

# DILEMAS NA CONSTRUÇÃO DE ESCALAS TIPO LIKERT: O NÚMERO DE ITENS E A DISPOSIÇÃO INFLUENCIAM NOS RESULTADOS?\*

DILEMMAS OF THE TYPE LIKERT SCALES CONSTRUCTION: DOES THE NUMBER OF ITEMS AND THE DISPOSITION INFLUENCE RESULTS?

MARLON DALMORO\*\*  
KELMARA MENDES VIEIRA\*\*\*

## RESUMO

No presente estudo buscou-se avaliar a influência do número de itens na escala tipo Likert e o efeito da disposição da escala nos resultados de uma mensuração. Juntamente, avaliou-se qual a escala preferida pelos respondentes quanto à facilidade, à velocidade e à precisão de resposta. Para isto, foram construídos quatro questionários, todos com as mesmas questões, mas com escalas distintas. Os três primeiros questionários apresentavam respectivamente escala tipo Likert de três, cinco e sete pontos. O quarto questionário apresentava a escala tipo Likert de cinco pontos, mas com a disposição invertida. Foram entrevistadas 211 pessoas, as quais responderam a questionários de maneira subsequente. Os resultados revelaram que a escala de três pontos é menos confiável e tem menos capacidade de demonstrar com precisão a opinião do entrevistado, mas foi considerada a escala mais fácil e veloz. A escala de cinco pontos teve, em média, a mesma precisão e mostrou-se mais fácil e mais rápida que a escala de sete pontos. Portanto, para este estudo a escala que se mostrou mais adequada foi a de cinco pontos. A inversão do formato da escala mostrou que alguns entrevistados mudaram de posição, apesar do efeito médio não ser significativo. Por fim, os resultados suportam algumas reflexões relevantes no processo de construção de escalas tipo Likert.

**Palavras-chave:** Mensuração de dados. Escalas tipo Likert. Número de itens. Efeito da disposição da escala.

## ABSTRACT

In this study, initially was intended to evaluate the influence of the number of items on the Likert scale and the effect of scale's arrangement in the results of a measurement. In addition, was evaluated what the scale preferred by the respondents as to the easy, speed and accuracy of response. For this, four questionnaires were built, all with the same issues but with different scales. The first three questionnaires were respectively Likert scale of three, five and seven points. The fourth questionnaire shows the Likert scale of five points, but with the reversed provision. Were interviewed 211 people who responded to questionnaires on a subsequent way. The results revealed that the scale of three points is less reliable and has less able to demonstrate exactly the view of the interviewee, but the scale was considered easier and fast. The scale of five points had, on average, the same precision and was easier and faster than the scale of seven points. So for this study the best scale was the five points. The reversal of the format of the scale showed that some respondents changed their position, despite the average effect is not significant. Finally, results support some relevant considerations in the process of Likert scales elaboration.

**Keywords:** Data measurement. Likert scale. Number of items. Scale arrangement effect.

\* Data de submissão: 10/06/2013. Data de aceite: 18/03/2014.

\*\* Doutor em Administração; Professor do Centro de Gestão Organizacional - UNIVATES

\*\*\* Doutora em Administração; Professora Adjunta da UFSM.

## 1. INTRODUÇÃO

Na área das ciências sociais é comum o uso de diversos instrumentos de medida para mensurar a realidade sobre um objeto em estudo. Para realizar essas mensurações, os pesquisadores precisam desenvolver instrumentos adequados, de forma que as medidas correspondam efetivamente ao que se deseja medir (possuir validade) e para que o erro amostral seja o menor possível (aumentar a confiabilidade) diante dos recursos disponíveis e, desta forma, obter resultados que sejam um reflexo da realidade (Cote; Buckley, 1988; Clark; Watson, 1995).

Para estas medições, os pesquisadores geralmente utilizam escalas de mensuração multi-item. A utilização de tais escalas tem sido apontada como requerente de cuidados essenciais para o alcance de resultados satisfatórios, confiáveis e que permitam conclusões apropriadas (Krosnick; Berent, 1993; Matos; Trez, 2012). Esta ressalva decorre do fato de que alguns pesquisadores e profissionais utilizam-se de instrumentos de coleta de dados que não medem aquilo que se pretende medir – ou por desconhecimento, ou por falta de instrução – inviabilizando, assim, todo o processo de pesquisa. Collings (2006) destaca que diversos autores têm argumentado a falta de atenção dada a certos aspectos no desenho de um questionário baseado no formato Likert, principalmente quanto à validade das alternativas de repostas. Destaca-se que a validação de uma questão não pode se dar sem considerar as opções de repostas como um componente essencial na elaboração de um questionário.

As falhas na medição das informações por meio de instrumentos de coleta de dados podem ocorrer devido a diversos fatores. O primeiro fator passa pela escala de mensuração utilizada, como exemplo, o número de pontos na escala, rotação dos pontos da escala e a força das âncoras. O segundo centra-se no respondente e nas suas características, como exemplos, necessidade de cognição, envolvimento, conhecimento. Enquanto diversos estudos abordaram esta segunda questão, poucos estudos analisaram o número de itens e a disposição destes em uma escala de mensuração (Weathers; Sharma; Niedrich, 2005).

Desde que Rensis Likert introduziu seu método em 1932, pesquisadores de áreas como psicologia, educação e marketing têm se valido do uso de diferentes formatos de escalas tipo Likert. Contudo, em uma revisão da literatura, Preston e Coleman (2000) colocam que, mesmo depois de décadas de pesquisa, não existe um consenso em termos do número de opções de repostas em uma esca-

la de mensuração. Esta demanda por uma melhor compreensão do tema é motivada pelo reconhecimento de que o número de opções, ou pontos na escala, podem produzir resultados conflitantes (Chang, 1994).

Desta forma, a literatura tem constantemente reforçado que uma das decisões-chave na construção do questionário - especialmente no caso de questionários tipo Likert -, recai sobre o número de categorias empregadas na escala de resposta. Esta demanda é relevante, visto o reconhecimento de que os diversos formatos de escalas tipo Likert utilizados têm distorcido instrumentos de mensuração e seus respectivos construtos, fruto de uma aparente displicência no uso de regras pelos pesquisadores na definição de suas escalas (Cummins; Gullone, 2000). A situação se agrava na medida em que os pesquisadores dispõem de opções de escalas com número de pontos de escolha que variam de dois até 100, escalas unidimensionais (exemplo: não satisfeito até completamente satisfeito), escalas bi-dimensionais (exemplo: completamente insatisfeito até completamente satisfeito), opção do uso do ponto neutro na escala, ou não, uso de âncoras verbais extremas (exemplo: terrível), ou âncoras médias (exemplo: insatisfeito), entre outras decisões relevantes na elaboração de uma escala.

Cada uma destas diferentes formas de estruturar uma escala de mensuração altera a avaliação psicométrica do respondente. Ainda que apontada pela literatura, esta avaliação não está totalmente compreendida, especialmente no que tange ao número de pontos na escala de mensuração (Wiswanathan; Sudman; Johnson, 2004; Weathers; Sharma; Niedrich, 2005; Collings, 2006). Diante desta contextualização, o presente estudo possui objetivos distintos. Inicialmente busca-se avaliar a influência do número de itens na escala tipo Likert e o efeito da disposição da escala nos resultados de uma mensuração. Juntamente, avalia-se qual a escala preferida pelos respondentes quanto à facilidade, à velocidade e à precisão de resposta.

