

**UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO  
PROGRAMA DE PÓS-GRADUAÇÃO**

**MARIANA PASSOS DE SOUZA**

**O TESTE DE PROGRESSO COMO FERRAMENTA DE AVALIAÇÃO  
LONGITUDINAL DE COORTES DE ESTUDANTES DE MEDICINA NA  
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO  
PAULO**

**RIBEIRÃO PRETO - SP**

**2023**

**O TESTE DE PROGRESSO COMO FERRAMENTA DE AVALIAÇÃO  
LONGITUDINAL DE COORTES DE ESTUDANTES DE MEDICINA NA  
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO  
PAULO**

Dissertação apresentada à Faculdade de Medicina  
de Ribeirão Preto para obtenção do título de Mestre em Ciências

Orientador: Prof. Dr. Marcelo Riberto

**Ribeirão Preto**

**2023**

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

### **FICHA CATALOGRÁFICA**

de Souza, Mariana Passos

**TÍTULO: O TESTE DE PROGRESSO COMO FERRAMENTA DE AVALIAÇÃO LONGITUDINAL DE COORTES DE ESTUDANTES DE MEDICINA NA FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO**

Ribeirão Preto, 2023.

112 pág, 30cm

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto para obtenção de título de Mestre em Ciências.

Orientador: Marcelo Riberto

1.Avaliação Longitudinal 2. Teste do Progresso 3. Feedback

Versão corrigida. A versão original encontra-se disponível tanto na Biblioteca da Unidade que aloja o Programa, quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD).

Nome: Mariana Passos de Souza

Título: O Teste de Progresso como ferramenta de avaliação longitudinal de coortes de estudantes de medicina na Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto para obtenção de título de Mestre em Ciências Médicas

Aprovado em: 11 de dezembro de 2023.

Banca Examinadora

Prof<sup>a</sup>.Dr<sup>a</sup> Angélica Maria Bicudo

Instituição: Faculdade de Ciências Médicas - UNICAMP

Julgamento: Aprovada

Prof. Dr. Valdes Roberto Bollela

Instituição: Faculdade de Medicina de Ribeirão Preto – USP

Julgamento: Aprovada

Prof. Dr. Marcelo Riberto

Instituição: Faculdade de Medicina de Ribeirão Preto - USP

Julgamento: Aprovada

## AGRADECIMENTOS

Agradeço a minha família por todo o apoio nos períodos de planejamento e escrita, na forma de risadas, lanchinhos e companhias. Aos meus pais, Luiz e Paula, obrigada por me oferecerem perspectivas para continuar a sonhar e por sempre me oferecer o amor mais gratuito do mundo, e por estarem cheios de orgulho independente do que eu faça. Ao meu irmão Gabriel, por sempre ser uma escuta disponível e amorosa, pelos presentinhos e por garantir que meu estoque de figurinhas sempre fosse apto a descrever meu estado de espírito nos dias em que não escrevia nada.

Agradeço ao meu orientador Prof. Riberto, pela eterna paciência e confiança no meu trabalho, por me escolher e me acompanhar desde a graduação nos mais variados projetos, e sempre me estimular a fazer voos corajosos e confiar no meu potencial. Obrigada por compartilhar essa jornada comigo, respeitar minhas decisões e me fazer sentir sabida, coisa que é difícil no processo da pós-graduação.

Aos meus amigos, novos e antigos, por sempre apoiarem minhas atividades de múltiplos turnos e sempre toparem conversar e sair comigo em horários aleatórios (você sabem quem são). Em especial, agradeço às minhas corretoras de texto não oficiais – Paula, minha mãe, e Têmise – que me salvaram de todos os meus erros gramaticais, à Marcela por ser a líder de torcida de todo o processo, quer as coisas estivessem funcionando ou não, e ao Max que sempre achou tempo para tirar minhas dúvidas de estatística e tabelas de Excel.

A todos os meus professores cuja didática e dedicação me inspiraram a iniciar meus aprendizados na educação e a me apaixonar pela área.

E ao Grupo PET-FMRP por ter sido origem de grandes amizades, sido fonte importante para trabalho coletivo e por ter sido a primeira casa de muitos dos meus projetos e aprendizados que seguem vivos até hoje, inclusive esse.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

## SUMÁRIO

<b>LISTA DE APÊNDICES</b> .....	8
<b>LISTA DE ANEXOS</b> .....	9
<b>LISTA DE TABELAS</b> .....	10
<b>LISTA DE FIGURAS</b> .....	11
<b>RESUMO</b> .....	13
<b>ABSTRACT</b> .....	15
<b>1. INTRODUÇÃO</b> .....	15
<b>1.1.Recomendações para elaborações curriculares dos cursos de medicina – A perspectiva na FMRP-USP</b> .....	15
<b>1.2 A importância da avaliação nas experiências educacionais em saúde</b> .....	16
<b>1.3. Contextualização Internacional do Teste de Progresso</b> .....	20
<b>1.4. Contextualização Histórica Nacional do Teste de Progresso</b> .....	21
<b>1.5 Ferramentas e Aplicações do Teste de Progresso</b> .....	23
<b>1.6. Justificativa</b> .....	23
<b>2. OBJETIVO</b> .....	26
<b>2.1.Objetivo Geral</b> .....	26
<b>2.2.Objetivos Específicos</b> .....	26
<b>3. MATERIAIS E MÉTODOS</b> .....	27
<b>3.1. Desenho do Estudo</b> .....	27
<b>3.2. Participantes do Estudo</b> .....	27
<b>3.3. Obtenção, apresentação e análise de dados</b> .....	28
<b>3.4. Aspectos Éticos</b> .....	28
<b>4. RESULTADOS</b> .....	30
<b>4.1. Uma breve descrição do TP na FMRP-USP</b> .....	30
<b>4.2. Apresentação de Dados aos Estudantes</b> .....	34
<b>4.3. Características Psicométricas do Teste de Progresso ao longo dos anos</b> .....	35
<b>4.4. Participantes Elegíveis e Adesão</b> .....	40
<b>4.5. Descrição Inicial da Progressão dos Resultados</b> .....	41
<b>4.6. Descrição e Comparação de resultados intra-Coortes</b> .....	43
<b>4.7. Comparação de resultados entre as Coortes</b> .....	61
<b>4.8. Relatos de Caso de Estudantes</b> .....	61
<b>4.9. Relação do teste com a prova de Residência</b> .....	62

<b>5. DISCUSSÃO</b> .....	70
<b>5.1. Explorando a Experiência da FMRP-USP</b> .....	70
<b>5.2. Resultados das Coortes</b> .....	75
<b>5.3 Comparação com a Prova de Residência</b> .....	77
<b>5.4 Exemplos de Uso Prático do Teste do Progresso</b> .....	77
<b>5.5. Devolutiva e Feedback dos Resultados aos Estudantes</b> .....	80
<b>5.6. Limitações do Estudo</b> .....	82
<b>5.7. Considerações Finais</b> .....	82
<b>6. CONCLUSÕES</b> .....	84
<b>BIBLIOGRAFIA</b> .....	85
<b>APÊNDICES</b> .....	88
<b>ANEXOS</b> .....	102



## **LISTA DE APÊNDICES**

Apêndice 1: Distribuição de Carga horária ao longo do curso

Apêndice 2: Exploratória média de resultados gerais por Coorte por ano

Apêndice 3: Detalhamento das Comparações por área e por coorte

Apêndice 4: Comparação entre resultados da prova inter-coortes, por ano

Apêndice 5: Detalhamento de movimentação de estudantes entre quartis

## **LISTA DE ANEXOS**

Anexo 1 – Autorização da Comissão de Graduação – FMRP/USP

Anexo 2 - Aprovação do Comitê de Ética – HCFMRP/USP

Anexo 3 – Carta de Aceite de trabalho na 2023 *AMEE Conference* como apresentação oral

## LISTA DE TABELAS

- Tabela 1 - Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova da Coorte de 2010
- Tabela 2 - Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova da Coorte de 2011
- Tabela 3 - Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova da Coorte de 2012
- Tabela 4 - Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova da Coorte de 2013
- Tabela 5 - Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova da Coorte de 2014
- Tabela 6 - Coeficiente de Confiabilidade das provas durante todo o período estudado
- Tabela 7 - Porcentagem de estudantes elegíveis e participantes em cada Coorte
- Tabela 8 - Comparação da média da progressão do desempenho geral no TP
- Tabela 9 - Comparação da média das provas de residência por coorte
- Tabela 10 - Taxas de Acerto e Distribuição por Quartis do TP e da prova de residência
- Tabela 11 - Movimentação de estudantes entre quartis
- Tabela 12 - Comparação de resultados da Área Básica na Coorte de 2010 \*
- Tabela 13 - Comparação de resultados da área de Saúde Coletiva na Coorte de 2010
- Tabela 14 - Comparação de resultados da área de Clínica Médica na Coorte de 2010
- Tabela 15 - Comparação de resultados da área de Cirurgia na Coorte de 2010
- Tabela 16 - Comparação de resultados da área de Pediatria na Coorte de 2010
- Tabela 17 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2010
- Tabela 18 - Comparação de resultados da Área Básica na Coorte de 2011
- Tabela 19 - Comparação de resultados da área de Saúde Coletiva na Coorte de 2011
- Tabela 20 - Comparação de resultados da área de Clínica Médica na Coorte de 2011
- Tabela 21 - Comparação de resultados da área de Cirurgia na Coorte de 2011
- Tabela 22 - Comparação de resultados da área de Pediatria na Coorte de 2011
- Tabela 23 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2011
- Tabela 24 - Comparação de resultados da área básica na Coorte de 2012
- Tabela 25 - Comparação de resultados da área de saúde coletiva na Coorte de 2012
- Tabela 26 - Comparação de resultados da área de Clínica Médica na Coorte de 2012
- Tabela 27 - Comparação de resultados da área de Cirurgia na Coorte de 2012
- Tabela 28 - Comparação de resultados da área de Pediatria na Coorte de 2012
- Tabela 29 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2012
- Tabela 30 - Comparação de resultados da área básica na Coorte de 2013
- Tabela 31 - Comparação de resultados da área de saúde coletiva na Coorte de 2013
- Tabela 32 - Comparação de resultados da área de Clínica Médica na Coorte de 2013
- Tabela 33 - Comparação de resultados da área de Cirurgia na Coorte de 2013
- Tabela 34 - Comparação de resultados da área de Pediatria na Coorte de 2013
- Tabela 35 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2013
- Tabela 36 - Comparação de resultados da área básica na Coorte de 2014
- Tabela 37 - Comparação de resultados da área de saúde coletiva na Coorte de 2014
- Tabela 38 - Comparação de resultados da área de Clínica Médica na Coorte de 2014
- Tabela 39 - Comparação de resultados da área de Cirurgia na Coorte de 2014
- Tabela 40 - Comparação de resultados da área de Pediatria na Coorte de 2014
- Tabela 41 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2014

(\*) – tabelas de 12 a 41 presentes em Apêndice 3

## LISTA DE FIGURAS

- Figura 1 - O Teste de Progresso e pirâmide de Miller
- Figura 2 - Descrição de evolução do TP ao longo do tempo
- Figura 3 - Exemplo de Apresentação de resultados do TP aos estudantes
- Figura 4 - Distribuição de Dificuldade das Questões - 6° Anos do Consórcio
- Figura 5 - Distribuição de Discriminação das Questões - 6° Anos do Consórcio
- Figura 6 - Adesão dos estudantes de diferentes Coortes ao longo dos anos
- Figura 7 - Representação de Desempenho de Coortes de estudantes ao longo de seu curso
- Figura 8 - Desempenho da Coorte de 2010 nas questões da Área Básica
- Figura 9 - Desempenho da Coorte de 2010 nas questões de Saúde Coletiva
- Figura 10 - Desempenho da Coorte de 2010 nas questões de Clínica Médica
- Figura 11 - Desempenho da Coorte de 2010 nas questões de Cirurgia
- Figura 12 - Desempenho da Coorte de 2010 nas questões de Pediatria
- Figura 13 - Desempenho da Coorte de 2010 nas questões de Ginecologia e Obstetrícia
- Figura 14 - Desempenho da Coorte de 2011 nas questões da Área Básica
- Figura 15 - Desempenho da Coorte de 2011 nas questões de Saúde Coletiva
- Figura 16 - Desempenho da Coorte de 2011 nas questões de Clínica Médica
- Figura 17 - Desempenho da Coorte de 2011 nas questões de Cirurgia
- Figura 18 - Desempenho da Coorte de 2011 nas questões de Pediatria
- Figura 19 - Desempenho da Coorte de 2011 nas questões de Ginecologia e Obstetrícia
- Figura 20 - Desempenho da Coorte de 2012 nas questões da Área Básica
- Figura 21 - Desempenho da Coorte de 2012 nas questões de Saúde Coletiva
- Figura 22 - Desempenho da Coorte de 2012 nas questões de Clínica Médica
- Figura 23 - Desempenho da Coorte de 2012 nas questões de Cirurgia
- Figura 24 - Desempenho da Coorte de 2012 nas questões de Pediatria
- Figura 25 - Desempenho da Coorte de 2012 nas questões de Ginecologia e Obstetrícia
- Figura 26 - Desempenho da Coorte de 2013 nas questões da Área Básica
- Figura 27 - Desempenho da Coorte de 2013 nas questões de Saúde Coletiva
- Figura 28 - Desempenho da Coorte de 2013 nas questões de Clínica Médica
- Figura 29 - Desempenho da Coorte de 2013 nas questões de Cirurgia
- Figura 30 - Desempenho da Coorte de 2013 nas questões de Pediatria
- Figura 31 - Desempenho da Coorte de 2013 nas questões de Ginecologia e Obstetrícia
- Figura 32 - Desempenho da Coorte de 2014 nas questões da Área Básica
- Figura 33 - Desempenho da Coorte de 2014 nas questões de Saúde Coletiva
- Figura 34 - Desempenho da Coorte de 2014 nas questões de Clínica Médica
- Figura 35 - Desempenho da Coorte de 2014 nas questões de Cirurgia
- Figura 36 - Desempenho da Coorte de 2014 nas questões de Pediatria
- Figura 37 - Desempenho da Coorte de 2014 nas questões de Ginecologia e Obstetrícia
- Figura 38 - Exemplo de Desempenho de estudantes
- Figura 39 - Representação de Desempenho de Coortes de estudantes ao longo de seu curso e com resultados de Residência Médica
- Figura 40 - Correlação do TP do 6° ano com prova de residência na coorte de 2010
- Figura 41 - Correlação do TP do 6° ano com prova de residência na coorte de 2011
- Figura 42 - Correlação do TP do 6° ano com prova de residência na coorte de 2012
- Figura 43 - Correlação do TP do 6° ano com prova de residência na coorte de 2013
- Figura 44 - Correlação do TP do 6° ano com prova de residência na coorte de 2014
- Figura 45 - Correlação do TP do 6° ano com prova de residência de todas as coortes

## **LISTA DE ABREVIATURAS**

FMRP: Faculdade de Medicina de Ribeirão Preto

TP: Teste de Progresso

USP: Universidade de São Paulo

CoC: Comissão Coordenadora do Curso

CG: Comissão de Graduação

ABEM: Associação Brasileira de Educação Médica

CAEG: Centro de Avaliação em Ensino de Graduação

GO: Ginecologia e Obstetrícia

CX: Cirurgia

SC: Saúde Coletiva

P: Pediatria

CM: Clínica Médica

BASC: Área Básica

HCRP: Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto

## RESUMO

DE SOUZA, M.P. **O teste de progresso como ferramenta de avaliação longitudinal de coortes de estudantes de medicina na Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.** 2023. Dissertação (Mestrado em Ciências - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2023)

O Teste de Progresso (TP) é uma avaliação longitudinal periódica da aquisição de conhecimentos dos alunos de graduação, que pode auxiliar na gestão curricular e na avaliação pessoal. Em 2005, a FMRP-USP passou a aplicá-lo em estudantes de medicina, tentando detectar potencialidades e lacunas no ensino, e permitir que os alunos avaliassem seu aprendizado ao longo dos anos. O estudo objetivou explorar a história do TP na FMRP-USP e os resultados dos testes de progresso de cinco coortes de estudantes, avaliando seu desempenho no teste e comparando com resultados da prova residência médica desta instituição.

Consistiu em um estudo descritivo retrospectivo avaliando o resultado do TP de cinco coortes de alunos (2010-2014) que concluíram o curso entre 2015 e 2019. Foram incluídos os alunos que não perderam mais de 4 provas de progresso e que concluíram o curso na sua turma de origem. O projeto foi aprovado pelo Comitê de Ética em Pesquisa.

Para avaliar as características dos testes, foram obtidos dados de discriminação, dificuldade e confiabilidade das questões de cada grande área. Em relação aos resultados das provas, analisamos as taxas de sucesso por área e comparamos com a progressão da Coorte no curso e nas áreas específicas, utilizando um modelo de regressão linear com efeitos mistos para análise das variáveis. Também, foi realizada a correlação entre os resultados dos alunos do último ano do curso e os respectivos resultados do teste de residência, por meio do coeficiente de Pearson.

As taxas de abstenção foram de até 20% nos primeiros anos, com menos de 10% entre os alunos do último ano. Houve ganho de conhecimento significativo ( $p < 0,001$ ) entre os alunos do 1º e 6º ano, com aumento de 30% a 40% nos acertos na comparação entre a média de ingressantes e concluintes, com estes tendo desempenho próximo de 70% da prova. O crescimento ininterrupto no ganho de conhecimento teve início variado entre as Coortes. O número de horas do currículo não teve um impacto significativo nos resultados das aulas: as disciplinas com mais horas não tiveram necessariamente os melhores resultados, mas a introdução das disciplinas esteve relacionada com o aumento do aproveitamento nessa área. A análise macro permite um melhor aproveitamento dos resultados, foi possível observar em quais momentos do curso ocorreram resultados inesperados. Separando os alunos em quartis com base nos resultados do

6º ano e nas notas dos testes de residência, quase metade dos estudantes permanece no mesmo quartil, especialmente os estudantes dos extremos de desempenho. Foi possível observar uma correlação moderada entre as provas do 6º ano e da prova de residência. O TP mostrou-se uma ferramenta muito útil para avaliar a progressão em termos cognitivos, os grupos observados tiveram comportamento semelhante quanto à evolução da nota média ao longo dos anos. O uso sistemático do TP pode auxiliar estudantes no autogerenciamento do desempenho do curso e no planejamento das metas e pode também permitir que gestores de currículo auditem a qualidade das atividades educacionais de suas escolas.

Palavras-chave: 1. Avaliação Longitudinal 2. Teste do Progresso 3. Feedback

## ABSTRACT

DE SOUZA, M.P. **Progress Testing as a longitudinal assessment tool for cohorts of medical students at the Faculty of Medicine of Ribeirão Preto of University of São Paulo.** 2023. Dissertation (Master of Science - Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2023)

Progress Testing (PT) is a periodic longitudinal assessment of undergraduate students' knowledge acquisition, which can assist in curriculum management and personal assessment. In 2005, FMRP-USP started to apply it to medical students, trying to detect potentialities and gaps in teaching, and allow students to evaluate their learning over the years. The study aimed to explore the history of PT at FMRP-USP and the results of the progress tests of five cohorts of students, evaluating their performance on the test and comparing them with the results of the medical residency test at this institution. It consisted of a retrospective descriptive study evaluating the PT results of five cohorts of students (2010-2014) who completed the course between 2015 and 2019. Students who did not miss more than 4 progress tests and who completed the course in the class were included source. The project was approved by the Research Ethics Committee. To evaluate the characteristics of the tests, data on discrimination, difficulty, and reliability of the questions in each major area were obtained. Regarding test results, we analyzed success rates by area and compared them with the class's progression in the course and in specific areas, using a linear regression model with mixed effects to analyze the variables. Also, a correlation was made between the results of students in the last year of the course and the respective results of the residency test, using the Pearson coefficient.

