

UNIVERSIDADE JOSÉ DO ROSÁRIO VELLANO - UNIFENAS

Gustavo Eugênio Martins Marinho

**AVALIAÇÃO PSICOMÉTRICA DA PROVA DE TÍTULO DE ESPECIALISTA EM
CARDIOLOGIA DA SOCIEDADE BRASILEIRA DE CARDIOLOGIA**

Belo Horizonte

2020

Gustavo Eugênio Martins Marinho

**AVALIAÇÃO PSICOMÉTRICA DA PROVA DE TÍTULO DE ESPECIALISTA EM
CARDIOLOGIA DA SOCIEDADE BRASILEIRA DE CARDIOLOGIA**

Dissertação apresentada ao curso de Mestrado Profissional em Ensino em Saúde da Universidade José do Rosário Vellano, para obtenção do título de Mestre em Ensino em Saúde.

Orientador: José Maria Peixoto

Belo Horizonte

2020

Ficha catalográfica elaborada pela Biblioteca Itapoã UNIFENAS
Conforme os padrões do Código de Catalogação Anglo Americano (AACR2)

61-051:159.938(043.3)

M337a Marinho, Gustavo Eugênio Martins.

Avaliação psicométrica da prova de título de especialista em cardiologia da Sociedade Brasileira de Cardiologia. [manuscrito] / Gustavo Eugênio Martins Marinho. -- Belo Horizonte, 2020.

65 f. : il.

Dissertação (Mestrado) - Universidade José do Rosário Vellano, Programa de Pós-Graduação Stricto Sensu em Ensino em Saúde, 2020.

Orientador : Prof. José Maria Peixoto.

1. Residência Médica. 2. Especialização médica. 3. Título de Especialista. 4. Análise Psicométrica. 5. Teoria Clássica dos Testes. 6. Teoria de Resposta ao Item. I. Peixoto, José Maria. II. Título.

Bibliotecária responsável: Jéssica M. Queiroz CRB6/3254



Presidente da Fundação Mantenedora - FETA

Larissa Araújo Velano

Reitora

Maria do Rosário Velano

Vice-Reitora

Viviane Araújo Velano Cassis

Pró-Reitor Acadêmico

Mário Sérgio Oliveira Swerts

Pró-Reitora Administrativo-Financeira

Larissa Araújo Velano

Pró-Reitora de Planejamento e Desenvolvimento

Viviane Araújo Velano Cassis

Diretora de Pesquisa e Pós-graduação

Laura Helena Órfão

Coordenador do Curso de Mestrado Profissional em Ensino em Saúde

Antonio Carlos de Castro Toledo Jr.

Coordenadora Adjunta do Curso de Mestrado Profissional em Ensino em Saúde

Maria Aparecida Turci

Certificado de Aprovação

AVALIAÇÃO PSICOMÉTRICA DA PROVA DE TÍTULO DE ESPECIALISTA EM CARDIOLOGIA DA SOCIEDADE
BRASILEIRA DE CARDIOLOGIA

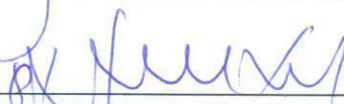
AUTOR: Gustavo Eugênio Martins Marinho

ORIENTADOR: Prof. Dr. José Maria Peixoto

Aprovado como parte das exigências para obtenção do Título de **Mestre Profissional em Ensino em Saúde** pela
Comissão Examinadora.



Prof. Dr. José Maria Peixoto



Prof. Dr. José Knopfholz



Prof. Dr. Marcus Vinicius Santos Andrade

Belo Horizonte, 02 de fevereiro de 2021.



Prof. Dr. Antonio Carlos de Castro Toledo Jr.

Coordenador do Mestrado Profissional

Em Ensino em Saúde

UNIFENAS

Dedico este trabalho a minha esposa e aos meus filhos.

AGRADECIMENTOS

A Deus, pela força e determinação que me guiaram até aqui.

À Marina, pelo apoio incondicional em tudo.

Aos meus filhos Samuel e Daniela, por compreenderem a minha ausência em momentos em que poderíamos estar juntos.

A minha mãe, pelo carinho constante, e ao meu pai, pelo apoio e herança na paixão pela cardiologia, baseada na ciência, na ética e no respeito ao próximo.

Ao meu orientador José Maria Peixoto, pela sabedoria em conduzir o meu trabalho.

Aos meus colegas de mestrado, em especial Igor, Ruy e Danielle, colegas inseparáveis dessa jornada.

Ao meu amigo e professor Carlos Alberto Castro, pela amizade e por sempre me incentivar na vida acadêmica.

Ao meu amigo Maurício de Rezende Barbosa, pela amizade incondicional e todo respeito profissional.

Agradeço a todos os membros da Comissão Julgadora do Título de Especialista, na figura de seu coordenador Marcus Vinicius Santos Andrade, e à Sociedade Brasileira de Cardiologia, por ter consentido o uso do banco de dados e por sempre apoiar a ciência.

RESUMO

Introdução: A certificação profissional está cada vez mais presente no currículo médico, seja por exigências do mercado, seja por demandas jurídicas. Após longa jornada inicial para a formação, o médico pode escolher entre dois caminhos para obter o título de especialista, uma residência ou uma especialização médica, desde que canceladas por entidades oficiais. Nesse último modelo, o médico precisa, após a conclusão da especialização, realizar uma prova para a obtenção do título de especialista. A Sociedade Brasileira de Cardiologia, com o auxílio da Comissão Julgadora do Título de Especialista em Cardiologia, promove, anualmente, uma prova com essa finalidade. Em relação à elaboração de itens, até o presente estudo, não havia sido feita nenhuma avaliação psicométrica da prova do Título de Especialista em Cardiologia (TEC), tendo em vista a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI). Essas teorias buscam, de forma técnica, avaliar a qualidade dos itens envolvidos em um teste e, de forma estatística, mostram um perfil do teste em relação a grau de dificuldade, poder de discriminação e possibilidade de acertos ao acaso (“chute”), o que nos permite uma análise crítica da qualidade da prova realizada. **Objetivos:** Avaliar as propriedades psicométricas da prova do TEC no ano de 2019, em relação aos parâmetros da TCT e da TRI. **Materiais e métodos:** Estudo observacional, com a análise psicométrica das 120 questões da prova realizada por 1120 (mil cento e vinte) candidatos para a obtenção do título de especialista em Cardiologia, no ano de 2019. **Resultados:** A análise pela TCT demonstrou que a prova apresenta boa distribuição dos itens em relação ao grau de dificuldade, com uma tendência para maior facilidade. Pela TRI, foi possível verificar que, das 118 questões avaliadas, apenas 49 (41,5%) possuem um fator satisfatório de discriminação. **Conclusão:** Os resultados deste estudo apresentam uma visão inédita da prova do TEC e, para as entidades oficiais que organizam a prova, poderão desencadear uma série de discussões e propostas para as futuras construções de provas no âmbito da cardiologia e das demais especialidades médicas.

Palavras-chave: Residência Médica. Especialização médica. Título de Especialista. Análise Psicométrica. Teoria Clássica dos Testes. Teoria de Resposta ao Item.

ABSTRACT

Introduction: Professional certification is increasingly present in the medical curriculum, either due to market demands or due to legal demands. After a long initial journey to training, doctors can choose between two paths to obtain the title of specialist, a residency program or a medical specialization, provided they are certified by official entities. In the latter, upon completing the specialization, the doctor must perform a test to obtain the title of specialist. The Brazilian Society of Cardiology, with the assistance of the Judging Panel on Cardiology Specialist Titles, promotes annually a test for this purpose. Regarding the elaboration of items, until the present study, there had been no psychometric evaluation of the test for the Title of Specialist in Cardiology (TSC), considering the Classical Test Theory (CTT) and the Item Response Theory (IRT). These theories seek, in a technical way, to evaluate the quality of the items involved in a test and, statistically, yield a profile of the test regarding the degree of difficulty, power of discrimination, and possibility of scoring due to guessing, which allows us to critically analyze the quality of the test performed. **Objectives:** To evaluate the psychometric properties of the TEC test in 2019, considering the CTT and IRT parameters. **Materials and methods:** Observational study, with the psychometric analysis of the 120 questions of the test performed by 1,120 (one thousand and one hundred and twenty) candidates to obtain the Title of Specialist in Cardiology, in the year 2019. **Results:** The analysis using CTT demonstrated that the test presents a good distribution of the items regarding the degree of difficulty, with a tendency towards ease. Through the IRT, it was possible to verify that, of the 118 questions evaluated, only 49 (41.5%) have a satisfactory discrimination parameter. **Conclusion:** The results of this study present an unprecedented view of the TEC test and, for the official entities that organize the test, they may trigger a series of discussions and proposals for the future elaboration of tests in the scope of Cardiology and other medical specialties.

Keywords: Medical Residency. Medical specialization. Specialist title. Psychometric Analysis. Classical Test Theory. Item Response Theory.

LISTA DE ILUSTRAÇÕES

Gráfico 1 -	Análise gráfica do item.....	22
Gráfico 2 -	Poder de discriminação.....	38
Gráfico 3 -	Dificuldade.....	38
Gráfico 4 -	Acerto ao acaso.....	39
Gráfico 5 -	Total da Nota - $T(\theta)$ - estimada pela TRI inicial, considerando-se 118 itens da prova, para cada candidato de acordo com sua aptidão (θ).....	43
Gráfico 6 -	Curva de Informação - $I(\theta)$ – e erro padrão da Curva de Informação gerados pela TRI inicial, considerando-se 118 itens da prova.....	44
Gráfico 7 -	Total da Nota - $T(\theta)$ - estimada pela TRI inicial, considerando-se 49 itens da prova, para cada candidato de acordo com sua aptidão (θ).....	44
Gráfico 8 -	Curva de Informação - $I(\theta)$ – e erro padrão da Curva de Informação gerados pela TRI inicial, considerando-se 49 itens da prova.....	45
Gráfico 9 -	Resultado da proficiência gerados pela TRI, considerando-se 49 itens da prova.....	49
Gráfico 10 -	Resultado da pontuação final gerada pela TRI.....	49

LISTA DE TABELAS

Tabela 1 -	Critérios de distribuição e classificação do grau de dificuldade do item pela TCT.....	20
Tabela 2 -	Critérios dos valores do poder de discriminação e classificação do item pela TCT.....	20
Tabela 3 -	Critérios dos valores do parâmetro de discriminação (“a”) e sua discriminação pela TRI.....	26
Tabela 4 -	Critérios de distribuição e classificação do grau de dificuldade (“b”) do item pela TRI.....	26
Tabela 5 -	Modelo da TRI – 3PL inicial, considerando-se 118 itens da prova.....	35
Tabela 6 -	Modelo da TRI – 3PL inicial, de acordo com motivos de retirada de itens.....	39
Tabela 7 -	Modelo da TRI – 3PL final, considerando-se 49 itens da prova.....	42
Tabela 8 -	Classificação dos parâmetros da TRI e TCT dos 118 itens da prova, no geral.	46
Tabela 9 -	Escore final (θ) e pontuação obtida pelos candidatos, baseando-se no modelo final da TRI com 49 itens, no geral.....	48
Tabela 10 -	Parâmetros de interesse dos modelos inicial e final da TRI utilizados para avaliação da precisão e validade.....	50
Tabela 11 -	Parâmetros da análise fatorial do modelo da TRI – 3PL inicial (118 Itens).....	53
Tabela 12 -	Parâmetros da análise fatorial do modelo da TRI – 3PL final (49 Itens).....	55

LISTA DE ABREVIATURAS, SÍMBOLOS E SIGLAS

AIC	<i>Akaike Information Criterion</i> (Critério de Informação Akaike)
AICc	<i>Akaike Information Criterion Correction</i> (Critério de Informação Akaike Corrigido)
AMB	Associação Médica Brasileira
CF	Carga Fatorial
CFI	Comparative Fit Index (Índice de ajuste comparativo)
CFM	Conselho Federal de Medicina
CJTEC	Comissão Julgadora do Título de Especialista em Cardiologia
CNRM	Conselho Nacional de Residência Médica
CRM	Regional de Medicina
ENADE	Exame Nacional de Desempenho dos estudantes
ENEM	Exame Nacional do Ensino Médio
gl	Grau de Liberdade
M2	Pacote estatístico R e Rstudio
MEC	Ministério da Educação e Cultura
OBMEP	Olimpíada Brasileira de Matemática das Escolas Públicas
PL	Parâmetro Logístico
RQE	Registro de Qualificação de Especialista
RMSEA	<i>Root Mean Square Error of Approximation</i> (Raio do erro quadrático médio aproximado)
SABIC	Medida de Avaliação de Ajuste Bayesiano
SBC	Sociedade Brasileira de Cardiologia
SRMSR	<i>Standardized Root Mean Square Residual</i> (Raiz quadrada média resíduo padronizada)
TCLE	Termo de Consentimento Livre e Esclarecido
TCUD	Termo de Consentimento de Utilização do Banco de dados
TEC	Título de Especialista em Cardiologia
TCT	Teoria Clássica dos Testes
TLI	<i>Tucker Lewis Index</i> (Índice Tucker Lewis)
TRI	Teoria de Resposta ao Item

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	Formação médica.....	14
1.2	Título de especialista.....	14
1.3	Sociedade Brasileira de Cardiologia.....	15
1.4	Avaliação psicométrica.....	18
1.5	Teoria Clássica dos Testes.....	19
1.6	Teoria de Resposta ao Item.....	23
2	JUSTIFICATIVA.....	29
3	OBJETIVOS	30
3.1	Objetivo geral	30
3.2	Objetivos específicos.....	30
4	MATERIAIS E MÉTODOS	31
4.1	Desenho do estudo	31
4.2	População/amostra.....	31
4.3	CrITÉrios de incluso.....	31
4.4	CrITÉrios de excluso.....	31
4.5	Amostra.....	31
4.6	Coleta de dados.....	32
4.7	Plano de anlise estatística.....	32
4.8	Aspectos éticos.....	32
4.9	Anlise estatística.....	33
5	RESULTADOS.....	35
6	DISCUSSO.....	57
7	CONCLUSO.....	62
	REFERNCIAS.....	63

1 INTRODUÇÃO

A formação dos médicos brasileiros implica em um período de seis anos de graduação em medicina, seguidos por mais alguns anos de treinamento supervisionado em Residência Médica ou Programa de Especialização Médica. A diversidade de especialidades médicas é enorme e contempla, atualmente, 55 especialidades e 59 áreas de atuação, de acordo com a resolução n. 2.162/2017, do Conselho Federal de Medicina-CFM. (CONSELHO FEDERAL DE MEDICINA, 2017).

Os programas de Residência Médica credenciados pelo Conselho Nacional de Residência Médica (CNRM) detém autonomia para a emissão do título de especialista, após a conclusão do curso. Já nos programas de especialização médica credenciados e conveniados pela Associação Médica Brasileira (ABM), para obtenção do título de especialista, é necessário a aprovação do candidato em uma prova organizada pela Sociedade Médica responsável pela especialidade em questão. Essa prova deve atender às normas constantes nos editais dos referidos concursos, que são homologados a cada edital liberado (ASSOCIAÇÃO MÉDICA BRASILEIRA, 2004).

A prova de Título de Especialista em Cardiologia (TEC) da Sociedade Brasileira de Cardiologia (SBC) é aplicada anualmente. A Comissão Julgadora do Título de Especialista em Cardiologia (CJTEC) é responsável pela coordenação dos trabalhos de aplicação do exame (planejamento, construção da matriz de objetivos, organização do item e aplicação). A cada ano, cerca de 1000 candidatos se inscrevem para a prova, uma vez que a aprovação no referido exame é um meio para a obtenção do título de especialista.

Há uma preocupação constante nas bancas de prova em relação às metodologias utilizadas na sua execução, bem como aos seus resultados. Isso tem levado a uma busca intensa por melhorias na qualidade e validação dos instrumentos utilizados para a elaboração dos itens das provas, para que essas reflitam, de forma mais real, a informação obtida com os resultados dos candidatos, tendo em vista os métodos utilizados para tais avaliações. Nesse sentido, de acordo com vários autores que estudam as referidas teorias, a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI) têm sido ferramentas importantes na análise e interpretação dos resultados, nos diversos cenários de provas e concursos. (VILARINHO, 2015).

