



**SOCIEDADE  
CRISE E RECONFIGURAÇÕES**

# **VII CONGRESSO PORTUGUÊS DE SOCIOLOGIA**

**19 a 22 Junho 2012**

**Universidade do Porto - Faculdade de Letras - Faculdade de Psicologia e Ciências da Educação**

---

ÁREA TEMÁTICA: Teorias e Metodologias

---

**O EFEITO DO FORMATO DAS RESPOSTAS NO TRATAMENTO DE ITENS E ESCALAS TIPO LIKERT**

---

BOTELHO, Maria do Carmo  
Doutoramento em Métodos Quantitativos  
ISCTE-IUL  
maria.botelho@iscte.pt

---

CALAPEZ, Teresa  
Doutoramento em Métodos Quantitativos  
ISCTE-IUL  
teresa.calapez@iscte.pt

---

RAMOS, Madalena  
Doutoramento em Educação  
ISCTE-IUL  
madalena.ramos@iscte.pt

---



### Resumo

Em 1932, para medir noções complexas como crenças, percepções ou opiniões, Likert preconizou a utilização de afirmações, cada uma delas a ser pontuada numa valoração de um a cinco, de acordo com o grau de concordância do respondente. Desde essa altura, várias alternativas de formatação das respostas foram sugeridas, diferindo quer no número de pontos a utilizar, quer nos descritores verbais incluídos e, mais recentemente, propondo a substituição de um conjunto finito de opções de resposta por uma linha contínua.

Nesta comunicação pretende-se analisar em que medida o formato de resposta oferecido condiciona os resultados obtidos. Mediram-se as percepções acerca da Matemática, em três dimensões (enquanto domínio masculino, importância de compreensão dos conceitos e utilidade), cada uma delas composta por seis itens. Construíram-se quatro tipos de questionários: em dois deles as afirmações são avaliadas através do posicionamento num segmento de recta, sendo o primeiro etiquetado nos extremos e no ponto central e o segundo apenas nos extremos; no terceiro são usados itens de Likert com 5 pontos, todos etiquetados (discordo completamente, discordo, não concordo nem discordo, concordo e concordo completamente); o último tipo de questionário difere do anterior por ter apenas os extremos etiquetados.

Serão apresentados os resultados da comparação das respostas item a item e também das escalas obtidas por soma dos respectivos itens teoricamente definidos.

### Abstract

In 1932, to measure complex concepts such as beliefs, perceptions or opinions, Likert advocated the use of statements, each one scored on a one to five graduation, according to the degree of agreement of the respondent. Since that time, several formatting alternatives have been suggested, including length, verbal descriptors or, in more recently, offering a visual continuum instead of discrete options for the responses.

In this communication we will analyze to what extent results are influenced by the response format provided. We measured the perceptions of mathematics in three dimensions (as a male domain, the importance of understanding the concepts and usefulness) of six items each. Four types of questionnaires were constructed: in two of them statements are evaluated by marking a point on a visual continuum while in the other two five discrete options were provided. Moreover, the first continuum type has verbal descriptors at both ends and at the middle mark, while in the second one no middle mark is provided (and no descriptor thereof). In what the two discrete types are concerned, differences refer to the number of verbal descriptors provided: all points versus only the extremes.

Analysis results will be presented, both at item level and scale level.

Palavras-chave: Itens tipo Likert; Visual Analogue Scales (VAS); Âncoras.

Keywords: Likert-type items; Visual Analogue Scales (VAS); Verbal descriptors

PAP1504



## 1. Introdução

Em muitas áreas das Ciências Sociais e afins, as pessoas são a origem da informação: as suas crenças, percepções, opiniões, são o próprio objecto de estudo. Estas noções, complexas, subjetivas e multidimensionais, são, à primeira vista, impossíveis de quantificar. Em 1932, Likert propôs um método de medição que mudou para sempre o papel dos métodos quantitativos na pesquisa social – e despoletou uma controvérsia que dura desde então. Likert preconizou a utilização de afirmações, relativas à atitude em estudo, cada uma delas a ser pontuada numa valoração de um a cinco, de acordo com o grau de concordância do respondente com a afirmação. A medida da atitude obtém-se pela soma das valorações dos itens elegidos, justificando a denominação de escalas adicionadas pela qual também são conhecidas. Desde essa altura, várias alternativas de formatação das respostas vieram a ser sugeridas. As propostas referem-se tanto ao número de pontos a utilizar, como aos descritores verbais incluídos (descrição, quantidade e posicionamento) e, mais recentemente, à substituição de um conjunto finito de opções de resposta por uma linha contínua (*Visual Analogue Scales*, VAS). Qual o número de pontos ideal? Deve ser uma escala com um número de pontos par ou ímpar? Os pontos devem ser todos etiquetados ou apenas os extremos? Estas são questões em aberto até hoje.

No que se refere ao número de pontos ideal, vários estudos tentaram responder a esta questão analisando a relação entre o número de categorias de resposta oferecido e a consistência interna, sendo, todavia, os resultados inconsistentes. Se, para alguns autores (Jacoby e Matell, 1971; Aiken 1983) o número de pontos não tem qualquer relação com a consistência interna, para outros (Alwin e Krosnick, 1991; Chang, 1994; Alwin, 1997; Dawes, 2008) a estabilidade e a consistência dos resultados dependem do número de categorias de resposta, fazendo, por isso mesmo algumas recomendações. Facilmente se encontra na literatura a defesa da utilização de escalas com apenas 2 ou 3 categorias (Jacoby e Matell, 1971) até 10 (Preston e Colman, 2000) ou mesmo 11 (Alwin, 1997).

Algumas das divergências encontradas nestes estudos podem explicar-se pelo tipo de análises que o investigador pretende efetuar: trata-se de análises item a item ou sobre as escalas teoricamente definidas e obtidas pela soma (ou média) dos itens? No primeiro caso, as recomendações são, regra geral, no sentido de um número de pontos elevado (ainda que este número nunca tenha sido genericamente definido), enquanto no segundo caso, poderão até ser itens dicotómicos.

Relacionado com o número de pontos, alguns estudos discutem a pertinência da existência de um ponto central (Presser e Schuman, 1980; Katlon et al., 1980, citado por Moors, 2007; Bishop, 1987; Adelson e McCoach, 2010). Existe a convicção generalizada de que ao oferecer um ponto central, correspondente a uma posição neutra, ele será frequentemente escolhido. Pelo contrário, Moors (2007), ao comparar as respostas obtidas com a utilização de uma escala com cinco pontos (com uma posição central), com uma de seis pontos, chega à conclusão de que são mais as semelhanças do que as diferenças. Alwin e Krosnick (1991) concluem também que a existência de um ponto central pode originar, nos formatos de resposta mais curtos, níveis de consistência interna mais baixos. Mais recentemente, Adelson e McCoach (2010) comparam a utilização de formatos com quarto ou cinco pontos e recomendam este último formato, argumentando igualmente com a maior consistência interna. Uma vez mais, também aqui não há consenso.

Uma outra questão, ligeiramente diferente, prende-se com a colocação de descritores verbais (etiquetas) em todos os pontos ou apenas nos extremos (e eventualmente no ponto central). A primeira solução poderá ser bastante difícil (ou meso impossível) se o número de categorias for elevado, já que será muito complicado encontrar etiquetas que traduzam a ideia de equidistância entre os pontos. Por outro lado, a solução de etiquetar apenas as categorias extremas pode confundir o entrevistado quanto ao significado de cada categoria, provocando uma escolha arbitrária ou mesmo recusa de resposta.

