



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

PAULO ROBERTO SILVA

SETE ANOS DE TESTE DE PROGRESSO NO CURSO DE MEDICINA DE UMA
ESCOLA PÚBLICA DO DF: QUAIS AS LIÇÕES APRENDIDAS?

BRASÍLIA – DF

JUNHO/2022

PAULO ROBERTO SILVA

SETE ANOS DE TESTE DE PROGRESSO NO CURSO DE MEDICINA DE UMA
ESCOLA PÚBLICA DO DF: QUAIS AS LIÇÕES APRENDIDAS?

Tese apresentada para obtenção do título de
Doutor em Ciências da Saúde pelo Programa de
Pós-Graduação em Ciências da Saúde da
Universidade de Brasília

Orientadora: Profa. Dra. Nilce Santos de Melo

BRASÍLIA – DF

JUNHO/2022

PAULO ROBERTO SILVA

SETE ANOS DE TESTE DE PROGRESSO NO CURSO DE MEDICINA DE UMA
ESCOLA PÚBLICA DO DISTRITO FEDERAL, QUAIS AS LIÇÕES APRENDIDAS?

Tese apresentada para obtenção do título de Doutor em Ciências da Saúde pelo
Programa de Pós-Graduação em Ciências da Saúde da Universidade de Brasília

BANCA EXAMINADORA

.....
Profª. Dra. Nilce Santos de Melo (UnB)
Presidente

.....
Pra. Dra. Lilian Marly de Paula (UnB)
Examinadora Interna

.....
Prof. Dr. Fábio Ferreira Amorim (ESCS)
Examinador Externo

.....
Prof. Dr. Wilton Ferreira Silva Santos (ESCS)
Examinador Externo

.....
Prof. Dr. André Ferreira Leite UnB)
Examinador Interno

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de ensino, estudo ou pesquisa, desde que citada a fonte.
Catalogação da Publicação

SS586s Silva, Paulo Roberto
SETE ANOS DE TESTE DE PROGRESSO NO CURSO DE MEDICINA DE
UMA ESCOLA PÚBLICA DO DF: QUAIS AS LIÇÕES APRENDIDAS? / Paulo
Roberto Silva; orientador Nilce dos Santos Melo. --
Brasília, 2022.
71 p.

Tese (Doutorado - Doutorado em Ciências da Saúde) --
Universidade de Brasília, 2022.

1. Educação médica. 2. Avaliação. 3. Teste de progresso.
4. Ações afirmativas. 5. Sistema de cotas. I. Melo, Nilce
dos Santos, orient. II. Título.

AGRADECIMENTOS

Agradeço a todos que de alguma forma, direta ou indiretamente, contribuíram nesta caminhada tornando possível a realização deste trabalho.

Agradeço a meus pais Baltazar Henrique da Silva, Ilca Braz de Queiroz e aos meus filhos Luisa Vieira Silva, Bruno Vieira Silva e Luca Fonseca de Castro Silva pela paciência e pelo amor que me dedicam.

Agradeço aos professores Sulani Silva de Souza, Ana Márcia Yunes Gaudard, Wilton da Silva Santos e Vanessa Viana Cardoso pela amizade e pelos conhecimentos compartilhados.

Agradeço aos professores Ana Tereza Bittencourt Guimarães, Girlene Ribeiro de Jesus, José Antônio Freitas, Maria Eduarda Tannuri-Pianto, Mauro Luiz Rabelo, Roberta Teodoro Santos pela colaboração na realização deste trabalho.

Agradeço à professora Nilce dos Santos Melo por não ter medido esforços para a realização deste trabalho.

Meus agradecimentos também aos professores Mourad Ibrahim Bellaciano e Maria Dilma Alves Teodoro pela inestimável experiência na gestão educacional.

Agradeço também aos colegas da Gerência de Avaliação da ESCS Evilásio Cortez, Marcelo Vinicius Pimenta e José Carlos Pires de Sá pelo companheirismo e por tornarem nosso cotidiano mais leve.

Finalmente, agradeço a todos os estudantes da ESCS que deram sentido a realização deste trabalho.

RESUMO

SILVA, Paulo Roberto. **Sete anos de teste de progresso no curso de medicina de uma escola pública do Distrito Federal, quais as lições aprendidas?** Brasília, 2022. Tese (Doutorado em Ciências da Saúde) – Faculdade de Ciências da Saúde, Universidade de Brasília, Brasília, 2022.

Introdução: O Teste de progresso (TP) é um instrumento de avaliação do tipo melhor resposta escolhida construído com o objetivo de avaliar, de maneira ampla, os conhecimentos esperados ao final de um curso de uma maneira global e sistematizada. A Escola aplica o teste regularmente desde 2005 e a partir de 2013 passou a realiza-lo em conjunto com as escolas participantes do Consórcio do Centro Oeste para o Teste de Progresso. **Objetivos:** 1) Analisar as características psicométricas dos testes pela Teoria Clássica dos Testes. 2) Analisar a progressão do desempenho dos estudantes correlacionando-o com a série cursada, gênero, participação nos programas de iniciação científica, reprovações no curso e forma de acesso. 3) Analisar a possibilidade da utilização do teste de progresso para feedback individual do estudante. 4) Descrever os processos de produção, aplicação, análise e feedback dos testes de progresso realizados pelo consórcio. 5) Relacionar a qualidade dos testes produzidos pelo consórcio com seu processo de produção. **Métodos:** Foi realizada uma análise retrospectiva dos dados existentes sobre o teste de progresso, correlacionando-os com as variáveis gênero, tipo de admissão, reprovação no curso e participação nos programas de iniciação científica. A confiabilidade dos testes foi calculada pelo alfa de Cronbach e a análise psicométrica dos itens foi realizada utilizando a teoria clássica dos testes (TCT). Para verificar a confiabilidade das informações dos testes em nível individual, foi realizada a análise pelo índice de mudança confiável ou *reliable change index* (RCI). Para analisar a progressão de conhecimentos dos estudantes foi realizada uma regressão linear múltipla clássica e calculado o tamanho do efeito de cada série sobre a progressão de conhecimentos pelo *d* de Cohen. Os processos de produção do teste, aplicação, análise dos resultados e feedback para os envolvidos foi analisada tendo como referencial teórico a estrutura proposta pelo *guideline* da *Association for Medical Education in Europe* (AMEE). **Resultados:** O alfa de Cronbach foi em média 0,88, a análise psicométrica revelou uma dificuldade média de 0,51 e discriminação média dos itens de 0,24. A

análise evidenciou um aumento de desempenho aproximadamente linear ao longo do curso, porém observou-se uma diminuição da taxa de crescimento no período correspondente ao estágio Curricular Obrigatório. Não foi observada correlação estatisticamente significativa entre gênero e desempenho no TP, os estudantes que participaram do programa de iniciação científica obtiveram um desempenho superior no teste quando comparados aos não participantes, enquanto a reprovação no Curso correlacionou-se a um pior desempenho. Os estudantes admitidos pelo sistema de cotas obtiveram um desempenho significativamente inferior ao dos seus colegas admitidos pelo sistema universal. Todas as etapas necessárias para realização do teste foram feitas de acordo com referencial teórico. O cálculo do RCI mostrou que a progressão individual de conhecimentos pode ser evidenciada com confiabilidade em apenas 14,7% dos casos. **Conclusões:** O teste de progresso mostrou-se válido e confiável para aferir a progressão de conhecimentos dos estudantes fornecendo informações relevantes para a gestão do Curso, no entanto, os resultados do teste não podem ser utilizados para o feedback individual. O Consórcio do Centro Oeste para o Teste de Progresso cumpriu adequadamente todas as etapas preconizadas para realização do teste de progresso de acordo com o *guideline* utilizado como referencial. A participação da Escola junto a consórcio do Centro Oeste mostrou-se uma estratégia efetiva e eficiente para a realização do Teste de Progresso propiciando a obtenção de importantes informações para a gestão educacional.

PALAVRAS CHAVE: Educação Médica, Graduação, Teste de Progresso, Conhecimento, Avaliação Educacional, Ações afirmativas.

ABSTRACT

SILVA, Paulo Roberto. **Seven years of progress testing in a public medical school in Federal District: What lessons learned?** Brasília, 2022. Thesis (Doctorate in Health Sciences) – Faculdade de Ciências da Saúde, Universidade de Brasília, Brasília, 2022.

Introduction: The Progress Test (PT) is an assessment tool of best single answer type constructed with the aim of broadly assess the knowledge expected at the end of the Course in a global and systematic way. The School has applied the test regularly since 2005 and from 2013 it began to carry it out in conjunction with the participating schools of the Midwest Consortium for the Progress Test. **Objectives:** 1) To analyze the psychometric characteristics of the tests using the Classical Test Theory. 2) Analyze the progression of students' performance correlating it with the grade taken, gender, participation in scientific initiation programs, failures in the course and type of access. 3) Analyze the possibility of using the progress test to provide individual student feedback. 4) Describe the production, application, analysis and feedback processes of the progress tests carried out by the consortium. 5) Relate the quality of tests produced by the consortium to its production process. **Methods:** A retrospective analysis of the existing data on the progress test was performed, correlating them with the variables gender, type of admission, failure in course and participation in scientific initiation programs. The reliability of the tests was calculated using Cronbach's alpha and the psychometric analysis of the items was performed using the classical test theory (TCT). In order to verify the reliability of the tests at an individual level, the analysis was performed using the reliable change index or reliable change index (RCI). To analyze the progression of students' knowledge, a classical multiple linear regression was performed and the size of the effect of each grade on the progression of knowledge was calculated using Cohen's d. The processes of production of the tests, application, analysis of results and feedback to stakeholders were analyzed using the guideline of the Association for Medical Education in Europe (AMEE) as a theoretical framework. **Results:** Cronbach's alpha averaged 0.88, psychometric analysis revealed an average difficulty of 0.51 and average item discrimination of 0.24. The analysis showed an approximately linear increase in performance throughout the

course, but a decrease in the growth rate was observed in the period corresponding to the internship. There was no statistically significant correlation between gender and performance in the TP, students who participated in the scientific initiation program performed better on the test when compared to non-participants, while failure in the Course was correlated with a worse performance. Students admitted through the affirmative action system performed significantly below than their peers admitted through the universal system. All the steps used to perform the test were carried out in accordance with the theoretical framework. The ICR calculation showed that individual progression of knowledge can be reliably evidenced in only 14.7% of cases.

Conclusions: The progress test proved to be valid and reliable to measure the progression of students' knowledge, providing information relevant to the course management, however, the test results cannot be used for individual feedback. The Midwest Consortium for the Progress Test adequately complied with all the recommended steps to carry out the progress test according to the guideline used as a reference. The School's participation in the Midwest consortium proved to be an effective and efficient strategy for carrying out the Progress Test, providing important information for educational management.

KEYWORDS: Undergraduate Medical Education, Progress Test, Knowledge, Educational Assessment, Affirmative action.

LISTA DE FIGURAS

Figura 1. Estrutura geral sistêmica para o teste de progresso.....28

Figura 2. Escores Z obtidos por cada série nos testes aplicados de 2013 a 2019,
Testes de 1 a 7.....31

LISTA DE TABELAS

Tabela 1. Média dos Escores Z e intervalos de confiança de primeira a sexta séries nos testes aplicados de 2013 a 2019.....	30
Tabela 2. Número de alunos, quantidade de itens válidos, confiabilidade dos testes, nível de dificuldade e discriminação dos itens nos testes aplicados de 2013 a 2019.....	32
Tabela 3. Número de itens e de distratores utilizados nos testes de 2013 a 2019	33
Tabela 4. Número de estudantes participantes por edição do teste e matrículas ativas de 2012 a 2019	33
Tabela 5. Número de participações dos estudantes no teste.....	34
Tabela 6. Número de participações dos estudantes no programa de iniciação científica.....	34
Tabela 7. Número de reprovações dos estudantes.....	35
Tabela 8. Resultados da regressão logística realizando a comparação dos escores por série do Curso, por gênero, pela ocorrência de reprovação, por participação no programa de iniciação científica e por forma de acesso.....	36
Tabela 9. Descrição das etapas realizadas pelo TPCO na realização do teste de progresso.....	41

LISTA DE GRÁFICOS

- Gráfico 1.** Desempenho por série nos testes de progresso de 2012 a 2019 e linha de projeção linear do crescimento.....36
- Gráfico 2.** Tamanho do efeito de cada série sobre o aumento de conhecimento.....37
- Gráfico 3.** Desempenho dos estudantes admitidos pelo sistema de cotas e pelo sistema universal nos testes de progresso aplicados no período de 2012 a 201938
- Gráfico 4.** Histograma de frequência do RCI para todos os PT consecutivos realizados pelo mesmo aluno39

LISTA DE ABREVIATURAS E SIGLAS

ABEM	Associação Brasileira de Escolas Médicas
AMEE	Association for Medical Education in Europe
ABP	Aprendizagem baseada em problemas
PBL	Problem based learning
CNE	Conselho Nacional de Educação
ECO	Estágio Curricular Obrigatório
ESCS	Escola Superior de Ciências da Saúde
FAMEMA	Faculdade de Medicina de Marília
PIBIC	Programa Institucional de Bolsas de Iniciação Científica
PIC	Programa de Iniciação Científica da ESCS
PISA	Programa Internacional de Avaliação de Estudantes
RCI	<i>Reliable change index</i>
SES-DF	Secretaria de Estado de Saúde do Distrito Federal
SISU	Sistema de Seleção Unificada
SUS	Sistema Único de Saúde
TCT	Teoria clássica dos testes
TP	Teste de Progresso
TPCO	Consórcio do Teste de Progresso das Escolas de Medicina do Centro-Oeste
TRI	Teoria de Resposta ao Item
UEL	Universidade Estadual de Londrina
UnB	Universidade de Brasília

Sumário

1. INTRODUÇÃO	11
2. JUSTIFICATIVA	21
3. OBJETIVOS	21
4. MÉTODOS	22
4.1 CONTEXTO DO ESTUDO.....	22
4.2 ANÁLISE DAS CARACTERÍSTICAS PSICOMÉTRICAS DOS TESTES PELA TEORIA CLÁSSICA DOS TESTES.....	25
4.3 ANÁLISE DO DESEMPENHO DOS ESTUDANTES NOS TESTES	26
4.4 ANÁLISE DA UTILIZAÇÃO DO TESTE PARA FEEDBACK INDIVIDUAL.....	27
4.5 ANÁLISE DOS PROCESSOS DE PRODUÇÃO, APLICAÇÃO, ANÁLISE DOS RESULTADOS E FEEDBACK DOS TESTES.....	27
5. ASPECTOS ÉTICOS	29
6. RESULTADOS	29
6.1 CARACTERÍSTICAS PSICOMÉTRICAS DOS TESTES.....	29
6.2 ANÁLISE DO DESEMPENHO DOS ESTUDANTES NOS TESTES	33
6.3 UTILIZAÇÃO DO TESTE PARA FEEDBACK INDIVIDUAL	38
6.4 DESCRIÇÃO DO PROCEDIMENTO UTILIZADO NA REALIZAÇÃO DOS TESTES 39	
7. DISCUSSÃO	42
7.1 CARACTERÍSTICAS DOS TESTES.....	42
7.2 DESEMPENHO DOS ESTUDANTES.....	45
7.3 UTILIZAÇÃO DO TESTE PARA FEEDBACK INDIVIDUAL	49
7.4 ANÁLISE DOS PROCEDIMENTOS UTILIZADOS PARA A REALIZAÇÃO DO TESTE.....	50
7.4.1 Construção do teste	50
7.4.2 Aplicação.....	52
7.4.3 Análise dos resultados e revisão (Controle de qualidade).....	52
7.4.4 Feedback para os envolvidos	53
8. CONCLUSÕES	55
9. REFERÊNCIAS	60

1. INTRODUÇÃO

É amplamente aceito no meio educacional que a avaliação dirige a aprendizagem (NEWBLE; JAEGER, 1983; NORMAN, 2010; NEWBLE, 2016). De uma maneira geral, os estudantes direcionam sua estratégia de estudo conforme as regras da avaliação, o que vai influenciar diretamente no processo de aprendizagem. Muito frequentemente, as avaliações da aprendizagem dos estudantes consistem em provas aplicadas durante e/ou ao final das disciplinas, de forma que esse formato de avaliação tende a induzir uma estratégia de estudo concentrada no período logo anterior à prova quando uma grande quantidade de informações é memorizada visando a avaliação (CILLIERS, 2012). Esta estratégia de estudo, baseada muito mais na memorização do que no raciocínio e na compreensão (AL-KADRI et al, 2012), faz com que boa parte do que foi memorizado para a prova seja esquecido num curto espaço de tempo. Portanto, as avaliações aplicadas durante e/ou ao final das disciplinas tende a aferir predominantemente a capacidade de memorização, avaliando principalmente os conteúdos armazenados na memória de curto prazo e não o conhecimento consolidado e baseado em redes explicativas.

As limitações deste formato de avaliação tornaram-se mais evidentes com o aparecimento de novas formas de organização curricular e metodologias de ensino, surgidas a partir do final dos anos 1960 como alternativa ao modelo Flexneriano, no ensino da medicina (Flexner, 1910). Um marco importante desta mudança no ensino médico ocorreu no final dos anos 1960 com a reforma curricular implementada no curso de medicina da Universidade de McMaster no Canadá (SPAUDING, 1969). Esta reformulação tinha dois objetivos fundamentais:

“1. Ajudar os estudantes a tornarem-se efetivos na resolução de problemas, preparando-os para entender os princípios essenciais da solução de problemas e ensinando-os a buscar e usar a informação necessária para sua solução”;

“2. Promover as atitudes que levem a um comportamento responsável como médicos e cientistas na sua relação com colegas e com a sociedade. Com comportamento marcado pela empatia e pela ação em benefício dos pacientes e na promoção do bem comum quando diante de decisões éticas.”

Estes dois objetivos seriam atingidos pelo contato com os pacientes desde o início do curso e pela discussão em pequenos grupos de problemas relevantes para prática, ao professor caberia o papel de facilitador do processo. Essa reforma curricular lançou as bases para o desenvolvimento da metodologia de Problem Based Learning (PBL) ou Aprendizagem Baseada em Problemas e impulsionou o desenvolvimento dos chamados currículos integrados (BRAUER; FERGUSON, 2015).

Os currículos integrados tem como princípios a quebra da barreira entre as ciências básicas e clínicas, a utilização de módulos interdisciplinares ao invés de disciplinas isoladas e a integração entre a teoria e a prática. Esta nova forma de organização curricular favoreceria a retenção do conhecimento e o maior desenvolvimento de habilidades clínicas.

Seguindo esta experiência pioneira da Universidade de MacMaster, as Universidades de Maastricht (Holanda), Universidade de Newcastle (Austrália) e a Universidade do Estado de Michigan (Estados Unidos) também passaram a utilizar o PBL ainda no início dos anos 1970 (HILLEN, SCHERPBIER, WIJNEN, 2010), esta nova proposta pedagógica se insere num grande movimento ocorrido no final do século XX, no sentido de propor reformas curriculares nos cursos graduação em medicina (AAMC, 1984; GMC,1993; CNE,2021).