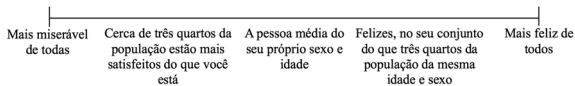
Para o alcance destes objetivos, o estudo está dividido em cinco partes, incluindo esta parte introdutória. O segundo tópico busca um entendimento sobre o tema, bem como a apresentação de resultados obtidos em estudos similares. Posteriormente, é detalhado o método utilizado para a realização da pesquisa empírica e no quinto tópico são apresentados os resultados. Por fim, destaca-se as considerações finais do estudo.

## 2. REFERENCIAL TEÓRICO

A história da construção de escalas de mensuração tem como trabalho seminal o estudo de

Rensis Likert, publicado em 1932. No entanto, é interessante destacar que anteriormente a esta data já se realizavam avaliações subjetivas com o uso de escalas, as quais eram mais sensíveis do que as utilizadas atualmente. Cummins e Gullone (2000) destacam a introdução do *Graphic rating method*, por Freyd, em 1923, utilizando uma escala no formato de 10 pontos ou de 100 pontos. A explicação para o uso destas escalas estava na facilidade de compreender o sistema de numeração de 0 a 10. O *Graphic rating method* deveria ser utilizado em conjunto com entrevistas, e o respondente deveria marcar um ponto apropriado em uma linha horizontal pontilhada. Poucos anos depois, Watson (1930) publicou uma escala de mensuração de felicidade, na qual o respondente marcaria um ponto em qualquer lugar na linha horizontal. Para análise dos dados, o autor recomendava a utilização de escores de 0 a 100, e a sua escala apresentava o formato conforme a Figura 01:

Figura 01: Modelo de escala desenvolvido por Watson (1930)



Fonte: Watson (1930)

Com base nestes formatos iniciais, em 1932, Likert desenvolveu sua forma de escala, conforme a Figura 02:

Figura 02: Modelo de escala desenvolvido por Likert (1932)



Fonte: Likert (1932)

Baseado nos modelos anteriores, Likert (1932) reduziu o número efetivo de pontos de escolha, preservando o sistema de medida contínuo. Na escala de Likert, os respondentes escolheriam somente um dos pontos fixos estipulados na linha, em um sistema de cinco categorias de resposta (pontos), partindo de “aprovo fortemente” até “desaprovo fortemente”. A escala de Likert (1932) também introduzia o caráter bidimensional da escala e com um ponto neutro no meio da escala.

Desde a publicação de sua obra, a escala formulada por Likert tem se tornado popular. As razões para isto incluem o tipo de psicometria utilizada na investigação, a dificuldade de generalizações com o uso de grande número de opções de marcação e a natureza complexa de escalas alternativas (Cum-

mins; Gullone, 2000). O trabalho de Likert (1932) deixava claro que sua escala centrava-se na utilização de cinco pontos, e não mencionava o uso de categorias de respostas alternativas na escala a ser utilizada. Embora o uso de escalas com outro número de itens, diferente de cinco, represente uma escala de classificação, quando esta não contiver cinco opções de resposta, não se configura uma escala Likert, mas sim do “tipo Likert”. No entanto, como Clason e Dormody (1994) ressaltam, muitos estudos têm usado diversas opções, paralelas à escala tradicional de cinco pontos, obtendo resultados satisfatórios. Neste caso, a escala se configura como do tipo Likert.

A variação no número de itens da escala originalmente proposta por Likert tem fomentado inúmeras discussões sobre a escolha da escala a ser utilizada. Um dos primeiros trabalhos a levantar esta questão foi o de Garner e Hake (1951) que, analisando a transmissão da informação, relataram que o aumento do número de categorias de respostas de uma escala aumenta o montante de informação transmitida pela escala. Isto impacta diretamente na forma como o entrevistado as interpretará. Ao analisar um objeto, o respondente processa mentalmente as informações disponíveis e suas respostas podem estar sujeitas às influências que comprometem a validade das medidas utilizadas. A complexidade na escolha do tamanho da escala surge em virtude de que, conforme aumenta o número de pontos na escala, aumenta a complexidade de escolha do respondente e a discriminação entre cada opção de respostas (Campell, 1988).

Tourangeau e Rasinski (1988) colocam que a resposta de uma escala envolve um processo mental de quatro estágios, nos quais o respondente: (1) interpreta o item, (2) recupera pensamentos e sentimentos relevantes, (3) formula um julgamento baseado nestes pensamentos e sentimentos, e (4) seleciona uma resposta. De acordo com a capacidade de processamento mental dos respondentes, este processo pode ser uma ação simples ou complexa. Caso seja complexa, os respondentes tendem a simplificar a tarefa com o uso de decisões heurísticas (Swait; Adarnowicz, 2001). Um dos modelos heurísticos que os respondentes tendem a utilizar é denominado de *status quo* heurístico (ou heurística habitual) em que a regra de decisão utilizada pelos respondentes é a de selecionar a opção de resposta que havia sido selecionada no item anterior (Weathers; Sharma; Niedrich, 2005). Este tipo de decisão está ligado ao fato de que diante de um processo de decisão complexo, os indivíduos tendem a utilizar a opção de *status quo* (Tversky; Shafir, 1992). Assim, o aumento no número de itens em uma escala de pontos tornaria a decisão

mais complexa, aumentando a possibilidade de os respondentes decidirem com base no *status quo*. Este efeito é minimizado quando os respondentes possuem ampla capacidade cognitiva ou de processamento, estando, assim, aptos a analisar um número maior de alternativas (Weathers; Sharma; Niedrich, 2005).

Nesta linha, um aspecto importante que deve ser levado em consideração na construção das escalas diz respeito ao limite da habilidade humana em fazer distinções. Miller (1956) detectou que o uso de aproximadamente sete categorias de resposta é o limite para a habilidade humana distinguir. Este seria o limite de categorias a que as pessoas estariam habilitadas para fazer julgamentos. Por outro lado, o autor coloca que escalas com poucas categorias de respostas podem não permitir discriminar suficientemente a opinião dos respondentes. Tourangeau e Rasinski (1988) também argumentam que deve ser levada em conta a sensibilidade de mensuração do respondente da escala na definição desta. Os autores colocam que, em uma medida na qual o respondente deve responder de 1 a 100, o resultado provavelmente será dado em número múltiplo de 10 ou de cinco. Desta forma, uma escala ampla terá uma precisão desnecessária.

Estas constatações conflitantes refletem a dificuldade em elaborar uma escala confiável. Porém, o uso de uma escala inadequada pode impactar na confiabilidade, validade e sensibilidade, propriedades básicas de uma escala tipo Likert (Cummins; Gullone, 2000). Masters (1974) detectou que o aumento do número de categorias utilizadas no questionário aumenta a consistência interna do instrumento, e um questionário com pequeno número de categoriais resulta em uma baixa variabilidade e confiabilidade, as quais aumentam à medida que se amplia o número de categorias de respostas. Churchill e Peter (1984) conduziram um estudo de meta-análise e identificaram que a confiabilidade de uma escala aumenta com o aumento do número de categorias de respostas.

Abordagens métricas também foram conduzidas para visualizar o número ótimo de categorias de respostas que afetam a confiabilidade das escalas com diferentes números de categorias. Estudo baseado no método de simulação Monte-Carlo realizado por Cicchetti, Showalter e Tyrer (1985) demonstrou evidências para a utilização de escalas entre dois e sete pontos, visto que de sete até 100 categorias de respostas não ocorrem incremento na confiabilidade do instrumento. Os pesquisadores concluíram que as diferenças das escalas entre sete e 100 pontos são insignificantes, e que uma escala ordinal de sete pontos apresenta a mesma funcionalidade que escalas com número maior de

itens. Resultados similares também foram encontrados por Oaster (1989), que, testando a confiabilidade por meio do teste-reteste e da consistência interna do instrumento, concluiu que a confiabilidade máxima é obtida com escalas de sete pontos.

