



Programa de Pós Graduação em
Ensino de Química
Instituto de Química
Universidade Federal do Rio de Janeiro

Elaboração, validação e aplicação de escalas psicométricas para professores do Ensino Básico



Daiane Cruz de Brito
Guilherme Cordeiro da Graça de Oliveira

VAMOS ELABORAR UMA ESCALA PSICOMÉTRICA?

Esse manual é uma iniciativa a ajudar professores e profissionais do Ensino Básico a construírem de forma prática instrumentos de medida de processos mentais.

O professor será capaz de construir seu próprio instrumento de avaliação válido para o traço latente¹ que ele deseja investigar em seus alunos.

SUMÁRIO

INTRODUÇÃO

Escalas psicométricas: o que são e para que servem?

1. HORA DE ELABORAR UMA ESCALA

Dimensionalidade do constructo.
Definição constitutiva e operacional do constructo.

Operacionalização em itens: 12 critérios. Quantidade de itens.

2. VAMOS TORNÁ-LA VÁLIDA?

Análise Semântica. Análise fatorial ou análise de juízes.

3. AGORA PODEMOS APLICAR

Público-alvo. Quantidade de respondentes.

4. POR FIM, ANALISAR OS RESULTADOS!

CONCLUSÃO

INTRODUÇÃO

Escalas psicométricas: o que são e para que servem?

A psicometria representa a teoria e a técnica de medida dos processos mentais. Ela é aplicada especialmente na área da Psicologia e da Educação. E se fundamenta na teoria da medida em ciências em geral, ou seja, do método quantitativo que tem, como principal característica e vantagem, o fato de representar o conhecimento da natureza com maior precisão do que a utilização da linguagem comum para descrever a observação dos fenômenos naturais.

Historicamente, a psicometria tem suas origens na psicofísica dos psicólogos alemães Ernst Heinrich Weber e Gustav Fechner. O inglês Francis Galton também contribuiu para o desenvolvimento da psicometria, criando testes para medir processos mentais; inclusive, ele é considerado o criador da psicometria. Foi, contudo,

¹ A teoria do traço latente se refere a uma família de modelos matemáticos que relaciona variáveis observáveis (itens de um teste) com traços hipotéticos não-

observáveis responsáveis pelo aparecimento das variáveis observáveis traduzíveis em respostas ou comportamentos emitidos pelo sujeito que são as variáveis observáveis.

Leon Louis Thurstone, o criador da análise fatorial múltipla, que deu o tom à psicometria, diferenciando-a da psicofísica. Esta foi definida como a medida de processos diretamente observáveis, ou seja, o estímulo e a resposta do organismo, enquanto a psicometria consistia na medida do comportamento do organismo por meio de processos mentais (lei do julgamento comparativo).

A psicometria procura explicar, através das escalas psicométricas, o sentido que têm as respostas dadas pelos sujeitos a uma série de tarefas, tipicamente chamadas de itens. Ou seja, uma explicação quantitativa para aspectos qualitativos.

Numa escala psicométrica, você apresenta ao sujeito um estímulo ou uma série de estímulos (tais como, itens de um teste) e ele responde aos mesmos. A partir das respostas dadas pelo sujeito, isto é, analisando as suas respostas aos itens especificados, pode-se inferir sobre o traço latente do sujeito, hipotetizando relações entre as respostas observadas deste sujeito com o nível do seu traço latente. Estas relações podem ser expressas por meio de equação matemática que descreve a forma de função que estas relações assumem.

O que precisamos saber com esta definição é que, ao se medirem os comportamentos (itens), que são a representação física do traço latente, está-se medindo o próprio traço latente (PASQUALI, 1998).

CAPÍTULO 1: HORA DE ELABORAR UMA ESCALA

Para elaborar uma escala psicométrica, precisamos passar por 3 etapas: a teórica, a experimental e a estatística.

A parte teórica é a hora de tornar clara a teoria sobre o traço latente que se pretende analisar (também chamado de construto) para conseguirmos representar bem em itens do instrumento que será construído.

Deve-se levantar toda a evidência empírica sobre o constructo psicológico que o instrumento pretende medir e procurar sistematizá-la e, assim, chegar a uma mini teoria sobre ele, que o possa guiar na elaboração do instrumento de medida para tal constructo. O que chamamos de dimensionalidade do constructo.