Com o advento das novas metodologias de ensino, surgiu também a necessidade de uma nova forma de avaliação, mais adequada aos currículos interdisciplinares, que fosse mais abrangente e não apenas limitada à conteúdos programáticos específicos. Neste contexto, no início dos anos 1970, a Universidade de Missouri-Kansas (EUA) criou o chamado Quarterly Profile Examination (QPE) (ARNOLD; WILLOUGHBY,1990). O QPE era um teste de melhor resposta escolhida, de conteúdo abrangente, aplicado quatro vezes ao ano, que substituíria a avaliação individualizada das disciplinas. Uma avaliação com características similares passou também a ser adotado no curso de medicina da Universidade de Limburg-Maastricht (Holanda), que já adotava um modelo de currículo integrado e o PBL (VLEUTEN; VERWIJNEN; WIJNEN;1996). Em Maastricht, este formato de avaliação foi chamado de Teste de Progresso (TP), denominação com a qual ficou conhecido mundialmente.

O TP foi introduzido no Brasil pelas primeiras escolas a utilizar o PBL no país. A Faculdade de Medicina de Marília (FAMEMA) foi a escola pioneira desta metodologia em 1997 (LIMA, 2013), no ano seguinte, a Universidade Estadual de Londrina (UEL) também adotou o PBL como metodologia de ensino. Estas escolas logo iniciaram a utilização do teste de progresso (TOMIC et al., 2005; SAKAI et al., 2008; PINHEIRO et al., 2015; ROSA et al., 2016) com o objetivo de avaliar os seus programas educacionais. A expressiva ampliação do número de escolas médicas no Brasil, iniciada a partir dos anos 2000 (ANTUNES; NUNES, 2019) e a necessidade de avaliação da qualidade do ensino dos novos cursos constituíram-se também num importante fator para a expansão do TP no país.

O TP tem como característica principal a quebra do vínculo entre o que está sendo ensinado no momento e a avaliação. Trata-se de uma prova de melhor resposta escolhida construída para avaliar de maneira ampla os conhecimentos esperados ao final do curso. A prova é aplicada simultaneamente a todos os estudantes de um mesmo curso, de maneira sistemática e regular e possibilita a obtenção de uma série temporal de dados sobre a progressão dos conhecimentos dos estudantes ao longo do curso. Esta característica do TP impossibilita a adoção de uma estratégia de estudos de memorização de conteúdos direcionado para prova (MCHARG et al, 2005), fazendo com que o teste avalie, sobretudo, o conhecimento mais profundo e significativo incorporado em redes explicativas e armazenado na memória de longo prazo (SCHUWIRTH; VLEUTEN, 2012), ou seja, o TP avalia predominantemente a capacidade de evocação dos conhecimentos retidos por um longo período de tempo (CUSTERS, 2010). Normalmente, o resultado observado é um padrão linear de progressão do conhecimento ao longo do curso (VLEUTEN; VERWIJNEN; WIJNEN; 1996; VERHOEVEN et al, 2002; SCHUWIRT; VAN DER VLEUTEN, 2012; ALBANESE; CASE, 2016; TIO et al, 2016).

Historicamente vinculado ao ensino da medicina, aos currículos integrados e ao PBL, o TP surgiu como um tipo de avaliação global e sistematizada capaz de aferir os conhecimentos consolidados (VERHOEVEN *et al.*, 2002) que formarão a base para o desenvolvimento das competências profissionais, o TP obteve ampla aceitação no meio acadêmico sendo hoje adotado por inúmeras instituições de ensino em todo o mundo, em diversas áreas e por diferentes abordagens metodológicas (FREEMAN et al, 2010).

O TP mostrou-se um instrumento versátil para avaliação de conhecimentos, o que possibilitou a ampliação de suas aplicações (ALBANESE, 2008; COOMBES et al, 2010; ALBANESE; CASE, 2016). Além do seu uso inicial como instrumento de avaliação somativo, seus resultados podem ser usados também de maneira formativa, fornecendo feedback para o estudante quanto a progressão de conhecimentos e/ou lacunas de aprendizagem (MUIJTJENS et al, 2010; GIVEN et al, 2016). O TP também é usado como feedback para as instituições de ensino, ao informar sobre o desenvolvimento dos programas educacionais (COOMBES et al, 2010), constituindo-se em mais um elemento para o controle de qualidade dos programas educacionais (VLEUTEN; VERWIJNEN; WIJNEN;1996), contribuindo dessa forma para a gestão acadêmica (PINHEIRO, 2015; ROSA, 2016). Com a formação de grupos de escolas para sua aplicação em conjunto, o TP possibilitou a realização de comparações entre diferentes instituições de ensino (MUIJTJENS et. al, 2008a; FREEMAN et. al., 2010) bem como o estabelecimento de padrões de desempenho (ALBANO, et *al.*, 1996; VAN DER VLEUTEN, FREEMAN, COLLARES, 2018).

A utilização do TP envolve o domínio de uma série complexa de etapas que são necessárias para garantir a validade e confiabilidade dos resultados. Segundo WRIGLEY *et al*, (2012), o processo pode ser dividido em quatro grandes etapas, a saber: construção dos testes, aplicação, análise dos resultados, *feedback* aos envolvidos.

A primeira etapa na construção do teste é a elaboração de um *blueprint* (RAYMOND; GRANDE, 2019) que represente adequadamente as diretrizes curriculares nacionais e que também contemple os objetivos educacionais estabelecidos no projeto pedagógico do curso. A elaboração de um *blueprint* adequado é uma etapa fundamental que servirá de base para a construção da série de testes e garantir a comparabilidade entre eles.

A elaboração de uma grande quantidade de bons itens necessária para manter uma série regular de testes é uma tarefa complexa e gera uma carga de trabalho considerável (JOZEFOWICZ et al. 2002; STAGNARO-GREEN; DOWNING, 2006). É necessário que sejam elaborados itens dentro das especificações estabelecidas, que sejam capazes de discriminar os diferentes níveis de conhecimento e que avaliem níveis taxonômicos elevados.

Existe uma grande variedade de tipos de itens utilizados no TP. Algumas escolas adotam o modelo de melhor resposta escolhida entre duas, três, quatro, ou até cinco alternativas, outras incluem uma alternativa “não sei” para evitar o acerto ao acaso (MUIJTJENS *et al.*,1999). Itens do tipo verdadeiro ou falso também são utilizados por algumas escolas. A escolha do número e do tipo de itens que compõem o teste e sua frequência de aplicação vai depender dos seus objetivos (RICKETTS *et al.* 2010). De uma forma geral, testes com objetivos somativos, ou que serão usados para feedback individual vão requerer um maior número de itens, enquanto testes com objetivo de avaliar grupos de estudantes necessitam uma amostragem menor.

A etapa de aplicação do teste vai depender da disponibilidade de recursos e dos seus objetivos. Existem escolas que optam pela aplicação em meio eletrônico em lugar de, abrindo a possibilidade também de aplicação remota. Testes com objetivos apenas formativos podem ser aplicados de forma assíncronica, enquanto testes com objetivos somativos necessitam aplicação síncronica. O cuidado na aplicação dos testes é um fator importante para a garantia dos objetivos do teste e da confiabilidade dos resultados obtidos (ALBANESE; CASE, 2016).

A etapa de análise e revisão dos resultados compreende o método de cálculo dos escores obtidos e o controle de qualidade do teste. Os resultados podem ser expressos de diferentes formas: em número ou percentual de acertos, em escalas relativas como escore Z, em escores definidos por fórmulas como, por exemplo, respostas certas menos respostas erradas, (CECILIO-FERNANDES *et al.*, 2017), em desvios padrão cumulativos (SCHAUBER; NOUNS, 2010) ou mesmo utilizando-se técnicas bayesianas (RICKETTS; MOYEED, 2011). Existem restrições quanto ao uso da Teoria da Resposta ao Item (TRI) para a análise dos resultados do TP, devido a necessidade dos pré-requisitos de unidimensionalidade e independência local (CONDE; LAROS, 2007).

Após a obtenção do resultado deve ser realizado o controle de qualidade do teste aplicado. Diversos parâmetros devem ser analisados, sendo que uma confiabilidade adequada é uma pré-condição necessária para a validade do teste. A análise do nível de dificuldade e da discriminação dos itens são fundamentais para o controle de qualidade e para a construção de um banco de itens. Caso haja a necessidade de realizar a comparação dos resultados entre os testes, a variação do nível de dificuldade entre eles pode ser corrigida por procedimentos de equiparação

de escores (LANGER; SWANSON, 2010). Finalmente, os itens aplicados devem ser armazenados num banco de itens (WRIGHT. BELL, 1984) para futura reutilização, diminuindo assim a carga de trabalho da criação de novos itens.

A última etapa consiste em prover um feedback que seja útil e relevante para os estudantes, docentes, elaboradores de itens, comissão revisora, e para a gestão educacional. O feedback para o estudante é um dos objetivos principais de qualquer sistema de avaliação, principalmente no contexto das metodologias ativas que propõem o desenvolvimento da capacidade de aprendizado (NORMAN et al, 2010) de maneira a promover a autonomia e a capacidade de autorregulação da aprendizagem. Idealmente, o teste deve fornecer ao estudante informações quanto as áreas em que objetivos educacionais foram atingidos e as áreas onde existem lacunas na aprendizagem (MUIJTJENS et al, 2010).

A gestão acadêmica e o corpo docente devem receber feedback quanto ao desempenho dos estudantes no sentido de identificar os que necessitam de recuperação ou estão sob o risco de reprovação, porém, o principal *feedback* para a gestão educacional é obtido pela análise das curvas de progressão de conhecimentos dos estudantes, que torna possível identificar possíveis falhas na estrutura curricular e/ou na execução dos programas educacionais.

Os elaboradores dos itens também devem receber o feedback sobre o funcionamento das questões construídas, de maneira a motivar e promover o seu desenvolvimento na elaboração de itens.

À organização central do teste, cabe por sua vez analisar os resultados obtidos no teste e todo o processo envolvido na sua realização, para tanto, deve contar com uma equipe capacitada nas áreas educacional, de avaliação, em análise psicométrica e em tecnologia da informação.

A realização do TP é trabalhosa e consome uma quantidade de recursos considerável, por esta razão, ao longo do tempo, escolas que aplicavam isoladamente o teste se uniram formando consórcios de escolas visando o compartilhamento de recursos e a divisão de tarefas (ALBANO *et al.*, 1996; VAN DER VLEUTEN *et al*, 2004; VERHOEVEN *et al*, 2005; ALBANESE, 2008; FREEMAN *al*, 2010; RICKETTS *et al*. 2010; FINDYARTINI *et al*, 2017; VAN DER VLEUTEN, FREEMAN, COLLARES, 2018; BICUDO *et al.*, 2019). Esta cooperação entre instituições, além das vantagens

logísticas, possibilita também o estabelecimento de padrões de qualidade e a adoção de políticas de *benchmark* entre as instituições (MUIJTJENS, 2008).

O presente trabalho foi realizado num contexto de uma escola que utilização metodologias ativas de ensino e aprendizagem, mais especificamente o PBL. O paradigma central destas metodologias é a aprendizagem centrada no estudante, colocando-o como protagonista da sua própria aprendizagem. Em contraposição ao ensino tradicional, que visa principalmente a entrega de conteúdos da maneira mais eficiente possível, o PBL objetiva desenvolver a capacidade de aprendizado do estudante; capacitando-o a conduzir o seu processo de aprendizagem em diversos contextos ao longo da vida. No PBL espera-se que o estudante desenvolva a capacidade de autorregulação do seu processo de aprendizagem, conduza seu próprio processo educacional e ganhe progressivamente autonomia, à medida que se desenvolve no curso (HMELO-SILVER, 2004.).

A formação médica requer a construção de uma extensa base de conhecimentos e constitui-se num dos principais objetivos dos cursos de graduação. Os estudantes devem desenvolver ao longo do curso uma rede de conhecimentos, estruturados e densamente conectados por relações temporais, causais ou hierárquicas, armazenadas na memória de longo prazo (REGEHR. NORMAN, 1996). A psicologia cognitiva denomina estas redes de conhecimentos densamente conectadas e compactadas de *scripts* (CUSTERS, 2015), trata-se de representações mentais de alto nível que são ativadas conforme a necessidade durante o enfrentamento dos problemas comuns da prática médica, dando-lhe sentido, orientando as ações e propiciando a agregação de novos conhecimentos. A construção destes scripts se dá pela experiência acumulada na aplicação do conhecimento em diversas situações (BORDAGE, 1994) e formará a base para o desenvolvimento das competências profissionais.

O Processo de aprendizagem que possibilitará a construção desta rede de conhecimentos é resultante de uma complexa interação entre fatores individuais, familiares, socioeconômicos e também do tipo e qualidade do ensino oferecido.

SCHUMACHER et al (2013), descreveram uma estrutura conceitual que correlaciona os diversos fatores que vão influenciar no desenvolvimento da capacidade de aprendizado dos estudantes.

A motivação é apontada como um dos fatores individuais mais importantes para o aprendizado (KISURKAR *et al*, 2012). Ela pode ser dividida em motivação intrínseca, derivada do interesse próprio em uma atividade e extrínseca, derivada de um ganho ou resultado esperado, como por exemplo ser aprovado numa disciplina. Segundo a teoria da autodeterminação (RYAN; DECI, 2000), a motivação intrínseca estaria relacionada a um melhor aprendizado e maior sensação de bem estar pelo estudante. Ainda segundo estes autores, a motivação intrínseca teria três determinantes principais:

- Senso de pertencimento, isto é a sensação de pertencimento a um grupo, comunidade ou profissão;
- Senso de autonomia, que diz respeito à sensação de que o estudante está agindo por vontade própria e não por obrigações externas e;
- Senso de competência que é a sensação de que o próprio conhecimento está aumentando ou novas habilidades estão sendo adquiridas.

Outro fator importante no processo de aprendizagem é a percepção da autoeficácia acadêmica, que se refere às crenças e atitudes do estudante em relação à sua capacidade para alcançar o sucesso acadêmico. A percepção da autoeficácia incorpora elementos tanto cognitivos quanto afetivos e é um dos fatores mais importantes no desempenho acadêmico (HAYAT *et al*, 2020).

O desenvolvimento da metacognição é outro elemento fundamental no processo de aprendizagem e um dos principais objetivos das metodologias ativas. Trata-se do desenvolvimento da capacidade de reflexão sobre o próprio processo de aprendizagem, tornando o estudante capaz de exercer um controle ativo sobre os processos cognitivos envolvidos na mesma (WINNE, 2017; LIVINGSTON, 2003).

A qualidade do ensino ofertado pelas escolas também é um fator determinante no aprendizado e vai muito além do aprendizado proporcionado pelas atividades acadêmicas previstas no currículo. De acordo com a teoria da cognição situada (BROWN *et al*, 2005) o aprendizado é fundamentalmente construído a partir das experiências proporcionadas pela escola, seu ambiente, o clima educacional, o tipo de interação entre professores, alunos e funcionários, a cultura e os valores cultivados na escola (GENN, 2001; ROFF; MCALEER, 2001).

Finalmente, e não menos importante, está a influência dos fatores socioeconômicos e familiares no aprendizado. Existe um longo debate que permeia a interface entre a sociologia e a educação sobre a inter-relação entre estes fatores e o desempenho acadêmico (COLEMANN, 1966; WHITE, 1982; HEYNEMAN; LOXLEY, 1983; SIRIN, 2005). Estes trabalhos clássicos, feitos no contexto dos Estados Unidos da América, indicavam que a família era um determinante muito mais importante no desempenho do estudante do que a escola. Esta conclusão teve grande repercussão tanto no meio acadêmico quanto nas políticas públicas nos Estados Unidos da América. Este argumento foi usado para questionar os resultados dos investimentos na melhoria do ensino, se o peso das escolas no aprendizado era bem menor do que o das condições familiares; o investimento em turmas pequenas, qualificação docente, equipamentos e outros teriam pouco impacto na aprendizagem; pois a melhora do aprendizado dependeria fundamentalmente da melhoria do nível socioeconômico da família.

Estudos posteriores realizados nos anos 1960 e 1970 mostraram uma correlação fraca entre desempenho acadêmico e nível socioeconômico (NSE); porém, havia uma correlação forte entre NSE e desempenho acadêmico quando a escola era usada como unidade de análise, ou seja, escolas com estudantes de NSE menos favorecidos tinham consistentemente piores médias do que escolas que atendem a estudantes de NSE mais altos (White, 1982).

Já no século XXI, mesmo com todas as mudanças sociais ocorridas na sociedade norte-americana, Sirin (2005) voltou a encontrar uma correlação fraca entre NSE e desempenho acadêmico, quando a unidade de análise era o estudante e uma correlação forte quando a unidade de análise era a escola. Estudos correlacionando NSE e desempenho acadêmico, comparando diversos países ampliaram os conhecimentos neste tema e encontraram uma realidade muito mais complexa. Um grande estudo educacional, patrocinado pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE), chamado Programa Internacional de Avaliação de Estudantes PISA (OCDE, 2019) trouxe vários elementos novos e importantes para este debate. O PISA é um estudo realizado a cada dois anos por uma rede mundial de avaliação do desempenho escolar que compara o desempenho de estudantes de quinze anos de idade em diversas áreas e países. Este estudo demonstra a quanto a relação entre NSE e desempenho acadêmico é complexa. Em países como o Brasil,

onde existe uma profunda desigualdade na distribuição de recursos entre o sistema educacional público e o privado, o nível socioeconômico é um forte preditor do desempenho acadêmico, existindo escolas que agrupam alunos de alto desempenho, normalmente privadas e escolas que concentram alunos de baixo desempenho, normalmente públicas (OCDE, 2019).

De uma maneira geral o estudo mostra uma correlação direta entre o desempenho dos estudantes e o nível de desenvolvimento do país; quanto mais pobre o país pior é o desempenho dos estudantes. Porém, neste próprio estudo (OCDE,2019), há elementos que chamam a atenção para o fato de que o nível socioeconômico não é o único determinante do desempenho, diversos países asiáticos de nível de desenvolvimento intermediários estão no topo do desempenho do estudo, indicando a importância de fatores culturais e familiares no aprendizado. Isto também se observa no Brasil onde, segundo o PISA, 10% dos alunos de baixo nível socioeconômico têm um desempenho no quartil superior em capacidade de leitura (OCDE,2019).

Apesar de a correlação entre NSE e desempenho acadêmico ser um tema bastante estudado no ensino fundamental e médio, no ensino superior esta correlação tem se mostrado fraca, duas metanálises (RICHARDSON; ABRAHAM; BOND, 2012; RODRÍGUEZ-HERNÁNDEZ et al., 2020) sobre o tema encontraram uma correlação fraca entre desempenho acadêmico e nível socioeconômico no ensino superior. Porém, como os sistemas educacionais variam muito nos diferentes países, mais estudos são necessários para se ter uma perspectiva mais clara deste tema dentro da realidade brasileira.

