

TEORIA DE RESPOSTA AO ITEM NA CONSISTÊNCIA DE TREINAMENTO EM ACADEMIAS DE GINÁSTICA

EVERTON, Adriana Nunes da Fonseca¹; SILVA, Luís Guilherme Pereira da²; BARRETO, Ana Cristina Lopes y Glória³; BRASIL, Roxana Macedo⁴; CARVALHO JUNIOR, Sergio⁵; BRITO, Diogo de Freitas^{3,6,7}; JUNIOR, Homero da Silva Nahum^{3,8}

188

Resumo

A investigação objetivou estimar um modelo de resposta ao item caracterizador da consistência de treinamento em de academia de ginástica. Utilizou-se um banco de dados com 593 pessoas das cidades de Rio de Janeiro, São Paulo e Belo Horizonte, tendo sido o Modelo de Crédito Parcial codificado em RStudio 2022. Constatou-se que os frequentadores tendiam a comportamentos extremos representados por baixa ou elevada consistência. Então, possível foi concluir que o objetivo foi adequadamente conquistado, dado que o modelo foi capaz a caracterizar a consistência de treinamento.

Palavras-chave: Planejamento. Inovação. Negócio. Cliente externo. Pesquisa de Mercado.

Abstract

This investigation aimed to estimate a response model to the item characterizing training consistency in a gym. A database of 593 people from the cities of Rio de Janeiro, São Paulo, and Belo Horizonte was used, and the Partial Credit Model was coded in RStudio 2022. It was found that gym-goers tended towards extreme behaviours represented by low or high consistency. Therefore, it was possible to conclude that the objective was adequately achieved, given that the model was able to characterize training consistency.

Keywords: Planning. Innovation. Business. External customer. Market research.

Introdução

A Teoria de Resposta ao Item (TRI) seira o conjunto de modelos matemáticos que relacionariam a probabilidade de determinada pessoa acertar um Item às características do indivíduo (traço latente) e do próprio item. Para tanto, consideraria que os itens seriam independentes das estimativas de habilidade (quantidade ou nível individual de traço

¹ Profissional de Educação Física da Adriana Nunes Consultoria de Corrida;

² Graduando do Curso de Educação Física do Centro Universitário Celso Lisboa;

³ Docentes do Curso de Educação Física do Centro Universitário Celso Lisboa;

⁴ Docente Ph.D. em Educação Física;

⁵ Pesquisador convidado Biodesa;

⁶ Docente do Curso de Gestão Desportiva e do Lazer do Centro Universitário Celso Lisboa;

⁷ Consultor Iceberg Business Academy;

⁸ Docente da Escola de Saúde da Universidade Cândido Mendes.

latente), assim como o seriam as características dos itens do grupo de respondentes, então, possível seria a comparação de indivíduos submetidos a diferentes conjuntos de itens (Araujo, Andrade e Bortolotti, 2009; Sartes e Souza-Formigoni, 2013). Portanto, extrapolaria a tradicional contagem de acertos.

Entretanto, o modelo requisitaria elevado número de respondentes ($n > 500$) para conquistar a estabilidade, e assumiria que 1) o único traço latente explicaria as respostas (unidimensionalidade); 2) dada a habilidade do indivíduo, as respostas aos itens seriam independentes (independência local); e 3) a função de resposta ao item seria monótona crescente. Por consequência, a complexidade computacional seria lugar comum, dado que as estimativas poderiam demandar métodos como I) *Expectation-Maximization* (EM) para estimação por máxima verossimilhança na existência de dados faltantes ou variáveis latentes (aquela oculta, não medida, mas explicativa de padrões nos dados), fornecendo solução eficiente e pontual; II) *Monte Carlo Markov Chain* (MCMC), baseado na cadeia de Markov, tratar-se-ia de conjunto de algoritmos adequados às distribuições complexas, possibilitando inferências bayesianas (combinação de distribuição a priori com as observações, gerando distribuição a posteriori). Isso ocorreria, porque, integrais intratáveis (aquelas carentes de soluções analíticas por funções elementares) surgiriam para estimar parâmetros de itens e habilidade, simultaneamente. O quadro apresentado exigiria o equilíbrio entre precisão e flexibilidade do modelo, em razão de recursos computacionais (Smith, 2000; Nunes e Primi, 2005; Smith e Plackner, 2009; Akour e Al Omari, 2013). As circunstâncias expostas tornariam a interpretação dos resultados demandadora de superiores esforços comparativamente à Teoria Clássica dos Testes (Pasquali e Primi, 2003; Primi, 2004; Sartes e Souza-Formigoni, 2013; Farias, 2013; Hauck Filho, Salvador-Silva e Teixeira, 2014; Lucian, 2017).

Originalmente, a TRI empregava o modelo Rasch para respostas dicotômicas, ou seja, na existência de duas Categorias como sim ou não, verdadeiro ou falso, e certo ou errado. Entretanto, o recorrente emprego de escalas Likert favoreceu o desenvolvimento do Modelo de Crédito Parcial (*Partial Credit Model* – PCM), cuja Função de Resposta ao Item, direta (I) ou parametrizada (II), permitiria estimar a probabilidade de ocorrência da Categoria k dada a habilidade θ , considerando a existência de $m + 1$ Categorias, como também a pontuação esperada (III) para determinado Item (Masters, 1982).

$$P_{ik}(\theta) = \frac{\exp [\sum_{j=0}^k (\theta - b_{ij})]}{\sum_{r=0}^m \exp [\sum_{j=0}^r (\theta - b_{ij})]} \quad (I)$$

onde: $P_{ik}(\theta)$: probabilidade de resposta na Categoria k dada habilidade θ ; b_{ij} : parâmetro de dificuldade de transição da Categoria $j-1$ para j ; $r = 0$: Categoria de base, comumente, aquela com pontuação mínima. Valendo a convenção se $k = 0$, então $\sum_{j=0}^0 (\theta - b_{ij}) = 0$, resultando em $\exp(0) = 1,00$.

$$P_{ik}(\theta) = \frac{\exp [\sum_{j=0}^k (\alpha_{ij} \cdot \theta - \beta_{ij})]}{\sum_{r=0}^m \exp [\sum_{j=0}^r (\alpha_{ij} \cdot \theta - \beta_{ij})]} \quad (II)$$

onde: α_{ij} : parâmetro de discriminação para a Categoria j do Item i ; β_{ij} : parâmetro de intercepto para a Categoria j do Item i .

$$E[X_i|\theta] = \sum_{k=0}^m k \cdot P_{ik}(\theta) \quad (III)$$

onde: $E[X_i|\theta]$: pontuação esperada para o Item i dado o nível de habilidade θ .

