



**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS DA SAÚDE
CURSO DE MEDICINA
NÚCLEO DE DOCENTE ESTRUTURANTE**

TUTORIAL

PLANEJAMENTO DE TESTES

CONSTRUÇÃO DA MATRIZ E DA FOLHA DE ESPECIFICAÇÕES

NÚCLEO DOCENTE ESTRUTURANTE

NOVEMBRO 2015

INTRODUÇÃO

O planejamento das atividades educacionais faz parte do ciclo de gerenciamento educacional e segue os princípios de gerenciamento de qualidade total. O ciclo PDCA (Plan - Do - Check - Act) é uma ferramenta muito útil para o desenvolvimento de sistemas de ensino e avaliação.¹

A avaliação é um processo, cujo resultado são os escores ou conceitos atribuídos aos examinandos. Estes escores devem refletir o nível de aptidão ou de competência dos alunos em determinado conteúdo educacional (teórico ou prático).

O planejamento de testes é a primeira fase no processo de avaliação, que compreende além desta, as fases de construção, administração, avaliação psicométrica e correções de itens ou de conteúdo.

Neste tutorial, serão abordados as duas ferramentas de planejamento de testes que contribuem para aumentar a eficiência avaliativa por garantirem a validade de conteúdo do teste: a matriz do teste e sua folha de especificações.

DEFINIÇÃO DE MATRIZ

Matriz é o modelo que guia a construção de um teste. Em inglês, é chamada *blueprint*, por analogia com as plantas arquitetônicas que eram impressas em papel azul. Assim, a matriz é a planta do teste, o mapa de itens do teste. Em matemática, matrizes são elementos que agregam vetores em duas ou mais dimensões. Analogamente, as matrizes de teste são representadas por tabelas multidimensionais.^{2,3}

OBJETIVOS

O objetivo primário de uma matriz de teste é assegurar a validade de conteúdo do teste. A validade de conteúdo é a propriedade psicométrica do teste que mede a congruência entre o conteúdo do teste e os objetivos de ensino ⁴. O conteúdo da avaliação é válido quando é congruente com os objetivos e as experiências de aprendizagem e congruência entre estes pilares da educação pode ser facilitada pela utilização de matrizes de especificações de itens na avaliação. ³

A matriz visa diminuir a ocorrência de duas ameaças à validade de conteúdo do teste:

- a. Sub-representação do conteúdo, que se refere à sub-amostragem ou amostragem tendenciosa do domínio do conteúdo ou do conteúdo do curso.
- b. Variância irrelevante do construto: erro sistemático causado pela inclusão de conteúdo não relacionado ao construto de interesse do teste (competência, aptidão em tópicos específicos e relacionados) ^{3,5}

CRITÉRIOS

Testes podem ter caráter formativo ou somativo. São de caráter formativo quando utilizados para prover ao aprendiz feedback com relação ao seu desempenho, estabelecer objetivos de desempenho futuro e tomada de ações corretivas. E a avaliação para o aprendiz. Testes adquirem caráter formativo quando visam definir a competência de um indivíduo em determinada área do conhecimento, definida pelo conteúdo educacional do curso, de um semestre ou de uma unidade de ensino completada. Os resultados de avaliações somativas definem se o indivíduo é capaz de avançar para o próximo nível do curso.

A extensão de um teste é o número de itens que o compõem. Um teste pode conter

apenas um número limitado de itens, ou seja, a extensão do teste é limitada. O número de itens é determinado pelos objetivos do teste (formativo ou somativo), pela abrangência do conteúdo e pelo nível de confiabilidade necessário.

Testes de caráter formativo podem conter um número menor de itens do que testes de caráter somativo, uma vez que estes necessitam de maior coeficiente de confiabilidade. A confiabilidade do teste depende basicamente dos índices de discriminação e do número de itens do teste. Quanto maior o índice de discriminação médio e maior a extensão do teste, maior a confiabilidade.⁶ Quando não se utilizam questões de banco de itens, o índice de discriminação dos itens não é conhecido. Por esta razão, o processo de construção de itens deve ter como objetivo a produção de itens com grande potencial discriminativo. Neste processo, além da redação clara, concisa e objetiva, o conteúdo do item assume papel fundamental, porque deve ser cuidadosamente direcionado a um objetivo educacional relevante e crítico, cuja resposta correta diferencie os indivíduos mais bem preparados dos indivíduos menos bem preparados.

CONSTRUÇÃO

A construção de uma matriz de teste é um exercício de reflexão. Algum tempo deve ser devotado a esta atividade. Este tempo será recuperado durante o processo de construção dos itens do teste.