1.1 Formação médica

A especialidade escolhida pelo médico poderá ser cursada em um programa de residência médica autorizado pelo Conselho Nacional de Residência Médica (CNRM) ou em um programa de especialização reconhecido pela Associação Médica Brasileira (AMB) (BRASIL, 1977).

No caso da residência médica credenciada pelo Ministério da Educação e Cultura (MEC), a programação teórica e prática a ser desenvolvida pelo médico residente varia de acordo com a área de especialidade pretendida, sendo definida por normas estabelecidas pelo CNRM que, junto às instituições responsáveis pela formação, fiscaliza e normatiza o funcionamento dos programas de residência médica no Brasil. Nesses programas, a carga horária de trabalho do médico residente, independente da especialidade, é de, no máximo, 60 horas semanais (BRASIL, 1981).

Após ter cumprido integralmente o programa de residência médica, o cursista receberá o certificado de conclusão do programa, conforme o artigo 2º da resolução do Conselho Nacional de Residência Médica (CNRM), n. 005/2003, de 23 de dezembro de 2003 (BRASIL, 2003).

1.2 Título de especialista

A busca pelo título de especialista tem se tornado uma constante dentre os médicos brasileiros, devido a questões que envolvem, desde o ganho de conhecimento, a esfera jurídica, a permissão para participar de concursos, até o ingresso em cooperativas médicas no mercado de trabalho. Dessa forma, a qualificação e a certificação por meio do título de especialista valoriza o profissional, aumentando seu prestígio e de sua especialidade.

A obtenção do título de especialista, historicamente, já passou por etapas distintas para a sua conquista. No passado, era concedido mediante a comprovação da experiência profissional do candidato, que buscava, no decorrer da sua carreira, certificados de cursos, congressos, estágios, trabalhos em clínicas e hospitais para obter tal titulação. Atualmente, com as mudanças técnicas e jurídicas no cenário profissional, é necessário cursar um programa credenciado pela CNRM ou obter a aprovação na prova de Título de Especialista da AMB para requerer a titulação.

Somente após a conclusão de uma das situações mencionadas, o médico terá o seu título e poderá solicitar o Registro de Qualificação de Especialista (RQE), junto ao Conselho Regional de Medicina (CRM), em determinada especialidade e/ou área de atuação, conforme estabelecido pelo Conselho Federal de Medicina (CONSELHO FEDERAL DE MEDICINA, 2018).

Cabe ressaltar que médicos que se apresentam como especialistas em determinada área e não possuem o RQE no CRM estão cometendo infração, conforme a resolução do Conselho Federal de Medicina (CFM), n. 2.217/2018, do Código de ética Médica que, em seu Art. 117, considera como ato infracional “Deixar de incluir, em anúncios profissionais de qualquer ordem, seu nome, seu número no Conselho Regional de Medicina, com o estado da Federação no qual foi inscrito e RQE quando anunciar a especialidade” (CONSELHO FEDERAL DE MEDICINA, 2018).

1.3 Sociedade Brasileira de Cardiologia

O crescimento e o destaque da Sociedade Brasileira de Cardiologia (SBC) no cenário nacional e internacional teve seu início oficial em 14 de agosto de 1943, quando então representada pelo Dr. Dante Pazzanese, que escreveu a primeira página de sua nobre história. Nos dias de hoje, conta com mais de 13.000 associados, que compõem a maior sociedade de cardiologia latino-americana, sendo subdividida em vários departamentos, cada qual com a sua relevância (SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2018).

O Título de Especialista em Cardiologia existe na Sociedade Brasileira de Cardiologia desde 1968, porém somente foi regulamentado pela AMB e pelo CFM por meio da Resolução n. 1286/89. Dessa forma, nos últimos anos, a SBC vem, por meio de diversos departamentos especializados e de grupos de estudos, aperfeiçoando seus métodos de avaliação. Nesse contexto, surge, em 1992, a CJTEC (SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2018).

Essa comissão é vinculada à diretoria da SBC, com a função de assessorá-la em assuntos referentes ao título de especialista, ao credenciamento de entidades formadoras e capacitadoras em cardiologia e ao curso nacional de reciclagem. É composta, atualmente, por 12 membros, todos detentores do TEC, com mandato de seis anos, sendo quatro representantes para cada

macrorregião (Sudeste; Sul; Centro-Oeste; Norte-Nordeste) e ainda pelo diretor científico da SBC, que atua como membro nato, de acordo com o estatuto. Por questões estatutárias, obrigatoriamente, quatro membros da comissão deverão ser substituídos a cada dois anos, em concordância com a nova gestão, que assume a diretoria. Cabe ressaltar que a composição constitui-se por cargos não remunerados e exercidos com elevados padrões éticos e científicos (SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2018).

A CJTEC, em conformidade com as normas da AMB, dentre outras funções, trabalha arduamente na organização e execução da prova para a obtenção do TEC (SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2018).

Com isso, a CJTEC analisou a necessidade de elaborar métodos e pré-requisitos para a obtenção do TEC, utilizando uma prova de proficiência cada vez mais bem fundamentada, devido às mudanças constantes no cenário da educação.

Em 2011, a CJTEC, com o intuito de uniformizar e dar maior qualidade à formação dos cardiologistas no Brasil, lançou a “I Diretriz da Sociedade Brasileira de Cardiologia sobre Processos e Competências para a Formação em Cardiologia no Brasil”, documento responsável por nortear os programas de Especialização em Cardiologia credenciados pela SBC, juntamente com a necessidade de padronização da prova de título (SOUSA *et al.*, 2011).

Dentre as várias decisões dessa diretriz, foram adotadas as normas do CFM para a obtenção do título de especialista, passando, então, a SBC a exigir que seus programas de formação em cardiologia seguissem as normas da Comissão Mista de Especialidades, formada pelo: CFM, AMB e CNRM, que preconiza que o tempo de formação na Cardiologia inclui como pré-requisito a Clínica Médica. É importante ressaltar que cursos teóricos de pós-graduação *lato sensu* servem apenas para reciclagem e podem ter outras finalidades, mas não são suficientes para formar e titular cardiologistas, sendo tacitamente rejeitados, também pelo CFM, como formadores (SOUSA *et al.*, 2011).

Em relação ao processo da prova, atualmente, após preenchidos os pré-requisitos do edital, o candidato terá, inicialmente, o seu currículo avaliado e pontuado, conforme preconizado pelo edital e, posteriormente, será submetido a uma avaliação teórica, com data e local determinados para tal.

A prova consta de 120 questões de múltipla escolha, com cinco alternativas cada, com apenas uma correta. É importante ressaltar que são aplicados modelos de provas diferentes, subdivididos em quatro subgrupos, com as mesmas questões, porém randomizadas de forma aleatória por um sistema eletrônico, para otimizar a segurança da sua aplicação, além de serem distribuídas de forma que os candidatos, no dia da prova, não fiquem próximos àqueles com o mesmo tipo de prova na sala.

As questões versam sobre temas da cardiologia em geral, que vão desde conhecimentos relativos à anatomia e à semiologia cardíaca ao diagnóstico e terapêutica das doenças cardiovasculares, com vistas à verificação da boa formação de um especialista em cardiologia nos mais variados assuntos dessa renomada especialidade. Em relação ao método de elaboração e confecção da prova, nos últimos anos, a CJTEC tem contratado uma empresa, que é a responsável desde a etapa da elaboração da prova, utilizando um banco de questões que são enviadas à CJTEC, até a organização e aplicação da prova em local preestabelecido.

No que tange às questões, elas são elaboradas por médicos cardiologistas titulares da SBC, que são convidados a elaborar algumas questões e recebem um valor previamente definido pelos itens enviados. O cardiologista que é convidado a elaborar as questões recebe algumas orientações em relação ao item a ser elaborado, tais como o número de alternativas de resposta para cada questão, a necessidade de não produzir enunciado muito longo, de fazer uso da referência bibliográfica prevista no edital, bem como de apresentar uma justificativa para o gabarito. No entanto, é desconhecido o perfil desses especialistas em relação ao conhecimento técnico específico em relação à elaboração de itens. Após a formação desse banco de questões, elas são avaliadas pela CJTEC que, em reuniões perante toda a comissão, analisa, julga e classifica os itens que serão selecionados para compor a prova.

Apesar da alta qualidade técnica daqueles que escolhem as questões, da importância da referida prova e da CJTEC trabalhar sempre buscando inovar e aperfeiçoar o nível das questões elaboradas, até o momento, nenhuma avaliação das características psicométricas dos itens da prova do TEC foi realizada. Em relação às questões selecionadas para a prova, a comissão classifica o grau de dificuldade que os itens poderão apresentar para os candidatos no dia da prova caracterizando-os como difíceis, médios e fáceis, no entanto, até o momento, essa classificação é feita de modo subjetivo, ou seja, de acordo com a opinião dos membros da CJTEC e não através de uma metodologia psicométrica, que avalia o grau de dificuldade para

quem faz a prova. Além disso, não é realizada nenhuma pré-testagem, por exemplo, das questões da prova do TEC, o que poderia trazer conhecimentos de parâmetros técnicos em relação a dificuldade, discriminação e acerto ao acaso dos itens. Portanto, importantes parâmetros psicométricos não são conhecidos no exame do TEC.

A partir de ações nesse sentido, melhorias de cunho pedagógico têm sido sugeridas e realizadas em diversas instituições, em relação à elaboração de provas, no intuito de alcançar resultados mais fidedignos aos tópicos propostos nos mais variados testes, aplicados em provas de qualquer natureza (VILARINHO, 2015).

Nesse contexto, a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI), utilizadas como avaliação psicométrica, têm sido ferramentas importantes na análise e na interpretação dos resultados nos diversos cenários de provas e concursos, de acordo com vários autores que estudam as referidas teorias. (VILARINHO, 2015).

1.4 Avaliação psicométrica

Segundo Bowman (1989), a utilização de testagem em larga escala, historicamente, é relatada desde a antiguidade, na China, há mais de 2200 anos, a exemplo da Dinastia Han (206 a.C.). No entanto, foi ao final do século XIX, na França, que a testagem psicológica moderna teve um impulso e começou a ganhar força e destaque nos meios acadêmicos. Dentre as diversas formas atuais de avaliação no campo de psicologia, uma área que vem apresentando destaque e ganhando cada vez mais interesse, principalmente na área da saúde, é a psicometria (HUTZ; BANDEIRA; TRENTINI, 2015).

Recentemente, diversos pesquisadores vêm lançando mão dos instrumentos psicométricos na área da saúde e, pela TCT e TRI, buscam obter avaliações mais robustas e fidedignas no que tange à qualidade dos testes educacionais aplicados e também com o intuito de corrigir os erros encontrados (SARTES; SOUZA-FORMIGONI, 2013).

Dessa forma, a psicometria busca encontrar respostas para uma pessoa ou um grupo, por meio de avaliações objetivas dos fenômenos psicológicos, e lança mão, para isso, de duas principais propriedades psicométricas importantes dos instrumentos: a validade e a confiabilidade (SARTES; SOUZA-FORMIGONI, 2013).

De acordo com Pasquali (2003) e Portney e Watkins (1993), o conceito de validade tem por base um instrumento com a sua capacidade de realmente medir aquilo a que foi proposto medir. E, conforme Armstrong, White e Saracci (1994), em relação a confiabilidade ou fidedignidade, seu conceito está relacionado à reprodutibilidade da medida, isto é, ao grau de concordância entre várias medidas de um mesmo sujeito inter e intra indivíduos (SARTES; SOUZA-FORMIGONI, 2013).

Para validar o processo, é necessário utilizar ferramentas que possam garantir certa confiabilidade. Para isso, o estudo da TCT e da TRI é de suma importância para a qualidade educacional das provas objetivas.

1.5 Teoria Clássica dos Testes

A Teoria Clássica dos Testes (TCT), de acordo com DeVellis (2006), utiliza-se de uma série de conceitos e técnicas como base para a formulação de vários instrumentos de medida, que são usados como referência nas medidas psicométricas. Dessa forma, a utilização de parâmetros descritivos contribui para a interpretação da distribuição das respostas de cada alternativa, ou seja, considera a prova como um todo e seus resultados são expressos no número total ou no percentual dos itens respondidos corretamente (SARTES; SOUZA-FORMIGONI, 2013). Assim, as propriedades psicométricas dos itens de uma prova, na TCT, relacionam-se aos seguintes parâmetros: índice de dificuldade; índice de discriminação e correlação ponto-bisserial.

No índice de dificuldade, o objetivo é calcular a proporção de acertos, ou seja, a razão entre o número de candidatos que responderam ao item de forma correta e o número total de candidatos submetidos ao item. Esse índice varia de 0 a 1 e, quando o resultado obtido é igual a 0 (zero), significa que ninguém respondeu ao item de forma correta. Por outro lado, quando o valor é 1, significa que todos acertaram o item. Sendo assim, quanto menor a porcentagem de acerto, maior será o grau de dificuldade. Com relação à classificação dos níveis de dificuldade dos itens e ao percentual de distribuição de itens por grau de dificuldade em uma prova, a TAB. 1 mostra, de forma simplificada, os critérios de distribuição e classificação que vêm sendo utilizados pela maioria dos autores (VILARINHO, 2015).

Tabela 1 - Critérios de distribuição e classificação do grau de dificuldade do item pela TCT

QUANTITATIVO IDEAL DE ITENS NA AVALIAÇÃO (% ESPERADO)	ÍNDICE DE DIFICULDADE DO ITEM	CLASSIFICAÇÃO DO ITEM EM RELAÇÃO AO ÍNDICE DE DIFICULDADE
10%	Superior a 0.9	Muito fáceis
20%	De 0.7 a 0.9	Fáceis
40%	De 0.3 a 0.7	Medianos
20%	De 0.1 a 0.3	Díficeis
10%	Até 0.1	Muito díficeis

Fonte: Vilarinho, 2015, p. 27.

Em relação ao índice de discriminação, trata-se de um parâmetro que permite diferenciar, para determinado item, os candidatos que obtiveram um maior número de acertos daqueles que obtiveram o pior desempenho. Ao efetuar esse cálculo, os candidatos serão alocados em três grupos: o grupo superior (27% dos candidatos com maiores pontuações), o grupo inferior (27% dos candidatos com menores pontuações) e os demais 46% dos candidatos, que compõem o grupo intermediário (VILARINHO, 2015).

Kelley, em 1939 (GREGOIRE; LAVEAULT, 2002), estabeleceu que o índice de discriminação pode ser definido pela diferença percentual entre os grupos de candidatos com maior e menor número de acertos e, dessa forma, quanto maior a diferença encontrada, maior será a o poder de discriminação do item elaborado. Nesse sentido, pode-se afirmar que um item com poder de discriminação $<0,20$ traduz-se como ineficiente e, por isso, deverá ser descartado. Enquanto uma discriminação $>0,40$ indica que estamos diante de um item de boa qualidade. Conforme a TAB. 2, pode-se classificar os itens de acordo com a sua discriminação (VILARINHO, 2015).

Tabela 2 - Critérios dos valores do poder de discriminação e classificação do item pela TCT

VALORES	CLASSIFICAÇÃO
Discriminação < 0.20	Item deficiente, deve ser rejeitado
$0.20 \leq$ Discriminação < 0.30	Item marginal, sujeito a reelaboração
$0.30 \leq$ Discriminação < 0.40	Item bom, mas sujeito a aprimoramento
Discriminação ≥ 0.40	Item bom

Fonte: Vilarinho, 2015, p 28.

Outro parâmetro utilizado para avaliar o poder de discriminação, na TCT, é o coeficiente de correlação ponto-bisserial, que diz respeito a uma medida que compara a nota total dos candidatos no teste com a nota total dos candidatos que acertaram um item específico no teste, ou seja, o número de acertos do item no teste.