Em razão do exposto, as aplicações alcançariam testes adaptativos computadorizados, possibilitando selecionar dinamicamente itens com base na habilidade estimada (Labarrere, Da-Silva e Costa, 2011; Silva Filho, Santos e Cardoso, 2015; Fernandes *et al.*, 2020); a comparação de pontuação em distintos testes, equalização de testes (Nunes e Primi, 2005; Freires *et al.*, 2017; Reberti *et al.*, 2020); o desenvolvimento de bancos de itens calibrados (Kim, 2006; Heck, 2018; Klein e Ricarte, 2025); a comparação longitudinal de resultados (Barbosa, 2014; Silva, 2014; Moro, 2015); o funcionamento diferencial do item para identificação de viés (Cavalcanti *et al.*, 2016; Sá Júnior, 2019; Garcia, 2021); e as avaliações em larga escala como Enem (Silva, 2024), Pisa (Capelato e Cúri, 2023) e Saeb (Andrade, Brandão e Santos, 2021). Apesar disso, a Educação Física, nos diversos domínios, destacadamente, do desempenho, da saúde e dos negócios, carente seria de investigações específicas, então, objetivou-se estimar um modelo de resposta ao item para caracterizar a consistência de treinamento de frequentadores de academia de ginástica.

Metodologia

Empregou-se um banco de dados com 321 mulheres e 272 homens, moradores das cidades de Rio de Janeiro, São Paulo e Belo Horizonte. O grupo era composto por indivíduos nas faixas etárias Adulto Jovem (20 a 29 anos, n = 156 pessoas), Adulto (30 a 39 anos, n = 147 pessoas), Meia-idade (40 a 59 anos, n = 154 pessoas) e Idoso (60 anos em diante, n = 136 pessoas). Os clientes externos eram classificados, quanto ao tempo de prática, em Iniciante (até seis meses, n = 167 pessoas), intermediário (seis a 18 meses, n = 172 praticantes) ou Avançado (\geq 19 meses, n = 254 respondentes), cujo turno de treinamento era Manhã (6h00 – 12h00, n = 247 frequentadores), Tarde (12h01 – 18h00, n = 102 frequentadores) ou Noite (18h01 – 23h00, n = 244 frequentadores), estabelecida a identificação pelo registro de entrada.

As variáveis consideradas foram Sexo, Idade, Experiência, Turno e as respostas às 12 questões do formulário de consistência (Apêndice A). O conjunto de perguntas fazia referência às últimas oito semanas e considerou a seguinte escala de respostas: 0 – Nunca (não ocorreu no período), 1 – Raramente (ocorreu em menos de 25,00% das sessões), 2 – Às Vezes (ocorreu entre 25,00% e 50,00% das sessões), 3 – Frequentemente (ocorreu entre 50,00% e 75,00% das sessões) e 4 – Sempre (ocorreu entre 75,00% e 100,00% das sessões).

Optou-se pelo PCM por lidar com itens politômicos (Masters, 1982), perguntas com mais de duas opções de Categorias ordenadas (respostas), portanto, possibilitando capturar a complexidade dos itens, pela atribuição de parâmetros específicos para cada transição entre Categorias (*threshold*). Mais claramente, uma pergunta (item) com quatro opções de respostas, a probabilidade de um indivíduo escolher cada Categoria dependeria do grau de dificuldade e da localização da pergunta, assim como da habilidade do respondente (traço/habilidade latente). O modelo foi codificado em RStudio 2022.07.1+554 for Windows, utilizando os pacotes *readxl* 1.4.5, *eRm* 1.0-10, *psych* 2.5.6, *TAM* 4.3-25, *WrightMap* 1.4, *ggplot2* 3.5.1 e *reshape2* 1.4.4.

Resultados e Discussão

Excetuando-se Meia-idade e Noite, as variáveis de caracterização apresentaram elevada variabilidade na pontuação, Coeficiente de Variação $>$ 20,00%, então, as estatísticas adequadas à descrição eram Mediana e Coeficiente de Variação, isso poderia influenciar a Consistência de Treinos, fragilizando o ajuste do modelo, mesmo na ausência

de diferenças estatisticamente significativas, valor- $p > 0,05$, entre as classes em cada variável (Tabela 1). Essa característica convergiu à expectativa originada na literatura (Prado *et al.*, 2019; Corrêa e Dias, 2020; Jesus *et al.*, 2025).

Tabela 1: Resultados Descritivos (s: desvio padrão; md: mediana; cv: coeficiente de variação) de Pontos por Variável com Teste de Mann-Whitney* ou Kruskal-Wallis, $\alpha = 0,05$.

Variável	Média	s	md	cv	Inferência
Sexo*					0,63
Feminino	22,59	4,53	23,00	20,08	
Masculino	22,78	4,73	23,00	20,79	
Idade					0,74
Adulto Jovem	23,16	4,68	23,00	20,21	
Adulto	22,18	4,89	22,00	22,07	
Meia-idade	22,69	4,29	23,00	18,90	
Idoso	22,63	4,62	23,00	20,41	
Experiência					0,53
Iniciante	22,26	4,46	22,00	20,05	
Intermediário	22,49	4,75	23,00	21,11	
Avançado	23,06	4,63	23,00	20,08	
Turno					0,82
Manhã	22,76	4,68	23,00	20,59	
Tarde	22,24	4,81	23,00	21,65	
Noite	22,77	4,49	23,00	19,71	

Fonte: Os Autores (2026).

O modelo apresentou $\log\text{-likelihood} = -9242,65$, essa seria medida de qualidade do modelo aos dados (nível de verossimilhança), a princípio, quanto mais próximo de zero o valor, melhor seria o ajuste (Pratt, 1981; Pinheiro e Bates, 1995; Huggins-Manley e Algina, 2015; Wijayanto *et al.*, 2021). Todavia, isoladamente não seria interpretável, pois a serventia residiria na comparação de modelos. Realizadas foram 40 iterações para conquistar a convergência, quantitativo adequado e indicativo de estabilidade das estimativas. Essas incluiriam 47 parâmetros, por restrição de identificação, impondo que o número de parâmetros livres seria o total de parâmetros reduzido em um, ou seja, para 12 itens com cinco opções, haveria quatro transições ou parâmetros a serem estimadas para cada pergunta, então Total de Parâmetros = $12 \cdot 4 = 48$ e Parâmetros Livres = $48 - 1 = 47$.

