

A MEDIDA NAS PESQUISAS EM EDUCAÇÃO: EMPREGANDO O MODELO RASCH PARA ACESSAR E AVALIAR TRAÇOS LATENTES

Amanda Amantes Ribeiro*
Geide Rosa Coelho**

RESUMO: Neste trabalho, fazemos uma reflexão acerca das potencialidades da associação de métodos qualitativos e quantitativos para responder a questões específicas, no sentido de obter maior coerência interna nas pesquisas da área educacional. Temos como foco discutir o modelo *Rasch* como ferramenta para acessar traços latentes, apresentando um exemplo de como esse modelo pode ser promissor para trabalharmos com medidas, assim como responder a questões de natureza causal e que se remetem à identificação de efeitos e mudanças.

Palavras-chave: Modelo *Rasch*. Traços latentes. Metodologia qualitativa-quantitativa.

MEDICIÓN EN LAS INVESTIGACIONES EN EDUCACIÓN: EL MODELO RASCH PARA ACCEDER Y EVALUAR RASGOS LATENTES

RESUMEN: En este trabajo reflexionamos acerca de las potencialidades de la asociación de métodos cualitativos y cuantitativos para contestar a cuestiones específicas, con el objetivo de obtener más coherencia interna en las investigaciones del área educacional. Tenemos como foco la discusión del modelo *Rasch* como herramienta para acceder rasgos latentes, presentando un ejemplo de cómo ese modelo puede ser promisor para trabajar con medidas, así como contestar cuestiones de naturaleza causal y que remiten a la identificación de efectos y cambios.

Palabras clave: Modelo Rasch. Rasgos latentes. Metodología cualitativa-cuantitativa

Doutora em Educação pela Universidade Federal de Minas Gerais (UFMG). Professora Adjunta da Faculdade de Educação da Universidade Federal da Bahia (UFBA).
Email: amandaamantes@gmail.com

Doutor em Educação pela Universidade Federal de Minas Gerais (UFMG). Atualmente é professor da área de ensino de Física no Centro de Educação da Universidade Federal do Espírito Santo (UFES).
Email: geidecoelho@gmail.com

Rafael Marinho - Professor do Instituto Federal de Minas Gerais (IFETMG).
Email: rafael.marinho@ifmg.edu.br

DOI - <http://dx.doi.org/10.1590/1983-21172015170306>

**MEASUREMENT IN TEACHING RESEARCH:
APPLYING RASCH MODEL TO ACCESS LATENT TRACES**

ABSTRACT: We report a discussion about the importance of combining qualitative and quantitative methods to lead specific questions in educational area. We point out the relevance of this approach in order to improve ours methods and obtain greater internal consistency. Rasch model is presented as a tool to access latent traces. We show how this model can be promising to work with measures, as well as answering questions of causality and questions which intend to identify effects and changes.

Keywords: Rasch model. Latent traits. Qualitative-quantitative analysis.

INTRODUÇÃO

O campo educacional tem sido demarcado por pesquisas que subtemem principalmente métodos de análise de natureza qualitativa. O foco em apresentar um quadro de pessoas e relações em um contexto particular, assim como descrever, de forma detalhada, o fenômeno, muito tem contribuído para investigar questões pertinentes sobre aprendizagem e outros processos a ela relacionados.

Entretanto, problemas de natureza causal, questões que se remetem à comparação de parâmetros e questões que requerem uma investigação de efeitos de atributos ou de elementos inerentes ao processo de ensino aprendizagem, têm se revelado cada vez mais essenciais para compreender esse processo. Tais problemas demandam metodologias que evidenciem mudanças no tempo, diferenças entre traços ou estados latentes e padrões de comportamento. Nesse sentido, a análise qualitativa, embora essencial em primeira instância, torna-se limitada, pois não fornece elementos suficientes para interpretar o processo: seus procedimentos não permitem mensurar e/ou comparar os atributos de maneira a atender critérios específicos de objetividade e confiabilidade, no sentido estatístico do termo.

Muitos pesquisadores da área defendem que é necessário empregar métodos mistos para que evidências específicas sejam construídas e interpretadas de maneira adequada. A triangulação de dados e de técnicas, assim como a utilização de metodologias quali-quantitativas de análise, são maneiras consideradas válidas para garantir uma maior adequação e coerência nas investigações (SHAFFER; SERLIN, 2004; GOLAFSHANI, 2003). Contudo, essa coerência demanda não só o emprego de tais métodos, mas também a conformidade da metodologia aos objetivos, questões, problemas, referenciais teóricos e resultados.

O emprego de métodos mistos pode fornecer elementos consistentes para lidar com muitos aspectos problemáticos de investigações, mas, para isso, deve-se atentar para a qualidade das medidas dos atributos que se pretende investigar. Na área de educação e psicologia, os objetos de estudo, na maioria das vezes, correspondem a atributos ou traços latentes, ou seja, que não são acessados ou mensurados diretamente (PASQUALI, 2008). Mensurar esses atributos, ou estabelecer uma escala para avaliá-los, é necessário quando se pretende fazer comparações entre diferentes indivíduos, entre diferentes grupos ou entre diferentes momentos ao longo do tempo (SINGER; WILLETT, 2003).

Nesse sentido, trabalhos na área da psicologia, mais precisamente a Psicomетria, usualmente empregam modelos probabilísticos que carregam em seus pressupostos teóricos a dimensão subjetiva dos traços latentes. Esse movimento em lidar mais objetivamente com tais traços também pode ser observado, ainda que de forma mais branda, na área de ensino.

Entre os diferentes modelos probabilísticos, o modelo Rasch (RASCH, 1960) tem se constituído como um procedimento promissor para interpretar parâmetros que não podem ser acessados diretamente (MEAD, 2008). Ele proporciona uma escala intervalar que permite interpretar as distâncias, em termos de

medidas, dos traços latentes de diferentes sujeitos, bem como avaliar características dos testes ou instrumentos aplicados.

O presente trabalho tem como propósito discutir, brevemente, o ato de medir e sua validade enquanto método no âmbito da pesquisa em Educação, assim como apresentar o modelo *Rasch* para acessar traços latentes e analisar dados dicotômicos, exemplificando a utilização das medidas obtidas para investigar fenômenos educacionais.

O ato de medir na área de Educação

A medida em Educação pode ter muitos significados a depender do contexto em que é incorporada. Muitos dos objetos de estudo da área correspondem a traços latentes, que são atributos abstratos, construídos (ou inventados) a partir de uma teoria (ou um conjunto de teorias), mas que podem ser relacionados a dados observáveis por meio dessa mesma teoria. Assim, traços latentes como entendimento, compreensão, emoção, percepção, atitudes, habilidades e tantos outros relacionados à aprendizagem podem ser avaliados desde que tenhamos uma teoria substancial que os associe a algum elemento explicitamente observado. Descrever qualitativamente esses atributos analisando os elementos a que se referem é de suma importância para avaliar os detalhes que lhes são característicos, assim como obter indícios para sustentar argumentos qualitativos em relação a esses atributos.

Contudo, quando lidamos com comparações (entre sujeitos, entre grupos, entre momentos diferentes) consideradas teoricamente distintas em uma escala hierárquica de valores (ainda que de cunho estritamente qualitativo), admitimos que há alguma diferença em relação a esses traços. Investigar essas diferenças não é uma tarefa de fácil execução, pois demanda estabelecer elementos confiáveis para que comparações sejam realizadas de maneira adequada. Isso implica a definição teórica dos traços, assim como a definição da mensuração de elementos correlatos e a determinação de métodos que evidenciem a mudança, como a construção de escalas, análises de progresso, testes de hipóteses e outros.

A medida da mudança é um elemento importante para avaliarmos o desenvolvimento humano, pois somente através dela podemos mapear e entender a aquisição de novas habilidades, o progresso no entendimento de novos conteúdos, o crescimento intelectual e físico, o desenvolvimento de interesses ao longo da vida (SINGER; WILLET, 2003). Nesse sentido, apropriação de métodos quantitativos se faz necessária, pois a constatação da mudança implica em mensuração e avaliação de diferenças subjacentes às características dos traços investigados.

Não defendemos, entretanto, a substituição de uma metodologia qualitativa por outra quantitativa. Acreditamos que lidamos, em primeira instância, com análises qualitativas, pois, como coloca Gorard (2002, p. 346), “todos os métodos na pesquisa educacional lidam com qualidades, mesmo que sejam contadas” e a depender dos propósitos da investigação ela é suficiente para obtermos resultados consistentes do ponto de vista acadêmico.

Alguns objetivos, contudo, demandam métodos estatísticos. Nesse ponto,

devemos “superar o falso dualismo ‘qualitativo’ e ‘quantitativo’” (PRING, 2000 citado por GORARD, 2002, p. 346) para podermos elaborar e utilizar metodologias válidas, que abarquem técnicas de ambas as vertentes, mas com a devida precaução em manter sempre a coerência interna da pesquisa.

Consideremos, por exemplo, uma pesquisa que tenha como objetivo a investigação do efeito de uma intervenção didática na aprendizagem dos alunos. Um caminho promissor é verificar se houve ou não aprendizagem avaliando a competência em diferentes momentos: antes, durante e depois da intervenção (AMANTES, 2009). Isso, por si só, já exige pelo menos uma estatística descritiva (e como defendemos, exige mais do que isso). Caso contrário, poderemos ter, no máximo, “indícios”, já que as diferenças encontradas podem ser fruto do acaso da escolha da amostra ou do acaso na interação do aluno com o instrumento de avaliação.

