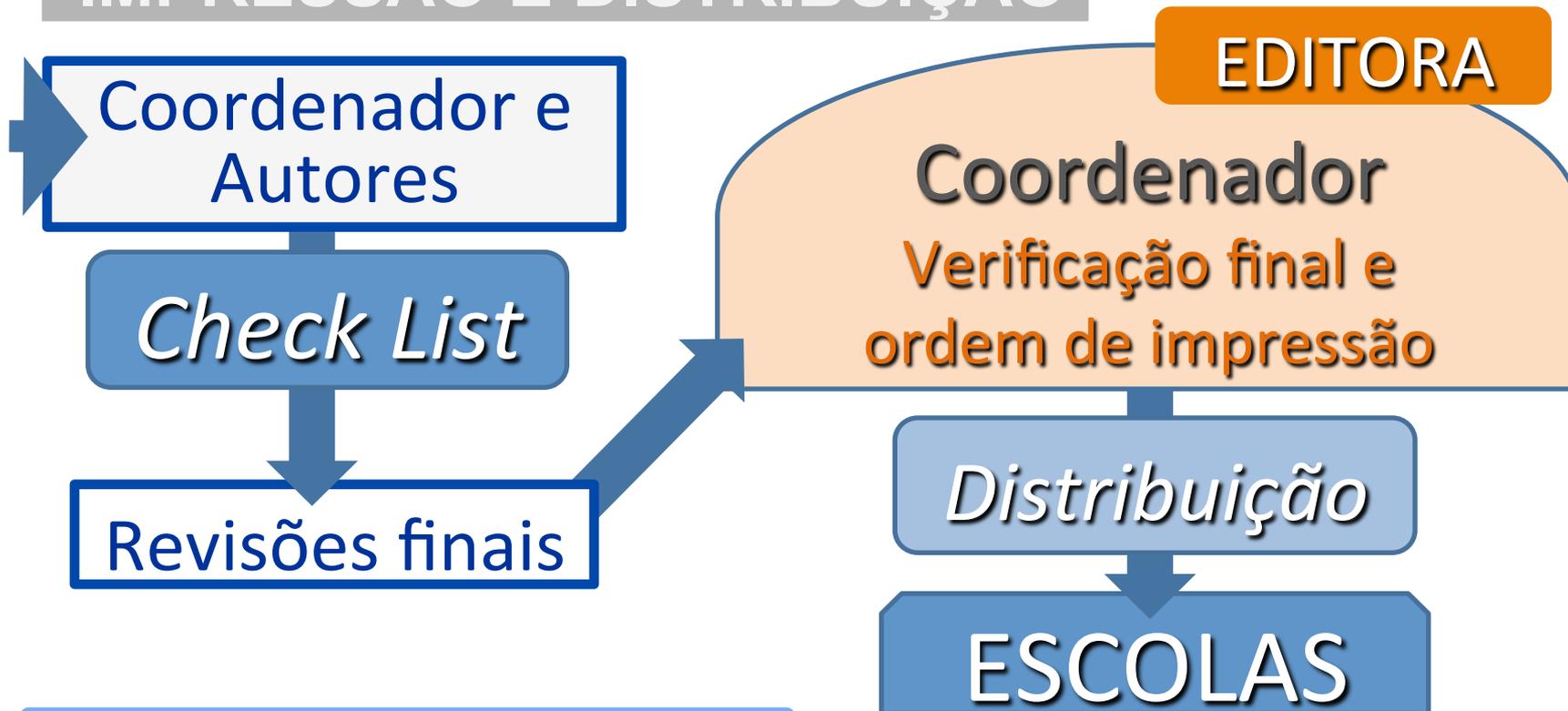








IMPRESSÃO E DISTRIBUIÇÃO



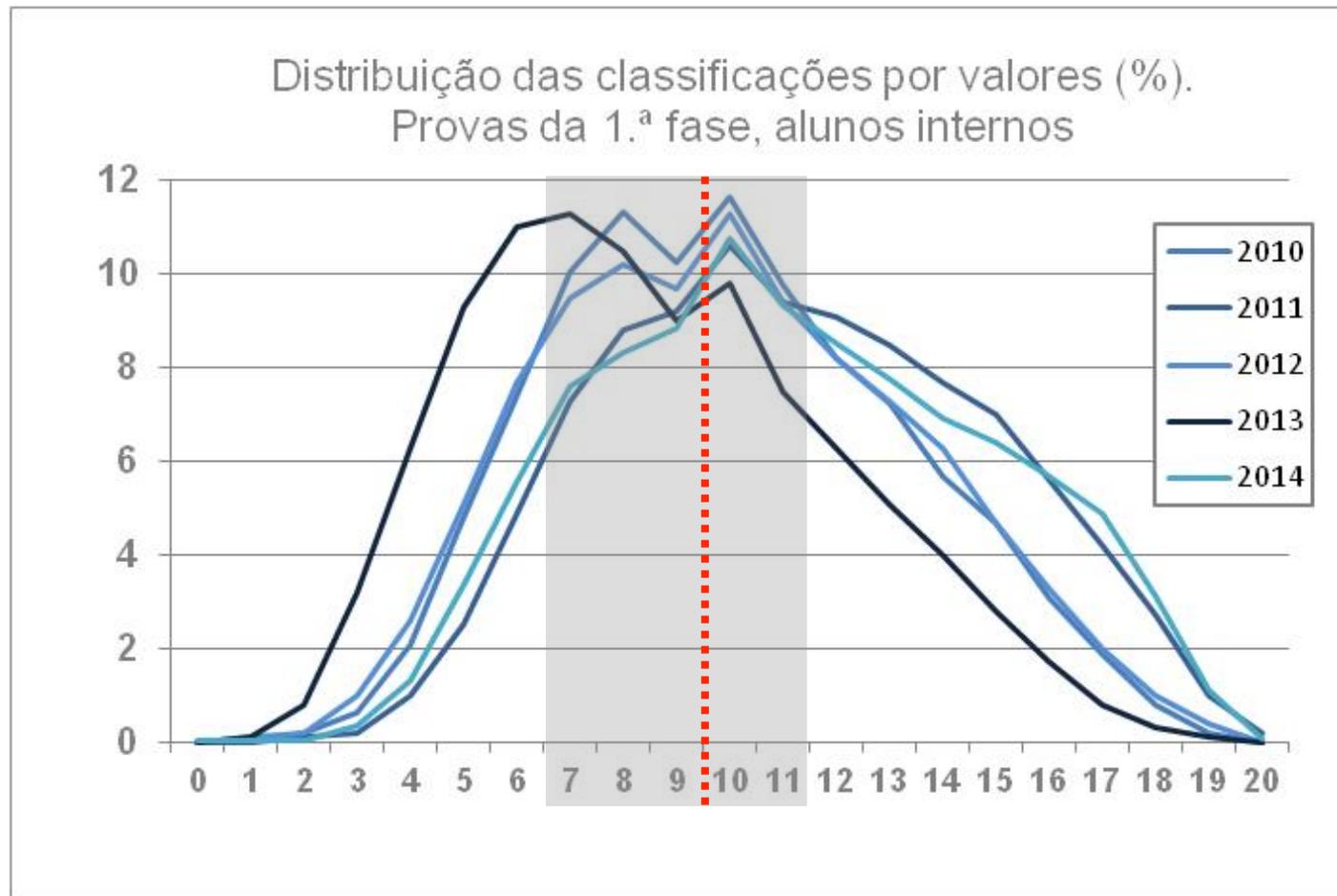
Tempo médio de execução de uma prova
14 a 16 semanas

Resultados – Validade, leitura, interpretação e uso

Não sendo a conceção de exames uma ciência exata, a validade dos exames enquanto instrumentos de medida depende também da forma como são lidos os resultados.

