

6. AVALIAÇÃO DOS MÉTODOS

A comparação de eficiência dos métodos para avaliar compreensibilidade de símbolos gráficos foi feita a partir da Técnica de Avaliação Binária por meio de critérios ponderados proposta por David Meister (1985, apud Moraes e Mont'Alvão, 2010).

Com essa tabela, estamos propondo um instrumento de avaliação comparativa de eficiência dos métodos de avaliação de compreensibilidade conhecidos pela bibliografia citada, suas características e adequações, para facilitar os designers e pesquisadores a escolherem o(s) método(s) que melhor lhe convenham em seus projetos ou pesquisas.

6.1 Quadro demonstrativo – 1ª Etapa

Na 1ª etapa foi feito um quadro demonstrativo especificando características de cada método usado nesta pesquisa. Não analisamos os métodos de pós-ocupação nem o método de avaliação AIGA, citados no final do capítulo 4 de métodos e técnicas, pois como não os aplicamos, não teríamos dados de cruzamento e verificação. A norma ISO não foi considerada como método à parte, pois os testes nela incluídos já estão contentados nos métodos aplicados.

Foram descritos os métodos:

1. Métodos de pré-seleção
2. Método de produção
3. Teste de Compreensão
4. Método de reidentificação
5. Teste de eleição
6. Teste de classe de adequação
7. Teste de distribuição de classes de adequação
8. Teste de estimativa de compreensibilidade
9. Avaliação Heurística
10. Método de correspondência.

Para o quadro demonstrativo, consideramos os critérios:

- Nome do método;
- Referência bibliográfica;
- Material necessário para realização do teste;
- Forma de tabulação;
- Perfil recomendado dos sujeitos;
- Número mínimo de sujeitos participantes do teste;
- Adaptações já feitas por pesquisadores;
- Nível de dificuldade de aplicação do teste pelos pesquisadores (1 = baixo / 2 = médio / 3 = alto);
- Nível de dificuldade para os respondentes (1 = baixo / 2 = médio / 3 = alto);
- Fatores que influenciam o teste;
- Nível de objetividade;
- Tempo de duração de cada teste por referente ou por vários (curto = 3 segundos a 3 minutos / médio = 3 a 6 minutos / longo = mais de 6 minutos);
- Resultado esperado ou objetivo de cada teste.

Este quadro (tabela 6.1) apresenta um panorama dos métodos aplicados nesta pesquisa com o objetivo.