Na Dificuldade das Categorias, a transição seria fácil, portanto, mesmo indivíduos com baixa habilidade conseguiria evoluir a resposta, se $\eta < 0,00$. Caso contrário, a transição

entre Categorias requisitaria habilidades elevadas. Além disso, convencionou-se que $0,00 < \eta < 1,00$ seria indicativo de exigência de habilidade moderada, e $\eta > 1,00$, demandaria elevada habilidade individual. Essa não foi exigida em qualquer das Categorias, independentemente do Item (Tabela 2). A transição Fácil se estabeleceu integralmente em Q2, Q3 e Q10, e parcialmente em Q4 (1, 2 e 3), Q6 (3 e 4), Q7 (1), Q9 (1, 3 e 4) e Q11 (1, 2 e 4). Enquanto que em todos os demais, a moderação foi identificada. Importante ressaltar que a estimativa de η indicaria se a transição entre Categorias se daria em níveis abaixo ($\eta < 0,00$) ou acima ($\eta > 0,00$) da média de habilidade. Outra análise relevante residiria no $IC_{95,00\%}$, porque informaria a significância da estimativa de η , quando o zero não pertencesse ao intervalo (extremos com o mesmo sinal algébrico), característica constatada em 28 Categorias, independentemente do nível de habilidade demandada.

Tabela 2: Estimativas para Dificuldade das Categorias (Cat): eta (η), erro padrão (ε) e $\alpha = 0,05$.

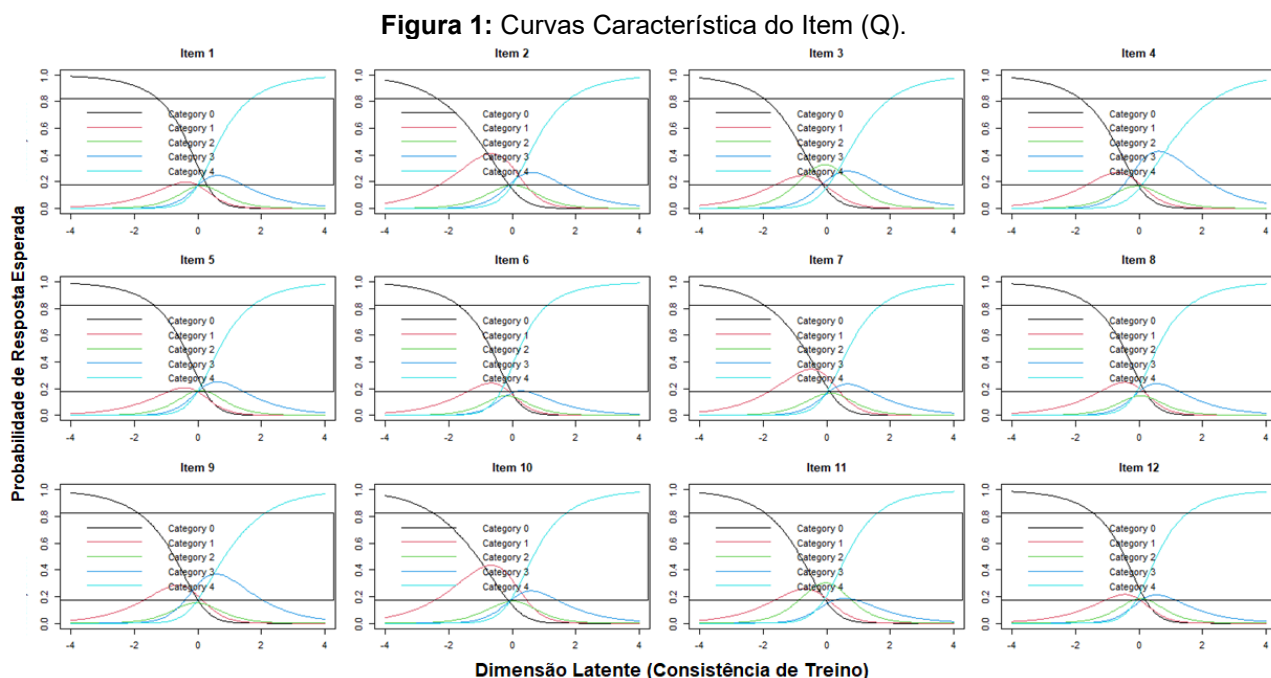
Cat	η	ε	$IC_{95,00\%}$	Cat	η	ε	$IC_{95,00\%}$	Cat	η	ε	$IC_{95,00\%}$
Q1 2	0,63	0,12	[0,39; 0,87]	Q5 1	0,47	0,12	[0,23; 0,70]	Q9 1	-0,17	0,12	[-0,42; 0,07]
3	0,62	0,13	[0,38; 0,87]	2	0,42	0,12	[0,18; 0,65]	2	0,13	0,14	[-0,14; 0,40]
4	0,58	0,13	[0,33; 0,83]	3	0,46	0,13	[0,22; 0,71]	3	-0,51	0,12	[-0,75; -0,28]
Q2 1	-0,76	0,12	[-0,99; -0,52]	4	0,45	0,13	[0,20; 0,70]	4	-0,01	0,14	[-0,28; 0,26]
2	-0,18	0,14	[-0,45; 0,09]	Q6 1	0,06	0,13	[-0,20; 0,32]	Q10 1	-0,88	0,12	[-1,12; -0,64]
3	-0,30	0,14	[-0,56; -0,03]	2	0,15	0,14	[-0,12; 0,43]	2	-0,21	0,14	[-0,48; 0,07]
4	-0,18	0,14	[-0,46; 0,10]	3	-0,02	0,14	[-0,29; 0,24]	3	-0,29	0,14	[-0,57; -0,02]
Q3 1	-0,17	0,14	[-0,44; 0,10]	4	-0,81	0,12	[-1,04; -0,58]	4	-0,36	0,14	[-0,64; -0,09]
2	-0,85	0,12	[-1,09; -0,61]	Q7 1	-0,31	0,11	[-0,52; -0,09]	Q11 1	-0,15	0,13	[-0,40; 0,11]
3	-0,40	0,13	[-0,66; -0,14]	2	0,26	0,13	[0,01; 0,52]	2	-0,64	0,12	[-0,87; -0,41]
4	-0,12	0,15	[-0,41; 0,16]	3	0,29	0,13	[0,03; 0,55]	3	0,09	0,14	[-0,19; 0,38]
Q4 1	-0,15	0,13	[-0,40; 0,11]	4	0,22	0,14	[-0,05; 0,48]	4	-0,27	0,13	[-0,53; 0,00]
2	-0,02	0,14	[-0,28; 0,25]	Q8 1	0,24	0,12	[0,01; 0,46]	Q12 1	0,39	0,12	[0,16; 0,62]
3	-0,72	0,12	[-0,95; -0,49]	2	0,59	0,13	[0,34; 0,85]	2	0,37	0,12	[0,14; 0,61]
4	0,09	0,15	[-0,20; 0,37]	3	0,39	0,13	[0,15; 0,64]	3	0,52	0,13	[0,27; 0,78]
				4	0,22	0,12	[-0,02; 0,46]	4	0,23	0,12	[-0,01; 0,47]

Fonte: Os Autores (2026).