Além disso, para identificarmos se a intervenção é a provável causa da aprendizagem, devemos estar cientes dos efeitos que outros fatores (além da intervenção) podem ter na aprendizagem dos alunos. Mas se o número de fatores que queremos cercar for grande (e, de fato, o é nos ambientes de aprendizagem), precisaremos de uma amostra grande de alunos, para que o efeito desses fatores não seja avaliado com base em casos particulares.

Queremos ressaltar, porém, que nem todo tipo de dado numérico é adequado ao tratamento por estatísticas paramétricas, como cálculos de médias e desvios amostrais e populacionais, testes de hipóteses, análises de variância, regressão, entre muitos outros (TRIOIA, 2008). Portanto, nos casos em que as perguntas de pesquisa só possam ser abordadas por métodos mistos (quali-quant), devemos tomar o cuidado de verificar se os indicadores observáveis (que se relacionam, por meio de uma teoria, ao construto latente) podem ser representados em uma escala intervalar.

Atualmente, pesquisas em Ciências Humanas têm se apropriado de metodologias e métodos para lidar com observáveis em termos de escalas intervalares de medidas. Modelos psicométricos, Teoria de Resposta ao Item e modelos Rasch são cada vez mais utilizados nesse sentido, e o intuito principal é obter uma escala que permita atribuir aos traços latentes um caráter de medida. No entanto, para que possamos considerar os valores estimados como “medidas”, em um sentido mais restrito do termo, devemos verificar se a relação entre eles e os observáveis atendem certas regras, denominadas de “características fundamentais de mensuração”. Por isso, é fundamental fazermos uma reflexão sobre o próprio ato de medir e o seu significado no âmbito das pesquisas em educação.

Quando pensamos no ato de medir, o instrumento mais simples que podemos imaginar é uma régua. As pessoas pensam na régua como um instrumento preciso (desde que bem calibrada), capaz de fornecer, por exemplo, a medida de comprimento de um objeto, sem maiores problemas, dentro de um limite de precisão.

Uma reflexão um pouco mais aprofundada nos faz pensar que nenhuma régua, por mais bem calibrada que seja, está isenta de problemas fundamentais.

Primeiramente, qualquer régua apresentará, em maior ou menor grau, uma calibração imperfeita. Em segundo lugar, o comprimento da régua depende de outros fatores, como, por exemplo, a temperatura em que ela se encontra (e em última instância até mesmo da pressão no meio em que ela está imersa). Além disso, o ato de medir o comprimento de um objeto com uma régua sempre dependerá de ações humanas, como posicionar corretamente a régua e fazer a leitura. A medição do comprimento de um objeto é, em última instância, uma interpretação subjetiva sobre a posição das extremidades do objeto em relação a uma referência (no caso, as marcas na régua). A régua é um dispositivo que operacionaliza, de forma imperfeita, a ideia abstrata de comprimento (WRIGHT, 1998).

Mas é claro que essas são ponderações que, apesar de filosoficamente relevantes, não tornam problemática a ideia de medir, por exemplo, a altura de uma criança e compará-la com a altura de outras crianças da população. Ou seja, mesmo com todas essas questões a respeito do processo de medida, podemos criar uma dimensão abstrata que chamamos “comprimento”, uma unidade abstrata que denominamos “metro” e procedimentos concretos que nos permitem operacionalizar as ideias de “comprimento” e “metro” na prática. Desde que sistemas de medidas produzidas. Utilizar outro elemento coesivo para fazer a retomada no texto) estejam funcionando bem, ou seja, desde que esse procedimento nos forneça informações coerentes e úteis, não há motivos para que nos preocupemos demais com as componentes subjetivas e aleatórias do processo de medida.

Quando analisamos o processo de medir temperaturas, a situação se torna um pouco mais problemática. Uma unidade de medida de temperatura, como o °C não é diretamente associada a algo concreto (como o metro pode ser associado a um bastão de determinado comprimento). Pensar em uma unidade de medida de temperatura nos faz perceber mais facilmente o quanto há de invenção humana na criação de uma variável mensurável, de uma dimensão de medida e de uma escala. Thurstone (1931, p. 257, tradução nossa) já ponderava que

There is a popular fallacy that a unit of measurement is a thing—such as a piece of yardstick. This not so. A unit of measurement is always a process of some kind which can be repeated without modification in the different parts of the measurement continuum..¹

Diante do exposto, podemos considerar que as Ciências Exatas e as Ciências Humanas compartilham alguns elementos no que diz respeito à mensuração de seus objetos de estudo. Como qualquer dimensão de medida (qualquer variável) é uma criação abstrata, assim como qualquer escala, o desafio para se fazer medidas em Ciências Humanas é o de elaborar construtos e unidades de medida que possam ser operacionalizados de forma mais sistemática.

Podemos idealizar um construto chamado “proficiência em física”, por exemplo, e construir instrumentos para medi-lo. No entanto, esse processo de mensuração deve obedecer a certas regras. Se isso for possível, poderemos obter medidas da “proficiência em física” e usá-las para comparar diferentes sujeitos ou

monitorar a evolução de um sujeito através do tempo (sempre em relação a esse construto). Se não for possível (ou improdutivo) operacionalizar a “proficiência em física”, podemos tentar modificar a forma que a definimos, como, por exemplo, separá-la em componentes menores. Nesse caso, o que nos permite concluir se a definição que criamos para um construto está coerente é a verificação de sua adequação ao que denominamos “características fundamentais de mensuração”.

O matemático dinamarquês Georg Rasch considera como característica fundamental de uma medida a independência entre instrumento e objeto a ser medido. Em outras palavras, as propriedades (temperatura, no caso dos termômetros) dos objetos não dependem das características dos instrumentos, e as propriedades dos instrumentos (leitura da marca) não dependem das características dos objetos. A essa ideia, Rasch chamou de “objetividade específica” (RASCH, 1977).

O mesmo raciocínio pode ser aplicado no caso em que se quer medir um atributo como “proficiência em física” (em vez de temperatura) através de itens de física (em vez de termômetros). Para que tenhamos uma medida válida, as proficiências estimadas dos sujeitos não podem depender de quais itens estão sendo usados para estimá-la, assim como as dificuldades dos itens não podem depender de quais sujeitos estão respondendo aos itens.

Assim, se o processo de medida ocorrer como esperado, temos uma operacionalização útil de um construto (abstrato e inventado) chamado “proficiência em física”. Felizmente, em muitos casos em Ciências Humanas, e na Educação em especial, é possível obter operacionalizações úteis de ideias abstratas, como podemos verificar em trabalhos de vários pesquisadores da área (BERGSTROM et al., 1992; BOND et al., 1997; BOND; KING, 2003; MASTERS, 2004; GRIFFIN, 2004; BOND, 2005; CRONIN et al., 2005; FOX, 2007; DRANEY, 2007). Essa operacionalização, contudo, não é direta e subtende uma complexidade com a qual devemos lidar. No caso das Ciências Humanas, muitos fatores desconhecidos estão envolvidos, ficando mais evidente a influência dos componentes aleatórios (inerente a qualquer ato de medida). Um modelo que relacione um observável a um construto teórico no contexto das Ciências Humanas é melhor delimitado em uma perspectiva probabilística.

Dessa forma, se quisermos criar medidas de atributos latentes baseados em observáveis, devemos assegurar que as medidas desses atributos atendam ao requerimento de objetividade específica em sua formulação probabilística. Os modelos psicométricos que nos permitem estimar medidas de traços latentes em escalas intervalares a partir de observáveis, respeitando esse requerimento, são chamados de modelos Rasch.

Modelo Rasch

George Rasch, na década de 1950, trouxe uma grande contribuição para o desenvolvimento de medidas nas ciências sociais. Segundo Wright e Linacre (1989), ele desenvolveu um modelo matemático probabilístico pautado na interação entre o objeto a ser medido (entendimento, competência, habilidade, etc.) e o

agente de medida (um teste, por exemplo). Seu trabalho se assentou na descrição do desempenho de sujeitos com dislexia ao interagirem com tarefas de leitura de textos. A “descoberta” do modelo se realizou a partir do estudo do modelo estocástico multiplicativo de Poisson (OLSEN, 2003) em conjunto com seus trabalhos anteriores, cujo foco estava em delimitar um Modelo de Crescimento Geral (RASCH, 1952). Uma característica da estratégia de Rasch no delineamento do modelo foi considerar como base de análise uma investigação criteriosa do dado (KÆRGÅRD, 2013) e isso, segundo alguns autores, diferencia o seu modelo dos Modelos da Teoria de Resposta ao Item, em termos de princípios e pressupostos (MEAD, 2008).