Tabela 6.1 - Quadro demonstrativo de características

Nome do Método	Referência bibliográfica	Material necessário	Método de tabulação	Perfil recomenda- do dos sujeitos	Nº mínimo de sujeitos recomendado	Adaptações	Nível de difi- culdade de aplicação do teste*	Nível de dificuldade para os responden- tes*	Fatores que influenciam o teste.	Nível de Objetividade	Tempo**	Resultado Esperado
1. Métodos de pré-seleção	Torgerson, 1960 Brugger 1994 Foster 1990/1	Coleta vasta de imagens colocadas em uma página A4 por referente. Caneta para marcação das imagens escolhidas. Ficha para tabulação.	Ficha A4 para cada referente, com todas as imagens correspondentes para contagem dos votos. Escolher de 3 a 10 imagens mais votadas.	Especialistas	mínimo 3	Pode-se usar os métodos 5, 6 ou 7, com suas respectivas tabulações.	1	2	Importância e hierarquia dos parâmetros gráficos	Opinião técnica objetiva com fatores subjetivos	Curto para cada referente	Avaliação técnica dos símbolos quanto à pertinência para a pesquisa.
2. Método de produção	Krampen 1969	1 ficha para os dados do sujeito + 1 ficha de papel branco, formato A5 ou A6, para cada referente com gramatura > 150g. Caneta futura ou lápis 2B.	Para cada referente, fazer uma tabela entre a variedade de imagens desenhadas pelos sujeitos, e a quantidade de vezes que elas se repetem durante o teste.	Usuários	15		2	2	Repertório e habilidade	Objetivo	Curto para cada referente. Tempo total médio	Repertório dos usuários das imagens relacionadas ao referente para desenvolver novo design.
3. Teste de Compreensão	Brugger 1994 ISO 9186	1 caderno grampeado com capa para os dados do sujeito + 1 folha de papel formato A5 ou A6, para cada referente com lugar para preencher. Não usar mais de uma imagem do mesmo referente para o mesmo sujeito. Lápis ou caneta.	É necessário pelo menos 3 avaliadores. Em tabela separada, cada avaliador atribui os pontos de cada imagem conforme a tabela de resposta de Brugger. Contagem de cada referente pela maioria dos pontos dos juizes.	Usuários	30	Caso não haja maioria entre os pontos dos juizes, pode-se calcular a média.	1	2	Repertório	Objetivo	Curto para todos os referentes	Grau de entendimento de cada imagem pelos usuários.
4. Método de reidentificação	Zwaga 1973 Stephen Carr 1973	2 tipos de baralho de papel com gramatura > 180 g. cada carta com uma imagem. Os baralhos terão todas as imagens. Um baralho com o significado e outro sem. Ficha para o entrevistador colocar as respostas.	Para cada referente, fazer uma ficha com as imagens correspondentes para contar a quantidade de acertos ou erros.	Usuários	30	Imagens projetadas	2	2	Memória visual	Objetivo	Médio para até 20 imagens	Facilidade de memorização das imagens.
5. Teste de eleição	Krampen e Sevray 1969 Brugger 1994		Em uma ficha A4, separar as imagens em fileiras por referente para tabular a contagem dos votos.	Usuários ou Especialistas	30		1	1		Subjetivo	Curto	
6. Teste de classe de adequação	Brugger 1994	Baralho de papel com gramatura > 180 g. cada carta com uma imagem e nome do referente. Lápis ou caneta.	Para cada imagem, separadas em tabelas por referente, deve-se anotar a ordem de preferência do sujeito segundo a escala Linkert.	Usuários ou Especialistas.	30	As imagens também poderão estar numa página reunidas por referente com espaço para a resposta.	1	1	Gosto, compreensão, comprometi-mento	Subjetivo	Curto	Teste de opinião baseado no entendimento e gosto do usuário.
7. Teste de distribuição de classes de adequação	Brugger 1994	Ficha para o entrevistador preencher as respostas.	Em uma tabela por referente com os campos: "Altamente Adequado", "Levemente Adequado" e "Não Adequado" fazer a contagem do quanto cada imagem aparece em cada um dos casos.	Usuários ou Especialistas.	30		1	2		Subjetivo	Curto	
8. Teste de estimativa de compreensibilidade	Zwaga 1989 ISO 9186 Brugger 1994 Olmstead 1994	Caderno de papel A4 com capa para dados e exemplo de resposta + uma página com as imagens de cada referente com espaço para a resposta. Lápis ou caneta.	Para cada imagem, coloca-se em uma tabela a porcentagem estimada pelas respostas dos entrevistados. Ao final calcula-se a média dessas porcentagens.	Usuários	30	Escolher o tipo de escala de respostas.	2	3	Comprometi-mento, avaliação dos outros	Subjetivo	Curto a médio dependendo do sujeito	Opinião do usuário quanto ao entendimento das outras pessoas adicionado ao pré-julgamento dos outros em comparação a si mesmo.
9. Avaliação Heurística	Jacob Nielsen e Rolf Molich	Página com a relação de perguntas para cada objeto avaliado. Caneta.	Cada entrevistado responde às perguntas segundo a Escala de Níveis de gravidade de usabilidade de Nielsen e Molich. Ao final calcula-se a média dos pontos.	Especialistas	Mínimo 5		1	2	Importância e hierarquia dos parâmetros gráficos	Opinião técnica objetiva com fatores subjetivos	Longo	Avaliação técnica dos folders e das imagens.
10. Método de correspondência	Easterby& Zwaga 1976	Fichas com quadro de imagens para escolha de uma delas relacionada com o referente escrito. Importante a escolha das imagens incluídas na ficha.	Para cada ficha só existe uma resposta certa. Para cada imagem relativa ao referente é computado o número de acertos.	Usuários	20	Imagens só da pesquisa ou uso de outras imagens que possam confundir	1	2	Repertório e percepção	Objetivo	Curto	Nível de distinção da imagem entre outras; relação da imagem ao referente.