Estes estudos realizados na década de 1980 contrariam estudos anteriores que defendiam o uso da escala de cinco pontos, proposta originalmente por Likert. Jenkins e Taber (1977) e Lissitz e Green (1975) concluíram, por meio de simulações, que escalas de cinco pontos são suficientes, visto que não foi observado um ganho de confiabilidade em escalas com mais que cinco itens. Em testes similares, porém, com escalas compostas por uma quantidade de itens múltiplos de dois, Green e Rao (1970) concluíram que o maior ganho com a escala gira em torno de seis categorias de respostas.

As contradições apresentadas pela literatura ficam ainda mais latentes com trabalhos como Rodriguez (2005), que, por meio de uma meta-análise, concluiu que uma escala com três opções de respostas é suficiente. O autor destaca que o efeito da diminuição do número de opções de escolha encolhe o teste, proporcionalmente aumenta a eficiência do teste para grandes quantidades de respondentes e diminui a eficiência para pequenas quantidades de respondentes. Em complemento, Rodriguez (2005) coloca que o tempo gasto na resposta do questionário é proporcional ao número total de alternativas, e o uso de três itens na escala diminui o tempo na coleta da informação. Apesar dos ganhos destacados por Rodriguez (2005), o uso de escalas com poucos itens tende a flutuar de amostra para amostra.

A escolha do número de itens na escala impacta também nos testes estatísticos realizados nas análises dos dados (Wiswanathan; Sudman; Johnson, 2004). O uso de uma escala com muitos pontos pode não prover uma base de dados válida para a realização de inferências estatísticas, visto que, de acordo com o tamanho da amostra, pode resultar em uma dispersão entre os respondentes, e curvas fora dos padrões normais, limitando o uso de alguns testes estatísticos. Contudo, uma escala que mensura um número máximo de discriminação deve ser mais fiável que uma escala com poucas categorias, e deve ser preferida para análises de correlação e regressão (Wiswanathan; Sudman; Johnson, 2004). Coelho e Esteves (2007) argumentam que escalas com poucos itens podem não fornecer uma boa discriminação das respostas (limitando a habilidade de encontrar diferenças significantes entre segmentos), o que limita o método de análise dos dados. Em adição, o uso de um número maior de pontos aumenta a base de dados, enriquecendo a análise dos dados e o cálculo da covariância entre

variáveis, usada em grande parte dos testes multivariados de dados.

Em uma linha de pesquisa diferente das pesquisas que buscam encontrar uma quantidade de categorias de respostas ideal, estudos como Barnette (2000) testaram a mudança na ordem da questão, utilizando questões de ordem positiva e negativa no mesmo instrumento. O autor detectou que estes procedimentos confundem os respondentes e, desta forma, são desaconselhados. Ao inverter a ordem da questão, a percepção dos itens pode não ser exatamente o oposto, o que resulta na redução da confiabilidade e da validade dos resultados. Contudo, nos testes realizados por Barnette (2000), a ordem da questão não resultou em perda de consistência interna e possibilitou detectar a consistência e o viés das respostas.

O uso da bidirecionalidade nas respostas tem sido apontado como uma alternativa, principalmente para evitar a inversão da ordem de questões (Robinson; Shaver; Wrightsman, 1991). A bidirecionalidade dos itens geralmente é disposta no sentido de direção oposta (concordo e não concordo). Uma alternativa pode ser a adoção de apenas uma mesma direção – todos positivamente orientados, sem a utilização de qualquer variação do “não” (Barnette, 2000). A direcionalidade possui um papel importante na elaboração da escala, visto que, de acordo com Cummins e Gullone (2000), somente 10% do composto de interpretação da escala pode ser atribuído à intensidade, e o restante está atribuído à direção.

Para fornecer o caráter de direcionalidade, as âncoras verbais que descrevem cada um dos pontos também merecem atenção. Apesar da recomendação de utilização de palavras indicando a intensidade da atitude num contínuo (Devellis, 1991), a utilização de descrições verbais em cada questão e categoria de resposta dificulta a resposta. Para isso, o uso de números para ancorar cada ponto de resposta é uma opção recomendada, visto que fornece uma percepção de contínuo, que não só contribui para ajudar o respondente a entender o que é requerido no item, mas também contribui para uma qualificação igualitária da escala. O uso de âncoras numéricas também facilita a tabulação e análise dos dados (Nunnally, 1978). Nesta questão, Churchill e Peter (1984) não encontraram diferenças no uso de números ou palavras na ancoragem das respostas. De maneira geral, Cummins e Gullone (2000) recomendam o uso do ponto final na escala baseado em palavras, ou seja, utilizam-se âncoras verbais nas extremidades da escala, visto que os respondentes tendem a analisar somente as extremidades da escala e não cada ponto isolado. Porém, como guia adicional para o significado de

cada ponto na escala, podem-se utilizar âncoras (números ou espaços pré-estabelecidos) indicando a intensidade do item, de maneira que o respondente visualize uma progressão contínua e direcional.

Cummins e Gullone (2000), além de levantarem a questão de ancorar a escala com palavras nas extremidades como “extremamente satisfeito” e “extremamente insatisfeito”, destacam a utilização da categoria central (ponto neutro), do tipo “nem satisfeito, nem insatisfeito”. Esta opção pode fazer os respondentes sentirem-se mais confortáveis em responder, mas, ao mesmo tempo, o ponto neutro pode gerar ambivalência e indiferença do respondente, destoando da verdadeira opinião do respondente (Collings, 2006). No entanto, Coelho e Esteves (2007) colocam que é possível que o respondente não tenha uma opinião ou experiência no tocante aos atributos específicos, e que a resposta neutra seria a mais indicada. Neste caso, a ambiguidade causada pelo ponto neutro pode ser sanada com a utilização de uma opção do tipo “sem condições de opinar”, não destoando à verdadeira opinião do respondente.

Por fim, como Halpin, Halpin e Arbet (1994) sugerem, a melhor opção de escolha para o número de itens na escala depende do conteúdo que a escala pretende mensurar. Coelho e Esteves (2007) ressaltam que respondentes com mais habilidade e experiência na resposta de escalas permitem o uso de escalas com maior número de pontos, visto que o uso rotineiro de escalas resultará em mais habilidade em identificar importantes relações e, por consequência, resultará em uma alta validade dos construtos e melhores testes das hipóteses tanto na teoria quanto na prática. Na busca de uma síntese dos estudos apresentados neste referencial teórico, a Tabela 01 resume os formatos de escala, ressaltando suas vantagens e desvantagens.

Tabela 01: Vantagens e desvantagens dos diferentes formatos de escala

Formato da Escala	Vantagens	Desvantagens
<b>10 pontos ou mais</b>	<ul style="list-style-type: none"> <li>- Facilidade de compreensão do sistema de numeração de 0 a 10;</li> <li>- Ganho de informação transmitida;</li> </ul>	<ul style="list-style-type: none"> <li>- Escolha e discriminação complexa por parte do respondente;</li> <li>- Estimula o uso da heurística habitual;</li> <li>- Gera uma prescrição desnecessária, pois não proporciona ganhos de confiabilidade;</li> <li>- Modelo de escala adotado inicialmente, mas em desuso;</li> </ul>

DILEMAS NA CONSTRUÇÃO DE ESCALAS TIPO LIKERT: O NÚMERO DE ITENS E A DISPOSIÇÃO INFLUENCIAM NOS RESULTADOS?