Krosnick defende em vários estudos (Alwin e Krosnick, 1991; Krosnick e Berent, 1993; Krosnick, 1999) que a especificação de todas as hipóteses de resposta com descritores verbais melhora a fiabilidade das escalas. Lam e Klockars (1982) sugerem que, desde que os pontos sejam reconhecidos como igualmente espaçadas, os resultados obtidos utilizando um formato ou outro são semelhantes. Dixon et al (1984) encontraram diferenças significativas, não ao nível das escalas e da sua fiabilidade, mas sim na variabilidade individual dos itens constitutivos de cada escala, com uma dispersão mais acentuada quando apenas existiam etiquetas nas categorias extremas. Com uma perspetiva diferente, Dixon, Bobo e Stevick (1984), Newstead e Arnold (1989), Landrum (1999) e Cummins e Gullone (2000) não encontraram diferenças entre os itens quando todas as categorias estão etiquetadas ou quando apenas os pontos extremos têm descritores verbais. Segundo estes autores, as diferenças entre os itens não são significativas e a utilização de um ou outro formato não influencia a consistência interna das escalas. Novamente, não há consenso quanto a esta questão.

Uma outra linha de discussão prende-se com o tipo de tratamento a dar aos itens tipo-Likert. A forma como os investigadores analisam e comparam as respostas obtidas com este tipo de itens nem sempre é idêntica. Alguns fazem-no recorrendo exclusivamente a medidas de estatística descritiva (média, desvio-padrão, frequências ou percentagens por categoria de resposta), outros usam técnicas de análise estatística não-paramétrica (por exemplo, teste do qui-quadrado ou análise de variância de *Kruskal-Wallis*) e outros ainda usam procedimentos paramétricos (testes t, análises de regressão, análise de variância). A opção por uma determinada estratégia analítica passa pela resposta à pergunta “Podem os itens tipo Likert ser considerados variáveis métricas?”.

A resposta a esta questão passa pela possibilidade (ou não) de presunção da adequabilidade de assumir iguais afastamentos entre categorias adjacentes nos itens tipo-Likert, implícita na sua quantificação usual – atribuição de inteiros consecutivos a sucessivas categorias. Este tem sido um assunto para o qual os argumentos, contra ou a favor, têm assumido, por vezes, contornos de fundamentalismo (Göb *et al*, 2007, Jamieson e Carifio e Perla, 2007).

Alguns defendem a ordinalidade dos dados recolhidos sob este formato (independentemente do número de pontos e das etiquetas atribuídas), ou seja, os dados têm uma ordem inerente mas não é possível assumir uma equidistância entre os pontos. Os códigos atribuídos às respostas, por exemplo 1-2-3-4-5, são isso mesmo, apenas códigos e não faz sentido somar uma resposta codificada com 2 (concordância) com uma resposta codificada com 3 (indeciso), obtendo um valor médio de resposta de 2,5, que não tem interpretação. Assim sendo, a utilização de técnicas paramétricas está fora de questão.

Outra perspetiva é aquela segundo a qual a utilização de itens tipo-Likert pressupõe que por detrás dos valores de respostas que refletem as suas atitudes, perceções ou opiniões dos indivíduos, existe uma variável latente contínua, que só por facilidade de medição é discretizada sob um formato tipo-Likert. Argumenta-se que, ainda que a variável latente seja métrica, as pessoas têm dificuldade em verbalizar o seu posicionamento com precisão. Ninguém diz “Em geral, estou 75% satisfeito com a minha vida”. Assim, itens tipo-Likert são métricos e podem ser tratados com recurso a procedimentos adequados a este tipo de dados.

A fim de evitar a classificação imediata como ordinal, novas formas de medição têm sido propostas, variantes dos formatos de resposta já referidos, entre elas as Visual Analogue Scales (VAS) que consistem na apresentação de uma linha com um comprimento previamente determinado, não necessariamente graduada, mas ancorada em cada extremo, traduzindo duas situações opostas de resposta (por exemplo, concordo completamente, discordo completamente). No entanto, os resultados não têm sido sempre mais favoráveis com a utilização deste formato de respostas (Johansson e Finizia, 2007; Svensson, 2000).

A nossa prática de investigação sugere que a decisão acerca da utilização de um determinado formato pode ter razões de ordem diversa. Algumas de ordem mais conceptual, como sejam as características do público-alvo do estudo, outras de ordem mais prática e que dizem respeito a aspetos como a formatação do questionário (condensação das perguntas de forma a maximizar o espaço; questões visuais ou estéticas...). Independentemente da não existência de consenso relativamente às questões anteriormente levantadas, uma

coisa parece evidente: o formato de resposta não pode ser ignorado. Ou, tal como referem Christian, Parsons, e Dillman (2009, pp. 421), “it is clear that the visual attributes of scalar questions also influences answers and cannot be ignored”.

Assim, nesta longa viagem para a descoberta do melhor formato para os itens tipo-Likert, serão comparadas as distribuições de respostas obtidas utilizando quatro tipos de formatos diferentes: dois em que as afirmações são avaliadas através do posicionamento num segmento de reta (*Visual Analogue Scale*), sendo um deles etiquetado nos extremos e no ponto central e o outro apenas nos extremos; dois em que são usados itens de Likert com 5 pontos, num dos casos com todas as categorias etiquetadas (discordo completamente, discordo, não concordo nem discordo, concordo e concordo completamente) e outro com apenas os extremos etiquetados.

Neste artigo serão apresentados os resultados da comparação das respostas obtidas através da aplicação dos quatro tipos de questionário item a item e também das escalas obtidas por soma dos respetivos itens teoricamente definidos. A avaliação do comportamento dos itens e das escalas será efetuada com recurso a indicadores de localização, comparação de distribuições e medição de consistência interna.

Pretende-se, com este estudo, alertar os investigadores que utilizam itens tipo-Likert para medir percepções, representações ou atitudes, para a possibilidade de obtenção de distribuições de respostas diferentes, e consequentemente conclusões distintas, dependendo do tipo de formato de resposta oferecido.

## **2. Método**

### **2.1. Participantes**

A população deste estudo é constituída pelos alunos inscritos pela primeira vez no primeiro ano de várias licenciaturas do ISCTE-IUL (ciências sociais – antropologia, ciência política, economia, história, psicologia, sociologia; ciências de gestão – gestão, finanças, marketing, gestão de recursos humanos, gestão e engenharia industrial; ciências tecnológicas – engenharia informática, informática e gestão de empresas, engenharia de telecomunicações e informática).

Participaram neste estudo 727 alunos, cujas idades estão compreendidas entre os 16 e os 56 anos, com uma idade média de 20,9 anos (DP=6,7), sendo os 18 anos a idade mais frequente. A maioria dos participantes é do sexo feminino (52,1%), vem da área de ciências do secundário (71%) e teve na sua esmagadora maioria (91,4%) disciplinas básicas ou avançadas de matemática anteriormente à entrada na universidade. Cerca de metade destes estudantes está em cursos da área das ciências sociais (45,8%), 39,3% em licenciaturas da área de gestão e os restantes 14,9% em cursos tecnológicos, o que reflete claramente o perfil da população estudantil deste Instituto Universitário.

### **2.2. Materiais e Procedimentos**

Este artigo insere-se no âmbito de um projeto mais amplo acerca de crenças sobre a Matemática. Pretende-se avaliar as percepções dos alunos do 1º ano de um conjunto de licenciaturas do ISCTE-IUL acerca da Matemática, em três dimensões: utilidade, enquanto domínio masculino e importância de compreensão dos conceitos.