Inicia então a formulação de definições detalhadas do traço ou constructo, derivadas da teoria psicológica encontrada na literatura existente. Essa etapa chamada de

definição constitutiva e operacional do constructo é importante para conseguirmos representá-lo em tarefas comportamentais (ou itens da escala).

A definição constitutiva consiste na definição do constructo por meio de outros constructos, situando assim o constructo, exata e precisamente, dentro da teoria, dando, portanto, as balizas e os limites que ele possui. Boas definições constitutivas vão permitir em seguida avaliar a qualidade do instrumento, que mede o constructo em termos do quanto de sua extensão semântica é coberta pelo instrumento.

Já a definição operacional do constructo é a passagem da teoria (abstrato) para o instrumento de medida (concreto). Nessa etapa o constructo deve ser definido não mais em termos de outros construtos, mas em termos de operações concretas, isto é, de comportamentos físicos através dos quais tal constructo se expressa.

Por exemplo, se você deseja investigar a motivação para os estudos entre seus alunos, inicialmente precisa ir à literatura especializada para construir um conhecimento sólido acerca do é motivação e do que se constitui num comportamento que se pode interpretar como motivado. Parâmetros observáveis de comportamento tais como a percepção

de competência, os vínculos afetivos ou autonomia (possibilidade de escolha) são expressões características de um comportamento motivado. Neste exemplo, sua escala poderá ter como referencial a abordagem organísmica da motivação (REEVE, 2006).

Apesar de nenhuma definição operacional esgotar a amplitude semântica de um constructo, ou seja, ser capaz de cobrir 100 % dele, esta deve ser o mais abrangente possível, pois quanto maior covariância existir entre constructo e definição operacional, maior qualidade deve-se atribuir a esta definição do constructo e, por consequência, maior chance terá o instrumento que de tal definição resulta ser superior em qualidade. É aqui que se fundamenta a validade de um instrumento - a legitimidade da representação empírica, referente ao comportamento, dos traços latentes (os constructos) e, portanto, quanto melhor e mais completa for esta especificação, melhor será a garantia de que o instrumento que resultar para a medida do constructo será válido e útil.

Agora sim pode ser feita a construção dos itens, que são a expressão da representação comportamental do constructo, a saber: as tarefas (os itens do instrumento) que

as pessoas deverão executar para que se possa avaliar a magnitude de presença do constructo (ou atributo). Nesta etapa, chamada de operacionalização em itens, os itens da escala são então preparados para se adequarem às definições do construto. Análises empíricas dos itens seguem, selecionando-se finalmente os itens mais eficazes (válidos) da amostra inicial de itens.

Para uma elaboração adequada dos itens deve-se obedecer a algumas regras ou critérios fundamentais que se aplicam à construção de cada item individualmente e ao conjunto dos itens que medem um mesmo constructo. São eles:

1. Critério comportamental: Os itens devem expressar um comportamento, não uma abstração ou constructo;
2. Critério de deseabilidade: Os itens devem expressar preferências. Não existem neste caso respostas certas ou erradas, e sim diferentes gostos, preferências, sentimentos e modos de ser;
3. Critério da simplicidade: Os itens devem expressar uma única ideia. Itens que introduzem explicações de termos ou oferecem razões ou justificativas são normalmente confusos porque

introduzem ideias variadas e confundem o respondente.

4. Critério da clareza: Os itens devem ser facilmente compreendidos por parte da população em estudo, até mesmo para o estrato mais baixo desta população. Para isso, deve-se utilizar frases curtas, com expressões simples e inequívocas. Frases longas e negativas incorrem facilmente na falta de clareza.

5. Critério da relevância ou pertinência: A expressão (frase) deve ser consistente com o traço (atributo, fator, propriedade psicológica) definido e com as outras frases que cobrem o mesmo atributo. O critério diz respeito à saturação que o item tem com o constructo, representada pela carga fatorial na análise fatorial e que constitui a covariância (correlação) entre o item e o fator (traço).

6. Critério da precisão: O item deve possuir uma posição definida no contínuo do atributo e ser distinto dos demais itens que cobrem o mesmo contínuo.