Para atingir as competências esperadas ao final da graduação o estudante deve construir uma ampla rede interconectada de conhecimentos, habilidades e atitudes. Este processo de aprendizagem depende de uma complexa interação entre fatores de ordem pessoal, familiar, social, econômica e da escola. O teste de progresso surgiu como uma alternativa mais adequada de avaliação global da evolução de conhecimentos, constituindo-se numa rica fonte de informação sobre o desenvolvimento da programação curricular e do processo de aprendizagem dos estudantes.

2. JUSTIFICATIVA

A realização do TP é trabalhosa e envolve uma quantidade considerável de recursos (VAN DER VLEUTEN, FREEMAN, COLLARES, 2018) e, em função disso, várias escolas médicas têm se associado, tanto no Brasil (BICUDO *et al*, 2019) quanto em outros países (NOUNS; GEORG, 2010), formando consórcios para sua realização, proporcionando uma diminuição dos custos, otimização de recursos e abrindo a possibilidade de comparação do desempenho dos estudantes com os de outras instituições. Apesar das vantagens evidentes existe uma lacuna na literatura quanto a qualidade dos testes produzidos em consórcio. A maior parte da literatura analisa os resultados da aplicação do teste de duas a quatro vezes ao ano existindo poucos trabalhos sobre qual feedback pode ser obtido a partir da realização do teste apenas uma vez ao ano.

Na tentativa de suprir esta lacuna realizamos o presente trabalho que analisa o teste de progresso realizado em uma escola pública que utiliza metodologia ativas de ensino e aprendizagem e participa do Consórcio do Centro Oeste para o teste de progresso desde início das atividades em 2013.

3. OBJETIVOS

O presente trabalho foi planejado tendo dois objetivos principais, analisar as características e a qualidade dos testes produzidos e qual o feedback pode ser obtido a partir dos seus resultados.

a. OBJETIVOS PRIMÁRIOS

- Analisar as características psicométricas dos testes pela Teoria Clássica dos Testes.
- Analisar a progressão do desempenho dos estudantes correlacionando-o com a série cursada, gênero, participação nos programas de iniciação científica, reprovações no curso e forma de acesso.

b. OBJETIVOS SECUNDÁRIOS:

- Analisar a possibilidade da utilização do teste de progresso para feedback individual do estudante.
- Analisar os processos de produção, aplicação, análise dos resultados e feedback dos testes de progresso realizados pelo consórcio.
- Relacionar a qualidade dos testes produzidos pelo consórcio com seu processo de produção.

4. MÉTODOS

Trata-se de um estudo retrospectivo, onde foram analisados os resultados de todos os testes de progresso aplicados no período de 2012 a 2019. Para análise das características psicométricas dos testes foram coletados os dados de 3135 aplicados entre 2013 e 2019. A análise do desempenho dos estudantes foi realizada a partir de uma amostra de 3548 testes realizados por 1040 estudantes entre 2012 e 2019. Como a participação no teste é optativa e em todos os anos existem tanto alunos ingressantes como concluintes, na amostra estudada, existem alunos que participaram de uma até oito vezes no teste. Foram incluídos todos os testes realizados pelos estudantes admitidos pelo sistema universal e pelo sistema de cotas. Foram excluídos da amostra todos os testes realizados pelos alunos admitidos por processos de transferência devido ao pequeno número de estudantes ingressantes nesta modalidade.

4.1 CONTEXTO DO ESTUDO

O trabalho foi conduzido em uma escola pública, localizada em Brasília, Distrito Federal, a Escola foi criada em 2001 e oferece anualmente oitenta vagas para o Curso de medicina com uma entrada única anual. O processo de seleção foi realizado pelo exame vestibular de 2001 a 2015; a partir de 2016 a entrada passou a se dar pelo Sistema de Seleção Unificada (SISU) do Ministério da Educação. A partir

de 2005, por força de uma Lei Distrital, 40% das vagas anuais passaram a ser reservadas para estudantes oriundos do ensino público do Distrito Federal (DISTRITO FEDERAL, 2004). A partir de 2013, o acesso pelo sistema de cotas foi estendido para estudantes do ensino público de todo o Brasil. Além do acesso diferenciado, os ingressantes pelo sistema de cotas recebem uma bolsa de estudos mensal durante todo o Curso.

O projeto pedagógico baseia-se nos princípios das metodologias ativas de ensino aprendizagem que priorizam o aprendizado a partir da prática, os alunos são inseridos nos serviços de saúde desde o início do curso. As experiências vivenciadas nos serviços de saúde e os problemas discutidos nas sessões de PBL servem de disparadores para os estudos e as reflexões em pequenos grupos (PPC-ESCS, 2021). O ensino é centrado no estudante e o professor tem o papel de facilitador do processo, à exceção de uma única palestra semanal, todas as atividades são realizadas em pequenos grupos. Trata-se da primeira Escola de Medicina do Brasil vinculada e mantida com recursos de uma Secretaria de Estado de Saúde. O corpo docente é integralmente constituído por profissionais de saúde atuantes nos serviços da Secretaria de Estado de Saúde do Distrito Federal (SES-DF). Estas características fizeram da escola uma experiência singular de integração entre o ensino e os serviços de saúde do SUS, um dos pilares das reformas propostas para ensino médico brasileiro, iniciada na primeira década dos anos 2000 (CNE,2001).

O Projeto Pedagógico do Curso visa, em essência, uma formação crítico-reflexiva e humanista e coloca o estudante no papel central de construção do próprio conhecimento (IBRAHIM BELACIANO, 2015; PPC-ESCS, 2021). O perfil desejado do egresso é o de um profissional com as competências necessárias para atuação no SUS, dentro de uma lógica da formação de redes de atenção. O currículo é seriado, com duração de seis anos e composto por quatro programas educacionais (PPC-ESCS, 2021):

- O programa de módulos temáticos e dinâmica tutorial que utiliza como metodologia o PBL, seguindo o modelo desenvolvido pela Universidade de Maastricht (SCHMIDT, 1983).

- O programa de interação ensino-serviço-comunidade que tem por objetivo a inserção na atenção primária e adota a metodologia da problematização (BERBEL, 1995).

- O programa Habilidades e Atitudes que ocorre de primeira à quarta séries e que utiliza diversas metodologias como simulações, discussão de casos e atividades nos serviços de saúde.

- O Estágio Curricular Obrigatório (ECO) que acontece nos dois últimos anos do Curso e consiste numa imersão nos serviços de saúde da SES-DF sob a supervisão dos profissionais da instituição, seja na qualidade de docentes ou de preceptores de ensino de graduação.

Existe também na ESCS um programa de iniciação científica que tem o objetivo de introduzir os estudantes no campo da pesquisa científica, sob a orientação de um pesquisador. A participação no programa é optativa e o ingresso é feito mediante um processo seletivo, na qual o estudante pode participar como bolsista ou voluntário. As bolsas de iniciação científica concedidas são oriundas do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) do Conselho Nacional de Desenvolvimento Científico e Tecnológico e do Programa de Iniciação Científica da ESCS (PIC). A duração é de 12 meses e o produto final da pesquisa é apresentado na mostra científica anual da ESCS.

O sistema de avaliação da Escola é critério referenciado, existindo apenas dois conceitos de desempenho, satisfatório ou insatisfatório. Os estudantes são avaliados por diferentes instrumentos de avaliação e existe um sistema de recuperação dos objetivos não atingidos. A decisão sobre a progressão no curso é anual e para ser aprovado o estudante deve obter o conceito satisfatório em todos os objetivos educacionais da série; a obtenção do conceito final insatisfatório em qualquer unidade educacional leva à reprovação na série.

O Consórcio do Centro Oeste para o teste de progresso foi criado em 2013 por um grupo de escolas da região centro oeste do Brasil, que se uniram aproveitando a experiência de uma das escolas que já aplicava o teste de maneira individual desde 2005.

Desde então, o número de escolas tem crescido, com a participação de instituições de outras regiões do país. O Consórcio tem a característica de ser aberto a qualquer instituição que queira participar, sendo necessário apenas indicar um representante, que vai participar da organização e revisão dos itens produzidos. As instituições participantes se comprometem também a produzir os itens

encomendados. Não existe contratação de qualquer serviço de terceiros e todas as etapas são realizadas com recursos próprios das escolas. O número de escolas participantes tem crescido a cada ano, sendo que a última edição do teste realizada em 2021 contou com a participação de 36 cursos de medicina de diversas regiões do Brasil.

4.2 ANÁLISE DAS CARACTERÍSTICAS PSICOMÉTRICAS DOS TESTES PELA TEORIA CLÁSSICA DOS TESTES.

Os testes aplicados, bem como os microdados correspondentes, foram obtidos a partir dos arquivos referentes aos testes de progresso existentes na escola.

As evidências de validade foram obtidas a partir da análise do processo de produção e aplicação e resultados dos testes, pelas curvas de desempenho dos alunos e pela correlação entre as reprovações e os resultados obtidos no teste.

A confiabilidade foi aferida pelo alfa de Cronbach, o nível de dificuldade dos itens foi calculado pelo percentual de estudantes que acertaram a questão e foi classificado em fácil, médio e difícil quando o percentual de acerto foi maior que 0,7, entre 0,3 e 0,7 e menor que 0,3, respectivamente (PASQUALI, 2003).

A discriminação foi calculada pela correlação ponto-bisserial e para efeitos da utilização da discriminação do item como indicador de sua qualidade foi utilizado o critério estabelecido no trabalho clássico de Ebell (1979) que estabelece que itens com discriminação entre 0,2 e 0,29 já demonstram algum poder de discriminação, no entanto necessitam ser revisados, e entre 0,3 e 0,39, como itens com boa discriminação e que necessitariam de pequenas ou nenhuma revisão e acima de 0,4 que são itens ótimos.

Foram considerados bons os itens com discriminação maior que 0,3. Por outro lado, com o objetivo de analisar a qualidade do conjunto dos itens produzidos, foram destacados também os itens com discriminação negativa, ou seja, itens em que os alunos de baixo desempenho obtiveram um resultado melhor do que os alunos de alto desempenho, o que constitui um indicador de problemas graves na construção do item.

Os distratores são as alternativas que apresentam as respostas incorretas para os itens. Como os testes apresentavam quatro opções de resposta, os itens

foram classificados segundo seu percentual de utilização dos distratores. Foram considerados não-funcionais, os distratores com percentual de escolha inferior a 5% (TARRANT, 2009). A efetividade dos distratores foi considerada: 0; 33,3; 66,6 e 100 para 0; 1; 2 e 3 distratores, respectivamente.

A descrição das etapas referentes ao processo de produção, aplicação, análise e feedback do teste de progresso foram feitas a partir da rotina utilizada pelo Consórcio do Centro Oeste para o teste de progresso, utilizando como referencial teórico o guia da *Association for Medical Education in Europe* (Wrigley et. al. 2012).

A análise da qualidade dos testes produzidos foi feita com base nos resultados da análise psicométricas dos itens e na confrontação dos processos utilizados pelo Consórcio do Centro Oeste com o referencial teórico proposto.

4.3 ANÁLISE DO DESEMPENHO DOS ESTUDANTES NOS TESTES

Para analisar o desempenho dos estudantes no teste foram coletados os dados existentes na ESCS sobre o desempenho individual dos estudantes nos testes realizados entre 2012 e 2019, estas informações foram relacionadas as variáveis série, gênero, tipo de admissão, número de reprovações e participação nos programas de iniciação científica coletadas nos registros acadêmicos da escola.

Para equalizar as diferenças decorrentes da variação na dificuldade entre os testes, os escores obtidos foram padronizados em escore Z relativo ao ano do teste.

Cada variável foi empilhada em uma variável única relacionada a série cursada, gênero, acesso, reprovação e participação em programas de iniciação científica. Os valores ausentes foram eliminados da base de dados.

Para correlacionar os diferentes padrões de progressão de conhecimentos de acordo com o tipo de acesso foi realizada uma regressão linear múltipla clássica (Fahrmeir et al. 2013) sem interação entre os termos e, posteriormente, para avaliar a possibilidade de interação entre as variáveis explicativas foi realizada uma regressão linear com interação entre os termos.

Foram incluídas no modelo, todos os testes diagnósticos do modelo linear, sendo testadas as hipóteses dos resíduos randômicos, homoscedasticidade, igualdade das variâncias e ausência de *outliers* alavancados. O modelo linear foi testado quanto às suas premissas. Somente a normalidade dos resíduos parece ter

sido levemente violada, porém, como os modelos lineares são robustos diante desses desvios, o modelo foi mantido.

Para quantificar ganho de conhecimento em cada série do Curso, foi calculado o *d* de Coehn (BECKER, 2000), que é uma estimativa do tamanho do efeito de cada série sobre o desempenho dos estudantes no teste. Coehn (1998) classificou arbitrariamente este efeito como baixo ($< 0,2$), médio (entre 0,2 e 0,8) e grande ($> 0,8$).

4.4 ANÁLISE DA UTILIZAÇÃO DO TESTE PARA FEEDBACK INDIVIDUAL

Para verificar a confiabilidade das informações dos testes para *feedback* a nível individual, realizamos o cálculo do índice de mudança confiável ou *reliable change index* (RCI), de acordo com Jacobson e Truax (1991). O RCI consiste em calcular a diferença de uma propriedade individual, em dois momentos do tempo, e padronizar essa diferença. Essa medida padronizada da diferença permite a discriminação entre situações em que o progresso individual não poderia ser atribuído ao acaso. Sendo “i” o ano de referência para toda a informação, o ganho é medido face ao desempenho do ano anterior “i-1” sempre que esta informação esteja disponível. Dito isto, o *RCI* pode ser definido da seguinte forma:

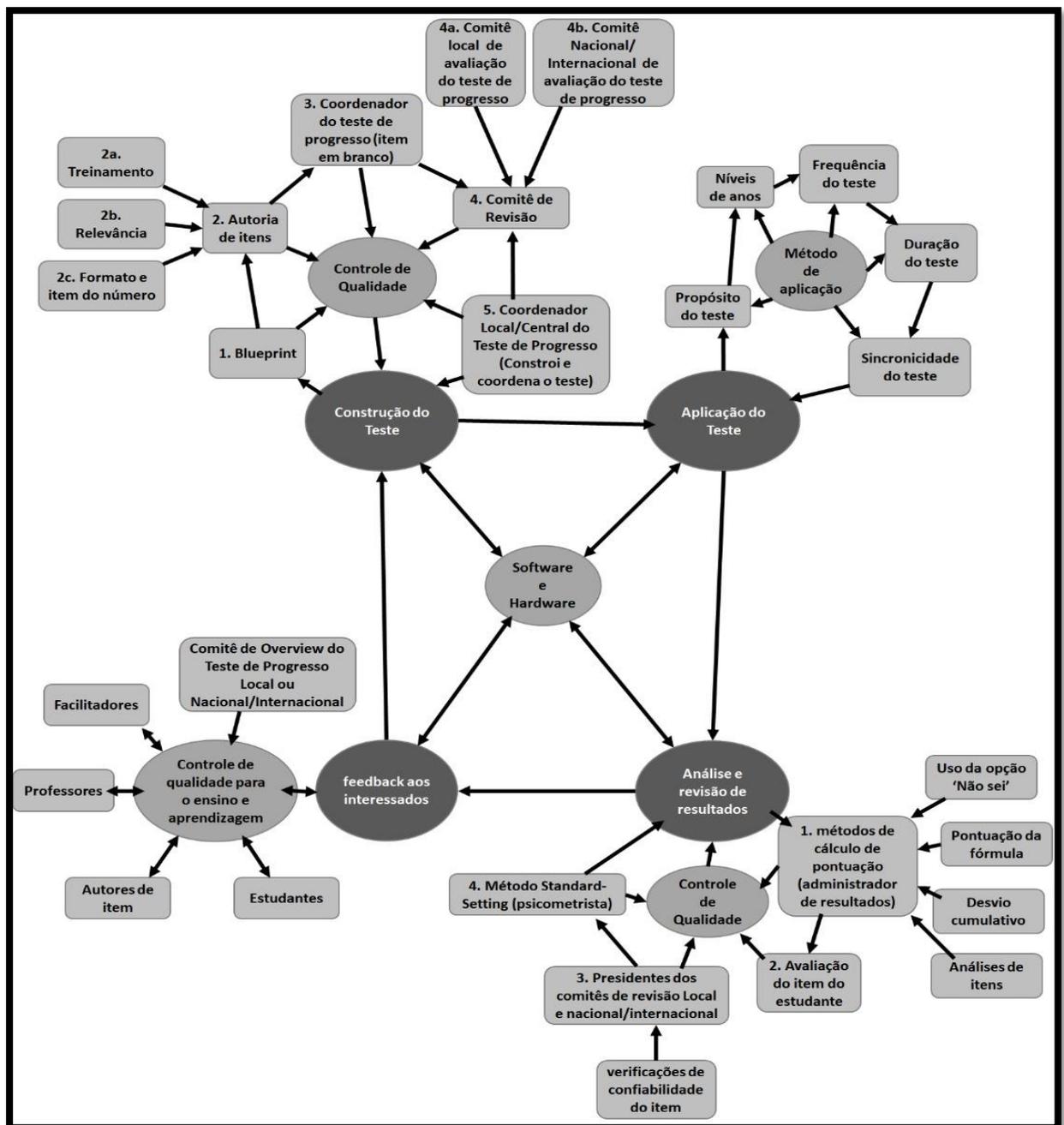
$$RCI = \frac{x_i - x_{i-1}}{\sqrt{2s_x^2(1 - r_{xx})}}$$

Onde s_x^2 é a variância das medidas de x, e r_{xx} é a correlação entre medidas consecutivas de x, avaliada pelo coeficiente alfa de Cronbach. Assumindo que a variável de teste x tem distribuição normal e assumindo a hipótese nula (que não houve ganho real), a diferença padronizada acima deve satisfazer uma distribuição z. Ou seja, deve ser estatisticamente significativo (com $p < 0,05$) quando o RC for maior que 1,96.

4.5 ANÁLISE DOS PROCESSOS DE PRODUÇÃO, APLICAÇÃO, ANÁLISE DOS RESULTADOS E FEEDBACK DOS TESTES.

Para a análise das etapas necessárias para a realização do teste de progresso foi utilizado como referencial teórico a estrutura proposta pelo guia da *Association for Medical Education in Europe* (Wrigley et al., 2012), (Figura 1) nele a realização teste é dividida em quatro etapas, construção do teste, aplicação, análise dos resultados e revisão e feedback aos envolvidos.

Figura 1. Estrutura geral sistêmica para o teste de progresso.



Fonte: Adaptado de WRIGLEY et al. *A systemic framework for the progress test: Strengths, constraints and issues: AMEE Guide No. 71. Medical Teacher*, v. 34, n. 9, p. 683–697, 2012.