Sumariante, transições fáceis ou não significativas indicariam que a mudança entre duas Categorias requisitaria níveis de habilidade irrelevantes ou demasiadamente próximos, por exemplo (Tabela 2), em Q2 (Realizo as sessões previstas sem adiar ou pular.) ocorreu $\eta_1 = -0,76$, portanto não seria possível diferenciar os praticantes que optaram por Nunca (marcaram 0) ou Raramente (marcaram 1). Em Q6, $\eta_2 = 0,15$ com $IC_{95,00\%} = [-0,12; 0,43]$, esse contendo o valor nulo, como efeito, estatisticamente a

estimativa (η_2) não seria diferente de zero. Tais ocorrências não determinariam a invalidação do formulário, mas sinalizariam a possibilidade de 1) redundância das Categorias envolvidas, e 2) funcionamento dicotômico do Item, independentemente da quantidade de opções de respostas. Tais características poderiam ser anuladas ou mitigadas com a remoção de Categorias detentoras de baixa ocorrência ou mudança nos descritores.

A Curva Característica do Item – ICC (Figura 1) apresentaria a probabilidade de um indivíduo atingir determinada Categoria (ou responder corretamente) em razão da função da dimensão latente, essa compreendida como o traço não observado diretamente, mas que o modelo objetivaria medir. Em geral, tratar-se-ia da habilidade que explicaria o padrão de respostas, ou seja, no caso estudado, a consistência de treino. Assim, haveria relação direta entre valor na escala latente e probabilidade de o indivíduo apresentar respostas indicativas de maior consistência. Aplicando no fenômeno avaliado, as curvas demonstraram como diferentes níveis de consistência de treino se relacionariam com a chance de responder em determinada Categoria de cada Item.



Fonte: Os Autores (2026).

O conjunto de indicadores de ajuste de Item apresentariam três nuances. O teste Qui-quadrado compara as respostas observadas e esperadas (pelo modelo), portanto, a ausência de significância (valor-p > 0,05) indicaria que os padrões de respostas reais e

preditas seriam compatíveis (sem discrepâncias significativas), por conseguinte, afastada estaria a hipótese de ajuste inadequado, conforme ocorrido (Tabela 3). Então, os itens seriam válidos dentro do modelo. O *Outfit MSQ* (*Mean Square Outfit*) avaliaria o ajuste, considerando repostas inesperadas distantes da habilidade estimada, assim, o Valor Esperado $\approx 1,00$. A ocorrência *Outfit MSQ* $< 0,70$ indicaria que o Item seria carente de informação ou pouco informativo, equivalendo a ser redundante, como consequência a predição dos resultados seria excessiva (“sobreajuste” ou “superajuste”). No extremo, *Outfit MSQ* $> 1,30$ informaria deficiência no ajuste, indicando a existência de ruído ou elevada variabilidade, logo, a discriminação do Item seria comprometida pela inconsistência. Essas possibilidades de interpretação se aplicariam ao *Infit MSQ* (*Mean Square Infit*), porém, tal métrica, consideraria repostas inesperadas, mas próximas da habilidade estimada. Por último, *Outfit-t* e *Infit-t*, como estatísticas padronizadas (z-score), indicariam a significância estatística das medidas, desde que pertencentes ao $[-2,00; 2,00]$, ratificando a aceitação do ajuste. Caso contrário, poderia indicar desvio do Item. No modelo proposto, todos os resultados indicaram o adequado ajuste, todavia, no Q7 houve desvio significativo do *Infit MSQ*, dado que *Infit-t* $< -2,35$, talvez a distinção das Categorias pelos respondentes não tenha ocorrido conforme a expectativa.

Tabela 3: Estimativas para Dificuldade dos Itens, $GI = 592,00$, $\alpha = 0,05$.

Item	χ^2	Valor-p	<i>Outfit MSQ</i>	<i>Infit MSQ</i>	<i>Outfit-t</i>	<i>Infit-t</i>
Q7	545,81	0,91	0,92	0,90	-1,77	-2,35
Q9	555,56	0,86	0,94	0,94	-1,55	-1,49
Q12	597,99	0,42	1,01	0,94	0,21	-1,44
Q10	565,83	0,77	0,95	0,95	-1,05	-1,20
Q8	569,15	0,74	0,96	0,96	-0,93	-0,99
Q5	562,39	0,80	0,95	0,96	-1,18	-0,94
Q1	583,18	0,59	0,98	0,97	-0,35	-0,82
Q4	574,97	0,68	0,97	0,97	-0,72	-0,78
Q11	584,45	0,58	0,99	0,98	-0,32	-0,51
Q6	605,68	0,34	1,02	1,02	0,52	0,43
Q2	607,54	0,32	1,03	1,01	0,57	0,13
Q3	604,36	0,35	1,02	0,99	0,44	-0,13

Fonte: Os Autores (2026).

Em síntese, os resultados estatísticos indicaram a possibilidade de explicar adequadamente as respostas. Característica necessária em qualquer pesquisa de mercado para a tomada de decisão de negócio ou pesquisa em saúde para caracterizar paciente ou

orientar ação profissional. Porém, não seria suficiente, pois não garantiria de forma plena a validade do formulário. Essa requisitaria avaliação de 1) representatividade do conteúdo, 2) interpretação das categorias pelos respondentes, 3) construto (abstração conceitual do fenômeno estudado), garantindo a exatidão entre aquilo que se objetivou medir e o que foi efetivamente medido (Silva, 2015; Gadenz *et al.*, 2019; Costa, 2020; Lanza, 2023; Menezes *et al.*, 2024; Ferreira, 2025; Messina e Ribeiro, 2025; Santos *et al.*, 2025; Oliveira e Garcia, 2025; Pascotini *et al.*, 2025; Torres e Santiago, 2025).