O método para a construção da matriz pode ser descrito nos seguintes passos ou etapas:

Passo 1- Trabalhe em colaboração com os colegas que participarão com itens

para o teste

A construção de uma matriz de teste pode ser um processo solitário somente se o conteúdo do teste não depender de múltiplos professores. No currículo integrado isto raramente acontece. Ao mesmo tempo, a construção de uma matriz de teste necessita de um *líder*, cuja função é a de coordenar os processos seguintes e compilar os resultados na forma final do teste. Portanto, recomenda-se que a matriz seja construída com a colaboração dos professores envolvidos no teste.

Passo 2 - Liste os conteúdos a serem testados

A matriz do teste normalmente será representada por uma tabela, cuja primeira coluna conterá a lista detalhada de objetivos educacionais a serem avaliados pelo teste (tópicos do conteúdo). Esta lista deverá ser extraída dos objetivos educacionais do projeto de ensino do curso, da disciplina, da aula, dependendo do foco do teste. Por exemplo, se o teste for referente a uma aula, então a lista conterá os objetos de aprendizado que foram abordados na aula. Se o teste for uma prova trimestral, o conteúdo será o administrado durante todo o trimestre, de cada disciplina, em cada unidade de ensino. E assim por diante.

Como regra geral, recomenda-se que o conteúdo listado seja detalhado o suficiente para facilitar o passo seguinte: a atribuição de peso aos tópicos do conteúdo.

Passo 3 - Estabeleça um peso para cada conteúdo

Os tópicos do conteúdo não possuem a mesma relevância ou impacto clínico. Dada a extensão limitada dos testes, é razoável que se abordem os conhecimentos sobre os tópicos mais relevantes do conteúdo educacional. Para determinar a importância

relativa de cada tópico, atribuem-se pesos a cada elemento da lista de tópicos. As colunas seguintes da matriz conterão os escores segundo os critérios utilizados para este processo e os pesos propriamente ditos.

Um dos métodos mais utilizados para estabelecer o peso dos tópicos do conteúdo clínico e cirúrgico é atribuir escores segundo sua frequência e seu impacto clínico. Neste método a frequência refere-se à prevalência ou à incidência do problema (I) e o impacto clínico a sua potencialidade letal ou sua urgência (F). Os escores são atribuídos segundo escalas de 3 pontos, em que 3 representa maior frequência e maior urgência/letalidade. O peso relativo de cada tópico é o produto da multiplicação dos dois escores $(I \times F)^2$.

Um outro critério para atribuição de peso aos conteúdos é o tempo de aula dedicado a ele. Neste caso, o peso de cada tópico será o tempo e seu percentual relativo ao tempo total do curso ou da aula.

Para ciências básicas, os critérios para atribuição de pesos aos tópicos do conteúdo são menos claros. Considerando que o objetivo das disciplinas de ciências básicas no curso de medicina é ensinar elementos que permitam o entendimento fisiopatológico dos problemas clínicos, sugere-se que os critérios centrados na relevância clínica do conhecimento sobre cada tópico do conteúdo curricular. Isto reforça a necessidade de integração entre os professores de ciências básicas e os de ciências clínicas na construção de testes no currículo integrado do curso de medicina.

Para avaliações formativas ou de conteúdos específicos de uma aula, em uma determinada disciplina, o critério para a atribuição de peso é a importância do tópico para o conhecimento do assunto da aula. Aos tópicos mais importantes, atribui-se

maior peso.

Os escores e pesos atribuídos a cada tópico são inseridos como colunas na matriz, que, a esta altura já possui as seguintes colunas:

Matriz para ciências clínicas e cirúrgicas

Tópico	Impacto (I)	Frequência (F)	Peso (I x F)

Matriz para Ciências básicas

Tópico	Peso (Implicação clínica)

Matriz baseada no tempo dedicado ao tópico

Tópico	Tempo	Percentual do tempo total

Passo 4 - Estabeleça o número de itens a serem incluídos no teste

A escolha do número de itens de um teste depende dos seguintes fatores:

- a. abrangência do teste - extensão do conteúdo
- b. tempo disponível para a construção, revisão e formatação do teste
- c. tempo disponível para a aplicação do teste. Testes longos são prejudicados por fadiga dos examinandos. Quando o conteúdo curricular é muito amplo, múltiplos pequenos testes são mais produtivos do que uma única prova.

