

Dilemas na Construção de Escalas Tipo Likert: o Número de Itens e a Disposição Influenciam nos Resultados?

Autoria: Kelmara Mendes Vieira, Marlon Dalmoro

Resumo:

No presente estudo, inicialmente 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, velocidade e 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 menor 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.

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 para 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 os resultados das medidas sejam um reflexo da realidade (MATTAR, 1999; COTE e BUCKLEY, 1988).

Para estas medições, os pesquisadores por diversas vezes utilizam escalas de mensuração multi-item. A utilização de tais escalas requer cuidados essenciais para o alcance de resultados satisfatórios, confiáveis e que permitam conclusões apropriadas (KROSNICK e BERENT, 1993). No entanto, 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. Destacando que a validação de uma questão não pode se dá sem considerar as opções de respostas 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, rotulação dos pontos da escala, força das âncoras. O segundo centra-se no respondente e nas suas características, como exemplo 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 e NIEDRICH, 2005).

Desde que Rensis Likert introduziu seu método em 1932, diversos estudos, provindo de diversas áreas como psicologia, educação e marketing têm investigado as implicações do uso de diferentes formatos de escalas tipo Likert. Contudo, em uma revisão da literatura, Preston e Coleman (2000) colocam que o número de opções de respostas ainda não está resolvido, mesmo depois de décadas de pesquisa. Chang (1994) complementa dizendo que há

uma demanda por pesquisas que examinem como o número de alternativas de respostas, ou pontos na escala, podem produzir resultados conflitantes. Desta forma, uma das decisões-chaves 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 questionamento de que os diversos formatos de escalas tipo Likert utilizados, tem 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 e 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, e o entendimento destas diferentes percepções ainda não está compreendido pela literatura. Weathers, Sharma e Niedrich (2005); Wiswanathan, Sudman e Johnson (2004) e Collings (2006) colocam que o número de pontos na escala que deve ser usada pelos pesquisadores ainda não está claro. Esta contextualização remete as seguintes questões problema: Qual a influência do número de opções de resposta na escala tipo Likert nos resultados de um estudo?

Diante desta problemática, 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, pretende-se avaliar qual a escala preferida pelos respondentes quanto à facilidade, velocidade e 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 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, são apresentadas as considerações finais do estudo.

2. Referencial Teórico

A história da construção de escalas de mensuração esta ligada ao trabalho seminal de Rensis Likert publicado em 1932. No entanto, é interessante destacar que anteriormente a esta data já se utilizavam avaliações subjetivas com o uso de escalas, as quais eram mais sensíveis do que as utilizadas atualmente. Cummins e Gullone (2000) destacam o trabalho de Freyd datados de 1923, no qual já se discutida as formas de escala disponíveis, baseadas 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. Em 1941, Ferguson também defendeu a utilização deste tipo de escala visto a vantagem de ter uma percepção da igualdade psicométrica da distância entre os pontos escala. Para ele, esta é uma suposição essencial quando tal envergadura é usada em combinação com estatísticas paramétricas.

Freyd, em 1923, foi um dos pioneiros na utilização de escalas para coleta de dados, quando introduziu o *Graphic rating method*. A escala deveria ser utilizada em conjunto com entrevistas, e deveria seguir como procedimento de resposta marcar o ponto apropriado em uma linha horizontal pontilhada. Poucos anos depois, Watson (1930) publicou uma escala de mensuração similar, onde 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:

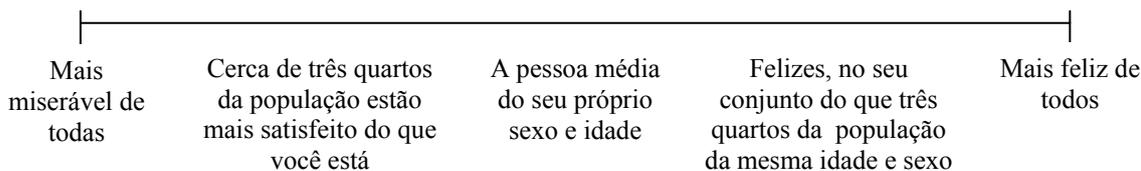


Figura 01: Modelo de escala desenvolvido por Watson (1930)