Por esse coeficiente, é possível indicar, por exemplo, dentre os candidatos com maior números de acertos, os itens respondidos de forma errada, identificando, dessa forma, um item com baixo poder de discriminação. Com uma variação no intervalo $[-1, 1]$, é esperado que o coeficiente de correlação ponto-bisserial apresente coeficiente $>0,30$, pois valores negativos ou próximos de zero mostram que candidatos com um bom desempenho total no teste estão marcando respostas incorretas. Essa análise permite concluir que o item não está exercendo a função de distinguir candidatos com uma grande habilidade daqueles com baixa habilidade (GOMES, 2012).

$$\rho_{pb} = \frac{S_p - S}{\sigma} \sqrt{\frac{p}{1-p}}$$

Em que: S_p é a nota média no teste, para os indivíduos que acertaram o item;
 S é a nota média no teste, para todos os indivíduos;
 σ é o desvio-padrão das notas obtidas no teste por todos os indivíduos;
 p é a proporção de acertos no item.

O ideal, nesse parâmetro da TCT, é que a opção correta de resposta apresente o coeficiente ponto-bisserial positivo (média dos indivíduos que acertaram o item maior que a média de todos os indivíduos que responderam à prova) e as opções restantes apresentem coeficientes negativos (acontecendo o inverso do caso anterior).

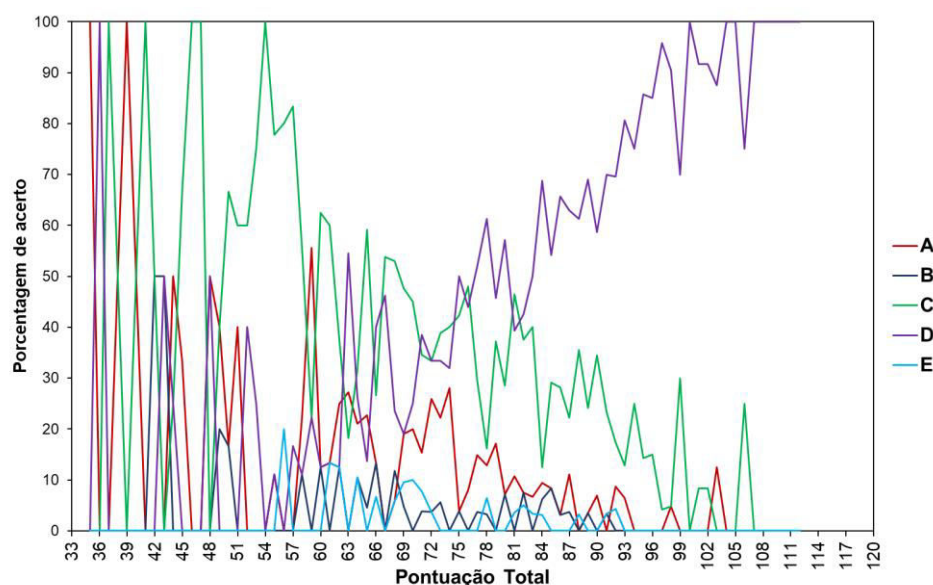
A importância prática desse coeficiente diz respeito ao que pode ser feito com determinado item, quando uma alternativa correta exibir um coeficiente negativo e/ou alguma alternativa errada tiver o coeficiente positivo. Nesse caso, há algum problema na construção do item, deixando evidente que ele não apresenta uma discriminação adequada, devendo ser rejeitado da prova ou, no mínimo, reformulado.

Em relação à composição gráfica na TCT, é possível fazer a Análise Gráfica do Item (AGI), por meio de um gráfico de linhas onde é realizado um esboço da proporção de respostas

efetuadas em cada opção de um determinado item em relação ao desempenho total dos candidatos que fizeram o teste (GOMES, 2012).

Do ponto de vista prático, conforme ilustrado no GRAF. 1, no eixo horizontal, tem-se as notas crescentes referentes ao teste respondido pelos candidatos, enquanto no eixo vertical, tem-se cinco linhas representando cada alternativa da questão. Conforme o percentual total de notas aumenta, o percentual de linhas das respostas escolhidas erradas irá diminuir, ficando evidente que a linha da resposta correta aumenta conforme o aumento das notas.

Gráfico 1 - Análise gráfica do item



Na busca por métodos para o aperfeiçoamento da confecção de uma prova, ao utilizar a análise das provas pela TCT, chamada também de análise clássica, utilizam-se métodos estatísticos que podem contribuir para essa finalidade e alcançar avaliações mais coerentes e fidedignas ao tema proposto. O objetivo é trazer aos candidatos testes com níveis superiores de excelência, se comparados a uma avaliação sem rigor científico e técnico que, ao final, pode não estar discriminando os bons candidatos daqueles com conhecimento abaixo do esperado.

Com o aprofundamento dos estudos na área da psicometria, tendo em vista a análise clássica dos itens, surgiram dificuldades e limitações em relação a essa técnica, com destaque para a estimação dos parâmetros dos itens e ao escore dos indivíduos, considerando situações nas quais o escore do sujeito depende e varia de acordo com a amostra utilizada no cálculo. Além disso, a TCT não permite, por exemplo, comparar candidatos que tenham realizado provas

diferentes e fazer um acompanhamento de um mesmo candidato (HUTZ; BANDEIRA; TRENTINI, 2015).

Essas situações levaram à busca de alternativas para responder, de forma mais completa e robusta, a essas novas demandas no campo da psicometria. Assim, a evolução dos princípios estatísticos da TCT permitiu o desenvolvimento da Teoria de Resposta ao Item (TRI).

1.6 Teoria de Resposta ao Item

Considerada pelos psicometristas e pesquisadores muito mais um modelo matemático do que uma teoria, a TRI surge no contexto das pesquisas não para substituir, mas para complementar informações não respondidas e transpor as limitações da TCT. Sustentada pela evolução de *softwares* na área de cálculos, permite uma gama maior de análises estatísticas para itens e escalas. Com essa evolução, os itens até então avaliados de forma geral nos testes passam a ser analisados de forma individual, não dependendo, então, da amostra como um todo, superando, dessa forma, uma importante limitação da TCT (HUTZ; BANDEIRA; TRENTINI, 2015).

Inclui-se, nesse novo contexto da TRI, a habilidade ou proficiência do indivíduo, denominada como traço latente individual, ou seja, toda probabilidade de resposta a determinado item inerente às características do sujeito serão representadas por esse traço, identificado nas fórmulas pela letra *theta* (θ). Esse modelo matemático relaciona variáveis observáveis (itens de um teste) e traços hipotéticos não observáveis ou aptidões. Conforme o nível *Theta* identificado em um indivíduo, que pode variar entre -4 e + 4, é possível, por exemplo, prever se ele irá responder de forma positiva ou negativa a determinado item, uma vez estabelecida uma probabilidade para cada item avaliado (SARTES; SOUZA-FORMIGONI, 2013).

Dessa forma, a TRI procura medir variáveis não observáveis (traço latente) que possam influenciar as respostas dadas aos itens, utilizando a aferição das variáveis observáveis (respostas aos itens). Portanto, estabelece uma relação entre a habilidade (proficiência) do respondente e os parâmetros do item com a probabilidade de acerto no item, de tal forma que quanto maior a proficiência de um candidato, maior é a sua probabilidade de responder corretamente ao item proposto no instrumento de avaliação.

Segundo Gomes (2012), “em relação à probabilidade, os parâmetros a serem analisados em cada item serão enquadrados em funções matemáticas tipicamente logísticas por *softwares*, conforme modelo proposto abaixo”.

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

com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, n$, onde:

- a) U_{ij} é uma variável dicotômica, que assume os valores 1, quando o indivíduo j responde corretamente ao item i ; ou 0, quando o indivíduo j não responde corretamente ao item i .
- b) θ_j representa a habilidade (traço latente) do j -ésimo indivíduo.
- c) $P(U_{ij} = 1|\theta_j)$ é a probabilidade de um indivíduo j , com habilidade θ_j , responder corretamente ao item i e é chamada de Função de Resposta do Item – FRI.
- d) b_i é o parâmetro de dificuldade (ou de posição) do item i , medido na mesma escala da habilidade.
- e) a_i é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional à inclinação da Curva Característica do Item — CCI no ponto b_i .
- f) c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente ao item i (muitas vezes referido como a probabilidade de acerto casual).
- g) D é um fator de escala, constante e igual a 1.

É importante ressaltar que os valores dos parâmetros a , b e c são calculados utilizando pré-testagens (calibragem dos itens), pelo método da máxima verossimilhança:

$$L(u_{1s}, u_{2s}, \dots, u_{ns} | \theta) = \prod_{i=1}^n P_i(\theta_s)^{u_{si}} Q_i(\theta_s)^{1-u_{si}}$$

A máxima verossimilhança (L) trabalha com as derivadas. Sua fórmula é a seguinte (L vem de *likelihood*, que significa verossimilhança):

Em que: $i = 1, 2, \dots, n$ itens

u_{is} = resposta do sujeito a cada item (1 = acertou, 0 = errou)

Baseado no cálculo acima, obtém-se a probabilidade, ou melhor, a verossimilhança de se obter um padrão de respostas a uma série de n itens ($u_{1s}, u_{2s}, \dots, u_{ns}$), levando-se em conta a aptidão/proficiência dos sujeitos (θ). Consiste simplesmente no produtório (Π) das probabilidades de acerto [$P_i(\theta)$] e pelas probabilidades de erro [$Q_i(\theta)$] de cada item individualmente.

Em relação ao cálculo da aptidão/proficiência de um candidato, é preciso achar o máximo da função acima. Para tanto, inicialmente, é preciso descobrir a probabilidade de acerto $[(P_i(\theta))]$ dos itens do teste de modo individual, por um dos três modelos da TRI (1LP, 2LP, 3LP. Após, de forma empírica, substitui-se os valores de θ , numa faixa de -5 a $+5$ ($-5,00 \leq \theta \leq +5,00$). Normalmente, utiliza-se $-3,00 \leq \theta \leq +3,00$) ou emprega-se o algoritmo de iteração de Newton-Raphson, para calcular o máximo da função de máxima verossimilhança (L). Esse máximo, com base nos valores de θ , é a aptidão/proficiência do candidato obtida no teste realizado.

Após a determinação do valor de habilidade (θ) de cada indivíduo, existe, pela TRI, a possibilidade de prever o desempenho de um candidato, tendo em vista os itens presentes em um teste a ser respondido, apontando a chance de ele acertar ou errar cada item.

A partir desse modelo, utilizando-se o nível de habilidade do indivíduo, bem como o grau de complexidade do item a ser analisado, poderão ser estimados três importantes parâmetros: discriminação, dificuldade e acerto ao acaso (HUTZ; BANDEIRA; TRENTINI, 2015).

O modelo baseado nos três parâmetros acima tem sido utilizado para avaliações psicométricas de itens, sendo adotado neste estudo. Em primeiro lugar, representado nos gráficos da TRI pela letra “a”, tem-se o índice de discriminação do item, que diz respeito à aptidão do item de discriminar pessoas com diferentes níveis do conhecimento, com relação, principalmente, à dificuldade do item que é correlacionada com a área de *Theta*.

Sua correlação gráfica indica valores de 0 até 3, correspondendo da menor para a maior discriminação de um item, ou seja, quanto maior, melhor será a discriminação do item. Sendo assim, esse parâmetro é o responsável por aumentar ou diminuir a diferença entre as probabilidades de candidatos com habilidades diferentes responderem corretamente ao mesmo item (HUTZ; BANDEIRA; TRENTINI, 2015).

De acordo com Rabelo (2013), a TAB. 3 representa uma classificação que é unanimidade entre os autores e é utilizada como referência.

Tabela 3 - Critérios dos valores do parâmetro de discriminação (“a”) e sua discriminação pela TRI

VALORES DO PARÂMETRO “a ”	DISCRIMINAÇÃO
$\alpha = 0$	Nenhuma
$0 < \alpha \leq 0,35$	Muita baixa
$0,35 < \alpha \leq 0,65$	Baixa
$0,65 < \alpha \leq 1,35$	Moderada
$1,35 < \alpha \leq 1,70$	Alta
$\alpha > 1,70$	Muito alta

Fonte: Rabelo (2013, p. 138).

O segundo parâmetro a ser analisado, representado nos gráficos da TRI pela letra “b”, corresponde à dificuldade do item e representa, de acordo com o nível de habilidade individual, a chance de se obter uma resposta correta maior que 50%, quando se adota esse modelo de três parâmetros. Dessa forma, itens com maior dificuldade irão exigir níveis *Theta* (θ) maiores, para que os indivíduos tenham a chance de acerto superior a 50%. Para tal, classifica-se esse parâmetro, em geral, de -3 até +3, indo dos itens fáceis aos difíceis, respectivamente, na pontuação apresentada. (HUTZ; BANDEIRA; TRENTINI, 2015).

De acordo com Rabelo (2013), a TAB. 4 representa uma classificação do parâmetro de dificuldade (b), que é unanimidade entre os autores e é utilizada como referência.

Tabela 4 - Critérios de distribuição e classificação do grau de dificuldade (“b”) do item pela TRI

DISTRIBUIÇÃO ESPERADA	ÍNDICE DE DIFICULDADE DO ITEM	CLASSIFICAÇÃO DO ITEM EM RELAÇÃO AO ÍNDICE DE DIFICULDADE
10%	Até - 1,28	Muito fáceis
20%	De - 1,27 a - 0,52	Fáceis
40%	De - 0,51 a 0,51	Medianos
20%	De 0,52 a 1,27	Difíceis
10%	1,28 ou mais	Muito difíceis

Fonte: Rabelo (2013, p. 143).

Por fim, representado nos gráficos da TRI pela letra “c”, nesse modelo, o último parâmetro a ser discutido diz respeito à probabilidade do indivíduo com baixo conhecimento acertar o item ao acaso, ou seja, o acerto por “chute”. A representação gráfica desse parâmetro permite dizer que a probabilidade de acerto para um item adequado fica com um valor igual a 0, ou bem próximo. Ao visualizar valores acima de 0, no ponto gráfico em que a curva passa pelo eixo das probabilidades, há o indício de que o acerto foi ao acaso. (HUTZ; BANDEIRA; TRENTINI, 2015).

No modelo proposto para o presente estudo, que envolve os três parâmetros (a = discriminação; b = dificuldade; c = acerto ao acaso), há uma curva que, de forma rápida e objetiva, informa os principais parâmetros sobre um determinado item e que é conhecida como Curva Característica do Item (CCI). Essa curva é um modelo matemático com a função de probabilidade e, dessa forma, trabalha a sua imagem no intervalo $[0, 1]$, na qual o número $P(U_{ij}=1 | \theta_j)$ pode ser identificado com a proporção de respostas corretas ao item I , no grupo de candidatos com habilidade θ . A sua apresentação se dá em forma sigmoide (curva em “S”), na qual o eixo horizontal está representado pela escala de habilidade e o eixo vertical pela probabilidade do candidato com uma habilidade θ_j acertar o item i (GOMES, 2012).

Completando as análises gráficas da TRI, temos também a curva de informação, que possibilita avaliar o quão bem o item prevê o critério na TCT ou o quão bem ele representa, de forma comportamental, o traço latente na TRI. Dessa forma, a função de informação da TRI corresponde ao método da TCT para verificar a validade dos itens, ou seja, ao cálculo do erro de estimação, que mostra o quanto o escore obtido pelo candidato num item se afasta do seu escore verdadeiro (PASQUALI, 2003).

A curva de informação do item ressalta para que níveis de conhecimento (θ) a curva do item traz informação adequada. Isso significa dizer que a curva nos mostra a amplitude do *Theta* para a qual o item apresenta uma informação confiável, dizendo que, fora dessa amplitude, o item produz mais informação errônea (erro) sobre o *Theta* do que informação correta. Sendo assim, a curva de informação cobre ou tem interface com ambos os parâmetros dos itens, isto é, validade e precisão, mas não se confunde com nenhum dos dois (PASQUALI, 2003).