- d. nível de confiabilidade desejada - testes com finalidades somativas, como provas trimestrais ou finais necessitam de maiores índices de confiabilidade (maiores que 70%). Testes de caráter formativo e com pequena consequência, admitem coeficientes de confiabilidade mais baixos (40 a 60%). A extensão do teste contribui para a consistência interna do teste, mantida a expectativa realística de índice de discriminação em torno de 0,2 a 0,3.
- e. Como regra geral, admite-se testes com 20- 40 itens para provas parciais. Para provas finais, no mínimo 60 itens são necessários. A confiabilidade dos testes não aumenta significativamente com a inclusão de mais de 100 itens em um teste.
- f. É importante que os conteúdos de cada teste sejam relacionados. Isto contribui para a obtenção de uma propriedade muito importante do teste: a unidimensionalidade, que é requisito para a consistência interna do teste.⁷

Passo 5 - Estabeleça o número de itens relativos a cada tópico

Nesta etapa, o número de itens referentes a cada tópico será proporcional ao peso atribuído. Para obter este número:

- a. Some o total dos pesos atribuídos a cada tópico
- b. Calcule o peso percentual de cada tópico (= peso do tópico/peso total)
- c. Multiplique o peso percentual de cada tópico pelo número total de itens do teste

Ao fim desta etapa, a matriz já está assim:

Matriz exemplo para um teste programado de 20 itens

Tópico	Impacto (I)	Frequência (F)	Peso (I x F)	Peso relativo	Numero de itens
Asma	3	2	6	0,23	5
Câncer	2	2	4	0,15	3
Mucoviscidose	1	1	1	0,04	1
DPOC	2	3	6	0,23	5
Pneumonias	3	3	9	0,35	7
Somatórios			26	1	20

Passo 6 - Estabeleça o que testar em cada item

A fase final da construção de uma matriz de teste consiste em estabelecer que faceta ou tipo de habilidade cognitiva será testada dentro de cada tópico.

Em questões sobre o conteúdo clínico, os aspectos de diagnóstico, tratamento e prevenção podem diferir em importância (relevância) entre os tópicos do conteúdo.

Abaixo, a matriz do teste, com pesos atribuídos aos aspectos de diagnóstico, tratamento e prevenção. Estes pesos são atribuídos em uma escala de 0 a 1 e devem somar à unidade. Quanto maior o peso, maior a importância atribuída. Por serem arredondados, os números de itens estimados nesta fase poderão ser maiores do que o total previamente programado. Ajustes podem ser feitos na etapa seguinte.

Matriz exemplo estabelecimento dos itens por aspectos do tópico

Tópico	I	F	(IxF)	Peso relativo	Nº de itens	Diag		Trat		Prev	
						P	n	P	n	P	n
Asma	3	2	6	0,23	5	0,3	2	0,7	4	0	0
Câncer	2	2	4	0,15	3	0,3	1	0,3	1	0,3	1

Mucoviscidose	1	1	1	0,04	1	1	1	0	0	0	0
DPOC	2	3	6	0,23	5	0,4	2	0,4	2	0,2	1
Pneumonias	3	3	9	0,35	7	0,6	4	0,4	3	0	0
Somas			26	1	20		10		10		2

Notas: I = impacto; F = frequência; Diag = diagnóstico; Trat = tratamento; Prev = prevenção P = peso; n = número de ítems

Passo 7 - Construa a folha de especificações do teste.

Nesta etapa, cria-se uma lista detalhada segundo o estabelecido na matriz. Por exemplo, na lista referente à matriz ilustrada acima, haverá dois ítems chamados Asma: diagnóstico e 4 ítems chamados Asma: tratamento. Na folha de especificações incluem-se as habilidades cognitivas a serem testadas em cada item da prova. Assim, os ítems podem elicitar os níveis lembrar, compreender, aplicar, analisar, avaliar ou criar da taxonomia de Bloom revisada⁸. Nesta fase estabelece-se também o tipo de ítems a serem incluídos no teste. Ítems de tipo resposta escolhida (múltipla escolha, por exemplo) são mais adequados para testar os três primeiros níveis do domínio cognitivo. Ítems tipo resposta construída (discursivos, por exemplo) são mais apropriados para testar níveis mais complexos, como análise, avaliação e criação. Lembre-se de que o tipo de item influencia a duração do teste.

A tabela abaixo é a folha de especificações do teste que criamos neste tutorial.