Abstention rates were up to 20% in the first years, with less than 10% among final year students. There was a significant gain in knowledge ( $p < 0.001$ ) among 1st and 6th year students, with an increase of 30% to 40% in correct answers when comparing the average of freshmen and seniors, with the latter performing close to 70% of the test. The uninterrupted growth in knowledge gain had a varied start between classes. The number of hours in the curriculum did not have a significant impact on class results: subjects with more hours did not necessarily have the best results, but the introduction of subjects was related to increased achievement in that area. Macro analysis allows better use of the results, it was possible to observe at which moments during the course unexpected results occurred. Separating students into quartiles based on 6th- year results and residency test scores, nearly half of students stay in the same quartile, especially students at the extremes of performance. It was possible to observe a



moderate correlation between the 6th-year tests and the residency test. The PT proved to be a very useful tool for evaluating progression in cognitive terms, the groups observed had similar behavior in terms of the evolution of the average score over the years. The systematic use of TP can assist students in self-managing course performance and planning goals and can also allow curriculum managers to audit the quality of their school's educational activities.

Key words: 1. Longitudinal Assessment 2. Progress Test 3. Feedback

## 1. INTRODUÇÃO

### 1.1. Recomendações para elaborações curriculares dos cursos de medicina – A perspectiva na FMRP-USP

As Diretrizes Curriculares Nacionais (DCNs) de 2014 dos cursos de Medicina (1) recomendam explicitamente as ações necessárias para a construção de uma formação adequada do futuro médico. Várias estratégias são oferecidas para qualificar o estudante, com o compromisso da construção de projetos de ensino contemporâneos e abrangentes.

No artigo 13º das DCNs, destaca-se que o estabelecimento de currículos deve ser sempre “acompanhado e permanentemente avaliado”, desde suas concepções até seus aperfeiçoamentos. O curso de medicina da FMRP-USP tem um currículo de medicina com abordagem tradicional, isto é, dividido majoritariamente em blocos de ensino: o ciclo básico, o ciclo clínico e o internato. Existem disciplinas obrigatórias em cada ciclo e há progressão de aulas majoritariamente expositivas nos primeiros anos para atividades de estágios clínicos nos últimos dois anos, e entre eles o ciclo clínico que busca fazer a ponte entre os conhecimentos da área básica e as habilidades da prática clínica. Durante sua criação e aperfeiçoamento, destaca-se uma reforma curricular importante no início dos anos 90, com foco em atingir uma maior integração de conteúdo, ampliando o internato de 1 para 2 anos e com mais atenção voltada à atenção primária e a formação humanística do estudante (2). Essa mudança levou a resultados que indicaram aumento de conhecimento em áreas mais gerais e melhora no desempenho de habilidades clínicas tidas como fundamentais (3). Porém alguns aspectos da reforma se perderam com o tempo e buscou-se uma renovação de suas intenções no início dos anos 2000, com a última reforma curricular na FMRP-USP implantada em 2008. Assim a turma de alunos ingressantes em 2010 foi a primeira a ter o curso inteiro com o currículo reformado, com os novos eixos de Formação Humanística, Atenção à Saúde da Comunidade, de Urgência e Emergência e de Formação Complementar, já integrados no currículo (2).

Quanto à avaliação do estudante, as DCNs destacam a importância de que as avaliações se baseassem nas competências dos currículos desenvolvidos, em consonância com o sistema de avaliação estabelecido. Manteve-se a importância de metodologias ativas e critérios claros de avaliação ao longo da graduação, que permitam “desenvolver instrumentos que verifiquem a estrutura, os processos e os resultados”. Por fim, destacou-se também a importância de processos de autoavaliação e o estabelecimento de exames formativos longitudinais,

complementares às avaliações tradicionais; inclusive com uma proposta de criar uma prova bianual nacional com essas características (1).

## **1.2 A importância da avaliação nas experiências educacionais em saúde**

A avaliação de estudantes é processo chave para uma efetiva experiência educacional e envolve uma sequência de atividades planejadas e sistemáticas a fim de fazer um juízo sobre a aprendizagem e o processo educacional em si. A avaliação se inicia na coleta de dados, isto é, medir ou testar os conhecimentos de uma maneira estruturada e planejada, seguida pela combinação desses dados, para a classificação e estratificação dos resultados, que pode ou não ter a intenção de ranquear ou aprovar aqueles que a realizaram. Por fim, para um processo avaliativo adequado, é necessária a realização de uma devolutiva baseada nas expectativas iniciais, isto é, nos objetivos de aprendizado, nos resultados esperados e no processo de avaliação escolhido (4).

Apesar do destaque para a estruturação do currículo como chave do aprendizado efetivo, uma avaliação condizente com os objetivos de aprendizagem de uma experiência educacional se mostra crucial nesse processo. Isso é importante especialmente nos ambientes educacionais contemporâneos em que o conteúdo de uma avaliação muitas vezes norteia e motiva os estudos dos alunos, a despeito de como é oferecida uma experiência educacional. Pensando na educação médica e no ensino superior de maneira geral, com uma grande carga de conteúdo, os estudantes tendem a priorizar seu tempo e estudar aquilo que é avaliado e não apenas aquilo que desperta seus interesses ou que considera mais importante. Ainda, se há uma avaliação sobre algum tema específico, há a tendência de se entender que aquele tópico é mais importante e focar seu tempo para entender tal competência. Se o aluno prioriza estudar os pontos que são avaliados, e é através do estudo que ele sedimenta e aprende novas informações e competências, faz sentido pensar que os pontos avaliados são os que tem mais chance de ser aprendidos por eles (5).

Pensando dessa maneira, vê-se a avaliação como um desfecho essencial do momento pedagógico e isso conversa com a percepção de Miller de que “a avaliação impulsiona o aprendizado”, pois ela objetiva garantir e sedimentar o aprendizado de um tema ou habilidade a partir dos resultados que o aprendizado gerou, que podem ser quantificados de diversas maneiras, criando um fluxo virtuoso que impulsiona boas experiências educacionais. Ainda, uma avaliação bem implementada seria capaz não só de aferir o conhecimento obtido, mas também aferir a efetividade da experiência educacional em si, isto é, ser um medidor do processo educacional e avaliativo, para além dos resultados por si só. Por isso, o processo de

avaliação deve ser estratégico, validando os objetivos traçados pelo currículo, movimentando aprendizado e oferecendo bases para feedbacks, reavaliação e correção de trajetórias educacionais malsucedidas.

Miller (6) ainda destacava a importância de buscar formas de avaliação que sejam condizentes com os objetivos de aprendizagem estabelecidos, a etapa de aprendizado e as competências a serem estabelecidas ao fim do processo. Assim, diferentes formatos de avaliação devem ser buscados para aferir bases cognitivas distintas: ‘saber’, ‘saber como fazer’, ‘mostrar como se faz’ e ‘fazer’ (6).

Cada nível da pirâmide de Miller requer instrumentos distintos de avaliação com crescente complexidade. Os dois níveis iniciais envolvem o conhecimento teórico, sendo o primeiro relacionado a conhecimento factuais que podem ser testados com perguntas diretas (testes ou dissertativas) e o segundo envolve não só saber o dado, mas o saber contextualizado, isto é como combiná-lo e interpretá-lo, podendo ser avaliado, portanto, com questões em que o conhecimento está aplicado a problemas ou casos clínicos. (5,6)

O terceiro nível implica em demonstração, isto é, o estudante deve adquirir habilidades técnicas para saber performar uma ação, envolve a realização de uma ação em um ambiente controlado, por isso mais bem avaliada com modelos de simulação. O último nível diz respeito à aplicação dos demais conhecimentos culminando em um comportamento na prática, o estudante então faz aquilo que aprendeu e então deve ser avaliado em um ambiente real (5,6).

Mais recentemente, um grupo de professores canadenses introduziu ainda um quinto nível de competência no topo da pirâmide: o “ser”. Relacionado à identidade profissional, isto é, à incorporação de valores e atitudes dos estudantes na sua prática e comportamento profissional, esse quinto nível demandaria tipos de avaliação que abarcassem, além das competências do nível anterior, questões de profissionalismo, valores e atitudes (7).

Nesse sentido, nas últimas décadas também se tem buscado desenvolver avaliações que possam ter um papel prognóstico das competências clínicas e ao mesmo tempo fornecer informações para remediação e formação durante o curso (5,8), a fim de garantir que o desfecho da formação seja um profissional competente para a prática clínica. Ainda, estudos apontam que mesmo avaliações relacionadas a desfechos de aprendizado cognitivo (bases da pirâmide de Miller) podem ter moderado potencial preditivo para os desfechos de aprendizagem de competências e habilidades (topo da pirâmide) e até mesmo desfechos após o fim da graduação de medicina, reforçando o papel prognóstico das avaliações e suas potencialidades como ferramenta de gestão do percurso acadêmico (9).

Outro ponto essencial do processo de avaliação é decidir seu objetivo, podendo ser classificado em somativo, diagnóstico ou formativo (10-12). Essa classificação diz respeito da intenção com a qual se aplica uma prova e como seus resultados serão utilizados no contexto de uma experiência educacional específica. A avaliação diagnóstica busca perceber a condição prévia de conhecimento e auxilia para perceber onde é necessário ajuste antes de se iniciar um novo processo de aprendizagem, pois busca um ponto de partida para ele (12).

Nas provas somativas, que são o tipo de avaliações preponderantes nos contextos educacionais de ensino superior, busca-se comprovar até que ponto um indivíduo cumpriu com os critérios mínimos de aprovação para um determinado objetivo e isso usualmente determina sua continuidade no seu processo educacional. Por envolver processos de aprovação/reprovação, tem a intenção de serem utilizadas para competências e conhecimento significativos para os objetivos da aprendizagem (12,13). Nesse tipo de avaliação é importante não perder de vista tais objetivos e buscar fazer uma devolutiva clara dos resultados, caso contrário corre o risco de ser apenas uma ferramenta classificatória sem muitos efeitos, focando no conteúdo ensinado apenas e não no desfecho ou expectativas iniciais (12,13).

As provas formativas não têm objetivo de “aprovação/reprovação”, mas sim fornecer informações sobre os processos pedagógicos e como orientá-los. Isso implica que seus resultados sejam analisados para descreverem qualitativamente o aprendizado e o seu contexto, se são “adequados/inadequados” para os objetivos propostos. Os resultados de uma avaliação formativa podem ser aplicados tanto para o estudante individualmente, se ele está cumprindo com as expectativas determinadas de um aprendizado, mas também se estão servindo como um diagnóstico para o processo educacional em si, se esse está adequado para garantir o aprendizado desejado, isto é, servir de guia para possíveis mudanças no percurso da experiência educacional (11,12). Vale destacar que os resultados das avaliações somativas também podem ser utilizados para avaliar o processo educacional em si, mas que não têm isso como parte dos objetivos de sua aplicação, como as provas formativas.

Nas experiências educacionais há espaço e demanda para todos os tipos de avaliações anteriormente descritas, pois, se bem planejadas, oferecem informações distintas sobre as experiências pedagógicas e diferentes perspectivas sobre o processo educacional (14). Todavia, é importante destacar que, apesar de ainda menos frequentes, as práticas formativas tendem a elevar o padrão de aprendizagem por permitir atuação em múltiplos espaços do processo educacional: ao estudante, observar lacunas em seu aprendizado sem a pressão da “aprovação/reprovação”; aos docentes, revisar e modificar suas estratégias pedagógicas se a

avaliação apontar alguma lacuna, e aos gestores fornecer espaço de correção do trajeto curricular escolhido (4,14).

Historicamente, ao longo do curso de medicina a avaliação focava no conhecimento factual (o ‘saber’) pois é crucial para a aquisição de competências, fornece as bases dos demais aprendizados. Atualmente, valoriza-se a importância de avaliar além desse aspecto, ou seja, saber aplicar o conhecimento de forma contextualizada (“saber como fazer”), utilizar o conhecimento para demonstrar habilidades (“mostrar como se faz”), e agir baseado nele (“fazer”) (6).

Assim, pensando no importante papel da avaliação nos cursos de saúde, estes critérios de qualidade foram levantados para sua melhor elaboração: validade/coerência, reprodutibilidade/consistência, equivalência, viabilidade, efeito educacional, efeito catalítico e aceitabilidade (4,15,16).

A validade, implica que o uso de algum tipo de avaliação para um tipo específico de objetivo de aprendizagem esteja ancorado em evidências coerentes. A consistência implica que o modelo escolhido pode ser replicado com resultados semelhantes em contextos educacionais distintos. A equivalência diz respeito à capacidade de uma avaliação levar a desfechos similares em diferentes ciclos de teste ou até mesmo outras instituições. A viabilidade, busca considerar as questões práticas que envolvem a sua aplicação, se ele é ou não realista no seu contexto (4).

O efeito educacional é justamente o entendimento destacado neste texto de que a avaliação deve ter um papel indispensável aos alunos no processo de educação, motivando-os a se preparar adequadamente para a avaliação. Já o efeito catalítico leva em consideração a capacidade de um teste levar a resultados que estimulem movimento naqueles que aplicaram e naqueles que realizaram uma avaliação. E, por fim, a aceitabilidade implica na aceitação de que os resultados obtidos através da avaliação têm credibilidade para todos os envolvidos (4).

Mais recentemente foi destacada a importância de se incluir a percepção de pacientes no processo de avaliação, além da importante correlação entre a avaliação, feedback e os sistemas nos quais ela se insere. Há também grande destaque para seu fator impulsionador na organização curricular (4).

Baseado nesses entendimentos de avaliação, e em busca de testes que pudessem atender as demandas da formação adequada de profissionais de saúde, além de superar limitações de avaliações tradicionais, foi elaborado o modelo de avaliação de Teste do Progresso (TP). Esse modelo buscou oferecer mais validade e confiabilidade aos resultados das avaliações, não estando direta ou especificamente conectado a nenhuma parte do currículo ou corpo docente,

mas buscando avaliar os objetivos de aprendizagem e competências definidas inicialmente para o curso.

O teste se destaca por ser uma aferição, formativa na sua concepção, que abrange as bases cognitivas de conhecimentos factuais e aplicação desse conhecimento (“saber” e “saber como fazer” respectivamente). Esse destaca por ser um teste abrangente, ao oferecer uma amostra de questões de todo o conteúdo; longitudinal e seriado, ao avaliar o ganho na aquisição de conhecimento dos estudantes ao longo dos anos aplicada a todas as turmas de um curso de medicina simultaneamente, independente do ano que se encontram; e preditivo, ao estabelecer prognóstico do resultado de estudantes em habilidades e competências ao longo do curso (17,18). A figura 1 apresenta um resumo da relação do TP com a pirâmide de Miller e suas potencialidades.

Figura 1 - O Teste de Progresso e pirâmide de Miller



Fonte: Feito pelo autor, adaptado das referências 6,7 e 9

### 1.3. Contextualização Internacional do Teste do Progresso

O Teste de Progresso (TP) foi primeiramente desenvolvido e aplicado na Faculdade de Medicina da Universidade de Missouri (EUA) e na Universidade de Limburg (Maastricht, Holanda), na década de 1970 (19). Em Maastricht, foi criado especialmente para atender a uma demanda de um teste que avaliasse o conhecimento em escolas com currículos com metodologia de ensino aprendido baseados em problemas (ABP) e os comparasse com o aprendizado de escolas com modelo tradicional de ensino. Em Missouri havia a necessidade de avaliar a qualidade de aprendizado após implementação de novo programa médico (20). O uso do TP foi crescente também em escolas médicas de currículo tradicional, isto é, cujo currículo não envolve, na sua maioria, estratégias curriculares e de ensino baseadas em problemas. Isso

porque o teste é capaz de avaliar bem currículos com diferentes metodologias, pois não avalia a metodologia em si, mas sim as bases de conhecimento teórico por trás dela (17,21).

Após cerca de meio século, o TP tem sido utilizado em diversas escolas médicas ao redor do mundo com excelentes resultados sobre sua efetividade para descrição do desempenho dos estudantes e avaliação do crescimento cognitivo longitudinalmente (17,18), com bons resultados inclusive com aplicação online (22).

Em países com mais experiência de aplicação do modelo, destacam-se as colaborações internacionais, como o projeto entre o *National Board of Medical Examiners* e quatro universidades inglesas, na qual as questões feitas para os testes americanos de acreditação foram aplicadas aos estudantes ingleses, a fim de avaliar a viabilidade e utilidade desse tipo de projeto.

Nele, destaca-se que é interessante usar um mesmo plano para elaboração de testes (“*blueprint*”), pois permite uso de questões de múltiplas fontes, aumentando o número de questões disponíveis e a comparabilidade entre resultados. Também destacam a importância de que quando aplicado em diferentes contextos de ensino, um teste tem mais potencial de ser efetivo desde que haja revisão do material por especialistas e que se respeitem diferenças culturais, de atendimento e terminologias (23).

Além de muito utilizado para graduação, já existem experiências de aplicação do TP para avaliação de programas de pós-graduação, com resultados semelhantes: ele é capaz de avaliar ganho de conhecimento e observar remediações necessárias (24,25).

Em países com implementações mais recentes, destaca-se a experiência de Moçambique, onde a experiência se mostrou frutífera para viabilizar o teste num país com menos recursos (26). Houve também a implementação do teste em uma universidade da Austrália, que identificou melhora dos resultados ao longo dos anos e com boa percepção dos estudantes em relação ao teste, referindo maior retenção do conhecimento após início da sua aplicação (27). Em 2016 houve o primeiro projeto piloto de aplicação do TP no leste asiático, que apesar de não apresentar dados longitudinais, mostrou que os estudantes favoreceram a avaliação tradicional ao fim das unidades de ensino, pois não viam objetivos tangíveis de avaliação, destacando a importância de explicar aos estudantes como o TP pode ser efetivo na avaliação do ganho de conhecimento e na autoavaliação longitudinal (28).

#### **1.4. Contextualização Histórica Nacional do Teste de Progresso**

No Brasil há registro de testes sistemáticos ou intenção de sua realização desde o fim dos anos 1990 (29) e a partir de 2001, um modelo de avaliação seriada foi implementado nas



faculdades de medicina da Universidade de São Paulo (USP), dentre as quais a Faculdade de Medicina de Ribeirão Preto (FMRP). Em 2005, foi criado o primeiro consórcio de escolas médicas para a aplicação do teste de progresso: o Núcleo Interinstitucional de Estudos e Práticas de Avaliação em Educação Médica (NIEPAEM), composto por: Universidade Estadual de Campinas, Faculdade de Medicina de São José do Rio Preto, Faculdade de Medicina de Marília, Universidade Regional de Blumenau, Universidade Federal de São Carlos, Universidade de São Paulo-Faculdade de Medicina de Ribeirão Preto (FMRP-USP), Universidade Federal de São Paulo, Universidade Estadual de São Paulo – Botucatu e Universidade Estadual de Londrina. Nesse mesmo ano, a FMRP-USP passou a integrar o grupo de escolas integrantes do consórcio e, no ano seguinte, a aplicar o TP sistematicamente para os alunos de Medicina (29).

Em 2015, a Associação Brasileira de Educação Médica (ABEM) organizou uma prova nacional do TP que envolveu 58 escolas médicas e 23.065 alunos e, apesar do modelo não se manter, foi importante para fomentar colaborações e aplicação mais disseminada do TP, porém questões logísticas impediram sua viabilidade a longo prazo (30).

Além da experiência do consórcio, a Faculdade de Medicina da USP também registrou a aplicação do TP entre 2001 e 2004 bianualmente, indicando que foi uma ferramenta interessante para avaliação de ganho de conhecimento ao longo do curso (31).

A Universidade Estadual de Londrina também registrou sua experiência do TP, cuja primeira implementação ocorreu em 1998, anteriormente à sua participação no consórcio, e esse estudo reforçou a capacidade de utilizar o TP para determinar fragilidades curriculares e avaliar o ganho cognitivo dos estudantes (32).