7. Critério da variedade: Deve-se variar a linguagem, pois o uso dos mesmos termos em todos os itens confunde as frases e dificulta diferenciá-las, além de provocar monotonia,

cansaço e aborrecimento. No caso de escalas de preferências recomenda-se também formular a metade dos itens em termos favoráveis e metade em termos desfavoráveis, para evitar erro da resposta estereotipada à esquerda ou à direita da escala de resposta.

8. Critério da modalidade: Deve-se formular frases com expressões de reação modal, isto é, não utilizar expressões extremadas. A intensidade da reação da pessoa deve ser dada na escala de resposta.

9. Critério da tipicidade: Deve-se formular frases com expressões condizentes (típicas, próprias, inerentes) com o atributo.

10. Critério da credibilidade (face validity): Os itens devem ser formulados de modo que não apareçam como ridículo, despropositado ou infantil. Isso pode afetar negativamente a resposta ao teste, pois o indivíduo respondente pode sentir-se ofendido, irritado ou algo similar.

11. Critério da amplitude: O conjunto dos itens referentes ao mesmo atributo deve cobrir toda a extensão de magnitude do contínuo desse atributo.

12. Critério do equilíbrio: Os itens do mesmo contínuo devem cobrir igual ou

proporcionalmente todos os segmentos (setores) do contínuo, devendo haver, portanto, itens fracos, moderados e extremos. A razão deste critério encontra-se no fato de que a grande maioria dos traços latentes se distribuem entre a população mais ou menos dentro da curva normal, isto é, a maioria das pessoas possuem magnitudes medianas dos traços latentes, sendo que uns poucos possuem magnitudes grandes e outros magnitudes pequenas.

Quanto à quantidade de itens que um instrumento de medida deve ter, não há um número definido de itens na literatura, deve-se elaborar uma quantidade que seja suficiente para cobrir a totalidade ou a maior parte ou, pelo menos, grande parte da extensão semântica do constructo, explicitada nas definições constitutivas.

Obedecendo ao critério de desejabilidade, um questionário para melhor medir comportamentos desejáveis (atitude) deve ser do tipo *Likert*. Uma escala do tipo Likert constitui um processo de escalonamento proposto por Rensis

Likert² é um tipo de instrumento de resposta psicométrica. Composta por um conjunto de frases (itens ou assertivas), para cada assertiva se pede ao sujeito respondente que manifeste o grau de concordância numa escala que pode ir, por exemplo: desde o “discordo totalmente” (normalmente nível 1) até o “concordo totalmente” (normalmente níveis 5, 7 ou 11). A escala do tipo Likert é uma escala de atitude, a atitude intervém em todas as relações do sujeito com o exterior: a nível físico, social, abstrato, e estas relações são observáveis. São estas relações que se tornam o objeto de medição quando se constrói uma escala de atitudes. Mede-se a atitude do sujeito somando, ou calculando a média, do nível selecionado para cada item.

Likert preconizava uma escala de 5 pontos, mas, atualmente, na mais diversa literatura preconiza-se a utilização de escalas de 3, 4, 7 ou 11 pontos, alegando a falta de poder discriminatório de cada sujeito quando a

escala tem muitas possibilidades de resposta, ou inversamente, alegando que só com muitos pontos a escala se assemelha ao continuum da opinião, ou traduzindo os resultados de diversas experiências que demonstram que há tendência para se responder na classe central, devendo por isso evitá-la.

Formulada a escala, a próxima etapa visa estabelecer a análise teórica dos itens.

CAPÍTULO 2: VAMOS TORNÁ-LA VÁLIDA

Para validarmos a escala, precisamos analisá-la de duas formas, ambas feitas por juízes, mas comportam dois tipos distintos de juízes: análise sobre a compreensão dos itens (análise semântica) e a pertinência dos itens ao atributo que se deseja medir (análise de constructo ou análise de juízes).