A família de modelos Rasch utiliza dados observáveis de forma qualitativa ou quantitativa (convertido em escore de um teste) para construir medidas intervalares de traços latentes (como a medida do entendimento de conceitos físicos) produzindo, dessa forma, medidas comparáveis. O mais familiar dos modelos da família Rasch é o modelo para a análise de dados dicotômicos. Ele é utilizado para analisar questões do tipo certo ou errado; verdadeiro ou falso; sim ou não; concordo ou não concordo. Esse modelo introduz uma expressão para estimar a probabilidade da resposta de um determinado sujeito ao item dicotômico em função de dois parâmetros: um que caracteriza a pessoa, e outro que caracteriza cada item, sendo ambos medidos em uma mesma escala logarítmica (a unidade de medida dessa escala é conhecida por *logit*, contração de *log odds unit*). Para entendimento da formulação do modelo *Rasch* na sua forma dicotômica, consideremos o raciocínio a seguir.

Suponha que dois itens dicotômicos, *i* e *j*, sejam respondidos por um grupo de sujeitos. Alguns acertarão os dois itens, outros errarão os dois itens. Haverá ainda os que acertarão o item *i* e errarão o *j* e os que acertarão o *j* e errarão o *i*.

Se um sujeito *A* acerta o item *i*, mas erra o *j*, há um indício de que o item *j* seja mais difícil que o *i*. Se um sujeito *B* acerta o item *j*, mas erra o *i*, há um indício de que o *i* seja mais difícil que o *j*. Os escores (0- errou; 1- acertou) dos sujeitos que acertaram ou erraram ambos não terão utilidade para que possamos comparar as dificuldades dos itens. Portanto, uma forma de comparar as dificuldades dos itens *i* e *j* é calcular a porcentagem de sujeitos que acertou cada um dos itens, descartando os sujeitos que acertaram ou erraram os dois.

Um exemplo: Suponha que, de um total de 200 respondentes: 35 acertaram os dois itens; 45 erraram os dois itens; 90 acertaram *i*, mas erraram *j*; e 30 acertaram *j*, mas erraram *i*. Descartando os 80 que tiveram o mesmo escore nos dois, sobram 120. Destes, 90 acertaram *i*, o que corresponde a 0,75. Outros 30 acertaram *j*, o que corresponde a 0,25. Podemos então pensar que o item *j* é três vezes ($0,75/0,25 = 3$) mais difícil que o item *i*.

Se, ao aplicarmos os itens para outros grupos de sujeitos, encontrarmos percentuais muito diferentes entre si, a comparação entre as dificuldades dos itens torna-se impossível, pois eles não estarão avaliando o mesmo tipo de habilidade. Em outras palavras, não é possível inventar uma escala de proficiência a partir

desses dois itens.

Mas se os resultados em outras populações forem semelhantes, isso indica que nossa comparação parece ser válida, pois há “algo” nos itens que parece não depender da população para a qual eles são aplicados. Se esses itens abordarem conteúdos de física, podemos pensar em chamar esse “algo” de “proficiência em física” e inventar uma escala útil para comparar sujeitos e itens.

Mesmo nos casos em que identificamos certa coerência nos percentuais de acerto entre diferentes populações e que, por isso, pudemos estimar que o item j (por exemplo) é mais difícil que o item i , sempre haverá sujeitos para, simultaneamente, acertarem o item mais difícil e errarem o mais fácil. Isso se deve ao fato de que qualquer processo de medida em ciências humanas (e em última instância, qualquer processo de medida) tem uma componente estocástica. Por isso, um modelo que nos permita relacionar proficiências, dificuldades e acertos, deve ser de natureza probabilística (e não determinística).

Considerando a probabilidade de um sujeito acertar o item i e errar o item j $P_n(x_i = 1/x_j = 0)$, temos que o valor dessa probabilidade é igual ao produto de duas probabilidades: a probabilidade de ele acertar o item i e a probabilidade de errar o item j .

$$P_n(x_i = 1/x_j = 0) = P_i (1 - P_j) \quad (1)$$

Já a probabilidade de o sujeito n , simultaneamente, errar o item i e acertar o item j é:

$$P_n(x_i = 0/x_j = 1) = (1 - P_i) P_j \quad (2)$$

Podemos comparar essas duas probabilidades (e, dessa forma, as dificuldades dos itens i e j) fazendo a razão de uma pela outra:

$$\frac{P_n(x_i = 1/x_j = 0)}{P_n(x_i = 0/x_j = 1)} = \frac{P_i (1 - P_j)}{P_j (1 - P_i)} \quad (3)$$

Onde a probabilidade do sujeito n acertar o item i deve ser uma função da habilidade do sujeito n e da dificuldade do item i $P_{ni} = f(n, i)$.

Mas, de acordo com o princípio da objetividade específica, as dificuldades dos itens devem ser comparadas independentemente das pessoas escolhidas para realizar esses itens. Dessa forma, a razão entre as probabilidades de acerto de cada item (condicionadas ao erro do outro item) deve ser a mesma para dois sujeitos n e m :

$$\frac{P_n(x_i = 1/x_j = 0)}{P_n(x_i = 0/x_j = 1)} = \frac{P_m(x_i = 1/x_j = 0)}{P_m(x_i = 0/x_j = 1)} \quad (4)$$

Ou

$$\frac{P_{\dot{n}}(1 - P_{\dot{j}})}{P_{\dot{j}}(1 - P_{\dot{n}})} = \frac{P_{\dot{m}}(1 - P_{\dot{j}})}{P_{\dot{j}}(1 - P_{\dot{m}})} \quad (5)$$

Para que se possa ter uma comparação objetiva de dois itens, a equação 5 deve ser satisfeita. Se ela não for satisfeita, ou seja, não se ajustar aos dados empíricos, os nossos itens não estão avaliando a mesma coisa. Nesse caso, não poderemos construir uma escala unidimensional abstrata e alocar as dificuldades dos itens sobre essa escala.

Como a equação 5 deve ser verdadeira para quaisquer sujeitos e quaisquer itens, podemos escolher, então, um sujeito referência, $m = o$, e um item referência, $j = o$, de forma que a probabilidade de o sujeito de referência acertar o item de referência seja igual a 50%, $P_o = 0,5$. A probabilidade de erro será: $(1 - P_o) = 0,5$.

Substituindo esses valores na equação 5, temos:

$$\frac{P_{\dot{n}}(1 - P_o)}{P_o(1 - P_{\dot{n}})} = \frac{P_o(1 - P_o)}{P_o(1 - P_o)} = \frac{P_o}{(1 - P_o)} \quad (6)$$

Ou, com um pouco de álgebra simples:

$$\frac{P_{\dot{n}}}{(1 - P_{\dot{n}})} = \frac{P_o}{(1 - P_o)} \times \frac{P_o}{(1 - P_o)} \quad (7)$$

A equação 7 nos indica que a razão das probabilidades de acerto e erro do sujeito n no item i depende, como era esperado, de características de n e i . No entanto, os fatores dependentes de n e i podem ser separados um do outro.

É importante ressaltar que a expressão $\frac{P_o}{(1 - P_o)}$ não depende do item i , mas apenas de n e do item de referência. Ou seja, essa expressão tem o mesmo valor, qualquer que seja o item respondido pelo sujeito. Da mesma forma, o valor da fração

$\frac{P_o}{(1 - P_o)}$ não depende de qual sujeito responde ao item i . Ela nada tem a ver com

qualquer atributo de n , mas apenas com atributos de i . Dessa forma, podemos escrever:

$$\frac{P_{ni}}{(1-P_{ni})} = g(n) \times h(i) \quad (8)$$

Para obter uma escala em que as diferenças entre medidas tenham significado prático (uma escala intervalar), devemos transformar o produto do segundo termo da equação em uma soma. Seremos, então, capazes de estimar a probabilidade de acerto com base em somas e diferenças de funções. Fazemos isso tomando o logaritmo natural em ambos os lados da equação.

$$\ln \left(\frac{P_{ni}}{(1-P_{ni})} \right) = \ln [g(n) \times h(i)] = \ln [g(n)] + \ln [h(i)] \quad (9)$$

Ou

$$\ln \left(\frac{P_{ni}}{(1-P_{ni})} \right) = G(n) + H(i) \quad (10)$$

Definindo a proficiência do sujeito como sendo $B_n = G(n)$ e a dificuldade do item como sendo $D_i = -H(i)$ (de modo que uma dificuldade maior diminua a probabilidade de acerto), podemos reescrever a equação acima como:

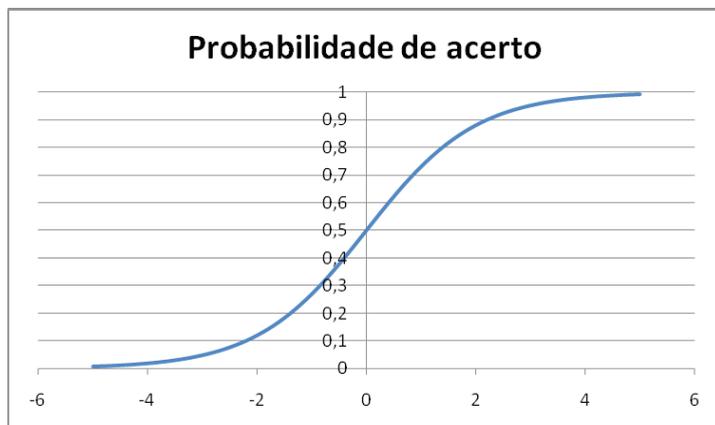
$$\ln \left(\frac{P_{ni}}{(1-P_{ni})} \right) = B_n - D_i \quad (11)$$

Ou, resolvendo para ,

$$P_{ni} = \frac{e^{(B_n - D_i)}}{1 + e^{(B_n - D_i)}} \quad (12)$$

Onde e é igual à base do logaritmo natural ($e = 2,7183\dots$). É importante notar que quanto maior for a diferença entre a proficiência do sujeito e a dificuldade do item, maior será a probabilidade de acerto. Ou seja, quanto maior for habilidade do sujeito, por exemplo, e menor a dificuldade do item, maior a chance desse sujeito acertar. O gráfico abaixo representa a probabilidade de acerto de um item em função dessa diferença.