Apesar de bastante utilizado nos espaços de graduação, o TP ainda é subutilizado nos espaços da residência médica, porém seu uso tem crescido pelo potencial de avaliação seriada do ganho de conhecimento de residentes e pelo uso potencial dos resultados tanto com intenção formativa (remediação, avaliação de ganho de conhecimento), quanto somativo (aprovação, pontuação na prova de título). Destaca-se a experiência da Federação Brasileira das Associações de Ginecologia e Obstetrícia (FEBRASGO) (33), que implementou o TP em todos os anos da residência anualmente a partir de 2018, após a estruturação de uma matriz de competências básicas para o curso e também a experiência da Sociedade Brasileira de Ortopedia e Traumatologia (SBOT) que aplica o TARO (Teste de Aptidão de Residentes de Ortopedia), com questões múltipla escolha que abordam as competências básicas dos residentes desta especialidade.

### **1.5 Ferramentas e Aplicações do Teste de Progresso**

O TP sedimentou-se em diversos cenários nacionais e internacionais como uma avaliação capaz de apontar o crescimento do estudante e comparar resultados individuais e institucionais (17,18). A proposta de aplicação do TP, na maioria das escolas de medicina internacionalmente, tem como propósito ser uma avaliação de caráter formativo, que estimule o aprendizado a longo prazo e reafirme o conhecimento já adquirido (34). Ele tem sido uma ferramenta consistente para demonstrar a fixação e o acréscimo de conhecimentos ao longo do curso e interferir positivamente no processo de ensino e aprendizagem (17,35).

Ainda, tem a vantagem de garantir que o resultado do teste não seja relacionado a um estudo intensivo ou revisão de temas para sua realização (36), pois não está relacionado temporalmente ao fim de unidades de ensino e não tem o objetivo principal de aprovar ou reprovar, estimulando um estudo e aprendizado a longo prazo (17,19,36). Talvez seja por esses motivos que o TP foi correlacionado com menores níveis de estresse pelos estudantes que outras avaliações (37). Os resultados devem ser disponibilizados aos estudantes, acompanhados de uma devolutiva, geralmente o gabarito e de uma resposta comentada, destacando também temas que os estudantes devem dedicar atenção em seus estudos atuais e futuros.

Isso porque, com a devolutiva, o teste ainda possibilita que o estudante exercite sua autoeficácia, isto é a capacidade de um indivíduo de se organizar e executar ações para atingir suas metas estabelecidas e avaliar se seus resultados estão condizentes com suas metas individuais e com as possíveis metas esperadas pela instituição, num dado período do curso (17). Pode subsidiar e estimular o estudante em direção a um aprendizado autodirigido, especialmente se o TP for aplicado com intenção formativa (36). Inclusive com o próprio estudante se torna capaz de identificar sua necessidade de atenção em um tema específico (34). A presença de uma devolutiva contemporânea e de fácil acesso às análises comparativas pode facilitar o engajamento dos estudantes com seus resultados (38).

O TP é boa ferramenta comparativa entre estudantes nacional e internacionalmente, especialmente de grupos com dados de longo prazo (23,39,40) sendo possível oferecer aos estudantes diferentes tipos de resultados para comparar seu desempenho e ser utilizado como fator de predição e auxiliar o estudante na elaboração de suas metas educacionais. Como exemplo, tem-se o consolidado consórcio de escolas holandesas, em que os resultados dos 4 testes anuais são apresentados aos estudantes, e comparados com sua turma, tem seu desempenho distribuído em gráficos de linhas, com áreas denominadas “adequadas” ou “inadequadas” para o período específico do aprendizado. Além disso, a maior frequência de

testes aumenta o potencial preditivo do TP, permitindo mostrar ao estudante, no seu próprio gráfico de curva, onde esse tende a estar nos próximos testes, caso nenhuma mudança no percurso de aprendizagem seja feita, auxiliando na autogestão de metas e planos durante a graduação (41).

Ainda quanto ao potencial preditivo, nota-se que o TP também tem boa validade preditiva em relação a desfechos de competências e bom desempenho profissional (42). Assim, pode ser ainda mais interessante como ferramenta de autogestão de aprendizado para os estudantes, quando se pensa em metas que vão além de resultados de testes, mas sim ao encontro da expertise profissional. No Brasil, estudos recentes apontam correlação interessante entre o TP e os resultados da prova teórica de residência médica (43,44).

Além disso, o TP pode ser muito útil aos gestores do currículo, pois os resultados do podem ser mais confiáveis e com menos viés na elaboração, por virem de uma avaliação externa aos blocos e disciplinas do curso. Dessa forma, permite avaliar as fortalezas e fragilidades do ensino, além de indicar possíveis ajustes para auxiliar na gestão de estratégias específicas que visem melhorar o ensino e o preenchimento de lacunas que possam existir no currículo da faculdade avaliada (16,45,46) garantindo assim um melhor alinhamento a faculdade com a matriz de competências do curso.

Em suma, os resultados gerados pelos TP podem ser de grande utilidade por diversos pontos: a avaliação da curva de aprendizado dos alunos durante o seu percurso na graduação, a comparação das pontuações com colegas de anos letivos e faculdades diferentes, a análise do desempenho para cada tema do currículo, ajuste de deficiências e correção de curso do aprendizado tanto para os estudantes quanto para gestores.

Dentre os desafios de estabelecimento do TP, está o custo de implementação, a demanda de elaboração e revisão de itens, criação de uma matriz de competências e temas para se basear as questões, além de demandar análise de dados. A criação de provas colaborativas, nos programas de consórcio, busca diminuir essas questões para facilitar a implementação (17,18).

Nesse contexto, o TP atualmente é opção complementar importante às avaliações somativas de fim de bloco pedagógico, oferecendo ferramentas para sedimentar aprendizado, avaliar longitudinalmente o desempenho individual e comparativo do aluno, além de fornecer informações para impulsionamento de uma melhor estrutura curricular.

### **1.6. Justificativa**

Até o momento não existe qualquer sistematização formal da experiência e uso do TP na FMRP-USP e muito poucas no Brasil. Este projeto justifica-se porque propõe essa documentação, visando, em um futuro breve, difundir na FMRP-USP o potencial de uso dos resultados do TP para estudantes, professores e gestores educacionais. Para isso, esse trabalho pretendeu descrever a criação, organização e resultados do TP enquanto ferramenta de avaliação formativa dos alunos e instituições participantes, no contexto da FMRP-USP.

## **2. OBJETIVO**

### **2.1. Objetivo Geral**

Descrever e avaliar os resultados do TP que é aplicado na FMRP-USP, considerando o processo de elaboração da prova, o desempenho de coortes de alunos participantes e finalmente a utilização que tem sido feita dos resultados da prova pela comunidade acadêmica da FMRP-USP (professores, alunos e gestores).

### **2.2. Objetivos Específicos**

- 2.1.1. Descrever o modelo de prova, como é organizado atualmente e como evoluiu no decorrer do tempo;
- 2.1.2. Descrever o contexto de criação e organização do consórcio de elaboração do teste de TP no qual a FMRP se insere;
- 2.1.3. Descrever a aplicação do TP na FMRP-USP;
- 2.1.4. Avaliação dos resultados da aplicação do TP na FMRP-USP:
  - Avaliar a porcentagem de participação anual no teste e a adesão ao TP no decorrer do tempo;
  - Avaliar o desempenho e progressão global de cada Coorte ao longo do curso, sob duas ópticas: período em que os estudantes se encontram na graduação (série) e as grandes áreas do conhecimento avaliadas pelo teste (Área Básica, Clínica Médica, Ginecologia e Obstetrícia, Pediatria, Cirurgia Geral, Saúde Coletiva e Ética);
  - Comparar a progressão das diferentes Coortes no mesmo período do curso;
- 2.1.5. Comparar resultados do TP dos estudantes no 6º ano de curso com seus desempenhos na prova de residência médica da FMRP e avaliar seu potencial preditivo

### **3. MATERIAIS E MÉTODOS**

#### **3.1. Desenho do Estudo**

##### **3.1.1. Revisão histórica e contextual do TP**

A fim de tecer um panorama geral sobre o tema e sobre o contexto no qual o TP se insere nesta faculdade, foram coletadas informações sobre a sua elaboração e aplicação, assim como da criação do consórcio de faculdades de medicina e seu funcionamento. Esses dados foram obtidos através de uma revisão histórica e análise de documentos. Durante o período de coleta e análise de dados quantitativos, foram feitas também revisão de documentos recebidos pela Comissão de Graduação (CG) e Comissão Coordenadora do Curso (CoC) referentes a análises estatísticas dos resultados, documentos de regimentos internos, reportagens do jornal da USP, documentos disponíveis da Associação Brasileira de Educação Médica (ABEM) e artigos científicos que descreviam a história do estabelecimento do TP no Brasil. Ainda, para auxiliar na contextualização, roteiros e ementas de disciplinas foram avaliados para contabilizar introdução da carga horária das grandes áreas ao longo da graduação. Também foram solicitadas as avaliações dos estudantes sobre as disciplinas, porém essa medida não era institucionalizada antes de 2019.

##### **3.1.2. Avaliação de Resultados de Aplicação do TP**

A avaliação dos resultados da aplicação do TP contou com a análise de planilhas com os resultados correspondentes aos dos alunos que ingressaram no curso de Medicina da FMRP USP entre 2010 e 2014, avaliando as notas por coorte e por área, além de comparar resultados do teste do progresso com resultados da prova de residência de cada coorte, a fim de avaliar possíveis correlações

#### **3.2. Participantes do Estudo**

Este estudo contou com resultados do TP de alunos das cinco coortes que ingressaram no curso de Medicina da FMRP USP entre 2010 e 2014 e que foram acompanhados até o final de seu curso. Os critérios de inclusão para ter os resultados do teste de progresso avaliados foram: ser estudante de medicina da FMRP-USP, com ingresso entre 2010 e 2014 e ter feito o TP. Os critérios de exclusão para ter os resultados do TP avaliados foram: não ser parte do grupo de interesse ou sendo parte do grupo de

interesse, ter mudado de Coorte ao longo do período ou ter faltado em mais de 4 avaliações do TP.

### 3.3. Obtenção, apresentação e análise de dados

O gabarito do TP é preenchido manualmente pelos alunos e um equipamento eletrônico lê as respostas, armazenando-as em tabelas. As análises iniciais (adesão, taxa de acerto, comparação entre universidades) e a avaliação de características psicométricas das questões das provas (dificuldade, discriminação e confiabilidade) são realizadas pela empresa EDUDATA, uma assessoria em Tecnologia da Informação que fornece serviços de estatística e que presta serviços para diversos processos seletivos e de avaliação no país, inclusive para o consórcio do NIEPAEM (Núcleo Interinstitucional de Estudos e Práticas em Avaliação em Educação Médica). Foram obtidos dados através dos documentos fornecidos pelo EDUDATA e apresentados como médias para a comparação entre os diferentes grupos. Os resultados da prova de residência foram fornecidos pela Comissão da Residência Médica (COREME) do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCRP).

Os resultados das provas dos estudantes foram obtidos através de registros brutos das respostas, sendo filtrados os dados dos estudantes que não terminaram o curso em sua coorte de origem. Os acertos foram descritos como taxa de acerto em todas as grandes áreas e não em número absoluto de questões, para facilitar as comparações entre anos de provas que tiveram número de questões distintas nas áreas.

Para as comparações entre anos e coortes foi utilizado modelo de regressão linear com efeitos mistos para a análise das variáveis. Este modelo considera um efeito aleatório por indivíduo, informando que o indivíduo possui medidas tomadas ao longo do tempo. E o efeito fixo representa a independência entre as medidas tomadas em cada indivíduo. São apresentadas as diferenças estimadas com seus respectivos valores p e intervalos de confiança 95%. Intervalos de confiança que não englobam o valor zero trazem evidências de diferença e seus limites mostram a magnitude desta diferença. Intervalos de confiança que englobam o valor zero, não trazem evidências de diferença.

São apresentados as médias, desvios-padrão, valor p do Teste t de *Student* para amostras independentes e intervalo de confiança 95% da diferença entre os resultados das diferentes coortes. A mesma observação sobre a inclusão do valor zero nos intervalos de confiança do parágrafo anterior é válida.

Com o objetivo de evidenciar possíveis correlações entre as notas obtidas nos anos 5 e 6 (internato) com as notas obtidas na prova teórica de residência médica são apresentados gráficos de dispersão com o coeficiente de correlação de Pearson e seu respectivo intervalo de confiança 95%. O coeficiente de correlação varia de -1 a 1, quanto mais próximo dos extremos (-1 ou 1) maior a magnitude da correlação e quanto mais próximo de zero menor a magnitude da correlação, sendo zero a evidência de ausência de correlação entre as variáveis. Valores positivos evidenciam uma relação direta, conforme uma variável aumenta de valor a outra também tende a aumentar de valor. Valores negativos evidenciam uma relação inversa na qual enquanto uma variável aumenta de valor para a outra ocorre diminuição. Quanto mais os valores calculados para essa correlação aproximam-se de zero, menor a correlação entre as variáveis, assim o valor zero representa a ausência absoluta de correlação, isto é, que enquanto uma variável muda de valor a outra permanece constante. Intervalos de confiança 95% que não contém o valor zero trazem evidência de associação entre as variáveis com 95% de confiança.

Ainda explorando os resultados entre as provas do 6° e da prova de residência, os estudantes que realizaram as duas provas foram categorizados em quartis conforme o desempenho na prova da sua própria coorte, a fim de se observar se o seu desempenho mudou nas diferentes provas. Os resultados foram divididos em 'Igual', 'Melhor' e 'Pior', sendo que foi considerado "Igual" o estudante que ficou no mesmo quartil na prova do 6° ano e na prova da residência, 'Melhor', o estudante que subiu de quartil na prova de residência e 'Pior' o estudante que apresentou um quartil na prova de residência inferior ao da prova do 6° ano.

Os softwares estatísticos utilizados para as análises foram: R Core Team (2016). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria (URL: <https://www.R-project.org>) e SAS Statistical Software (version 9.3; SAS Institute, Inc. Cary, NC).

### **3.4. Aspectos Éticos**

A maioria dos dados utilizados foram agregados, mas quando dados divididos por estudantes foram utilizados, suas identidades foram preservadas por codificações, sem exposição de nome ou números para garantir sua anonimidade ao longo da análise de dados. O projeto foi aprovado pelo Comitê de Ética em Pesquisa (CEP), a submissão ocorreu dentro de um projeto de educação em saúde com múltiplos focos, (nº:5.302.819). O termo de consentimento livre e esclarecido foi desnecessário, uma vez que todo o estudo se deu por análise de dados secundários e análise documental.



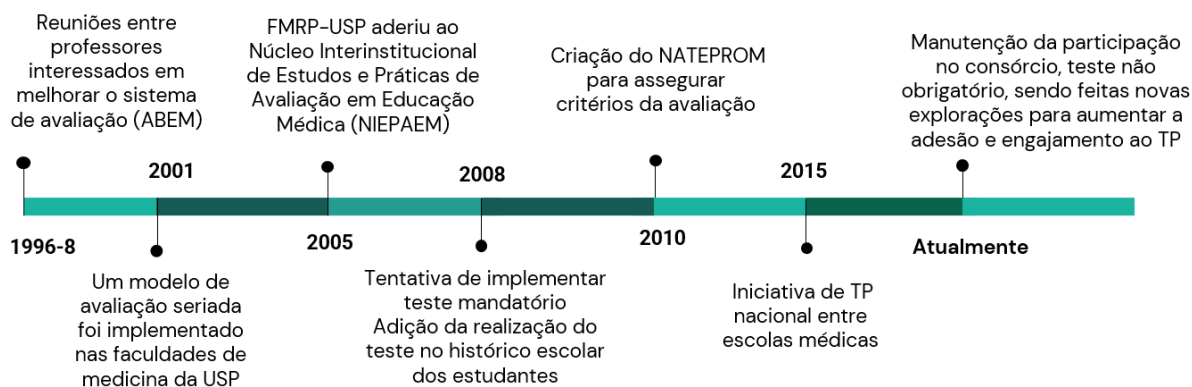
## 4. RESULTADOS

### 4.1. Uma breve descrição do TP na FMRP-USP

Com a análise de documentos durante a revisão bibliográfica, foi possível tecer um panorama geral sobre o tema e sua contextualização nas Faculdades de Medicina da USP.

A Figura 2 sumariza uma linha do tempo que situa o teste do progresso na FMRP-USP, indicando que há algumas décadas há interesse de professores da instituição sobre o tema e que foram tomadas condutas ativas para sua implementação, melhora e manutenção desde então.

Figura 2- Descrição de evolução do TP ao longo do tempo



#### 4.1.1. Desenvolvimento do Consórcio na FMRP-USP

O Teste de Progresso (TP) está consolidado na estrutura curricular do curso de Medicina da FMRP-USP. A aplicação acontece anualmente no final do segundo semestre letivo e tem caráter formativo, sendo aplicado, simultaneamente, para todos os estudantes do 1º ao 6º ano do Curso de Medicina.

A ideia de elaborar um teste de progresso nas escolas médicas brasileiras originou-se do interesse em melhorar seu sistema de avaliação. Eventos da ABEM no final dos anos 1990 sediaram reuniões entre professores interessados nesse tema que levaram ao fomento de tais ideias na década seguinte (28). A partir de 2001, um modelo de avaliação seriada foi implementado nas Faculdades de Medicina da Universidade de São Paulo (USP), dentre as quais a Faculdade de Medicina de Ribeirão Preto.

A discussão sobre elaboração do teste envolveu decisões da Comissão de Graduação (CG), da Congregação e da Comissão Coordenadora de Curso Medicina (CoC). A função da CG é traçar diretrizes e garantir a execução dos programas determinados pela estrutura curricular dos cursos de graduação. A Congregação é órgão consultivo e deliberativo da FMRP-USP com representantes de diversas instâncias do curso, inclusive da Diretoria da Unidade. A Comissão Coordenadora de Curso Medicina (CoC) é responsável pela articulação com outras instâncias, gestão acadêmica do curso, para o aperfeiçoamento e planejamento estratégico para garantir a qualidade do curso, intervenções para adesão, formação de professores e discussão de novas iniciativas curriculares.

O modelo que mais se aproxima do atual foi o criado a partir de 2005, por iniciativa de nove escolas médicas que propuseram a realização desse teste em conjunto. Com aprovação das instâncias descritas no parágrafo anterior, a FMRP-USP aderiu ao NIEPAEM, juntamente com as demais escolas. Nesse mesmo ano, a FMRP-USP passou a colaborar para a elaboração do teste e, a partir de 2006, passou a aplicar o TP sistematicamente para todos os alunos do curso de Medicina.

Para adequado andamento dos trabalhos desse grupo, a FMRP-USP constituiu oficial e legalmente o Núcleo de Apoio ao Teste de Progresso para Medicina (NATEPROM) em 2010. Esse núcleo buscava dar suporte aos trabalhos da CoC Medicina na elaboração e seleção de questões, divulgação e aplicação da prova, confecção de relatório, análise dos resultados, discussão dos resultados nos diferentes departamentos e acompanhamento desenvolvimento dos estudantes que apresentem deficiências no ganho de conhecimento durante o processo de formação.

Vale destacar que em 2015 houve uma tentativa de unificação das provas, realizando um teste de progresso nacional. Nessa prova, por incluir escolas de modelo PBL (*problem based learning*) que não dividem os conteúdos em ciclos fixos anuais, a área básica não foi incluída diretamente, sendo abordada nas outras áreas de maneira mais contextualizada. Foram aplicadas 24 questões das áreas: clínica geral, pediatria, ginecologia e obstetrícia, cirurgia e saúde coletiva. O modelo não se sustentou nos anos seguintes, mantendo-se as seis áreas no teste e envolvendo apenas as escolas do consórcio, mas serviu como ponto de partida para elaboração de novos consórcios pelo país e novas iniciativas quanto ao TP.