Para a análise semântica, os juízes são sujeitos da própria população em estudo. Nessa fase, a principal preocupação é de verificar se os itens

² Nascido nos Estados Unidos em 1903, Rensis Likert faleceu em 1981. Filho de engenheiro, seguiu inicialmente os passos do pai realizando seu treinamento profissional em engenharia. Influenciado pelos conflitos que pode observar durante a grande greve de 1922, resolveu estudar as organizações e o comportamento humano. Likert recebeu seu bacharelado em Sociologia pela Universidade de Michigan em 1926. O campo da sociologia

na década de 1920 foi altamente experimental e incorporou muitos aspectos da psicologia moderna. Em 1932 recebeu seu Ph.D. em psicologia pela Universidade de Columbia. Em sua tese de doutorado, elaborou uma escala de pesquisa como um meio de medir atitudes, mostrando que são captadas mais informações do que outros métodos utilizados na época.

são inteligíveis para toda a população meta. De forma alguma a dificuldade na compreensão dos itens deve ser um fator complicador na resposta dos indivíduos. Uma técnica que se tem mostrado mais eficaz na avaliação da compreensão dos itens consiste em checá-los com pequenos grupos de pessoas (3 ou 4) numa situação de "brainstorming". Essa técnica funciona da seguinte forma: constitui-se um grupo de até 4 pessoas do estrato mais simples da população-meta. A este grupo é apresentado item por item, pedindo que ele seja reproduzido pelos membros do grupo. Se a reprodução do item não deixar nenhuma dúvida, o item é corretamente compreendido. Se surgirem divergências na reprodução do item ou se o pesquisador perceber que ele está sendo entendido diferentemente do que ele, pesquisador julga que deveria ser entendido, tal item tem problemas. Dada esta situação, o pesquisador então explica ao grupo o que ele pretendia dizer com tal item. Normalmente, neste caso, as próprias pessoas do grupo irão sugerir como se deveria formular o item para expressar o que o pesquisador quer dizer com ele; e aí está o item reformulado como deve ser.

Para a análise de juízes ou de constructo, às vezes chamada de análise de conteúdo, deve ser propriamente chamada de análise de constructo, dado que precisamente procura verificar a adequação da representação comportamental do(s) atributo(s) latente(s). Nessa análise, os juízes devem ser peritos na área do constructo, pois sua tarefa consiste em ajuizar se os itens estão se referindo ou não ao traço em questão, com base nas definições constitutivas dos constructos/fatores para os quais se criaram os itens. Uma concordância de, pelo menos, 80% entre os juízes pode servir de critério de decisão sobre a pertinência do item ao traço a que teoricamente se refere. A análise de construto pode ser feita como alternativa ou como forma de confirmação através de uma análise fatorial.

Completa-se então os procedimentos teóricos da construção de uma escala psicométrica que antecede os procedimentos experimentais (os quais consistem em coletar informação empírica válida) e os procedimentos analíticos ou estatísticos (os quais consistem em analisar os dados obtidos e a consistência do instrumento de medida).

CAPÍTULO 3: AGORA PODEMOS APLICAR A ESCALA

Definido o público-alvo da escala, deve-se tomar alguns cuidados quando a pesquisa envolve coleta de opiniões de estudantes, principalmente menores de idade. Além de possuir a anuênciada(s) escola(s) que se pretende(m) aplicar, o pesquisador deve possuir também a autorização para coleta de dados dos participantes da pesquisa e, se menores de idade, dos seus pais ou responsáveis. Essa autorização é obtida através do termo de assentimento livre esclarecido (TALE) e do termo de consentimento livre esclarecido (TCLE), respectivamente. Exemplos de TALE e de TCLE podem ser encontrados na literatura³.

É importante também, que o pesquisador submeta sua pesquisa a um Comitê de Ética em Pesquisa⁴, tendo em vista a responsabilidade da pesquisa aplicada a seres humanos. Além do mais o pesquisador está criando um instrumento de medida do “zero” e não está apenas utilizando uma

escala já validada e aprovada quanto as questões éticas do estudo.

Antes dos estudantes iniciarem o preenchimento da escala, é fundamental que sejam informados sobre a proposta da pesquisa e a confidencialidade ou não das respostas, havendo ou não a necessidade de identificação pessoal. Além disso, é importante assegurá-los de que a atividade não vale nota, bem como de que não há prejuízo nas notas ou qualquer problema, caso não desejem participar, e que não há respostas certas ou errados, pois a intenção é colher as preferências individuais de cada um.

Usualmente, aplica-se a pesquisa a um número de respondentes no mínimo 5 vezes maior que a quantidade de itens estipulada pelo pesquisador, a fim de que o resultado obtido em pesquisa seja o mais próximo possível da realidade.