Gráfico 1 – Probabilidade de acerto de um item em função da diferença entre a proficiência do sujeito e a dificuldade do item



Fonte: Coelho (2011)

As estimativas para as proficiências e para as dificuldades são feitas baseadas na contagem do número de acertos de cada sujeito e de cada item. Um método de interação e não interação é seguido até que as estimativas feitas pelo modelo estejam de acordo com os dados observados. Existem também outros modelos, derivados deste, que atendem aos requerimentos de mensuração discutidos, mas adaptados ao contexto de itens politômicos (Polytomous Rasch Models) (BOND; FOX, 2007; MEAD, 2008). Esses modelos, entretanto, não serão discutidos neste artigo.

As medidas de estimativas dos parâmetros dos sujeitos e dos itens, obtidas a partir de modelos como o *Rasch*, são mais confiáveis para serem utilizadas enquanto indicativo do traço latente do que o escore bruto. A nota ou o escore bruto obtido em um teste produz uma escala ordinal. Esse escore bruto permite ter alguma noção do espaçamento entre valores na escala (em uma escala de 0 a 100, escores de 20 e 25 descrevem situações muito mais próximas que escores de 60 e 90). No entanto, o significado de uma dada diferença (por exemplo, 5 pontos) não é a mesma em diferentes regiões da escala (geralmente, as diferenças de escores em regiões centrais da escala correspondem a diferenças menores nas proficiências e habilidades no mesmo ponto).

Quando usamos os dados obtidos em testes para estimar proficiências de sujeitos e dificuldades de itens de acordo o modelo *Rasch*, obtemos parâmetros cujas diferenças têm um significado que se mantém em diferentes regiões da escala: um acréscimo de uma unidade (que chamamos de 1 *logit*) de proficiência ou de dificuldade do item implica em um aumento de 2,718 vezes (número de Euler, e) na razão entre a probabilidade de acerto e de erro de um item. Isso é válido para qualquer região da escala. Podemos dizer, portanto, que uma análise Rasch produz medidas em uma escala intervalar.

Uma escala intervalar traz, pelo menos, duas grandes vantagens em relação

a escalas ordinais. A primeira é a facilidade de interpretação das diferenças. Quando estamos interessados em medir a aprendizagem de conteúdos, por exemplo, estamos interessados na diferença entre a proficiência inicial e a final dos sujeitos². Essa diferença só tem significado se as proficiências estiverem representadas em uma escala intervalar. Uma segunda vantagem de uma escala intervalar é a possibilidade de usar estatísticas paramétricas para analisar os dados. Dados que estejam representados em escalas ordinais (como escores em testes) nem sempre são adequados a esse tipo de tratamento, exigindo estatísticas não paramétricas (que são menos poderosas). Muitas vezes, pesquisas na área de Humanas usam métodos paramétricos, como teste t e ANOVA, em dados provenientes de escores brutos, o que é um procedimento equivocado. Isso porque tais testes têm como pressuposto que a variável analisada é de natureza intervalar; se esse pressuposto é violado, produzem-se estatísticas enviesadas, não susceptíveis a interpretações plausíveis.

Modelo *Rasch* na Pesquisa em Educação

Investigações na área de ensino/educação têm se apropriado de modelos probabilísticos para apurar atributos relacionados ao processo de aprendizagem. Thomas (2002), por exemplo, submete itens sobre leitura e matemática à avaliação de professores e analisa seus resultados a partir da Taxonomia SOLO (BIGGS; COLLIS, 1982) e modelagem *Rasch*. Dawson (2006) modela os dados relativos a respostas de estudantes a testes virtuais para identificar níveis de compreensão sobre o conceito de Energia. Planinic, Ivanjek e Susac (2010) utilizam o modelo *Rasch* para mapear o entendimento conceitual de 1676 estudantes da Croácia sobre os conteúdos de Mecânica, utilizando o FCI (Force Concept Inventory, HESTENES; WELLS; SWACKHAMER, 1992). Morris et al. (2012) discutem as propriedades do FCI enquanto instrumento de acesso ao conhecimento conceitual aplicando a Teoria de Resposta ao Item (TRI), cujo primeiro modelo é considerado, por alguns autores, como o modelo *Rasch* dicotômico. Segundo esses últimos autores, tal procedimento é importante porque pode auxiliar professores a ter ferramentas consistentes para avaliar níveis de conhecimento intuitivo e conhecimento científico. Tais como esses, muitos pesquisadores incorporam esses modelos para acessar, mensurar, avaliar e interpretar traços psicológicos e cognitivos.

Uma forma que tem se mostrado promissora para avaliar aprendizagem é a associação de análises categóricas a escalas que podem ser modeladas nessa perspectiva probabilística. Uma técnica adotada é a elaboração de sistemas de rubricas, já realizada há algum tempo por diferentes pesquisadores na área, como podemos encontrar nos trabalhos de Moskal e Leydens (2000), Roblyer e Ekhaml (2000), Bodzin e Beerer (2003), Hafner, J. e Hafner, P. (2003), Clark e Sampson (2008), e a associação desse sistema a escalas numéricas, como em Amantes (2009) e Coelho (2011).

Como exemplo de aplicação do modelo *Rasch*, apresentamos a análise de dados categóricos realizados a partir de um sistema do tipo rubrica. Esse sistema, denominado TCE (Taxonomia da Complexidade do Entendimento) (AMANTES et al., 2013) é uma ferramenta para acessar e analisar o entendimento sobre conteúdos científicos a partir de uma perspectiva teórica cognitiva de aprendizagem.

O parâmetro teórico que fundamenta a TCE é a aceção de que traços subjacentes à aprendizagem podem se apresentar em diferentes níveis de complexidade, a depender de diversos fatores, tais como o tempo, maturidade, estado emocional, etc. Esse é um ponto de concordância em algumas das teorias relativas ao desenvolvimento cognitivo, como a Teoria de Habilidades Dinâmicas (FISCHER, 1980, 2006) e a Teoria de Complexidade Hierárquica de Commons (COMMONS; PEKKER, 2004). Ferramentas dessa natureza baseiam-se na possibilidade de classificação dos atributos em termos de diferentes níveis de complexidade no momento em que o sujeito realiza a tarefa.

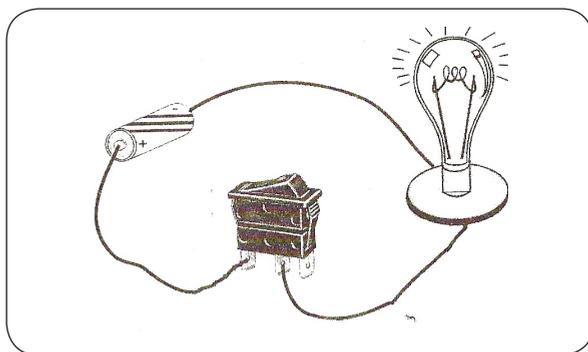
Neste trabalho, reportamos dados categorizados a partir de uma TCE elaborada no estudo desenvolvido por Coelho (2011). Os dados foram modelados para identificar o entendimento dos sujeitos em relação ao conteúdo de eletricidade e corresponderam a respostas fornecidas por 240 estudantes da terceira série do Ensino Médio de uma escola pública federal a testes de conhecimento. Os testes compreendiam questões abertas sobre conceitos tais como diferença de potencial, corrente elétrica, resistência elétrica, etc. As respostas foram classificadas segundo esse sistema e foram, posteriormente, tabuladas para modelagem *Rasch*.

Na TCE existem demarcações de conteúdos por temas, que, por sua vez, podem ser reportados em diferentes níveis de complexidade. Essas demarcações de temas e categorias de complexidade carregam uma perspectiva docente ou acadêmica de avaliação, em que o maior nível corresponde a um entendimento mais completo, sistematizado e correto do ponto de vista científico. Dessa forma, há uma associação da escala categórica qualitativa com a escala do tipo Guttman (Guttman, 1940).

O quadro 1 representa a categorização para níveis de um tema, que corresponde ao entendimento dos estudantes sobre corrente elétrica, ao responderem a questão:

Uma ação cotidiana e corriqueira é apertar um interruptor e acender uma lâmpada, no teto ou no abajur. A figura 1 mostra um modelo mais simples dessa situação: uma pilha comum está ligada a um interruptor e a uma lâmpada de lanterna. Ao pressionar o interruptor, a lâmpada acende. Redija um texto explicando, de forma mais clara possível, tudo o que ocorre na pilha, fios, interruptor e na lâmpada quando ela está acesa.