#### 4.1.2. Características do Teste e Elaboração das Questões

A prova é aplicada com duração de 5 horas, oferecendo 150 segundos em média

para cada questão. É composto por 120 questões de múltipla escolha, com cinco alternativas cada, dividida em 6 áreas do conhecimento: ciências básicas, clínica geral, pediatria, ginecologia e obstetrícia, cirurgia e saúde coletiva, dentro desta última existem aquelas que envolvem o tema de ética e bioética. No início da elaboração do teste, ética e bioética tinham sua área separada no teste compondo 6 questões, com o restante das áreas tendo 19 questões, mas a partir de 2016 o tema foi incorporado e contextualizado dentro das questões de saúde coletiva.

As questões do TP são formuladas pelos professores das diversas escolas que participando consórcio. O modelo mais adotado de formulação das questões dos TP internacionais é o de teste de múltipla escolha, com uma única alternativa correta (*One-best answer*). Como modelo proposto para o NIEPAEM, também foi adotado esse padrão, com quatro alternativas possíveis. A alternativa “não sei” é ausente nas provas do consórcio, assim como a pontuação negativa a cada resposta errada. Inicialmente, o modelo para elaboração de questões envolve o uso de uma matriz de competências que detalha quais seriam os temas mais relevantes para formação do estudante no final do curso de medicina. São então elaboradas questões das grandes áreas por diversos professores do consórcio e, após isso, um núcleo revisor discute sua relevância, formato, se o tema já havia sido contemplado em outra questão, se a prova estava repetitiva e se havia ambiguidade nas respostas.

Para que tal processo funcionasse, o núcleo concebia diretrizes de elaboração de questões de múltipla escolha (*blueprint*) e solicitava aos professores não só a questão com a resposta, mas também sua fonte da informação. Mais recentemente, há também uma matriz de questões dentro de cada área, considerando o nível de complexidade. Assim, são definidos os cenários e a porcentagem de questões que aparecem em cada área: nos cenários de atenção primária, média complexidade, alta complexidade e cenário de urgência e emergência. Isso está distribuído com uma porcentagem, conforme a proporcionalidade epidemiológica de aparecimento na prática médica.

Outra estratégia interessante presente na elaboração do teste são as “questões-âncora”, questões que são repetidas ano a ano, seja em sua totalidade ou em estruturação temática. Essas questões permitem avaliar a frequência de erros dos estudantes nessas questões repetidas, ajudando a inferir se os estudantes estão utilizando a devolutiva da prova para revisar a prova realizada, de acordo com a taxa de acerto da questão a cada

ano.

#### 4.1.3. O Contexto do Teste na FMRP-USP

Em relação ao currículo no qual o teste de progresso se insere e avalia, a faculdade apresentava no período relacionado às coortes de estudantes avaliados, um currículo tradicional de 12 semestres, dividido em três ciclos: um de ciências básicas (anatomia, biologia celular, histologia, fisiologia e outras disciplinas); um de introdução às ciências clínicas; e finalmente o internato. No internato acontece o rodízio em estágios na Atenção Básica, nos ambulatórios e enfermarias, Centro Cirúrgico de hospitais gerais (Hospital Estadual, Hospital Estadual Américo Brasiliense) e especializados (Hospital das Clínicas de Ribeirão Preto/Unidade de Emergências), na maternidade, sala de emergência, sempre rodando nas especialidades básicas da formação médica: Medicina Interna, Pediatria, Cirurgia, Medicina de Família e Comunidade, Obstetrícia e Ginecologia.

A carga horária do curso variou um pouco entre os anos estudados, com carga horária próxima de 10.000 horas, dentre elas mais de 4.000 horas de estágios práticos. Ao explorar as disciplinas presentes no curso, foi possível dividir a maioria dentro de alguma das 5 grandes áreas da prova, salvo aquelas que discutem competências práticas, comunicação ou ética (apêndice 1). Uma estimativa de horas de atividades teóricas de cada disciplina foi feita a partir dos roteiros de disciplinas de cada ano. Há distribuição similar entre os anos de horas para as áreas, com maior quantidade de conteúdo em Clínica Geral e da Área Básica.

Quando se pensa no contexto avaliativo, todas as disciplinas da FMRP-USP têm atividades avaliativas ao seu final, e todas têm provas com caráter somativo, ainda que em algum momento complementada por atividades formativas. Ao longo do curso, as avaliações mudam de majoritariamente teóricas nos primeiros anos do curso (testes de múltipla escolha, questões dissertativas, seminários), para avaliações ao final do curso que além do componente cognitivo do saber avaliam também componentes de aquisição das habilidades e demonstração de atitudes (estações práticas, consulta observada, simulações).

Ainda, em 2014 foi retomada uma Avaliação Prática Terminal, prova aplicada ao final do 5º e 6º anos do curso. É composta por estações práticas das grandes áreas do ensino médico que analisam habilidades, atitudes e competências do estudante. Nesse mesmo ano, foi criado o Centro de Avaliação do Estudante de Graduação (CAEG) que ficou responsável por auxiliar a CG na aplicação e utilização dos dados do TP e pela aplicação e expansão da Avaliação Prática Terminal para além dos anos do internato.

Nesse contexto, a aplicação do TP acontece anualmente no final do segundo semestre letivo na FMRP-USP para todos os alunos, em caráter formativo, a nota é baseada na teoria clássica dos testes, isto é, cada resposta correta consiste em um ponto, sem diferente proporção entre os itens e sem punição por questões erradas. A participação no teste não é obrigatória, porém desde 2007, dois anos após iniciada a aplicação do teste com participação no consórcio, passou a ser associado a uma disciplina optativa e começou a ter importância no histórico escolar do estudante. Nesse caso, não há uma nota que é acrescentada ao currículo ou a sua média ponderada, mas sim o registro da presença ou não, o que pode ser considerado nas seleções de residências médicas do país.

Quanto à distribuição de resultados, notou-se que eles eram apresentados assim que liberados aos professores, chefes de departamentos, mas não era sempre que discutiam a fundo o que cada resultado significava. Também não foram encontrados relatórios que permitissem avaliar a qualidade dos dados enviados aos departamentos. Por outro lado, os relatórios referentes à qualidade de questão e erros em “questões-âncora” eram amplamente discutidos entre os professores membros do NATEPROM e do Consórcio e repassado depois à comunidade de professores da instituição.

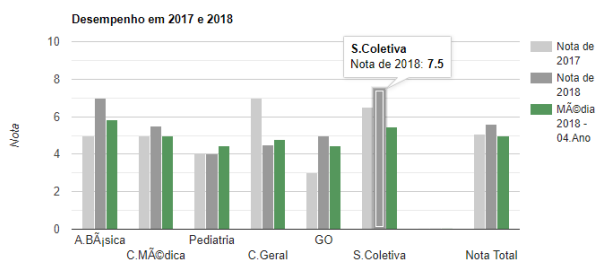
Nas reuniões da CoC, outro ponto levantado nas discussões foi a necessidade de incluir os discentes nos processos relacionados ao TP, que não podem participar da elaboração de questões, mas que deveriam participar das revisões sobre os resultados, formato e frequência.

Alguns professores que não participavam do NATEPROM ou necessariamente da elaboração da prova tiveram reuniões sobre o TP em encontros mediados pela ABEM em 2012 e indicaram que tiveram novas perspectivas dos resultados a partir de discussões mais aprofundadas, sendo favoráveis a discussões mais aprofundadas sobre os resultados de sua instituição rotineiramente.

#### **4.2. Apresentação de Dados aos Estudantes**

A apresentação dos dados aos estudantes, após a realização da prova e computação dos resultados, ocorre como apresentado abaixo (figura 3), recebendo via e-mail e podendo ser checado em site próprio. Além disso, a devolutiva do conteúdo teste de progresso na FMRP ocorre através de um gabarito comentado das questões que é disponibilizado online e por e-mail.

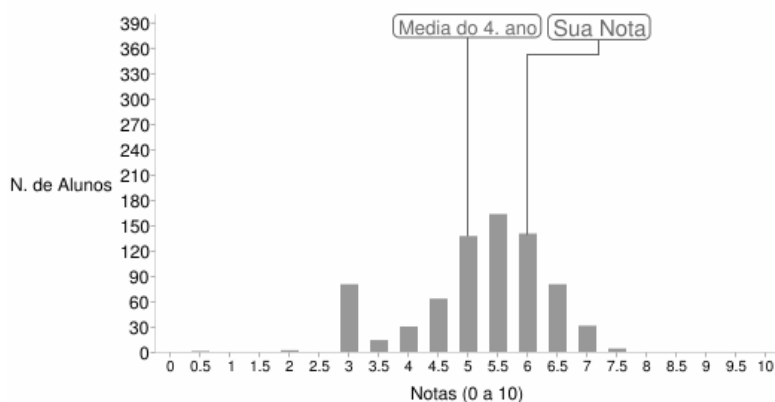
Figura 3- Exemplo de Apresentação de resultados do TP aos estudantes



A tabela a seguir mostra o seu desempenho por tipo de prova (em percentual e graficamente), permitindo que você verifique a evolução de seu progresso em 2018, em comparação com o ano de 2017.

	Área Básica	Clinica Médica	Pediatría	Clinica Cirúrgica	Ginecologia e Obstetria	Saúde Coletiva	Total
Desempenho (%)	+40	0	0	-42.86	+66.67	+16.67	0
Progresso	↑	—	—	↓	↑	↑	—

Histograma das Notas do 4. ano de todas as faculdades participantes



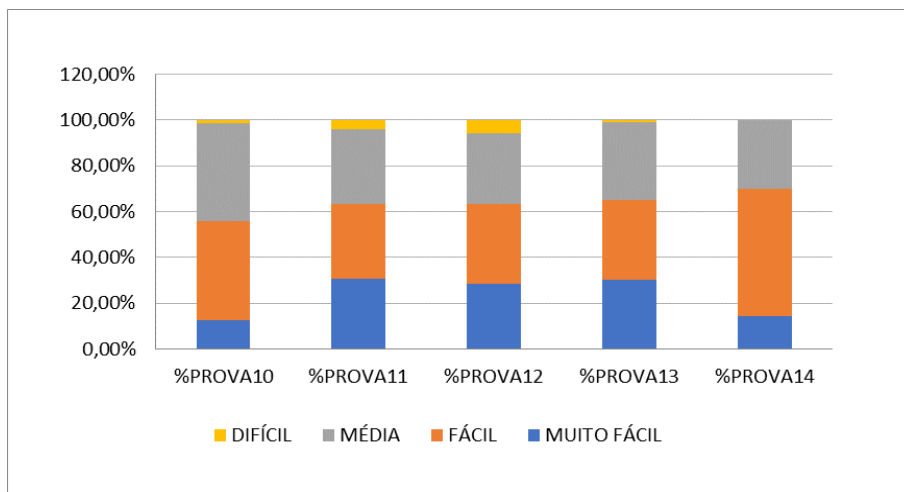
Nota-se que se oferece o desempenho em acertos absolutos e compara-se a média com o resultado dos estudantes do consórcio. Ainda, permite ao estudante comparar suas notas do ano atual com o ano anterior, tanto no aspecto geral como por cada grande área. Porém, não há informação que permita dizer se o progresso ou regresso do desempenho são estatisticamente significativos ou não. Também não são oferecidas informações sobre a construção da prova em si: dificuldade das questões por área, acerto/erro de questões-âncora das provas, isto é, as perguntas chave que detêm os conhecimentos mais prevalentes ou essenciais.

#### 4.3. Características Psicométricas do Teste de Progresso ao longo dos anos

Baseados em análises fornecidas pelo EDUDATA, as respostas das questões foram analisadas em relação a sua dificuldade (Figura 4) e discriminação (Figura 5) em diferentes grupos de estudantes. Foram utilizados os resultados dos alunos do 6º ano de todo o consórcio

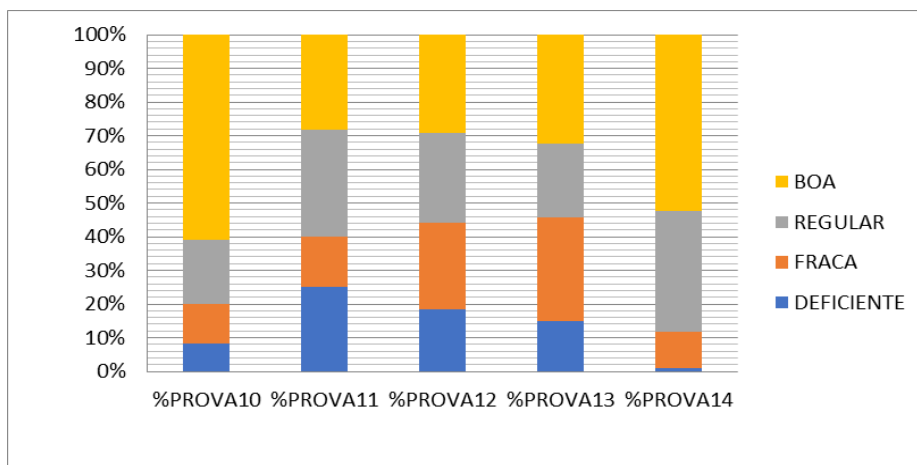
para essa análise, temos os dados das 5 coortes de estudantes e as provas que realizaram no 6º ano correspondem às avaliações de 2015 a 2019.

Figura 4 - Distribuição de Dificuldade das Questões - 6º Anos do Consórcio



Legenda: PROVA10: prova realizada pelo 6º da coorte de 2010; PROVA11: prova realizada pelo 6º da coorte de 2011; PROVA12: prova realizada pelo 6º da coorte de 2012; PROVA13: prova realizada pelo 6º da coorte de 2013; PROVA14: prova realizada pelo 6º da coorte de 2014

Figura 5 - Distribuição de Discriminação das Questões - 6º Anos do Consórcio



Legenda: PROVA10: prova realizada pelo 6º da coorte de 2010; PROVA11: prova realizada pelo 6º da coorte de 2011; PROVA12: prova realizada pelo 6º da coorte de 2012; PROVA13: prova realizada pelo 6º da coorte de 2013; PROVA14: prova realizada pelo 6º da coorte de 2014

Em relação à distribuição da dificuldade das questões, as provas da coorte de 2010 e 2014 apresentaram-se equilibradas, com cerca de 85% das questões entre “fácil” e “média”, enquanto as provas de 2011 a 2013 apresentaram entre 65-70% das questões nessa classificação, sendo que as provas de 2011 e 2012 ainda tinham mais questões nos extremos (muito fácil ou

difícil) cerca de 35%, com a maior parte sendo “muito fácil”. Nas coortes de 2013 e 2014 não existiram questões difíceis.

Quanto à discriminação das questões, as provas com as questões que melhor discriminaram seus estudantes foram 2010 e 2014, pois quando se avaliar a prevalência de questões regulares ou boas as provas dessas coorte, elas somam 80% ou mais de questões com discriminação “regular” ou “boa”, enquanto as provas das coortes de 2011, 2012 e 2013 tem entre 55 e 60% das questões com essa classificação. As provas com mais questões classificadas como “deficiente” também são as provas de 2011, 2012 e 2013 (25%, 18% e 15% respectivamente).

Tabela 1 – Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova da Coorte de 2010

	DIFICULDADE (%)				DISCRIMINAÇÃO (%)			
	Muito Fácil	Fácil	Média	Difícil	Deficiente	Fraca	Regular	Boa
GO	0	54	38	8	4,2	8,3	12,5	75
CX	12,5	41,7	45,8	0	8,3	12,5	41,7	37,5
SC	25	20,8	54,2	0	16,7	20,8	8,3	54,2
P	8,3	50	41,7	0	4,2	8,3	12,5	75
CM	16,7	50	33,3	0	8,3	8,3	20,8	62,6

LEGENDA: GO: Ginecologia e Obstetrícia; CX: Cirurgia; SC: Saúde Coletiva; P: Pediatria; CM: Clínica Médica; BASC: Área Básica

Na prova 2010, as questões consideradas difíceis só estiveram presentes na parte relativa a GO. CX apresentou a menor porcentagem questões classificadas como “boa” quanto a discriminação. SC apresentou mais questões “deficientes”. Estas duas áreas também apresentaram o maior número de questões classificadas como “fáceis” ou “muito fáceis” (tabela 1)

Tabela 2 – Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova de Coorte de 2011

	DIFICULDADE (%)				DISCRIMINAÇÃO (%)			
	Muito Fácil	Fácil	Média	Difícil	Deficiente	Fraca	Regular	Boa
GO	10	30	55	5	20	5	10	65
CX	35	45	10	10	25	15	45	15
SC	40	25	30	5	5	30	35	30
P	35	45	20	0	25	10	40	25
CM	20	40	35	5	45	10	45	0
BASC	45	10	45	0	30	20	15	35



LEGENDA: GO: Ginecologia e Obstetrícia; CX: Cirurgia; SC: Saúde Coletiva; P: Pediatria; CM: Clínica Médica; BASC: Área Básica

Na prova da coorte de 2011, apenas GO apresentou maioria das questões com dificuldade “média”, sendo também a única grande área com questões com maioria de discriminação das questões “boa”. Por outro lado, GO também apresentou um quinto das questões como deficientes. As áreas de CM e BASC também apresentaram questões com menos discriminação, com cerca de metade das questões consideradas “fraca” ou “deficiente” (tabela 2).

Tabela 3 – Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova da Coorte de 2012

	DIFICULDADE (%)				DISCRIMINAÇÃO (%)			
	Muito Fácil	Fácil	Média	Difícil	Deficiente	Fraca	Regular	Boa
GO	5	30	60	5	5	35	25	35
CX	35	40	20	5	30	15	30	25
SC	40	20	25	15	25	30	15	30
P	25	30	40	5	15	35	20	30
CM	45	35	20	0	20	30	25	25
BASC	20	55	20	5	15	10	45	30

LEGENDA: GO: Ginecologia e Obstetrícia; CX: Cirurgia; SC: Saúde Coletiva; P: Pediatria; CM: Clínica Médica; BASC: Área Básica

Para a coorte de 2012 a prova de GO foi a que mais apresentou questões de dificuldade “média”, enquanto CM e SC foram as áreas com mais questões classificadas como “muito fácil”. Quando se observar a discriminação, CX e SC apresentaram maior porcentagem de questões deficientes, entre 25-30%, mas todas as grandes áreas se aproximaram em proporção de questões com boa discriminação (tabela 3).

Particularmente nesse ano, houve muita discussão entre os professores do consórcio sobre a qualidade da prova, questões específicas foram questionadas, pois foram obtidos menos de 50% de acerto entre estudantes do 6º ano, apontando possíveis problemas de elaboração.

Tabela 4 – Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova da Coorte de 2013

	DIFICULDADE (%)				DISCRIMINAÇÃO (%)			
	Muito Fácil	Fácil	Média	Difícil	Deficiente	Fraca	Regular	Boa
GO	25	35	40	0	15	20	10	55
CX	35	35	30	0	15	35	15	35

SC	40	40	15	5	20	30	20	30
P	35	25	40	0	10	35	25	30
CM	15	45	40	0	15	20	30	35
BASC	30	30	40	0	15	45	30	10

LEGENDA: GO: Ginecologia e Obstetrícia; CX: Cirurgia; SC: Saúde Coletiva; P: Pediatria; CM: Clínica Médica; BASC: Área Básica

Nota-se que na prova da coorte de 2013, o número de questões difíceis é mínimo e que não houve predomínio de questões de uma dificuldade nas áreas de GO, CX, P e BASC, com variações discretas, apenas CM com proporção menor que 20% de questões “muito fáceis”. Quanto à discriminação, há distribuição homogênea de questões deficientes, com apenas GO tendo um predomínio de questões “boas” (tabela 4).

