

Escalas avaliativas e classificatórias em exames seletivos: uma abordagem utilizando a
Teoria da Resposta ao Item

Edilan de Sant Ana Quaresma¹

Carlos Tadeu dos Santos Dias²

Pedro Alberto Barbetta³

RESUMO:

O uso de provas elaboradas na forma de itens, em processos de avaliação para classificação, é uma herança história dos séculos XVI e XVII, ainda em uso nos dias atuais tanto na educação formal quanto em processos seletivos, a exemplo de exames vestibulares para admissão de estudantes em cursos de graduação no ensino superior. Elaboradas para mensurar conhecimentos, traços latentes que não podem ser medidos diretamente, as provas costumam ser corrigidas considerando unicamente o escore obtido pelo sujeito avaliado, sem considerar informações importantes relacionadas aos itens das mesmas. O presente trabalho teve como objetivos (i) utilizar a modelagem baseada na Teoria da Resposta ao Item (TRI) para construir escalas do conhecimento para a prova da Fundação Universitária para o Vestibular (FUVEST) e (ii) classificar os candidatos aos seis cursos de graduação oferecidos pela Escola Superior de Agricultura “Luiz de Queiroz”, unidade da Universidade de São Paulo, com base na escala construída. A hipótese utilizada no corpo do trabalho admitiu que o uso da TRI classifica de forma diferente os candidatos que os atuais métodos utilizados pela FUVEST. Foram utilizados padrões de respostas de 2326 candidatos submetidos à prova, para que uma análise unidimensional fosse utilizada, sob o enfoque da TRI, gerando uma escala de proficiências. Análises comparativas entre os critérios de classificação utilizados pela FUVEST e TRI foram realizadas, identificando discordância entre os mesmos. O trabalho apresenta propostas de interpretação pedagógica para as escalas unidimensional e indica a TRI como critério complementar para classificação dos candidatos, valorizando informações individuais dos itens e, portanto, utilizando uma avaliação classificatória mais abrangente.

Palavras chave: Classificação, escalas de proficiências, Teoria da Resposta ao Item, traços latentes

¹ Professor Associado da Universidade Federal do Oeste do Pará. (E-mail: edilan.quaresma@ufopa.edu.br)

² Professor Titular da Universidade de São Paulo – USP (E-mail: ctsdias@usp.br)

³ Professor Titular da Universidade Federal de Santa Catarina – UFSC (E-mail: pedro.barbetta@ufsc.br)

ABSTRACT:

The use of elaborate exams in the form of items, in evaluation procedures for classification, is a historical legacy of the 16th and 17th centuries, still in use today both in formal education and in selective cases such as entrance examinations. Designed to measure knowledge, latent trait that can not be measured directly, the exams are usually corrected considering only the score obtained by the subject, without including important information related to the items of it. This study aimed to: (i) use the modeling approach unidimensional item response theory (IRT), to build knowledge scales of the entrance examination Fundação Universitária para o Vestibular (FUVEST); (ii) classifying candidates for the 6 undergraduate courses offered by the "Luiz de Queiroz " College of Agriculture, unit of the University of São Paulo, based on the scale then. The hypothesis supposes that the use of IRT ranked candidates differently than current methods used by FUVEST. We used the patterns of responses of 2326 candidates submitted to the test, so that a one-dimensional analysis was performed under the IRT approach, generating a range of proficiencies. A proposal for classification of the candidates is presented, based on the weighted average of the individual proficiencies by the factor loadings diagnosed by modeling. Comparative analysis of the classification criteria used by FUVEST and IRT were performed by identifying discrepancies between them. This work presents the proposals of the pedagogical interpretation for one-dimensional and multidimensional scales and indicates the MIRT as additional criteria for the candidates, to valorize individual information of the items and therefore using a more comprehensive classification review.

Key words: Rank; Item response theory; Scale of proficiency

Introdução

A complexidade que envolve a qualidade da educação no ensino superior, com reflexos diretos na qualidade da gestão, perpassa pela integração de elementos que, em rede, contribuem coletivamente para o processo. Um desses elementos é a qualidade dos alunos ingressantes nos cursos de graduação, e no reconhecimento de suas potencialidades, aqui traduzidas como habilidades e competências.

Na sociedade moderna o processo de ingresso de alunos em Instituições de Ensino Superior vem sendo realizado por meio de processos seletivos, que utilizam como ferramenta de avaliação a Teoria Clássica dos Testes (TCT), baseada unicamente no total de questões respondidas corretamente pelos alunos, criando-se o que se conhece como escore do aluno.