Legendas *1 = Baixo 2 = Médio 3 = Alto

** Curto - 3 seg. a 3 min. Médio - 3 a 6 min. Longo - mais de 6 min.

Tabela 6.1 – Quadro demonstrativo de características dos métodos de avaliação de compreensibilidade e usabilidade

6.2 Técnica de avaliação binária através de critérios ponderados

Essa técnica apresentada por David Meister (1985, apud Moraes e Mont'Alvão, 2010) objetiva tornar menos subjetivas as decisões, para selecionar a alternativa mais eficiente comparando as alternativas planejadas segundo um sistema adequado de critérios. Propomos uma adaptação desta técnica considerando os métodos como sendo as alternativas e os critérios como sendo suas características quantificáveis.

O primeiro passo da técnica consiste em explicitar os critérios que o pesquisador, ergonomista ou designer, considera importantes pra avaliar as alternativas, o que já foi feito no item anterior.

O segundo passo é a definição do peso/importância de cada critério, que é um julgamento inteiramente subjetivo, pois é uma questão de valor.

“O procedimento apresentado para determinar pesos quantitativos para estes critérios é fundamentado em Hagen (1967) e apresentado na Tabela 1”... “Na situação real de projeção, poucos designers quantificam seus julgamentos, o que propicia infundáveis discussões e pouca objetividade.” (Moraes e Mont'Alvão, 2010)

6.2.1 Explicitação dos critérios – 1ª etapa

Consideramos, dentre os critérios descritos no quadro demonstrativo, os itens mais importantes e quantificáveis para levarmos em consideração na avaliação de eficiência dos métodos, são eles:

- Objetividade;
- Tempo de duração de cada teste;
- Dificuldade de aplicação do teste;
- Dificuldade do teste para os respondentes;
- Número de sujeitos.

6.2.2 Distribuição de pesos - 2ª etapa

Atribui-se então, com base em comparações binárias – 0 ou 1 -, aos pares de critérios, de pesos bons (1) ou ruins (0). Todos os critérios são comparados entre si de dois em dois. O processo continua até que todos os critérios tenham sido comparados entre si.

Conforme a tabela 6.2, podemos observar que a objetividade foi o critério com peso mais importante que todos os outros, seguido pelo critério: nível de dificuldade do teste para os respondentes.

Somam-se todos os valores 1 de cada critério para se ter o total de cada, que é dividido pelo número total de valores. Com isso teremos um coeficiente de peso para cada critério que é um valor relativo e não um valor absoluto. Na tabela a seguir, o coeficiente de peso do critério objetividade foi estipulado em 0,4 (maior) enquanto o tempo de duração de cada teste apurou um coeficiente 0 (menor valor).

TABELA 6.2		Distribuição de pesos para critérios escolhidos											
CRITÉRIOS	REGISTRO DE ESCOLHAS										TOTAL	COEFICIENTE DE PESO	
	Objetividade	1	1	1	1								
Tempo de duração de cada teste	0				0	0	0					0	0
Dificuldade de aplicação do teste		0			1			0	0			1	0,1
Dificuldade do teste para os respondentes			0			1		1		1		3	0,3
Número de sujeitos				0			1		1	0		2	0,2
												10	1

Tabela 6.2 – Distribuição de pesos para critérios escolhidos

Temos então um ranking de coeficientes que somam sempre 1 que é a totalidade. Por ordem de importância apuramos:

- Objetividade = 0,4
- Dificuldade do teste para os respondentes = 0,3
- Número de sujeitos = 0,2
- Dificuldade de aplicação do teste = 0,1
- Tempo de duração de cada teste = 0

A partir desses coeficientes serão calculados os valores de eficiência para cada teste no cruzamento final mostrado na tabela 6.8.