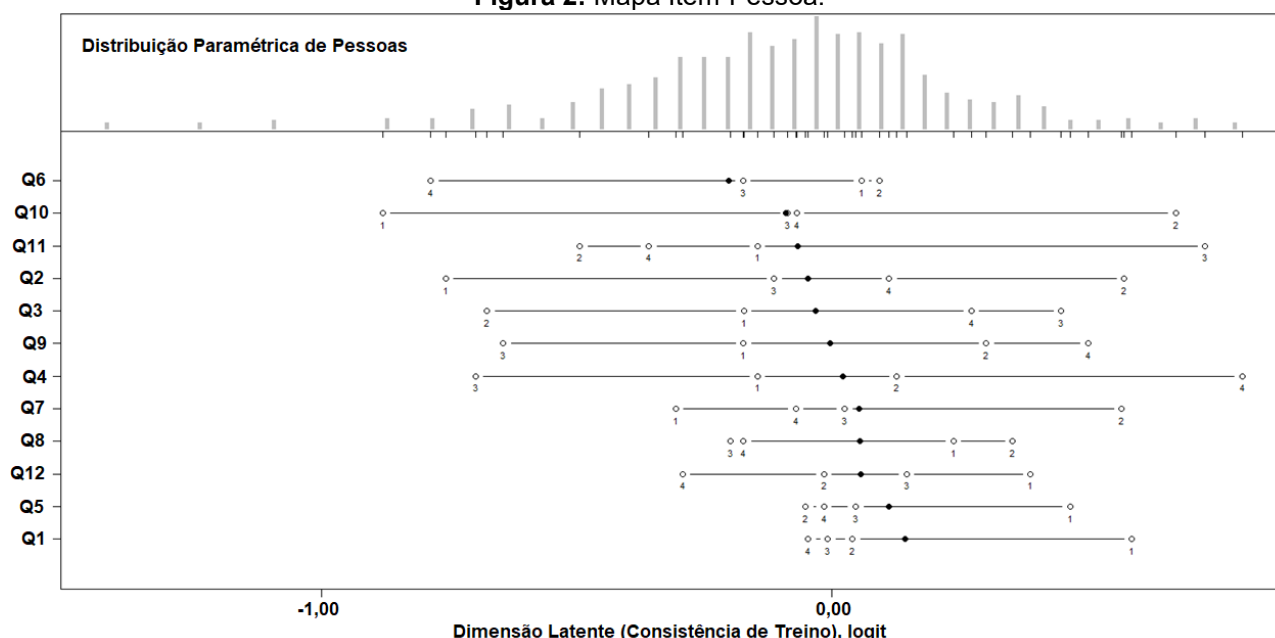
As análises de Categorias e Itens coadunadas revelariam que a consistência de treinos tenderia a ser dicotomizada, ou seja, possivelmente o comportamento dos clientes externos se posicionaria nas gradações extremas da escala, ignorando as sutilezas de avaliação, o que poderia ser uma peculiaridade do público frequentador de academia de ginástica (Ciampo *et al.*, 2010; Porto *et al.*, 2015; Prado *et al.*, 2019; Corrêa e Dias, 2020; Jesus *et al.*, 2025), ou seja, a concentração de determinados padrões de resposta seria a tendência realística. Cientificamente, esse comportamento seria relevante, pois indicaria que o fenômeno ora estudado teria tendência de distribuição dicotômica.

Essas considerações foram corroboradas pelo índice de confiabilidade (*Separation Reliability* = 0,3817 \approx 0,38), no qual valores próximos à unidade indicariam elevada capacidade de discriminação dos indivíduos em razão dos níveis de habilidade, mas, o modelo não foi capaz de distinguir os respondentes. Isso, porque, os participantes apresentaram respostas demasiadamente semelhantes, ou seja, a dispersão nos níveis de consistência de treino foi baixa, conforme refletiu a variância observada (*Observed Variance*) nas pontuações estimadas dos indivíduos ($s^2 = 0,0766 \approx 0,08$), a qual se aproximou da variância do erro do modelo (*Mean Square Measurement Error* = 0,0474 \approx 0,05), então a maior parte da variabilidade nas pontuações seria explicada pelo erro de medida, e não por diferenças reais entre os respondentes.

A compreensão da distribuição integrada de Indivíduos e Itens pela Dimensão Latente seria possível pelo Mapa de Wright (*Wright Map*) ou Mapa Item-Pessoa (*Item-Person Map*), nele o nível de habilidade seria representado em escala logit no eixo horizontal. No vertical, os Itens seriam apresentados decrescentemente em razão da dificuldade relativa (exigência da dimensão latente), assim o item mais acima seria o mais exigente quanto ao nível de habilidade para ser respondido em Categorias superiores (Frequentemente ou Sempre). À cada Item corresponderia um grupo de linhas horizontais com a distribuição dos respectivos pontos de transição entre as Categorias, essas

apresentadas em ordem crescente de probabilidade de escolha dominante, e representadas por pontos claros. A distância entre pontos consecutivos seria o *threshold*, então, longo comprimento significaria maior separação entre as Categorias, exigindo maior nível na Dimensão Latente para que ocorresse a transição (Rojas *et al.*, 2019 ; Farida *et al.*, 2024; Cessa *et al.*, 2025). Exemplificando, no Item Q6, a transição de Sempre (Categoria 4) para Frequentemente (Categoria 3) exigiria maior esforço na Consistência de Treino, comparativamente, à evolução de Raramente (Categoria 1) para Às Vezes (Categoria 2). Logo, a primeira transição teria maior capacidade de discriminação do nível de habilidade (Figura 2).

Figura 2: Mapa Item-Pessoa.



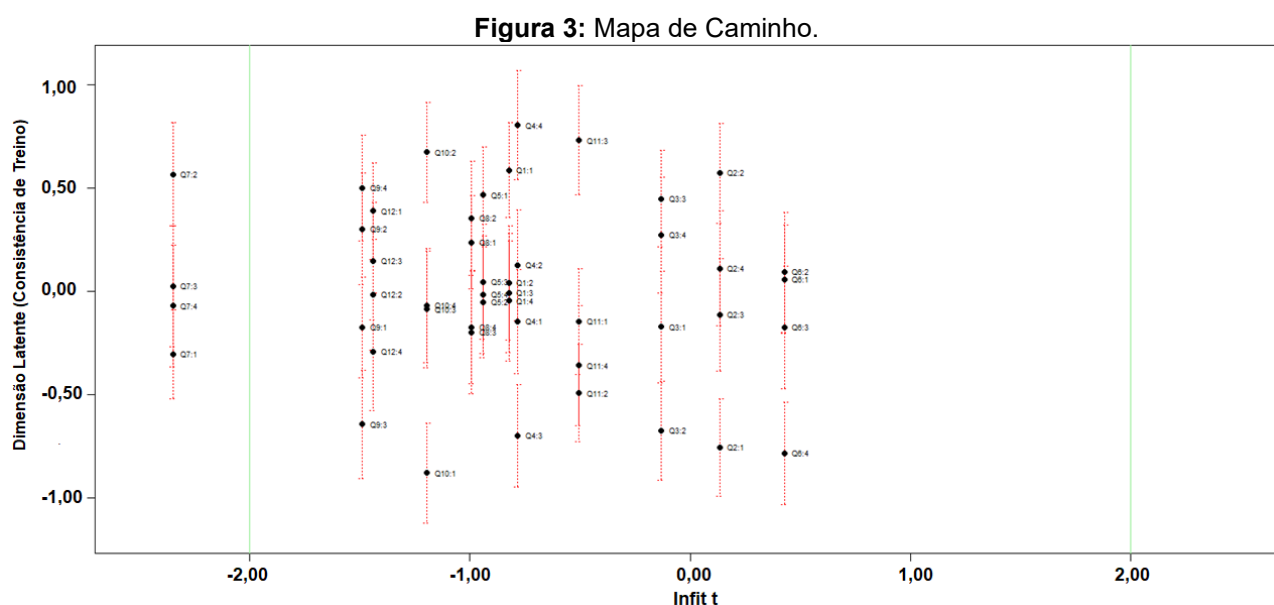
Fonte: Os Autores (2026).