Ainda em uso expansivo nos processos de avaliação, a TCT tem contribuído significativamente para o avanço do processo avaliativo, entretanto seu método tem mostrado sérias limitações na consideração individual dos itens que compõem um teste. Associado a essas limitações, tem-se percebido a necessidade de mensurar o conhecimento de alunos por meio da geração de uma medida de proficiência que viabilize também comparar alunos, ainda que submetidos a provas diferentes, estimar uma distribuição de proficiências de uma população e de subgrupos, além de possibilitar o monitoramento do progresso de um sistema educacional, visando a sua qualidade.

Estas limitações fomentaram a busca pelo desenvolvimento de uma teoria que viabilizasse a mensuração de variáveis latentes (habilidades ou proficiências), considerando informações importantes presentes nos itens de uma prova, e não somente o número de questões respondidas corretamente. Na década de 1950 surgiu uma teoria que resolvia as limitações apresentadas anteriormente, ficando conhecida com Teoria da Resposta ao Item (TRI).

A TRI é uma metodologia que surgiu no campo da psicometria, objetivando mensurar habilidades ou traços latentes de sujeitos submetidos a um instrumento de coleta de dados, como uma prova, por exemplo. Esta teoria sugere formas de representar a relação entre a probabilidade de um indivíduo responder corretamente um item, em função das habilidades do sujeito avaliado e de certas características do item, como a dificuldade, a discriminação e a probabilidade do sujeito acertar o item, sem que tenha habilidades para tal na área do conhecimento avaliada.

No Brasil a TRI vem sendo usada desde 1995 na análise dos dados do Sistema Nacional de Ensino Básico (SAEB), permitindo que o desempenho de alunos de séries diferentes pudessem ser comparados e colocados em uma escala única de conhecimento. Outro marco importante do uso da TRI no Brasil ocorreu a partir de 2009 com sua utilização no Exame Nacional do Ensino Médio (ENEM).

Desde a década de 1970 as universidades brasileiras vêm utilizando o vestibular como forma de ingresso dos alunos nos cursos de graduação. É o caso, por exemplo, do exame vestibular coordenado pela Fundação Universitária para o Vestibular (FUVEST), para candidatos aos seis cursos de graduação da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP).

O trabalho aqui apresentado objetivou utilizar a TRI como uma forma complementar de avaliação no exame vestibular da FUVEST, analisando os itens na abordagem da TCT e da TRI, além de propor escalas de proficiências pedagogicamente interpretáveis. O método de avaliação aqui proposto, considera informações importantes não contempladas atualmente no referido processo avaliativo, podendo ser utilizado como instrumento potencial em prol da qualidade do processo educacional no ensino superior brasileiro.

Referencial teórico

O termo avaliação é definido no contexto acadêmico, como o processo de apreciação e julgamento do rendimento acadêmico dos discentes, com o objetivo de acompanhar, diagnosticar e melhorar o processo de ensino e aprendizagem, bem como a habilitação do discente em cada componente curricular.

No início do Século XX a concepção de avaliação começou a incluir algumas complementações. Começou-se a questionar, para além dos objetivos da avaliação, o que ela de fato buscava mensurar e quais as dificuldades para tal. Essa percepção do objeto final de um processo avaliativo deu origem ao que hoje conhecemos como variável latente. Uma definição bem restritiva de variável latente esta associada àquela que não pode ser observada diretamente, necessitando de outras variáveis que podem ser observadas e que são utilizadas para determinar o seu valor (da variável latente).

Após a I Guerra Mundial, consolidou-se os procedimentos altamente formalizados sobre a teoria clássica dos testes (Gulliksen, 1950). Nesse período os pesquisadores começaram a perceber a importância do item como unidade de análise, principalmente após o trabalho de Loevinger (1947). Segundo Coimbra (2005), Loevinger percebeu que o teste deveria ser pensado como uma coleção de itens, todos medindo o mesmo traço latente geral, e assim, introduziu os conceitos de homogeneidade do teste e de escalonabilidade de itens.

Na década de 50 o físico George Rasch (1901-1980) desenvolveu pesquisas de novas técnicas para extrair a informação diretamente dos itens e em 1960, chegou aos resultados de que as características de itens e de pessoas podem ser medidas em uma mesma dimensão ou escala, e o encontro da pessoa com o item pode ser modelado por uma função aditiva, caracterizando a definição do primeiro modelo da Teoria da Resposta ao Item (TRI).