Figura 1 – Representação de um circuito elétrico simples



Fonte: Coelho (2011)

Quadro 1 – Exemplo do tema “corrente elétrica, entendimentos e respostas típicas do sistema categórico sobre a Física envolvida no funcionamento de circuitos elétricos simples”

Tema	Categoria	Exemplo
<p>Corrente elétrica como fluxo de cargas elétricas FCE1- O estudante entende corrente elétrica como algo que flui pelo circuito.</p>	<p>FCE1- O estudante entende corrente elétrica como algo que flui pelo circuito.</p>	<p>“(…) quando pressionamos o interruptor, fechamos o caminho da corrente, assim deixando que ela alcance a lâmpada.” (Estudante 111)</p>
	<p>FCE2- O estudante entende corrente elétrica como fluxo de carga elétrica no circuito.</p>	<p>“Quando o interruptor é acionado, ele fecha um caminho (...). Assim como esse caminho existe, as cargas vão circular pelo sistema (...).” (Estudante 103)</p>
	<p>FCE3- O estudante entende corrente elétrica como fluxo de elétrons no circuito.</p>	<p>“Ao acionarmos o interruptor (...), ocorre o fechamento do circuito, o que permitirá que haja circulação de elétrons pelo fio.” (Estudante 176)</p>
	<p>FCE4- O estudante entende corrente elétrica como fluxo de elétrons livres do circuito.</p>	<p>“(…) O interruptor pressionado faz com que os elétrons livres, presentes no fio que o liga à pilha, migrem em direção ao polo positivo da mesma.” (Estudante 139)</p>
	<p>FCE5- O estudante entende corrente elétrica como fluxo de elétrons livres do circuito e como meio de transporte de energia.</p>	<p>“(…) Essa energia está contida nos elétrons livres do circuito.” (Estudante 135)</p>

Fonte: Coelho (2011)

Outro conceito que podemos encontrar nas respostas para a mesma questão, considerado como outro tema da TCE, é a resistência elétrica. O quadro 2 apresenta os níveis de complexidade nesse caso.

Quadro 2 – Exemplo do tema “resistência elétrica, entendimentos e respostas típicas do sistema categórico sobre a Física envolvida no funcionamento de circuitos elétricos simples”

Tema	Categoria	Exemplo
Resistência Elétrica	R1- O estudante sugere que existe uma resistência elétrica no circuito.	“Quando essa energia chega à lâmpada, ela encontra uma dificuldade em continuar sendo transmitida pelos elétrons (...)” (Estudante 208)
	R2- O estudante considera que existe uma resistência elétrica no circuito.	“(...) ao passar pelo resistor da lâmpada, a energia elétrica se transforma em energia luminosa.” (Estudante 126)
	R3- O estudante considera que existe uma resistência elétrica no circuito e a associa à oposição ao fluxo da corrente elétrica.	“A lâmpada é um componente predominantemente resistivo, ou seja, dificulta a passagem da corrente elétrica por ele (...)” (Estudante 109)
	R4- O estudante considera que existe uma resistência elétrica no circuito, associando-a a oposição ao fluxo da corrente elétrica e dependente das propriedades do material e da densidade de carga elétrica.	“Quando a corrente elétrica passa pelo filamento da lâmpada (elemento resistivo), esta feita com materiais como de tungstênio, essa lâmpada gera uma queda de tensão devido a sua resistividade alta (...)” (Estudante 105)

Fonte: Coelho (2011)

Os dados obtidos a partir das TCEs correspondem à sistematização de um sistema de rubricas que carrega em si uma hierarquia relacionada a níveis da qualidade do entendimento sobre um conteúdo, nesse caso sobre corrente elétrica e resistência elétrica. Essa matriz pode ser utilizada de diferentes maneiras no intuito de atender a objetivos distintos, dentre eles:

- a) Analisar qualitativamente a configuração do entendimento dos estudantes em cada momento ao longo de uma instrução;
- b) Fazer estudos de caso sobre a evolução do entendimento de sujeitos escolhidos aleatoriamente, a partir da descrição das frequências de categorias em cada momento de medida;
- c) Transformar a matriz de dados categóricos em uma matriz de dados numéricos, o que favorece a aplicação do modelo Rasch para construção de uma escala de proficiência;
- d) Descrever trajetórias de aprendizagem a partir de medidas obtidas com métodos quantitativos que descrevem a matriz de respostas codificadas, em termos de dados dicotômicos.

Na Tabela 1, está apresentado um exemplo de codificação qualitativa transformada em dado dicotômico para posterior modelagem:

Tabela 1 – Exemplo da transformação de dados categóricos em um sistema de dados dicotômicos

Estudante Estudante	Tema: Corrente elétrica como fluxo de cargas elétricas				
	Exemplo				
	FCE1	FCE2	FCE3	FCE4	FCE5
176	1	1	1	0	0

Fonte: Coelho (2011)

Nesse caso, o estudante 176 apresenta a descrição para corrente elétrica como fluxo de cargas da seguinte maneira: “Ao acionarmos o interruptor [...] ocorre o fechamento do circuito, o que permitirá que haja circulação de elétrons pelo fio”. A resposta foi classificada como categoria FCE3, ou seja, no terceiro nível de complexidade. A construção dos itens dicotômicos consiste em atribuir o valor 1 para a categoria de maior complexidade alcançada pelo aluno e também para todas as categorias precedentes. Para todas as outras categorias mais altas não atingidas, o valor 0 é atribuído (FCE4 e FCE5).

Feito isso para todos os temas e para todos os níveis de cada tema, temos uma matriz de resposta dicotômica, que podemos vislumbrar a partir do recorte feito na Tabela 2, dos alunos 175 a 181.

Tabela 2 – Matriz de dados dicotômicos para o tratamento Rasch

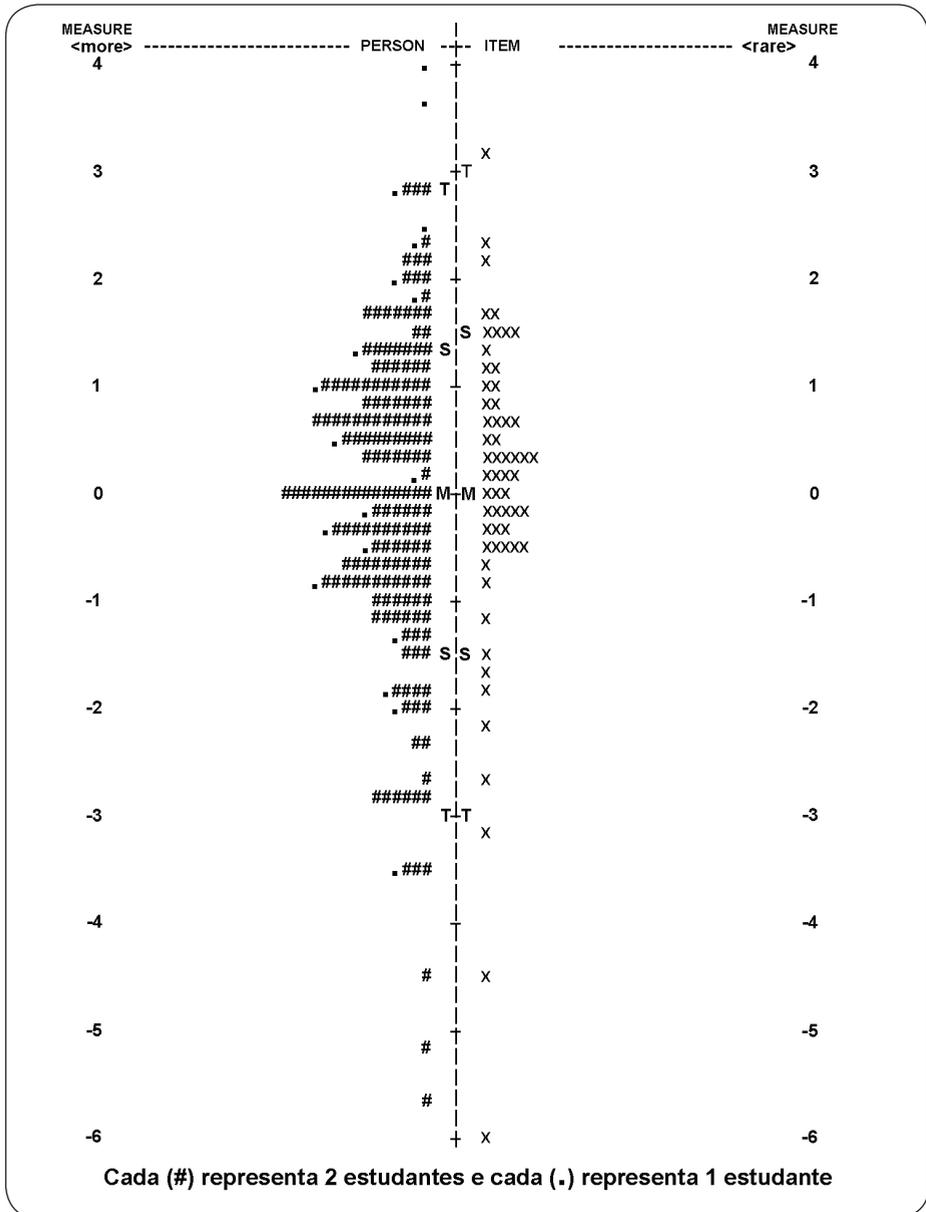
Student	FCE1	FCE2	FCE3	FCE4	FCE5	FDD1	FDD2	FDD3	FAP1	FAP2	FAP3
.											
.											
.											
175	1	1	1	1	0	1	1	1	0	0	0
176	1	1	1	0	0	1	1	0	1	1	1
177	1	1	0	0	0	1	0	0	1	1	1
178	1	0	0	0	0	1	1	0	0	0	0
179	0	0	0	0	0	1	1	0	1	1	0
180	1	1	0	0	0	0	0	0	1	1	0
181	1	1	1	1	1	1	0	0	0	0	0
.											
.											
.											