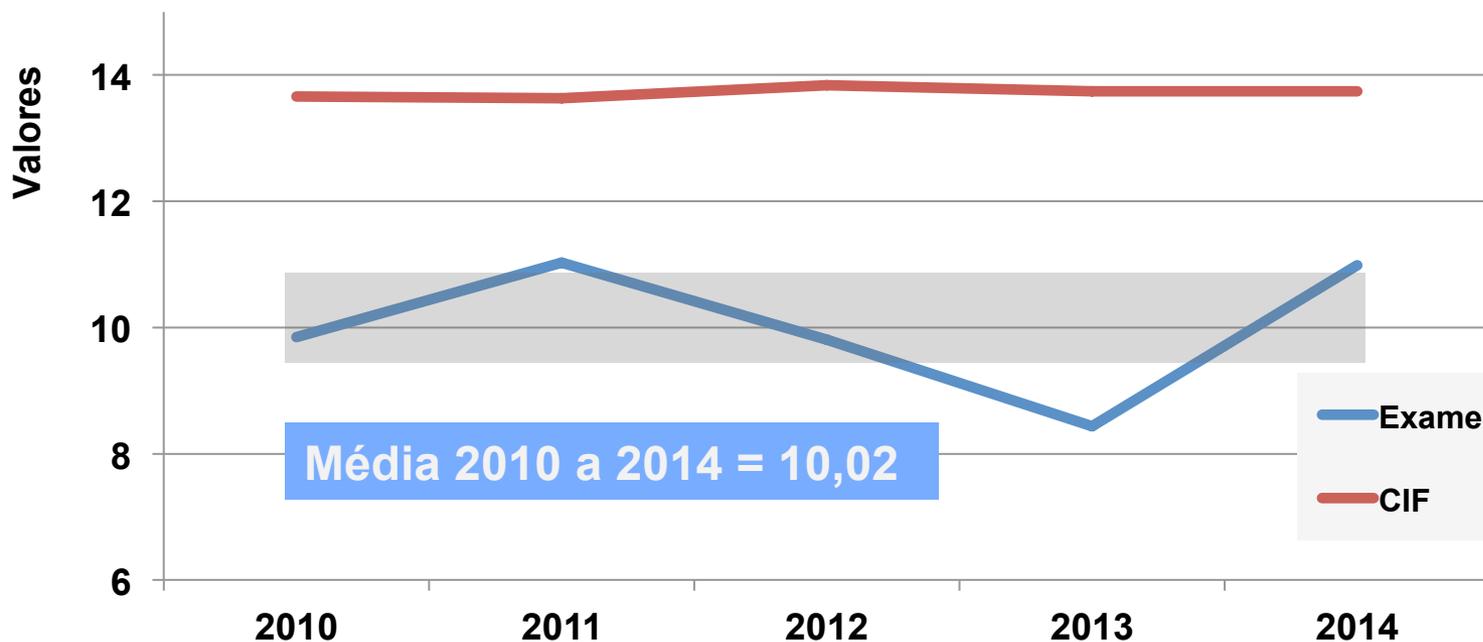
Segundo Gordon Stobart, **“um teste bem construído torna-se inválido se os resultados são mal entendidos ou interpretados de forma errada”**.

A normalidade da «anormal» curva de distribuição normal



A variação interanual de resultados e a sua relação com as classificações internas de frequência (CIF)

Evolução dos resultados – CIF e CE



As leituras e os interesses dos *stakeholders*: algumas considerações

Resultados globais (macro):

- significados da variabilidade interanual (o resultado da prova é a soma dos resultados dos itens)
- para que servem e a quem interessa a sua divulgação

A problemática da «comparabilidade» das provas a partir da leitura dos resultados globais

- o que são «provas fáceis» ou «provas difíceis»?
- as variações interanuais expressam alterações na qualidade do sistema educativo ou são apenas um efeito estatístico do comportamento da variável?

O eterno problema do acesso ao ensino superior

A qualidade da informação gerada pelos resultados e o seu uso para melhorar a qualidade da aprendizagem

Resultados por item (micro): como são usados os relatórios estatísticos por escola e qual o seu potencial?

Que explicações para a cristalização de resultados que medem o desempenho ao nível dos processos cognitivos avaliados nas provas?