As análises estatísticas foram realizadas usando o Microsoft Excel, IBM SPSS e o R Statistical Software onde foram utilizados os seguintes pacotes: MASS - Modern Applied Statistics with S); ggfortify (Data Visualization Tools for Statistical Analysis Results); Hmisc: (Harrell Miscellaneous); dplyr (A Grammar of Data Manipulation).

5. ASPECTOS ÉTICOS

O estudo foi aprovado pelo comitê de ética em pesquisa da Fundação de Ensino e Pesquisa em Ciências da Saúde (FEPECS) parecer número 1.656.421, nenhum estudante foi identificado em qualquer uma das etapas da análise. Houve dispensa do termo de consentimento livre e esclarecido e o acesso aos dados e a divulgação dos resultados do teste de progresso da escola foram autorizados pela instituição.

6. RESULTADOS

6.1 CARACTERÍSTICAS PSICOMÉTRICAS DOS TESTES.

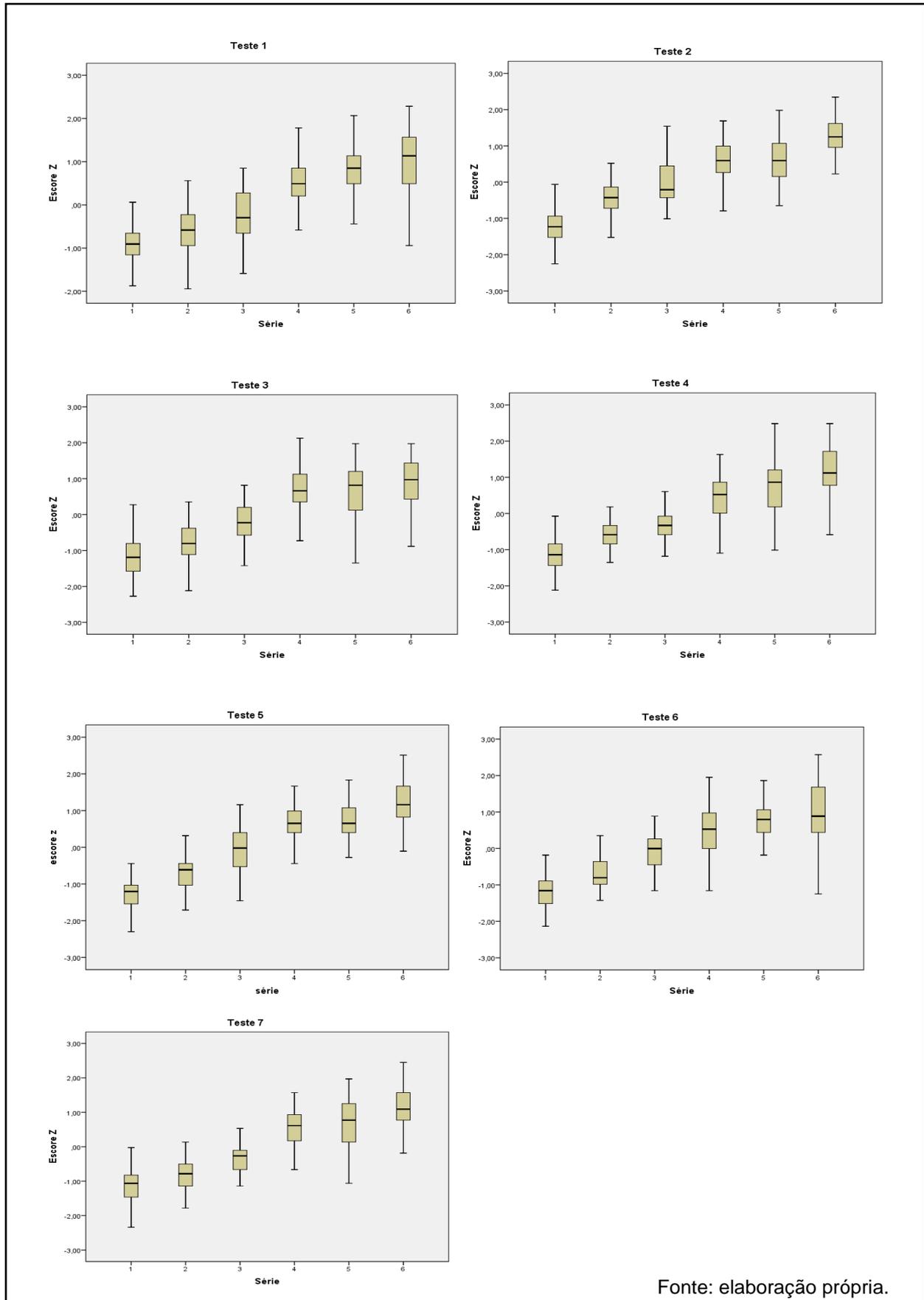
Foram analisados 3135 testes aplicados entre 2013 e 2019. A progressão de conhecimentos dos estudantes ficou evidenciada em todos os anos conforme disposto na Tabela 1 e na figura 2.

Tabela 1. Média dos Escores Z e intervalos de confiança de primeira a sexta séries nos testes aplicados de 2013 a 2019

ANO	ESCORE Z		SÉRIE					
			1	2	3	4	5	6
2013	Média		-0,10	-0,62	-0,24	0,49	0,59	1,04
	I.C. 95%	Lim. Inf.	-1,13	-0,77	-0,39	0,38	0,36	0,49
		Lim. Sup.	-0,86	-0,47	-0,09	0,61	0,84	1,03
2014	Média		-1,20	-0,49	-0,49	0,52	0,60	1,00
	I.C. 95%	Lim. Inf.	-1,30	-0,63	-0,23	0,35	0,42	0,77
		Lim. Sup.	-1,11	-0,36	0,13	0,68	0,77	1,23
2015	Média		-1,15	-0,75	-0,23	0,70	0,63	0,82
	I.C. 95%	Lim. Inf.	-1,28	-0,87	-0,35	0,53	0,46	0,65
		Lim. Sup.	-1,02	-0,63	-0,10	0,85	0,80	0,99
2016	Média		-1,16	-0,66	-0,28	0,46	0,74	1,07
	I.C. 95%	Lim. Inf.	-1,27	-0,76	-0,39	0,33	0,57	0,87
		Lim. Sup.	-1,04	-0,56	-0,17	0,59	0,92	1,26
2017	Média		-1,25	-0,70	-0,06	0,62	0,70	1,14
	I.C. 95%	Lim. Inf.	-1,35	-0,81	-0,19	0,50	0,57	0,95
		Lim. Sup.	-1,14	-0,59	0,06	0,74	0,83	1,33
2018	Média		-1,19	-,068	-0,08	0,47	0,78	0,99
	I.C. 95%	Lim. Inf.	-1,29	-0,77	-0,19	0,31	0,63	0,79
		Lim. Sup.	-1,10	-0,58	0,03	0,63	0,93	1,20
2019	Média		-1,12	-0,83	-0,30	0,55	0,68	1,10
	I.C. 95%	Lim. Inf.	-1,23	-0,93	-0,41	0,42	0,52	0,93
		Lim. Sup.	-1,01	-0,73	-0,19	0,67	0,85	1,26
TOTAL	Média		-1,16	-0,68	-0,18	0,54	0,68	0,97
	I.C. 95%	Lim. Inf.	-1,20	-0,72	-0,23	0,49	0,62	0,90
		Lim. Sup.	-1,11	-0,64	-0,13	0,60	0,74	1,05

fonte: elaboração própria.

Figura 2. Escores Z obtidos por cada série nos testes aplicados de 2013 a 2019, Testes de 1 a 7 respectivamente.



A confiabilidade dos testes mostrou-se estável ao longo dos anos, com um alfa de Cronbach médio de 0,88 (mínimo de 0,84 máximo de 0,90). A dificuldade média dos testes foi de 0,51 variando entre 0,49 e 0,53. 18% dos itens aplicados foram fáceis, 66% tiveram dificuldade média, e 16% foram difíceis. A discriminação média dos testes foi de 0,24, em média 61% dos itens tiveram discriminação igual ou maior que 0,20; 36% dos itens atingiram discriminação maior ou igual a 0,3; 4,6% dos itens tiveram discriminação negativa (tabela 2).

Tabela 2. Número de alunos, quantidade de itens válidos, confiabilidade dos testes, nível de dificuldade e discriminação dos itens nos testes aplicados de 2013 a 2019.

		TESTE						MÉDIA	
		2013	2014	2015	2016	2017	2018		2019
NÚMERO DE TESTES APLICADOS		411	385	466	466	457	479	471	448
QUANTIDADE DE ITENS VÁLIDOS		98	100	100	99	96	99	100	98,9
ALFA DE CRONBACH		0,9	0,88	0,88	0,86	0,88	0,84	0,86	0,88
DIFICULDADE MÉDIA DO TESTE		0,50	0,49	0,53	0,50	0,54	0,51	0,51	0,51
DISCRIMINAÇÃO MÉDIA DO TESTE		0,27	0,26	0,25	0,22	0,24	0,21	0,23	0,24
DIFICULDADE DOS ITENS (%)	≥ 0,7	15	11	23	19	26	18	16	18
	> 0,3 e < 0,7	73	75	64	64	57	62	69	66
	≤ 0,3	12	14	13	17	17	20	15	16
DISCRIMINAÇÃO DOS ITENS (%)	> 0,2	71	65	66	52	61	49	61	61
	> 0,3	46	39	45	33	43	31	36	39
	< 0	4	8	5	3	5	7	9	4,6

fonte: elaboração própria

Quanto ao uso dos distratores, em média, em um item por teste não houve uso dos distratores, ou seja, todos os estudantes assinalaram a resposta correta. Somente um distrator foi utilizado em média em nove itens, dois em vinte e quatro itens. Em média, em 65 questões todos os distratores foram utilizados (Tabela 3).

Tabela 3. Número de itens e de distratores utilizados nos testes de 2013 a 2019

NÚMERO DE DISTRATORES UTILIZADOS	TESTE							Média
	2013	2014	2015	2016	2017	2018	2019	
0	3	0	0	2	2	2	1	1
1	4	7	14	10	6	11	9	9
2	22	19	20	25	32	18	30	24
3	69	74	66	62	56	68	60	65

fonte: elaboração própria

6.2 ANÁLISE DO DESEMPENHO DOS ESTUDANTES NOS TESTES

Foram coletados dados relativos a um total de 1040 estudantes, 540 mulheres (51,9%) e 500 homens (48,1%). Quanto à forma de ingresso, 644 (61,92%) dos estudantes da amostra ingressaram pelo sistema universal, enquanto 391 (38,07%) ingressaram pelo sistema de cotas.

Os estudantes da amostra realizaram 3548 testes no período de 2012 até 2019. No ano de 2020, em função da pandemia COVID-19, o teste não foi realizado. O número de estudantes participantes por teste está disposto na tabela 4. Observa-se que em média a participação no teste foi de 87,9% variando de 77,6 a 92,3%.

Tabela 4. Número de estudantes participantes por edição do teste e matrículas ativas de 2012 a 2019

Ano do teste	Nº de estudantes participantes	Número de matrículas ativas	Percentual de estudantes que fizeram o teste
2012	413	482	85,7
2013	411	478	86,0
2014	385	496	77,6
2015	466	516	90,3
2016	466	505	92,3
2017	457	505	90,5
2018	479	529	90,5
2019	471	523	90,0

fonte: elaboração própria

O número de participações dos estudantes no teste foi distribuído conforme a tabela 5. Observa-se que dos 1040 estudantes que compuseram a amostra, apenas 164 realizaram seis ou mais testes.

Tabela 5. Número de participações dos estudantes no teste.

Número de testes realizados	Número de estudantes	Percentual de estudantes da amostra
1	195	18,8
2	195	18,8
3	144	13,9
4	169	16,3
5	172	6,5
6	154	14,8
7	9	0,9
8	1	0,1

fonte: elaboração própria

Do total de estudantes do estudo, 485 participaram de programas de iniciação científica. O número participações no programa de iniciação científica dos estudantes da amostra está disposto na tabela 6.

Tabela 6. Número de participações dos estudantes no programa de iniciação científica.

Número de participações no programa de iniciação científica	Número de estudantes da amostra	Percentual de estudantes da amostra (%)
1	195	18,7
2	142	13,6
3	98	9,4
4	38	3,7
5	11	1,1
6	1	0,1

fonte: elaboração própria

Em relação à reprovação, os dados mostram que 127 (12,2%) dos estudantes da amostra tiveram pelo menos uma reprovação durante o curso, o número de reprovações dos estudantes da amostra está disposto na tabela 7.

Tabela 7. Número de reprovações dos estudantes

Número de reprovações no curso	Número de estudantes da amostra	Percentual de estudantes da amostra (%)
1	88	8,5
2	23	2,2
3	12	1,2
4	4	4

fonte: elaboração própria

Nossos resultados mostraram um aumento significativo no desempenho dos estudantes em cada série do curso. Não foi observada diferença de desempenho entre os gêneros ($p=0,37$). Os estudantes que tiveram reprovação em alguma série tiveram em média um desempenho significativamente menor no teste de progresso que correspondeu a 0,29 desvios padrão ($p < 0,01$). Os estudantes que participaram do programa de iniciação científica obtiveram em média um desempenho 0,06 desvio padrão ($p=0,03$) acima dos seus colegas que não participaram do programa. Os estudantes admitidos pelo sistema de cotas obtiveram um desempenho significativamente inferior ($p < 0,01$) equivalente a em média 0,30 desvio padrão quando comparados com seus colegas admitidos pelo sistema universal (Tabela 8).

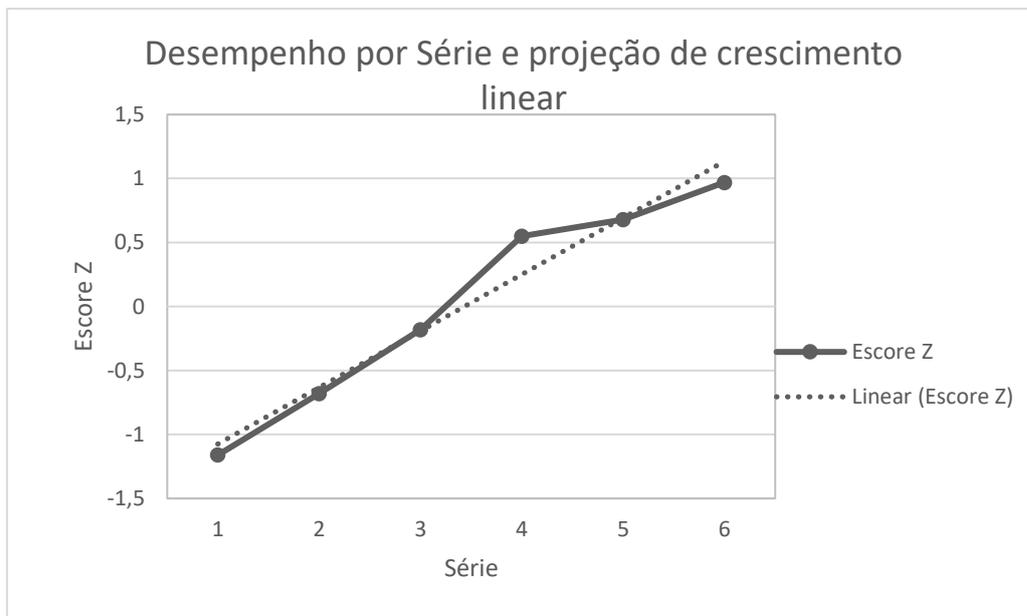
Tabela 8. Resultados da regressão logística realizando a comparação dos escores por série do Curso, por gênero, pela ocorrência de reprovação, por participação no programa de iniciação científica e por forma de acesso

VARIÁVEL		MÉDIA	DESVIO PADRÃO	MEDIANA	ESCORE Z MÉDIO	SIGNIFICÂNCIA (p)
SÉRIE	1	35	6,7	35	-1,16	-----
	2	41	6,6	41	-0,68	< 0,01
	3	46	6,9	46	-0,18	< 0,01
	4	55	7,9	56	0,54	< 0,01
	5	57	9,2	58	0,68	< 0,01
	6	61	11,1	62	0,97	< 0,01
GÊNERO	Masculino	47	10,3	47	- 0,08	0,37
	Feminino	48	9,6	48	- 0,04	
REPROVAÇÃO	Sim	45	9,8	44	- 0,27	< 0,01
	Não	48	9,9	48	- 0,03	
INICIAÇÃO CIENTÍFICA	Sim	51	9,0	50	0,21	0,03
	Não	44	9,8	44	- 0,31	
ACESSO	Universal	49	10,0	50	0,10	< 0,01
	Cotas	44	9,0	44	- 0,34	

Fonte: elaboração própria.

A progressão de desempenho dos estudantes por série do curso e a projeção de crescimento linear podem também ser observados no gráfico 1.

Gráfico 1. Desempenho por série nos testes de progresso de 2012 a 2019 e linha de projeção linear do crescimento



Fonte: elaboração própria.

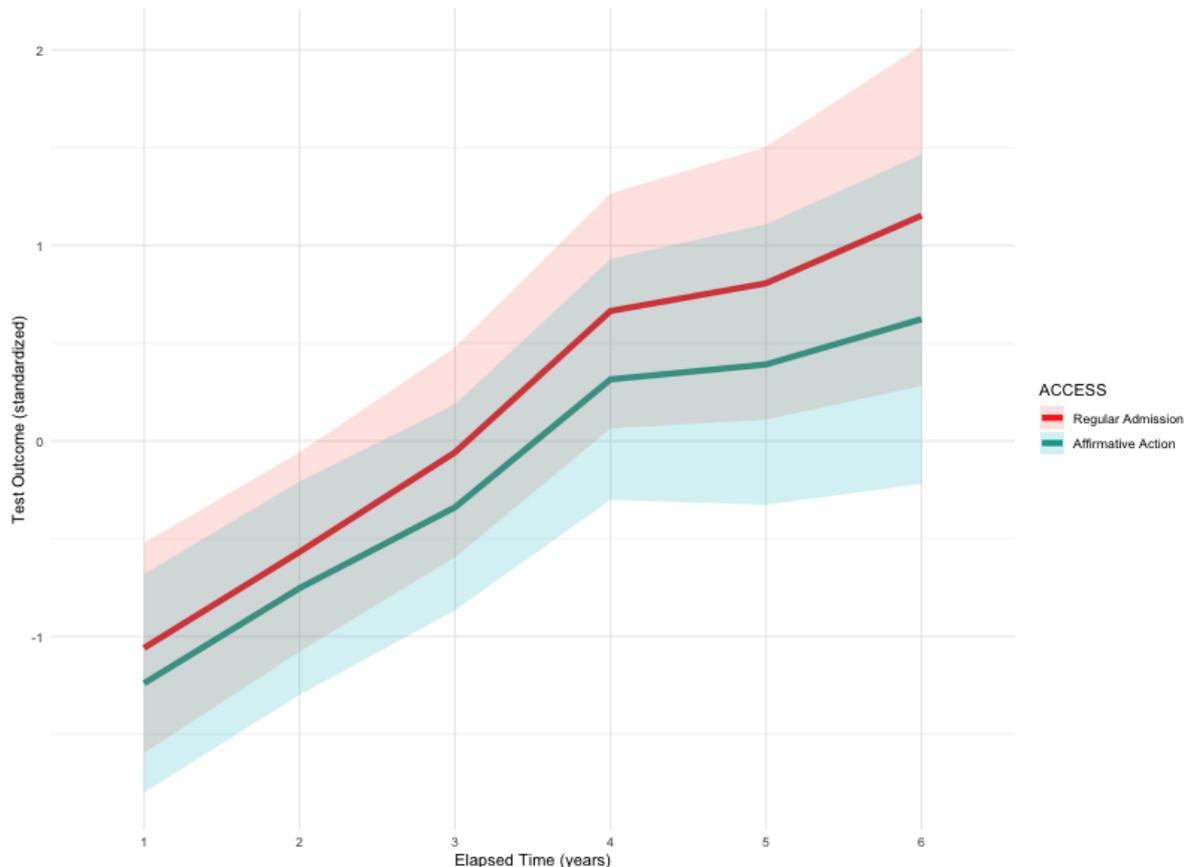
O tamanho do efeito da progressão do desempenho, calculado pelo d de Coehn, evidenciou o ganho de conhecimento incorporado por cada série do Curso (Gráfico 2). Observou-se um aumento na taxa de progressão na quarta série, que, posteriormente apresenta uma diminuição na quinta e sextas séries.