<b>7 pontos</b>	<ul style="list-style-type: none"> <li>- Limite da habilidade humana de distinção;</li> <li>- Permite melhor discriminação;</li> <li>- Ganho de consistência interna e confiabilidade;</li> <li>- Boa discriminação da covariância;</li> <li>- Se ajusta bem a estatísticas multivariadas;</li> </ul>	<ul style="list-style-type: none"> <li>- Exige uma grande quantidade de casos (respondentes) para inferências;</li> </ul>
<b>5 pontos</b>	<ul style="list-style-type: none"> <li>- Ponto neutro (decorrente das escalas ímpares);</li> <li>- Nível de confiabilidade adequado;</li> <li>- Se ajusta aos respondentes com diferentes níveis de habilidade;</li> </ul>	<ul style="list-style-type: none"> <li>- Não oferece a discriminação da escala de sete pontos;</li> <li>- Mais longa que a escala de três pontos;</li> </ul>
<b>3 pontos</b>	<ul style="list-style-type: none"> <li>- Opções de respostas suficientes;</li> <li>- Se ajusta a pequenas amostras;</li> <li>- Demanda pouco tempo de resposta;</li> </ul>	<ul style="list-style-type: none"> <li>- Baixa variabilidade e confiabilidade;</li> <li>- Maior flutuação entre diferentes amostras;</li> <li>- Pouca discriminação;</li> </ul>
<b>Disposição de ordem inversa</b>	<ul style="list-style-type: none"> <li>- Detecta a consistência interna;</li> <li>- Verifica vieses nas respostas.</li> </ul>	<ul style="list-style-type: none"> <li>- Confunde os respondentes;</li> <li>- Percepção da escala invertida pode não ser exatamente o oposto;</li> <li>- Reduz confiabilidade e validade.</li> </ul>

Fonte: Elaborado pelos autores com base no referencial teórico

A partir da Tabela 01, é possível observar que os estudos divergem em termos da quantidade de itens e disposição da escala. Ressalta-se que a utilização de escalas longas, com mais de 10 itens, é desencorajada (Cummins; Gullone, 2000), pois a sua complexidade desestimula o respondente e não fornece mais informações que escalas de até sete pontos. Desta forma, serão alvo de investigação empírica neste estudo as escalas de sete, cinco e três pontos, bem como a disposição de ordem inversa da escala.

### 3. MÉTODO DO ESTUDO

Ao nível epistemológico, este estudo tem um cunho quantitativo, com corte transversal, por meio de uma *survey*. Esta opção metodológica demonstrou ser a mais adequada, pois a operacionalização do estudo demandava obter respostas de um grande número de pessoas (Hair Jr. et al, 2005). Para a operacionalização do estudo, buscou-se a utilização de questionário com uma base de questões comum, com diferentes escalas para respostas, ou seja, um mesmo conjunto de questões variando apenas o formato da escala. Foram construídos

quatro modelos de questionários, todos com as mesmas questões, mas com escalas distintas. Os três primeiros questionários apresentavam respectivamente escala do tipo Likert de três, cinco e sete pontos. O quarto questionário apresentava a escala tipo Likert de cinco pontos, mas com a disposição invertida. Foram utilizadas âncoras numéricas para cada opção de resposta, complementadas com âncoras verbais nas extremidades, sendo: discordo totalmente e concordo totalmente. As Figuras 03, 04, 05 e 06 representam as escalas inseridas nos instrumentos.

Figura 03: Escala tipo Likert com cinco pontos invertida

Fonte: Elaborado pelos autores

<b>Discordo Totalmente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>Concordo Totalmente</b>
----------------------------	----------	----------	----------	----------------------------

Figura 04: Escala tipo Likert com três pontos

Fonte: Elaborado pelos autores

<b>Discordo Totalmente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Concordo Totalmente</b>
----------------------------	----------	----------	----------	----------	----------	----------------------------

Figura 05: Escala tipo Likert com cinco pontos

Fonte: Elaborado pelos autores

<b>Discordo Totalmente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>Concordo Totalmente</b>
----------------------------	----------	----------	----------	----------	----------	----------	----------	----------------------------

Figura 06: Escala tipo Likert com sete pontos

Fonte: Elaborado pelos autores

<b>Concordo Totalmente</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>Discordo Totalmente</b>
----------------------------	----------	----------	----------	----------	----------	----------------------------

Cada questionário era composto por vinte e uma questões em escala tipo Likert que versavam sobre o tema “dinheiro”. A escolha do tema justificava-se pela necessidade de este ser de domínio dos entrevistados, para evitar vieses referentes à profundidade de conhecimento do tema. Conforme Coelho e Esteves (2007), o grau de conhecimento do tema objeto de estudo pode interferir na capacidade dos entrevistados discernirem entre escalas com diferentes números de itens. De forma complementar, todos os questionários possuíam três questões referentes à facilidade de uso, à velocidade de uso e à precisão do instrumento. Essas questões visavam mensurar a opinião do respondente sobre cada uma das escalas. Para isso, havia um campo de resposta no qual os respondentes poderiam atribuir um nota que variava de zero a dez, sendo um (muito ruim) a dez (muito bom).

Para a aplicação dos questionários, adotou-se o seguinte procedimento: inicialmente foi distribuída uma etiqueta numerada a cada entrevistado e foi explicado que para cada questionário recebido

existia um campo chamado “código de identificação”, no qual deveria ser fornecido o número constante na etiqueta. Tal procedimento visou garantir que posteriormente fosse possível identificar quais questionários pertenciam a um mesmo entrevistado. Em seguida, cada entrevistado preencheu os quatro questionários subsequentemente, ou seja, após o preenchimento do primeiro questionário, este era recolhido e era entregue o segundo questionário e assim sucessivamente até que o entrevistado completasse a série de quatro questionários. A opção pelo recolhimento teve como objetivo evitar que o entrevistado visualizasse a resposta dada no questionário anterior no momento de decidir sobre a nova resposta.

A amostra do estudo foi do tipo não-probabilística, selecionada por conveniência, formada por estudantes do curso de Administração da Universidade Federal de Santa Maria. Este tipo de amostra é caracterizada pela facilidade de acesso do pesquisador aos entrevistados (Malhotra, 2006) e se demonstrou adequada ao estudo, pois não se pretendia mensurar opiniões e comportamentos de um grupo específico, mas sim testar os diferentes tipos de questionários. Foram entrevistados todos os alunos presentes em salas de aula de diferentes disciplinas no mesmo dia, perfazendo um total de 211 entrevistados. Para a análise dos resultados, foram utilizados o Alpha de Cronbach, o coeficiente de correlação de Pearson, estatísticas descritivas e testes de diferença de média, por meio da utilização do software estatístico *Statistical Package for the Social Sciences* – SPSS 16.0.

Inicialmente, buscou-se avaliar a diferença de confiabilidade dos quatro instrumentos. A confiabilidade indica o grau de consistência interna entre os múltiplos indicadores de um construto, referindo-se à extensão na qual um mesmo instrumento de medida produz resultados coerentes a partir de diversas mensurações. Para mensurar a confiabilidade, utilizou-se o Alfa de Cronbach que, de acordo com Kline (2000), indica o coeficiente de confiabilidade interna de um instrumento de coleta de dados. Para Hair Jr. et al (2002), o valor aceitável do alfa deve ser superior a 0,7.

Em seguida, utilizou-se o Coeficiente de Correlação de Pearson, para avaliar a relação entre as respostas nos diversos instrumentos. O Coeficiente de Correlação de Pearson indica a força de associação entre quaisquer duas variáveis (Hair Jr. et al., 2005). No caso deste estudo, se as respostas dos entrevistados não são influenciadas pelo número de itens da escala, espera-se que a correlação entre as respostas para a mesma questão em dois instrumentos com escalas diferentes tenha correlação perfeita.