Cada dimensão é composta por seis itens, três com conotação positiva e três negativa (adaptações de Fennema-Sherman, 1976 e Kloosterman e Stage, 1992). Cada item consiste numa afirmação, para a qual se pede o registo da intensidade de concordância (Tabela 1).

Foram construídos quatro tipos de questionários: nos dois primeiros (A e B) as afirmações são avaliadas através do posicionamento num segmento de recta com 5 centímetros de comprimento (*Visual Analogue Scale* – VAS), sendo o A etiquetado nos extremos e no ponto central e o B apenas nos extremos; no terceiro (C) são usados itens tipo-Likert com 5 pontos, todos etiquetados (discordo completamente, discordo, não

concordo nem discordo, concordo e concordo completamente); o questionário do tipo D difere do anterior por ter apenas os extremos etiquetados.

Tabela1. Itens por dimensão

<b>Matemática como um Domínio Masculino</b>		
	Acrónimo <sup>a)</sup>	Descrição
1	DM1_I	É difícil acreditar que uma mulher possa ser um génio a matemática
2	DM2_I	As raparigas que gostam de estudar matemática são um pouco estranhas
3	DM3	Os rapazes não são naturalmente melhores do que as mulheres na matemática
4	DM4	As raparigas podem ser tão boas como os rapazes em matemática
5	DM5	Confitaria igualmente numa mulher e num homem para fazer cálculos importantes
6	DM6_I	Espero de uma mulher matemática que tenha um ar masculino
<b>Compreensão dos Conceitos é Importante na Matemática</b>		
	Acrónimo	Descrição
1	CC1_I	Na matemática, não interessa se se percebeu ou não um determinado problema, mas sim acertar na resposta
2	CC2	É tão importante encontrar a resposta certa a um problema matemático como saber porque é que se acertou.
3	CC3	Na matemática, o tempo que se gasta a perceber porque é que uma resolução de um problema faz sentido é sempre tempo bem gasto.
4	CC4_I	Não é importante saber porque é que um certo procedimento matemático funciona, desde que se obtenha a resposta certa
5	CC5	Uma pessoa que não entende porque é que determinada resposta a um problema matemático está correta, não resolveu verdadeiramente o problema.
6	CC6_I	Na matemática, obter um resultado correto é mais importante do que perceber como foi obtido
<b>Utilidade da Matemática</b>		
	Acrónimo	Descrição
1	U1_I	Estudar matemática é uma perda de tempo
2	U2	A matemática é um assunto necessário e importante
3	U3	Saber matemática vai ajudar-me a arranjar um emprego
4	U4_I	A matemática não tem relevância para a minha vida
5	U5_I	A matemática não vai ser importante para o meu trabalho futuro
6	U6	Estudo matemática porque sei quão útil é

Em todas as dimensões procedeu-se à inversão dos itens cuja conotação era negativa de forma a que valores mais altos indicassem: i) menor concordância com a matemática como um domínio masculino, ii) maior concordância com a importância da compreensão dos conceitos e iii) maior concordância com a utilidade da matemática.

Os questionários foram aplicados sistematicamente em cada turma, tendo-se conseguido uma distribuição equilibrada por tipo de questionário (Tabela 2). Os grupos podem ser considerados homogéneos atendendo ao sexo ( $\chi^2_{(3)}=5,385$ ,  $p = 0,146$ ), idade ( $\chi^2_{(15)}=14,815$ ,  $p = 0,465$ ) e área do curso ( $\chi^2_{(6)}=0,303$ ,  $p = 0,999$ ).

Tabela 2. Tipo de Questionário

	N	%
A – VAS com marcação e etiqueta no ponto central e extremos	184	25,3
B – VAS com marcação e etiqueta só nos extremos	179	24,6
C – Tipo-Likert com etiquetas em todos os pontos	187	25,7
D – Tipo-Likert com etiquetas só nos extremos	177	24,3
Total	727	100,0

As respostas foram comparadas a dois níveis: i) item a item, e ii) comparação das escalas obtidas por soma de itens. A avaliação do comportamento dos itens e das escalas fez-se com recurso a indicadores de localização, assimetria, dispersão, comparação de distribuições, medição de consistência interna e testes de independência de *Kolmogorov-Smirnov*.

### 3. Resultados

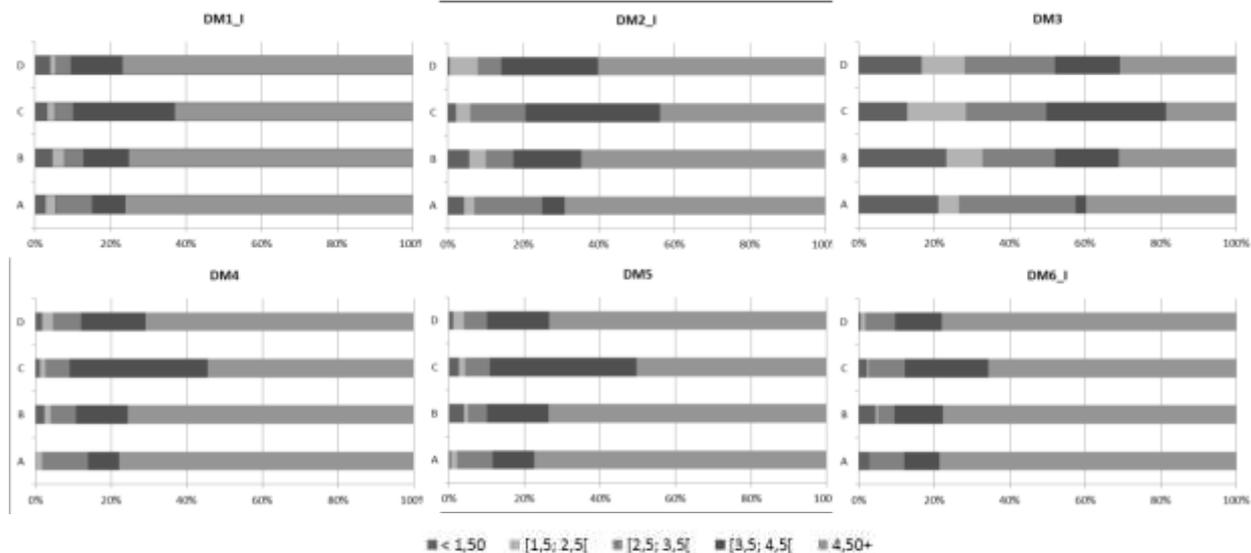
#### 3.1. Análise item a item

Como se pode observar na Figura 1 e corroborar pelas medidas de enviesamento presentes na Tabela 3, grande parte dos itens, qualquer que seja o seu formato, exhibe distribuição altamente enviesada, com 75% das respostas, ou mais, acima de 3,5.