Fonte: Construída pelos autores a partir dos dados de Coelho (2011):

A matriz reportada na Tabela 2 foi inserida como fonte de dados no programa WISTEPS (LINACRE; WRIGHT, 2000) com programação específica para obtermos os valores das medidas dos sujeitos e dos itens a partir do modelo Rasch Dicotômico Unidimensional. Como saída, obtemos indicadores de adequação do modelo aos dados empíricos (INFIT e OUTFIT), curvas características dos itens, medidas da proficiência, medidas dos itens, mapas dos itens e sujeitos, além de outras estatísticas que nos permitem avaliar o grau de convergência das estimativas, assim como seu histórico e outras propriedades que não discutiremos aqui.

A figura 2 apresenta a régua resultante do tratamento Rasch para um conjunto de dados dicotômicos. A medida do entendimento dos estudantes (do modelo) é mostrada no lado esquerdo da figura, e a medida da dificuldade dos itens (corresponde à estimativa do parâmetro β) é mostrada no lado direito da figura. Cada x no lado direito da figura denota um item da escala. A marca M representa o valor médio da dificuldade dos itens e do entendimento dos estudantes, S indica que os parâmetros estão deslocados da média em um desvio padrão e T indica que os parâmetros estão deslocados da média em dois desvios padrão. A escala apresenta um intervalo que vai de -6 logits a 4 logits.

Figura 2 – Mapa descrevendo nível de entendimento dos estudantes e a complexidade dos itens

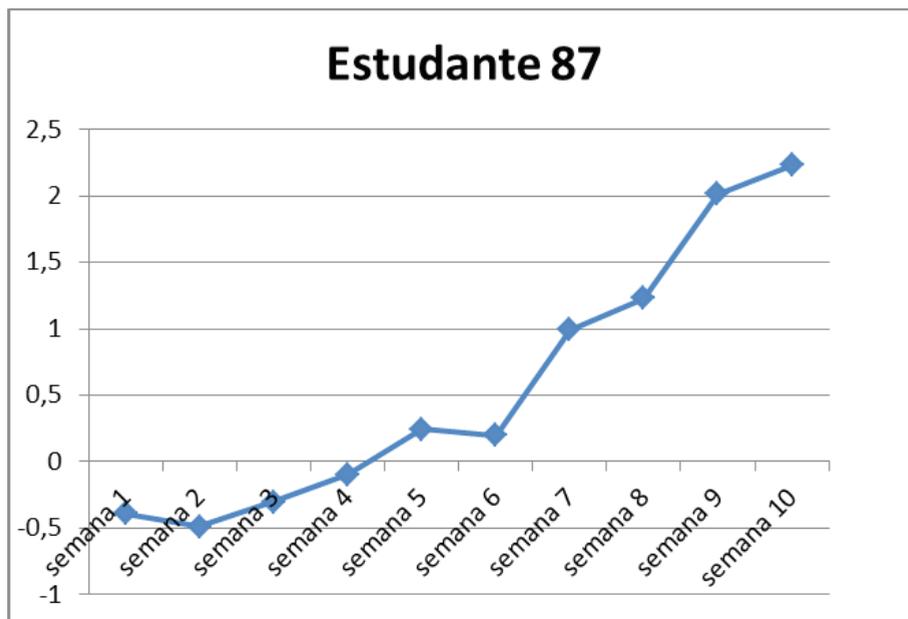


Fonte: Coelho (2011)

Com os valores referentes à medida de cada sujeito, podemos estabelecer, entre outros procedimentos, comparações em termos de diferenças de médias:

a) Dos mesmos estudantes em momentos distintos da instrução: isso nos indica se o entendimento de cada estudante evoluiu no tempo, fica estagnado ou oscila. Esse procedimento também permite obter trajetórias de aprendizagem e, através dessas trajetórias, avaliar como o entendimento se desenvolve. O gráfico 2 é um exemplo de trajetória de aprendizagem para o aluno 87.

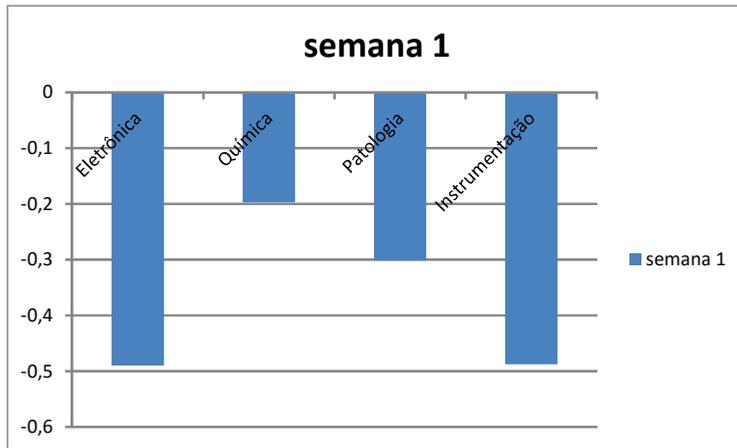
Gráfico 2 – Exemplo de como evolui o entendimento sobre um conteúdo em dez semanas, a partir da medida obtida pelo modelamento *Rasch*



Fonte: Construído pelos autores a partir dos dados de Coelho (2011):

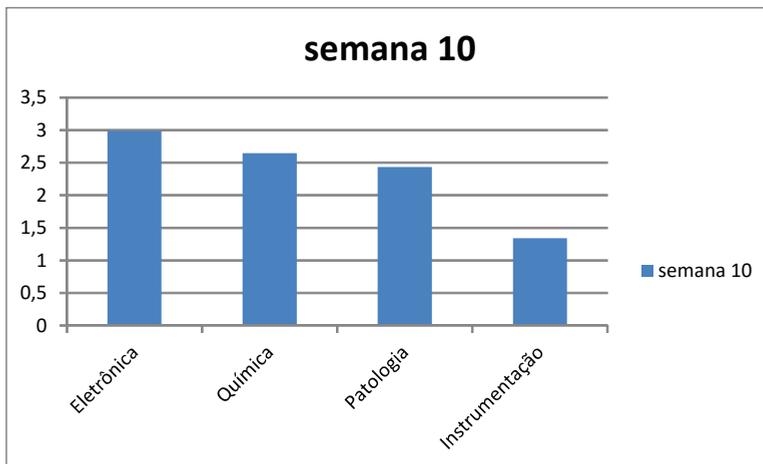
b) De estudantes pertencentes a grupos diferentes, no mesmo instante: esse tipo de comparação é desejável se intencionamos verificar as diferenças que estudantes pertencentes a determinado grupo demonstram no entendimento sobre o conteúdo. A escala intervalar nos permite fazer esse tipo de comparação. Um exemplo pode ser visualizado nos gráficos 3 e 4, em que comparamos, em um determinado momento da instrução, o entendimento dos alunos pertencentes a cursos técnicos diferentes.

Gráfico 3 – Médias dos alunos de cada curso na primeira semana



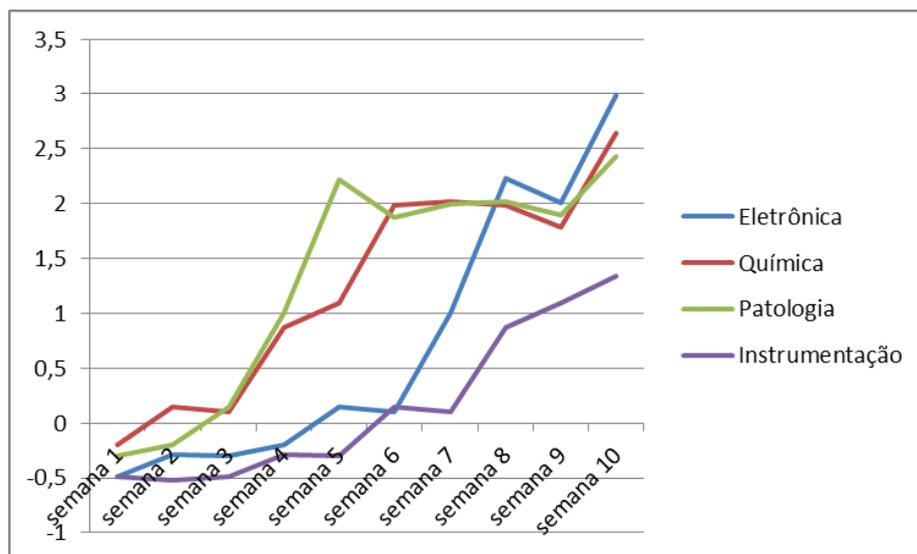
Fonte: Construído pelos autores a partir dos dados de Coelho (2011):

Gráfico 4 – Médias dos alunos de cada curso na décima semana



Fonte: Construído pelos autores a partir dos dados de Coelho (2011):

c) De estudantes pertencentes a grupos diferentes, em momentos distintos: Com esse procedimento, podemos comparar os grupos amostrais em termos de como os sujeitos se comportam em média e se existe diferença entre eles em termos gerais. Isso nos permite avaliar se há diferença, por exemplo, entre a trajetória de aprendizagem de meninos e meninas submetidos a uma mesma instrução. O gráfico 5 mostra um exemplo de trajetórias de aprendizagem de estudantes de cursos diferentes.