A diferença entre os instrumentos também foi avaliada a partir do teste t emparelhado. O teste t é apropriado para comparar dois conjuntos de dados quantitativos, em seus valores médios. Neste caso, utilizou-se o teste t emparelhado, porque se deseja comparar as respostas de um mesmo indivíduo em dois instrumentos distintos (Hair Jr. et al, 2002). A hipótese nula do teste afirma que a resposta média dos entrevistados não se altera com a mudança no número de itens da escala. Para a avaliação das diferenças de médias, faz-se necessária a padronização das escalas para que possam ser comparadas dentro de uma mesma ordem de valores. Assim, optou-se por transformar as escalas dos instrumentos de cinco e sete pontos, para que todos os instrumentos apresentassem mínimo igual a um e máximo igual a três. As Tabelas 02 e 03 apresentam os valores utilizados para a transformação das escalas de cinco e sete pontos.

Tabela 02: Valores utilizados para a padronização das escalas para o instrumento com a escala de cinco pontos e para o instrumento com a escala de cinco pontos invertida

Escala Original Padronizada	Valores				
	1	2	3	4	5
	1	1,5	2	2,5	3

Fonte: Elaborado pelos autores

Tabela 03: Valores utilizados para a padronização das escalas para o instrumento com a escala de sete pontos

Escala Original Padronizada	Valores						
	1	2	3	4	5	6	7
	1,000	1,333	1,666	2,000	2,333	2,666	3,000

Fonte: Elaborado pelos autores

Os resultados das variáveis que mensuravam a opinião dos entrevistados acerca de cada escala foram avaliados inicialmente com o cálculo da média e desvio-padrão. Isto permitiu conhecer qual a tendência central dos respondentes quanto à avaliação da facilidade de uso, velocidade e possibilidade de expressar opinião em cada escala. Complementarmente, foi aplicado teste t e de significância de médias, visando comparar a avaliação dos respondentes entre os diferentes formatos de escala testados.

#### 4. ANÁLISE DOS RESULTADOS

Inicialmente, cada uma das quatro escalas foi avaliada quanto à confiabilidade. Para tanto se utilizou o Alpha de Cronbach, que avalia a extensão

DILEMAS NA CONSTRUÇÃO DE ESCALAS TIPO LIKERT: O NÚMERO DE ITENS  
E A DISPOSIÇÃO INFLUENCIAM NOS RESULTADOS?

em que os itens formam uma escala internamente consistente. A Tabela 04 apresenta os resultados do Alpha de Cronbach.

Tabela 04: Valores do Alpha de Cronbach para cada um dos instrumentos

Escala	Alpha de Cronbach
Três Pontos	0,66
Cinco Pontos	0,77
Sete Pontos	0,80
Cinco Pontos Invertida	0,78

Fonte: Elaborado pelos autores

Hair Jr. et al (2005) recomendam que para que a consistência interna seja considerada satisfatória, os valores devem ser iguais ou superiores a 0,7. Os resultados da Tabela 03 mostram que o instrumento com a escala de três pontos não atingiu o valor sugerido. Já os instrumentos de cinco e sete pontos atingiram os limites aceitáveis sendo que, do ponto de vista da consistência interna, o instrumento de sete pontos mostrou-se mais confiável do que o de cinco pontos. A inversão da escala de cinco pontos não promoveu alterações significativas na confiabilidade do instrumento.

A validade convergente mede a extensão em que a escala se correlaciona positivamente com outras medidas do mesmo construto (Malhotra, 2006). Neste caso, a extensão em que cada escala tipo Likert mede o mesmo construto pode ser avaliada a partir da correlação de cada item nas diferentes escalas. Assim, quanto maior a correlação entre duas escalas diferentes, mais as duas medidas convergem para o mesmo resultado.

Tabela 05: Coeficientes de Correlação de Pearson para a comparação entre os instrumentos dois a dois

Variável	3 versus 5 Correlação	3 versus 7 Correlação	5 versus 7 Correlação	5 versus 5 Invertida Correlação
1	0,813	0,766	0,857	0,856
2	0,819	0,758	0,887	0,777
3	0,703	0,700	0,781	0,696
4	0,647	0,536	0,844	0,792
5	0,495	0,478	0,757	0,698
6	0,608	0,548	0,802	0,800
7	0,619	0,584	0,793	0,790
8	0,742	0,715	0,873	0,858
9	0,632	0,666	0,830	0,759
10	0,673	0,608	0,827	0,764
11	0,791	0,783	0,878	0,858
12	0,678	0,579	0,824	0,801
13	0,627	0,621	0,776	0,749

14	0,725	0,700	0,821	0,754
15	0,482	0,530	0,796	0,739
16	0,509	0,409	0,565	0,566
17	0,632	0,631	0,761	0,792
18	0,722	0,600	0,809	0,762
19	0,543	0,548	0,762	0,767
20	0,788	0,736	0,865	0,848
21	0,573	0,587	0,803	0,755

Fonte: Elaborado pelos autores

Todas as correlações apresentadas na Tabela 05 são significativas ao nível de 1%. Observa-se que, de maneira geral, as variáveis apresentam correlações altas, o que sugere que os instrumentos apresentam validade convergente dois a dois. Os maiores coeficientes são observados na comparação dos instrumentos com as escalas de cinco e sete pontos. Por outro lado, as correlações mais baixas são obtidas na comparação entre os instrumentos com escalas de três e sete pontos. Nesta ótica, pode-se afirmar que, ao aumentar a diferença entre as escalas (três para sete) a convergência dos resultados diminui quando comparada a escalas mais próximas (três e cinco pontos). Por outro lado, se a inversão da escala não implicasse nenhuma alteração nas respostas dos entrevistados, os coeficientes de correlação entre a escala de cinco pontos e a escala de cinco pontos invertida deveriam ser iguais ao valor '1'. No entanto, a coluna das correlações entre a escala de cinco pontos e a escala de cinco pontos invertida apresentam coeficientes de correlações abaixo de um e, em geral, os mesmos ficaram abaixo dos coeficientes apresentados na comparação entre as escalas de cinco e sete pontos.

Foi aplicado um teste t emparelhado para diferença de média entre as correlações obtidas para a comparação cinco versus sete e as correlações obtidas para cinco versus cinco invertida. O valor de t calculado foi de 4,55, valor este significativo ao nível de 1%. Tal resultado indica que há menor convergência na inversão de escala do que no aumento do número de itens da escala de cinco para sete. Portanto, na inversão de escala, os entrevistados mudaram mais de posição do que na ampliação do número de itens de cinco para sete.

Para avaliar mais detalhadamente a possibilidade de mudança de posição por parte dos entrevistados, optou-se pela realização de um teste t emparelhado para diferença de médias entre cada par de escalas. A Tabela 06 apresenta os valores do teste t e o valor da significância para cada variável.