Gráfico 2. Tamanho do efeito de cada série sobre o aumento de conhecimento



A regressão linear com interação entre os termos mostrou que a diferença de desempenho entre os ingressantes pelo sistema de cotas e seus colegas do acesso universal permaneceu constante do primeiro ao quarto ano e foi em média de 0,21 desvio padrão; no quinto e sexto ano, a diferença entre os dois grupos aumenta para 0,35 unidades de desvio padrão (Gráfico 3). Os resultados da regressão logística estão expressos na Tabela 2:

Gráfico 3. Desempenho dos estudantes admitidos pelo sistema de cotas e pelo sistema universal nos testes de progresso aplicados no período de 2012 a 2019.



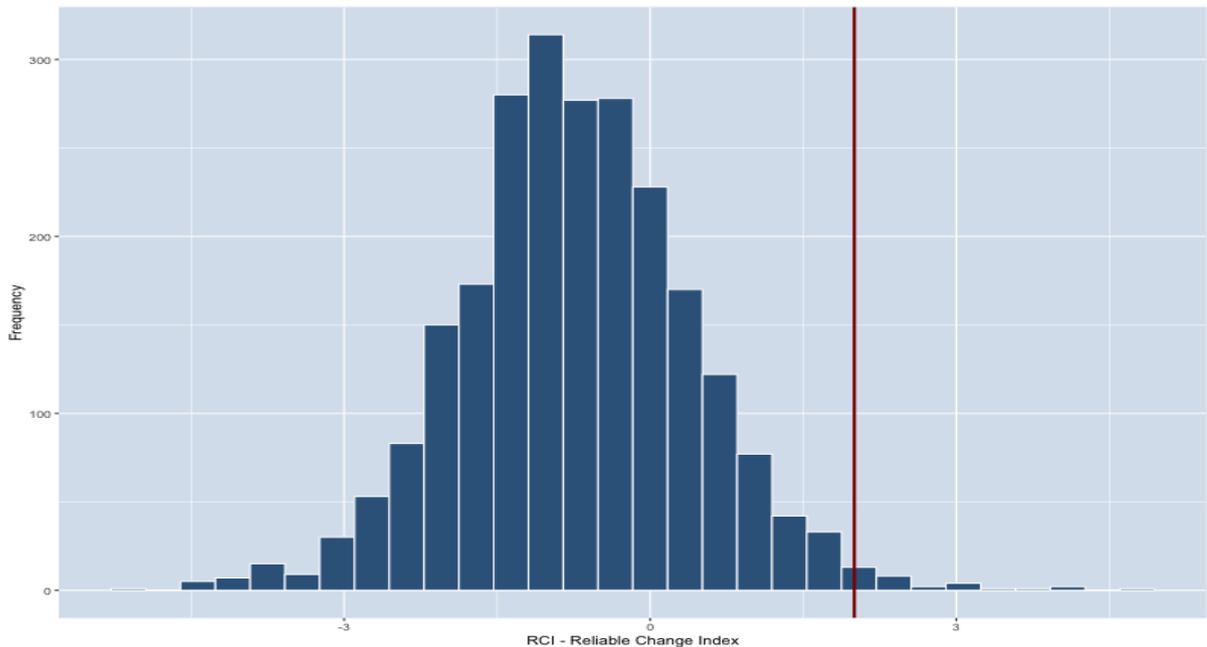
Fonte: elaboração do autor

6.3 UTILIZAÇÃO DO TESTE PARA FEEDBACK INDIVIDUAL

O RCI foi calculado para todos os testes consecutivos feitos pelo mesmo aluno. Os resultados permitiram apontar o número de casos em que houve evidência de progresso. Os resultados são mostrados no gráfico 4.

No conjunto, a distribuição do RCI deslocou-se para a direita (0,73 unidades de desvio padrão), confirmando a tendência geral de evolução. No entanto, a maioria das aplicações ocorreu na região onde $RCI < 1,96$. Ou seja, não corresponderam a diferenças que pudessem ser consideradas estatisticamente significativas (com $p < 0,05$). Nota-se que para 14,7% dos casos considerados, podemos afirmar que houve evidência de evolução (com $p > 0,05$). No outro extremo (de regressão em vez de progressão), os casos estatisticamente significativos eram mais improváveis de aparecer.

Gráfico 4. Histograma de frequência do RCI para todos os PT consecutivos realizados pelo mesmo aluno. O lado direito da linha vermelha indica os casos em que houve evidência de progresso individual.



fonte: elaboração do autor

6.4 DESCRIÇÃO DO PROCEDIMENTO UTILIZADO NA REALIZAÇÃO DOS TESTES

Segundo Wrigley et al. (2012) o processo envolvido na realização do teste de progresso pode ser dividido em quatro etapas: Produção, aplicação, análise dos resultados e feedback.

A etapa inicial para a produção do teste é a construção de um *blueprint*. O TPCO adota um *blueprint* baseado na opinião de *experts*. A matriz contém, num dos seus eixos, as diversas disciplinas que constituem os currículos e no outro os sistemas corporais. No momento da construção do teste, o grupo de coordenação central identifica os cruzamentos mais relevantes e a proporção de questões em cada cruzamento. Após esta etapa é acrescentado o processo mais relevante para este cruzamento. Este procedimento gera uma lista de encomendas de questões que são igualmente distribuídas para as escolas participantes do consórcio. A lista de encomendas contém três descritores: o sistema envolvido, a disciplina e o processo, por exemplo: sistema neuroendócrino (sistema), Patologia (disciplina), diagnóstico (processo). Todos os itens construídos devem ser de nível taxonômico de aplicação

do conhecimento que deve ser contextualizado por meio de vinhetas ou situações problema.

As encomendas são enviadas para os representantes locais de cada escola que distribui as encomendas para os docentes especialistas em cada área, os itens elaborados devem vir acompanhados de comentários que embasem a resposta correta e as respectivas referências.

Os itens elaborados retornam para a comissão revisora que os analisa conforme diretrizes para elaboração de itens (PANIAGUA; SWYGERT, 2016). A comissão revisora aprova os itens considerados adequados para serem incluídos no teste, os itens considerados falhos podem receber alterações pela comissão ou retornam para os elaboradores com comentários e sugestões de ajustes num processo dialógico visando o seu aperfeiçoamento. Após este processo o grupo realiza a seleção dos itens que irão compor o teste.

O teste é aplicado de maneira presencial e em papel, simultaneamente em todas as escolas participantes. O estudante leva os cadernos de prova e o gabarito com os comentários são divulgados logo após o término da prova. Os estudantes tem um período definido para interpor recursos contra o gabarito. Caso o recurso seja deferido o item é eliminado da prova. Os cartões de resposta são processados e os dados enviados para uma das instituições participantes que centraliza e divulga os resultados finais expressos em percentual de acertos.

A análise dos resultados e revisão das características psicométricas dos itens e do teste é realizada pela coordenação central, são construídas as curvas de progressão por série e é fornecido para cada escola seu resultado individual e o resultado geral do consórcio. As escolas participantes também recebem os microdados correspondentes aos seus alunos.

O procedimento de revisão dos resultados do TP é feito rotineiramente pela coordenação central, são realizados o cálculo da confiabilidade do teste e a análise das características psicométrica dos itens. Posteriormente, os responsáveis pela revisão dos resultados realizam reuniões com os elaboradores dos itens e com a coordenação local de cada escola onde é dado o feedback sobre o desempenho das questões no teste.

Todo o suporte de tecnologia da informação para a operacionalização do teste, como inscrição dos alunos, distribuição nas salas, geração e correção dos cartões de resposta, recebimento de recursos, análise e divulgação de resultados e feedback, é fornecida por uma das instituições participantes sem custo para as demais escolas.

Os processos envolvidos e maneira que são realizados pelo TPCO estão sumarizados na tabela 9.

Tabela 9. Descrição das etapas realizadas pelo TPCO na realização do teste de progresso

Processo	Componente	Realização pelo Consórcio
Construção	Blueprint	Baseado na opinião de experts, composto por 17 sistemas orgânicos e 19 disciplinas
	Autoria dos itens	Identificada pela escola elaboradora
	Formato e número de itens	Cem itens de melhor resposta escolhida entre quatro alternativas, com enunciado em formato de vinheta.
	Relevância	Definida pela relevância e prevalência de cada tema definido pelas áreas do blueprint
	Treinamento	Os coordenadores locais de cada instituição recebem as orientações sobre a construção dos itens e orientam os elaboradores
	Banco de itens	Administrado pela coordenação central
	Comissão revisora	Composta por representantes das escolas participantes
	Coordenação central do teste	Composta por um grupo de docentes das escolas participantes
	Controle de qualidade	Realizado pela Comissão revisora
Aplicação	Objetivo do teste	Formativo, para controle de qualidade do processo de ensino e aprendizagem
	Séries	Todas
	Frequência do teste	Uma vez ao ano
	Duração	Quatro horas
	Sincronia	Todas as escolas participantes aplicam o teste simultaneamente
	Método de aplicação	Em papel
Análise dos resultados e revisão	Método de cálculo do escore	Percentual de acertos
	Avaliação do item pelo estudante	O estudante recebe o caderno de provas e o gabarito comentado, qualquer discordância com relação ao gabarito é passível de recurso
	Comissão revisora	Realizado pela Coordenação central
	Standard-Setting	Não há
Feedback para os envolvidos	Controle de qualidade do processo de ensino e aprendizagem	As escolas participantes recebem as curvas de progressão da sua escola e do consórcio. São também fornecidos os

		microdados para análises específicas de interesse da escola
	Estudantes	Recebem o caderno de provas, o gabarito com as questões comentadas e as referências relativas a cada questão. Cada estudante recebe seu resultado e a comparação com os colegas da mesma série na escola e no consórcio
	Elaboradores de itens	Realizam encontros com o comitê revisor onde são discutidas as características psicométricas dos itens e sua relação com a construção do item
	Professores	Recebem as curvas de desempenho do grupo de estudantes
	Escola	Recebe as curvas de progressão comparando o seu resultado com o consórcio e os microdados
	Comitê revisor	Realiza análise do processo e a avaliação da prova realizada pelos alunos.
Software e Hardware		Fornecido sem custo para o consórcio por uma das escolas participantes

7. DISCUSSÃO

7.1 CARACTERÍSTICAS DOS TESTES

Aferir a qualidade de um teste não é tarefa fácil (KARTHIKEYAN *et al*, 2019), além da validade e confiabilidade vários outros aspectos devem ser levados em consideração como: as características psicométricas dos itens, a funcionalidade dos distratores, o nível taxonômico dos itens e sua adequação ao nível dos estudantes avaliados.

A validade refere-se ao suporte teórico e às evidências empíricas que apoiam as interpretações dos resultados obtidos em um teste relacionando-os aos objetivos propostos (PLESSAS, 2015). A validade de um teste é sempre uma hipótese que requer múltiplas fontes de evidência que apoiem a interpretação dos resultados (DOWNING, 2003 B).

O alinhamento do *blueprint* com o currículo (RAYMOND; GRANDE, 2019), o processo de elaboração dos itens que passam pela revisão do Coordenador local e, posteriormente, é submetido à análise por uma comissão revisora com expertise em elaboração de itens que atua no sentido de minimizar as falhas de construção

constituem-se numa evidência de validade do processo de construção da prova. Outra evidência é obtida a partir dos resultados obtidos pelos estudantes: A demonstração sistemática da progressão de conhecimentos em todos os testes aplicados e a correlação entre os resultados obtidos e a ocorrência de reprovação podem ser também elencadas como evidências de validade dos testes. Este conjunto de evidências suportam a hipótese de que os testes aplicados foram capazes de aferir a progressão de conhecimento dos grupos de estudantes.

Todas as edições do teste apresentaram um alfa de Cronbach acima de 0,8 evidenciando a confiabilidade dos testes produzidos pelo TPCO. A confiabilidade ou fidedignidade de um teste diz respeito a sua capacidade de aferir com o menor erro possível o construto avaliado. O alfa de Cronbach é uma análise de consistência interna do teste que reflete o grau de covariância dos itens entre si. São considerados confiáveis os testes com alfa de Cronbach igual ou maior que 0,8 (PASQUALI, 2009).

A discriminação é o parâmetro que informa sobre a capacidade do item em diferenciar os estudantes que detém o conhecimento avaliado dos que não atingiram este objetivo. É frequentemente considerada como a principal característica psicométrica determinante da qualidade do item (COLLARES; LOGULO; GREC, 2012). Na amostra estudada apenas 39% dos itens alcançaram discriminação igual ou superior a 0,3, sendo considerados itens adequados (QUAIGRAIN; ARHIN, 2017). SAKAI et al. (2011) ao analisar quatro testes aplicados numa universidade estadual brasileira, encontrou 37,2% de questões com discriminação maior ou igual a 0,3, resultado muito próximo do alcançado pelo teste na Escola.

Por outro lado, apenas 4,6% dos itens obtiveram discriminação negativa, ou seja, as questões nas quais os estudantes de baixo desempenho tem um resultado melhor que os estudantes de alto desempenho, normalmente, são questões com falhas graves na sua construção. SWANSON *et al*, (2010) ao analisar um teste de progresso realizado em consórcio por várias escolas inglesas construído por questões que foram aplicadas no exame de licenciamento médico nos estados Unidos da América (USLME) encontraram 3% de itens com discriminação negativa, valor muito próximo do encontrado no nosso trabalho.

Apesar de ter encontrado o mesmo resultado de outros trabalhos quanto ao percentual de questões com bom poder de discriminação, chama a atenção o fato de que, apesar da capacitação dos envolvidos na elaboração das questões ser realizada rotineiramente ao longo dos anos, ela não teve impacto na melhoria das questões ao longo do tempo.

O nível de dificuldade apresentado pelos itens foi de 18% fáceis, 66% médias e 16% difíceis. Sabe-se que o nível de dificuldade ideal dos itens depende dos objetivos do teste, dessa forma, testes destinados a processos seletivos, que têm como objetivo selecionar os candidatos de alto desempenho, devem ter um número maior de questões difíceis, enquanto testes com objetivos formativos devem ser mais abrangentes e conter um número equilibrado de questões fáceis, médias e difíceis. O fato de o teste ter apresentado itens com dificuldade adequada para todas as faixas de desempenho pode-se considerar que os testes aplicados podem ser considerados adequados para os objetivos propostos (PASQUALI, 2003).

Outro parâmetro que pode ser usado como indicador da qualidade de um teste é a número de distratores utilizados nos itens, quanto mais distratores forem utilizados maior será a discriminação dos itens (TARRANT; WARE; MOHAMMED, 2009). Na nossa amostra, em média 65% dos itens por teste utilizaram todos os distratores. Haladyna e Downing (1993) relatam que ao analisar quatro testes de diferentes áreas encontraram apenas de 1,1 a 8,4% dos itens utilizando todos os distratores.

Os itens utilizados nos testes foram baseados em vinhetas e/ou envolveram aplicação do conhecimento, sendo considerados, portanto, itens de nível taxonômico elevado (KRATHWOHL, 2002) e que exigem um processamento cognitivo mais complexo (CECILIO-FERNANDES *et al*, 2018). Este tipo de item é apontado como o mais adequado para avaliar os estudantes do final do curso, por ser mais efetivo na indução da consolidação do conhecimento (JENSEN *et al*, 2014).

É importante considerar que no presente trabalho a discriminação foi calculada com base nos resultados obtidos por todas as séries e não somente do sexto ano, lembrando que o teste é construído com o objetivo de avaliar os concluintes do curso. Como os itens que compõem a prova são de nível taxonômico elevado e, portanto, muito difíceis para os estudantes das séries iniciais do curso a discriminação

dos itens nestas séries é muito baixa. Assim sendo, o tipo de questão adotado pode também ter contribuído para o baixo percentual de itens discriminativos.

7.2 DESEMPENHO DOS ESTUDANTES

A interdisciplinaridade e a aprendizagem em espiral, em que o mesmo tema é abordado diversas vezes ao longo do currículo numa ordem crescente de complexidade, são alguns dos princípios pedagógicos que norteiam o processo de aprendizagem no PBL. (BARROWS, 1996; HMELO; GOTTERER; BRANSFORD, 1997; HMELO-SILVER, 2004) Espera-se, portanto, que o PBL favoreça o ganho progressivo e constante de conhecimentos ao longo do Curso. O teste de progresso, pelas suas características, seria capaz de aferir esta progressão.

Como esperado, os resultados deste estudo evidenciaram um padrão geral de progressão de conhecimentos aproximadamente linear da primeira à sexta série do curso. Este mesmo padrão já foi relatado em diversos trabalhos utilizando o teste de progresso (BLAKE *et al*, 1996; VAN DER VLEUTEN; VERWIJNEN; WIJNEN, 1996; ALBERS *et al*, 1989; VERHOEVEN *et al*, 2002). Desvios deste padrão são normalmente atribuídos a fatores curriculares (MUIJTJENS, *et al*. 2008a; MUIJTJENS *et al*. 2008b; ALBANO, *et al*. 1996), daí a importância da utilização do TP como instrumento de avaliação do desenvolvimento dos programas educacionais e seu uso.

O ganho médio de conhecimento acrescentado por cada série do curso, calculado pelo d de Coehn, foi de 0,69 por ano. GÖRLICH e FRIEDERICHS (2021) ao utilizar o teste de progresso relataram um acréscimo médio de médio de 0,53 por ano em um curso de medicina. Um desvio da linearidade da curva foi evidenciado quando comparadas as séries iniciais do Curso com as duas séries finais, correspondentes ao Estágio Curricular Obrigatório (ECO). Enquanto o ganho médio de conhecimento foi de 0,95 por série de primeira à quarta série enquanto no ECO este índice médio foi de 0,31.