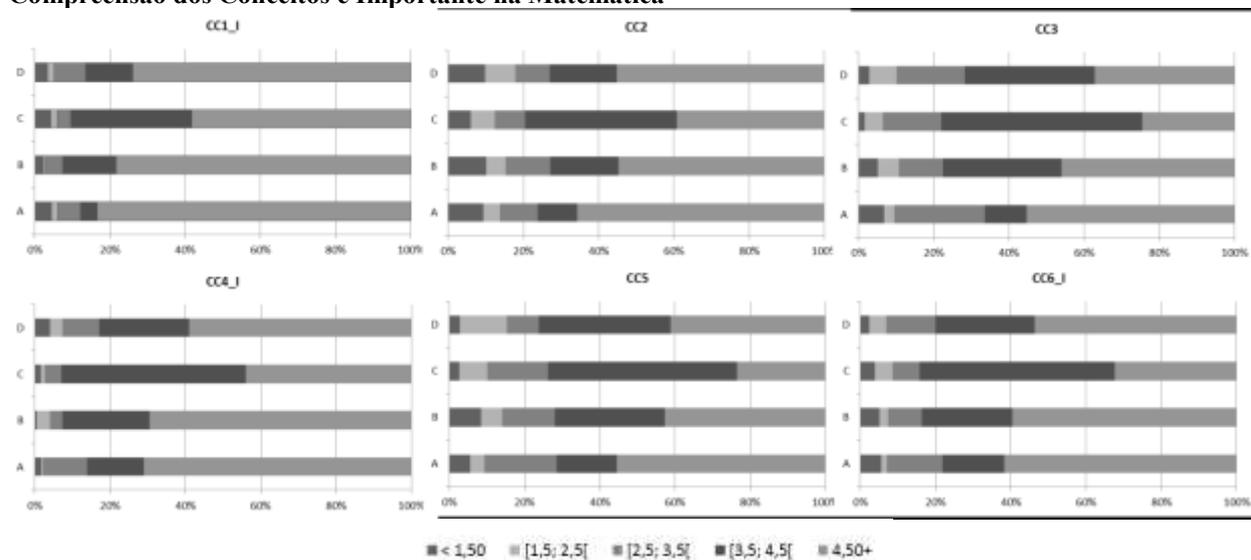
Tabela 3. Coeficientes de enviesamento por item e tipo de questionário.

	A - VAS (centro e extremos)		B - VAS (extremos)		C - Likert (todas)		D - Likert (extremos)	
	Skewness	Skewness /S.E. Skew	Skewness	Skewness /S.E. Skew	Skewness	Skewness /S.E. Skew	Skewness	Skewness /S.E. Skew
<b>Matemática como um Domínio Masculino</b>								
DM1_I	-2,185	-12,198	-2,325	-12,771	-2,140	-12,041	-2,626	<b>-14,381</b>
DM2_I	-1,514	-8,451	-1,777	-9,757	-1,179	-6,599	-1,535	-8,386
DM3	-0,286	-1,584	-0,245	-1,344	-0,365	<b>-2,055</b>	-0,328	<b>-1,785</b>
DM4	-1,912	-10,590	-2,511	-13,755	-1,652	-9,298	-2,106	-11,472
DM5	-2,349	-13,006	-2,548	-13,993	-1,797	-10,058	-2,249	-12,317
DM6_I	-2,561	-14,257	-2,815	-15,502	-2,002	-11,265	-2,322	-12,720
<b>Compreensão dos Conceitos é Importante na Matemática</b>								
CC1_I	-2,663	-14,748	-3,100	-16,934	-2,120	-11,927	-2,238	-12,257
CC2	-1,432	-7,931	-1,256	-6,879	-1,293	-7,257	-1,149	-6,239
CC3	-1,182	-6,511	-1,341	-7,363	-0,974	-5,483	-0,898	-4,919
CC4_I	-2,090	-11,540	-2,537	-13,894	-1,698	-9,530	-1,674	-9,141
CC5	-1,189	-6,599	-1,212	-6,658	-0,947	-5,330	-1,018	-5,576
CC6_I	-1,643	-9,150	-1,918	-10,532	-1,441	-8,064	-1,345	-7,324
<b>Utilidade da Matemática</b>								
U1_I	-1,689	-9,379	-2,151	-11,848	-1,160	-6,508	-1,405	-7,698
U2	-1,180	-6,550	-1,213	-6,623	-1,201	-6,756	-0,981	-5,326
U3	-0,372	-2,073	-0,539	-2,926	-0,412	-2,304	-0,468	-2,544
U4_I	-1,302	-7,250	-1,510	-8,272	-1,082	-6,090	-0,746	-4,062
U5_I	-1,021	-5,670	-1,378	-7,571	-0,802	-4,515	-0,883	-4,838
U6	-0,806	-4,501	-0,728	-3,996	-0,867	-4,867	-0,599	-3,282

## Matemática como um Domínio Masculino



## Compreensão dos Conceitos é Importante na Matemática



## Utilidade da Matemática

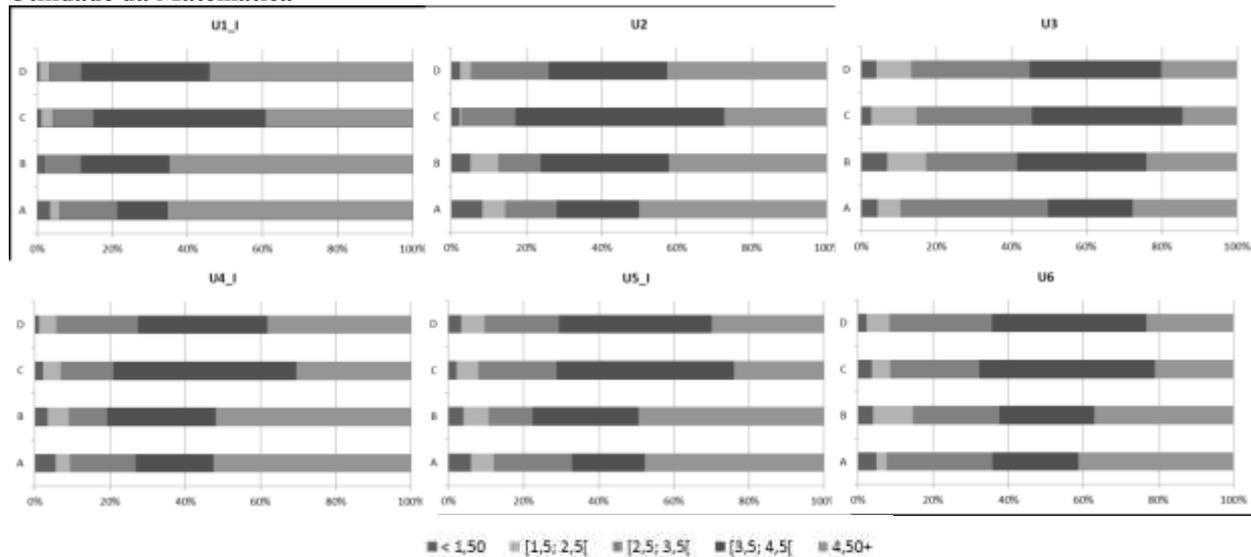


Figura 1. Distribuição de frequências dos itens de cada escala, por tipo de questionário.

Esse enviesamento é mais notório no caso dos itens que compõem a escala Matemática como Domínio Masculino (5 dos 6, em qualquer uma das apresentações), mas também pode ser observada na distribuição

dos itens que compõem a escala Compreensão dos Conceitos é Importante em Matemática, (em 4 dos 6 itens, para as apresentações VAS e para o tipo D, em 5 dos 6 na apresentação tipo C – discreta com todas as etiquetas). Mesmo no caso dos itens que correspondem à escala Utilidade da Matemática – com distribuição muito menos enviesada – os rácios, em valor absoluto, do enviesamento pelo respetivo erro padrão encontram-se genericamente acima do valor de referência de 2.