Andrade, Tavares e Valle (2000) definem a TRI como um conjunto de modelos matemáticos que procuram representar a probabilidade de um sujeito responder um certo item (pergunta, questão) como função dos parâmetros do item e da(s) habilidade(s) do respondente, de tal forma que, quanto maior a habilidade, maior a probabilidade do sujeito acertar aquele item. Dentre os modelos matemáticos propostos pela TRI para mensurar o traço latente de sujeitos, encontra-se o modelo logístico de três parâmetros, definido por:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, \quad (1)$$

com $i=1, \dots, I$, e $j=1, \dots, n$.

Em que, para n indivíduos submetidos a I itens:

1. U_{ij} : é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente o item i ;
2. θ_j : representa a proficiência (traço latente) do j -ésimo indivíduo;
3. b_i : é o parâmetro de dificuldade do item i , medido na mesma escala da habilidade θ_j , de tal forma que valores pequenos indicam itens com baixa dificuldade, e alternativamente, valores grandes indicam itens com alta dificuldade;
4. c_i : é o parâmetro do item que representa a probabilidade de indivíduos com habilidades extremamente baixas responderem corretamente o item i (muitas vezes referido como probabilidade de acerto casual);
5. $P(U_{ij} = 1|\theta_j)$: é a probabilidade condicional de um indivíduo j , com habilidade θ_j , responder corretamente o item i .

Desde o ano de 2009 o Ministério da Educação do Brasil vem utilizando a TRI no Exame Nacional do Ensino Médio (ENEM), por meio do modelo apresentado na

equação (1), inicialmente objetivando avaliar o aproveitamento dos alunos ao fim do terceiro ano do ensino médio para subsidiar reformas no sistema e comparar resultados entre diversos ciclos da educação formal. Seus resultados têm também sido utilizados por universidades como critério de classificação de candidatos aos seus cursos de graduação. No ano de 2013, 169 instituições de ensino superior, dentre faculdades, universidades, institutos federais e centros universitários, adotaram os resultados do ENEM, total ou parcialmente, como critério de ingresso nas instituições.

Percurso metodológico

Os dados utilizados na pesquisa aqui proposta são provenientes do exame vestibular coordenado pela Fundação Universitária para o Vestibular (FUVEST) e realizado no ano 2012, para candidatos a ingressar na Escola Superior de Agricultura “Luiz de Queiroz”(ESALQ), unidade acadêmica da Universidade de São Paulo (USP), em um dos seus seis cursos de graduação, num montante de 390 vagas, para o qual tiveram 2537 candidatos inscritos.

Nesta tese foram analisados os dados da primeira fase do exame vestibular 2012 para a ESALQ/USP. Como cada uma das 89 questões válidas do exame vestibular foi constituída por itens dicotômicos, em que as respostas dos candidatos aos itens foram do tipo certo ou errado, o modelo da TRI proposto para ser utilizado foi o de três parâmetros, definido pelo modelo de equação (1).

No processo de análise unidimensional, quando se utilizou o software BILOG, as respostas dicotomizadas dos candidatos foram disponibilizadas na extensão DAT, ao passo que, quando se utilizou o software R, os dados foram formatados em planilha eletrônica Excell, salvo com formato CSV.

Foi realizada uma análise dos itens utilizando a abordagem da teoria clássica dos testes, contemplando a consistência interna da prova para verificar a qualidade daquele instrumento na mensuração das proficiências dos candidatos, posteriormente o índice de dificuldade de cada item e, em seguida, o poder discriminativo do item que avalia a contribuição de um item na diferenciação de sujeitos. Posteriormente uma análise dos itens foi feita utilizando a abordagem da teoria da resposta ao item, identificando itens que efetivamente contribuíam com a mensuração das proficiências dos alunos, e viabilizando a construção de uma escala de habilidades/proficiências interpretadas pedagogicamente.

Resultados

A primeira ideia na análise dos dados sob a perspectiva da abordagem clássica, foi avaliar a qualidade do instrumento utilizado, no que diz respeito à precisão do mesmo, ou seja, à sua consistência interna. O procedimento clássico com esta finalidade é a análise do coeficiente alfa de Cronbach, que variam de 0,8094 a 0,8176, conforme se exclui determinados itens, mas quando se considera todos os itens o coeficiente obtido é 0,8152, indicando uma boa consistência interna do instrumento utilizado.