Tabela 06: Valores do teste t e significância para cada variável, na comparação dos instrumentos dois a dois

Variável	3 pontos versus 5 pontos		3 pontos versus 7 pontos		5 pontos versus 7 pontos	
	Valor t	Significância	Valor t	Significância	Valor t	Significância
1	-0,345	0,731	-0,615	0,539	-0,570	0,569
2	1,206	0,229	0,518	0,605	-1,132	0,259
3	1,589	0,114	3,691	0,000	2,924	0,004
4	1,333	0,184	1,489	0,138	0,672	0,503
5	5,242	0,000	6,177	0,000	1,252	0,212
6	2,890	0,004	2,354	0,019	-0,718	0,474
7	1,282	0,201	1,211	0,227	0,147	0,883
8	-1,964	0,051	-3,039	0,003	-1,738	0,084
9	0,355	0,723	1,303	0,194	1,505	0,134
10	2,253	0,025	2,979	0,003	1,486	0,139
11	-0,165	0,869	0,164	0,870	0,367	0,714
12	1,727	0,086	1,276	0,203	-0,620	0,536
13	0,947	0,345	2,416	0,017	1,961	0,051
14	1,769	0,078	0,851	0,396	-1,392	0,165
15	3,173	0,002	4,035	0,000	1,243	0,215
16	-1,053	0,294	-1,641	0,102	-1,104	0,271
17	-1,597	0,112	-2,513	0,013	-1,378	0,170
18	-0,488	0,626	-2,074	0,039	-2,056	0,041
19	0,146	0,884	0,098	0,922	-0,300	0,765
20	0,905	0,367	-0,214	0,831	-1,312	0,191
21	-2,574	0,011	-3,594	0,000	-1,710	0,089

Fonte: Elaborado pelos autores

Para a maioria das variáveis, em média, não há diferença entre as respostas dos dois instrumentos. Os maiores efeitos são observados na comparação entre a escala de três pontos e a escala de sete pontos, em que, das 21 questões, 10 apresentaram diferenças médias significativas pelo menos ao nível de 5%. Complementando a análise das correlações que já havia indicado que os valores mais baixos são apresentados na comparação entre as escalas de três e sete pontos, o teste t indica em quais das variáveis a mudança de posição foi mais proeminente.

Para avaliar o efeito da inversão do formato da escala de cinco pontos foram realizadas duas análises. Na primeira (Tabela 07) calculou-se para cada variável o percentual de entrevistados que marcaram resultados diferentes nos dois instrumentos, ou seja, mudaram de posição. Na segunda (Tabela 08), realizou-se um teste t para verificar se, em média, a inversão da escala alteraria os resultados.

Caso os entrevistados não sofressem nenhuma influência da formatação da escala, as respostas aos dois instrumentos seriam idênticas. No entanto, como observou-se na Tabela 07, em torno de 33% dos entrevistados marcaram respostas diferentes entre o instrumento com a escala de cinco pontos

(discordo totalmente para concordo totalmente) e aquele com a escala de cinco pontos cujo formato foi invertido (concordo totalmente para discordo totalmente).

Para avaliar se tais mudanças afetam os resultados médios, construiu-se a Tabela 08. Apenas para quatro variáveis, a média da escala de cinco pontos difere significativamente da média da escala de cinco pontos invertida. Analisando as duas tabelas, percebe-se que, apesar de aproximadamente um terço dos entrevistados mudarem suas posições, a inversão da escala não implicou em mudança significativa dos resultados médios para a maioria das questões.

Tabela 07: Percentual de entrevistados que mudaram de posição e que mantiveram sua posição na comparação das respostas entre os instrumentos com escalas de cinco pontos e cinco pontos invertida

Variável	Percentual de Entrevistados	
	Mantiveram a mesma posição	Mudaram de posição
1	66,20	33,80
2	64,60	35,40
3	58,50	41,50
4	67,10	32,90
5	65,70	34,30
6	71,80	28,20
7	60,30	39,70
8	59,30	40,70
9	65,60	34,40
10	68,30	31,70
11	77,60	22,40
12	77,50	29,50
13	61,70	38,30
14	67,10	32,90
15	70,30	29,70
16	68,30	31,70
17	77,00	23,00
18	58,00	42,00
19	78,30	21,70
20	77,50	22,50
21	57,60	42,40

Fonte: Elaborado pelos autores

DILEMAS NA CONSTRUÇÃO DE ESCALAS TIPO LIKERT: O NÚMERO DE ITENS E A DISPOSIÇÃO INFLUENCIAM NOS RESULTADOS?

Tabela 08: Valores do teste t e significância para cada variável, na comparação dos instrumentos com escala de cinco pontos e escala de cinco pontos invertida

Variável	5 Pontos versus 5 Pontos Invertida	
	Valor t	Significância
1	0,497	0,620
2	-0,907	0,366
3	2,356	0,019
4	1,078	0,282
5	1,229	0,220
6	1,079	0,282
7	-0,350	0,727
8	-2,673	0,008
9	-0,561	0,575
10	1,429	0,155
11	-0,562	0,575
12	-0,717	0,474
13	0,890	0,374
14	-0,872	0,384
15	-0,105	0,917
16	-0,961	0,338
17	-1,079	0,282
18	-2,580	0,011
19	-1,846	0,066
20	-0,584	0,560
21	-2,803	0,006

Fonte: Elaborado pelos autores

Numa segunda etapa da análise dos resultados, buscou-se avaliar a opinião dos entrevistados quanto a três aspectos da escala: (1) Facilidade de Uso; (2) Velocidade de Uso e (3) Possibilidade de Expressar sua Opinião com Precisão. Para cada um dos critérios, os entrevistados poderiam atribuir valores de um (muito ruim) a dez (muito bom).

Tabela 09: Média e desvio padrão das variáveis Facilidade de Uso, Velocidade de Uso e Possibilidade de Expressar sua Opinião com Precisão, nas três escalas.

Variável	3 Pontos		5 Pontos		7 Pontos	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
Facilidade	8,68	1,33	8,24	1,48	7,46	1,84
Velocidade	8,73	1,39	7,91	1,67	7,06	2,04
Precisão	6,89	1,85	7,81	1,44	7,83	1,78

Fonte: Elaborado pelos autores

Observa-se que existe uma relação inversa entre o tamanho da escala e a Facilidade de Uso. Apesar de os três instrumentos apresentarem médias altas, o instrumento com a menor escala apresentou a

maior média (8,68) e aquele com a maior escala, a menor média (7,46). O mesmo comportamento pode ser observado para a variável Velocidade de Uso, comprovando que o aumento do número de itens na escala diminui a Velocidade de Uso. Tal resultado está em linha com os argumentos de que, ao aumentar o número de itens, o entrevistador está multiplicando o número de respostas possíveis, o que implica aumentar a complexidade do processo de tomada de decisão e, consequentemente, aumentar o tempo de resposta.

Por outro lado, a escala com um menor número de itens apresenta, em média, menor capacidade de expressar a opinião do entrevistado. Observa-se ainda que o aumento da precisão ao se passar de uma escala de três pontos para uma escala de cinco pontos é de quase um ponto. Já o aumento médio da precisão ao se passar da escala de cinco pontos para a de sete pontos é de apenas 0,2. Tais resultados sugerem que o ganho de precisão não é diretamente proporcional ao aumento no número de itens, pois um aumento de dois itens numa escala com pequeno número de itens (três pontos para cinco pontos) é muito maior do que o mesmo aumento em escala com mais itens (cinco pontos para sete pontos). Para avaliar se as alterações nestas variáveis são significativas, optou-se pela aplicação de um teste t emparelhado para diferença de médias.

Tabela 10: Valor do teste t e significância para as variáveis Facilidade de Uso, Velocidade de Uso e Possibilidade de Expressar sua Opinião com Precisão

Variável	3 Pontos versus 5 Pontos		3 Pontos versus 7 Pontos		5 Pontos versus 7 Pontos	
	Valor t	Significância	Valor t	Significância	Valor t	Significância
Facilidade	5,002	0,000	10,228	0,000	8,593	0,000
Velocidade	8,607	0,000	11,575	0,000	7,759	0,000
Precisão	-7,762	0,000	-5,518	0,000	-0,279	0,780

Fonte: Elaborado pelos autores

Todos os testes de diferença de média apresentaram significância ao nível de 1%, exceto para a diferença de precisão entre as escalas de cinco pontos e sete pontos. Tais resultados confirmam o comportamento levantado na Tabela 08, na qual um aumento na escala passando de cinco para sete pontos implica uma diminuição significativa da Facilidade e da Velocidade de Uso, mas não gera um aumento de Precisão. Do ponto de vista da Precisão, estes resultados indicam para o pesquisador uma preferência pela adoção da escala de cinco pontos, pois esta se mostrou ser bem mais precisa do que a de três pontos, além de não se diferenciar da de sete pontos. Tomando os três aspectos em conjunto, a escala de cinco pontos também deve

ser preferida em relação à de sete pontos, já que a mesma é tão precisa quanto à de sete pontos, mas apresenta mais Facilidade e Velocidade de Uso. Por outro lado, a escolha entre a escala de três e cinco pontos não pode ser determinada pelos resultados encontrados, pois, apesar da maior precisão, a escala de cinco pontos apresenta menor Facilidade e Velocidade do que a de três pontos. Neste caso, se o pesquisador considerar a precisão como fator decisivo, optará pela escala de cinco pontos em detrimento da de três. Já se a facilidade ou a velocidade forem os fatores preponderantes, a escala de três pontos pode ser escolhida.