O padrão bastante enviesado das respostas pode ter origem na forma como as afirmações foram escritas, bem como na vontade por parte dos respondentes em indicar respostas socialmente aceitáveis, mas as razões desse comportamento não são objeto de estudo deste artigo. Na comparação entre os 4 formatos utilizados, verifica-se que as apresentações contínuas geraram – com a exceção de 2 itens – distribuições mais enviesadas (vd Tabela 3). Em geral, a apresentação discreta com todos os pontos etiquetados (C) origina distribuições ligeiramente menos enviesadas. É ainda de notar que, em termos de enviesamento, a apresentação discreta com etiquetas apenas nos extremos (tipo D) se aproxima das apresentações VAS. Comparando agora as duas apresentações VAS, aparentemente a não indicação do ponto intermédio (tipo B) está associada a maiores rácios (enviesamento/erro padrão).

Tabela 4. Associação entre a distribuição dos itens e o tipo de questionário

(Testes de *Kolmogorov-Smirnov* com correção de *Bonferroni*)

<b>Matemática como um Domínio Masculino</b>												
	A/B		A/C		A/D		B/C		B/D		C/D	
	K-S	sig										
DM1_I	2,011	0,001	1,826	0,003	1,981	0,001	2,700	0,000	3,957	0,000	1,310	0,065
DM2_I	1,824	0,003	2,842	0,000	1,246	0,089	3,000	0,000	2,970	0,000	1,562	0,015
DM3	1,118	0,164	2,073	0,000	0,894	0,402	2,148	0,000	1,464	0,028	1,154	0,139
DM4	1,685	0,007	2,929	0,000	1,524	0,019	2,934	0,000	3,149	0,000	1,551	0,016
DM5	2,094	0,000	3,330	0,000	1,721	0,005	3,185	0,000	3,797	0,000	2,204	0,000
DM6_I	1,856	0,002	1,873	0,002	2,471	0,000	3,192	0,000	4,298	0,000	1,162	0,134
<b>Compreensão dos Conceitos é Importante na Matemática</b>												
	A/B		A/C		A/D		B/C		B/D		C/D	
	K-S	sig										
CC1_I	2,258	0,000	2,782	0,000	1,461	0,028	2,836	0,000	3,697	0,000	1,499	0,022
CC2	2,091	0,000	3,120	0,000	1,568	0,015	2,773	0,000	2,840	0,000	1,510	0,021
CC3	1,304	0,067	3,524	0,000	2,280	0,000	3,821	0,000	2,574	0,000	1,210	0,107
CC4_I	2,005	0,001	3,647	0,000	2,182	0,000	4,087	0,000	3,057	0,000	1,427	0,034
CC5	2,254	0,000	4,072	0,000	2,339	0,000	3,708	0,000	2,298	0,000	1,689	0,007
CC6_I	1,438	0,032	3,860	0,000	1,794	0,003	4,295	0,000	2,618	0,000	2,018	0,001
<b>Utilidade da Matemática</b>												
	A/B		A/C		A/D		B/C		B/D		C/D	
	K-S	sig										
U1_I	1,152	0,141	3,508	0,000	2,045	0,000	3,935	0,000	2,473	0,000	1,428	0,034
U2	1,499	0,022	3,396	0,000	1,897	0,002	3,408	0,000	2,171	0,000	1,448	0,030
U3	1,155	0,139	2,426	0,000	1,868	0,002	2,101	0,000	1,715	0,006	0,523	0,948
U4_I	1,131	0,155	3,585	0,000	2,788	0,000	4,097	0,000	3,298	0,000	0,742	0,641
U5_I	1,321	0,061	3,599	0,000	2,993	0,000	4,140	0,000	3,530	0,000	0,561	0,912
U6	1,447	0,030	2,897	0,000	2,652	0,000	3,144	0,000	2,899	0,000	0,318	1,000

Nota: Estão assinalados com sombreado os casos para os quais as diferenças são significativas, tendo-se ajustado o alfa com a correção de *Bonferroni* para  $\alpha/c$ , em que  $c$  é o número de combinações possíveis ( $\alpha/6=0,008$ ).

A relação entre cada item e os quatro tipos de questionário foi analisada com recurso ao teste de independência de *Kolmogorov-Smirnov* (Tabela 4). Como é patente, em todas as dimensões analisadas, foram encontradas diferenças significativas na grande maioria dos casos: nas dimensões Matemática como Domínio Masculino e Importância da Compreensão dos Conceitos em cerca 3/4 dos casos foi encontrada uma associação com o tipo de questionário; na dimensão Utilidade da Matemática foram encontradas diferenças significativas em aproximadamente 2/3 dos testes efetuados.

Relativamente ao tipo de diferenças encontradas, saliente-se o facto dos questionários A e B (VAS) se distinguirem claramente dos restantes, em todas as dimensões analisadas. Igualmente de realçar é também o facto terem sido encontradas menos diferenças entre os itens quando comparados os formatos de questionário C e D (escalas tipo-Likert). Na quase totalidade dos casos, as diferenças entre estes dois tipos de formato de resposta não são significativas, seja qual for a dimensão considerada. Já no que se refere aos formatos A e B, quando comparados entre si, só na dimensão Utilidade da Matemática não foram encontradas diferenças significativas.

Foram também realizados testes não paramétricos de *Kruskal-Wallis* para aferir da diferença de distribuição de cada um dos itens, consoante o questionário. Os resultados podem ser consultados na Tabela 5. Tal como nos testes anteriormente efetuados, os itens que compõem a escala Utilidade da Matemática são os que denotam menos diferenças de distribuição entre tipos de questionário (apenas dois casos).

Tabela 5. Associação entre a distribuição dos itens e o tipo de questionário  
(Testes de *Kruskal-Wallis*)

<b>Matemática como um Domínio Masculino</b>						
	DM1_I	DM2_I	DM3	DM4	DM5	DM6_I
Qui-quadrado	32,068	14,182	1,568	15,783	31,21	38,839
Sig (aprox)	0,000	0,003	0,667	0,001	0,000	0,000
<b>Compreensão dos Conceitos é Importante na Matemática</b>						
	CC1_I	CC2	CC3	CC4_I	CC5	CC6_I
Qui-quadrado	23,059	10,435	5,158	9,441	14,979	14,148
Sig (aprox)	0,000	0,015	0,161	0,024	0,002	0,003
<b>Utilidade da Matemática</b>						
	U1_I	U2	U3	U4_I	U5_I	U6
Qui-quadrado	9,868	3,264	1,213	5,769	8,86	4,043
Sig (aprox)	0,020	0,353	0,750	0,123	0,031	0,257

Nota: Estão assinalados com sombreado os casos para os quais as diferenças são significativas

Em termos das comparações múltiplas efetuadas (Tabela 6), é de notar a existência de diferenças significativas entre os dois tipos discretos de itens, em especial no caso dos que compõem a escala Matemática como Domínio Masculino. O tipo contínuo A apresenta mais frequentemente diferenças significativas com o tipo B do que com os tipos discretos, em especial o D. Aparentemente, a utilização de descritores nos extremos e a marcação explícita do centro nesse tipo contínuo origina distribuições de respostas relativamente similares ao caso discreto, D, onde apenas os extremos possuem tais descritores, talvez porque nos tipos discretos a categoria intermédia seja facilmente identificável, possua ou não descritores verbais.