Em relação à definição estatística da Função de Informação, se essa fosse uma medida ótima de um dado traço latente, então, ela deveria trazer informação alta para todos os níveis desse

traço, resultando numa reta sobre todos os níveis. No entanto, um item normalmente traz melhor informação sobre alguns níveis de proficiência (θ) do que outros, sendo a representação dessa informação do item semelhante a uma curva de tipo normal (curva na forma de “sino”) (PASQUALI, 2003).

2 JUSTIFICATIVA

Em virtude da importância e da relevância do exame do TEC, e uma vez que, até o momento, não foi feita nenhuma avaliação psicométrica da prova, é imprescindível conhecer se esse modelo de avaliação fornece uma medida coerente e confiável do ponto de vista técnico, capaz de identificar de forma adequada os candidatos merecedores da aprovação na prova do TEC e, com isso, contribuir para o aprimoramento desse instrumento.

3 OBJETIVOS

3.1 Objetivo geral

- Avaliar as propriedades psicométricas da prova de Título de Especialista em Cardiologia realizada no ano de 2019, em relação aos parâmetros clássicos (índice de dificuldade; índice de discriminação e coeficiente correlação ponto-bisserial) e à Teoria de Resposta ao Item.

3.2 Objetivos específicos

- Verificar, de acordo com a TCT, a distribuição dos níveis de dificuldade dos itens da referida prova, o percentual de itens com bom poder discriminatório, o percentual de itens que apresentam boa correlação bisserial.
- Verificar, de acordo com a TRI, a distribuição dos níveis de dificuldade dos itens da referida prova, o percentual de itens com bom poder discriminatório e a probabilidade do indivíduo acertar o item ao acaso. Averiguar a qualidade dos itens da prova de Título de Especialista, elaborada pela Comissão Julgadora do Título de Especialista em Cardiologia (CJTEC), da Sociedade Brasileira de Cardiologia.

4 MATERIAIS E MÉTODOS

4.1 Desenho do estudo

A presente proposta configura-se como um estudo observacional, com a análise psicométrica das 120 questões da prova realizada por 1120 (mil e cento e vinte) candidatos, para a obtenção do título de especialista em Cardiologia, no dia 27 de outubro de 2019, das 13:00 às 18:00 horas, na Universidade Paulista (UNIP) localizada na rua Vergueiro, 1211, bairro Paraíso, na cidade de São Paulo-SP.

4.2 População/amostra

A amostra desse estudo foi constituída pelos gabaritos de 120 questões, entregues e respondidas pelos médicos que realizaram a prova para a obtenção do título de especialista em cardiologia, realizada pela CJTEC no ano de 2019.

4.3 Critérios de inclusão

Foram considerados critérios de inclusão todos os gabaritos entregues pelos candidatos que realizaram a prova aplicada no ano de 2019, que constava de 120 questões, para obtenção do TEC.

4.4 Critérios de exclusão

Foram excluídas duas questões, após a fase de recursos, sendo essas eliminadas da fase de análise estatística, além da prova de um candidato que respondeu apenas duas questões.

4.5 Amostra

Foram incluídas para avaliação 118 questões da prova de título de especialista em cardiologia, que foram respondidas e entregues pelos 1120 candidatos.

4.6 Coleta de dados

Os dados dos candidatos foram extraídos diretamente do banco de dados da empresa responsável pela elaboração da prova (Segmento Farma Editores Ltda., em parceria com a Simples Detalhe Assessoria, Planejamento e organização de Eventos Ltda. e a Picsis informática Indústria e comércio Ltda.), os quais foram alocados em planilhas Microsoft Excel®.

A partir desses dados, foram geradas planilhas separadas para os dados de identificação e para os dados referentes às notas de cada prova. Os nomes dos candidatos foram excluídos das planilhas, com o objetivo de manter o sigilo da pesquisa, sendo a identificação de cada candidato feita por um número. As planilhas foram integradas num único banco de dados, com base no número de cada candidato.

4.7 Plano de análise estatística

O desfecho do estudo foi a avaliação psicométrica das questões da prova de Título de Especialista em Cardiologia, em relação a índice de dificuldade, índice de discriminação e correlação bisserial pela análise clássica dos testes, além da análise pela Teoria de Resposta ao Item. Dessa forma, a análise clássica e a análise realizada a partir da TRI dos itens da prova de título em Cardiologia foram baseadas em seus parâmetros descritivos, os quais auxiliam a interpretação da distribuição das respostas para cada alternativa.

4.8 Aspectos éticos

Foram utilizados, nesse estudo, bases de dados secundárias, sem identificação individual dos participantes, de modo que não houve necessidade de utilização de Termo de Consentimento Livre e Esclarecido (TCLE). Assim, o pesquisador assume o compromisso de manter a confidencialidade dos dados e não realizar identificação individual dos participantes.

No entanto, para a obtenção do banco de dados, foi firmado um Termo de Consentimento de Utilização do Banco de Dados (TCUD), o qual foi encaminhado, inicialmente, para a SBC e, posteriormente, ao comitê de Ética em Pesquisa (CEP), sendo aprovado pelas duas entidades, conforme o CAAE: 31226819.6.0000.5143. Instituição proponente: Universidade José Rosário

Vellano / UNIFENAS. Patrocinador principal: Financiamento próprio. Número do parecer: 4.030.702.

4.9 Análise estatística

Uma avaliação psicométrica das questões foi realizada, utilizando os métodos da TCT e da TRI.

O modelo mais utilizado é a TCT, que obtém os resultados de uma prova e considera os escores brutos dos indivíduos que fazem o teste. Nesse caso, o resultado é dado comparando a quantidade de itens respondidos corretamente por cada indivíduo.

A TCT não permite comparar indivíduos que tenham respondido a testes diferentes, nem fazer o acompanhamento de um mesmo indivíduo em várias etapas do processo, ou seja, os resultados dependem do conjunto de itens selecionados para a construção da referida prova.

Como forma de diminuir essas limitações, houve um aumento na procura das técnicas da TRI, que utiliza modelos estatísticos para medir traços latentes de um candidato, isto é, características que variam de candidato para candidato e não podem ser aferidas ou observadas de forma direta, como altura, peso, idade, entre outras variáveis.

A presente análise estatística apresenta elementos qualitativos que fazem da TRI uma teoria capaz de acabar com algumas deficiências na TCT, como a comparação entre provas aplicadas em diferentes anos ou para diferentes grupos de candidatos.

Para um melhor entendimento da TRI, é importante apresentar alguns fatores nela contidos. O primeiro fator a ser tratado é a proficiência (θ), ou seja, o conhecimento do candidato, que mostra o nível de aptidão de um indivíduo para responder corretamente a um conjunto de itens. Essa habilidade é denominada como traço latente (proficiência) e é o que se quer conhecer do candidato com a aplicação da TRI. O segundo fator refere-se aos parâmetros do item, que são conhecidos por:

a) Parâmetro de Discriminação (a) - consiste na avaliação da aptidão do item em distinguir candidatos com habilidades diferentes.

b) Parâmetro de Dificuldade (b) - trata-se da habilidade mínima que um candidato precisa para ter uma grande probabilidade de dar a resposta correta.

c) Parâmetro de Acerto ao Acaso (c) - é a probabilidade de um candidato com baixa proficiência responder de forma correta ao item.

Nesse último parâmetro (c), Acerto ao Acaso (“chute”), valores superiores a 20% são considerados altos para itens com cinco categorias de resposta (1 de 5 = 20%, é a probabilidade de um candidato de acertar um item ao respondê-lo dando um “chute”).

Na presente análise, foi adotado o critério de, pelo menos, 25% ($c > 0,25$) de chute como insatisfatório para um determinado item da prova. Ou seja, como foram aplicadas 1.120 provas, 5% a mais de chute além do esperado (20%) é considerado muito alto e, portanto, o item avaliado apresenta algum problema em sua formulação ou nas opções de respostas. Isto é, será preciso considerar a presença de um item de muito fácil solução ou o fato de as opções de respostas permitirem, por eliminação, que o candidato acerte facilmente o item proposto na prova.

O “chute” é demonstrável pela falta de coerência do candidato, por exemplo, em errar itens fáceis ou muito fáceis e, de modo contraditório, acertar itens difíceis ou muitos difíceis sem proficiência suficiente, teoricamente, para tal. Está aqui uma das grandes vantagens da TRI sobre a TCT, que é avaliar a coerência de um candidato qualquer em responder aos itens da prova, ou seja, acertar mais itens fáceis do que difíceis, por exemplo.

5 RESULTADOS

Os resultados apresentados baseiam-se nos estudos das técnicas discriminadas, representados, na TCT, pela distribuição dos níveis de dificuldade dos itens, o percentual com poder discriminatório e o percentual de itens que apresentam boa correlação bisserial, bem como nos parâmetros estudados na TRI.

De acordo com o estudo das respostas dos candidatos em relação ao número das questões, a TAB. 5 apresenta um resumo dos parâmetros de avaliação dos itens, com base no modelo logístico unidimensional de três parâmetros (3PL) da TRI.

Tabela 5 - Modelo da TRI – 3PL inicial, considerando-se 118 itens da prova

(Continua)

ITEM	A	B	C	ITEM	A	B	C	ITEM	A	B	C
<i>1</i>	2,024	-0,161	0,368	<i>42</i>	0,990	-1,123	0,772	<i>83</i>	1,481	0,608	0,503
<i>2</i>	3,522	1,168	0,557	<i>43</i>	1,461	-1,755	0,193	<i>84</i>	0,851	-1,441	0,447
<i>3</i>	0,787	-1,257	0,562	<i>44</i>	0,640	0,280	0,002	<i>85</i>	0,241	-1,186	0,024
<i>4</i>	1,219	1,903	0,056	<i>45</i>	0,459	-0,317	0,006	<i>86</i>	1,544	1,469	0,269
<i>5</i>	1,444	0,076	0,071	<i>47</i>	1,194	-0,641	0,181	<i>87</i>	1,969	1,049	0,156
<i>6</i>	0,848	-1,825	0,594	<i>48</i>	1,286	1,863	0,497	<i>88</i>	1,132	-0,558	0,211
<i>7</i>	1,005	1,013	0,039	<i>49</i>	0,770	1,269	0,001	<i>89</i>	1,885	-0,448	0,165
<i>8</i>	1,575	2,700	0,272	<i>50</i>	1,100	-1,377	0,492	<i>90</i>	0,491	4,217	0,001
<i>9</i>	0,935	-0,455	0,043	<i>51</i>	1,013	-0,095	0,376	<i>91</i>	0,872	-1,148	0,018
<i>10</i>	0,897	-1,995	0,008	<i>52</i>	0,803	-1,644	0,021	<i>92</i>	0,336	1,052	0,005
<i>11</i>	0,753	-0,960	0,039	<i>53</i>	0,868	1,190	0,198	<i>93</i>	0,876	-4,737	0,364
<i>12</i>	0,293	3,079	0,003	<i>54</i>	1,638	0,135	0,185	<i>94</i>	1,371	0,841	0,343

Tabela 5 - Modelo da TRI – 3PL inicial, considerando-se 118 itens da prova

(continuação)

ITEM	A	B	C	ITEM	A	B	C	ITEM	A	B	C
<i>13</i>	1,089	-1,670	0,034	<i>55</i>	1,251	-0,778	0,762	<i>95</i>	1,072	-2,926	0,047
<i>14</i>	2,724	1,925	0,193	<i>56</i>	0,275	5,139	0,014	<i>96</i>	0,947	0,704	0,295
<i>15</i>	2,023	1,153	0,268	<i>57</i>	1,392	0,356	0,416	<i>97</i>	1,637	2,539	0,260
<i>16</i>	0,870	-2,253	0,055	<i>58</i>	0,296	0,953	0,026	<i>98</i>	0,253	-0,372	0,049
<i>17</i>	1,864	0,213	0,656	<i>59</i>	1,501	-0,963	0,659	<i>99</i>	0,825	-0,201	0,155
<i>18</i>	1,064	-0,777	0,166	<i>60</i>	0,667	-2,201	0,036	<i>100</i>	1,492	-0,462	0,257
<i>19</i>	1,060	-2,013	0,012	<i>61</i>	1,012	-3,812	0,245	<i>101</i>	1,071	-0,744	0,349
<i>20</i>	1,675	-0,827	0,405	<i>62</i>	1,076	-2,485	0,045	<i>102</i>	1,175	-1,064	0,019
<i>21</i>	2,110	-1,126	0,356	<i>63</i>	1,533	0,804	0,575	<i>103</i>	1,079	0,560	0,037
<i>22</i>	1,190	-1,963	0,177	<i>64</i>	1,258	1,055	0,283	<i>104</i>	0,750	-1,409	0,006
<i>24</i>	2,071	-0,730	0,121	<i>65</i>	0,620	-1,810	0,007	<i>105</i>	0,967	0,449	0,420
<i>25</i>	1,165	-0,738	0,124	<i>66</i>	1,113	-0,351	0,205	<i>106</i>	1,487	1,501	0,461
<i>26</i>	1,781	0,336	0,602	<i>67</i>	0,168	3,826	0,044	<i>107</i>	0,983	0,250	0,005
<i>27</i>	0,325	-2,722	0,021	<i>68</i>	1,437	0,352	0,310	<i>108</i>	1,805	-0,711	0,495
<i>28</i>	0,899	0,399	0,075	<i>69</i>	1,025	-2,444	0,040	<i>109</i>	1,106	1,422	0,459
<i>29</i>	0,507	-2,855	0,014	<i>70</i>	1,042	-1,738	0,018	<i>110</i>	-0,174	-0,907	0,010
<i>30</i>	1,646	0,805	0,283	<i>71</i>	1,265	-0,882	0,003	<i>111</i>	0,723	-0,798	0,533
<i>31</i>	0,683	0,869	0,002	<i>72</i>	1,752	-1,343	0,397	<i>112</i>	1,357	-0,589	0,478

Tabela 5 - Modelo da TRI – 3PL inicial, considerando-se 118 itens da prova

(Conclusão)											
ITEM	A	B	C	ITEM	A	B	C	ITEM	A	B	C
32	1,680	0,561	0,071	73	1,509	-0,606	0,214	113	0,862	0,620	0,181
33	1,690	-1,010	0,815	74	1,396	0,488	0,385	114	0,483	-1,990	0,010
34	0,767	-0,750	0,322	75	1,367	1,506	0,230	115	1,114	-0,530	0,296
35	0,924	-0,759	0,473	76	0,625	-2,361	0,025	116	1,518	-1,903	0,265
36	0,529	0,650	0,013	77	0,152	5,640	0,042	117	0,978	0,138	0,201
37	2,239	-0,764	0,693	78	1,109	-0,397	0,846	118	1,147	0,149	0,723
38	0,498	-2,433	0,049	79	0,750	0,173	0,492	119	1,465	-2,029	0,569
39	0,567	-2,892	0,034	80	0,922	-0,402	0,715	120	1,123	-1,730	0,239
40	0,191	-1,734	0,046	81	0,177	5,816	0,009				
41	0,698	-1,724	0,008	82	1,486	-0,407	0,222				

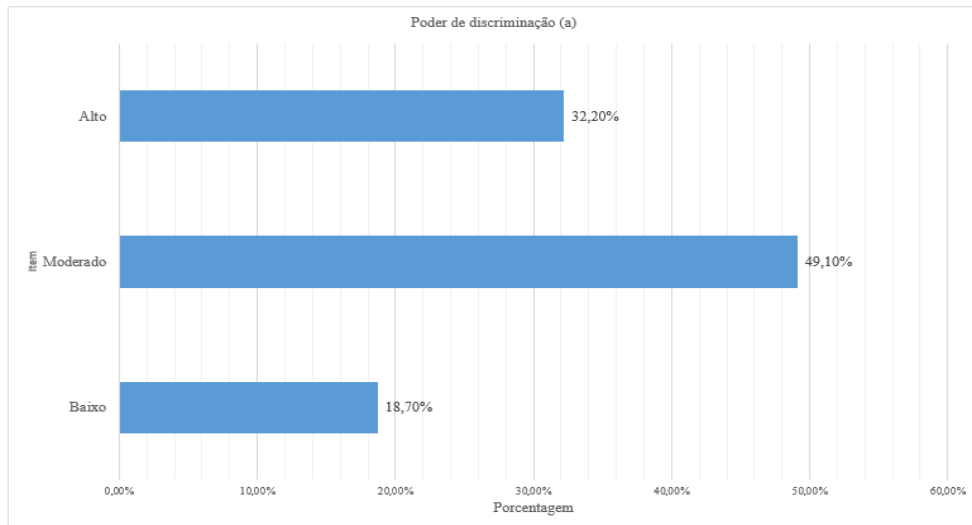
Fonte: Elaborado pelo autor.