Folha de especificações do teste:

		Lem	Ent	Apli	Ana	Ava	Cri	Tipo de item
1	Asma: diagnóstico		X					ME
2	Asma: diagnóstico				X			ME

3	Asma: tratamento			X				ME
4	Asma: tratamento		X					ME
5	Asma: tratamento	X						ME
6	Asma: tratamento			X				ME
7	Câncer: diagnóstico	X						ME
8	Câncer: tratamento		X					ME
9	Câncer: prevenção						X	DI
10	Mucoviscidose: diagnóstico		X					ME
11	DPOC: diagnóstico				X			ME
12	DPOC: diagnóstico		X					ME
13	DPOC: tratamento			X				ME
14	DPOC: tratamento		X					ME
15	DPOC: prevenção			X				ME
16	Pneumonias: diagnóstico	X						ME
17	Pneumonias: diagnóstico		X					ME
18	Pneumonias: diagnóstico		X					ME
19	Pneumonias: diagnóstico		X					ME
20	Pneumonias: tratamento			X				ME
21	Pneumonias: tratamento			X				ME
22	Pneumonias: tratamento		X					ME

Notas: Lem = lembrar; Ent - entender; Apl = aplicar; Ana = analisar; Ava = avaliar; Cri = criar

ME = múltipla escolha; DI = discursivo

Passo 8 - Estabeleça a ordem para os itens no teste.

Considera-se apropriado organizar os itens de um teste segundo o nível de complexidade cognitiva exigida do aluno, ou seja, do esforço mental e físico que o

examinando deverá investir para responder ao item. Outro fator a ser considerado é a impressão inicial que o teste provoque no examinando. Uma impressão de que o teste está muito acima da sua capacidade poderá aumentar o nível de ansiedade relacionada ao teste e diminuir o desempenho cognitivo do examinando ^{9,10}. Considerando estes dois fatores, uma boa regra a seguir durante a criação da ordem dos itens no teste é iniciar com itens menos complexos e aumentar gradativamente a complexidade até a metade do teste (estratégia para diminuir a ansiedade de teste), quando a complexidade deve ser reduzida novamente (estratégia para enfrentamento da fadiga relacionada ao teste).

Compõem a complexidade do item (a) o seu nível de exigência cognitiva (taxonomia de Bloom, por exemplo) e (b) o seu formato do item (resposta escolhida versus resposta construída). Classificar os itens apenas como fáceis, médios ou difíceis tem-se provado ineficiente e muito impreciso. Estes atributos só podem ser verificados após a realização do teste.

Por exemplo, na matriz acima, os itens poderiam ser assim dispostos:

Disposição dos itens segundo o nível de complexidade.

Cognição	Lem	Ent	Apl	Ana	Cri	Ana	Apl	Ent	Lem
Tipo de item	ME	ME	EM	DI	ME	EM	EM	ME	ME
Itens da matriz	5,6	1,4,8, 10,12	3,6,13	2	9	11	15,20, 21	14,17, 18,19, 22	16

Notas: Lem = lembrar; Ent - entender; Apl = aplicar; Ana = analisar; Ava = avaliar; Cri = criar

CONCLUSÃO

Uma vez construída a matriz e sua folha de especificações, inicia-se a construção dos itens.

Esta etapa requer técnica. Felizmente existem regras que, se seguidas, asseguram a validade de face dos itens e eliminam ameaças à confiabilidade e discriminação dos itens.¹¹

REFERÊNCIAS

1. Sallis, E. & Edward, S. *Total Quality Management in Education*. (1988).
2. Coderre, S., Woloschuk, W. & McLaughlin, K. Twelve tips for blueprinting. *Med. Teach.* **31**, 322–324 (2009).
3. Patil, S. Y., Gosavi, M., Bannur, H. B. & Ratnakar, A. Blueprinting in assessment: A tool to increase the validity of undergraduate written examinations in pathology. *Int J Appl Basic Med Res* **5**, S76–9 (2015).
4. Downing, S. M. Validity: on the meaningful interpretation of assessment data. *Med. Educ.* **37**, 830–837 (2003).
5. Hamdy, H. & Hossam, H. Blueprinting for the assessment of health care professionals. *Clin. Teach.* **3**, 175–179 (2006).
6. Ebel, R. L. THE RELATION OF ITEM DISCRIMINATION TO TEST RELIABILITY¹. *Journal of Educational Measurement* **4**, 125–128 (1967).
7. Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297–334 (1951).
8. Anderson, L. W. in *Encyclopedia of Educational Theory and Philosophy* (2014).
9. Munz, D. C. & Smouse, A. D. Interaction effects of item-difficulty sequence and achievement-anxiety reaction on academic performance. *J. Educ. Psychol.* **59**, 370–374 (1968).
10. Chen, H. & Huilin, C. The Moderating Effects of Item Order Arranged by Difficulty

on the Relationship between Test Anxiety and Test Performance. *Creative Education* **03**, 328–333 (2012).

11. Case, S. M. & Swanson, D. B. *Constructing written test questions for the basic and clinical sciences*. (1998).