Estatísticas para a proporção de acerto dos itens identificam, para os 89 itens, média de acerto igual a 43; 73% com desvio padrão igual a 18; 52% revelando maior concentração de valores na cauda esquerda da curva formada pelas frequências dos dados (assimetria = 0,18) o que identifica, de forma geral, que o maior número de indivíduos se concentra abaixo da média de acertos.

Outro procedimento importante na análise exploratória de dados, dentro do enfoque da teoria clássica dos testes, diz respeito ao estudo da capacidade de discriminação do item em relação ao resultado do teste. Para este procedimento utiliza-se a correlação bisserial para verificar se uma determinada variável apresenta correlação significativa com o escore bruto produzido pelo conjunto dos itens. Esse passo é crucial para a escolha de itens que de fato apresentam consistência interna e se associam bem ao escore que será produzido.

Foram identificados os itens 11, 32, 52, 64, 85 e 87 com correlações negativas dos quais, os itens 85 e 64 apresentam respectivamente os maiores valores, em módulo. Estes resultados permitem concluir que tais itens não apresentam uma boa correlação com o escore bruto produzido pelo conjunto dos itens, não apresentando, portanto, boa discriminação para o índice que se pretende produzir, devendo ser excluídos de qualquer análise posterior.

Após a exclusão dos itens que pouco contribuíam com a medida de proficiência, objeto desse estudo, e dos itens com problemas nas estimativas dos parâmetros, foram selecionados 72 itens para mensurar as habilidades/proficiências dos candidatos.

Os valores estimados para os parâmetros dos itens viabilizaram obter as probabilidades de resposta correta, para diferentes níveis do traço latente, que puderam ser utilizadas para encontrar diferentes níveis de proficiência e, conseqüentemente, os

itens âncora e quase âncora. Essa escala de proficiência sintetiza o domínio dos conteúdos e habilidades alcançados pelos candidatos, o que permite inferir o nível de domínio das competências avaliadas.

Os níveis de traço latente permitem situar as capacidades dos sujeitos que responderam às questões da prova em quatro níveis de proficiência (I: abaixo do básico, II: básico, III: adequado e IV: avançado), definidos a partir das expectativas de aprendizagens (conteúdos, competências e habilidades) estabelecidos para o ensino médio, e que necessitam de avaliação pedagógica para serem interpretados.

Para a descrição do nível I, por exemplo, que reúne as proficiências mais simples esperadas de um candidato a ingressar na ESALQ, devem ser utilizados os itens quase âncora 47, 48, 54, 57, 74 e 78, que apresentam competências cognitivas semelhantes, cujos parâmetros são mostrados na Tabela 1.

Tabela 1: Itens âncora (A) e quase âncora (QA) por conteúdos e parâmetros

Nível (Escala)	Item	Tipo	Conteúdo	a	b	c	% acerto
I (0)	47	QA	Líng Portuguesa	1,220	-0,213	0,328	69,8
	48	QA	Líng Portuguesa	1,483	-0,086	-0,086	66,0
	54	QA	Inglês	1,156	-0,239	0,266	67,3
	57	QA	Inglês	1,392	-0,311	0,263	69,0
	74	QA	História	1,599	-0,108	0,249	64,9
	78	QA	História	1,301	-0,515	0,274	72,7
II (1)	13	QA	Biologia	1,332	0,528	0,258	53,4
	17	QA	Biologia	1,142	1,078	0,329	51,1
	24	QA	Química	1,012	1,117	0,335	52,1
	41	QA	Líng Portuguesa	0,841	0,693	0,227	51,7
	43	QA	Ling Portuguesa	1,432	0,547	0,240	51,5
	45	QA	Geografia	1,354	0,492	0,227	52,1
	70	QA	História	1,517	0,702	0,266	50,1
III (2)	19	A	Biologia	2,092	1,525	0,146	24,5
	9	QA	Geografia	1,155	1,636	0,180	32,4
	12	QA	Biologia	1,198	1,461	0,256	40,3
	18	QA	Biologia	1,197	1,341	0,160	34,2
	25	QA	Química	1,658	1,623	0,301	39,0
	26	QA	Química	1,215	1,581	0,251	38,3
	28	QA	Química	1,755	2,013	0,302	35,3
	50	QA	Ling Portuguesa	0,886	1,540	0,161	35,6
	75	QA	História	1,418	1,358	0,177	33,3
IV (3)	30	A	Química	1,426	2,960	0,218	24,1
	60	A	Matemática	1,708	2,704	0,072	9,7
	84	A	Física	1,638	2,520	0,114	14,8
	61	QA	Matemática	1,393	2,401	0,207	25,8
	63	QA	Matemática	1,243	2,428	0,132	19,8
	68	QA	Matemática	1,045	2,775	0,280	33,3

Uma proposta de interpretação pedagógica, utiliza as habilidades apresentadas nos Parâmetros Curriculares Nacionais para o Ensino Médio (2000), descritas como: (i) Linguagens, códigos e suas Tecnologias; (ii) Ciências da Natureza, Matemática e suas Tecnologias e (iii) Ciências Humanas e suas Tecnologias. Nessa perspectiva, a proposta de interpretação pedagógica é apresentada nos tópicos a seguir.