Tabela 5 – Características de Dificuldade e Discriminação em cada área do Conhecimento da Prova da Coorte de 2014

	DIFICULDADE (%)				DISCRIMINAÇÃO (%)			
	Muito Fácil	Fácil	Média	Difícil	Deficiente	Fraca	Regular	Boa
GO	20	60	20	0	0	15	25	60
CX	20	65	15	0	0	5	35	60
SC	35	35	30	0	5	25	35	35
P	5	70	25	0	0	10	45	45
CM	0	75	25	0	0	5	45	50
BASC	5	30	65	0	0	5	30	65

LEGENDA: GO: Ginecologia e Obstetrícia; CX: Cirurgia; SC: Saúde Coletiva; P: Pediatria; CM: Clínica Médica; BASC: Área Básica

Na prova da coorte de 2014, não houve questões difíceis, e a maior parte das áreas apresentou principalmente questões “fáceis”, com apenas a área Básica tendo a maioria de questões de dificuldade média. Na discriminação, na maior parte das áreas prevaleceu as questões regulares ou boas, com mínimas questões deficientes (tabela 5).

#### 4.3.1 Coeficiente de fidedignidade

Ainda foi possível registrar o coeficiente de fidedignidade de todas as provas realizadas no período, baseada nos resultados dos estudantes do 6º ano do consórcio (tabela 6). Esse coeficiente que busca oferecer uma estimativa da estabilidade dos itens que compõem o teste e se o item diferencia os examinados, quando eles apresentam níveis diversos de desempenho.

Tabela 6 – Coeficiente de Confiabilidade das provas durante todo o período estudado

<b>Ano da prova do 6º ano</b>	<b>BASC</b>	<b>CX</b>	<b>GO</b>	<b>PED</b>	<b>CM</b>	<b>SC</b>	<b>TOTAL</b>
2010	0,31	0,59	0,62	0,46	0,49	0,43	0,51
2011	0,4	0,61	0,64	0,42	0,33	0,56	0,51
2012	0,6	0,51	0,64	0,57	0,66	0,64	0,60
2013	0,39	0,7	0,62	0,63	0,48	0,47	0,58
2014	0,57	0,62	0,49	0,5	0,32	0,55	0,49
2015	-	0,65	0,73	0,74	0,74	0,67	0,70
2016	0,51	0,58	0,58	0,5	0,47	0,64	0,55
2017	0,53	0,47	0,45	0,47	0,53	0,62	0,50
2018	0,29	0,52	0,6	0,53	0,44	0,62	0,54
2019	0,61	0,69	0,68	0,61	0,62	0,67	0,65

LEGENDA: GO: Ginecologia e Obstetrícia; CX: Cirurgia; SC: Saúde Coletiva; P: Pediatria; CM: Clínica Médica; BASC: Área Básica

#### 4.4. Participantes Elegíveis e Adesão

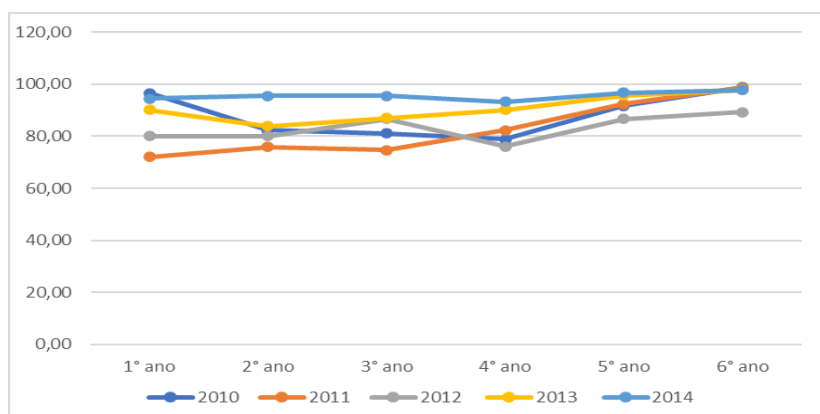
As análises foram realizadas com os dados dos estudantes que preencheram os critérios e elegibilidade, a fim de obter análises mais fidedignas e maior controle dos dados (tabela 7). Foram utilizados dados de 81,4% dos estudantes, com melhor aproveitamento das últimas duas coortes (2013 e 2014).

Tabela 7 – Porcentagem de estudantes elegíveis e participantes em cada Coorte

<b>Coortes</b>	<b>N coorte original</b>	<b>N participantes (%)</b>
2010	102	85 (83%)
2011	107	79 (73%)
2012	98	75 (76%)
2013	104	92 (88%)
2014	103	90 (87%)

Em relação à adesão, ela tende a diminuir após o primeiro ano de prova, com abstenções chegando até 20%, e crescer novamente nos últimos anos do curso, com abstenções ficando menor que 10% em todos os grupos de estudantes (figura 6).

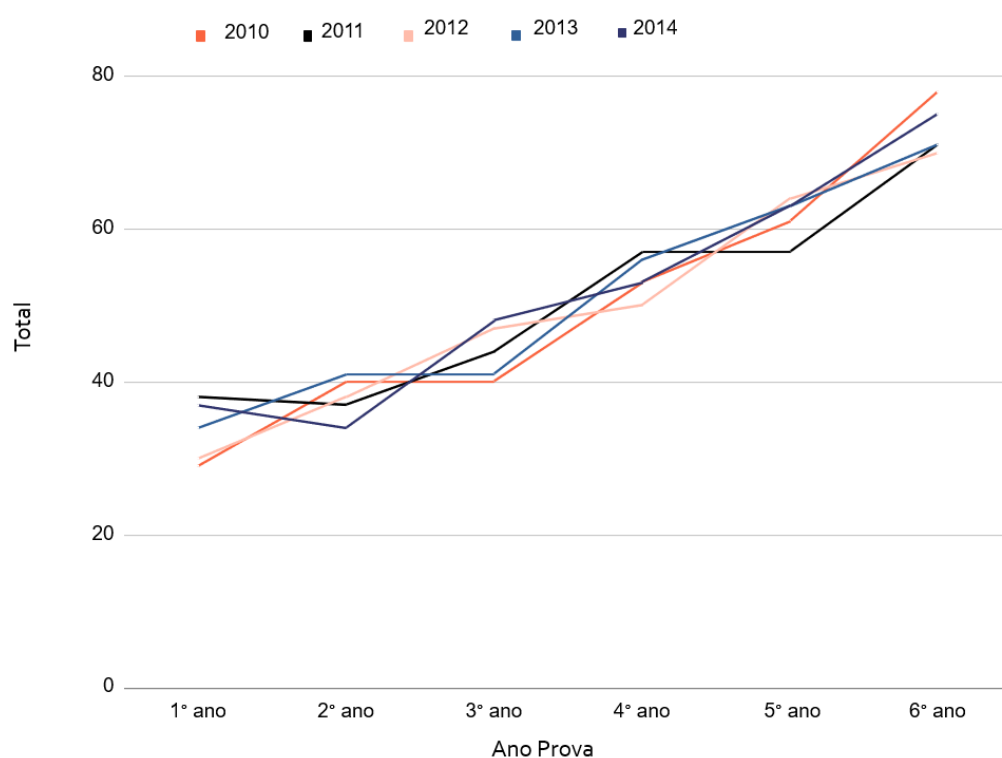
Figura 6 - Taxa de adesão dos estudantes ao longo dos anos



#### 4.5. Descrição Inicial da Progressão dos Resultados

As médias de desempenho global das Coortes crescem em termos nominais entre o primeiro e o último ano do curso, com as médias do último ano ficando todas em torno dos 70% de acerto (2010: 70,5%; 2011: 71,1%; 2012: 70,6%; 2013: 71%; 2014: 75%); com desvio padrão menor que 10% em todos os grupos. O crescimento nominal total aumenta do 1º ao 6º ano em todos os grupos (figura 7), entre 30-40% (2010: 41,3%; 2011: 33%; 2012: 40,3%; 2013:37%; 2014: 37,5%), com uma média de crescimento nominal de 37,8%.

Figura 7 – Representação de Desempenho de Coortes de estudantes ao longo de seu curso



Legenda: Total: acerto geral das provas, AnoProva: ano do curso

Houve diferença significativa de resultados entre o desempenho do 1º e do 6º ano em todas as coortes. Esse crescimento (diferença significativa entre resultados) também ocorreu a cada ano em quase todas as coortes, havendo pontos específicos nos quais não foram observadas mudanças significativas (tabela 8). Na coorte de 2010, houve queda não significativa do desempenho entre o 2º e 3º ano da prova, assim como no desempenho entre o 1º e 2º ano e entre 4º e 5º anos na coorte de 2011. A coorte de 2012 foi a única que apresentou crescimento significativo (diferença entre o desempenho) entre todas as provas da coorte. Na coorte de 2013, só não houve diferença significativa entre os resultados entre o 2º e 3º anos e na coorte de 2014, houve queda significativa do desempenho entre a 1ª e a 2ª prova, com desempenho crescente nos demais anos.

Não houve padrão para o período da graduação em que houve maior salto de desempenho comparado, sendo maior entre o 3º e 4º anos nas coortes de 2010 e 2013, entre a 4ª e a 5ª na de 2012 e entre a 5ª e 6ª na de 2011.

Tabela 8 - Comparação da média da progressão do desempenho geral no TP

Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
	Ano	Ano			Limite Inferior	Limite Superior
2010	1	2	-0,1112	0,0001	-0,1306	-0,0918
2010	2	3	0,0066	0,5186	-0,0135	0,0268
2010	3	4	-0,1300	0,0001	-0,1505	-0,1095
2010	4	5	-0,0852	0,0001	-0,1051	-0,0652
2010	5	6	-0,0920	0,0001	-0,1107	-0,0733
<b>2010</b>	<b>1</b>	<b>6</b>	<b>-0,4117</b>	<b>0,0001</b>	<b>-0,4302</b>	<b>-0,3933</b>
2011	1	2	0,0106	0,3473	-0,0116	0,0328
2011	2	3	-0,0751	0,0001	-0,0971	-0,0531
2011	3	4	-0,1342	0,0001	-0,1556	-0,1127
2011	4	5	-0,0020	0,8468	-0,0224	0,0184
2011	5	6	-0,1449	0,0001	-0,1643	-0,1255
2011	1	6	-0,3455	0,0001	-0,3665	-0,3246
2012	1	2	-0,0789	0,0001	-0,1008	-0,0570
2012	2	3	-0,0883	0,0001	-0,1096	-0,0669
2012	3	4	-0,0283	0,0106	-0,0499	-0,0066
2012	4	5	-0,1466	0,0001	-0,1683	-0,1248
2012	5	6	-0,0612	0,0001	-0,0820	-0,0405
<b>2012</b>	<b>1</b>	<b>6</b>	<b>-0,4033</b>	<b>0,0001</b>	<b>-0,4246</b>	<b>-0,3819</b>
2013	1	2	-0,0685	0,0001	-0,0874	-0,0497
2013	2	3	-0,0060	0,5374	-0,0250	0,0130
2013	3	4	-0,1472	0,0001	-0,1659	-0,1286

2013	4	5	-0,0764	0,0001	-0,0946	-0,0582
2013	5	6	-0,0730	0,0001	-0,0908	-0,0552
<b>2013</b>	<b>1</b>	<b>6</b>	<b>-0,3711</b>	<b>0,0001</b>	<b>-0,3892</b>	<b>-0,3529</b>
2014	1	2	0,0363	0,0001	0,0182	0,0545
2014	2	3	-0,1405	0,0001	-0,1586	-0,1224
2014	3	4	-0,0491	0,0001	-0,0673	-0,0309
2014	4	5	-0,1009	0,0001	-0,1191	-0,0827
2014	5	6	-0,1211	0,0001	-0,1390	-0,1031
<b>2014</b>	<b>1</b>	<b>6</b>	<b>-0,3753</b>	<b>0,0001</b>	<b>-0,3934</b>	<b>-0,3572</b>

#### 4.6. Descrição e Comparação de resultados Intra-Coortes

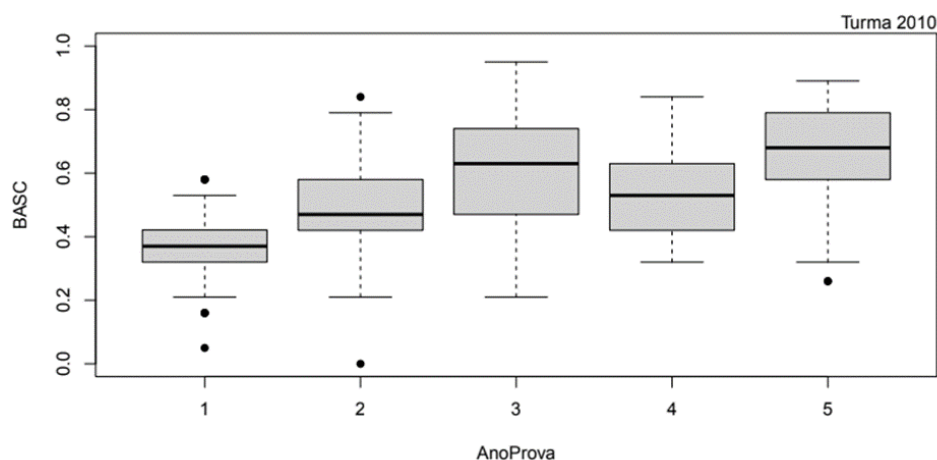
As tabelas com os dados gerais de cada Coorte, por ano e por área podem ser encontradas com todos os dados disponíveis no apêndice 3.

##### 4.6.1. Descrição dos Dados da Coorte de 2010

###### 4.6.1.1. Avaliação de resultados da área Básica

Quanto ao desempenho da coorte de 2010 nas questões da área básica, houve crescimento positivo ao longo dos anos (figura 8). No 6º ano dessa coorte não houve área básica na prova, pois foi uma prova unificada pela ABEM. Na tabela de comparações, é possível perceber que em todos os anos houve crescimento significativo a cada ano, sem nenhuma queda.

Figura 8 – Desempenho da Coorte de 2010 nas questões da área básica

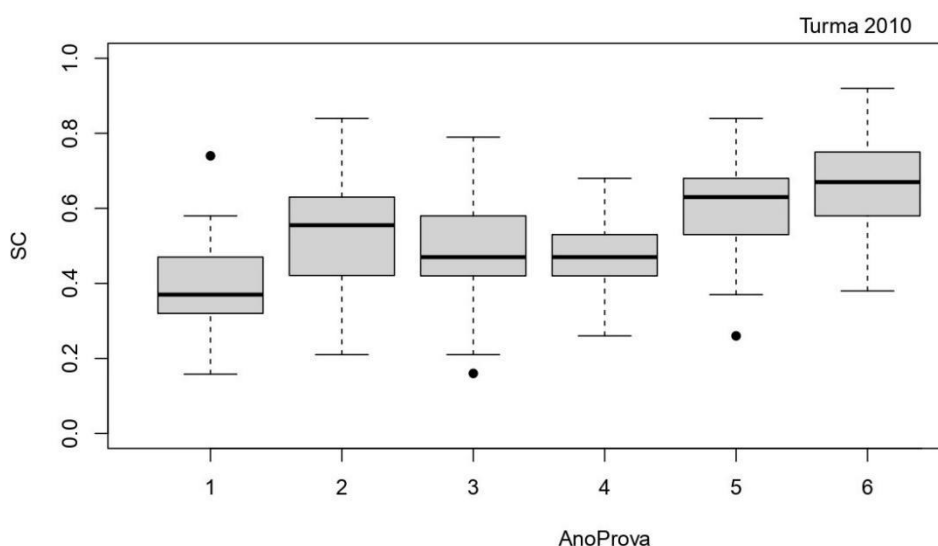


Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

###### 4.6.1.2. Avaliação de resultados da área de Saúde Coletiva

Nas questões de saúde coletiva, houve um crescimento positivo ao longo dos anos (figura 9). Na tabela de comparações, é possível perceber que houve leve queda dos resultados entre o 2º e 3º anos e não houve crescimento entre o 3º e 4º anos, mas houve ganho significativo em seguida. Também houve ganho quando se considera o crescimento global da coorte.

Figura 9 – Desempenho da Coorte de 2010 nas questões da Saúde Coletiva

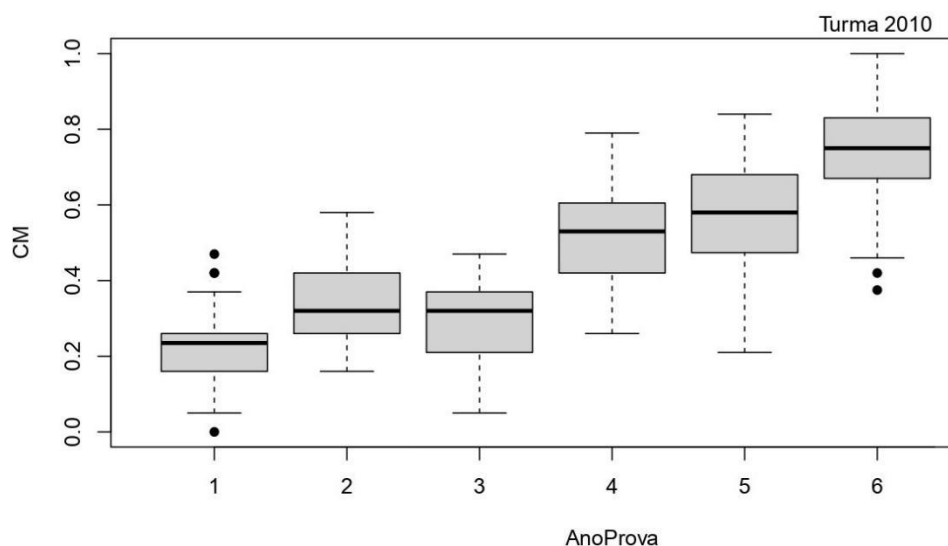


Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica outlier

#### 4.6.1.3. Avaliação de resultados da área de Clínica Médica

No desempenho da coorte de 2010 nas questões de clínica médica, houve crescimento progressivo da proporção de acertos ao longo dos anos (figura 10), com leve queda entre o 2º e o 3º ano, mas recuperação significativa após. O maior crescimento pode ser observado entre os anos de 3º e 4º anos.

Figura 10 – Desempenho da Coorte de 2010 nas questões da Clínica Médica

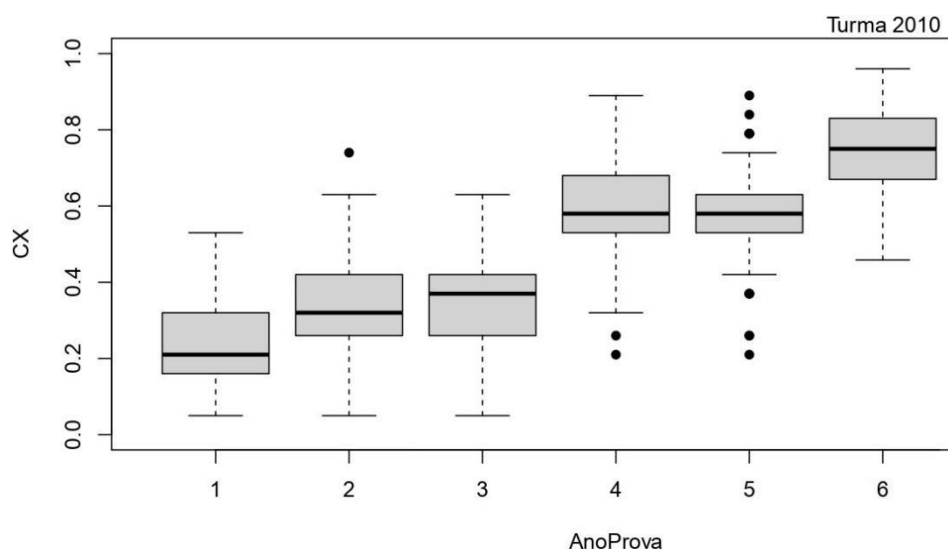


Legenda: CM: área clínica médica da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica outlier

#### 4.6.1.4. Avaliação de resultados da área de Cirurgia

Nas questões de cirurgia, houve crescimento positivo geral ao longo dos anos na coorte de 2010 (figura 11), com exceção dentre as provas do 2º e 3º anos do curso e entre 4º e 5º. O maior crescimento pode ser observado entre as provas do 3º e 4º anos do curso.