Finalizada a coleta de dados, o próximo passo é trabalhar e analisar os resultados obtidos.

³ <https://www.ufmg.br/bioetica/coep/tale/>

⁴ É um colegiado interdisciplinar e independente, com “múnus público”, que deve existir nas instituições que realizam pesquisas envolvendo seres humanos no Brasil, criado para defender os interesses dos participantes de pesquisa em sua integridade

e dignidade e para contribuir no desenvolvimento da pesquisa dentro dos padrões éticos (Normas e Diretrizes Regulamentadoras da Pesquisa Envolvendo Seres Humanos – Resolução CNS n.º 466/12). Pode ser acessado através do preenchimento da base nacional Plataforma Brasil.

CAPÍTULO 4: POR FIM ANALISAR OS RESULTADOS

Para a análise dos resultados é necessário a transcrição dos dados obtidos para um editor de planilhas. Sugerimos o Programa de dados Excel para realização de separação dos dados, como exemplo, uma separação em grupos caso queira realizar uma posterior comparação entre grupos, além da facilidade de realizar procedimentos matemáticos de análise. Para uma análise estatística dos dados, sugerimos os Programas estatísticos Biostat versão 2008 e o Statistical Package for the Social Science – SPSS, versão 22.0.

A análise dos resultados é feita através de estatística descritiva (cálculo das médias e desvios padrões) e estatística inferencial (comparação das variáveis entre os grupos através de um teste estatístico apropriado). Inicialmente, aderência dos dados à distribuição normal deve ser testada através de um teste de normalidade. A partir da quantidade de amostras e a dependência ou não entre elas, deve-se escolher o teste adequado.

Na internet existem vários tutoriais gratuitos que orientam quanto a escolha do melhor teste a se aplicar.

Particularmente, recomendamos o site: <<https://eaulas.usp.br/portal/video.action?idItem=8721>>.

Por exemplo, na comparação entre grupos independentes com distribuição que não seguem a normalidade, a significância das respostas dos grupos pode ser obtida através de um teste não paramétrico, a exemplo o teste de Mann-Whitney. O teste de Mann-Whitney é aplicado quando estão em comparação dois grupos independentes através de uma variável dependente qualitativa ordinal (um atributo que possui uma ordenação tal como a escala de Likert). São criadas, portanto, hipóteses nulas H_0 onde H_0 é testada para um nível de significância (α) geralmente utilizado na literatura o de 5,0%. Dessa forma, quando a probabilidade de significância $p > \alpha$ mantém-se H_0 , enquanto, quando $p < \alpha$, H_0 é refutada.

CONSIDERAÇÕES FINAIS

Realizada todas as etapas, você criou um instrumento de medida psicológico validado semanticamente e em constructo, e pode utilizá-lo para investigar diversos aspectos que podem interferir no processo de ensino e aprendizagem.

Com sua escala psicométrica será possível gerar uma resposta a cada estudante e ao conjunto, sobre determinados comportamentos (procedimentos) específicos que você deseja analisar. E, através desse levantamento, planejar melhor suas atividades frente ao grupo de estudantes a se trabalhar, identificar o perfil dos estudantes ou até mesmo comparar grupos de diferentes espécies, a exemplo, gêneros, idades, momentos de escolarização, diferentes escolas, dentre outros.

Caso queira, você poderá utilizar uma escala psicométrica já existente que se adapte ao perfil do

público que se deseja investigar. Nossa grupo de pesquisa do Instituto de Química da UFRJ vem trabalhando na elaboração de escalas psicométricas e, parcerias para a aplicação de nossas escalas podem ser solicitadas. Duas escalas encontram-se disponíveis para utilização por parte desse grupo de trabalho: uma que avalia o perfil motivacional de estudantes que cursam Química e uma que avalia o perfil afetivo do professor de Ciências ou Química entre estudantes do Ensino Básico. Para maiores informações entrar em contato com Daiane Brito através do e-mail daianecruzbrito@yahoo.com.br.

REFERÊNCIAS

PASQUALI. Princípios de elaboração de escalas Psicológicas, *Rev. psiquiatr. clín.*, São Paulo, 25(5):206-13, 1998.

REEVE, Johnmarthal. Motivação e emoção. 4. Ed. Rio de Janeiro: LTC, 2006.