Ainda nas linhas horizontais associadas aos Itens, os respectivos pontos escuros indicariam o *Threshold* estimado central, que seria o valor na escala latente de ocorrência da transição de probabilidade entre duas categorias, consequentemente, a distância entre esse e cada ponto claro indicaria o intervalo de confiança da estimativa (Gómiz-Aragón, Aragón-Méndez e Oliva-Martínez, 2024 ; Cessa *et al.*, 2025; Cessa^a *et al.*, 2025). Por exemplo, em Q10, o *Threshold* central seria aproximadamente coincidente com a Categoria 3, então a probabilidade de permanecer nela seria inferior a alcançar a Categoria 4. Situação que permitiria entender a Categoria 3 como zona de instabilidade probabilística. Portanto, se Frequentemente o indivíduo adotar o comportamento de não compensar

desordenadamente a perda de uma sessão de treino, a tendência seria de Sempre se comportar dessa forma (Figura 2).

A Distribuição Paramétrica de Pessoas permitiria identificar respondentes com baixos níveis de traço latente em relação à média (Berger *et al.*, 2025), pois 1) valores positivos indicariam indivíduos com níveis acima da média; 2) entre 0,00 e -1,00, pessoas pouco abaixo da média; 3) valores inferiores a -1,00, traço latente abaixo da média, mas dentro da distribuição esperada; e 4) abaixo de -3,00, indivíduos divergentes. Dentre os avaliados, havia pessoas abaixo de -1,00, ou seja, com baixa Consistência de Treino em relação ao grupo, porém sem comprometimento do instrumento ou modelo.

O Mapa de Caminho (*Pathway Map*) ilustraria mudança na probabilidade de transição em função da habilidade latente (Consistência do Treino), assim, possível seria verificar os níveis de habilidade nos quais determinada Categoria seria mais provável. No eixo horizontal residiriam as estimativas padronizadas de *infit-t* ($\bar{\mu} = 0,00; s = 1,00$), permitindo localizar o ajuste do Item ao modelo, e no vertical, o nível de habilidade padronizado, logo, ratificando o local de cada Categoria frente a Dimensão Latente (Figura 3).



Fonte: Os Autores (2026).

Essas informações replicariam as condições e a interpretação expostas na discussão da Tabela 3 (Sadjadi *et al.*, 2012; Rusch, Mair e Hatzinger, 2013; Gwathmey *et al.*, 2018; Sattelmayer *et al.*, 2020; Wang *et al.*, 2021). Dessarte, os Itens próximos ao

centro, à média ($x \approx 0,00$) seriam representativos da Consistência de Treino, pois o ajuste seria adequado, conforme ocorrido na maioria dos casos (Figura 3). Mas, Itens à direita, $x > 0,00$, indicariam comportamento não consistente com o modelo, o que poderia ocorrer por deficiências na formação ou existência de ambiguidade ou multidimensionalidade. Isso não foi constatado no modelo desenvolvido. Por outro lado, se $x < 0,00$, deslocamento à esquerda, o Item poderia ser redundante, não acrescentando variabilidade e favorecendo o excesso de previsibilidade, conforme ilustrado pelo Q7.

Considerações Finais

O objetivo foi modelar resposta ao item para caracterizar a consistência de treino de frequentadores de academia de ginástica. O modelo revelou comportamento tendencioso aos extremos da escala, ou seja, os respondentes detinham elevada ou baixa consistência. Por conseguinte, o objetivo foi adequadamente conquistado.

A avaliação deve ser replicada, porém com realização de ajustes na escala, o que pode favorecer a captura de nuances de respostas ou confirmar o comportamento polarizado dos respondentes. O modelo deve ser segmentado por dicotomia sexual, faixa etária, turno e experiência, utilizando a análise de Funcionamento Diferencial por Item (DIF), porque grupos distintos podem apresentar Itens com comportamento diferentes, mesmo com similar nível de habilidade. Investigar as correlações entre os Itens permitirá reduzir a dimensionalidade do fenômeno, o que pode favorecer o ajuste.

Referências

- AKOUR, M; AL OMARI, H. Empirical investigation of the stability of IRT item-parameters estimation. **International Online Journal of Educational Sciences**, v. 5, n. 2, p. 291–301, 2013.
- ANDRADE, WM; BRANDÃO, JC; SANTOS, MJC. Análise dos Parâmetros de Itens de Matemática à luz da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI). **Perspectivas da Educação Matemática**, v. 13, n. 32, p. 1-21, 2020. DOI: <https://doi.org/10.46312/pem.v13i32.9962>.
- ARAUJO, EAC; ANDRADE, DF; BORTOLOTTI, SLV. Teoria da resposta ao item. **Revista da Escola de Enfermagem da USP**, v. 43, n. especial, p. 1000-1008, 2009.
- BARBOSA, RJ. **Comparabilidade das informações disponíveis nos Censos (1960-2010) e PNADs (1976, 1985, 1995 e 2005)**. São Paulo: Centro de Estudos da Metrópole, 2014.
- BERGER, C *et al.* Design and analysis of a scale to measure pro-environmental behaviours among adolescents/Diseño y análisis de una escala de conductas proambientales para adolescentes. **PsyEcology**, v. 16, n. 2, p. 152-175, 2025.

CAPELATO, É; CÚRI, M. Proficiência em Letramento Financeiro: estudo a partir de uma escala e de um Teste Adaptativo Computadorizado baseados na Teoria de Resposta ao Item. **Bolema: Boletim de Educação Matemática**, v. 37, n. 77, p. 1063-1086, 2023.

CAVALCANTI, TM *et al.* Escala de Avaliação da Fadiga: funcionamento diferencial dos itens em regiões brasileiras. **Avaliação Psicológica: Interamerican Journal of Psychological Assessment**, v. 15, n. 1, p. 105-113, 2016.

CESSA, AG *et al.* Diseño y Validación de un Instrumento para Medir la Percepción de la Seguridad Industrial en PyMEs Manufactureras de Huatusco, Veracruz. **Ciencia Latina Revista Científica Multidisciplinar**, v. 9, n. 2, p. 4426-4446, 2025.

CESSA^a, AG *et al.* Diseño y Validación de un Instrumento para Medir la Cultura Organizacional en PyMEs Manufactureras de región Huatusco, Veracruz. **Ciencia Latina Revista Científica Multidisciplinar**, v. 9, n. 2, p. 6624-6647, 2025.