Essa diminuição da taxa de progressão de conhecimento no ECO pode estar relacionada a diversos fatores: O objetivo principal do estágio é o treinamento prático supervisionado e nesta etapa do curso o estudante realiza uma imersão no serviço de saúde, com uma redução expressiva de atividades teóricas e, alinhado com este

objetivo, todas as avaliações realizadas no ECO são práticas (PPC-ESCS,2021). É possível que esta diminuição na progressão de conhecimentos identificada pelo TP esteja relacionada à menor prioridade dada aos aspectos cognitivos neste período do curso. Outra hipótese é a de que falhas na estruturação do estágio estejam dificultando a integração e correlação entre as experiências práticas vivenciadas e os conhecimentos teóricos correspondentes.

Outro desvio da linearidade da curva é observado na quarta série do Curso o que pode estar relacionado a um aumento no conteúdo cognitivo nesta série desproporcional em relação às demais séries.

Não foi observada diferença de desempenho entre os gêneros, resultado que está de acordo com outros estudos utilizando o teste de progresso (FINUCANE *et al.*, 2013; TOMIC *et al.*, 2005; ROSA *et al.*, 2016). Nossos dados também mostraram uma correlação fraca entre a participação nos programas de iniciação científica e desempenho no TP. Como a participação nos programas de iniciação científica é opcional, é possível que este achado se explique pela maior probabilidade de estudantes de maior desempenho procurarem este tipo de atividade.

Foi observada uma diferença significativa de desempenho no TP entre o grupo de estudantes que tiveram alguma reprovação e seus colegas que não tiveram reprovação no curso. Este achado constitui-se em mais uma evidência de validade do TP em avaliar a progressão de conhecimentos dos alunos.

Foi evidenciado também uma diferença significativa de desempenho entre estudantes admitidos pelo sistema de cotas e seus colegas admitidos pelo sistema universal. Esta diferença existe desde o início do curso, segue constante até a quarta série e aumenta nos dois últimos anos. A diferença é pequena, de em média 0,3 desvio padrão, com uma larga faixa de sobreposição entre os dois grupos.

A reserva de 50% das vagas nas universidades públicas federais para estudantes oriundos do ensino público, proporcionalmente à composição racial da população de cada estado, foi estabelecida, no Brasil, por meio da lei federal 12.711 (Brasil, 2012). A instituição do sistema de cotas no ensino superior brasileiro logo provocou um intenso debate de fundo ideológico, discussões sobre equidade, meritocracia, justiça social e se o acesso de estudantes de menor rendimento

acadêmico não iria diminuir o padrão dos egressos. Até hoje, o sistema de cotas é objeto de debate acalorado e tema de muitos estudos. Os estudos brasileiros que analisaram o desempenho de estudantes admitidos por ações afirmativas no Brasil obtiveram resultados diversos:

CUNHA (2006) relatou um índice de 71% de cotistas com desempenho alto contra 48% de desempenho alto entre não cotistas no primeiro semestre do curso de medicina, já CARDOSO (2008), não encontrou diferença de desempenho entre cotistas e não cotistas no primeiro semestre do curso de medicina na mesma instituição. VELLOSO (2009) também na mesma instituição, analisou o desempenho de estudantes em três turmas consecutivas utilizando a média das notas nas disciplinas. Seus resultados referentes ao curso de medicina foram: na primeira turma não houve diferença entre os grupos, nas outras duas o desempenho dos estudantes admitidos pelo sistema universal foi maior.

MENDES JUNIOR (2014) analisou o desempenho entre cotistas e não cotistas na coorte dos ingressantes no vestibular de 2005, relatou um desempenho menor no grupo dos cotistas quando comparado aos admitidos pela ampla concorrência; também observou que esta diferença era significativamente maior nos cursos de alta dificuldade relativa, porém sua amostragem não incluiu o curso de medicina.

QUEIROZ (2015) estudou o rendimento no primeiro semestre de alunos ingressantes segundo a modalidade de cota e a área de cada curso. Na área de ciências biomédicas não houve diferença entre alunos ingressantes pelas ações afirmativas quando comparados à ampla concorrência.

PEIXOTO (2016) relatou um desempenho geral superior dos alunos admitidos pelo sistema de cotas, porém na área de biológicas/saúde encontrou um maior número de cursos com desempenho superior de não cotistas, no entanto, segundo o autor, foi excluído da amostra o curso de medicina por “apresentar uma concorrência muito alta e com isso distorcer os resultados”.

VALENTE e BERRY (2017) ao comparar o desempenho no ENADE de 2010 no grupo das escolas públicas encontrou, no curso de medicina, quando corrigidos para todos os outros fatores, um desempenho de apenas 2,39 pontos menor no grupo dos cotistas.

WAINER e MELGUIZO (2018) também ao utilizar os resultados agregados do ENADE para todos os cursos não encontrou diferenças de desempenho entre alunos admitidos pelas ações afirmativas e os admitidos pelo acesso universal.

GOLGHER (2021), utilizando dados agregados de todos os cursos, encontrou que a diferença média das notas existente a favor dos não cotistas no primeiro semestre não era mais observada no terceiro semestre.

MARCOMINI NETO et al. (2022) ao analisar a média aritmética das notas finais das disciplinas do primeiro ao terceiro ano de um curso de medicina relata que no primeiro ano as notas dos ingressantes pela concorrência universal era mais alta do que os ingressantes pelas ações afirmativas, esta diferença não apareceu no segundo ano e no terceiro ano houve um reaparecimento da diferença, no entanto menor do que a observada na série inicial.

As diferenças nos indicadores, nos cursos estudados, nas metodologias de estudo empregadas e nos tipos de ações afirmativas são provavelmente responsáveis pela diversidade de resultados encontrados.

Diferentemente dos nossos resultados, a maioria dos trabalhos parece indicar que, de uma maneira geral, não existe diferença de desempenho acadêmico entre estudantes admitidos pelo sistema de cotas e os do sistema universal. Novamente, parece adequado supor que as diferenças metodológicas entre os trabalhos possam explicar os resultados discrepantes. Pelo nosso conhecimento, o presente trabalho é o primeiro no Brasil a realizar esta comparação utilizando testes padronizados, num modelo de estudo longitudinal a longo prazo.

O fato de nossos resultados indicarem que a diferença de desempenho já existente desde a admissão persistir até o final do curso sugere que fatores independentes da escola podem estar contribuindo para os resultados. A diferença de nível socioeconômico entre os estudantes admitidos pelo sistema de cotas e pelo sistema universal na escola relatado por BIAZOTTO (2022) é evidenciada pela renda familiar dos estudantes admitidos pelo sistema de cotas sendo cerca de 20% da dos estudantes admitidos pelo sistema universal. Na amostra estudada por SILVA (2013) na mesma escola, evidenciou-se que somente 25% dos pais dos ingressantes pelo sistema de cotas concluíram o ensino superior, enquanto no grupo de acesso universal este número é de aproximadamente 70%.

Existem evidências de que fatores socioeconômicos têm pouca influência no desempenho acadêmico no ensino superior (RODRÍGUEZ-HERNÁNDEZ, CASCALLAR, KYNDT, 2020), no entanto, é possível que o conjunto de fatores de ordem pessoal, familiar e socioeconômico que são determinantes do desempenho acadêmico possam explicar, pelo menos em parte, porque o desempenho entre os cotistas e não cotistas não sejam iguais ao longo do curso.

Além da diferença de desempenho persistir durante todo o curso observou-se que ela aumenta nos dois últimos anos. É possível que diferenças socioeconômicas e de perspectivas quanto ao futuro profissional possam estar relacionadas a este resultado. No Brasil, após a graduação, os egressos dos cursos de medicina têm a opção de seguir os estudos por meio de uma especialização, sendo a forma mais comum a residência médica, ou já ingressar diretamente no mercado de trabalho sem uma especialização. Um estudo realizado na escola (BIAZOTTO et al., 2022) encontrou um percentual significativamente menor acesso a residência médica entre os egressos admitidos pelo sistema de cotas quando comparados com seus pares.

O número limitado de vagas nos programas de especialização/residência médica faz com que o processo seletivo de entrada, baseado em provas de alta concorrência, seja muito disputado. Uma parte dos estudantes inicia a preparação para estas provas já durante os últimos anos da graduação por meio de cursos preparatórios que têm um custo elevado, o que pode ser um fator restritivo para estudantes de menor renda.

A diferença de progressão de conhecimentos observada entre estudantes admitidos pelo sistema de cotas e pelo acesso universal não significam, necessariamente, diferenças na competência profissional dos futuros médicos. Além da diferença entre os dois grupos ter sido pequena, existindo uma larga faixa de sobreposição no desempenho dos dois grupos, é importante observar que a competência profissional, além de ser contexto dependente, engloba outros aspectos além do conhecimento como por exemplo a capacidade de comunicação, emoções, habilidades técnicas e valores, entre outras (EPSTEIN; HUPERT, 2000).

7.3 UTILIZAÇÃO DO TESTE PARA FEEDBACK INDIVIDUAL

Além de fornecer feedback para a instituição sobre o desenvolvimento dos programas educacionais é desejável que o teste seja capaz de fornecer para o estudante um feedback detalhado sobre a progressão de seus conhecimentos identificando inclusive áreas em que os objetivos já foram alcançados e áreas que ainda requerem atenção.

Para analisar a confiabilidade dos resultados a nível individual foi calculado o Índice de Confiabilidade Mudança (RCI) (ZAHRA et al, 2016), que nos informa basicamente qual o grau de confiabilidade de que uma diferença entre dois resultados represente uma mudança real na variável de estudo ou se deve ao acaso.

Os resultados indicam que, nas condições do estudo, o aumento no desempenho entre os testes só pode ser interpretado de maneira confiável como ganho de conhecimento para apenas 14% dos estudantes, fato que se confirma com relatos da literatura que indicam que testes com o objetivo de analisar o desempenho individual dos estudantes ou com objetivos somativos requerem amostragens maiores (MUIJTJENS, *et al.* 1998). Dessa forma, os resultados obtidos indicam que para atingir este objetivo seria necessário um aumento da amostragem seja por meio do aumento do número de itens e/ou aumento do número de provas no ano.

7.4 ANÁLISE DOS PROCEDIMENTOS UTILIZADOS PARA A REALIZAÇÃO DO TESTE

7.4.1 Construção do teste

A rotina de construção do teste obedeceu às etapas preconizadas: As questões do teste foram definidas conforme o blueprint adotado, o coordenador local de cada escola recebe sua lista de questões e distribui para os elaboradores locais especialistas no assunto. Existe relato na literatura de que este processo de elaboração de itens por diferentes instituições poderia constituir um viés na construção da prova (MUIJTJENS, 2007). Após a elaboração as questões são revisadas por uma comissão revisora que pode aceitar o item, fazer alterações ou retornar para o elaborador com recomendações para melhorias.

O processo de produção dos itens é uma etapa fundamental para a qualidade do teste, gerar um grande número de itens qualidade regularmente é reconhecidamente uma tarefa desafiadora (JOZEFOWICZ et al. 2002; STAGNARO-GREEN; DOWNING, 2006). A inclusão de itens com erros de construção introduz no teste variáveis estranhas e não controladas produzindo a chamada variância irrelevante de construto (DOWNING, 2005). Itens falhos, como por exemplo, itens em que existe uma “pista” que permite aos estudantes assinalar a resposta correta sem deter o conhecimento avaliado, podem produzir uma distorção nos resultados, reduzindo a precisão e dificultando a interpretação.

A função do grupo revisor é tentar garantir que os itens produzidos sigam as regras estabelecidas nos guias para construção de testes (PANIAGUA; SWYGERT, 2016). Vários trabalhos destacam o papel fundamental da comissão que realiza a revisão dos itens (WRIGLEY *et al.*, 2012) e da capacitação dos elaboradores (NAEEM; VAN DER VLEUTEN; ALFARIS, 2012); HAMAMOTO FILHO; BICUDO, 2020) na garantia da qualidade do teste.

Apesar do treinamento em elaboração dos itens propiciado pelo feedback dado para os envolvidos, em média, menos da metade dos itens apresentou discriminação satisfatória. Mesmo com a capacitação sendo realizada anualmente a discriminação dos itens não apresentou melhoria ao longo do tempo. Outros fatores como a baixa motivação e a pequena valorização dada a deste tipo de tarefa, tanto pelos elaboradores quanto pelas instituições, podem ser apontados como dificultadores para a construção de itens de boa qualidade (KARTHIKEYAN, O'CONNOR, HU; 2019).

Pela sua característica de ser um consórcio aberto à entrada de novos membros, com grande diversidade de instituições participantes, faz com que ocorra uma grande rotatividade e diversidade entre os elaboradores dos itens. É possível que estas características do consórcio também estejam contribuindo para a dificuldade na melhoria da qualidade dos testes ao longo do tempo.

O tipo de item definido pelo TPCO, na sua grande maioria no formato de casos clínicos, avalia principalmente a capacidade de raciocínio e a aplicação do conhecimento. Por necessitarem de um maior processamento cognitivo este tipo de

item influencia positivamente na retenção do conhecimento na memória de longo prazo (JENSEN *et al.*, 2014). Por outro lado, a taxonomia elevada dos itens (KRATHWOHL, 2002) fazem com que o teste tenha baixa discriminação para avaliar as séries iniciais do Curso (CECILIO-FERNANDES *et al.*, 2016), uma vez que é necessário conhecer fatos e mecanismos e entender as relações entre eles antes de conseguir aplicar o conhecimento.

O TPCO não realiza um procedimento de *standard setting* para a construção testes, isto pode introduzir uma dificuldade de comparabilidade dos resultados e de uma possível inadequação dos itens ao nível dos estudantes, no entanto, nossos resultados indicam um nível de dificuldade média do teste que permaneceu notavelmente constante em torno de 50%, o que mostra a adequação dos itens ao nível dos estudantes da Escola.

7.4.2 Aplicação

Em média, 88% dos estudantes do Curso participaram do teste, mesmo sendo a participação de caráter optativo, o que indica a valorização do TP pelos estudantes. O objetivo principal do TP é analisar o desenvolvimento dos programas educacionais, para tanto, a frequência de aplicação e número de questões do teste (RICKETTS, *et al.*, 2010) devem estar de acordo com este objetivo. A confiabilidade média alcançada, de 0,88 nos permite concluir que o número e a qualidade dos de itens foram adequados para o objetivo proposto. Mesmo com objetivos apenas formativos a aplicação simultânea e presencial dos testes são mais um fator de garantia da confiabilidade e validade dos resultados obtidos.

7.4.3 Análise dos resultados e revisão (Controle de qualidade)

O TPCO expressa os resultados tanto em escores absolutos (percentual de acertos) quanto em escores relativos (posição do estudante com relação ao seu grupo). Existe uma discussão na literatura quanto a maneira mais adequada de expressar o resultado do TP. A utilização de escores absolutos tem se mostrado pouco defensável em função da variabilidade no nível de dificuldade dos testes, por outro lado, o resultado expresso de forma normoreferenciada não deixa claro para o estudante qual o desempenho considerado adequado (MUIJTJENS *et al.*, 1998).

O uso de fórmulas como por exemplo pelo número de respostas certas menos as respostas erradas se mostram mais adequados para itens onde a probabilidade de acerto ao acaso é elevada como por exemplo itens de verdadeiro ou falso.

A tentativa de inibir o acerto ao acaso com o uso de apenação (respostas erradas diminuindo o escore das respostas certas) ou com a inclusão da alternativa “não sei” (MUIJTJENS *et al.*, 1999) introduzem no resultado uma outra variável que é a tendência ou não do estudante em correr o risco (DOWNING, 2003 A; CECILIO-FERNANDES *et al.*, 2017) o que pode diminuir a confiabilidade do teste. Portanto a opção de expressão dos resultados realizada pelo TPCO parece adequada em função dos objetivos e características da prova.

A ausência de um procedimento de *standard setting* (DOWNING; TEKIAN; YUDKOWSKY, 2006; RICKETTS; FREEMAN; COOMBES, 2009) pelo TPCO e a não realização da equiparação de escores (LANGER; SWANSON, 2010) tornam difíceis a comparação dos desempenhos entre os diferentes testes que podem ser muito úteis no acompanhamento de longo prazo da qualidade dos programas educacionais.

Os microdados fornecidos para cada escola permitiram o desenvolvimento de *expertise* própria em análise de testes. A triangulação dos resultados do TP com outras informações, como por exemplo as realizadas neste trabalho, são bastante úteis como feedback para a gestão educacional. Outro subproduto valioso produzido pelo TPCO é a possibilidade da construção de um grande banco de itens, pré-condição necessária para a adoção de novas tecnologias educacionais como os testes adaptativos (COLLARES; CECILIO-FERNANDES, 2018).

7.4.4 Feedback para os envolvidos

O TPCO cumpre a etapa de fornecer *feedback* para os envolvidos no teste. O estudante recebe o caderno de provas, o gabarito com as questões comentadas e as respectivas referências para cada item, eventuais discordâncias do gabarito são passíveis de recurso por parte do estudante. Ao final do processo o estudante também recebe um relatório individual onde pode comparar seu resultado com a média dos colegas da mesma série, da própria escola e também do consórcio. Esta rotina possibilita ao estudante aprender com o teste e posicionar seu desempenho com

relação aos colegas da mesma série na escola e entre diferentes escolas, porém, a efetividade desta estratégia como indutora do aprendizado ainda está por ser determinada.

Idealmente, o TP deveria prover feedback para os estudantes com o objetivo de estimular a identificação de eventuais lacunas de aprendizagem e estimular a aprendizagem autogerida, todavia, o cálculo do RCI indica que existe uma baixa confiabilidade dos resultados a nível individual, o que limita o fornecimento deste tipo de feedback. A possibilidade de o teste passar a fornecer feedback a nível individual está ligada principalmente ao aumento das quantidades de testes e itens aplicados, como todo o ciclo necessário para a utilização do teste é trabalhoso e consome muitos recursos seria necessária uma análise da relação entre a ampliação dos objetivos do teste e os custos envolvidos.

O principal *feedback* recebido pelas instituições são as curvas de progressão gerais e pelas áreas especificadas no *blueprint*, de posse destes dados, as escolas podem analisar o desenvolvimento dos seus programas educacionais e identificar áreas que eventualmente requeiram atenção. Além disso, os microdados podem ser utilizados para outras análises. Conforme demonstrado no presente estudo, a correlação com outras variáveis educacionais foi capaz de fornecer importantes informações para a gestão educacional.

Os resultados dos TP das escolas participantes do Consórcio não são tornados públicos pois existe o risco de que um ranqueamento entre as escolas leve ao esvaziamento do consórcio e que a competição entre elas produza uma distorção no objetivo eminentemente formativo do teste. O único *benchmark* que é divulgado é o resultado geral de todas as escolas em cada série.