Figura 11 – Desempenho da Coorte de 2010 nas questões de Cirurgia



Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica outlier

#### 4.6.1.5. Avaliação de resultados da área de Pediatria

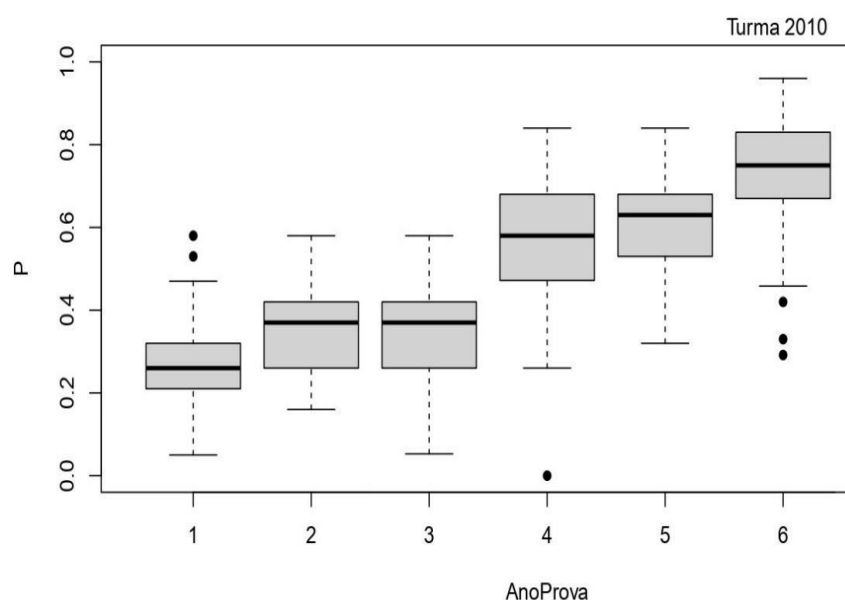
No desempenho da coorte de 2010 nas questões de pediatria houve crescimento geral positivo ao longo dos anos (figura 12), sem crescimento entre o 2º e 3º anos do curso, e entre a



4° e 5°, mas com recuperação significativa após. O maior crescimento pode ser observado entre as provas do 3° e 4° anos.

No ano de ingresso dessa coorte houve uma discussão sobre problemas de elaboração de questões de pediatria, relacionados ao não seguimento das diretrizes de elaboração. Essa discussão fez parte do processo que levou à criação do NATEPROM, buscando um grupo que pudesse participar ativamente no processo de melhoria da avaliação e garantissem a qualidade e padronização das questões realizadas por membros da instituição para o consórcio.

Figura 12 - Desempenho da Coorte de 2010 nas questões de Pediatria

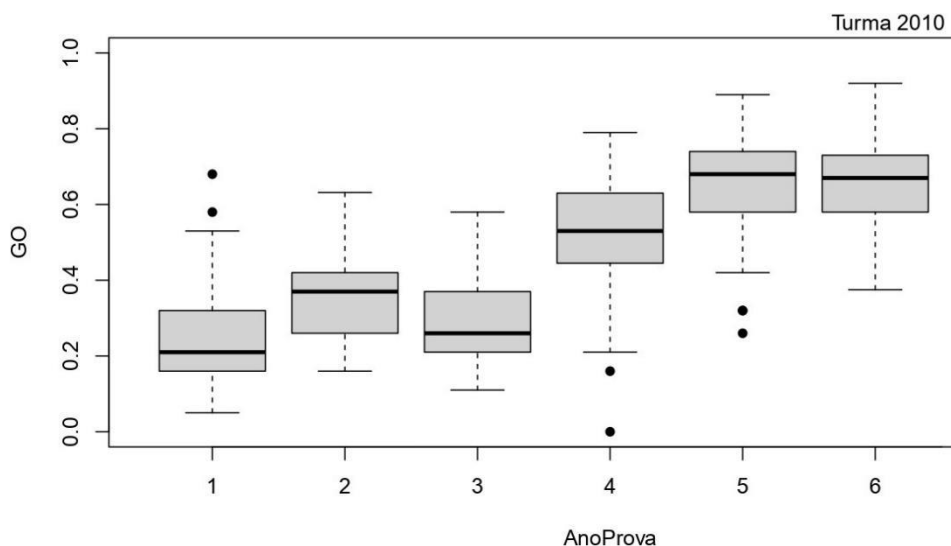


Legenda: P: área de pediatria prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.1.6. Avaliação de resultados da área de Ginecologia e Obstetrícia

No desempenho da coorte de 2010, quanto às questões de GO houve crescimento geral positivo ao longo dos anos (figura 13), com queda entre as provas do 2° e 3° anos, mas com recuperação significativa após, sem crescimento entre o 5° e 6° anos. O maior crescimento pode ser observado entre os 3° e 4° anos.

Figura 13 - Desempenho da Coorte de 2010 nas questões de Ginecologia e Obstetrícia



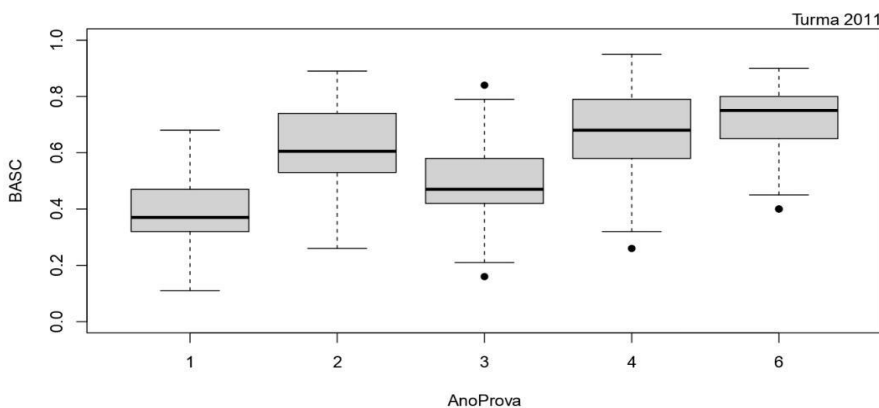
Legenda: GO: área de Ginecologia e Obstetrícia da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.2. Descrição dos Dados da Coorte de 2011

##### 4.6.2.1. Avaliação de resultados da área Básica

O desempenho da coorte de 2011 nas questões da área básica é positivo ao longo dos anos (Figura 14). No 5º ano dessa coorte não houve área básica na prova, pois foi uma prova unificada pela ABEM. Em quase todos os anos houve crescimento significativo a cada ano, menos entre o 4º e 6º ano, porém não houve queda no resultado.

Figura 14 – Desempenho da Coorte de 2011 nas questões da Área Básica

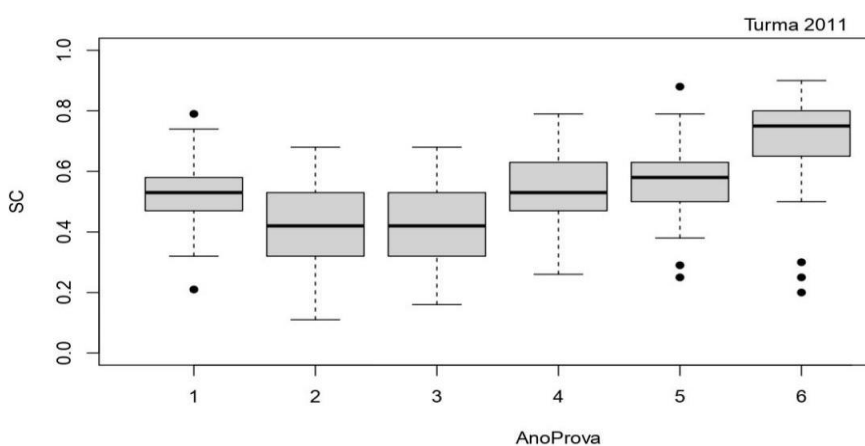


Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.2.2. Avaliação de resultados da área de Saúde Coletiva

Ao observar o desempenho da coorte de 2011 nas questões de saúde coletiva, é possível observar um crescimento positivo ao longo dos anos (figura 15), com maior salto de crescimento entre o 5º e 6º anos. Não houve crescimento cognitivo entre as provas do 2º e 3º ano e 4º e 5º ano, mas é possível perceber, na tabela de comparações, crescimento significativo quando se considera o início e término do curso.

Figura 15 – Desempenho da Coorte de 2011 nas questões de Saúde Coletiva

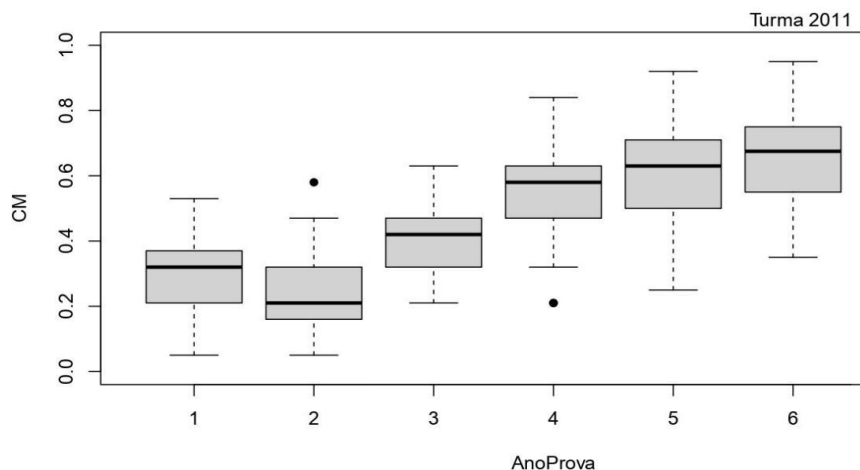


Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica outlier

#### 4.6.2.3. Avaliação de resultados da área de Clínica Médica

No desempenho da coorte de 2011 nas questões de Clínica médica, houve crescimento geral positivo ao longo dos anos (figura 16), a partir do 2º ano de prova, sem quedas após esse momento. O maior crescimento pode ser observado entre as provas do 2º e 3º anos.

Figura 16- Desempenho da Coorte de 2011 nas questões de Clínica Médica

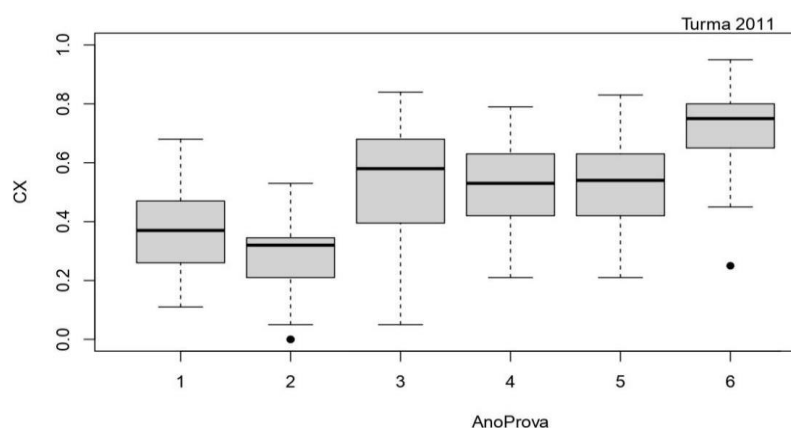


Legenda: CM: área de clínica médica da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica outlier

#### 4.6.2.4. Avaliação de resultados da área de Cirurgia

No desempenho da coorte de 2011 nas questões de cirurgia houve crescimento geral positivo ao longo dos anos (figura 17), porém com queda entre as provas do 1° e 2° anos, e sem crescimento entre o 3° e 4° ano, mas com recuperação significativa após. O maior crescimento pode ser observado entre as provas do 2° e 3° anos.

Figura 17- Desempenho da Coorte de 2011 nas questões de Cirurgia

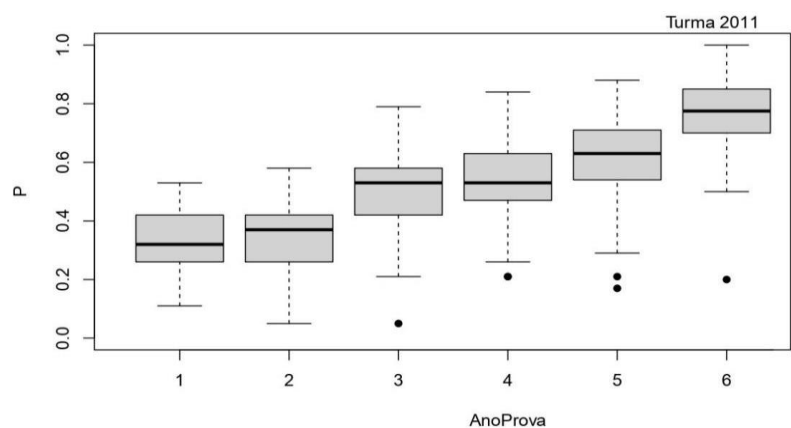


Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.2.5. Avaliação de resultados da área de Pediatria

O desempenho da coorte de 2011 nas questões de pediatria houve crescimento geral positivo ao longo dos anos (figura 18), sem crescimento significativo entre o 1° e 2° anos do curso ou entre o 3° e 4°, mas com recuperação significativa após. O maior crescimento pode ser observado entre as provas do 2° e 3° anos.

Figura 18 - Desempenho da Coorte de 2011 nas questões de Pediatria

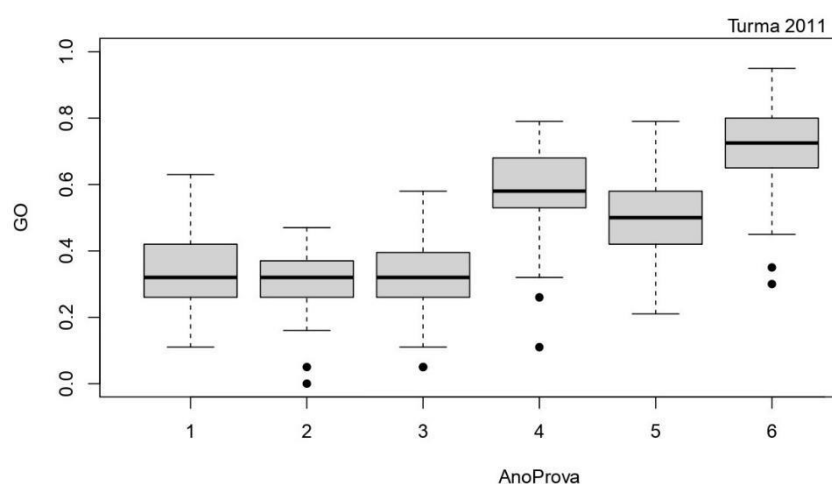


Legenda: P: área de Pediatria da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.2.6. Avaliação de resultados da área de Ginecologia e Obstetrícia

O desempenho da coorte de 2011 nas questões de GO houve crescimento geral positivo ao longo dos anos (figura 19). Apesar da queda entre o 1º e 2º ano, houve crescimento significativo a partir do 3º ano. O maior crescimento pode ser observado entre as provas do 3º e 4º anos, mas houve uma queda significativa também entre o 4º e o 5º ano.

Figura 19 - Desempenho da Coorte de 2011 nas questões de Ginecologia e Obstetrícia



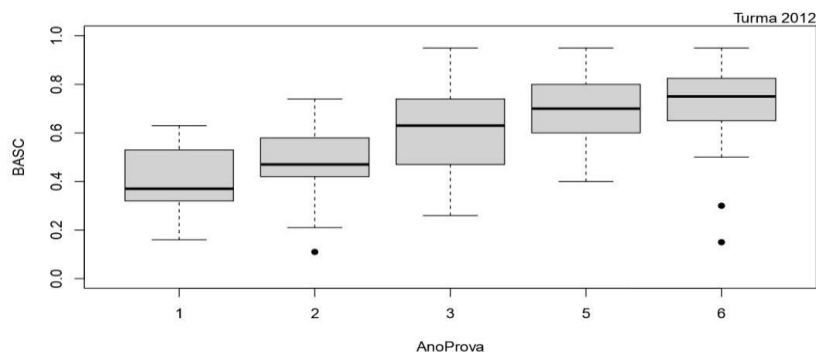
Legenda: GO: área de Ginecologia e Obstetrícia prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

### 4.6.3. Descrição dos Dados da Coorte de 2012

#### 4.6.3.1. Avaliação de resultados da área Básica

Ao observar o desempenho da coorte de 2012 nas questões da área básica, também é possível observar um crescimento positivo ao longo dos anos (figura 20). No 4º ano dessa coorte não houve área básica na prova, pois foi uma prova unificada pela ABEM. Na tabela de comparações, é possível perceber que em quase todos os anos houve crescimento significativo a cada ano, menos entre o 5º e 6º ano, porém não houve queda no resultado.

Figura 20 – Desempenho da Coorte de 2012 nas questões da Área Básica

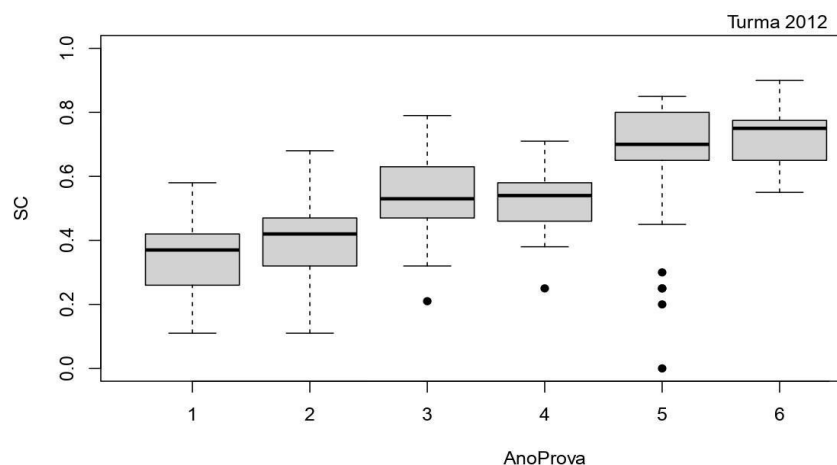


Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.3.2. Avaliação de resultados da área de Saúde Coletiva

O desempenho da coorte de 2012 nas questões de saúde coletiva, é possível observar um crescimento positivo ao longo dos anos (figura 21), com maior crescimento entre o 4º e 5º anos. Não houve crescimento entre as provas do 3º e 4º anos, mas é possível perceber, na tabela de comparações, crescimento significativo quando se considera o início e término do curso.

Figura 21 – Desempenho da Coorte de 2012 nas questões de Saúde Coletiva

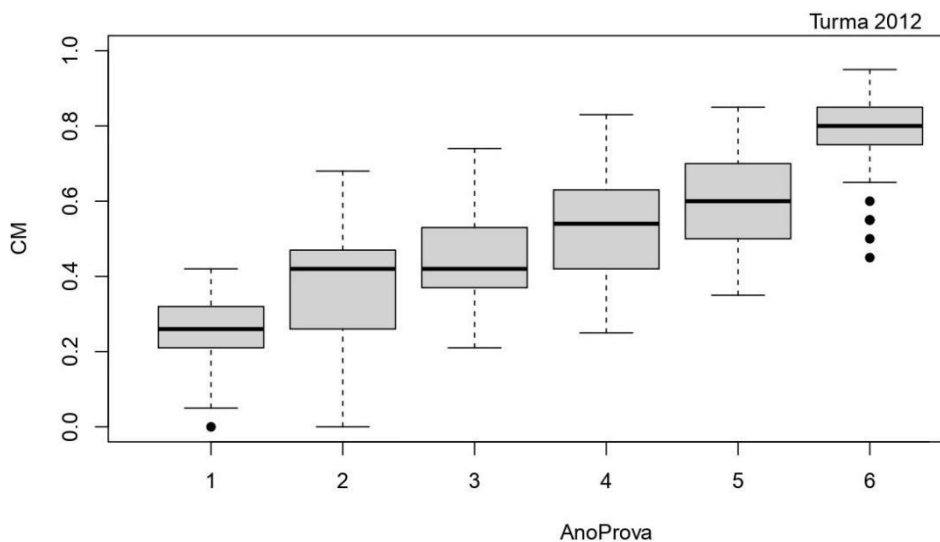


Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.3.3. Avaliação de resultados da área de Clínica Médica

No desempenho da coorte de 2012 nas questões de clínica médica houve crescimento geral positivo ao longo dos anos (figura 22). Houve crescimento entre todos os anos de prova ( $p < 0,001$ ). O maior crescimento pode ser observado entre as provas do 5º e 6º anos.