Fonte: Watson (1930)

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

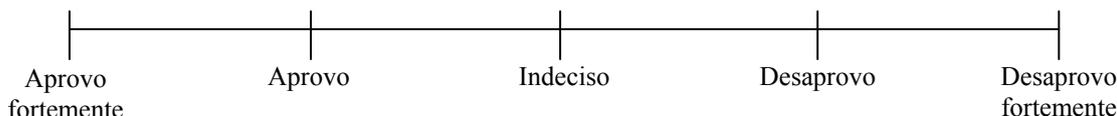


Figura 02: Modelo de escala desenvolvido por Likert (1932)

Fonte: Likert (1932)

Baseado nos modelos anteriores, Likert (1932) realizou uma redução no número efetivo de pontos de escolha, visto que inicialmente o sistema de medida era contínuo. Na escala de Likert, os respondentes precisavam marcar somente os pontos fixos estipulados na linha, em um sistema de cinco categorias de resposta (pontos) que vão de “aprovo totalmente” a “desaprovo totalmente”. Likert (1932) também introduziu a escala bidimensional 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 psicométrica utilizada na investigação, a dificuldade de generalizações com o uso de grande número de opções de marcação, e a natureza complexa de escalas alternativas (CUMMINS e GULLONE, 2000).

O trabalho de Likert (1932) deixa claro que a sua escala centrava-se na utilização de cinco pontos, e não mencionou 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, representem uma escala de classificação, quando esta não conter cinco opções de resposta, não se configura uma escala Likert. No entanto, Clason e Dormody (1994) afirmam que muitos estudos têm usado diversas classificações paralelas à classificação tradicional de cinco pontos de maneira satisfatória. Neste caso, a escala se configura como tipo Likert.

A variação no número de itens da escala surgida após a criação de Likert tem fomentado inúmeras discussões sobre a escolha da escala a ser utilizada. O problema da escolha de escalas está relacionado à 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 é um processo de quatro estágios em que 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. Este processo pode ser uma tarefa difícil, e os respondentes tendem a simplificar a tarefa com o uso de heurísticas. Os diferentes modos de uso da heurística podem refletir na opção de resposta selecionada pelo respondente.

O aumento da complexidade das opções de respostas tende a estimular o uso de decisões heurísticas (SWAIT e ADARNOWICZ, 2001). Um dos modelos heurísticos que os

respondentes tendem a utilizar é denominado de *status quo* heurística (ou heurística habitual) onde a regra de decisão utilizada pelos respondentes é a de selecionar a opção de resposta que havia sido selecionada no item anterior (WEATHERS, SHARMA e NIEDRICH, 2005). Tversky e Shafir (1992) demonstram que quando a opção de resposta é mais complexa, alguns indivíduos passam a utilizar a opção de *status quo*, desta forma, o aumento no número de itens em uma escala de pontos faz a decisão mais complexa, aumentando por sua vez a expectativa do uso da opção de *status quo*.

No entanto, Weathers, Sharma e Niedrich (2005) colocam que o número significativo de alternativas de respostas muda em diferentes categorias de pessoas. Pessoas com maior capacidade cognitiva ou de processamento devem estar aptas a analisar mais alternativas, e usar menos do *status quo*, do que pessoas com menor capacidade.

Estas diferentes visões fomentam ainda mais o surgimento de literatura sobre o número ótimo de opções de respostas utilizado em uma escala. Um dos primeiros trabalhos a levantar esta questão foi de Garner e Hake (1951), onde por meio de uma abordagem teórica quanto à transmissão da informação, os autores relataram o número de categorias de respostas de uma escala para o montante de informação transmitida por ele, onde quanto mais itens na escala forem utilizados, maior será a informação transmitida pela escala.