Nota: 2 itens da prova anulados (n^{os} 23 e 46). Base de Dados: 1.120 candidatos.

O resultado mostra que somente um item (Questão n° 110, $a = -0,174$) apresentou nível negativo para o parâmetro de Discriminação, ou seja, o item indica que quanto maior o nível de conhecimento do candidato, menor a chance de acertá-lo. Portanto, trata-se de um resultado completamente incoerente com o objetivo do parâmetro, de modo que o item foi retirado da prova em análise.

Ressalta-se que, em relação à discriminação, 18,7% dos itens são considerados com “Muito Baixo” ou “Baixo” poder de Discriminação ($a \leq 0,65$). Além disso, 49,1% deles mostram “Moderada” capacidade de Discriminação ($0,65 < a \leq 1,70$) e 32,2% dos itens demonstram “Alto” ou “Muito alto” poder de Discriminação ($a > 1,70$), conforme a GRAF. 2.

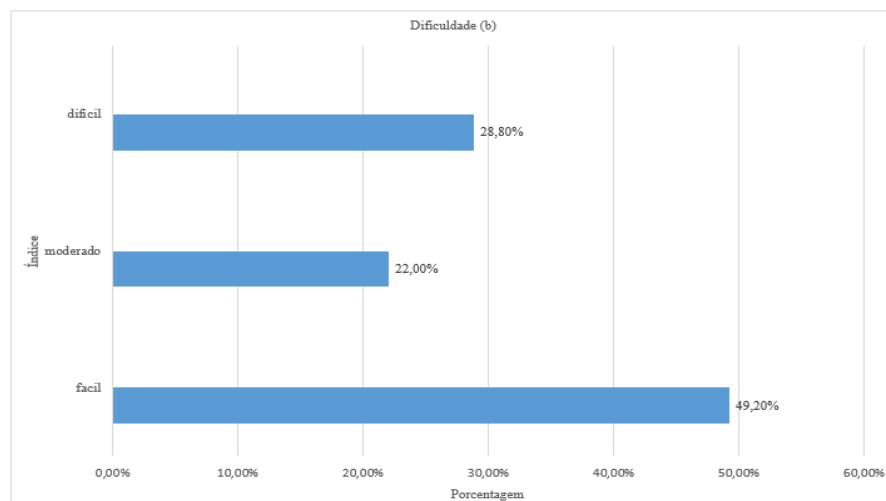
Gráfico 2 - Poder de discriminação



Fonte: Elaborado pelo autor

Em relação ao parâmetro Dificuldade (b), os resultados mostraram que 49,20% dos itens apresentam níveis de dificuldade considerados “Muito fáceis” ou “Fáceis” ($\leq -1,28$: Muito fácil ou de $-1,27$ a $-0,52$: Fácil); 22,0% dos itens apresentaram nível de dificuldade “Mediano / Moderado” (De $-0,51$ a $0,51$) e o restante, 28,80% dos itens, apresentou níveis de dificuldade considerados “Difíceis” ou “Muito difíceis” (De $0,52$ a $1,27$: “Difíceis” e $\geq 1,28$: “Muito difíceis”), conforme GRAF. 3.

Gráfico 3 - Dificuldade



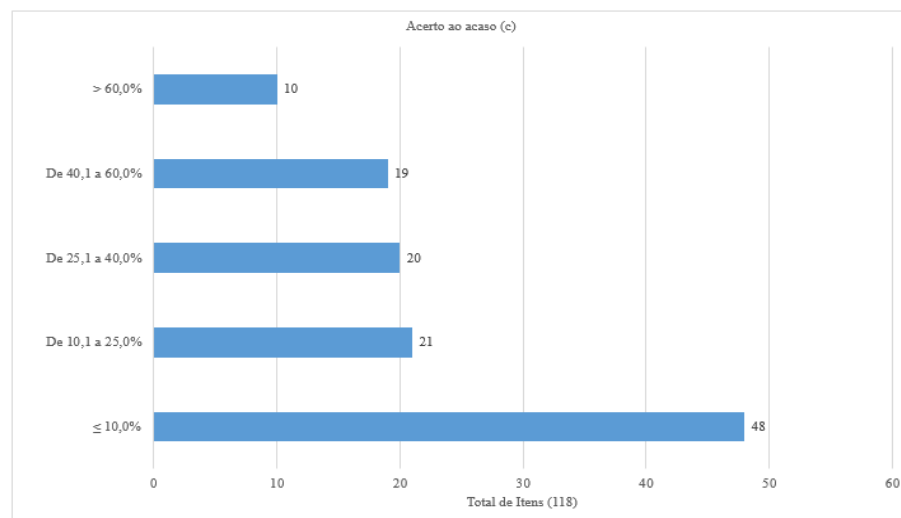
Fonte: Elaborado pelo autor

Considera-se como uma prova equilibrada e adequada quanto ao parâmetro de Dificuldade (b) aquela que adota a seguinte distribuição para os itens: 10% de “Muito fáceis”, 20% de Fáceis,

40% de itens “Moderados”, 20% de Díficeis e 10% de itens “Muito difíceis”. Ressalta-se que esse parâmetro, em 99,7% das vezes, varia numa escala de -3 a +3 pontos (faixa de normalidade de uma variável padronizada).

A distribuição dos acertos ao acaso está representada no GRAF. 4.

Gráfico 4 - Acerto ao acaso



Fonte: Elaborado pelo autor

A TAB. 6 apresenta as marcações dos itens retirados da prova, com base no modelo logístico unidimensional da TRI com três parâmetros (3PL). As marcações dos X's, destacados em negrito e maiores em tamanho, indicam os itens com problemas, de acordo com os três parâmetros do modelo TRI (a, b ou c) gerados para a prova proposta.

Tabela 6 - Modelo da TRI – 3PL inicial, de acordo com motivos de retirada de itens

(Continua)

ITEM	A	B	C	ITEM	A	B	C	ITEM	A	B	C
1			X	42			X	83			X
2			X	43				84			X
3	X	X	X	44				85	x	X	X
4				45	X	X	X	86			X

Tabela 6 - Modelo da TRI – 3PL inicial, de acordo com motivos de retirada de itens

(continuação)

ITEM	A	B	C	ITEM	A	B	C	ITEM	A	B	C
5				47				87			
6	X	X	X	48			X	88			
7				49				89			
8			X	50			X	90	x	X	X
9				51			X	91			
10				52				92	x	X	X
11				53				93	x	X	X
12	X	X	X	54				94			X
13				55			X	95			
14				56	X	X	X	96			X
15			X	57			X	97			X
16				58	X	X	X	98	x	X	X
17			X	59			X	99			
18				60				100			X
19				61		X		101			X
20			X	62				102			
21			X	63			X	103			
22				64			X	104			
24				65				105			X
25				66				106			X
26			X	67	X	X	X	107			
27	X	X	X	68			X	108			X
28				69				109			X
29	X	X	X	70				110	x	X	X
30			X	71				111	x	X	X

Tabela 6 - Modelo da TRI – 3PL inicial, de acordo com motivos de retirada de itens

(conclusão)

ITEM	A	B	C	ITEM	A	B	C	ITEM	A	B	C
31				72			X	112			x
32				73				113			
33			X	74			X	114	x	X	x
34			X	75				115			x
35			X	76				116			x
36				77	X	X	X	117			
37			X	78	X	X	X	118			x
38	X	X	X	79	X	X	X	119			x
39	X	X	X	80	X	X	X	120			
40	X	X	X	81	X	X	X				
41				82							

Fonte: Elaborada pelo autor.

Nota: 2 itens da prova anulados (n^{os} 23 e 46). Base de dados: 1.120 candidatos.

Cabe ressaltar que os itens marcados com 3 X's demonstram baixa capacidade de gerar informação (Gráfico de Informação da TRI para o conjunto dos 118 itens da prova) para a prova, portanto, deverão ser retirados do conjunto dos itens a serem utilizados na composição da nota final do candidato.

A TAB. 7 apresenta os parâmetros do modelo final da TRI, considerando-se somente os 49 itens selecionados para avaliação dos candidatos após ajuste final, que descartou 69 itens da prova inicialmente aplicada.

Ao observar a tabela, verifica-se que as 49 questões analisadas possuem um fator satisfatório, de acordo com a análise estatística. O GRAF. 5 mostra a Curva Característica do Item, considerando-se o conjunto de 118 itens avaliados pelo modelo inicial da TRI de três parâmetros (Discriminação – a, Dificuldade – b e Acerto ao Acaso / “Chute” – c).

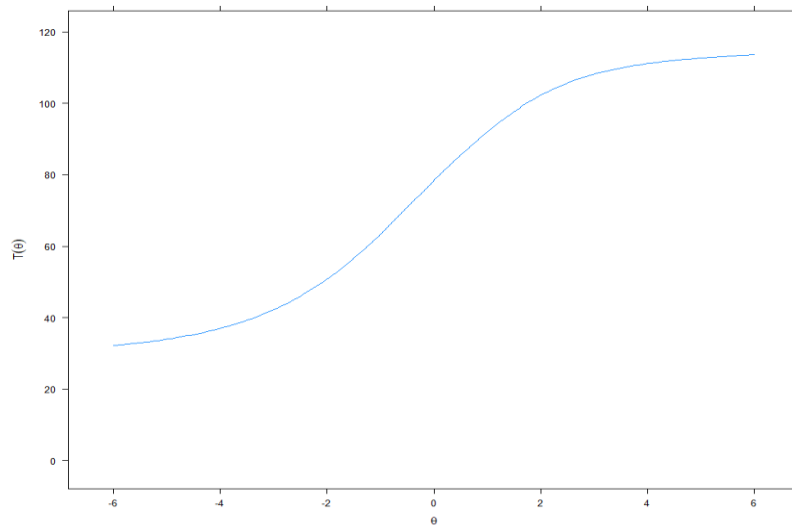
Tabela 7 - Modelo da TRI – 3PL final, considerando-se 49 itens da prova

ITEM	A	B	C	ITEM	A	B	C
4	1,291	1,825	0,055	60	0,646	-2,301	0,016
5	1,424	0,034	0,052	62	1,104	-2,403	0,073
	0,954	0,987	0,024	65	0,590	-1,894	0,005
9	0,944	-0,433	0,052	66	1,183	-0,308	0,218
10	0,905	-1,989	0,003	69	1,074	-2,403	0,006
11	0,822	-0,667	0,134	70	1,022	-1,746	0,029
13	1,133	-1,577	0,069	71	1,297	-0,787	0,059
14	2,606	1,988	0,195	73	1,594	-0,593	0,212
18	1,104	-0,609	0,235	75	1,056	1,494	0,191
19	1,038	-2,020	0,028	76	0,585	-2,542	0,007
22	1,150	-2,205	0,024	82	1,475	-0,461	0,195
24	1,986	-0,778	0,094	87	1,863	1,032	0,147
25	1,173	-0,631	0,178	88	1,074	-0,748	0,133
28	0,806	0,283	0,032	89	2,065	-0,456	0,148
31	0,696	0,851	0,001	91	0,860	-1,175	0,010
32	1,676	0,580	0,079	95	1,030	-3,041	0,020
36	0,530	0,611	0,005	99	0,830	-0,209	0,152
41	0,707	-1,712	0,005	102	1,122	-1,120	0,004
43	1,445	-1,705	0,237	103	1,062	0,505	0,018
44	0,628	0,280	0,001	104	0,743	-1,424	0,005
47	1,037	-0,998	0,031	107	0,987	0,243	0,003
49	0,778	1,257	0,000	113	0,714	0,304	0,084
52	0,824	-1,637	0,005	117	0,892	-0,048	0,141
53	1,036	1,200	0,228	120	1,128	-1,679	0,265
54	1,502	0,068	0,156				

Fonte: Elaborada pelo autor.

Nota: 2 itens da prova anulados (n^{os} 23 e 46). Base de Dados: 1.120 candidatos

Gráfico 5 - Total da Nota - $T(\theta)$ - estimada pela TRI inicial, considerando-se 118 itens da prova, para cada candidato de acordo com sua aptidão (θ).



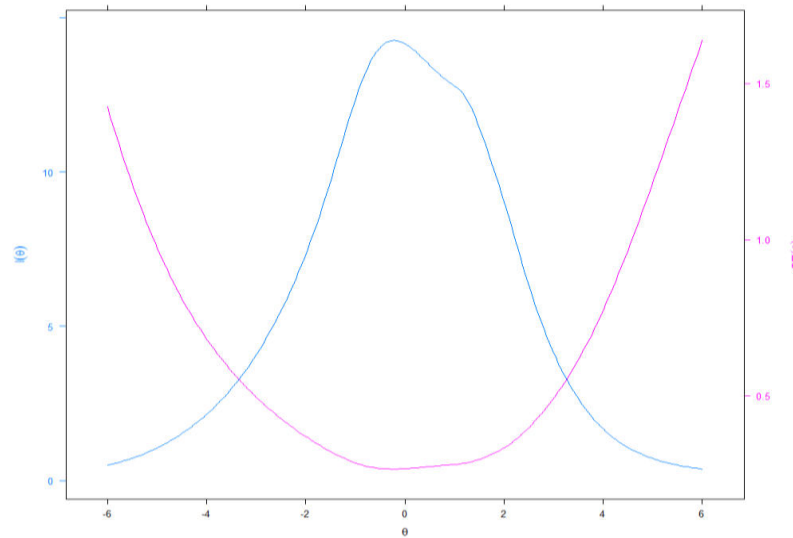
Fonte: Elaborado pelo autor.

Em se tratando de um teste de proficiência, o valor de *theta* (θ) do candidato é expresso pelo número de itens que ele acertou, isso na TCT. Em se tratando da TRI, entretanto, não se pergunta quantos itens o candidato acertou e, sim, por que ele acertou ou errou cada item individualmente. Dessa forma, a TRI está interessada em descobrir qual é o valor de *theta* (θ) que o candidato deve ter para poder acertar cada um dos itens de uma prova. Portanto, o problema a se resolver consiste em saber como o item sinaliza o valor da proficiência (valor de θ do candidato).

Para responder a essa questão, utiliza-se a Curva Característica do Item (CCI). É razoável supor que um candidato que tenha maior proficiência (θ) tenha uma probabilidade maior de acertar esse item do que um sujeito com nível menor de aptidão (θ).

A Curva de Informação, representada pelo GRAF. 6, mostra que a quantidade de informação máxima recuperada sobre o raciocínio analógico do candidato, pelo modelo inicial da TRI (118 itens), encontra-se em torno da mediana da proficiência, ou seja, valor de θ próximo de 0 (zero). Além disso, para os valores extremos dos níveis de *Theta* (θ), onde a curva “rosa” corta os pontos da curva “azul” (curva em forma de “sino”), a prova produz mais erro de informação do que informação legítima, portanto, para valores de proficiência variando entre -3,2 a +3,1 ($-3,2 \leq \theta \leq 3,1$), encontra-se o máximo de informação gerada pelo modelo inicial da TRI.