Gráfico 5 – Trajetórias de aprendizagem dos estudantes de diferentes cursos técnicos

Construído pelos autores a partir dos dados de Coelho (2011):

Com os valores das medidas dos itens, podemos estabelecer a análise em termos de diferenças de médias:

- a) Das medidas dos parâmetros de itens referentes a um mesmo conceito, em diferentes testes de conhecimento: Com essa ação, verificamos se um determinado conceito está sendo mais aprendido na medida em que o parâmetro do item diminui, ou seja, os itens vão ficando “mais fáceis” ao longo da instrução.
- b) De medidas de itens que se remetem a conceitos diferentes: ao fazermos comparações dessa natureza, podemos ter evidências de quais conteúdos estão sendo mais aprendidos, uma vez que, para esses conteúdos, as medidas dos itens vão diminuindo em valor de *logits* ao longo do tempo.

Além desses procedimentos, podemos aplicar outros testes clássicos aos dados de saída do modelagem *Rasch*, como Análise Fatorial para avaliar consistência interna dos instrumentos, verificar dimensionalidade do traço cognitivo reportado em uma tarefa e validar ferramentas. Se a análise fatorial indicar que nosso instrumento é composto por mais de uma dimensão é composto por mais de um fator, então não poderemos tratá-lo como unidimensional. Temos duas saídas: ou eliminamos (ou trocamos) os itens que carregaram em outro fator, ou analisamos o teste como sendo composto de dois “testes menores” cada um sendo unidimensional.

Podemos, ainda, com as medidas estabelecidas para a proficiência, estabelecer o “ganho” no atributo do sujeito (no nosso caso, no entendimento sobre eletricidade) entre dois ou mais instantes diferentes ao longo da instrução. Esse procedimento nos fornece a possibilidade de verificar, através de testes estatísticos clássicos (como

Análise de Regressão Múltipla e Multinível), a influência de preditores no processo de ensino-aprendizagem, como efeito do professor, efeito do material aplicado, efeito do engajamento do estudante, efeito do gênero, maturidade, especificidade de curso técnico e outras variáveis contextuais que se mostrarem relevantes para o estudo.

Em nosso exemplo, os valores obtidos para o ganho no entendimento foram analisados a partir da análise de regressão multinível (COELHO, 2011). Entretanto, antes de realizar a regressão, foi desenvolvido um estudo exploratório para analisar o efeito de diferentes preditores como gênero, engajamento comportamental e cognitivo, e nível socioeconômico sobre o ganho no entendimento dos estudantes³. Desses preditores somente o engajamento cognitivo apresentou efeito sobre a evolução dos estudantes sendo, portanto, uma variável contextual importante para o estudo.

O indicador de engajamento cognitivo foi definido a partir das trajetórias individuais dos estudantes, tomando como base o escore bruto desses estudantes nos testes aplicados durante a intervenção educacional de eletricidade. Três diferentes grupos foram identificados: (1) o primeiro grupo agrupava os estudantes que mantiveram a estabilidade no engajamento cognitivo (engajcogest); (2) o segundo grupo aglutinaria as categorias com os estudantes que apresentaram instabilidade durante todo o curso ou aumentaram a instabilidade ao final das aulas (categorias engajcogoh e engajcogof); (3) o terceiro grupo corresponderia à categoria na qual os estudantes aumentaram a estabilidade em relação ao engajamento cognitivo ao final do curso (engajcogoa). Os resultados da Tabela 03 indicam que existem diferenças em relação ao ganho dos grupos ($F(2,83) = 5,950$; $p=0,004$) sendo que o maior ganho foi do grupo constituído pelos estudantes que apresentaram engajamento estável ($\alpha = 2,536$; $SD = 1,653$).

Tabela 3 – Resultados do procedimento ANOVA para os ganhos dos diferentes grupos de engajamento cognitivo

Categoria	Ganho (inclinação)		
	Média/ Desvio Padrão (logits)	F	Significância estatística (p)
Engajamento cognitivo estável (engajcogest)	2,536 (1,653)	5,950	0,004
Engajamento cognitivo oscilatório e forçado (engajcogoh e engajcogof)	1,644 (1,235)		
Engajamento cognitivo amortecido (engajcogoa)	1,434 (1,385)		

Fonte: Coelho (2011)

A partir da análise exploratória, Coelho (2011) reconheceu que o preditor engajamento cognitivo deveria ser inserido nos modelos a serem testados na regressão multinível. Diferentes modelos foram testados e foi verificado que o fato de os estudantes pertencerem a uma turma específica e mantiverem o engajamento cognitivo ao longo do curso influenciou na aprendizagem sobre circuitos elétricos. Além de verificar quais são esses fatores, a análise fornece o modelo que melhor explica como eles influenciam o ganho (ou evolução) no entendimento, ou seja, especificam a taxa de evolução (representado pelo valor da inclinação da regressão) referente a cada um dos preditores. No nosso caso, o modelo que melhor descreveu a aprendizagem sobre circuito elétrico pode ser descrito pela equação:

$$Y_j = -0,917 + 1,887(\text{tempo}_j) - 0,822(\text{turma}2_i) + 0,741(\text{engajcogest}_i \text{ } \& \text{ tempo}_j) \quad (13)$$

Por essa equação, temos que a manutenção do engajamento cognitivo ao longo do curso (engajcogest) influenciou no ganho dos estudantes. Como a taxa evolutiva foi positiva (igual a 0,741), podemos dizer que os estudantes desse grupo apresentaram um ganho maior do que os estudantes dos outros grupos de engajamento cognitivo. Nessa equação, temos a variável dependente (ganho no entendimento- Y_{ij}) em função do tempo (tempo $_{ij}$), e podemos perceber que pertencer à turma 2 (estudantes que cursavam somente o ensino médio) influencia o ponto de partida, uma vez que essa parte da equação diz respeito ao intercepto da reta de regressão. Isso quer dizer que os alunos dessa turma apresentam um entendimento inicial significativamente menor que o restante da amostra, mas o fato de pertencer a essa turma não altera a evolução do entendimento em relação aos outros estudantes no decorrer da instrução.

Esses são alguns dos exemplos de como as medidas obtidas pelos modelos Rasch podem servir de dados para análises mais consistentes do ponto de vista estatístico e, ao mesmo tempo, conservar a natureza relativa à qualidade do atributo investigado. Acreditamos que congregando métodos qualitativos e quantitativos, como no caso reportado, ampliamos o leque de questões a serem investigadas, assim como conferimos mais objetividade e coerência metodológica quando lidamos com avaliação de mudanças nos ambientes de ensino.

CONSIDERAÇÕES FINAIS: LIMITES E POSSIBILIDADES DE MEDIDAS EM EDUCAÇÃO

Neste trabalho, apresentamos uma breve discussão sobre a medida em ensino; ressaltamos algumas justificativas para a integração entre metodologias de caráter qualitativo e quantitativo de análise; apresentamos sucintamente o modelo *Rasch* dicotômico e apresentamos um exemplo nesse sentido. Consideramos como ponto essencial a busca da qualidade em nossas pesquisas e, para tanto, defendemos a necessidade de maior coerência interna.

A coerência se realiza a partir do rigor que estabelecemos nas escolhas referentes aos pressupostos teóricos, objetivos e métodos de análise. Para lidar com problemas e questões de diferentes naturezas, é preciso construir um desenho que possa

atender à especificidade dessas questões, ao mesmo tempo em que garanta generalidade aos apontamentos. Desse ponto de vista, devemos reconhecer as dificuldades e limitações de um único método, em decorrência, sobretudo, da complexidade dos objetos com os quais lidamos no âmbito educacional. O reconhecimento de que é necessário ir além de descrições do fenômeno para lidar com problemas causais, de efeitos e de comparações, tem norteado a busca por métodos que possam garantir maior objetividade no tratamento dos dados.

A medida nas Ciências Humanas subtende a consideração de que existem atributos associados a observáveis que podem ser acessados e, em alguma instância, mensurados. Isso nos permite pensar em métodos que nos forneçam parâmetros para lidar com esses atributos em uma escala que tenha propriedades intervalares, o que se constitui em ferramenta poderosa de análise e amplia as possibilidades de investigar fenômenos educacionais.

Devemos ressaltar, entretanto, que a utilização de metodologias que compreendem medidas e análises estatísticas não é suficiente para lidar com todos os aspectos dos fenômenos educacionais, devido, sobretudo, à complexidade inerente aos objetos de estudo dessa área. Os testes estatísticos fornecem parâmetros confiáveis para interpretarmos fenômenos em termos de comparações de atributos que podem ser generalizados, mas isso pode diminuir a riqueza de detalhes que são significativos e essenciais para entender os processos de maneira mais pontual. Nesse sentido, devemos estar cientes de que qualquer que seja a metodologia empregada, ela apresentará limites e possibilidades de aplicação e abrangência e, por isso, a integração de diferentes perspectivas de análise se mostra pertinente quando realizada de forma adequada.