Cummins e Gullone (2000) destacam que as propriedades básicas de uma escala tipo Likert são confiabilidade, validade e sensibilidade. Esta questão foi levantada por Masters (1974) ao encontrar 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. Abordagens métricas também foram conduzidas para visualizar o número ótimo de categorias de respostas que afetam a fidelidade das escalas com diferentes números de categorias (WISWANATHAN, SUDMAN e JOHNSON, 2004). Churchill e Peter (1984) conduziram um estudo de meta-análise e acharam que a confiabilidade de uma escala aumenta com o aumento do número de categorias de respostas.

No entanto, Miller (1956) e posteriormente Wiswanathan, Sudman e Johnson (2004) detectaram que o uso de aproximadamente sete categorias de resposta é o limite para a habilidade humana distinguir, o que implica no limite de categorias 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. Estas constatações conflitantes refletem a dificuldade em elaborar uma escala confiável com uma métrica que fornece informações adequadas. Neste sentido, Wiswanathan, Sudman e Johnson (2004) colocam como regra geral que o número de categorias de respostas utilizadas em uma escala deve ser entre dois e sete pontos.

Corroborando com os autores supra citados, o estudo baseado no método de simulação Monte-Carlo realizado por Cicchetti, Showalter e Tyrer (1985) encontraram 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 de itens superior. Resultados similares também foram encontrados por Oaster (1989) testando a confiabilidade por meio do teste-retest e da consistência interna do instrumento, concluiu assim que a confiabilidade máxima é obtida com escalas de sete pontos.

Outra questão importante levantada por Tourangeau e Rasinski (1988) é a capacidade de distinção do respondente. Os autores usam como exemplo, o caso de uma medida onde 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.

Desta forma, Tourangeau e Rasinski (1988) recomendam que o pesquisador considere a sensibilidade de mensuração do respondente da escala na definição desta.

Nesta mesma linha, anteriormente ao trabalho de Tourangeau, alguns autores já defendiam o uso da escala proposta por Likert, mantendo os cinco pontos. 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 resposta é suficiente. O autor destaca que o efeito da diminuição do número de opções de escolha encolhe o teste, e 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.

Contudo, o uso de escalas com poucos itens tende a flutuar de amostra para amostra. Desta forma Cummins e Gullone (2000) desencorajam o uso de escala com menos pontos do que cinco, e recomendam como melhor alternativa a escala de 10 pontos que apresentam uma melhor estabilidade indiferente da amostra.

Na mesma linha, Coelho e Esteves (2007) questionam o uso de poucos itens na escala de mensuração, visto que esta pode não fornecer uma boa discriminação das respostas (limitando a habilidade de achar diferenças significantes entre segmentos), o que deve limitar o método de análise dos dados. O uso de um número maior de pontos aumenta a base de dados, enriquecendo a análise dos dados e a cálculo da covariância entre variáveis, usada em grande parte dos testes multivariados de dados. Wiswanathan, Sudman e Johnson (2004) também demonstram a preocupação da relação entre a escala e os testes estatísticos, visto que a definição do número de itens na escala vai afetar os testes estatísticos a serem realizados. 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, 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 e JOHNSON, 2004).

Em uma linha de pesquisa diferente das pesquisas que buscam encontrar uma quantidade de categorias de respostas ideal, Barnette (2000) trabalhou com 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 desaconselhadas. Contudo, no 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. No entanto, 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.

O uso da bidirecionalidade nas respostas tem sido apontada como uma alternativa, principalmente para evitar a inversão da ordem de questões (ROBINSON, SHAVER e WRIGHTSMAN, 1991). A bidirecionalidade dos itens pode utilizar apenas uma mesma direção – por exemplo, todos positivamente orientados, sem a utilização de qualquer variação do não – variando apenas as respostas (BARNETTE, 2000).

Nesta mesma temática, Cummins e Gullone (2000) colocam que quando uma pessoa responde uma escala de mensuração de atitude tipo Likert, o resultado deriva da combinação

da direção de resposta (concordo e não concordo) com a intensidade da resposta (extensão do concordar e não concordar). No entanto, na visão do autor, de fato somente 10% deste composto pode ser atribuído à intensidade, e o restante está atribuído à direção.