CIAMPO, LAD *et al.* Percepção corporal e atividade física em uma coorte de adultos jovens brasileiros. **Journal of Human Growth and Development**, v. 20, n. 3, p. 671-679, 2010.

CORRÊA, MA; DIAS, ACG. Avaliações autorreferentes e práticas esportivas: Comparações entre universitários praticantes e não praticantes. **Psicologia: Teoria e Prática**, v. 22, n. 2, p. 444-457, 2020.

COSTA, KP. **Construção e validação de formulário para o registro de enfermagem na parada cardiorrespiratória em paciente oncológico**. Dissertação (Mestrado Profissional em Práticas de Saúde e Educação) - Programa de Pós-Graduação em Saúde e Sociedade. Escola de Saúde. Universidade Federal do Rio Grande do Norte. Natal (RN), 2020.

FARIAS, ACOMO. **Aplicação da teoria de resposta ao item na avaliação educacional e seu processo de informatização**. Dissertação (Mestrado em Engenharia de Eletricidade) – pós-graduação em Engenharia de Eletricidade. Departamento de Engenharia de Eletricidade. Universidade Federal do Maranhão. São Luís (MA), 2013.

FARIDA, F *et al.* Validación de una Herramienta de Evaluación Basada en el Modelo Rasch para Medir la Resolución Creativa de Problemas en Estudiantes Mediante el Uso de TIC. **Pixel-Bit: Revista de medios y educación**, n. 71, p. 83-106, 2024.

FERNANDES, KP *et al.* Cross-cultural adaptation and validation of the universal Portuguese-version of the Pediatric Functional Assessment of Chronic Illness Therapy–Fatigue (pedsFACIT-F). **Jornal de Pediatria (Versão em Português)**, v. 96, n. 4, p. 456-463, 2020.

FERREIRA, SAC. **Consulta farmacêutica de pacientes em uso de terapia com antineoplásicos orais: construção e validação de formulário**. Dissertação (Mestrado Profissional em Práticas de Saúde e Educação) - Programa de Pós-Graduação em Saúde e Sociedade. Escola de Saúde. Universidade Federal do Rio Grande do Norte. Natal (RN), 2025.

FREIRES, LA *et al.* Ensino da avaliação psicológica no Norte brasileiro: analisando as ementas das disciplinas. **Avaliação Psicológica: Interamerican Journal of Psychological Assessment**, v. 16, n. 2, p. 205-214, 2017.

GADENZ, SD *et al.* Elaboração e validação de uma medida para avaliar o conhecimento de médicos de atenção primária do Brasil sobre recomendação nutricional para controle da hipertensão. **Cadernos Saúde Coletiva**, v. 27, n. 4, p. 404-411, 2019.

GARCIA, ML. **TRIDIF: um aplicativo para teoria de resposta ao item e funcionamento diferencial do item**. Trabalho de Conclusão de Curso (Bacharelado em Estatística) – Departamento de Estatística. Instituto de Matemática e Estatística. Universidade Federal do Rio Grande do Sul. Porto Alegre (RS), 2021.

GÓMIZ-ARAGÓN, M; ARAGÓN-MÉNDEZ, MDM; OLIVA-MARTÍNEZ, JM. Saberes del alumnado de secundaria sobre el modelo de inmunidad y vacunas. **Tecné, Episteme y Didaxis: TED**, n. 55, p. 100-116, 2024.

GWATHMEY, KG. *et al.* Validation of a simple disease-specific, quality-of-life measure for diabetic polyneuropathy: CAPPRI. **Neurology**, v. 90, n. 23, p. e2034-e2041, 2018.

HAUCK FILHO, N; SALVADOR-SILVA, R; TEIXEIRA, MA. Análise de teoria de resposta ao item de um instrumento breve de avaliação de comportamentos antissociais. **Psico**, v. 45, n. 1, p. 120-125, 2014.

HECK, MF. Sistema nacional de avaliação da educação básica (SAEB): compreendendo a realidade educacional. **Revista Eletrônica Científica Ensino Interdisciplinar**, v. 4, n. 11, p. 410-422, 2018.

HUGGINS-MANLEY, AC; ALGINA, J. The partial credit model and generalized partial credit model as constrained nominal response models, with applications in M Plus. **Structural Equation Modeling: A Multidisciplinary Journal**, v. 22, n. 2, p. 308-318, 2015.

JESUS, IV *et al.* Autoavaliação da saúde e prática de exercício físico: uma abordagem de idosos frequentadores de academias. **Brazilian Journal of Health Review**, v. 8, n. 2, e78842, 2025.

KIM, S. A comparative study of IRT fixed parameter calibration methods. **Journal of Educational Measurement**, v. 43, n. 1, p. 355-381, 2006.

KLEIN, R; RICARTE, TA. Calibración con parámetros fijos. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 33, n. 127, e0255144, 2025.

LABARRERE, JG; DA-SILVA, CQ; COSTA, DR. Testes adaptativos computadorizados. **Revista Brasileira de Biometria**, v. 29, n. 2, p. 229-261, 2011.

LANZA, VE. **Construção e validação de formulário para avaliação da indicação e manejo do cateter intravenoso periférico curto**. Dissertação (Mestrado em Enfermagem Fundamental) – Escola de Enfermagem de Ribeirão Preto. Universidade de São Paulo. Ribeirão Preto (SP), 2023.

LUCIAN, R. Uso da teoria de resposta ao item para determinar a confiabilidade de escalas. **PMKT–Revista Brasileira de Pesquisas de Marketing, Opinião e Mídia (PMKT on-line)**, v. 10, n. 1, p. 64-73, 2017.

MASTERS, GN. A Rasch model for partial credit scoring. **Psychometrika**, v. 47, n. 2, p. 149–174, 1982. DOI: 10.1007/BF02296272.

MENEZES, RM *et al.* Desenvolvimento e validação de questionário sobre o uso de antimicrobianos na atenção primária à saúde. **Revista de Epidemiologia e Controle de Infecção**, v. 14, n. 2, 2024. DOI: <https://doi.org/10.17058/reci.v14i2.19361>.

MESSINA, SR; RIBEIRO, MGL. Construção e validação de questionário para avaliar a saúde mental de professores que atendem crianças com Transtorno do Espectro Autista. **Educitec - Revista de Estudos e Pesquisas sobre Ensino Tecnológico**, v. 11, n. jan./dez., e258525, 2025.

MORO, KGC. **Avaliação externa de escolas e avaliação externa de aprendizagens no contexto institucional: um estudo paralelo sobre Brasil e Portugal**. Dissertação (Mestrado em Ciências da Educação) – Instituto de Educação. Universidade do Minho. Minho (Portugal), 2015.

NUNES, CHSS; PRIMI, R. Impacto do tamanho da amostra na calibração de itens e estimativa de escores por teoria de resposta ao item. **Avaliação Psicológica**, v. 4, n. 2, p. 141-153, 2005.