Tabela 6. Testes de comparações múltiplas associados aos testes *Kruskal-Wallis* significativos

		A/B	A/C	A/D	B/C	B/D	C/D
<b>Matemática como um Domínio Masculino</b>							
DM1	teste (z)	2,834	-0,023	-2,874	-3,071	-5,657	-2,659
	<i>sig ajust.</i>	0,028	1,000	0,024	0,013	0,000	0,047
DM2	teste (z)	2,035	1,880	-1,183	-0,017	-3,186	-3,043
	<i>sig ajust.</i>	0,251	0,360	1,000	1,000	0,009	0,014
DM4	teste (z)	2,265	1,613	-1,359	-0,679	-3,597	-2,969
	<i>sig ajust.</i>	0,141	0,640	1,000	1,000	0,002	0,018
DM5	teste (z)	3,015	2,687	-1,797	-0,355	-4,787	-4,478
	<i>sig ajust.</i>	0,015	0,043	0,434	1,000	0,000	0,000
DM6	teste (z)	2,762	-0,902	-3,443	-3,674	-6,164	-2,567
	<i>sig ajust.</i>	0,034	1,000	0,003	0,001	0,000	0,062
<b>Compreensão dos Conceitos é Importante na Matemática</b>							
		A/B	A/C	A/D	B/C	B/D	C/D
CC1	teste (z)	3,206	1,662	-1,324	-1,582	-4,503	-2,967
	<i>sig ajust.</i>	0,008	0,579	1,000	0,682	0,000	0,017
CC2	teste (z)	2,638	2,227	0,338	-0,441	-2,276	-1,864
	<i>sig ajust.</i>	0,050	0,156	1,000	1,000	0,137	0,374
CC4	teste (z)	1,878	2,726	0,478	0,822	-1,392	-2,229
	<i>sig ajust.</i>	0,362	0,038	1,000	1,000	0,984	0,155
CC5	teste (z)	2,329	3,630	0,995	1,265	-1,324	-2,603
	<i>sig ajust.</i>	0,119	0,002	1,000	1,000	1,000	0,055
CC6	teste (z)	0,820	3,089	-0,329	2,244	-1,138	-3,384
	<i>sig ajust.</i>	1,000	0,012	1,000	0,149	1,000	0,004
<b>Utilidade da Matemática</b>							
		A/B	A/C	A/D	B/C	B/D	C/D
U1	teste (z)	0,589	2,508	-0,380	1,906	-0,963	-2,873
	<i>sig ajust.</i>	1,000	0,073	1,000	0,340	1,000	0,024
U5	teste (z)	-0,325	2,245	1,479	2,559	1,794	-0,740
	<i>sig ajust.</i>	1,000	0,149	0,835	0,063	0,437	1,000

Nota: Estão assinalados com sombreado os casos para os quais as diferenças são significativas

### 3.2. Análise das escalas

As três escalas foram construídas tendo sido analisada a sua consistência através do *Alpha* de *Cronbach*, para cada tipo de questionário. Os valores apresentaram-se elevados nas escalas Domínio Masculino e Utilidade da Matemática, semelhantes aos encontrados em estudos anteriores (Fennema and Sherman, 1976; Mulher and Rae, 1998; Kloosterman and Stage, 1992; Mason, 2003). Na escala Importância da Compreensão dos Conceitos os valores são mais baixos, para os questionários tipo A e B (VAS), com valores inferiores a 0,6 (Tabela 7). Verifica-se que, no tipo de questionário A (segmento com etiquetagem no centro e nos extremos) o item CC2, “É tão importante encontrar a resposta certa a um problema matemático como saber porque é que se acertou”, surge muito pouco correlacionado com os restantes. Este efeito surge acentuado apenas nesta escala. No tipo de questionário B, a questão que apresenta baixa correlação com a restante é agora a CC5, “ Uma pessoa que não entende porque é que determinada resposta a um problema matemático está correta, não resolveu verdadeiramente o problema”. Não existe assim um padrão, um item que se revele inconsistente no conjunto de questionários cuja opção de resposta é dado por um segmento de reta.

Pode também ser verificado que todas as distribuições apresentam acentuado enviesamento, sendo a escala Domínio Masculino a que apresenta um rácio (skewness/standard error) mais desfavorável, mais afastado da referência de simetria. A escala com menor enviesamento é a Utilidade da Matemática. Para cada escala, os níveis de assimetria são semelhantes para os diferentes tipos de questionários, no entanto surgem regularmente valores absolutos mais baixos no tipo B, VAS com etiqueta no centro.

Tabela 7. Análise de consistência e assimetria

Escala e Tipo de questionário	N	Média	D.P	Alpha	Skewness	Skewness /S.E. Skew
<b>Matemática como um Domínio Masculino<sup>(a)</sup></b>						
A - VAS (centro e extremos)	184	19,36	2,441	0,732	-1,680	-9,23
B - VAS (extremos)	179	19,09	2,433	0,808	-1,637	-8,89
C - Likert (todas)	187	18,85	2,560	0,796	-1,711	-9,52
D - Likert (extremos)	177	19,40	2,348	0,776	-1,727	-9,38
<b>Compreensão dos Conceitos é Importante na Matemática</b>						
A - VAS (centro e extremos)	184	25,63	3,691	0,582	-1,157	-6,14
B - VAS (extremos)	179	25,35	3,164	0,473	-,552	-2,95
C - Likert (todas)	187	24,56	3,580	0,711	-1,722	-9,59
D - Likert (extremos)	177	25,02	3,849	0,618	-,772	-4,16
<b>Utilidade da Matemática</b>						
A - VAS (centro e extremos)	184	23,85	4,386	0,736	-,725	-3,97
B - VAS (extremos)	179	24,07	3,908	0,724	-,535	-2,86
C - Likert (todas)	187	23,36	4,025	0,835	-,969	-5,39
D - Likert (extremos)	177	23,80	3,752	0,737	-,405	-2,17

(a) O item MD3 – Os rapazes não são naturalmente melhores do que as mulheres na matemática – foi excluído da escala devido à baixa consistência da escala com a sua inclusão.

A avaliação e comparação das distribuições de cada escala por tipo de questionário foi feita graficamente, através de histogramas (Figura 2) e através da realização de testes de comparação de distribuições de *Kolmogorov-Smirnov* e de *Kruskal-Wallis*, com associação a testes de comparação múltipla. Para detetar diferenças nas médias das distribuições foram usados testes *t*. Sempre que necessário foram efetuadas correções de *Bonferroni* para a tomada de decisão (Tabelas 8 e 9). Devido à acentuada assimetria das escalas os resultados destes últimos testes devem ser cautelosamente interpretados.

Na escala Matemática como Domínio Masculino verifica-se que a distribuição obtida com o questionário tipo C, Likert com etiquetas em todas as categorias, difere das restantes distribuições, apresentando uma assimetria negativa mais evidente e maior dispersão de valores. Verificaram-se também diferenças significativas entre a forma das distribuições para os tipos B e D, com forma de resposta por segmento ou por categorias, ambas etiquetadas apenas nos extremos. A apresentação da forma de resposta em categorias gerou uma distribuição com uma assimetria negativa mais acentuada que a distribuição gerada pela apresentação de um segmento de reta (Figura 2 e Tabela 8). Nesta escala não foram encontradas diferenças entre os valores médios, obtidos para os diferentes tipos de questionários.