Defendemos que uma melhor compreensão dos aspectos principais relacionados à mensuração no campo educacional nos instrumentaliza a traçar planos e estratégias investigativas que explorem melhor nossos dados e, conseqüentemente, nos forneça evidências mais bem fundamentadas. Com isso, estaremos nos valendo de metodologias cada vez mais robustas e, como resultado, teremos uma ascensão do rigor e qualidade nas nossas pesquisas.

NOTAS

¹Inserir texto original na nota de rodapé.

²Não estamos defendendo a posição de que a evolução das proficiências seria o único – nem o principal – objetivo educacional, nem negando as dificuldades de se criar definições de proficiências operacionalizáveis e com significado substancial.

³O procedimento ANOVA foi utilizado para investigar a diferença média do ganho no entendimento dos diferentes grupos de cada um dos preditores utilizados no referido estudo.

REFERÊNCIAS

- AMANTES, A. (2009). Contextualização no ensino de Física: Efeitos sobre a evolução do entendimento dos estudantes. Tese de Doutorado, 275p. Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.
- BERGSTROM, B. A. Ability measure equivalence of computer adaptive and pencil and paper tests: A research synthesis. *American Educational Research Association*, San Francisco, CA, 1992.
- BIGGS, J.; COLLIS, K. *Evaluating the quality of learning: the SOLO taxonomy*. New York: Academic Press, 1982.
- BODZIN, A. M.; BEERER, K. M. Promoting Inquiry-Based Science Instruction: The Validation of the Science Teacher Inquiry Rubric (STIR). *Journal of Elementary Science Education*. Publisher: Springer Netherlands p. 39-49, 2003.
- BOND, L.; SMITH, T. W.; BAKER, W. K.; HATTIE, J. A.. *The certification system of the National Board for Professional Teaching Standards: a construct and consequential validity study* (Research Report). Greensboro, NC: University of North Carolina at Greensboro, Center for Educational Research and Evaluation, 2000.
- BOND, G. T.; FOX, C. M. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 2. ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2007. 340p.
- CLARK, D.; SAMPSON, V. Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, Michigan State University, n.45, v. 3, p. 293-321, 2008.
- COELHO, G.R. (2011). A evolução do entendimento dos estudantes em eletricidade: um estudo longitudinal. Tese de Doutorado, 173p. Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil.
- COMMONS, M. L.; PEKKER, A. *Hierarchical complexity: A Formal Theory*. 2004. Disponível em: <<http://www.dareassociation.org/papers.php>>. Acesso em: 07 jul. 2013.
- CRONIN, J.; KINGSBURY, G.G.; MCCALL, M.S.; BOWE, B. *The Impact of the No Child Left Behind Act on Student Achievement and Growth*. Northwest: Evaluation Association, 2005.
- DAWSON, T. L. Stage-like patterns in the development of conceptions of energy. In: LIU, X.; BOONE, W. (Ed.). *Applications of Rasch measurement in science education* Maple Grove, MN: JAM Press, 2006. p. 111-136.
- DAWSON, T. L.; STEIN, Z. Cycles of research and application in education: Learning pathways for energy concepts. *Mind, Brain e Education*, Cambridge, Massachusetts, v. 2, n. 2, p. 90-103, 2008.
- DESIGN-BASED RESEARCH COLLECTIVE. Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, Flórida, USA, v. 32, n. 1, p. 5-8, Jan./Fev. 2003.
- DRANEY, K.; WILSON, M. Application of the Saltus model to stage-like data: Some applications and current developments. In: VON DAVIER, M.; CARSTENSEN, C. H. (Ed.). *Multivariate and mixture distribution Rasch models: Extensions and applications*. New York: Springer, 2007.
- FISCHER, K. W. A theory of cognitive development: the control and construction of hierarchies of skills. *Psychological Review*, United States, v. 87, p. 477-531, 1980.
- FISCHER, K. W. Dynamic cycles of Cognitive and Brain development. In: BATTRO, A. M.; FISCHER, K. W. (Ed.). *The educated brain*. Cambridge, U.K.: Cambridge University Press, 2008.
- GOLAFSHANI, N. Understanding reliability and validity in qualitative research. *The Qualitative Report*, Canadá, v. 8, n. 4, p. 597-607, 2003.

- GORARD, Stephen. Can we overcome the methodological schism? Four models for combining qualitative and quantitative evidence. *Research Papers in Education*, University of Exeter, UK, v. 17, n. 4, p. 345-361, 2002.
- GRIFFIN, P. The comfort of competence and the uncertainty of assessment. *Hong Kong School Principals' Conference*, Hong Kong, 2004.
- GUTTMAN, L. A basis for scaling qualitative data. *American Sociological Review*, Vanderbilt University, USA, n. 9, v. 2, p. 139-150, 1944.
- HAFNER, J.; HAFNER, P. Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, UK, n. 25, v. 12, p. 1509-1528, 2003.
- HESTENES D.; WELLS M.; SWACKHAMER G., *Force concept inventory, Physics*. Teacher. n. 30, v. 141, p. The Physics Teacher, Vol. 30, March 1992, p. 141-158
- KÆRGÅRD, N. Professor Georg Rasch (1901-1980) and Modern Econometrics, *Nordic Econometric Meeting*, 7, Bergen, p. 17-19, 2013.
- KENNEDY, C. A. Models and Tools for Drawing Inferences from Student Work. *American Education Research Association*, Montreal, Canada, 2005.
- LINACRE, J. M.; WRIGHT, B. D. *Winsteps: Rasch analysis for all two facet models*. Chicago: MESA, 2000.
- MASTERS, G. N. Continuity and Growth: Key Considerations in Educational Improvement and Accountability. *ACE and ACEL National Conference*, Perth, 2004.
- MEAD, R. A. Rasch Primer: The Measurement Theory of Georg Rasch Psychometrics services research memorandum 2008-001. Maple Grove, MN: Data Recognition Corporation, 2008.
- MORRIS G. A. et al. An item response curves analysis of the Force Concept Inventory, *Am. J. Phys*, USA, v. 80, p. 8-25, 2012.
- MOSKAL, B. M.; LEYDENS J. A. Scoring rubric development: validity and reliability. *Practical Assessment, Research & Evaluation*, USA, v. 7, n. 10, 2000.
- OLSEN, L. W. *Essays on Georg Rasch and his Contributions to Statistics*. 2003. 237p. Ph.d. (Dissertation Department of Economics) – University of Copenhagen. 2003.
- PLANINIC, M.; IVANJEK, L.; SUSAC, A. Rasch model based analysis of the Force Concept Inventory. *Physical Review Special Topics. Physics Education Research*, USA, v. 6, n. 1, p1-11, 2010.
- PRING, R. Editorial conclusion: a philosophical perspective. *Oxford Review of Education*, Oxford, Reino Unido (UK), v. 26, n. 3 e 4, p. 495-501, 2000.
- RASCH, G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Studies in Mathematical Psychology I. Danmarks pædagogiske Institut Copenhagen, 1960.
- RASCH, G. On specific objectivity: an attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, Dinamarca, v. 14, p. 58-94, 1977.
- RYAN, G.; BERNARD, H. Data Management and Analysis Methods. In: DENSZIN, N.; LINCOLN, Y. (ed.) *Handbook of qualitative research*, 2nd ed. Thousand Oaks, CA: Sage Publications, 2000. p. 769-802.
- ROBLYER, M. D.; EKHAML, L. How Interactive are YOUR Distance Courses? A Rubric for Assessing Interaction in Distance Learning. *The Online Journal of Distance Learning Administration*, University of West Georgia-USA, v. 3, n. 2, 2000.

- BEERER, K.; BODZIN, A. Promoting Inquiry-Based Science Instruction With the Science Teacher Inquiry Rubric (STIR). *Journal of Elementary Science Education*, v. 15, Issue 2, p. 39-49, fall 2003.
- SAMPSON, V.; CLARK, D. Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, Madison-USA, v. 92, n. 3, p. 447-472, 2008.
- SHAFFER, D. W.; SERLIN R. C. What Good are Statistics That don't Generalize? *Educational Researcher*, University of Wisconsin, Madison, v. 9, n. 33, p. 14-25, Dec. 2004.
- SINGER, J. D.; WILLETT, J. B. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Nova York: Oxford University Press, 2003. 644p.
- THOMAS, G. et al. *Numeracy item signature study*: A theoretically derived basis. Auckland, NZ: University of Auckland, Project asTTL, 2002.
- THURSTONE, L. L. Measurement of social attitudes. *Journal of Abnormal and Social Psychology*, Emory University- USA, v. 26, p. 249-269, 1931.
- TRIOLA, M. F. *Introdução à Estatística*. 10. ed. Rio de Janeiro: LTC, 2008. 696p.
- WRIGHT, B. D. Where do dimensions come from? *Popular Measurement*, University of Chicago-USA, v. 1, n. 1, p. 32, 1998.
- WRIGHT B. D.; LINACRE J. M. Observations are always ordinal; measurement, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, USA, v. 70, p. 857-860, 1989.

Data do recebimento: 05/05/2014

Data de aprovação: 13/10/2015

Data Versão Final: 12/11/2015

Contato:

Amanda Amantes Ribeiro

Rua Prof. Sabino Silva, 919, apto 501, Jardim Apipema, Salvador, BA, Brasil

Cep: 40155-250