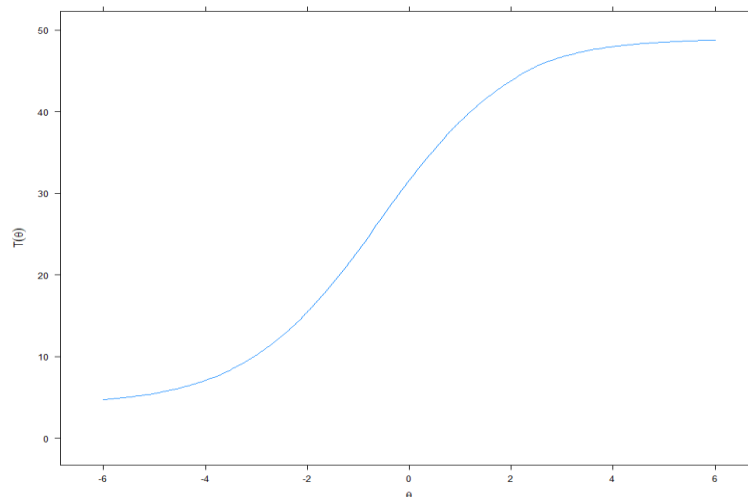
Gráfico 6 - Curva de Informação - $I(\theta)$ – e erro padrão da Curva de Informação gerados pela TRI inicial, considerando-se 118 itens da prova



Fonte: Elaborado pelo autor

O GRAF. 7 mostra a CCI do conjunto de itens estimado pelo modelo da TRI, considerando-se somente os 49 itens da prova (modelo final). Da mesma forma que o modelo TRI inicial, o resultado mostra que quanto maior a proficiência (θ) do candidato, maior será o número de itens com resposta correta.

Gráfico 7 - Total da Nota - $T(\theta)$ - estimada pela TRI inicial, considerando-se 49 itens da prova, para cada candidato de acordo com sua aptidão (θ)



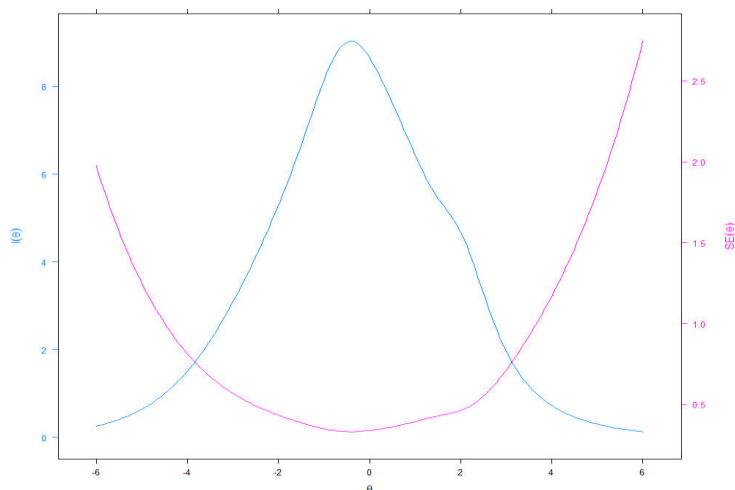
Fonte: Elaborado pelo autor

Sendo que um candidato apresenta proficiência igual a 0 ($\theta = 0$ – aptidão mediana, θ entre -1 e +1), é esperado que ele acerte, aproximadamente, 32 dos 49 itens da prova (65,3% de acertos). E, de um candidato com muito baixo nível de aptidão ($\theta < -4,0$), é esperado que acerte pelo menos 4 dos 49 itens da prova [(8,2%) - onde a curva CCI corta o eixo da ordenada do gráfico

$T(\theta)$]. Portanto, para o modelo final da TRI com 49 itens, os candidatos precisarão de um nível de proficiência (θ) maior do que os candidatos do modelo em que são considerados os 118 itens da prova, para obter o mesmo percentual de pontuação de partida.

A Curva de Informação, representada no GRAF. 8, mostra que a quantidade de informação máxima recuperada sobre o raciocínio analógico do candidato, pelo modelo final da TRI (49 itens), encontra-se, também, em torno da mediana da proficiência, ou seja, valor de θ próximo de 0 (zero). Além disso, para os valores extremos dos níveis de *Theta* (θ), onde a curva “rosa” corta os pontos da curva “azul” (Curva em forma de “sino”), a prova produz mais erro de informação do que informação legítima. Portanto, para valores de proficiência variando entre -4,0 a +3,2, encontra-se o máximo de informação gerada pelo modelo final da TRI.

Gráfico 8 - Curva de Informação - $I(\theta)$ – e erro padrão da Curva de Informação gerados pela TRI inicial, considerando-se 49 itens da prova



Fonte: Elaborado pelo autor

Cabe destacar, na comparação entre os modelos inicial e final da TRI, o valor da redução do percentual de candidatos, de 30,5% no modelo inicial para 8,2% no modelo final que, mesmo com baixíssima proficiência (θ), acertariam itens da prova. Há uma redução significativa desse percentual devido à diminuição de itens, por exemplo, respondidos corretamente com base no acerto ao acaso (Parâmetro c da TRI).

A TAB. 8 apresenta, de forma resumida, a classificação dos parâmetros da TRI e da TCT dos 118 itens da prova aplicada aos candidatos ao Título de Especialista em Cardiologia, em 2019, da Sociedade Brasileira de Cardiologia.

Tabela 8 - Classificação dos parâmetros da TRI e TCT dos 118 itens da prova, no geral

(Continua)

	Frequência	
	n	%
TRI (Teoria de Resposta ao Item)		
Discriminação (a)		
$\leq 0,35$ (Muito baixa)	12	10,2
De 0,351 a 0,650 (Baixa)	10	8,5
De 0,651 a 1,350 (Moderada)	58	49,1
De 1,351 a 1,700 (Alta)	25	21,2
$> 1,700$ (Muito Alta)	13	11,0
TOTAL	118	100,0
Dificuldade (b)		
$\leq -1,28$ (Muito fácil)	31	26,3
De -1,27 a -0,52 (Fácil)	27	22,9
De -0,51 a 0,51 (Mediana)	26	22,0
De 0,52 a 1,27 (Difícil)	19	16,1
$\geq 1,28$ (Muito difícil)	15	12,7
TOTAL	118	100,0
Acerto ao acaso / “Chute” (c)		
$\leq 10,0\%$	48	40,7
De 10,1 a 25,0%	21	17,8
De 25,1 a 40,0%	20	16,9
De 40,1 a 60,0%	19	16,1
$> 60,0\%$	10	8,5
TOTAL	118	100,0
Curva Característica do Item (CCI)		
Satisfatória	49	41,5
Insatisfatória	69	58,5
TOTAL	118	100,0
Curva de Informação		
Satisfatória	93	78,8
Insatisfatória	25	21,2
TOTAL	118	100,0

Tabela 8 - Classificação dos parâmetros da TRI e TCT dos 118 itens da prova, no geral (conclusão)

		Frequência	
		n	%
• TCT (Teoria Clássica do Teste)			
% de acerto			
	≤ 10,0%	0	0
	De 10,1 a 30,0%	6	7,6
	De 30,1 a 50,0%	21	17,8
	De 50,1 a 70,0%	24	20,3
	De 71 a 90,0%	52	44,1
	> 90,0%	12	10,2
	TOTAL	118	100,0
Correlação Bisserial			
	<i>Satisfatória</i>	82	69,5
	<i>Insatisfatória</i>	36	30,5
	TOTAL	118	100,0
Discriminação			
	< 20,0% (<i>Item deficiente, deve ser rejeitado</i>)	30	25,4
	De 20 a 29,9% (<i>Item marginal, sujeito à reelaboração</i>)	31	26,3
	De 30 a 39,9% (<i>Item bom, sujeito a aprimoramento</i>)	29	24,6
	≥ 40,0% (<i>Item bom</i>)	28	23,7
	TOTAL	118	100,0
Gráfico AGI			
	<i>Satisfatória</i>	53	44,9
	<i>Insatisfatória</i>	65	55,1
	TOTAL	118	100,0

Fonte: Elaborada pelo autor.

Nota: 2 itens da prova anulados (n^{os} 23 e 46). Base de Dados: 1.120 candidatos.

A TAB. 9 apresenta a distribuição da proficiência (θ) dos respondentes, de acordo com a modelagem final (49 itens) da TRI, para a prova com 118 itens aplicada aos candidatos ao Título de Especialista em Cardiologia, em 2019, da Sociedade Brasileira de Cardiologia.

Tabela 9 - Escore final (θ) e pontuação obtida pelos candidatos, baseando-se no modelo final da TRI com 49 itens, no geral

	Frequência	
	N	%
• Escore final: Proficiência (θ)		
$\theta < -3,00$	8	0,7
$-3,00 \leq \theta < -2,00$	37	3,3
$-2,00 \leq \theta < -1,00$	134	12,0
$-1,00 \leq \theta < 0,00$	355	31,7
$0,00 \leq \theta < +1,00$	381	34,0
$1,00 \leq \theta < +2,00$	170	15,2
$2,00 \leq \theta \leq +3,00$	28	2,5
$\theta > 3,00$	7	0,6
TOTAL	1.120	100,0
Pontuação final: 0 a 1.000 pontos		
$Y < 200$	8	0,7
$200 \leq Y < 300$	37	3,3
$300 \leq Y < 400$	134	12,0
$400 \leq Y < 500$	355	31,7
$500 \leq Y < 600$	381	34,0
$600 \leq Y < 700$	170	15,2
$700 \leq Y \leq 800$	28	2,5
$Y > 800$	7	0,6
TOTAL	1.120	100,0
Aprovados pela TRI e TCT, conjuntamente		
<i>Sim</i>	382	83,8
<i>Não</i>	74	16,2
TOTAL	456	100,0

Fonte: Elaborada pelo autor.

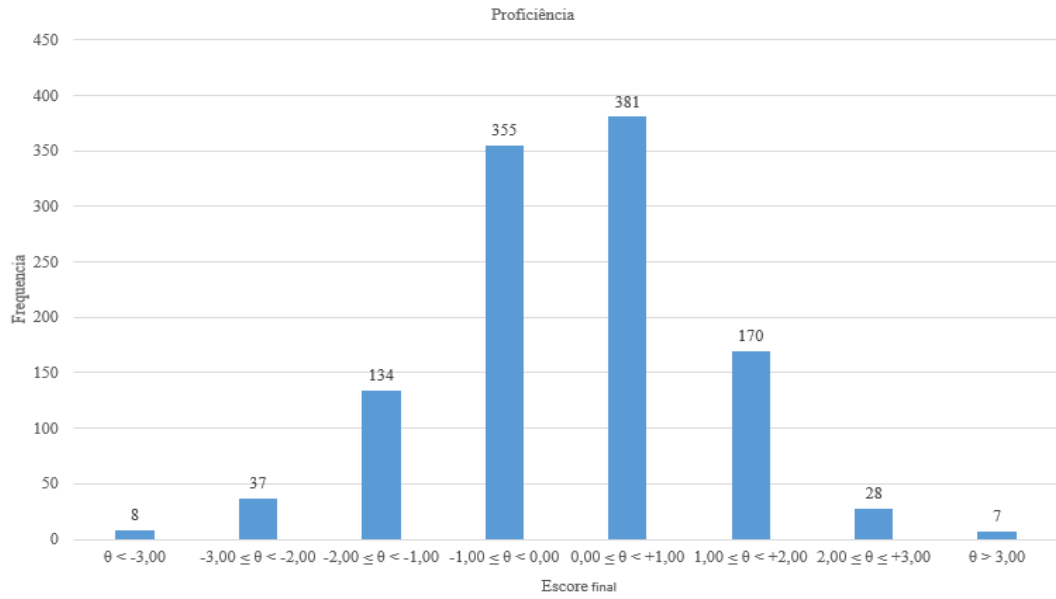
Nota: 2 itens da prova anulados (n^{os} 23 e 46). Base de Dados: 1.120 candidatos (456 candidatos foram aprovados).

Escore final (θ): pode variar de $-\infty$ a $+\infty$, normalmente, varia de -3 a +3, portanto, $\theta \sim N(0, 1)$.

Pontuação final: $Y = 500 + 100 \times X$, onde, X é o valor de θ , portanto, $Y \sim N(500, 100)$

O GRAF. 9 mostra os resultados da proficiência.

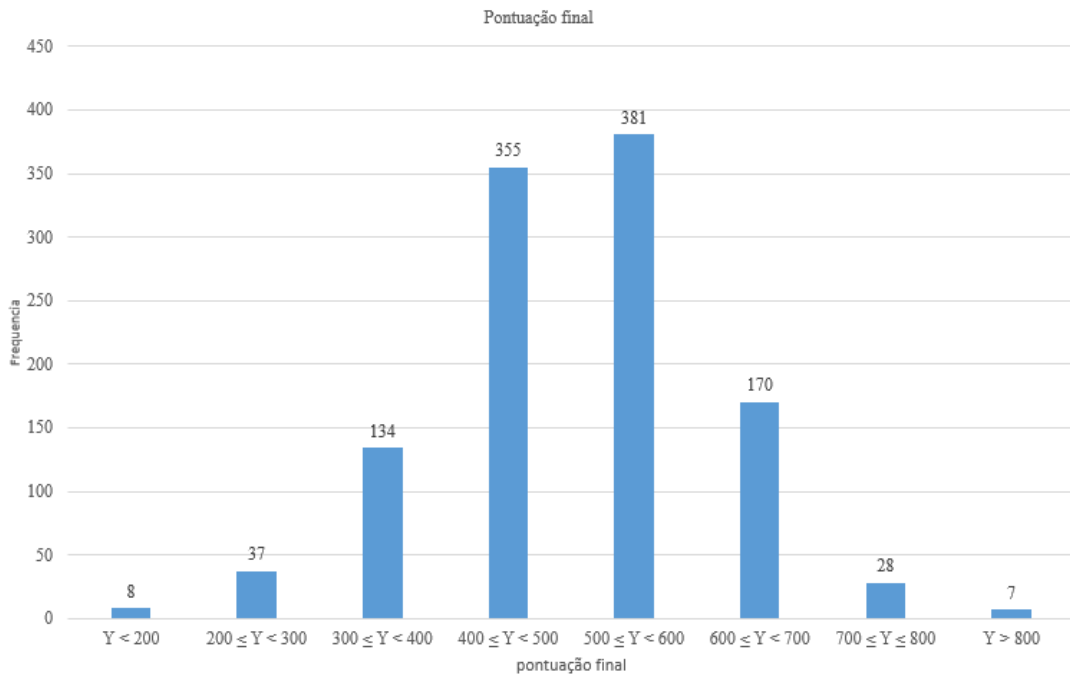
Gráfico 9 - Resultado da proficiência gerados pela TRI, considerando-se 49 itens da prova



Fonte: Elaborado pelo autor

O GRAF. 10 mostra a pontuação final gerada pela TRI, considerando-se 49 itens da prova.

Gráfico 10 - Resultado da pontuação final gerada pela TRI



Fonte: Elaborado pelo autor

Os resultados mostram que 65,7% dos candidatos obtiveram uma nota mediana na prova ($1,00 \leq \theta < +1,00$), ou em pontuação transformada: $400 \leq Y < 600$ pontos, numa escala de 0 a 1.000 pontos. Apenas 3,1% dos candidatos apresentaram nível de proficiência igual ou superior a 2,00 ($\theta \geq 2,00$, ou em pontuação: $Y \geq 600$ pontos).

No geral, cabe destacar que a pontuação utilizando modelagem TRI tende a variar de 200 a 800 pontos (ou, em termos de proficiência: $-3,00 \leq \theta \leq +3,00$). O que acontece, às vezes, é que alguns poucos candidatos podem apresentar um desempenho muito superior ou muito inferior ao esperado, com base no conjunto de casos estudados. Dessa forma, um percentual pequeno de candidatos pode extrapolar a pontuação / aptidão (θ) fora do padrão. E, se fosse adotada a metodologia da TRI como critério de aprovação dos candidatos, 16,2% dos aprovados seriam outros (74 candidatos).

As medidas a seguir, conforme mostradas nas tabelas 10, 11 e 12, são parâmetros básicos utilizados na validação e avaliação da precisão do modelo TRI gerado para os dados estudados, os 118 itens válidos da prova para o título de especialista.