O uso de âncoras verbais também é discutido na literatura. DeVellis (1991) recomenda a utilização de palavras indicando a intensidade da atitude que se está mensurando, sempre seguindo um contínuo. No entanto, Collings (2006) coloca que a utilização de descrições verbais em cada questão e categoria de resposta dificulta a resposta. Para isso o uso de números para ancorar cada opção 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, e também facilita o uso de numerais na 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) destacam o uso do ponto final na escala baseado em palavras, ou seja, se utiliza 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, pode-se utilizar âncoras (números ou espaços pré-estabelecidos), indicando a intensidade do item de maneira que o respondente visualize uma progressão.

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), tipo nem satisfeito, nem insatisfeito. Collings (2006) coloca que o ponto neutro é existente em escalas com opções de respostas ímpar. Esta opção, defendida por alguns autores, pode fazer os respondentes sentirem-se mais confortáveis em responder, e defendida por alguns autores. Porém o autor destaca que por outro lado, o ponto neutro pode gerar ambivalência e indiferença do respondente, destoando à verdadeira opinião do respondente. 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. Cummins e Gullone (2000) complementam dizendo que a ambigüidade causada pelo ponto neutro, pode ser sanada com a utilização da uma opção do tipo sem condições de opinar e desta forma não destoa à verdadeira opinião do respondente.

Por fim, Halpin, Halpin e Arbet (1994) sugerem que a melhor opção de escolha para o número de itens na escala depende do conteúdo que a escala pretende mensurar. Coelho e Esteves (2007) colocam que respondentes com maior 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 maior habilidade em identificar importantes relações, e por consequência resultará em uma alta validade dos construtos e melhores testes das hipóteses tanto na teoria como na prática.

3. Método do Estudo

Ao nível epistemológico, o presente estudo possui um cunho quantitativo, com corte transversal, por meio de uma *survey*. Como instrumento de coleta de dados optou-se pela utilização de questionários. 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. As Figuras 03, 04, 05 e 06 representam as escalas inseridas nos instrumentos.

Discordo	1	2	3	Concordo
Totalmente				Totalmente

Figura 03: Escala tipo Likert com três pontos
Fonte: Elaborado pelos autores

Discordo	1	2	3	4	5	Concordo
Totalmente						Totalmente

Figura 04: Escala tipo Likert com cinco pontos
Fonte: Elaborado pelos autores

Discordo	1	2	3	4	5	6	7	Concordo
Totalmente								Totalmente

Figura 05: Escala tipo Likert com sete pontos
Fonte: Elaborado pelos autores

Concordo	5	4	3	2	1	Discordo
Totalmente						Totalmente

Figura 06: Escala tipo Likert com cinco pontos invertida
Fonte: Elaborado pelos autores

Cada questionário era composto por vinte e uma questões em escala tipo Likert além de três questões referentes à facilidade de uso, velocidade de uso e precisão do instrumento. As questões tinham como tema “dinheiro”. A escolha do tema justifica-se pela necessidade de o mesmo 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.

Para a aplicação dos questionários adotou-se o seguinte procedimento. Inicialmente foi distribuída uma etiqueta enumerada a cada entrevistado e foi explicado que para cada questionário recebido no campo “código de identificação” 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 recebeu um questionário de cada vez, ou seja, após o preenchimento do primeiro questionário o mesmo 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 é do tipo não-probabilística, selecionada por conveniência, formada por estudantes do curso de administração da Universidade Federal de Santa Maria. Foram entrevistados todos os alunos presentes em sala de aula nos dias 25, 26 e 27 de março de 2008, perfazendo um total de 211 entrevistados.

Para a análise dos resultados foram utilizados o Alpha de Crombach, o coeficiente de correlação de Pearson, estatísticas descritivas e testes de diferença de média.

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 Crombach que, de acordo com Hair *et al.* (2005), deve possuir um valor 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 *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. 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 as mesmas possam ser comparadas. Assim, optou-se por transformar as escalas dos instrumentos de cinco e sete pontos, para que todos os instrumentos apresentassem mínimo igual a 1 e máximo igual a três. As Tabelas 01 e 02 apresentam os valores utilizados para a transformação das escalas de cinco e sete pontos.