Biologia e Geologia

Resultados por Prova (globais e por item) - Escola, Região, País

	Cotação máxima (% de alunos)	Cotação nula (% de alunos)	Pontos (média)	Pontos (%)	Agrup.	NUTS III	NUTS II	Nacional
Nacional								84
Centro							84	-
Baixo Mondego						88	+4	+4
Escola Secundária					83	-5	-1	-1
Escola Secundária			83		-	-5	-1	-1
I-1	38.3	61.7	1.9	38.3	-	49.3% -11.0%	49.0% -10.7%	49.4% -11.1%
I-2	83.0	17.0	4.1	83.0	-	90.0% -7.0%	89.1% -8.1%	89.3% -8.3%
I-3	72.3	27.7	3.6	72.3	-	78.7% -6.4%	78.4% -8.1%	78.5% -8.2%
I-4	61.7	38.3	3.1	61.7	-	67.5% -5.8%	65.1% -3.4%	65.5% -3.8%
I-5	78.7	21.3	3.9	78.7	-	82.6% -3.9%	81.5% -2.8%	82.3% -3.6%
I-6	-	48.9	1.9	18.7	-	22.2% -3.9%	19.9% -1.2%	20.4% -1.7%
I-7	2.1	80.9	1.1	11.3	-	13.4% -2.1%	11.9% -0.8%	12.5% -1.2%
II-1	29.8	70.2	1.5	29.8	-	24.9% +4.9%	23.5% +6.3%	23.1% +6.7%
II-2	51.1	48.9	2.6	51.1	-	58.9% +4.9%	59.1% +6.7%	59.2% +6.7%

Variáveis que afetam a validade dos resultados: alguns exemplos

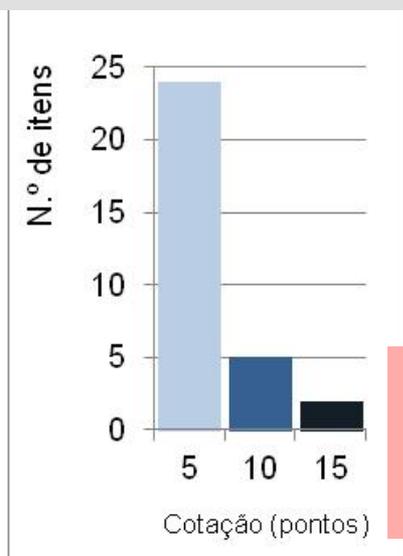
A natureza pública das provas face às opções técnicas que a aplicação de provas não públicas oferece

- As fragilidades ou impossibilidade de pré-testagem dos itens
 - A ausência de calibração dos itens
 - A produção anual de mais de 2000 itens novos
 - A impossibilidade de normalização dos resultados
 - O modelo de previsão de resultados baseado nos resultados e na manutenção da propriedades técnicas dos itens
-

Os esquemas de classificação : três exemplos a partir dos resultados da prova 702 de 2013 (1.ª fase)

Esquema tradicional

A cotação dos itens cresce na razão direta da sua dificuldade – a dupla discriminação dos itens



Resultado (real)
84,3 pontos

■ ←
Difícil

A cotação total é distribuída uniformemente pelos itens da prova

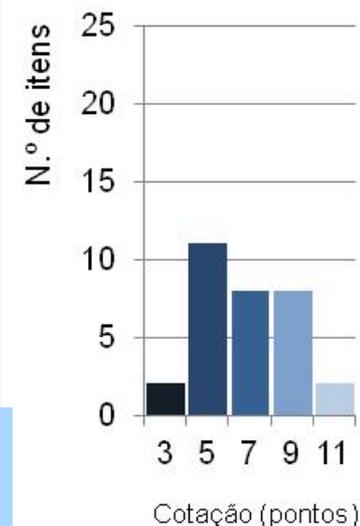
31 itens com 6,451613 pontos

Resultado (simulação)
93,1 pontos

Resultado (simulação)
105,1 pontos

→ ■
Fácil

A cotação dos itens cresce na razão inversa da sua dificuldade*



* Classificações finais ajustadas ao valor de referência - 200 pontos

Muito obrigado