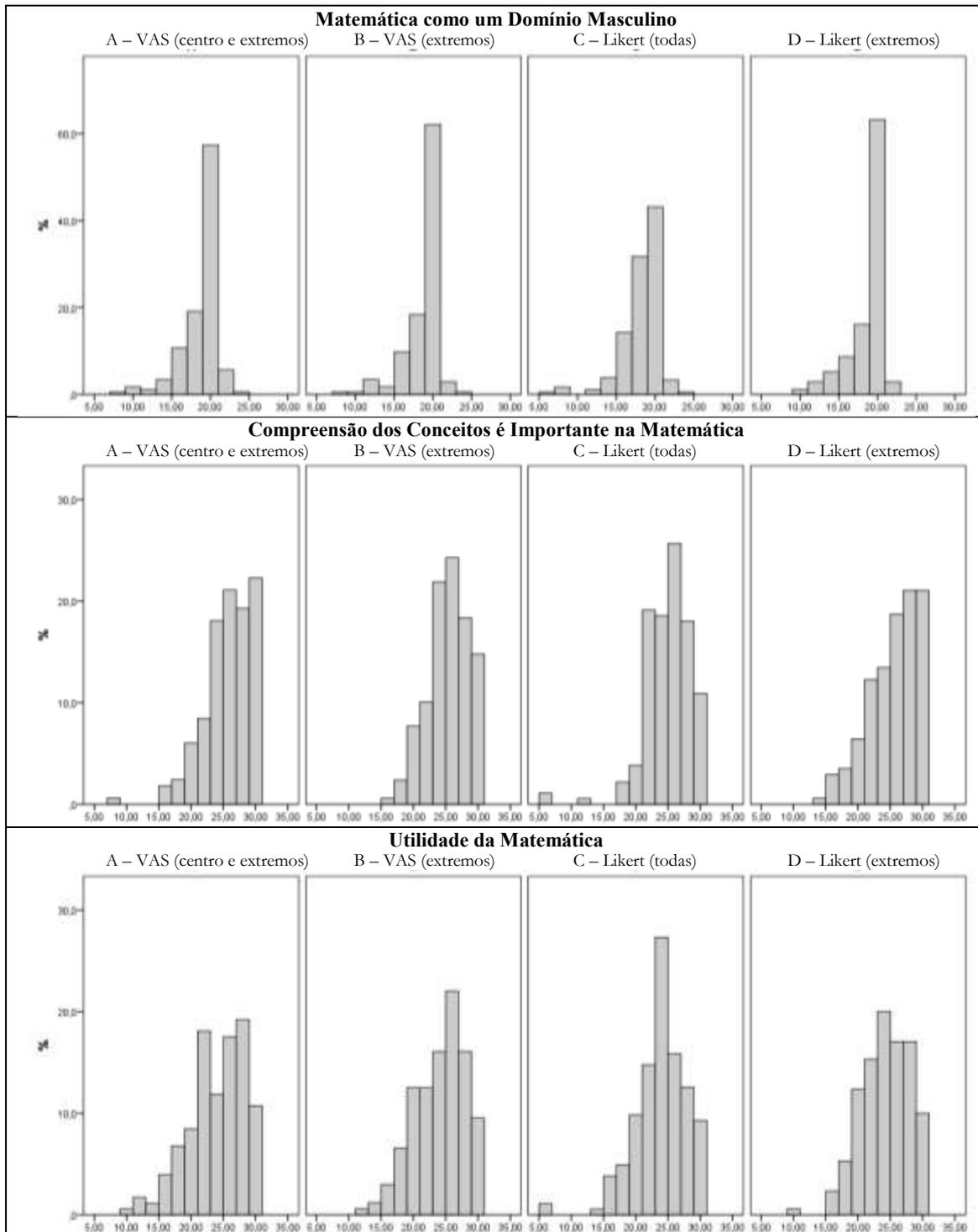


Figura 2. Histograma de cada escala por tipo de questionário.

Na escala Importância da Compreensão dos Conceitos, apenas a distribuição obtida com o questionário tipo A difere da distribuição para o tipo C, respectivamente VAS e Likert com maior número de etiquetas. Verifica-se que a apresentação em categorias todas etiquetadas (C) possibilitou uma distribuição de valores com assimetria negativa mais acentuada que a apresentação em segmento. Em consequência da forma da distribuição, as médias também surgiram diferentes, com o tipo C a apresentar menor valor médio.

A escala Utilidade da Matemática apresentou apenas diferenças entre as distribuições obtidas com os questionários tipo B (VAS-extremos) e C (Likert-todas). Para o tipo C a distribuição surge com uma assimetria negativa mais acentuada. As médias das escalas surgiram semelhantes para todos os tipos de questionários.

Tabela 8. Testes para as diferenças das distribuições e médias das escalas por tipos de questionários

	A/B	A/C	A/D	B/C	B/D	C/D
<b>Matemática como um Domínio Masculino</b>						
<b>K-S (z)</b>	1,427	2,322	1,088	2,047	2,144	1,804
<b>sig</b>	0,034	0,000	0,187	0,000	0,000	0,003
<b>t</b>	1,035	1,918	-0,174	0,893	-1,223	-2,112
<b>sig</b>	0,301	0,056	0,862	0,373	0,222	0,035
<b>Compreensão dos Conceitos é Importante na Matemática</b>						
<b>K-S (z)</b>	1,004	2,094	1,133	1,534	0,954	0,236
<b>sig</b>	0,266	0,000	0,153	0,018	0,323	0,094
<b>t</b>	0,758	2,764	1,485	2,191	0,853	-1,180
<b>sig</b>	0,449	0,006	0,138	0,029	0,394	0,239
<b>Utilidade da Matemática</b>						
<b>K-S (z)</b>	0,523	1,599	1,078	1,763	1,143	0,602
<b>sig</b>	0,948	0,012	0,196	0,004	0,147	0,862
<b>t</b>	-0,480	1,116	0,114	1,674	0,637	-1,072
<b>sig</b>	0,631	0,265	0,909	0,095	0,525	0,285

Nota: Estão assinalados com sombreado os casos para os quais as diferenças são significativas, tendo-se ajustado o alfa com a correção de *Bonferroni* para alfa/c, em que *c* é o número de combinações possíveis ( $\alpha/6=0,008$ ).

Ao comparar as distribuições das escalas nos 4 tipos de questionários (teste de *Kruskal-Wallis*), no que refere à escala Matemática como Domínio Masculino, verificou-se que a diferença surge principalmente entre os tipos de questionários C e D, onde as opções de resposta são tipo Likert, sugerindo que a apresentação de etiquetas em todas as categorias possibilita um menor valor médio de ordenações do que quando etiquetada apenas nos extremos. A semelhança entre o tipo D e os A e B indica que a distribuição obtida por opção de resposta tipo Likert apenas com etiquetas nos extremos se assemelha a uma opção de VAS, como se pode observar na Tabela 9.

Tabela 9. Testes de comparação de distribuições das escalas nos 4 tipos de questionários

(Teste *Kruskal-Wallis* e testes de comparações múltiplas)

<b>Matemática como um Domínio Masculino <math>K-W_{(3)}=12,295;Sig=0,006</math></b>						
	A/B	A/C	A/D	B/C	B/D	C/D
teste (z)	1,515	2,596	-0,549	1,056	-2,053	-3,134
sig ajust.	0,779	0,057	1,000	1,000	0,241	0,010
<b>Compreensão dos Conceitos é Importante na Matemática <math>K-W_{(3)}=11,246;Sig=0,01</math></b>						
	A/B	A/C	A/D	B/C	B/D	C/D
teste (z)	1,189	3,301	1,536	2,098	0,344	-1,753
sig ajust.	1,000	0,006	0,748	0,215	1,000	0,477
<b>Utilidade da Matemática <math>K-W_{(3)}=3,767;Sig=0,288</math></b>						

Nota: Estão assinalados com sombreado os casos para os quais as diferenças são significativas

Em relação à escala Importância da Compreensão dos Conceitos, as diferenças encontradas referem-se às distribuições geradas com os questionários tipo A (VAS – centro e extremos) e C (Likert – todas), ou seja, as distribuições geradas por apresentações em segmento ou categorias, mesmo que tenham maior número de identificações com etiquetas não são iguais. A distribuição obtida com o tipo de questionário A apresenta maior média de ordenações, que a distribuição obtida com o questionário tipo C, a mediana é menor nesta situação.