Se por um lado é compreensível o receio de que a identificação das escolas envolvidas e o ranqueamento decorrente leve a uma distorção do objetivo formativo do teste, este receio torna os dados das diferentes escolas praticamente inacessíveis aos pesquisadores. Esse fator limitador enfraquece o potencial de pesquisa acadêmica que poderia ser fornecido pelo teste.

Apesar do teste atingir os objetivos propostos de maneira confiável em todo o período, o *feedback* para os elaboradores e revisores não parece ter surtido efeito na melhoria da discriminação dos testes produzidos ao longo do período analisado. Este

resultado é particularmente importante uma vez que a qualificação dos elaboradores é fundamental na produção de itens de qualidade (JOZEFOWICZ *et al.* 2002; HAMAMOTO FILHO; BICUDO, 2020). Mais estudos são necessários para estabelecer as causas e as ações necessárias para a melhoria da qualidade dos itens produzidos.

8. CONCLUSÕES

O presente trabalho realizou uma análise dos testes de progresso produzidos em consórcio pelo grupo de escolas participantes do TPCO e aplicado num curso de medicina em uma escola pública do Distrito Federal. O objetivo principal do teste era avaliar o desenvolvimento dos programas educacionais da instituição.

Os resultados mostram que os testes aplicados foram válidos e confiáveis para atingir o objetivo proposto. A análise psicométrica dos itens mostrou que o nível de dificuldade do teste foi adequado, porém, menos da metade dos itens apresentaram discriminação adequada, corroborando os relatos na literatura sobre a dificuldade de produção de itens de boa qualidade. Apesar de os fatores que levam a dificuldade em produzir itens de qualidade são múltiplos e ainda mal compreendidos (KARTHIKEYAN, O'CONNOR e HU, 2019) levantamos a hipótese de que o tipo de treinamento utilizado, a baixa motivação dos elaboradores e o tipo de item definido para ser utilizado no teste podem ter contribuído para resultados encontrados. Nossos resultados indicam que existe a necessidade de o consórcio analisar o processo de produção dos itens no sentido de determinar as causas e as ações específicas necessárias para sua melhoria.

A participação maciça no teste, mesmo sendo optativa, evidenciou sua valorização pelos estudantes. Como esperado, os testes evidenciaram a progressão significativa de conhecimento dos estudantes a cada série. Esta progressão foi aproximadamente linear, no entanto, quando analisado o tamanho do efeito de cada série sobre o aumento de conhecimento, observou-se uma diminuição da taxa de progressão nos dois últimos anos do curso, período correspondente ao Estágio Curricular. É possível que essa diminuição na taxa de progressão esteja relacionada à maior ênfase em atividades e avaliações práticas ou que falhas existentes na estrutura curricular do ECO estejam contribuindo para o resultado. Estes resultados

demonstram a importância do feedback oferecido pelo TP à instituição, possibilitando o controle de qualidade do ensino oferecido.

O teste não mostrou diferença de desempenho entre homens e mulheres, a participação no programa de iniciação científica mostrou uma correlação fraca com o melhor desempenho no TP, provavelmente este resultado está relacionado ao maior interesse dos alunos de maior desempenho pelo programa. Os resultados também evidenciaram uma correlação significativa entre o desempenho no teste e as reprovações no curso, fornecendo mais uma evidência de validade do teste de progresso.

As ações afirmativas têm como objetivo promover o acesso de estudantes de classes socioeconômicas menos favorecidas ao ensino superior. Há evidências de que este objetivo foi alcançado na escola onde além de ter um nível socioeconômico mais baixo os estudantes admitidos pelo sistema de cotas têm uma menor evasão do que os admitidos pelo sistema universal (SILVA, 2013; LIMA et al., 2022).

A adoção do sistema de cotas para acesso ao ensino superior suscitou um forte debate de fundo ideológico na sociedade e também no meio acadêmico e que se refletiram nos trabalhos existentes na literatura. Dúvidas sobre o desempenho dos ingressantes pelo sistema de cotas são frequentemente levantadas. Para contribuir para o debate, o presente trabalho comparou o desempenho de estudantes admitidos pelo sistema de cotas e pelo sistema universal no teste de progresso.

Nossos resultados mostraram uma diferença significativa de desempenho entre os estudantes admitidos pelo sistema de cotas e seus colegas admitidos pelo sistema universal durante todo o curso. A diferença observada, apesar de estatisticamente significativa foi pequena, havendo uma larga faixa de sobreposição entre os dois grupos. A diferença foi constante da primeira à quarta série e aumentou nos dois últimos anos do curso que correspondem ao estágio curricular obrigatório. Diferenças no acesso aos cursos preparatórios para a residência que são privados e de alto custo e/ou diferenças nas perspectivas de inserção no mercado de trabalho podem estar envolvidas no aumento da diferença observado no final do Curso.

Os resultados de estudos comparando o desempenho de ingressantes pelo sistema de cotas e seus colegas que acessaram o curso pela concorrência universal

variam muito, eles diferem quanto ao indicador de desempenho, tipo de estudo, local do estudo, tipo de curso entre outros, no entanto, de uma maneira geral, a maioria dos trabalhos indica que o desempenho entre os dois grupos ou é igual ou vai se igualando durante o curso. Diferentemente destes trabalhos, nossos resultados mostram que a diferença, apesar de pequena, existe, persiste durante todo o curso e aumenta nos dois últimos anos. É possível que diferenças nas metodologias de estudo, nos cursos, tipo de cotas e locais estudados sejam responsáveis pela discrepância. Pelo nosso conhecimento este é o primeiro trabalho brasileiro a fazer esta comparação utilizando testes padronizados num desenho longitudinal.

Esta diferença de desempenho entre cotistas e não cotistas pode estar relacionado à complexa interação de fatores de ordem pessoal, familiar, socioeconômica e da escola que são determinantes do desempenho acadêmico. Os diversos fatores que diferenciam os dois grupos podem explicar porque a escola não foi capaz de equalizar o desempenho no decorrer do curso. É possível que a diferença observada entre os ingressantes pelo sistema de cotas e seus colegas admitidos pelo sistema universal, possa ser diminuída com o fortalecimento do apoio aos estudantes com desempenho insuficiente, cabendo avaliar se é necessário a adoção de novas medidas de suporte para além das já existentes.

É muito provável que esta diferença, apesar de estatisticamente significativa, não tenha magnitude suficiente para influenciar no alcance das competências esperadas ao final do curso, porém, mais trabalhos são necessários para esclarecimento desta hipótese. Importante ressaltar que a diferença de progressão de conhecimentos observada entre estudantes admitidos pelas ações afirmativas e pelo acesso universal não significa necessariamente diferenças na competência profissional. Nossos resultados vêm se somar ao longo debate existente no meio educacional sobre o peso dos fatores pessoais, socioeconômicos e da escola no desempenho acadêmico.

Todo o processo envolvido na realização do teste pelo Consórcio do Centro Oeste foi analisado de acordo com o referencial teórico publicado pela Associação para a Educação Médica da Europa (Wrigley et. al. 2012). A construção do teste seguiu as etapas preconizadas A construção dos itens, realizada em conjunto pelas instituições participantes, seguiu as normas preconizadas com exceção do

procedimento de *standard setting*. A aplicação simultânea e presencial em todas as escolas participantes garantiu a confiabilidade dos resultados. A análise dos resultados se mostrou adequada ao tipo e ao objetivo do teste, o processo de revisão possibilitou o controle de qualidade dos testes e favoreceu o desenvolvimento de expertise própria em análise psicométrica e no controle de qualidade em avaliação.

O feedback para os estudantes foi realizado por meio do fornecimento do resultado acompanhado das questões comentadas e das respectivas referências e pelo fornecimento da comparação do resultado individual com o resultado dos colegas da mesma série na escola e no consórcio. A eficácia deste feedback fornecido atualmente para os estudantes em promover o aprendizado ainda está por ser determinada. Seria desejável que fosse fornecido um feedback mais específico indicando com confiabilidade a progressão de conhecimentos individual do estudante, porém, o cálculo do RCI indicou que esta informação só pode ser dada com confiabilidade para uma pequena parcela dos estudantes.

O objetivo principal das metodologias ativas é desenvolver a capacidade de aprendizado do estudante fazendo com que ele adquira autonomia e controle sobre seu desenvolvimento profissional, prover este feedback deve ser um dos objetivos fundamentais a ser buscado pelo sistema de avaliação da instituição, portanto, um feedback eficaz constitui-se num dos pilares do processo de aprendizado (NORCINI, 2010). Idealmente o teste de progresso deveria prover um feedback mais específico para os estudantes, informando sobre as áreas em que os objetivos educacionais já foram atingidos e áreas que ainda requerem mais estudo, porém a amostragem de cem itens aplicados anualmente pelo TPCO mostrou-se insuficiente para tal. A ampliação dos objetivos do teste no sentido de fornecer este feedback mais específico dependeria do aumento do número de itens e/ou do número de testes no ano, o que acarretaria na necessidade maiores recursos investidos. O compartilhamento de custos pelo consórcio pode ser uma estratégia capaz de prover os recursos necessários para atingir este objetivo.

O TPCO realiza rotineiramente o teste de progresso utilizando exclusivamente recursos próprios das instituições envolvidas, não existindo a contratação de qualquer serviço externo nem qualquer obrigação financeira das instituições envolvidas. A realização do teste pela escola em conjunto com o Consórcio do Centro Oeste

mostrou-se uma estratégia eficiente, eficaz e de baixo custo para a avaliação dos programas educacionais da escola.

A participação da escola no TPCO mostrou-se uma estratégia capaz atingir os objetivos propostos, tornando possível realizar adequadamente todas as etapas do processo com otimização de recursos. Para além da questão dos custos, cabe destacar que os resultados exitosos são frutos de uma abordagem cooperativa, agregadora e capaz de promover o desenvolvimento institucional, tanto pela análise da progressão de conhecimento dos estudantes quanto pela possibilidade de comparação com o referencial externo. Os microdados gerados também possibilitaram análises de diversos temas de interesse da instituição e o desenvolvimento de expertise em avaliação. Nossos resultados demonstram a importância da realização do teste de progresso em consórcio com outras instituições para viabilizar o desenvolvimento institucional.

Nosso estudo analisou as características da curva de crescimento de conhecimento dos estudantes correlacionando-a a algumas variáveis e levantando hipóteses explicativas para os resultados encontrados, mais estudos são necessários para testar as hipóteses levantadas.

O presente trabalho tem a limitação de ser um estudo de caso, cujos resultados devem ser interpretados somente no contexto da instituição estudada e tendo em vista suas peculiaridades: ser um curso de alta concorrência, numa escola pública da capital do país, que usa metodologia ativas de ensino aprendizagem e que adota um sistema de cotas baseado em critérios socioeconômicos.

Os dados gerados pelo conjunto de escolas participantes do consórcio demonstraram ser uma rica fonte de informações que poderiam ser utilizadas com fins de pesquisa educacional, contribuindo para o desenvolvimento da educação médica do país. No entanto, as regras atuais, que condicionam o acesso a estas informações a permissão de cada instituição, sob o receio de que ocorra seu uso para fins diversos à pesquisa não permitem este desenvolvimento. O estabelecimento pelo consórcio de novas regras que permitam aos pesquisadores o acesso aos dados e que ao mesmo tempo garantam o sigilo sobre a identificação de sua origem abririam a perspectiva de um maior desenvolvimento e a valorização do teste de progresso.

9. REFERÊNCIAS

1. AAMC. Physicians for the twenty-first century. Report of the Project Panel on the General Professional Education of the Physician and College Preparation for Medicine. **J Med Educ**. v.59, n.11, p. 1-208, 1984
2. ALBANESE, Mark. *Benchmarking progress tests for cross-institutional comparisons: Every road makes a difference and all of them have bumps*. **Medical Education**, v. 42, n. 1, p. 4–7, 2008.
3. ALBANESE, Mark; CASE, Susan M. *Progress testing: Critical analysis and suggested practices*. **Advances in Health Sciences Education**, v. 21, n. 1, p. 221–234, 2016.
4. ALBANO, M G *et al.* An international comparison of knowledge levels of medical students: the Maastricht Progress Test. **Medical education**, [s. l.], v. 30, n. 4, p. 239–245, 1996.
5. ALBERS, W. *et al.* A stochastic growth model applied to repeated tests of academic knowledge. **Psychometrika**, v. 54, n. 3, p. 451–466, 1989.
6. AL-KADRI, Hanan M.; AL-MOAMARY, Mohamed S.; ROBERTS, Chris; VAN DER VLEUTEN, Cees P.M. *Exploring assessment factors contributing to students' study strategies: Literature review*. **Medical Teacher**, v. 34, n. SUPPL. 1, 2012.
7. ANTUNES DOS SANTOS, Renato; NUNES, Maria do Patrocinio Tenorio. Medical education in Brazil. **Medical Teacher**, v. 0, n. 0, p. 1–6, 2019.
8. ARNOLD, Louise; WILLOUGHBY, Lee. *The quarterly profile examination*. **Academic Medicine**, v. 65, n. 8, p. 515–516, 1990.
9. BARROWS, Howard S. Problem-Based Learning in Medicine and Beyond : A Brief Overview. **New Directions for Teaching and Learning**, [s. l.], v. winter, n. 68, p. 3–12, 1996.
10. BECKER, Lee A. Effect Size (ES). 2000.available at [EffectSizeBecker.pdf \(uv.es\)](http://EffectSizeBecker.pdf(uv.es))
11. BERBEL, Neusi Aparecida Navas. Metodologia da problematização: uma alternativa metodológica apropriada para o ensino superior. **Semina: Ciências Sociais e Humanas**, v. 16, n. 3, p. 09, 1995.
12. BIAZOTTO, Marize *et al.* *Comparison Between Students Admitted Through Regular Path and Affirmative Action Systems in a Brazilian Public Medical*

- School. **Advances in Medical Education and Practice***, v. 13, n. March, p. 251–263, 2022.
13. BICUDO, A.M. HAMAMOTTO FILHO, P.T.; ABBADE, J.F.; HAFNER, M.L.M.B.; MAFFEI, C.M.L. Teste de Progresso em Consórcios para Todas as Escolas Médicas do Brasil. **Revista Brasileira de Educação Médica**, v. 43, n. 4, p. 151–156, 2019.
 14. BLAKE, Jennifer M. et al. Introducing progress testing in McMaster University's problem-based medical curriculum. **Academic Medicine**, v. 71, n. 9, p. 1002–7, 1996.
 15. BORDAGE, Georges. Elaborated Knowledge, *A Key to Successful Diagnostic Thinking*. **Academic Medicine**, v. 69, n. 11, p. 883–885, 1994.
 16. BRASIL. Lei Nº 12.711, de 29 de agosto de 2012 - **Diário Oficial da União**. v.1 p.3 2012.
 17. BRAUER, David G.; FERGUSON, Kristi J. The integrated curriculum in medical education: AMEE Guide No. 96. **Medical Teacher**, [s. l.], v. 37, n. 4, p. 312–322, 2015.
 18. BROWN, John Seely; COLLINS, Allan; DUGUID, Paul. Situated cognition and the culture of learning. **Educational Researcher**, [s. l.], n. jan-feb, p. 32–42, 1989.
 19. CARDOSO, Claudete Batista. **Efeitos da política de cotas na Universidade de Brasília: uma análise do rendimento e da evasão**. 2008. 134 f. dissertação de Mestrado. Universidade de Brasília, 2008. Disponível em: <<http://repositorio.unb.br/handle/10482/1891>>.
 20. CECILIO-FERNANDES, Dario et al. *Development of cognitive processing and judgments of knowledge in medical students: Analysis of progress test results*. **Medical Teacher**. v. 38, n. 11, p. 1125–1129, 2016.
 21. CECILIO-FERNANDES, Dario et al. *Comparison of formula and number-right scoring in undergraduate medical training: A Rasch model analysis*. **BMC Medical Education**, v. 17, n. 1, p. 1–9, 2017.
 22. CILLIERS, Francois J. **The Pre-assessment Learning Effects of Consequential Assessment: Modelling how the Examination Game is Played**. 2012. 209 f. Tese de Doutorado. Maastricht University, 2012.
 23. COHEN, Jacob. **Statistical power analysis for the behavioral sciences**. Routledge, 2013.

24. COLEMAN, J S. CAMPBELL, E Q. HOBSON, C J. MCPARTLAND, J. MOOD, A M. WEINFELD, F D; YORK, R L. *Equality of Educational Opportunity*. Washington, DC: **US Department of Health, Education & Welfare**. Office of Education (OE-38001 and supp.), 1966. 548 p.
25. COLLARES, Carlos Fernando; LOGULO, Waldir; GREC, Parizoto. Psicometria na garantia de qualidade da educação médica: conceitos e aplicações. **Science in Health**, [s. l.], v. 3, n. 1, p. 33–49, 2012.
26. COLLARES, Carlos Fernando; CECILIO-FERNANDES, Dario. When I sat... computerised adaptative testing. **Medical Education**, [s. l.], n. 53, p. 115–116, 2018.
27. CONDE, Frederico Neves; LAROS, Jacob Arie. Unidimensionalidade e a propriedade de invariância das estimativas da habilidade pela TRI. **Avaliação Psicológica**, v. 6, n. 2, p. 205–215, 2007.
28. CONSELHO NACIONAL DE EDUCAÇÃO (CNE) . **RESOLUÇÃO CNE/CES Nº 4, DE 7 DE NOVEMBRO DE 2001**. Disponível em: <<http://portal.mec.gov.br/cne/arquivos/pdf/CES04.pdf>>. , 2001
29. COOMBES, Lee et al. *Beyond assessment: Feedback for individuals and institutions based on the progress test*. **Medical Teacher**, v. 32, n. 6, p. 486–490, 2010.
30. CUNHA, Egláisa Micheline Pontes. **Sistema universal e sistema de cotas para negros na Universidade de Brasília: Um estudo de desempenho**. 2006. 98. Dissertação de mestrado. Universidade de Brasília, 2006
31. CUSTERS, Eugene J F M. *Long-term retention of basic science knowledge: A review study*. **Advances in Health Sciences Education**, [s. l.], v. 15, n. 1, p. 109–128, 2010.
32. CUSTERS, Eugene J F M. *Thirty years of illness scripts: Theoretical origins and practical applications*. **Medical teacher**, v. 37, n. 5, p. 457–462, maio 2015.
33. DISTRITO FEDERAL. lei nº 3.361, de 15 de junho de 2004 – **diário oficial do distrito federal** nº 114 de 17/06/2004. 2004.
34. DOWNING, Steven M (A). *Guessing on selected-response examinations*. **Medical Education**, v. 37, n. 8, p. 670–671, 2003.
35. DOWNING, Steven M. (B) *Validity: on the meaningful interpretation of assessment data*. **Medical Education**, n. 37, p. 830–837, 2003.
36. DOWNING, Steven M. *The Effects of Violating Standard Item Writing Principles on Tests and Students: The Consequences of Using Flawed Test Items on*

- Achievement Examinations in Medical Education. **Advances in Health Sciences Education**, n. 10, p. 133–143, 2005.*
37. DOWNING, Steven M.; TEKIAN, Ara; YUDKOWSKY, Rachel. *Procedures for Establishing Defensible Absolute Passing Scores on Performance Examinations in Health Professions Education, Teaching and Learning in Medicine. **Teach Learn Med**, v. 18, n. 1, p. 50–57, 2006.*
 38. EBEL, R. L. (1979). **Essentials of Educational Measurement** (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
 39. EPSTEIN, Ronald M; HUNDERT, Edward M. Defining and assessing professional competence. **JAMA**, [s. l.], v. 287, n. 2, p. 226–235, 2002.
 40. FAHRMEIR, L. et al. (2013) **Regression: models, methods and applications**. Berlin: Springer.
 41. FLEXNER A. **Medical Education in the United States and Canada**. New York: Carnegie Foundation for The Advancement of Teaching; 1910. (Bulletin, 4)
 42. FINDYARTINI, A et al. Collaborative progress test (cPT) in three medical schools in Indonesia: the validity, reliability and its use as a curriculum evaluation tool. **Medical Teacher**, v. 37, n. February 2017, p. 366–373, 2015.
 43. FINUCANE, Paul et al. *Demographic attributes and knowledge acquisition among graduate-entry medical students. **Medical Teacher**, v. 35, n. 2, p. 134–138, 2013.*
 44. FREEMAN, A et al. *Progress testing internationally. **Medical Teacher**, v. 32, n. April, p. 451–455, 2010.*
 45. FRANCIS, Andrew M; TANNURI-PIANTO, Maria. Using Brazil ' s Racial Continuum to Examine the Short-Term Effects of Affirmative Action in Higher Education. **The Journal of Human Resources**, v. 47, n. 3, p. 754–784, 2013.
 46. GENN, J. M. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education - A unifying perspective. **Medical Teacher**, [s. l.], v. 23, n. 5, p. 445–454, 2001.
 47. GIVEN, Karen; HANNIGAN, Ailish; MCGRATH, Deirdre. *Red, yellow and green: What does it mean? How the progress test informs and supports student progress. **Medical Teacher**, v. 38, n. 10, p. 1–8, 2016.*
 48. GMC, “**Tomorrow’s Doctors. Recommendations on Undergraduate Medical Curriculum**,” General Medical Council, London, 1993.