Tabela 10 - Parâmetros de interesse dos modelos inicial e final da TRI utilizados para avaliação da precisão e validade

(Continua)

Medidas de qualidade do modelo	Modelo da TRI	
	Inicial	Final
<i>Total de itens</i>	118	49
<i>AIC</i>	133.615	57.749
<i>AICc</i>	133.943	57.794
<i>SABIC</i>	134.268	58.020
<i>M2</i>	7.639	1.192
<i>g.l.</i>	6.667	1.078
<i>P</i>	$3,3306 \times 10^{-6}$	0,00826
<i>RMSEA</i>	0.01142	0,00974
<i>RMSEA 5</i>	0.01009	0,00535
<i>RMSEA 95</i>	0.01264	0,01296
<i>SRMSR</i>	0.03052	0,02851
<i>TLI</i>	0.98251	0,99405

Tabela 10 - Parâmetros de interesse dos modelos inicial e final da TRI utilizados para avaliação da precisão e validade

(conclusão)

Medidas de qualidade do modelo	Modelo da TRI	
	Inicial	Final
<i>Total de itens</i>	118	49
<i>CFI</i>	0,98311	0,99455
<i>Auto valor total</i>	34,5	14,2
<i>Variância explicada (%)</i>	29,2	29,0

Fonte Elaborada pelo autor.

Nota: 2 itens da prova anulados (nos 23 e 46). Base de Dados: 1.120 candidatos.

g.l.: Graus de liberdade do modelo.

A TAB. 10 apresenta algumas das importantes medidas de avaliação de qualidade de um modelo TRI. Em relação às medidas AIC, AICc e SABIC, os resultados mostram uma redução importante entre os modelos inicial (118 itens) e final (49 itens). O Critério de informação de Akaike (AIC), o AIC corrigido (AICc) e a Medida de Avaliação de Ajuste Bayesiano (SABIC) têm como objetivo comparar dois modelos, porém, para modelos que tenham o mesmo número de itens.

A redução expressiva dessas medidas, em modelos com número de itens diferentes, mostra não apenas um bom ajuste de covariâncias e correlações observadas *versus* aquelas previstas pelo modelo, como, também, um modelo não propenso ao “super-ajustamento” dos dados. Os três critérios apresentados, apesar de serem conceitualmente diferentes, acerca do modelo da TRI em avaliação, adotam a mesma interpretação estatística desses parâmetros, ou seja, o máximo da função de verossimilhança como medida do ajustamento. Entretanto, definem valores críticos diferentes. Com o teste da razão de verossimilhança, considera-se, por hipótese, que o modelo mais simples é o de melhor ajuste, até que se observem, dado um nível de significância, diferenças estatísticas para um modelo mais completo.

A estatística M2 (Pacote estatístico R e RStudio) tem o mesmo objetivo das medidas estatísticas apresentadas anteriormente, ou seja, avaliar o ajuste do modelo TRI aos itens da prova aplicada nos candidatos ao Título de Especialista em Cardiologia da Sociedade Brasileira de Cardiologia. O resultado mostra que mais importante que a redução da medida de M2 do modelo inicial para o modelo final é a probabilidade de significância (p) dos dois modelos

avaliados. Observa-se que o ajuste do modelo inicial mostra um valor muito menor do que 0,01 (1%), indicando, portanto, um mau ajuste para o modelo da TRI com 118 itens. Resultado esperado, uma vez que os parâmetros do modelo TRI (3PL: modelo logístico unidimensional de três parâmetros da TRI) apresentaram, por exemplo, altos percentuais de chute (parâmetro c) para vários dos itens aplicados na prova e níveis de Dificuldades (b) em vários itens consideradas muitos fáceis ($b < -3$) ou muito difíceis ($b > 3$), o que implicaria em baixa capacidade de Discriminação desses itens (a).

Conforme mostrado nas medidas do modelo final reduzido (49 itens), o valor da probabilidade de significância é muito próximo de 1%, ou seja, o modelo mostra um ajuste muito melhor quando comparado ao modelo inicial, com 118 itens ($p = 1\%$ é aceitável devido ao número muito grande de itens a serem avaliados pela TRI, dado o número total de candidatos participantes da prova).

Além disso, todos os *Standardized Root Mean Square Residual* (RMSEA), ou Raiz quadrada média resíduo padronizada, e o *Standardized Root Mean Square Residual* (SRMSR), ou Raiz quadrada média resíduo padronizada, são resíduos avaliados e estão abaixo de 5%. Isso demonstra um bom ajuste das estimativas apresentadas pelo modelo gerado pela TRI, em relação ao modelo considerado como verdadeiro, esperado para o conjunto de itens avaliados com base no traço latente (aptidão / proficiência / conhecimento do candidato) apresentado pelos candidatos para o conjunto de itens avaliados. Isto é, os resíduos avaliam as discrepâncias entre a matriz observada de dados e a modelada pela TRI e quanto menor os resíduos, melhor o ajuste do modelo aplicado aos dados.

Os índices *Comparative Fit Index* (CFI) e *Tucker Lewis Index* (TLI) calculam o ajuste relativo do modelo observado, ao compará-lo com um modelo base, cujos valores acima de 0,95 indicam ótimo ajuste (NORONHA; PINTO; OTTATI, 2016).

Para determinar o número de variáveis latentes (conjunto de itens que medem uma dimensão de interesse, ou seja, a proficiência sobre o tema proposto na prova) são utilizadas as técnicas de determinação com base nas medidas de autovalores e na porcentagem de variância explicada. O objetivo é obter um melhor ajuste das variáveis latentes identificadas, bem como determinar um número mínimo de fatores que respondam pela máxima variância nos itens pesquisados.

Na modelagem da TRI, o objetivo é identificar uma única dimensão que consiga explicar o traço latente (θ) do candidato submetido a uma prova de avaliação, ou seja, certificar-se de que o conjunto de itens aplicados esteja medindo o conhecimento proposto para obtenção do Título de Especialista em Cardiologia da Sociedade Brasileira de Cardiologia. Conforme mostrado na tabela, 6, 29,2% do conjunto de itens (dimensão única) avaliados explicam o conhecimento (θ) do candidato sobre o tema proposto pela prova aplicada, baseando-se no total do autovalor obtido na avaliação do modelo inicial da TRI.

Cabe ressaltar que o aspecto mais importante dessa medida é que o percentual de variância explicado do modelo final da TRI, contendo apenas 49 itens, é praticamente o mesmo (29,0%) do modelo inicial da TRI com os 118 itens da prova (29,2%). Portanto, não há perda de generalidade das notas obtidas pelos candidatos com a apresentação de um modelo de TRI com três parâmetros, com redução no número total de itens a serem utilizados na avaliação do escore de proficiência (θ) do candidato.

Tabela 11 - Parâmetros da análise fatorial do modelo da TRI – 3PL inicial (118 Itens)

(Continua)

ITEM	CF	h ₂	ITEM	CF	h ₂	ITEM	CF	h ₂
<i>1</i>	0,765	0,586	<i>42</i>	0,503	0,253	<i>83</i>	0,656	0,431
<i>2</i>	0,900	0,811	<i>43</i>	0,651	0,424	<i>84</i>	0,447	0,200
<i>3</i>	0,420	0,176	<i>44</i>	0,352	0,124	<i>85</i>	0,140	0,020
<i>4</i>	0,582	0,339	<i>45</i>	0,260	0,068	<i>86</i>	0,672	0,451
<i>5</i>	0,647	0,419	<i>47</i>	0,574	0,330	<i>87</i>	0,757	0,572
<i>6</i>	0,446	0,199	<i>48</i>	0,603	0,363	<i>88</i>	0,554	0,307
<i>7</i>	0,509	0,259	<i>49</i>	0,412	0,170	<i>89</i>	0,742	0,551
<i>8</i>	0,679	0,461	<i>50</i>	0,543	0,294	<i>90</i>	0,277	0,077
<i>9</i>	0,482	0,232	<i>51</i>	0,511	0,262	<i>91</i>	0,456	0,208
<i>10</i>	0,466	0,217	<i>52</i>	0,427	0,182	<i>92</i>	0,194	0,038
<i>11</i>	0,405	0,164	<i>53</i>	0,454	0,206	<i>93</i>	0,458	0,209
<i>12</i>	0,170	0,029	<i>54</i>	0,694	0,481	<i>94</i>	0,627	0,393
<i>13</i>	0,539	0,291	<i>55</i>	0,592	0,351	<i>95</i>	0,533	0,284
<i>14</i>	0,848	0,719	<i>56</i>	0,159	0,025	<i>96</i>	0,486	0,236
<i>15</i>	0,765	0,585	<i>57</i>	0,633	0,401	<i>97</i>	0,693	0,481
<i>16</i>	0,455	0,207	<i>58</i>	0,171	0,029	<i>98</i>	0,147	0,022

Tabela 11 - Parâmetros da análise fatorial do modelo da TRI – 3PL inicial (118 Itens)

(Continua)

ITEM	CF	h ₂	ITEM	CF	h ₂	ITEM	CF	h ₂
17	0,739	0,545	59	0,662	0,438	99	0,436	0,190
18	0,530	0,281	60	0,365	0,133	100	0,659	0,434
19	0,529	0,280	61	0,511	0,261	101	0,533	0,284
20	0,701	0,492	62	0,534	0,285	102	0,568	0,323
21	0,778	0,606	63	0,669	0,448	103	0,535	0,287
22	0,573	0,328	64	0,595	0,353	104	0,404	0,163
24	0,773	0,597	65	0,342	0,117	105	0,494	0,244
25	0,565	0,319	66	0,547	0,300	106	0,658	0,433
26	0,723	0,523	67	0,098	0,010	107	0,500	0,250
27	0,188	0,035	68	0,645	0,416	108	0,728	0,529
28	0,467	0,218	69	0,516	0,266	109	0,545	0,297
29	0,286	0,082	70	0,522	0,273	110	-0,102	0,010
30	0,695	0,483	71	0,597	0,356	111	0,391	0,153
31	0,373	0,139	72	0,717	0,514	112	0,623	0,389
32	0,702	0,493	73	0,663	0,440	113	0,452	0,204
33	0,705	0,497	74	0,634	0,402	114	0,273	0,075
34	0,411	0,169	75	0,626	0,392	115	0,548	0,300
35	0,477	0,228	76	0,345	0,119	116	0,666	0,443
36	0,297	0,088	77	0,089	0,008	117	0,498	0,248
37	0,796	0,634	78	0,546	0,298	118	0,559	0,312
38	0,281	0,079	79	0,403	0,162	119	0,652	0,426
39	0,316	0,100	80	0,476	0,227	120	0,551	0,303
40	0,112	0,013	81	0,103	0,011			
41	0,379	0,144	82	0,658	0,432			

Fonte: Elaborada pelo autor.

Nota: 2 itens da prova anulados (n^{os} 23 e 46). Base de Dados: 1.120 candidatos.

CF : Carga Fatorial h₂ : Comunalidade.

Tabela 12 - Parâmetros da análise fatorial do modelo da TRI – 3PL final (49 Itens)

ITEM	CF	h ₂	ITEM	CF	h ₂
<i>4</i>	0,604	0,365	<i>60</i>	0,355	0,126
<i>5</i>	0,642	0,412	<i>62</i>	0,544	0,296
<i>7</i>	0,489	0,239	<i>65</i>	0,327	0,107
<i>9</i>	0,485	0,235	<i>66</i>	0,571	0,326
<i>10</i>	0,469	0,220	<i>69</i>	0,534	0,285
<i>11</i>	0,435	0,189	<i>70</i>	0,515	0,265
<i>13</i>	0,554	0,307	<i>71</i>	0,606	0,367
<i>14</i>	0,837	0,701	<i>73</i>	0,684	0,467
<i>18</i>	0,544	0,296	<i>75</i>	0,527	0,278
<i>19</i>	0,521	0,271	<i>76</i>	0,325	0,106
<i>22</i>	0,560	0,314	<i>82</i>	0,655	0,429
<i>24</i>	0,759	0,576	<i>87</i>	0,738	0,545
<i>25</i>	0,567	0,322	<i>88</i>	0,534	0,285
<i>28</i>	0,428	0,183	<i>89</i>	0,772	0,595
<i>31</i>	0,378	0,143	<i>91</i>	0,451	0,203
<i>32</i>	0,702	0,492	<i>95</i>	0,518	0,268
<i>36</i>	0,297	0,089	<i>99</i>	0,438	0,192
<i>41</i>	0,384	0,147	<i>102</i>	0,550	0,303
<i>43</i>	0,647	0,419	<i>103</i>	0,529	0,280
<i>44</i>	0,346	0,120	<i>104</i>	0,400	0,160
<i>47</i>	0,520	0,271	<i>107</i>	0,502	0,252
<i>49</i>	0,416	0,173	<i>113</i>	0,387	0,150
<i>52</i>	0,436	0,190	<i>117</i>	0,464	0,215
<i>53</i>	0,520	0,270	<i>120</i>	0,553	0,305
<i>54</i>	0,662	0,438			

Fonte: Elaborada pelo autor.

Nota: 2 itens da prova anulados (n^{os} 23 e 46). Base de Dados: 1.120 candidatos.

CF :Carga Fatorial h₂: Comunalidade.

As TAB. 11 e 12 mostram, respectivamente, as medidas das cargas fatoriais (CF) das comunalidades (h₂) do modelo TRI de três parâmetros, inicial (118 itens) e final (49 itens), com base nos parâmetros da análise fatorial apresentados no final da TAB. 6. Os resultados mostram que existem Cargas Fatoriais com valores abaixo ou muito abaixo de 0,400 (CF <

0,40, valor de referência, considerando-se o total de candidatos avaliados), que podem variar de -1 a +1. Essas cargas representam as correlações entre a variável latente (traço latente (θ), conhecimento do candidato na TRI) e cada uma das variáveis (itens da prova) que a compõem, sendo que valores grandes, em módulo, indicam que os traços latentes estão estreitamente correlacionados.

Já a medida de comunalidade de cada um dos itens tem o objetivo avaliar a adequabilidade do fator (traço latente) gerado, ou seja, o quanto cada um dos itens originais compartilham com os demais itens, com o conjunto de itens (118) avaliados no modelo inicial da TRI. Valores muito abaixo ou abaixo de 0,30 ($h^2 < 0,30$) indicam itens com baixa capacidade de gerar informação para o traço latente (θ) obtido pelo candidato. Resumindo, valores de Cargas Fatoriais próximos ou abaixo de 0,40 e valores de comunalidade (h^2) próximos ou abaixo de 0,30 são indicativos de que o item pouco ou nada contribui para o desempenho do candidato no escore (nota: proficiência do candidato) final obtido na prova.

Esses parâmetros apresentados nas tabelas 10, 11 e 12 são medidas básicas utilizadas na validação e avaliação da precisão do modelo TRI gerado para os dados estudados, os 118 itens válidos da prova para o Título de Especialista em Cardiologia, aplicada em 2019, pela Sociedade Brasileira de Cardiologia.

6 DISCUSSÃO

O objetivo principal do presente estudo foi avaliar os itens da prova para obtenção do Título de Especialista em Cardiologia, aplicada no ano de 2019, em relação aos seus parâmetros psicométricos (TCT e TRI) que, até então, eram desconhecidos. Utilizaram-se 118 itens validados (dos 120 propostos inicialmente dois foram anulados na fase de recurso), que foram respondidos por 1120 candidatos que fizeram a prova. Em relação aos parâmetros psicométricos estudados, o único cenário, até então, avaliado pela CJTEC e que se aproximava de um desses parâmetros era o grau de dificuldade dos itens. No entanto, eles foram julgados de forma subjetiva, como fáceis, médios ou difíceis, tendo por base o conhecimento prévio e a experiência de cada integrante da comissão que participou da formatação da prova. Essa forma de avaliar os itens mostrou-se totalmente diferente do modelo de avaliação técnica proposto pelos principais autores da literatura psicometrista.

Esse novo cenário de análise psicométrica trouxe resultados inéditos e importantes para o ambiente técnico da elaboração de itens na prova de título da SBC, assim como já estudado e confirmado em outras provas avaliadas a partir desses parâmetros, considerando o cenário nacional e outras áreas do conhecimento. Até onde se sabe, este é o primeiro estudo a avaliar as características psicométricas de uma prova de título de especialista da AMB e seus resultados contribuirão para reflexões e aprimoramentos desses importantes instrumentos certificadores. Por esse motivo, não foram localizadas referências bibliográficas que permitissem comparar os resultados encontrados com os de outras sociedades de especialidades, mas há publicações disponíveis em outros cenários.