Não se verificaram diferenças entre as distribuições geradas por diferentes tipos de questionários, na escala Utilidade da Matemática

#### 4. Conclusão e Discussão

Na presente investigação foi realizado um estudo com os alunos do primeiro ano de diferentes licenciaturas (ciências sociais, de gestão e tecnológicas). Foram aplicados quatro tipos de questionários com diferentes formas de apresentação de resposta: segmentos de reta com etiquetas no centro e nos extremos; segmento de reta com etiquetas apenas nos extremos; itens tipo-Likert com 5 pontos, todos etiquetados e itens tipo-Likert com 5 com apenas os extremos etiquetados.

Este estudo tem como principal objectivo analisar em que medida o formato de resposta oferecido condiciona os resultados obtidos, sendo comparadas as distribuições obtidas, item a item ou escalas, nos diferentes tipos de questionários.

Numa análise item a item é possível concluir, que independentemente do formato de resposta, a maior parte dos itens apresenta distribuições muito enviesadas, com a maioria das respostas situadas acima de 3,5. Comparando os quatro tipos de questionário, as apresentações contínuas, em geral, produziram distribuições mais enviesadas. A apresentação com todos os pontos etiquetados (C) apresenta distribuições ligeiramente menos enviesadas. Destaca-se ainda a semelhança entre a apresentação discreta com etiquetas apenas nos extremos (tipo D) e as apresentações VAS.

No que respeita à forma das distribuições, aferidas com recurso a testes de *Kolmogorov-Smirnov*, esta depende do tipo de formato de resposta utilizado, principalmente nas dimensões Matemática como Domínio Masculino e Importância da Compreensão dos Conceitos (as mais enviesadas). Mais especificamente, as apresentações contínuas distinguem-se das restantes e entre si (com exceção dos itens referentes à Utilidade da Matemática). Os formatos discretos não apresentam diferenças significativas na grande generalidade dos casos.

Foram também realizados testes não paramétricos de *Kruskal-Wallis* mais sensíveis à ordenação das observações, que, tal como os anteriores, detetam menos diferenças nos itens que compõem a escala Utilidade da Matemática. Contudo, no que se refere à comparação das apresentações discretas, é de notar a existência de diferenças significativas entre os dois tipos discretos de formato, em especial no caso dos itens que compõem a escala Matemática como Domínio Masculino. Também os formatos contínuos revelaram comportamentos diferenciados entre si. Uma conclusão também importante é o facto de os formatos mais semelhantes serem o A e o D, ou seja, quando apenas os extremos possuem descritores verbais e o ponto central da escala está facilmente identificado.

Os resultados obtidos por comparação das escalas mostram um acentuado enviesamento em todas as dimensões, sendo o formato contínuo com etiquetas apenas nos extremos o que apresenta menor assimetria.

Genericamente, não se verificaram diferenças na forma e média das distribuições segundo o formato de resposta; a existirem, são mais frequentes na escala mais enviesada (Matemática como Domínio Masculino) e associadas ao formato de resposta discreto com etiquetas em todas as categorias (C).

A discussão sobre a correta forma de analisar e comparar respostas em formatos como os aqui analisados nasceu com a proposta de Likert em 1932 e ainda hoje se mantém. A utilização de formatos VAS surge aliás como uma tentativa de obtenção de variáveis métricas sobre as quais a utilização de métodos paramétricos é mais consensual.

Com este trabalho quisemos alargar a discussão á comparação de formatos discretos e contínuos. O comportamento genericamente similar, em termos distribucionais, entre itens com apresentação VAS que incluem a identificação do ponto central (A) e entre itens com apresentação discreta de extremos etiquetados (D), leva-nos a defender a utilização deste último formato se operacionalmente a utilização de VAS não for adequada mas se pretenda atingir a proporcionalidade de afastamentos.

## References:

Adelson, J.L. e McCoach (2010). Measuring the Mathematical Attitudes of Elementary Students: The Effects of a 4-Point or 5-Point Likert-Type Scale, *Educational and Psychological Measurement*, 70, 796-807.

Aiken, L. R. (1983). Number of response categories and statistics on a teacher rating scale. *Educational and Psychological Measurement*, 43, 397-401.

Alwin, D.F. e Krosnick, J.A (1991). The reliability of survey attitude measurement: the influence of question and respondent attributes. *Sociological Methods e Research*, 20(1), 139-181.

Bishop, G.F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51, 220-232.

Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18(3), 205-215.

Cummins, R.A e Gullone, E. (2000). Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. *Proceedings, Second International Conference on Quality of Life in Cities*. Singapore, National University of Singapore, 74-93.

Dawes, J. (2008), Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales, *International Journal of Market Research*, 50(1), 61-77.

Dixon, P.N, Bobo, M. e Stevick, R.A. (1984). Response differences and preferences for all-category defined and end-defined Likert formats. *Educational and Psychological Measurement*, 44, 61-66.

Fennema, E. H. e Sherman, J.A. (1976). Fennema-Sherman mathematics attitudes scales: instrument designed to measure attitudes toward mathematics. *Journal for Research in Mathematics Education*, 7(5), 324-326.

Göb, R., McCollin C. e Ramalhoto M.F (2007). Ordinal methodology in the analysis of Likert Scales. *Quality e Quantity*, 41, 601-626.

Jacoby, J. e Matell, M.S. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research*, VIII, 495-500.

Kloosterman, P. e Stage, F.K. (1992). Measuring beliefs about mathematical problem solving. *School Science and Mathematics*, 92, 109-115.

Krosnick, J.A. e Berent, M.K. (1993). Comparisons of party identification and policy preferences: the impact of survey questions format. *American Journal of Political Science*, 37(3), 941-964.

Landrum, R.E. (1999). Scaling issues in faculty evaluations. *Psychological Report*, 84, 178-180.

Likert, R.A. (1932/1966). "The Method of Constructing an Attitude Scale", appendix of "A Technique for the Measurement of Attitudes", *Archives of Psychology*, 140 (Jun 1932), included in *Readings in Attitude Theory and Measurement*, Fishbein (ed), 1966, 90-95, John Wiley e Sons, Inc.

Moors, G. (2007). Exploring the effect of a middle response category on response style in attitude measurement. *Quality e Quantity*, 42(6), 779-794.

Presser, S. e Schuman, H. (1980). The measurement of a middle position in attitude surveys. *Public Opinion Quarterly*, 4, 70-85.

Preston C.C e Colman, A.M (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1-15.

Newstead, S.E. e Arnold, J. (1989). The effect of response format on ratings of teaching. *Educational and Psychological Measurement*, 49, 33-43.

Svensson, E. (2000). Comparison of the quality of assessments using continuous and discrete ordinal rating scales. *Biometrical Journal*, 42(4), 417-434.