Tabela 01: 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	Valores				
Original	1	2	3	4	5
Padronizada	1	1,5	2	2,5	3

Fonte: Elaborado pelos autores

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

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

Fonte: Elaborado pelos autores

4. Análise dos Resultados

Inicialmente cada uma das quatro escalas foi avaliada quanto à confiabilidade. Para tanto se utilizou o Alpha de Crombach, que avalia a extensão em que os itens formam uma escala internamente consistente. A Tabela 03 apresenta os resultados do Alpha de Crombach.

Tabela 03: Valores do Alpha de Crombach para cada um dos instrumentos

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

Fonte: Elaborado pelos autores

Hair *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, 2001). 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 04: 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 04 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 escala 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 em 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 um. 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 05 apresenta os valores do teste t e o valor da significância para cada variável.

Tabela 05: 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, onde 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 06) calculou-se para cada variável o percentual de entrevistados que marcaram resultados diferentes nos dois instrumentos, ou seja, mudaram de posição. Na segunda (Tabela 07) realizou-se um teste t para verificar se, em média, a inversão da escala alterada os resultados.

Caso os entrevistados não sofressem nenhuma influência da formatação da escala, as respostas aos dois instrumentos seriam idênticas. No entanto, como observado na Tabela 06, em torno de 33% dos entrevistados marcaram respostas diferentes entre o instrumento com a escala de cinco pontos (discordo totalmente para concordo totalmente) e o 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 07. 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 06: 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

Variáveis	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	70,50	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

Tabela 07: 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 <i>t</i>	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 08: 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 dos três instrumentos apresentarem médias altas, o instrumento com a menor escala apresentou a maior média (8,68) e o 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 em aumentar a complexidade do processo de tomada de decisão, e conseqüentemente 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 (3 pontos para cinco pontos) e muito maior do que o mesmo aumento em escala com mais itens (5 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 09: 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 onde um aumento na escala passando de cinco para sete pontos implica em 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 a mesma mostrou-se ser bem mais precisa do que a de três pontos e não se diferencia 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 maior 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

É muito comum 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 valor "x". Tal fenômeno pode sugerir inicialmente que a escolha de uma escala de medida não seja um fator decisivo para o resultado da pesquisa. 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 em discutir a importância do ponto neutro; (2) quanto ao uso de âncoras, se elas serão verbais, se serão unidirecionais ou bidirecionais, se aparecerão apenas nos extremos. 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. 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.

Desde que Likert 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 até meta-análises. Independentemente do método utilizado, foram encontrados resultados diferentes, indicando deste 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 do 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, pode-se sugerir que:

- 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 maiores números de itens geralmente são mais indicado quando os entrevistados dominam o assunto objeto de estudo ou quando o objeto de estudo tem muitos atributos;
- Deve-se evitar o uso de escala com formatos diferentes, principalmente com inversão do sentido num mesmo instrumento, já que alguns entrevistados mudam de posição. 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 700. Cabe destacar ainda que alguns artigos argumentam que quando o número de opções é muito grande o indivíduo pode apresentar uma predisposição à 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 a vontade no momento de expressar sua opinião. Caso a escala seja “par”, a literatura tem sugerido a inclusão da opção “sem condições de opinar”;
- O tamanho da amostra e os testes estatísticos que se pretende usar na análise dos resultados, influencia 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. 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 poluindo 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, e os itens em específico podem ser ancorados por números.

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 elaboração de instrumentos de coleta de dados, como a ordem das questões, a sua formulação e a disposição no questionário, contribuem no processo de reflexão acerca das questões metodológicas, que são uma demanda constante da academia.