Como exemplo, temos a análise dos itens da prova realizada na primeira fase da Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP), aplicada em uma escola de Joinville, no Estado de Santa Catarina, no Brasil, no ano de 2016. Por se tratar de uma prova específica na área de matemática, além de se caracterizar como uma competição nacional, a expectativa inicial em relação à prova era de que ela continha itens bem distribuídos e com bom poder de discriminação para o evento proposto. No entanto, após a análise psicométrica, constatou-se que, na prova de vinte itens, eles não estavam distribuídos de forma adequada em relação ao grau de dificuldade sugerido pelos principais autores psicometristas que utilizam a TCT. A prova em questão apresentava um maior número de itens com graus de dificuldade "muito fácil" e "fácil". Por fim, foi constatado que a prova apresentava 11 itens deficientes, tendo em

vista os parâmetros psicométricos. (KNÜPFER; AMARAL; HENNING, 2016). Pasquali (2003) recomenda que a distribuição ideal dos itens de uma prova, conforme o grau de dificuldade, na TCT, apresente a seguinte proporção: 10% de “muito fáceis”, 20% de “fáceis”, 40% de itens “moderados”, 20% de “difíceis” e 10% de itens “muito difíceis” (VILARINHO, 2015).

Neste estudo, a análise, pela TCT, do grau de dificuldade nos mostra uma proporção próxima à ideal, porém com muitos itens fáceis e uma proporção de até 50% de itens medianos, abaixo do considerado ideal. Essa distribuição nos leva a refletir sobre a necessidade da construção de uma prova mais balanceada, segundo um dos parâmetros da TCT. Esse achado é interessante, uma vez que a impressão geral dos candidatos e de alguns integrantes da SBC é de que a prova do TEC esteja sendo construída com elevado grau de dificuldade, o que justificaria a baixa aprovação em alguns concursos. Entretanto, a análise pela TCT não confirma essa opinião e, ao contrário, demonstra que o instrumento está relativamente bem balanceado em relação à distribuição de dificuldade dos itens, com uma tendência em direção a um instrumento de avaliação mais fácil do que se pensava.

Outro componente importante de uma prova é o poder de discriminação de seus itens. Oliveira (2017), por exemplo, utilizou a análise psicométrica da medida do componente de formação geral da prova do Exame Nacional de Desempenho dos Estudantes (ENADE), nos anos de 2010, 2011 e 2012, e pôde identificar questões com baixo poder de discriminação, contribuindo de forma técnica para possíveis melhorias na elaboração futura de novos itens para as provas do ENADE (OLIVEIRA, 2017).

Neste estudo, quando avaliado o parâmetro de discriminação dos itens pela TCT, pôde-se observar uma alta porcentagem de itens com indicação para serem substituídos e/ou revisados. Nesse sentido, ressaltamos a importância do emprego da análise dos parâmetros psicométricos em uma prova, os quais apresentam medidas matemáticas, contribuindo para que melhorias possam ser feitas, favorecendo o aprimoramento do instrumento de avaliação.

Apesar de trazer informações importantes, a TCT tem as limitações já descritas, como, por exemplo, o fato de não permitir a comparação de candidatos que tenham realizado provas diferentes ou o acompanhamento de um mesmo candidato. Além disso, há limitações em relação à estimação dos parâmetros dos itens e ao escore dos candidatos. Sendo assim, agrega-

se, no estudo, a análise dos resultados pela TRI que, por sua complexidade, trouxe respostas que não seriam possíveis somente com a TCT.

Conforme análise estatística em relação ao poder de discriminação, de acordo com a TRI, a prova de 2019 da SBC apresenta-se com 32,2% de itens com bom poder de discriminação e 18,7% de itens insuficientes. Quando comparados os resultados das duas teorias (TCT e TRI), no presente estudo, foi possível verificar, claramente, que alguns itens, apesar de serem adequados pela TCT, quando analisados de forma conjunta com a TRI, apresentaram um resultado ruim, de modo geral, sugerindo, por exemplo, sua reformulação ou exclusão da prova.

Com a associação da análise pela TRI, permitiu-se, ao final, concluir, estatisticamente, quais candidatos, baseados no traço latente, seriam ou não aprovados, caso essa teoria fosse considerada como método de ajuste para a correção e a classificação final da prova, conforme a população estudada. Cabe ressaltar que essa análise estatística foi feita após a realização da prova pelos candidatos, embora o ideal fosse pré-testar os itens da prova antes da avaliação da proficiência dos candidatos ao Título de Especialista em Cardiologia.

Essa possibilidade de pré-testar permite construir uma prova com itens “calibrados” e adequados em relação aos parâmetros psicométricos, incluindo, dessa forma, itens com bom poder de discriminação, com distribuição mais adequada em relação à dificuldade e baixa probabilidade de acerto ao acaso (chute). A presente reflexão convoca para a discussão do modelo de confecção e montagem da prova atualmente, que reflete uma situação em que os itens são elaborados por um conjunto heterogêneo de pessoas, que não discutiram a prova como um único instrumento, ou seja, o instrumento avaliativo não foi pensado como uma história com princípio, meio e fim.

Apesar dos itens que constituem o banco de dados deste estudo terem sido confeccionados por diversos autores e submetidos à avaliação e ao julgamento da CJTEC, a retirada, por exemplo, de 69 itens, pelo modelo da TRI, da prova realizada no ano de 2019, demonstra a oportunidade que este estudo representa, no sentido de contribuir para reflexão e aprimoramento da prova do TEC.

Quando analisado o parâmetro de acerto ao acaso, de forma geral, existe redução do percentual de candidatos que, mesmo com baixíssima proficiência, acertariam itens da prova. A redução

foi de 30,5% no modelo inicial (com 118 itens) para 8,2% no modelo final, após ajuste (com 49 itens). A redução significativa desse percentual deve-se à diminuição de itens respondidos corretamente com base no acerto ao acaso ("chute"), sendo um resultado relevante para o modelo final da TRI.

Esses dados contribuem para que a CJTEC possa, inclusive, avaliar o número de questões que são necessárias na prova do TEC, uma vez que, pela TRI, um modelo ajustado com 49 itens apresentou os mesmos resultados certificadores. A vantagem da possibilidade de redução do número de questões de uma prova, desde que orientada por métodos estatísticos como a TRI/TCT, é o resultado final de uma avaliação capaz de discriminar com maior precisão os candidatos merecedores do TEC e oferecer ao candidato um instrumento de avaliação menos cansativo, reduzindo a carga cognitiva para quem faz a prova e favorecendo, inclusive, um melhor desempenho dos candidatos que estejam bem preparados.

Por fim, cabe destacar que, se fosse adotada a metodologia da TRI como critério final de aprovação dos candidatos, 16,2% dos aprovados (74 candidatos em 456 aprovados pela pontuação total obtida pelo candidato em 120 itens) não seriam os mesmos. Ou seja, pelo ajuste baseado na TRI, que busca resultados mais coerentes, se considerados candidatos com um traço latente maior, aqueles que passaram acertando, por exemplo, questões ao acaso, não seriam aprovados, dando lugar a candidatos que responderam aos itens de forma mais coerente em relação aos parâmetros estudados.

Com base nos dados encontrados neste estudo, seguindo também as tendências de outros estudos envolvendo análises psicométricas, essa metodologia pode incrementar, de forma impactante, a qualidade dos testes aplicados atualmente nas diversas instituições, seja no âmbito das provas de título de especialista, seja nos diversos testes aplicados do ensino médio ao superior.

É importante destacar que, ainda hoje, a TRI está em contínuo desenvolvimento e, após ser adotada no principal exame de avaliação educacional do país, o Exame Nacional do Ensino Médio (ENEM), ela tem atraído o olhar de inúmeros pesquisadores que estudam e trabalham com processos de avaliação em larga escala. A AMB não possui estudos de provas de títulos com essa metodologia para comparação, sendo este um dos primeiros estudos realizados, o que reforça a necessidade de otimizar a construção de provas, de forma técnica e abrangente, no

sentido de trazer itens bem elaborados, conforme os parâmetros psicométricos.

A adoção desses parâmetros permitirá a elaboração de itens com bom poder para aprovar aqueles candidatos merecedores de um título de especialista, ou seja, candidatos bem avaliados e que detém o conhecimento esperado para determinada área. Nessa busca, os avanços na psicometria têm sido cada vez mais incorporados na elaboração e análise dos itens, de modo que grandes instituições têm se esforçado para ampliar seu horizonte, com novos recursos e métodos de avaliação.

Cabe ressaltar que o presente estudo apresenta algumas limitações, tais como o fato de que, para a obtenção de melhores resultados pela TRI, é desejável a existência de um banco de dados com itens previamente utilizados e calibrados. Isso não foi possível, uma vez que este é o primeiro estudo realizado em uma prova do TEC e, provavelmente, também de um exame para concessão do título de especialista da AMB. Outra limitação diz respeito ao banco de dados avaliado. O presente estudo considerou a prova realizada no ano de 2019, entretanto cada uma das edições da prova do TEC é construída de forma independente e, apesar de seguirem a mesma metodologia de elaboração, a equipe objetiva que as provas utilizem questões inéditas, de modo que não é possível afirmar que os resultados aqui apresentados possam ser extrapolados para os concursos anteriormente realizados. Apesar disso, o estudo apresenta importantes contribuições para que a Sociedade Brasileira de Cardiologia, representada pela CJTEC, e a Associação Médica Brasileira possam aprimorar os instrumentos certificativos para concessão de títulos de especialistas no Brasil, com repercussões diretas na qualidade da assistência médica brasileira.

Como perspectivas futuras, o trabalho proporciona a oportunidade de que a SBC, uma das mais importantes sociedades de especialidades do Brasil, com toda a sua grandeza e busca contínua por inovações e melhorias, possa, a partir dos resultados obtidos, aprimorar a metodologia de elaboração da prova do TEC. O debate sobre a questão pode favorecer a promoção de capacitação para o grupo de especialistas envolvidos na formulação e confecção dos itens da prova, bem como instrumentalizar o trabalho da distinta CJTEC, de modo a permitir que a prova do TEC, instrumento responsável pela certificação do Especialista em Cardiologia no Brasil, com impactos diretos na saúde pública desse país, possa ser elaborada dentro dos parâmetros psicométricos modernos e validados pelos grandes estudos sobre avaliação.

7 CONCLUSÃO

O presente estudo permitiu verificar que, em relação às propriedades psicométricas da prova do TEC de 2019, tanto em relação à TCT quanto à TRI, o exame apresentou questões de moderado a fácil grau de dificuldade, de forma geral. Baseado na TRI, mostrou cerca de um terço de itens com alto poder de discriminação, estando os demais itens sujeitos à readequação, com possibilidades de melhoria em relação a sua elaboração, pois apresentaram alto índice de itens com probabilidade de acerto ao acaso.

REFERÊNCIAS

- ARMSTRONG, B. K.; WHITE, E.; SARACCI, R. **Principles of exposure measurement in epidemiology**. Oxford: Oxford University Press, 1994.
- ASSOCIAÇÃO MÉDICA BRASILEIRA. **Normativa de Regulamentação para Obtenção de Título de Especialista ou Certificado de Área de Atuação**. São Paulo, 2004.
- BRASIL. **Decreto Presidencial nº 80.281**, de 5 de setembro de 1977. Regulamenta a Residência Médica, cria a Comissão Nacional de Residência Médica e dá outras providências. Diário Oficial, Brasília, DF, 06/set/1977.
- BRASIL. **Lei 6.932**, de 7 de julho de 1981. Dispõe sobre as atividades do médico residente e dá outras providências. Brasília, 1981.
- BRASIL. Ministério da Educação. **Resolução Conselho Nacional de Residência Médica nº 005/2003**, de 23 de dezembro de 2003. Registro dos certificados de Programas de Residência Médica, 2003.
- BOWMAN, M. L. Testing individual differences in ancient China. **American Psychologist**, [S.l.], n. 44, n. 3, p. 576-578, 1989.
- CONSELHO FEDERAL DE MEDICINA. **Resolução CFM Nº 2.162/2017**. Homologa a Portaria Comissão Mista de Especialidades nº 1/2017 que atualiza a relação de especialidades e áreas de atuação médicas aprovadas pela Comissão Mista de Especialidades. Brasília, 2017.
- CONSELHO FEDERAL DE MEDICINA. **Código de ética médica: resolução CFM nº 2.217/2018**, de 01 de novembro de 2018. Conselho Federal de Medicina, 2018.
- DEVELLIS, R. F. Classical Test Theory. **Medical Care**, [S.l.], v. 44, n. 11, p. 50-59, 2006.
- GOMES, L. D. S. **Teoria de resposta ao item na avaliação em larga escala: um estudo sobre o Exame Nacional de Acesso do Mestrado Profissional em Matemática em Rede Nacional – PROFMAT. 2012. 95f. Dissertação (Mestrado profissional em Matemática) - Instituto Nacional de Matemática Pura e Aplicada, Universidade Federal de Campina verde. Rio de Janeiro, 2012.**
- GRÉGOIRE, J.; LAVEAULT, D. **Introdução às teorias dos Testes em Ciências Humanas**. Porto: Porto Editora, 2002.
- HUTZ, C. S.; BANDEIRA, D. R.; TRENTINI, C. M. **Psicometria**. Porto Alegre: Artmed, 2015.
- KNÜPFER, R. E. N.; AMARAL, A.; HENNING, E. Análise Clássica de Testes: uma proposta de análise de desempenho dos estudantes na primeira fase da OBMEP. *In: Colóquio Luso-brasileiro de Educação. 2., 2016. Joinville. Anais [...]. Joinville: UFSC, 2016.*
- NORONHA, A. P.; PINTO, P. P.; OTTATI, F. Análise fatorial confirmatória da Escala de Aconselhamento Profissional. **Periódicos eletrônicos em Psicologia**, Rio de Janeiro, v. 68, n. 1, p. 62-71, abr., 2016. Disponível em http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S180952672016000100006&lng=pt&nrm=iso. Acessos em: 20 nov. 2020.

OLIVEIRA, A. L. S. **Avaliação psicométrica da medida do componente de formação geral da prova do exame nacional de desempenho de estudantes (ENADE) de 2010, 2011 E 2012.** 2017. 107f. Dissertação (Mestrado em Gestão em Avaliação) – Universidade Federal de Santa Catarina, Florianópolis, 2017.

PASQUALI, L. **Psicometria: Teoria dos Testes na Psicologia e na Educação.** Petrópolis: Vozes, 2003.

PORTNEY, L. G.; WATKINS, M. P. **Power analysis and determination of sample size: foundations of clinical research: applications to practice.** Norwalk, CT: Applenton & Lange, 1993.

RABELO, M. **Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro.** Rio de Janeiro: SBM, 2013.

SARTES, L. M. A.; SOUZA-FORMIGONI, M. L. O. **Avanços na Psicometria: Da Teoria Clássica dos Testes à Teoria de Resposta ao Item.** São Paulo, 2013.

SOCIEDADE BRASILEIRA DE CARDIOLOGIA. **Regimento da Comissão de Julgamento do Título de Especialista em Cardiologia da Sociedade Brasileira de Cardiologia CJTEC.** Rio de Janeiro, 2018. Disponível em: <https://www.cardiol.br/Portal/Regimento>. Acesso em: 10 set. 2019.

SOUSA, M. R. *et al.* I Diretriz da Sociedade Brasileira de Cardiologia sobre Processos e Competências para a Formação em Cardiologia no Brasil. **Arquivos Brasileiros de Cardiologia**, Rio de Janeiro, v. 96, n. 5, Supl.1, Maio 2011.

VILARINHO, A. P. L. **Uma proposta de análise de desempenho dos estudantes e de valorização da primeira fase da OBMEP.** 2015. 98f. Dissertação (Mestrado profissional em Matemática) – Instituto de Ciências Exatas, Universidade de Brasília, Brasília 2015.