49. GOLGHER, André Braz. *Comparisons of counterfactual affirmative action policies at the Federal University of Minas Gerais*. **Revista Brasileira de Ciências Sociais**, v. 36, n. 105, 2021.
50. GÖRLICH, Dennis; FRIEDERICHS, Hendrik. Using longitudinal progress test data to determine the effect size of learning in undergraduate medical education—a retrospective, single-center, mixed model analysis of progress testing results. **Medical Education Online**, [s. l.], v. 26, n. 1, 2021. Disponível em: <https://doi.org/10.1080/10872981.2021.1972505>.
51. HALADYNA, Thomas M.; DOWNING, Steven M. How many options is enough for a multiple-choice test item?. **Educational and Psychological Measurement**, [s. l.], n. 53, p. 999–109, 1993.
52. HAMAMOTO FILHO, Pedro Tadao; BICUDO, Angélica Maria. *Improvement of Faculty's Skills on the Creation of Items for Progress Testing Through Feedback to Item Writers: a Successful Experience*. **Revista Brasileira de Educação Médica**, v. 44, n. 1, 2020.
53. HAYAT, Ali Asghar et al. *Relationships between academic self-efficacy, learning-related emotions, and metacognitive learning strategies with academic performance in medical students: A structural equation model*. **BMC Medical Education**, v. 20, n. 1, p. 1–11, 2020.
54. HEYNEMAN, S. P.; LOXLEY, W. A. (1983). *The effect of primary-school quality on academic achievement across twenty-nine high and low income countries*. **American Journal of Sociology**, 88(6), 1162–1194.
55. HILLEN, H.; SCHERPBIER, A.; WIJNEN, W., *History of problem-based learning in medical education*. In VAN BERKEL, H., SCHERPBIER, A., HILLEN, H.; VAN DER VLEUTEN, C. (Eds.). (2010). **Lessons from problem-based learning**. OUP Oxford
56. HMELO, Cindy E.; GOTTERER, Gerald S.; BRANSFORD, John D. A theory-driven approach to assessing the cognitive effects of PBL. **Instructional Science**, [s. l.], v. 25, n. 1, p. 387–408, 1997.
57. HMELO-SILVER, Cindy E. Problem-based learning: What and how do students learn?. **Educational Psychology Review**, [s. l.], v. 16, n. 3, p. 235–266, 2004.
58. IBRAHIM BELACIANO, Mourad. **Uma forma curricular Notas para uma Teoria de Currículo para a Educação Médica**. Tese de doutorado 2015. 478 f. Universidade de Brasília, 2015. Disponível em: <https://repositorio.unb.br/bitstream/10482/22548/3/2015_MouradIbrahimBelaciano.pdf>.

59. JACOBSON, Neil S.; TRUAX, Paula. Clinical Significance: A statistical approach to defining meaningful change in psychotherapy research. ***Journal of Consulting and Clinical Psychology***, v. 59, n. 1, p. 12–19, 1991.
60. JENSEN, Jamie L. et al. *Teaching to the Test...or Testing to Teach: Exams Requiring Higher Order Thinking Skills Encourage Greater Conceptual Understanding*. ***Educational Psychology Review***, v. 26, n. 2, p. 307–329, 2014.
61. JOZEFOWICZ, Ralph F. et al. *The quality of in-house medical school examinations*. ***Academic Medicine***, v. 77, n. 2, p. 156–161, 2002
62. KARTHIKEYAN, Sowmiya; O'CONNOR, Elizabeth; HU, Wendy. *Barriers and facilitators to writing quality items for medical school assessments - A scoping review*. ***BMC Medical Education***, v. 19, n. 1, p. 14–17, 2019.
63. KRATHWOHL, David R. *A Revision of Bloom's Taxonomy: An Overview*. ***Theory Into Practice***. v. 41, n. 4, p. 212–218, 1 nov. 2002.
64. KUSURKAR, Rashmi A. et al. *Have motivation theories guided the development and reform of medical education curricula? a review of the literature*. ***Academic Medicine***, v. 87, n. 6, p. 735–743, 2012.
65. LANGER, Michelle M; SWANSON, David B. *Practical considerations in equating progress tests*. ***Medical teacher***, v. 32, p. 509–512, 2010.
66. LIMA, E.V. **Estudantes de medicina em metodologias ativas: Desafios da Aprendizagem Baseada em Problemas**. Dissertação de mestrado 2013. 89p. f. 2013.
67. LIVINGSTON, J. A. **Metacognition: an overview** U.S. Department of Education, Educational Resources Information center. [S. l.: s. n.], 2003. Disponível em: <http://gse.buffalo.edu/fas/shuell/CEP564/Metacog.htm>.
68. MARCOMINI NETO, Atílio *et al.* Flutuação do desempenho acadêmico de alunos de medicina selecionados por políticas afirmativas (PIMESP). ***Revista Brasileira de Educação Médica***, [s. l.], v. 46, n. 1, p. 1–10, 2022.
69. MCHARG, Jane et al. *Assessment of progress tests*. ***Medical Education***, v. 39, n. 2, p. 221–227, 2005.
70. MENDES JUNIOR, Alvaro Alberto Ferreira. Uma análise da progressão dos alunos cotistas sob a primeira ação afirmativa Brasileira no ensino superior: O caso da universidade do estado do rio de janeiro. ***Ensaio: Avaliação Política Pública Educacional***, [s. l.], v. 22, n. 82, p. 31–56, 2014.

71. MUIJTJENS, A.M.M. et al. *Relative or Absolute Standards in Assessing Medical Knowledge Using Progress Tests*. **Advances in health sciences education : theory and practice**, v. 3, n. 2, p. 81–87, 1998.
72. MUIJTJENS, A. M.M. et al. *The effect of a “don’t know” option on test scores: Number-right and formula scoring compared*. **Medical Education**, v. 33, n. 4, p. 267–275, 1999.
73. MUIJTJENS, Arno M.M. et al. *Origin bias of test items compromises the validity and fairness of curriculum comparisons*. **Medical Education**, v. 41, n. 12, p. 1217–1223, 2007.
74. MUIJTJENS, Arno M.M. et al. *Benchmarking by cross-institutional comparison of student achievement in a progress test*. **Medical Education**, v. 42, n. 1, p. 82–88, 2008a.
75. MUIJTJENS, Arno M.M. et al. *Differences in knowledge development exposed by multi-curricular progress test data*. **Advances in Health Sciences Education**, v. 13, n. 5, p. 593–605, 2008b.
76. MUIJTJENS, Arno M. M. et al. *Flexible electronic feedback using the virtues of progress testing*. **Medical Teacher**, v. 32, n. 6, p. 491–495, 2010.
77. NAEEM N, VAN DER VLEUTEN C, ALFARIS E.A. Faculty development on item writing substantially improves item quality. **Adv Health Sci Educ Theory Pract**. V. 17, n. 3, p. 369–76, 2012.
78. NEWBLE, David; JAEGER, Kerry. *The effect of assessments and examinations on the learning of medical students*. **Medical Education**, v. 17, p. 165–171, 1983.
79. NEWBLE, David. Revisiting ‘*The effect of assessments and examinations on the learning of medical students*’. **Medical Education**, v. 50, n. 5, p. 498–501, 2016.
80. NORCINI, John. The power of feedback. **Medical Education**, [s. l.], v. 44, n. 1, p. 16–17, 2010.
81. NORMAN, Geoff et al. *Assessment steers learning down the right road: Impact of progress testing on licensing examination performance*. **Medical Teacher**, v. 32, n. 6, p. 496–499, 2010.
82. NOUNS, Zineb M; GEORG, Waltraud. *Progress testing in German speaking countries*. **Medical teacher**, v. 32, p. 467–470, 2010.
83. OECD (2019), PISA 2018 Results (Volume I): *What Students Know and Can Do*. **PISA, OECD Publishing**, Paris.

84. PANIAGUA, Miguel A; SWYGERT, Kimberly A. *Constructing Written Test Questions For the Basic and Clinical Sciences*. **National Board of Medical Examiners**. 4ª ed. 2016.
85. PASQUALI, Luiz. **Psicometria. Teoria dos Testes na Psicologia e na Educação**. 4ª edição ed.Vozes Petrópolis: 2003.
86. PASQUALI, Luiz. Psicometria. **Revista da Escola de Enfermagem da USP**. v. 43, n. esp, p. 992–999, 2009.
87. PEIXOTO, Adriano de Lemos Alves et al. Cotas e desempenho acadêmico na UFBA: um estudo a partir dos coeficientes de rendimento. Avaliação. **Revista da Avaliação da Educação Superior** (Campinas), v. 21, n. 2, p. 569–592, 2016.
88. PINHEIRO, Osni Lázaro et al. *Progress test: an evaluative tool for academic management*. **Rev Bras Educ Méd**, v. 39, n. 1, p. 68–78, 2015.
89. PLESSAS, Anastasios. Validity of progress testing in healthcare education. **International Journal of Humanities Social Sciences and Education**, [s. l.], v. 2, n. 8, p. 23–33, 2015.
90. PPC-ESCS, 2021. **Projeto pedagógico do curso de medicina ESCS**, 2021. disponível em projeto_pedagogico_do_curso_de_medicina_normalizado_bce_2021__2_.pdf (escs.edu.br)
91. QUAIGRAIN, Kennedy; ARHIN, Ato Kwamina. Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. **Cogent Education**, [s. l.], v. 4, n. 1, 2017.
92. QUEIROZ, Zandra et al. A lei de cotas na perspectiva do desempenho acadêmico na Universidade Federal de Uberlândia. **Revista Brasileira de Estudos Pedagógicos**, v. 96, n. 243, p. 299–320, 2015.
93. RAYMOND, Mark R.; GRANDE, Joseph P. *A practical guide to test blueprinting*. **Medical Teacher**, v. 41, n. 8, p. 854–861, 2019.
94. REGEHR, Glenn; NORMAN, Geoffrey R. Issues in cognitive psychology: Implications for professional education. **Academic Medicine**, v. 71, n. 9, p. 988–1001, 1996.
95. RICHARDSON, Michelle; ABRAHAM, Charles; BOND, Rod. Psychological correlates of university students' academic performance: A systematic review and meta-analysis. **Psychological Bulletin**, v. 138, n. 2, p. 353–387, 2012.

96. RICKETTS, Chris; FREEMAN, Adrian C.; COOMBES, Lee R. *Standard setting for progress tests: Combining external and internal standards.* **Medical Education**, v. 43, n. 6, p. 589–593, 2009.
97. RICKETTS, Chris et al. *Difficult decisions for progress testing: How much and how often?* **Medical Teacher**, v. 32, n. 6, p. 513–515, 2010.
98. RICKETTS, Chris; MOYEED, Rana. *Improving progress test score estimation using Bayesian statistics.* **Medical Education**, v. 45, n. 6, p. 570–577, 2011.
99. RODRÍGUEZ-HERNÁNDEZ, Carlos Felipe; CASCALLAR, Eduardo; KYNDT, Eva. *Socio-economic status and academic performance in higher education: A systematic review.* **Educational Research Review**, v. 29, n. September 2019, p. 1-24, 2020.
100. ROFF, S.; MCALEER, S. What is educational climate?. **Medical Teacher**, [s. l.], v. 23, n. 4, p. 333–334, 2001.
101. ROSA, Maria Inês et al. *Progress testing as an indicator for improvements in a medical School.* **Revista Brasileira de Educação Médica**, v. 41, n. 1, p. 58–68, 2016.
102. RYAN, Richard M.; DECI, Edward L. *Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.* **American Psychologist**, v. 55, n. 1, p. 68–78, 2000.
103. SCHAUBER, Stefan; NOUNS, Zineb M. Using the cumulative deviation method for cross-institutional benchmarking in the Berlin progress test. **Medical Teacher**, v. 32, n. 6, p. 471–475, 2010.
104. SCHMIDT, Henk G. *Problem-based learning: Rationale and description.* **Medical Education**, v. 17, n. 1, p. 11–16, 1983.
105. SCHUMACHER, Daniel J.; ENGLANDER, Robert; CARRACCIO, Carol. *Developing the Master Learner.* **Academic Medicine**, v. 88, n. 11, p. 1635–1645, 2013.
106. SCHUWIRTH, Lambert W T; VAN DER VLEUTEN, Cees P M. *The use of progress testing.* **Perspectives on medical education**, v. 1, n. 1, p. 24–30, 2012.
107. SPAULDING, W. B. *The undergraduate medical curriculum (1969 model): McMaster university.* **Canadian Medical Association Journal**, v. 100, n. 14, p. 659–664, 1969.

108. SAKAI, Marcia Hiromi *et al.* Teste de progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina. **Revista Brasileira de Educação Médica**, [s. l.], v. 32, n. 2, p. 254–263, 2008.
109. SILVA, Paulo R. ***Affirmative action and academic performance : Brazilian medical students at Escola Superior de Ciências da Saúde.*** 2013. 27 p. Dissertação de mestrado (MHPE-Brasil) - Maastricht University, 2013.
110. Sirin, S. R. (2005). *Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research.* **Review of Educational Research**, 75(3), 417–453.
111. STAGNARO-GREEN, Alex S.; DOWNING, Steven M. *Use of flawed multiple-choice items by the New England Journal of Medicine for continuing medical education.* **Medical Teacher**, v. 28, n. 6, p. 566–568, 2006.
112. SWANSON, D. B. *et al.* Collaboration across the pond: The multi-school progress testing project. **Medical Teacher**, [s. l.], v. 32, n. 6, p. 480–485, 2010.
113. TARRANT, Marie; WARE, James; MOHAMMED, Ahmed M. *An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis.* **BMC Medical Education**, v. 9, n. 1, p. 1–8, 2009.
114. TIO, René A *et al.* The progress test of medicine: the Dutch experience. **Perspectives on Medical Education**, v. 5, n. 1, p. 51–5, 2016.
115. TOMIC, Eliane R *et al.* Progress testing: evaluation of four years of application in the school of medicine, University of São Paulo. **Clinics (Sao Paulo, Brazil)**, v. 60, n. 5, p. 389–396, 2005.
116. VALENTE, Rubia R.; BERRY, Brian J.L. *Performance of students admitted through affirmative action in Brazil.* **Latin American Research Review**, v. 52, n. 1, p. 18–34, 2017.
117. VAN DER VLEUTEN, C P M; VERWIJNEN, G M; WIJNEN, W.H.F.W. *Fifteen years of experience with progress testing in a problem-based learning curriculum.* **Medical Teacher**, v. 18, n. 2, p. 103, 1996
118. VAN DER VLEUTEN, C. P M *et al.* *Cross institutional collaboration in assessment: A case on progress testing.* **Medical Teacher**, v. 26, n. 8, p. 719–725, 2004.
119. VAN DER VLEUTEN, Cees; FREEMAN, Adrian; COLLARES, Carlos Fernando. *Progress test utopia.* **Perspectives on Medical Education**, p. 2–4, 2018.

120. VELLOSO, Jacques. Cotistas e não-cotistas: rendimento de alunos da Universidade de Brasília. **Cadernos de Pesquisa**, v. 39, n. 137, p. 621–644, 2009.
121. VERHOEVEN, B. H. et al. *Growth of medical knowledge*. **Medical Education**, v. 36, n. 8, p. 711–717, 2002.
122. VERHOEVEN, B.H. et al. The versatility of progress testing assessed in an international context: a start for benchmarking global standardization? **Medical Teacher**, v. 27, n. 6, p. 514–520, 2005.
123. VIEIRA, Renato Schwambach; ARENDS-KUENNING, Mary. *Affirmative action in Brazilian universities: Effects on the enrollment of targeted groups*. **Economics of Education Review**, v. 73, n. July, p. 101931, 2019.
124. WHITE, K. R. (1982). *The relation between socioeconomic status and academic achievement*. **Psychological Bulletin**, 91(3), 461–481.
125. WAINER, J.; MELGUIZO, T. *Inclusion policies in higher education: Evaluation of student performance based on the Enade from 2012 to 2014*. **Educação e Pesquisa**, v. 44, n. 1, p. 1–15, 2018.
126. WINNE, Philip H. *Cognition and metacognition within self-regulated learning*. **Handbook of self-regulation of learning and performance (2nd ed.)**, n. 10644, p. 36–48, 2017.
127. WRIGHT, BENJAMIN D.; BELL, SUSAN R. *Item Banks: What, Why, How*. **Journal of Educational Measurement**, v. 21, n. 4, p. 331–345, 1984.
128. WRIGLEY, William et al. *A systemic framework for the progress test: Strengths, constraints and issues: AMEE Guide No. 71*. **Medical Teacher**, v. 34, n. 9, p. 683–697, 2012.
129. ZAHRA, Daniel et al. *Accounting for test reliability in student progression: the reliable change index*. **Medical Education**, v. 50, n. 7, p. 738–745, 2016.