Figura 22- Desempenho da Coorte de 2012 nas questões de Clínica Médica

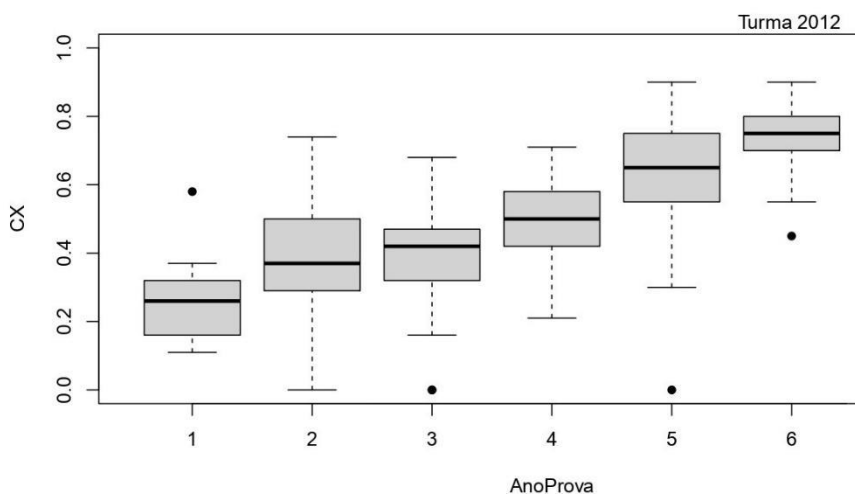


Legenda: CM: área de Clínica Médica na prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.3.4. Avaliação de resultados da área de Cirurgia

No desempenho da coorte de 2012 nas questões de cirurgia houve crescimento geral positivo ao longo dos anos (Figura 23), sem crescimento significativo entre as provas do 2º e 3º anos, mas com recuperação significativa após. O maior crescimento pode ser observado entre as provas do 4º e 5º anos.

Figura 23- Desempenho da Coorte de 2012 nas questões de Cirurgia

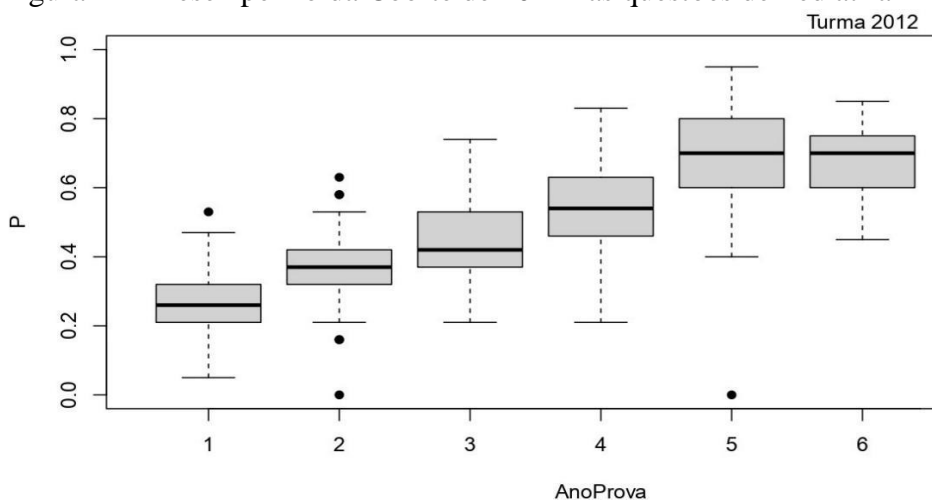


Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada ponto: indica *outlier*

#### 4.6.3.5. Avaliação de resultados da área de Pediatria

No desempenho da coorte de 2012 nas questões de pediatria houve crescimento geral positivo ao longo dos anos (figura 24), sem crescimento entre os resultados do 5º e 6º anos apenas. O maior crescimento pode ser observado entre as provas do 4º e 5º anos.

Figura 24 - Desempenho da Coorte de 2012 nas questões de Pediatria

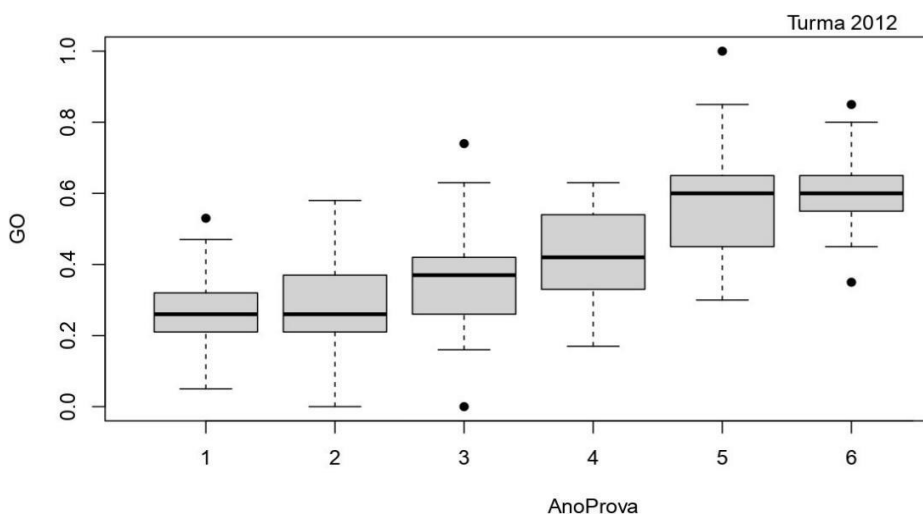


Legenda: P: área de Pediatria da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.3.6. Avaliação de resultados da área de Ginecologia e Obstetrícia

No desempenho da Coorte de 2012 nas questões de GO houve crescimento geral positivo ao longo dos anos (figura 25), sem crescimento significativo entre as 2 primeiras provas ou entre os resultados do 5º e 6º anos. O maior crescimento pode ser observado entre as provas do 3º e 4º anos.

Figura 25 - Desempenho da Coorte de 2012 nas questões de Ginecologia e Obstetrícia



Legenda: GO: área de ginecologia e obstetrícia prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

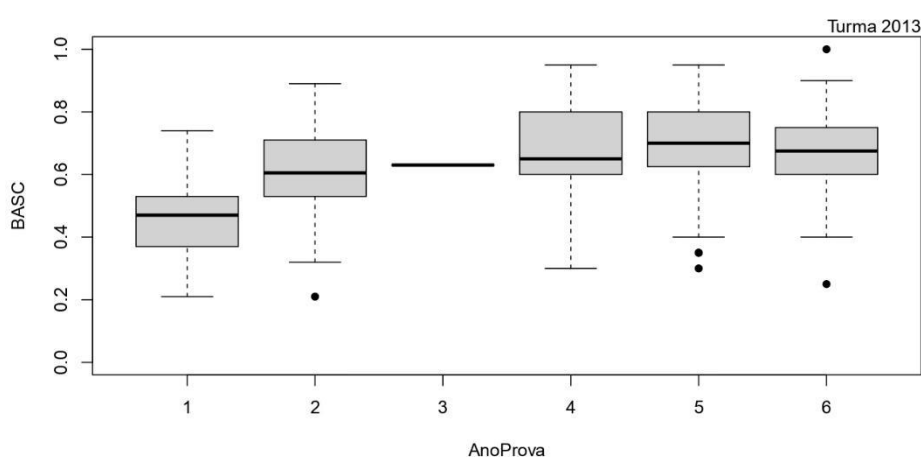


#### 4.6.4. Descrição dos Dados da Coorte de 2013

##### 4.6.4.1. Avaliação de resultados da área Básica

Ao observar o desempenho da coorte de 2013 nas questões da área básica, também é possível observar um crescimento positivo ao longo dos anos (figura 26). No 3º ano dessa coorte não houve área básica na prova, pois foi uma prova unificada pela ABEM. Em quase todos os anos houve crescimento significativo a cada ano, menos entre o 5º e 6º ano, porém não houve queda no resultado.

Figura 26 – Desempenho da Coorte de 2013 nas questões da Área Básica

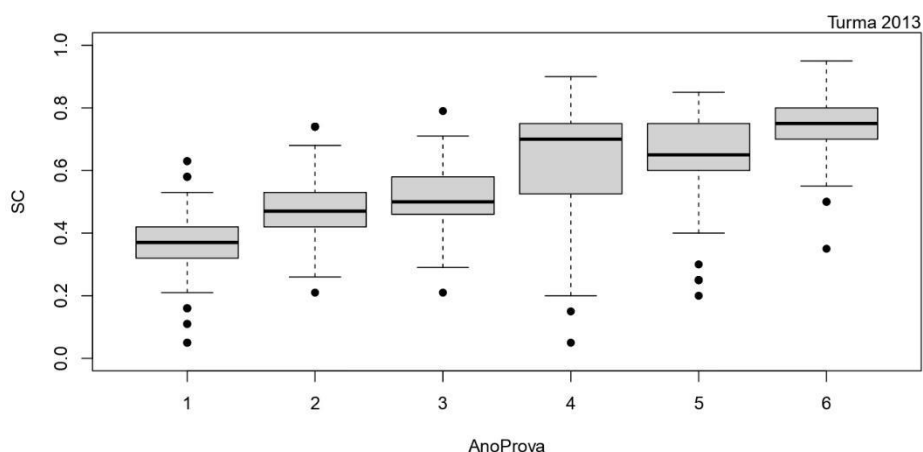


Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

##### 4.6.4.2. Avaliação de resultados da área de Saúde Coletiva

Ao observar o desempenho da coorte de 2013 nas questões de saúde coletiva, também é possível observar um crescimento global significativo ao longo dos anos, com maior salto de conhecimento entre o 3º e 4º anos (figura 27). Houve leve queda dos resultados entre o 2º e 3º anos e não houve crescimento entre o 3º e 4º anos, mas houve ganho significativo após e quando se considera o crescimento global da coorte.

Figura 27 – Desempenho da Coorte de 2013 nas questões de Saúde Coletiva

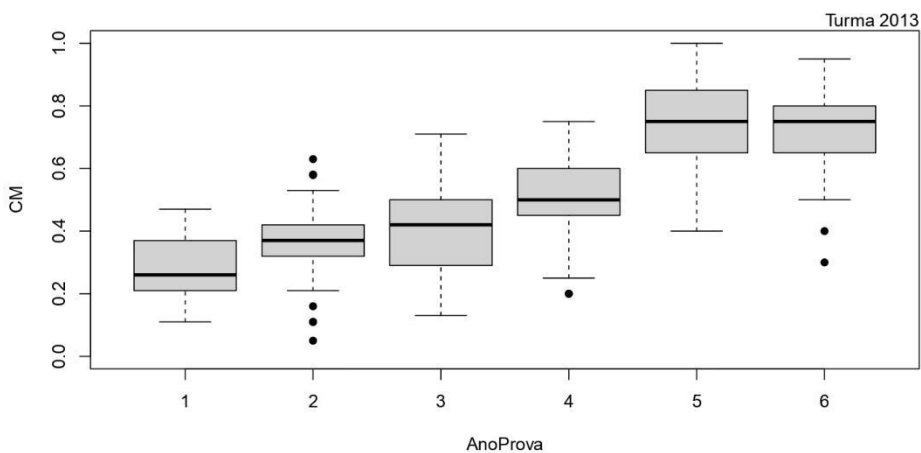


Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica outlier

#### 4.6.4.3. Avaliação de resultados da área de Clínica Médica

No desempenho da coorte de 2013 nas questões de clínica médica houve crescimento geral positivo ao longo dos anos (figura 28), sem crescimento significativo entre as provas do 2 e 3º ano, mas com crescimento significativo após e leve queda de desempenho entre o resultado das provas do 5 e 6º. O maior crescimento pode ser observado entre as provas do 4º e 5º anos.

Figura 28- Desempenho da Coorte de 2013 nas questões de Clínica Médica

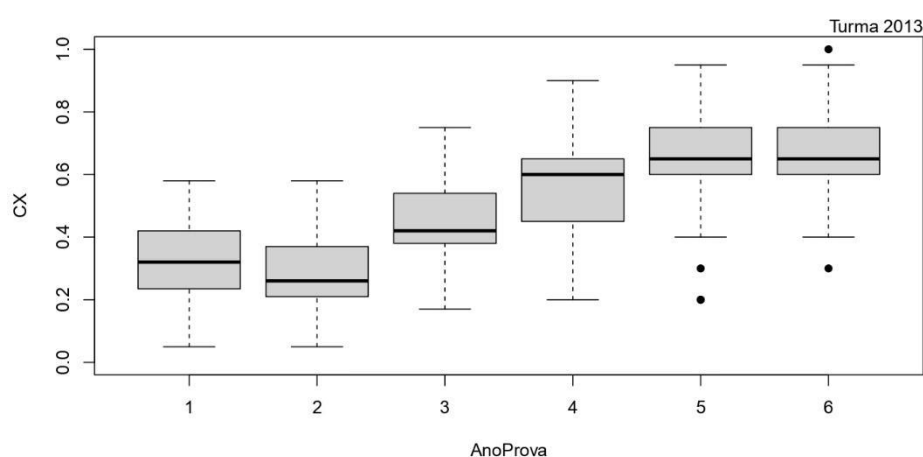


Legenda: CM: área de clínica médica da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica outlier

#### 4.6.4.4. Avaliação de resultados da área de Cirurgia

No desempenho da coorte de 2013 nas questões de cirurgia houve crescimento geral positivo ao longo dos anos (figura 29), sem crescimento significativo entre os resultados das duas primeiras provas da coorte e entre o 5º e 6º anos. O maior crescimento pode ser observado entre as provas do 2º e 3º anos.

Figura 29- Desempenho da Coorte de 2013 nas questões de Cirurgia

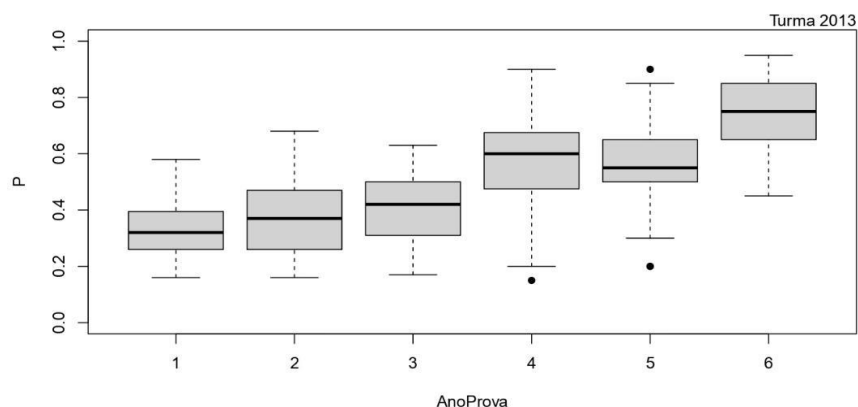


Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.4.5. Avaliação de resultados da área de Pediatria

No desempenho da coorte de 2013 nas questões de pediatria houve crescimento geral positivo ao longo dos anos (figura 30), sem crescimento entre o 4º e 5º ano, mas com recuperação significativa após. O maior crescimento pode ser observado entre as provas do 5º e 6º anos.

Figura 30 - Desempenho da Coorte de 2013 nas questões de Pediatria

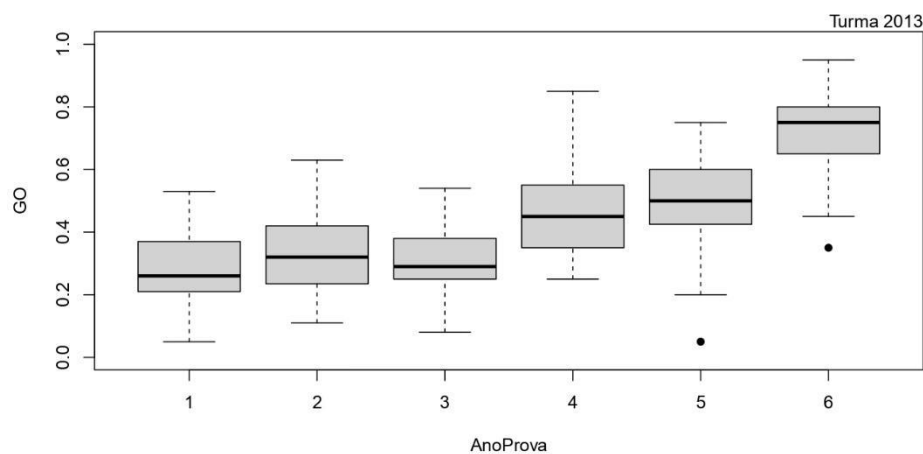


Legenda: P: área de Pediatria da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.4.6. Avaliação de resultados da área de Ginecologia e Obstetrícia

No desempenho da coorte de 2013 nas questões de GO houve crescimento geral positivo ao longo dos anos (figura 31), significativo entre todas os anos. O maior crescimento pode ser observado entre as provas do 5º e 6º anos.

Figura 31 - Desempenho da Coorte de 2013 nas questões de Ginecologia e Obstetrícia



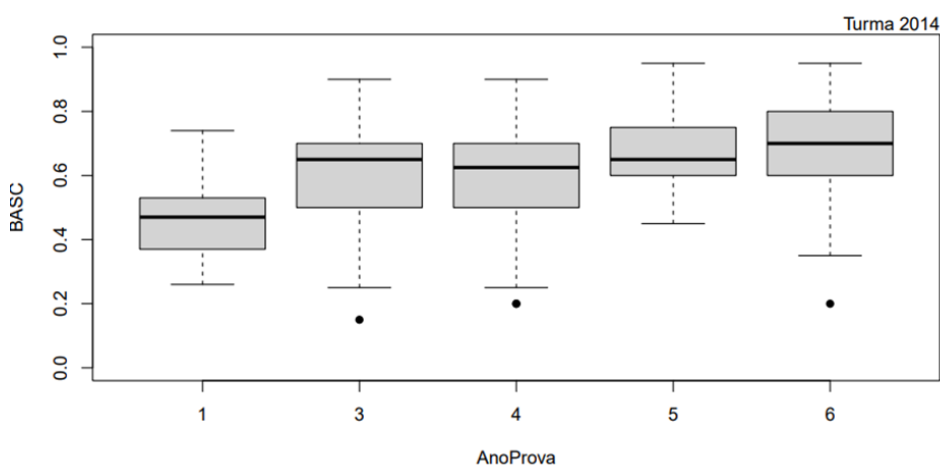
Legenda: GO: área de ginecologia e obstetrícia prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.5. Descrição dos Dados da Coorte de 2014

##### 4.6.5.1. Avaliação de resultados da área Básica

Ao observar o desempenho da coorte de 2014 nas questões da área básica, também é possível observar um crescimento positivo ao longo dos anos (figura 32). No 2º ano dessa coorte não houve área básica na prova, pois foi uma prova unificada pela ABEM. Na tabela de comparações, é possível perceber que em quase todos os anos houve crescimento significativo a cada ano salvo entre 3º e 4º ano, com leve queda da média, e entre o 5º e 6º ano.