6.2.3 Avaliação binária - 3ª etapa

Na terceira etapa, é possível comparar cada método com todos os outros com base nos critérios estabelecidos. O melhor método recebe o valor 1 e o pior, o valor 0 como ocorreu na tabela de coeficientes de peso. Podemos entender o procedimento observando as tabelas 6.3 a 6.7. Cada uma delas se refere a um critério estabelecido, de acordo com a listagem já mencionada. Temos assim a avaliação binária entre todos os métodos. A soma dos valores apurados para cada método é dividido pelo total de valores, resultando assim o coeficiente de escolha para cada método com base nos critérios apurados.

Nas páginas seguintes apresentamos as tabelas relativas aos critérios.

TABELA 6.3		Comparação dos métodos pelos critérios																																				
		CRITÉRIOS																																				
MÉTODOS	Objetividade																																			Total de valores	Coeficiente de escolha	
1. Métodos de pré-seleção	0 0 0 1 1 1 1 0 0																																			4	0,0889	
2. Método de produção	1							0 0 1 1 1 1 0 0																													5	0,1111
3. Teste de Compreensão		1																																			9	0,2000
4. Método de reidentificação			1																																		8	0,1778
5. Teste de eleição				0																																	3	0,0667
6. Teste de classe de adequação					0																																2	0,0444
7. Teste de distribuição de classes de adequação						0																															1	0,0222
8. Teste de estimativa de compreensibilidade							0																														0	0,0000
11. Avaliação Heurística																																					6	0,1333
12. Método de correspondência																																					7	0,1556
																																					45	1

Tabela 6.3 – Comparação de alternativas de configuração – critério: objetividade

TABELA 6.4		Comparação dos métodos pelos critérios																																					
		CRITÉRIOS																																					
MÉTODOS	Tempo																																			Total de valores	Coeficiente de escolha		
1. Métodos de pré-seleção	0 0 1 0 0 0 0 1 0																																				2	0,0444	
2. Método de produção	1							0 1 0 0 0 1 1 0																														4	0,0889
3. Teste de Compreensão		1																																				9	0,2000
4. Método de reidentificação			0																																			1	0,0222
5. Teste de eleição				1																																		7	0,1556
6. Teste de classe de adequação					1																																	5	0,1111
7. Teste de distribuição de classes de adequação						1																																6	0,1333
8. Teste de estimativa de compreensibilidade							1																															3	0,0667
11. Avaliação Heurística								0																														0	0,0000
12. Método de correspondência									1																													8	0,1778
																																						45	1

Tabela 6.4 – Comparação de alternativas de configuração – critério: tempo

6.2.4

Matriz para seleção final de eficiência dos métodos - 4ª etapa

A quarta e última etapa é a construção de uma matriz com as alternativas (métodos) em colunas e os critérios em fileiras (tabela 6.8). Os valores nas colunas abaixo dos métodos representam a multiplicação do critério do coeficiente de peso (tabela 6.2) pelo coeficiente de escolha do método com base em cada critério. Os resultados classificam os métodos em um ranking de eficiência.

O nível comparativo de eficiência de cada método foi então resultado da avaliação com base em cinco critérios utilizando a técnica de avaliação binária através de critérios ponderados proposta por Meister (1985).