OLIVEIRA, T; GARCIA, R. Desenvolvimento e validação de tecnologia cuidativo-assistencial para qualificação de demanda de pacientes com diabetes tipo 2. **Metodologias e Aprendizado**, v. 8, p. 394-412, 2025. <https://orcid.org/0000-0001-9454-6810>

PASCOTINI, MR *et al.* Instrumento de Conhecimento, Atitudes e Práticas (CAP) para enfrentamento de epidemias e pandemias. **Saúde em Debate**, v. 49, n. 144, e9792, 2025.

PASQUALI, L; PRIMI, R. Fundamentos da teoria da resposta ao item: TRI. **Avaliação Psicológica: Interamerican Journal of Psychological Assessment**, v. 2, n. 2, p. 99-110, 2003.

PINHEIRO, JC; BATES, DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. **Journal of computational and Graphical Statistics**, v. 4, n. 1, p. 12-35, 1995.

PORTO, DB *et al.* Fatores associados à autoavaliação do peso corporal em mulheres praticantes de academia. **Revista Brasileira de Cineantropometria & Desempenho Humano**, v. 17, n. 2, p. 175-185, 2015.

PRADO, RCR *et al.* Exercício com intensidade autosselecionada para idosos: implicações do afeto em aulas comunitárias. **Revista Brasileira de Atividade Física & Saúde**, v. 24, p. 1-7, 2019. DOI: <https://doi.org/10.12820/rbafs.24e0089>

PRATT, JW. Concavity of the log likelihood. **Journal of the American Statistical Association**, v. 76, n. 373, p. 103-106, 1981.

PRIMI, R. Avanços na interpretação de escalas com a aplicação da Teoria de Resposta ao Item. **Avaliação Psicológica: Interamerican Journal of Psychological Assessment**, v. 3, n. 1, p. 53-58, 2004.

REBERTI, AG *et al.* Teste de Progresso na escola médica: uma revisão sistemática acerca da literatura. **Revista Brasileira de Educação Médica**, v. 44, n. 1, e014, 2020.

ROJAS, S *et al.* Análisis Rasch de la medida de cooperación del paciente ortodóncico en adolescentes de Medellín. **CES Psicología**, v. 12, n. 1, p. 43-53, 2019.

RUSCH, T; MAIR, P; HATZINGER, R. **Psychometrics with R: A review of CRAN packages for item response theory**. Discussion Paper Series. Vienna (Austria): Center for Empirical Research Methods, 2013.

SÁ JÚNIOR, AR. **Aplicação da teoria de resposta ao item na análise de diferença de gênero de sintomas depressivos na população universitária brasileira**. Tese (Doutorado em Ciências) – Departamento de Psiquiatria. Faculdade de Medicina. Universidade de São Paulo. São Paulo, 2019.

SADJADI, R *et al.* Psychometric evaluation of the myasthenia gravis composite using Rasch analysis. **Muscle & nerve**, v. 45, n. 6, p. 820-825, 2012.

SANTOS, TSS *et al.* Self-efficacy for weight loss: questionnaire and criteria validation to guide the care of individuals with obesity in the SUS. **Revista de Nutrição**, v. 38, e240058, 2025.

SARTES, LMA; SOUZA-FORMIGONI, MLO. Avanços na psicometria: da teoria clássica dos testes à teoria de resposta ao item. **Psicologia: Reflexão e Crítica**, v. 26, p. 241-250, 2013.

SATTELMAYER, KM *et al.* The assessment of procedural skills in physiotherapy education: a measurement study using the Rasch model. **Archives of physiotherapy**, v. 10, n. 9, 2020. <https://doi.org/10.1186/s40945-020-00080-0>

SILVA FILHO, AS; SANTOS, MAC; CARDOSO, RTN. Aplicação de testes adaptativos computadorizados em modelos de desdobramento graduado generalizados. **Blucher Mathematical Proceedings**, v. 1, n. 1, p. 766-778, 2015.

SILVA, AP. A incompatibilidade da Teoria de Resposta ao Item com processos seletivos públicos: estudo de caso do Exame Nacional do Ensino Médio. **Journal of Law and Regulation**, v. 10, n. 2, p. 220-250, 2024.

SILVA, FCB. **Construção e validação de instrumento de coleta de dados do casal grávido no âmbito da atenção básica**. Tese (Doutorado em Enfermagem na Atenção à Saúde) – Programa de Pós-graduação em Enfermagem. Departamento de Enfermagem. Universidade Federal do Rio Grande do Norte. Natal (RN), 2015.

SILVA, FG. **A inclusão de ciências naturais nos sistemas de avaliação da educação: o caso do avalia BH**. Dissertação (Mestrado Profissional em Gestão e Avaliação da Educação Pública) – Faculdade de Educação. Universidade Federal de Juiz de Fora. Juiz de Fora (MG), 2014.

SMITH, RM. Fit analysis in latent trait measurement models. **Journal of Applied Measurement**, v. 1, n. 2, p. 199–218, 2000.

SMITH, RM; PLACKNER, C. The family approach to assessing fit in Rasch measurement. **Journal of Applied Measurement**, v. 10, n. 4, p. 424-437, 2009.

TORRES, RV; SANTIAGO, LM. Autoavaliação da satisfação profissional com a consulta em medicina geral e familiar: a reflexão do Patient Satisfaction Questionnaire. **Revista Portuguesa de Medicina Geral e Familiar**, v. 41, n. 5, p. 386-95, 2025.

WANG, X *et al.* Supporting digitally enhanced learning through measurement in higher education: Development and validation of a university students' digital competence scale. **Journal of Computer Assisted Learning**, v. 37, n. 4, p. 1063-1076, 2021.

WIJAYANTO, F *et al.* Semi-automated Rasch analysis using in-plus-out-of-questionnaire log likelihood. **British Journal of Mathematical and Statistical Psychology**, v. 74, n. 2, p. 313-339, 2021.

APÊNDICE A: QUESTIONÁRIO DE CONSISTÊNCIA DE TREINO EM ACADEMIA

Objetivo: avaliar a regularidade e a aderência ao treino em academia nas últimas 6–8 semanas.

Tempo de resposta: 4–6 minutos.

Participante (Opcional): _____

Data: ___/___/___ **Idade (anos):** _____ **Sexo:** M F **Experiência em treino (anos):** _____

Escala de resposta

Código	Categoria	Descritor
0	Nunca	Não ocorreu nas últimas semanas
1	Raramente	Menos de 25% das sessões
2	Às vezes	25–50% das sessões
3	Frequentemente	50–75% das sessões
4	Sempre	75–100% das sessões

Nº Item	0	1	2	3	4
1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Total (0 – 48)					