## 5. CONSIDERAÇÕES FINAIS

São muito comuns em estudos no campo da administração pesquisas que apresentam no método a expressão *utilizou-se uma escala tipo Likert de "x" pontos*, mas é raro o aparecimento de justificativas para a escolha do número "x". Tal fenômeno pode sugerir inicialmente que a escolha de uma escala de medida não seja um fator decisivo para o resultado da pesquisa, gerando uma falta de atenção para este item (Cummins; Gullone, 2000; Collings, 2006). No entanto, quando o pesquisador se abdicar de usar o senso comum e passa a se questionar sobre os aspectos da construção de uma escala, descobre que o desenvolvimento de uma escala de medida para um instrumento de pesquisa é uma tarefa complexa.

A complexidade decorre principalmente dos múltiplos aspectos envolvidos na construção da escala. Além da questão-chave de quantos itens deverá ter uma escala, é necessário tomar uma série de decisões: (1) quanto ao uso de números ímpares ou pares, o que implica discutir a importância do ponto neutro (Coelho; Esteves, 2007); (2) quanto ao uso de âncoras – se elas serão verbais ou numéricas (Churchill; Peter, 1984), se serão unidirecionais – formando um contínuo – ou bidirecionais (Devellis, 1991), se aparecerão apenas nos extremos (Cummins; Gullone, 2000). E estas decisões dependerão de aspectos específicos do objeto de estudo como, por exemplo, (1) a complexidade do tema; (2) e o número de questões/variáveis a serem inseridas no instrumento (Campbell, 1988; Weathers; Sharma; Niedrich, 2005). A combinação destas opções cria uma gama enorme de possibilidades, sobre as quais o pesquisador é obrigado a fazer a sua opção, já que, geralmente, apenas uma escala é apresentada no instrumento. A decisão acerca da escala utilizada impacta diretamente em aspectos como confiabilidade, validade, sensibilidade do instrumento (Cummins; Gullone, 2000), na sua consistência interna (Masters, 1974).

Assim, desde que Likert (1932) desenvolveu sua escala de cinco pontos, diversos autores questionaram qual seria o número ideal de pontos para uma escala, utilizando-se para tanto desde modelos matemáticos – como exemplo o estudo de Cicchetti, Showalter e Tyrer, (1985) – até meta-análises – como exemplo o estudo de Churchill e Peter (1984). Independentemente do método utilizado, foram encontrados resultados diferentes, indicando desde o uso de três pontos até o uso de escala maiores, com dez ou mais pontos. Portanto, é possível encontrar na literatura artigos que defendam o uso de diferentes tamanhos de escala. Neste contexto, não há um argumento teórico único capaz de ser utilizado para defender o uso de "x" pontos. O que se tem até o momento são indicadores que sugerem os aspectos a serem levados em consideração na escolha do "x".

Dentre os diversos aspectos envolvidos na decisão, este trabalho se dedicou principalmente ao estudo do número de itens da escala, sendo testadas escalas com três, cinco e sete pontos. Os testes de diferença de média demonstraram que, para a maioria das questões, o resultado médio não é alterado pelo número de itens da escala. Os resultados revelaram, ainda, que a escala de três pontos é menos confiável e tem menor capacidade de demonstrar com precisão a opinião do entrevistado. A escala de cinco pontos teve, em média, a mesma precisão e mostrou-se mais fácil e mais veloz no uso que a escala de sete pontos. Portanto, para este estudo, a escala que se mostrou mais adequada foi a de cinco pontos.

Com relação à inversão do formato da escala, o estudo mostrou que alguns entrevistados mudaram de posição, apesar de o efeito médio não ser significativo. Como o nível de confiabilidade das duas escalas é muito semelhante, não podemos afirmar que uma seja mais consistente que a outra. No entanto, tais resultados sugerem que a criação de um único instrumento com escalas invertidas pode confundir os entrevistados.

Considerando as evidências apresentadas pela literatura sobre o tema, e os resultados deste estudo, propõe-se alguns pontos de reflexão relevantes na construção de escalas tipo Likert:

- Em termos de capacidade para expressar a opinião com precisão, a escala com três itens apresenta os piores resultados, devendo ser preterida em relação às escalas de cinco e sete pontos;
- As escalas de cinco e sete pontos são muito semelhantes em termos de resultados médios. A escolha pode depender de fatores subjetivos, como complexidade do tema e quantidade de questões;

- Escalas com mais itens geralmente são mais indicadas quando os entrevistados dominam o assunto objeto de estudo ou quando o objeto de estudo tem muitos atributos – reforçando a compreensão de Coelho e Esteves (2007);
- Deve-se evitar o uso de escala com formatos diferentes, principalmente com inversão do sentido num mesmo instrumento, visto que isto pode provocar a mudança de posição de alguns respondentes. Além disto, em consonância com Devellis (1991), as escalas também devem apresentar claramente um contínuo de direção;
- Devido a algumas evidências de que a complexidade da tomada de decisão pode interferir nos resultados, em questionários que envolvem um grande número de questões, deve-se optar por escalas menores para diminuir o número total de opções a serem analisadas. Por exemplo, num instrumento de 20 questões, se a escala tiver cinco pontos, serão 100 opções a serem analisadas. Já se a escala for de sete pontos o número de opções que o indivíduo terá que analisar sobe para 140. Assim, quanto maior a escala, maior a necessidade de processamento mental por parte dos respondentes. Cabe destacar ainda que alguns artigos (Swait; Adarnowicz, 2001; Weathers; Sharma; Niedrich, 2005) argumentam que quando o número de opções é muito grande, o indivíduo pode apresentar uma pré-disposição a manter a mesma resposta ao longo do instrumento (efeito status quo);
- A utilização do ponto neutro é defendida por ser uma opção que deixa o respondente mais à vontade no momento de expressar sua opinião. Caso a escala seja “par”, a literatura tem sugerido a inclusão da opção “sem condições de opinar” (Cummins; Gullone, 2000; Coelho; Esteves, 2007);
- O tamanho da amostra e os testes estatísticos que se pretende usar na análise dos resultados influenciam a definição do número de itens, visto que, em uma amostra pequena, um grande número de itens pode não dar uma base adequada para as análises estatísticas (Wiswanathan; Sudman; Johnson, 2004). Do mesmo modo, escalas com poucos itens podem limitar as análises, bem como gerar flutuações na normalidade dos dados;
- Uso de âncoras verbais clarificam as opções de escolha para os respondentes, no entanto, ancorar cada item pode acabar poluin-

do a escala. Desta forma, as âncoras verbais devem ser usadas somente nas extremidades, visto que os respondentes tendem a analisar somente as extremidades de uma escala (Cummins; Gullone, 2000). Já os itens específicos podem ser ancorados por números, gerando uma percepção de contínuo com distribuição igualitária entre os itens (Nunnally, 1978).