Referências Bibliográficas

- BARNETTE, J. Jackson. 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**, n. 13, v. 1, p. 40-52, 1988.
- CLASON, Dennis L. e DORMODY, Thomas 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. e TYRER, P.J. The effect of number of rating scale categories on levels of interater reliability: A Monte Carlo investigation. **Applied Psychological Measurement**. n. 9, p. 31-36, 1985.
- CHANG, L. A. A psychometric evaluation of 4-point and 6-point Likert-type scale in relation to reliability and validity. **Appl Psychol Meas**. n. 18, v. 2, p. 05-15, 1994.
- CHURCHILL, Gilbert A. e PETER, J. Paul. Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis, **Journal of Marketing Research**. n. 21, v. 4, p. 360-375, 1984.
- COELHO, Pedro S. e ESTEVES, Susana 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, David P. **Selecting a questionnaire response scale for student feedback surveys: a comparison of psychometric properties and student preferences among three alternatives**. Master dissertation. Murdoch University. 2006.
- COTE, J.A. e BUCKLEY, M.R. Measurement error and theory testing in consumer research: an illustration of the importance of construct validation. **Journal Consumer Research**. n. 14, p. 579-582, 1988.
- CUMMINS, Robert A. e GULLONE, Eleonora. Why we should not use 5-point Likert scales: the case for subjective quality of life measurement. In. **Second International Conference on Quality of Life in Cities**, Singapore, 2000.
- DEVELLIS, R. F. **Scale development: theory and applications**. Newbury Park: Sage, 1991.
- FREYD, M. The graphic rating scale. **Journal of Educational Psychology**. n. 14, p. 83-102, 1923.
- GARNER, W. R. e HAKE, H. W. The amount of information in absolute judgments. **Psychological Review**. n. 58, p. 446-459, 1951.
- GREEN, P. E. e RAO, V. R. Rating Scales and Information Recovery – How Many Scales and Response Categories to Use. **Journal of Marketing**, v. 34, p. 33-39, 1970.
- HALPIN, G., HALPIN, G., e 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., F. *et al.* **Fundamentos de Métodos de Pesquisa em Administração**. Porto Alegre: Bookman, 2005.
- JENKINS, G. D. e TABER, T. D. A Monte Carlo Study of Factors Affecting Three Indices of Composite Scale Reliability. **Journal of Applied Psychology**. v.62, p.392-398, 1977.
- KROSNICK, Jon A. e BERENT, Matthew K. Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format. **American Journal of Political Science**. n. 37, v. 3, p. 941-964, 1993.
- LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**. n. 140, p. 44-53, 1932.
- LISSITZ, R.W. e GREEN, S.B. Effect of the number of scale points on reliability: A Monte Carlo approach. **Journal of Applied Psychology**. n. 60, p. 10-13, 1975.
- MALHOTRA, N. K. **Pesquisa de Marketing: Uma Orientação Aplicada**. 3.ed. Porto Alegre: 2001.
- MASTERS, James 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.
- MATTAR, F. N. **Pesquisa de Marketing**. 5. ed. São Paulo: Editora Atlas, 1999.
- MILLER, George 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**. n. 68, p. 539-550, 1989.
- PRESTON, Carolyn C. e COLEMAN, Andrew M. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. **Acta Psychologica**. n. 104, p. 1-15, 2000.
- ROBINSON, J. P.; SHAVER, P. R. e WRIGHTSMAN, L.S. **Measures of personality and social psychological attitudes**. San Diego, CA: Academic Press, 1991.
- RODRIGUEZ, Michael C. Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. **Educational Measurement: Issues and Practice**. Summer, 2005.
- SWAIT, J Swait e 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. e RASINSKI, K.A. Cognitive processes underlying context effects in attitude measurement. **Psychology Bull.** n. 103, p. 299-314, 1988.
- TVERSKY, A. e SHAFIR, E. Choice under conflict: the dynamics of deferred decision, **Psychology Science**. n. 6, p. 358-361, 1992.
- WATSON, G.B. Happiness among adult students of education. **Journal of Educational Psychology**. n. 21, p. 79-109, 1930.
- WEATHERS, Danny; SHARMA, Subhash e NIEDRICH, Ronald 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**. n. 58, p. 1516-1524, 2005.
- WISWANATHAN, M.; SUDMAN, S. e JOHSON, M. Maximum versus meaningful discrimination in scale response: implications for validity of measurement of consumer perception about products. **Journal of Business Research**. n. 57, p. 08-24, 2004.