RESULTADO DO CRUZAMENTO DOS CRITÉRIOS										
CRITÉRIOS	MÉTODOS									
	1. Métodos de pré-seleção	2. Método de produção	3. Teste de Compreensão	4. Método de reidentificação	5. Teste de eleição	6. Teste de classe de adequação	7. Teste de distribuição de classes de adequação	8. Teste de estimativa de compreensibilidade	9. Avaliação Heurística	10. Método de correspondência
Objetividade	0,089	0,111	0,200	0,178	0,067	0,044	0,022	0,000	0,133	0,156
Tempo de duração de cada teste	0,044	0,089	0,200	0,022	0,156	0,111	0,133	0,067	0,000	0,178
Dificuldade de aplicação do teste	0,089	0,022	0,200	0,000	0,156	0,133	0,111	0,044	0,067	0,178
Dificuldade do teste para os respondentes	0,111	0,022	0,133	0,044	0,200	0,178	0,156	0,000	0,089	0,067
Número de sujeitos	0,200	0,156	0,000	0,022	0,067	0,089	0,111	0,044	0,178	0,133
TOTAL	0,5333	0,4000	0,7333	0,2667	0,6444	0,5556	0,5333	0,1556	0,4667	0,7111
LUGAR	5º	7º	1º	8º	3º	4º	5º	9º	6º	2º

Tabela 6.8 – Matriz para seleção final dos métodos

6.3 Conclusões da técnica de avaliação binária

Propomos por meio do Método de Meister um instrumento comparativo de eficiência dos métodos de avaliação da compreensibilidade da informação podendo ser replicado para outros métodos de avaliação ergonômica informacional de outros objetos de estudos ou casos exemplares e ainda com escolha de critérios diferentes.

A partir da análise do método de Meister e da tabela resultante poderemos discutir sobre os critérios utilizados. Será que deveríamos usar mais critérios ou outros critérios? Pode-se perceber que o critério tempo influenciou bastante a colocação dos métodos de reidentificação e avaliação heurística. Pensamos que os critérios deverão ser escolhidos de acordo com os objetivos da pesquisas e que novos critérios poderão ser escolhidos.

A partir da experiência com métodos de avaliação de compreensibilidade dos pesquisadores entrevistados e da própria autora o ranking de eficiência dos métodos provavelmente traria uma ordem diferente do resultado demonstrado. Questões como objetividade e facilidade de aplicação e tabulação dos métodos nos predispõem a favorecê-los. Estaríamos assim elegendo-os de forma subjetiva. Por isso foi escolhido o método de avaliação binária de Meister objetivando as escolhas.

Vale ressaltar alguns aspectos da tabulação final e confrontá-los com nossa experiência.

A maior eficiência do **método de compreensão** é sem dúvida inquestionável. Ele é objetivo, rápido, reconhecido internacionalmente, e por isso recomendado pelas normas ISO. Tem tabulação sofisticada com sete (7) parâmetros e três (3) juízes. (explicações do teste no cap. 4.5.2 p.111)

Vale a pena repetir aqui a resposta de o professor/pesquisador C com o qual foi feita a entrevista sobre métodos aplicados. (cap. 5.2 p. 129) Ele se refere à confiabilidade dos testes de compreensão e correspondência:

“O **teste de avaliação de correspondência** me pareceu bastante confiável, assim como o de **compreensão** com resposta aberta. Este último, contudo, deve ser realizado com um rigor metodológico talvez maior que os anteriormente citados aqui, pois é o que dá maior margem para interações entre participantes (se o teste for realizado em grupo) e, dependendo do material testado, sua condução deve ser acompanhada de perto pelo pesquisador de modo a (1) esclarecer possíveis dúvidas no decorrer do teste; (2) tentar extrair do participante uma quantidade de

informação maior que a que ele esteja disposto a fornecer; ou (3) não deixar que o participante responda às perguntas de maneira inadequada.”

O **método de correspondência** (2º lugar) é trabalhoso na sua confecção e pode ser tendencioso de acordo com a escolha feita para os símbolos vizinhos que estarão na mesma ficha do símbolo correto. Porém, dependendo do contexto da pesquisa e símbolos escolhidos ele poderá apresentar resultados surpreendentes. (explicações do teste no cap. 4.5.4 p. 115) Esse lugar no ranking confirma as observações feitas pelo professor/pesquisador C.