- **Nível I:** Este nível reúne as proficiências mais simples esperadas de um candidato a estudos de nível universitário, descritas como *“Analisar, interpretar e aplicar os recursos expressivos das linguagens, relacionando textos com seus contextos, mediante a natureza, função, organização e estrutura das manifestações de acordo com as condições de produção e recepção.”*;
- **Nível II:** Além da proficiência do Nível I, neste nível o candidato deve saber *“Apropriar-se dos conhecimentos da Química e da Biologia, e aplicar esses conhecimentos para explicar o funcionamento do mundo natural, planejar, executar e avaliar ações de intervenção na realidade natural.”*, além disso, deve também *“Compreender o desenvolvimento da sociedade como processo de ocupação de espaços físicos e as relações da vida humana com a paisagem, em seus desdobramentos político-sociais, culturais, econômicos e humanos.”*;
- **Nível III:** Além das proficiências dos Níveis I e II, neste nível o candidato deve *“Compreender a produção e o papel histórico das instituições sociais, políticas e econômicas, associando-as às práticas dos diferentes grupos e atores sociais, aos princípios que regulam a convivência em sociedade, aos direitos e deveres da cidadania, à justiça e à distribuição dos benefícios econômicos.”*;
- **Nível IV:** Além das proficiências dos Níveis I, II e III, neste nível o candidato deve *“Entender e aplicar métodos e procedimentos próprios das Ciências Naturais.”*

Pode-se observar que as interpretações pedagógicas dos níveis de proficiência são traduzidas como um processo em que o conhecimento é acumulativo, o que é justificável, uma vez que o modelo matemático proposto parte do princípio de que, quanto maior a proficiência do respondente, maior sua probabilidade de acerto. O traço latente do indivíduo representa a proficiência mínima que um respondente deve possuir para que sua probabilidade de acerto seja alta.

Atendido o primeiro objetivo da proposta do trabalho, partiu-se para o atendimento do segundo, que buscava classificar os sujeitos avaliados, com base nos pressupostos da Teoria da Resposta ao Item.

Tabela 2: Classificação de 11 alunos, com base na Teoria Clássica dos Testes e na Teoria da Resposta ao Item.

Candidato	Teoria Clássica dos Testes	Teoria da Resposta ao Item
50	67(1º)	2,1310(1º)
60	63(2º)	2,0942(2º)
35	55(7º)	1,7894(3º)
17	60(4º)	1,6385(4º)
41	62(3º)	1,6184(5º)
65	56(6º)	1,4291(6º)
77	58(5º)	1,3650(7º)
28	53(9º)	1,3090(8º)
4	52(11º)	1,2900(9º)
49	55(8º)	1,2883(10º)
79	48(18º)	1,2795(11º)

A Tabela 2 apresenta os resultados classificatórios de onze alunos submetidos ao processo seletivo vestibular, utilizando a teoria clássica dos testes e a teoria da resposta ao item. Lembrando que a Teoria Clássica dos Testes considera apenas os itens respondidos corretamente pelos candidatos, e que a Teoria da Resposta ao Item considera outros parâmetros como a dificuldade do item, a chance de acerto ao acaso e o poder de discriminação de cada item, é possível observar divergências nos processos classificatórios. Observa-se, por exemplo, que enquanto o candidato de número 35 obteve a 7ª colocação no processo seletivo, que utiliza a teoria clássica dos testes, o mesmo candidato passaria a ocupar a posição 3ª no ranking de classificação do certame.

A análise dos itens sob o enfoque unidimensional da TRI, feita nessa seção mostrou evidências consideradas importantes como, por exemplo, itens correlacionados negativamente com o escore obtido pelos candidatos, caracterizados provavelmente como itens mal formulados ou itens que medem um traço latente não considerado no referido processo de avaliação e ainda, itens prejudicados no processo de calibração pela ausência de candidatos melhor preparados para respondê-los.