Fechar estudos como este se torna uma tarefa difícil, visto que a reflexão sobre os métodos utilizados em pesquisas é um passo essencial para o avanço da ciência como um todo, contribuindo para que as pesquisas que utilizam escalas tipo Likert tenham um poder de captação do real ainda mais significativo. Desta forma, estudos que permeiam a qualificação do processo de elaboração de instrumentos de coleta de dados, fornecendo subsídios aos pesquisadores para embasar suas escolhas, são uma demanda constante da academia. Novas abordagens, testando diferentes formatos de escalas, especialmente em termos de ancoragem dos itens no contexto brasileiro contribuem no processo de reflexão acerca dos dilemas na construção de escalas.

## REFERÊNCIAS

- BARNETTE, J.J. Effects of Stem and Likert response option reversals on survey internal consistency: if you feel the need, there is a better alternative to using those negatively worded stems. **Educational and Psychological Measurement**. v. 60, n. 3, p. 361-370, 2000.
- CAMPELL, D.J. Task complexity: a review and analysis. **Academic Management Review**. v. 13, n. 1, p. 40-52, 1988.
- CLASON, D.L.; DORMODY, T.J. Analyzing data measured by individual Likert-type items. **Journal of Agricultural Education**. v. 35, n. 4, p. 54-71, 1994.
- CICCHETTI, D.V.; SHOWALTER, D.; TYRER, P.J. The effect of number of rating scale categories on levels of interater reliability: a Monte Carlo investigation. **Applied Psychological Measurement**. v. 9, n.1, p. 31-36, 1985.
- CLARK, L.E.; WATSON, D. Constructing validity: basic issues in objective scale development. **Psychological Assessment**. v. 7, n. 3, p. 309-319, 1995.
- CHANG, L.A. A psychometric evaluation of 4-point and 6-point Likert-type scale in relation to reliability and validity. **Applied Psychological Measurement**. v. 18, n. 2, p. 05-15, 1994.

- CHURCHILL, G.A.; PETER, J.P. Research design effects on the reliability of rating scales: a meta-analysis, **Journal of Marketing Research**. v. 21, n. 4, p. 360-375, 1984.
- COELHO, P.S.; ESTEVES, S.P. **The choice between a 5-point and a 10-point scale in the framework of customer satisfaction measurement**. Lisboa: ISEGI - Instituto Superior de Estatística e Gestão de Informação - New University of Lisbon, 2007.
- COLLINGS, D.P. **Selecting a questionnaire response scale for student feedback surveys: a comparison of psychometric properties and student preferences among three alternative**. Perth, Australia: Murdoch University, 2006.
- COTE, J.A.; BUCKLEY, M.R. Measurement error and theory testing in consumer research: an illustration of the importance of construct validation. **Journal Consumer Research**. v. 14, n. 4, p. 579-582, 1988.
- CUMMINS, R.A.; GULLONE, E. Why we should not use 5-point Likert scales: the case for subjective quality of life measurement. In: International Conference on Quality of Life in Cities, 2., 2000, Singapore. **Proceedings...** Singapore, 2000.
- DEVELLIS, R.F. **Scale development: theory and applications**. Newbury Park: Sage, 1991.
- FREYD, M. The graphic rating scale. **Journal of Educational Psychology**. v. 14, n. 2, p. 83-102, 1923.
- GARNER, W.R.; HAKE, H.W. The amount of information in absolute judgments. **Psychological Review**. v. 58, n. 6, p. 446-459, 1951.
- GREEN, P. E.; RAO, V. R. Rating scales and information recovery: how many scales and response categories to use. **Journal of Marketing**. v. 34, n.3, p. 33-39, 1970.
- HALPIN, G.; HALPIN, G.; ARBET, S. Effects of number and type of response choices on internal consistency reliability. **Perceptual and motor skills**. v. 79, n. 2, p. 928-930, 1994.
- HAIR JR., J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W. C. **Multivariate data analysis**. 5 ed. Englewood Cliffs, NJ: Prentice Hall, 2002.
- HAIR JR., J.F.; BABIN, B.; MONEY, A.H.; SAMOUEL, P. **Fundamentos de métodos de pesquisa em administração**. Porto Alegre: Bookman, 2005.
- JENKINS, G.D.; TABER, T.D. A Monte Carlo study of factors affecting three indices of composite scale reliability. **Journal of Applied Psychology**. v.62, n. 4, p. 392-398, 1977.
- KLINE, P. **The handbook of psychological testing**. Routledge: London, 2000.
- KROSNICK, J.A.; BERENT, M.K. Comparisons of party identification and policy preferences: the impact of survey question format. **American Journal of Political Science**. v. 37, n. 3, p. 941-964, 1993.
- LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**. v. 22, n. 140, p. 44-53, 1932.
- LISSITZ, R.W.; GREEN, S.B. Effect of the number of scale points on reliability: a Monte Carlo approach. **Journal of Applied Psychology**. v. 60, n. 1, p. 10-13, 1975.
- MALHOTRA, N.K. **Pesquisa de marketing: uma orientação aplicada**. 4.ed. Porto Alegre: Bookman, 2006.
- MASTERS, J.R. The relationship between number of response categories and reliability of Likert-type questionnaires. **Journal of Educational Measurement**. v. 11, n. 1, p. 49-53, 1974.
- MATOS, C.; TREZ, G. A influência da ordem das questões nos resultados de pesquisas surveys. **Revista de Administração FACES**. v. 11, n. 1, p. 151-172, 2012.
- MILLER, G.A. The magical number seven, plus or minus two some limits on our capacity for processing information. **Psychological Review**. v. 101, n. 2, p. 343-352, 1956.
- NUNNALLY, J. C. **Psychometric theory**. New York: McGraw Hill, 1978.
- OASTER, T.R.F. Number of alternatives per choice point and stability of Likert-type scales. **Perceptual and Motor Skills**. v. 68, n. 2, p. 539-550, 1989.
- PRESTON, C.C.; COLEMAN, A.M. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. **Acta Psychologica**. v. 104, n. 1, p. 1-15, 2000.
- ROBINSON, J.P.; SHAVER, P.R.; WRIGHTSMAN, L.S. **Measures of personality and social psychological attitudes**. San Diego, CA: Academic Press, 1991.
- RODRIGUEZ, M.C. Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. **Educational Measurement: Issues and Practice**. v. 24, n. 2, p. 3-13, 2005.
- SWAIT, J.S.; ADAMOWICZ, W. The influence of task complexity on consumer choice: a latent class model of decision strategy. **Journal of Consumer Research**. v. 21, n. 1, p. 189-199, 2001.
- TOURANGEAU, R.; RASINSKI, K.A. Cognitive processes underlying context effects in attitude measurement. **Psychology Bulletin**. v. 103, n. 3, p. 299-314, 1988.
- TVERSKY, A.; SHAFIR, E. Choice under conflict: the dynamics of deferred decision, **Psychology Science**. v. 3, n. 6, p. 358-361, 1992.

DILEMAS NA CONSTRUÇÃO DE ESCALAS TIPO LIKERT: O NÚMERO DE ITENS  
E A DISPOSIÇÃO INFLUENCIAM NOS RESULTADOS?

WATSON, G.B. Happiness among adult students of education. **Journal of Educational Psychology**. v. 21, n. 2, p. 79-109, 1930.

WEATHERS, D.; SHARMA, S.; NIEDRICH, R.W. The impact of the number of scale points, dispositional factors, and the status quo heuristic on scale reliability and response accuracy. **Journal of Business Research**. v. 58, n. 11, p. 1516–1524, 2005.

WISWANATHAN, M.; SUDMAN, S.; JOHSON, M. Maximum versus meaningful discrimination in scale response: implications for validity of measurement of consumer perception about products. **Journal of Business Research**. v. 57, n. 2, p. 08–24, 2004.