Os **testes de opinião (eleição, classe de adequação e distribuição de classes de adequação)** que tiveram 3º, 4º e 5º lugares no ranking estão coerentes com o resultado seqüencial já que são muito semelhantes na sua forma e aplicação. Eles trazem subsídios importantes principalmente se forem usados em conjunto com outros métodos, nunca isoladamente. Tivemos, em outra pesquisa, resultados que contrapõem opinião e compreensão dos usuários. Às vezes eles elegem um símbolo, porém no método de compreensão o mesmo símbolo não é compreendido por outros usuários de mesmo perfil. (explicações dos testes nos cap. 4.5.5; cap. 4.5.6; cap. 4.5.7 p. 116 a 119)

O **método de pré-seleção** empatou em 5º lugar com o **teste de distribuição de classes de adequação** muito por conta de usarmos testes de opinião como pré-seleção, porém com especialistas. Considero o resultado coerente. (explicações dos testes no cap. 4.4.1 p. 103)

A **avaliação heurística** fornece bons subsídios para a projeção. Talvez pelo fato de ser um método qualitativo sofisticado, feito com especialistas, mas de difícil aplicação e lento, pois demanda de boa vontade de profissionais em colaborar. Talvez por esses fatores tenha sido colocado em 6º lugar, pois discordo do valor de sua eficiência, eu o colocaria em 2º lugar. (explicações do método no cap.4.4.2 p. 104 a 108)

O **método de produção** ficou localizado em 7º lugar no ranking de eficiência pois é subjetivo, porém muito rico na captação de repertório dos usuários. É um método que nos traz idéias e muitas vezes imagens inesperadas. O usuário pode interpretar o referente de uma forma diferente caso não explicitemos o contexto. Por exemplo: para o referente “Enxague”, um usuário desenhou um enxague de roupa, quando o contexto era cabelo. Pode ter sido um erro do

entrevistador ou foi o que realmente veio à cabeça do usuário naquele instante. Considero um método muito importante e eficiente. Pensaria em colocá-lo pela minha experiência nos primeiros lugares. (explicações do método no cap.4.5.1 p. 109 a 111) Vale a pena repetir aqui trecho da entrevista feita com o professor/pesquisador C:

“O **teste de produção** me pareceu bastante útil e preciso desde que seja utilizado para identificação dos recursos de articulação da mensagem pictórica empregados pelos participantes e não para fornecer informações acerca de estilo de representação, uma vez que as habilidades dos participantes podem variar.”

O **método de reidentificação** ocupou o 8º lugar, porém ele pode ser bem aproveitado em diversas ocasiões. Quando temos um conjunto de símbolos e queremos testar a memorização de cada um ou seu nível de aprendizagem, ele pode ser bastante útil. Realmente, como método de avaliação de compreensibilidade ele pode ser falho, mas mesmo um símbolo ruim pode ficar retido na mente por sua estranheza. (explicações do método no cap.4.5.3 p. 113 e 114)

Em último lugar em nível de eficiência ficou o **método de estimativa de compreensibilidade**, talvez por ser o método mais subjetivo de todos. Ele é baseado nas opiniões dos respondentes acerca do entendimento dos outros. Por toda a bibliografia lida, incluindo as normas ISO, ele é um método muito considerado e relevante na Europa, porém no Brasil ele tem valores mais baixos. Isso pode ser consequência do perfil dos sujeitos da nossa pesquisa: mulheres das classes C/D e com baixa escolaridade. (explicações do método no cap.4.5.8 p. 120 a 123) Vale a pena repetir aqui trecho da entrevista feita com o professor/pesquisador C:

“O **Teste de Estimativa** me pareceu o menos confiável, uma vez que os participantes têm dificuldade para compreender sua participação na pesquisa e tendem a subestimar dificuldades de interpretação de representações pictóricas que outros públicos possam ter.”

Podemos afirmar que o **Método de Meister** teve resultados coerentes com nossa experiência em pesquisas de avaliação de compreensibilidade de símbolos e imagens pictóricas e que recomendamos sua aplicação com outros objetos de estudo confirmando ou não sua confiabilidade para que sirva de referência em pesquisas de eficiência de métodos de compreensibilidade.