Considerações finais

A avaliação do conhecimento é um tema bastante discutido por profissionais da educação, com o consenso de que o conhecimento é o referente teórico que deve dar sentido global ao seu processo (Álvarez Méndez, 2002). Historicamente sistematizada

na forma de exames e provas nos séculos XVI e XVII em colégios católicos da Ordem Jesuítica (Luckesi, 2003), a avaliação chegou aos dias atuais, como uma atividade desprovida da neutralidade do avaliador (Chureiri, 2008 e Sordi, 2001), e muitas vezes reduzida a uma medida, embora se saiba que “a avaliação não é uma medida pelo simples fato de que o avaliador não é um instrumento, e porque o que é avaliado não é um objeto no sentido imediato do termo”(Hadji, 2001).

A necessidade de medir algo que não pode ser mensurado diretamente, como o conhecimento, induziu a reflexões sobre a forma como processos avaliativos vinham sendo implementados, surgindo a teoria da resposta ao item como uma possibilidade concreta em avaliações de larga escala, por valorizar informações provenientes de cada item individualmente, e dessa forma, propor uma avaliação mais abrangente.

Neste trabalho o termo avaliação foi usado para situações que exigem classificação, como nos concursos e vestibulares, mas também pode ser uma ferramenta útil para o diagnóstico como recurso de acompanhamento, comparação e reorientação da aprendizagem, defendido por Luckesi (2003).

Os resultados mostraram que o processo atualmente utilizado para classificar alunos no exame vestibular da Fundação Universitária para o Vestibular (FUVEST) da Universidade de São Paulo (USP), nos cursos ofertados pela Escola Superior de Agricultura “Luiz de Queiroz”, ao não considerar informações importantes fornecidas pelos itens da prova (dificuldade, discriminação e acerto ao acaso), classifica os candidatos de forma menos robusta que aquele utilizado pela Teoria da Resposta ao Item (TRI). Adicionalmente, foi possível construir uma escala de proficiências dos candidatos, com interpretação pedagógica.

A extensa abrangência de possibilidades de abordar a qualidade da educação superior considera questões político-sociais, onde se enquadram a expansão de oportunidades educacionais, até a mensuração da aprendizagem dos alunos, sua formação para a cidadania e para o mundo do trabalho. Neste trabalho foi feita a opção de abordar as proficiências ou habilidades de alunos ingressantes nos cursos de graduação, não sendo demonstrada preocupação com o entrelaçamento de fragmentadas evidências da realidade.

Este trabalho apresenta possibilidades de reflexões acerca da qualidade da educação superior, entendida como um processo multirreferencial e complexo, oriunda da análise de uma das suas inúmeras variáveis influenciáveis, a saber, a qualidade do aluno ao ingressar na instituição. Defende-se esta variável como potencial instrumento para (re)direcionamento das escolhas dos cursos, em conformidade com as habilidades/proficiências demonstradas pelos alunos durante o processo seletivo para ingresso na instituição de ensino superior.

Referências bibliográficas

Gulliksen, H. (1950). *Theory of mental testes*. New York: Wiley, 487 p.

Loevinger, J. (1947). A sistemática approach to the construction and evaluation of testes of ability. *Psychological Monographs*. Iowa City, v. 61, n. 4, 49 p.

Coimbra, C.A.Q. (2005). Modelos não lineares em avaliação nas ciências sociais: estimação por aproximação estocástica uma MCMC frequentista. 131p. Tese (Doutorado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

Andrade, D.F., Tavares, H.R., Valle, R. C. (2000). *Teoria da resposta ao item: conceitos e aplicações*. São Paulo: SINAPE, 154p.

Álvarez Méndez, J.M. (2002). *Avaliar para conhecer: examinar para excluir*. Porto Alegre: Artmed, 144p.

Luckesi, C.C. (2003). *Avaliação da aprendizagem na escola: reelaborando conceitos e recriando a prática*. Salvador: Malabares Comunicação e Eventos, 87p.

Chureiri, M.S.F. (2008, jan-abr). Concepções sobre a avaliação escolar. *Estudos em avaliação educacional*. São Paulo, v.19, n.39, p.49-64

Sordi, M.R.L. (2001). Alternativas propositivas no campo da avaliação: por que não? In: Castanho S.E.M., Castanho, M.E. (orgs). *Temas e textos em metodologia do ensino superior*. Campinas: Papirus, 182p.

Hadji, C. (2001). *A avaliação desmistificada*. Porto Alegre: Artmed, 133p.