Figura 32 – Desempenho da Coorte de 2014 nas questões da Área Básica

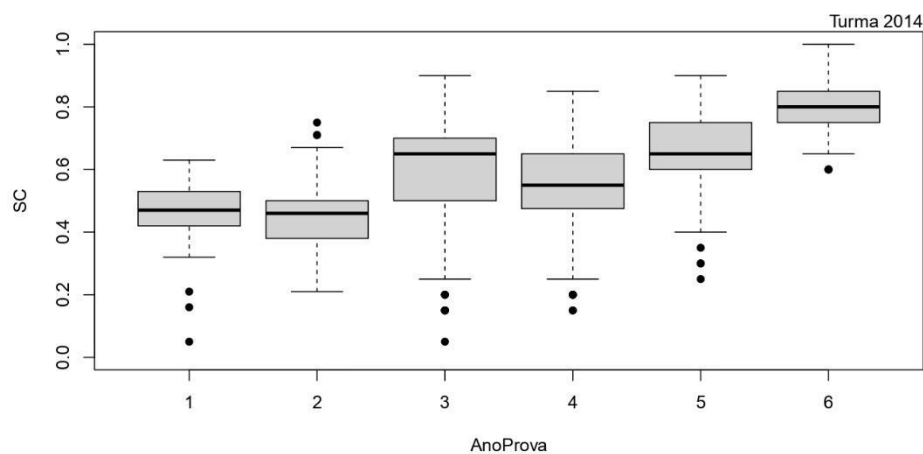


Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

##### 4.6.5.2. Avaliação de resultados da área de Saúde Coletiva

No desempenho da coorte de 2014 nas questões de Saúde Coletiva houve crescimento geral positivo ao longo dos anos (figura 33), significativo a partir do segundo ano. O maior crescimento pode ser observado entre os desempenhos do 5º e 6º anos.

Figura 33 – Desempenho da Coorte de 2014 nas questões de Saúde Coletiva

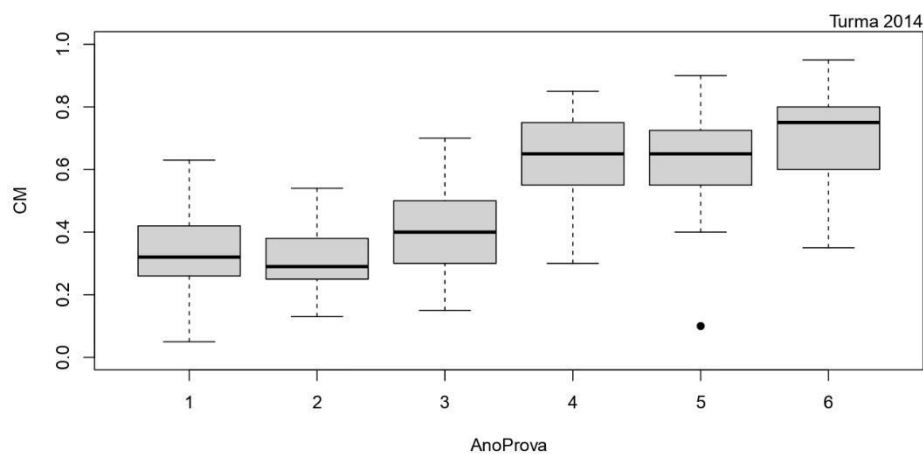


Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica outlier

#### 4.6.5.3. Avaliação de resultados da área de Clínica Médica

No desempenho da coorte de 2014 nas questões de Clínica Médica houve crescimento geral positivo ao longo dos anos (figura 34), significativo a partir da segunda prova, com exceção entre o 4º e 5º ano. O maior crescimento pode ser observado entre os resultados do 3º e 4º anos.

Figura 34 - Desempenho da Coorte de 2014 nas questões de Clínica Médica

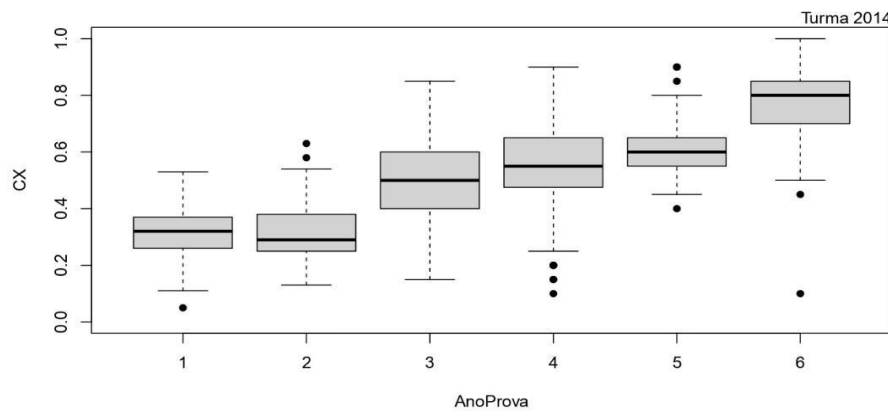


Legenda: CM: área de clínica médica da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica outlier

#### 4.6.5.4. Avaliação de resultados da área de Cirurgia

No desempenho da coorte de 2014 nas questões de Cirurgia houve crescimento geral positivo ao longo dos anos (figura 35), significativo a partir da segunda prova. O maior crescimento pode ser observado entre as provas do 2º e 3º anos.

Figura 35- Desempenho da Coorte de 2014 nas questões de Cirurgia

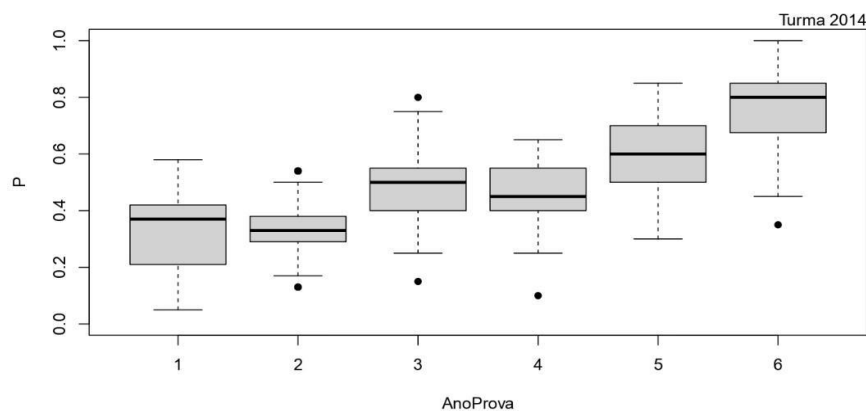


Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.5.5. Avaliação de resultados da área de Pediatria

No desempenho da coorte de 2014 nas questões de Pediatria houve crescimento geral positivo ao longo dos anos (figura 36), significativo a partir da segunda prova, com exceção de entre o 3º e 4º anos. O maior crescimento pode ser observado entre as provas do 5º e 6º anos.

Figura 36 - Desempenho da Coorte de 2014 nas questões de Pediatria

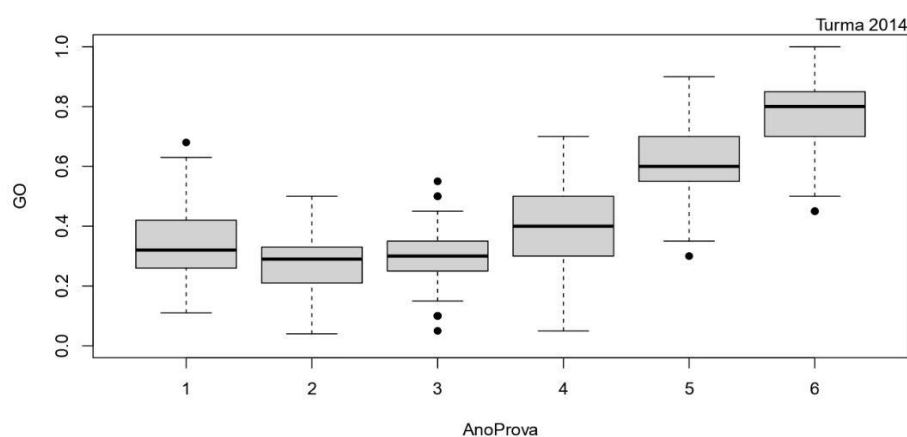


Legenda: P: área de Pediatria da prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.6.5.6. Avaliação de resultados da área de Ginecologia e Obstetrícia

No desempenho da coorte de 2014 nas questões de GO houve crescimento geral positivo ao longo dos anos (figura 37), com queda entre os resultados do 1º e 2º ano, e crescimento significativo apenas a partir da terceira prova. O maior crescimento pode ser observado entre as provas do 4º e 5º anos.

Figura 37 - Desempenho da Coorte de 2014 nas questões de Ginecologia e Obstetrícia



Legenda: GO: área de ginecologia e obstetrícia prova; AnoProva: ano do curso em que a prova foi realizada; ponto: indica *outlier*

#### 4.7. Comparação de resultados entre as Coortes

Ao comparar os resultados gerais das provas intercoortes, a maioria das provas apresenta resultados não equivalentes (com diferença estatística) nos dois primeiros terços do curso (Apêndice 4). Por outro lado, nas provas do 5º e 6º anos as diferenças tornam-se menos acentuadas. No 5º ano, metade das coortes não apresentaram diferença significativa entre seus resultados e não apresentaram diferença significativa entre si. Nas provas feitas pelos 6ºs anos, apenas o resultado da coorte de 2014 apresentou diferença de desempenho em relação às demais.

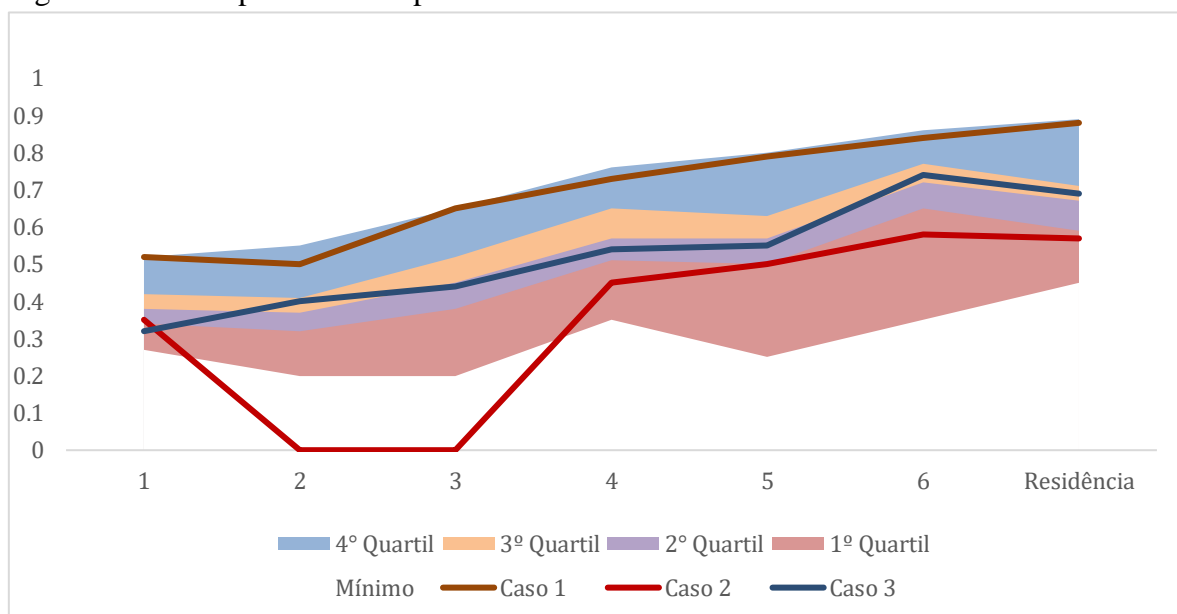
#### 4.8. Relatos de Caso de Estudantes

Na figura 38 foi elaborado um exemplo de progressão de desempenho de estudantes da coorte de 2011, baseado nos quartis que ocupavam em sua coorte. No caso 1 apresenta-se um estudante que desde o início da graduação esteve nos melhores quartis de sua coorte e manteve esses resultados ao fim da graduação e na prova de residência. No caso 2 apresentamos um



estudante que no início da graduação estava com classificação mais baixa, com faltas e manteve-se nos piores resultados ao fim da graduação e residências. No caso 3 apresentamos um estudante que no início da graduação estava em um quartil baixo, mas que melhorou seu desempenho ao fim do curso.

Figura 38 – Exemplo de Desempenho de Estudantes

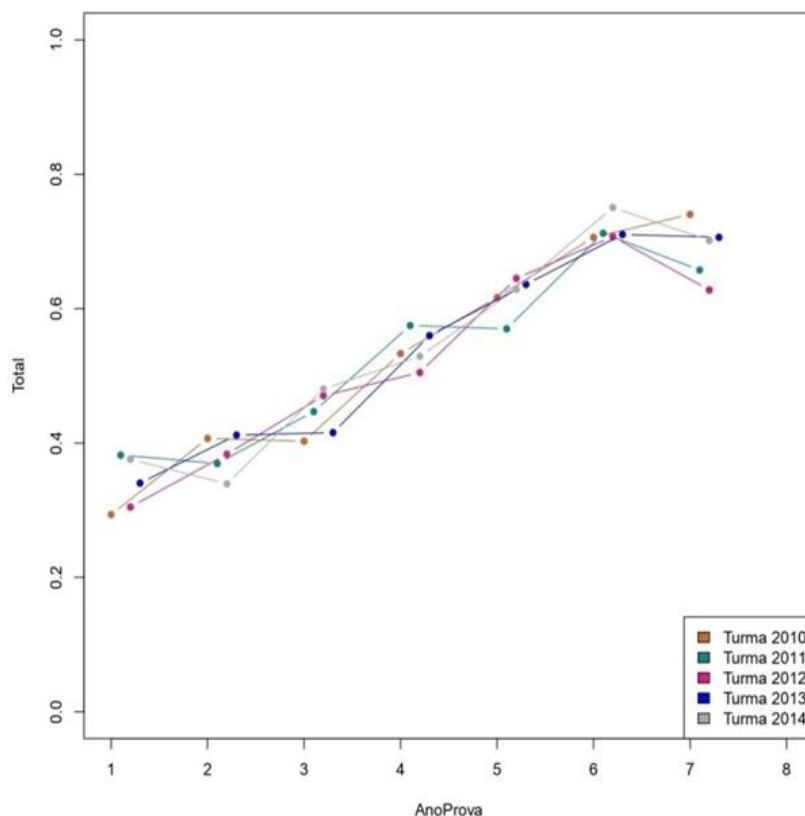


## 4.9. Relação do teste com a prova de Residência

### 4.9.1. Comparação de crescimento cognitivo entre Provas do TP e Prova de Residência Médica

As notas médias da prova de residência médica da instituição foram menores em todas as coortes de estudantes a partir de 2011 (figura 39), e a diferença entre as médias dos resultados da prova do TP do 6º ano e da prova de residência foi pequena, porém variada entre as Coortes. (2010: 3,4%; 2011: 5,4%; 2012: 7,8%; 2013: 0,42%; 2014: 4,9%).

Figura 39– Representação de Desempenho de Coortes de estudantes ao longo de seu curso e com resultados de Residência Médica



Legenda: Total: acerto geral das provas, AnoProva: ano do curso

Há diferença significativa entre os resultados das provas do 6º ano do TP e da prova de residência em todas as coortes menos na de 2013 ( $p = 0,1927$ ), indicando que apesar de próximos, os resultados das provas não podem ser considerados equivalentes.

Ainda, observa-se diferença significativa da pontuação média da prova de residência entre a maioria dos resultados ( $p < 0,001$ ), exceto nas comparações entre 2011-2012 e 2013-2014 (tabela 9).

Tabela 9 - Comparação da média das provas de residência por coorte

Grande Área	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
	Coorte	Coorte			Limite Inferior	Limite Superior
<b>Total</b>	2010	2011	0,0839	0,0001	0,0581	0,1097
<b>Total</b>	2010	2012	0,1071	0,0001	0,0797	0,1346
<b>Total</b>	2010	2013	0,0412	0,0016	0,0156	0,0668

---

<b>Total</b>	2010	2014	0,0416	0,0018	0,0155	0,0677
<b>Total</b>	2011	2012	0,0232	0,107	-0,0050	0,0514
<b>Total</b>	2011	2013	-0,0427	0,0015	-0,0691	-0,0163
<b>Total</b>	2011	2014	-0,0423	0,002	-0,0692	-0,0154
<b>Total</b>	2012	2013	-0,0659	0,0001	-0,0939	-0,0378
<b>Total</b>	2012	2014	-0,0655	0,0001	-0,0941	-0,0370
<b>Total</b>	2013	2014	0,0004	0,9795	-0,0264	0,0271

---

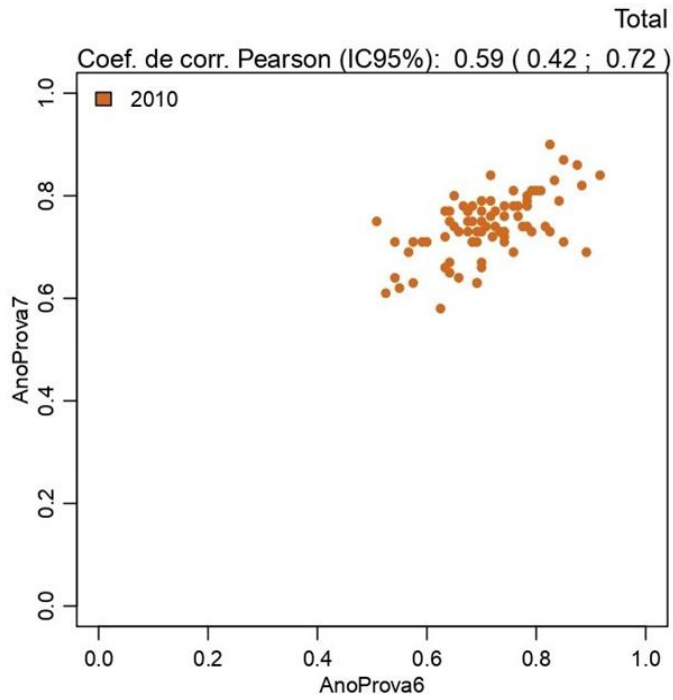
#### 4.9.2. Correlação entre resultados do TP e da Prova de Residência

Comparado o resultado entre as provas do 5° e 6° anos do TP (provas 5 e 6 respectivamente) com a prova de residência, de maneira geral, é possível observar uma correlação moderada entre as provas.

Quando comparado resultado da prova de residência com o resultado da prova realizada no 5° ano, manteve-se entre todas as coortes correlação moderada entre os resultados das provas do TP com resultados melhores nos anos 2010 (0,57), 2011 (0,62) e 2012 (0,59) e piores nos anos de 2013 (0,5) e 2014 (0,4).

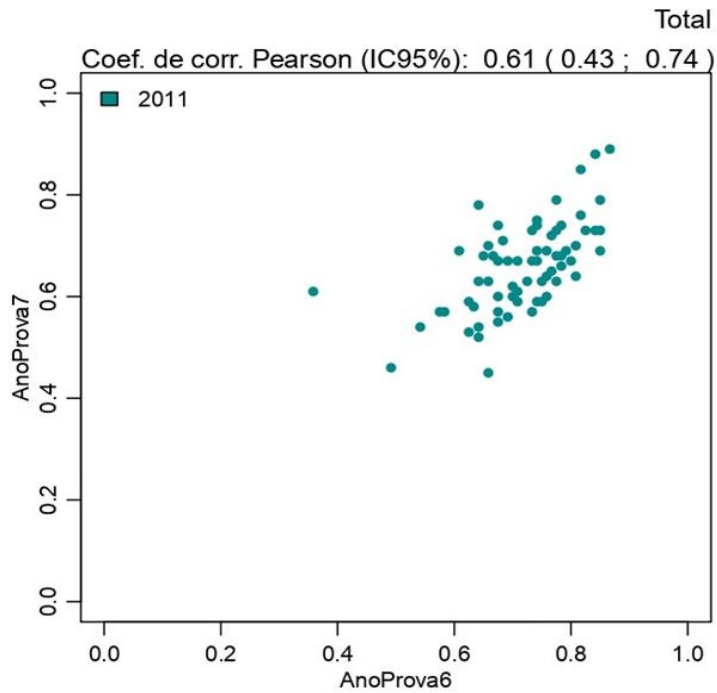
Houve moderada correlação entre a nota da prova de residência a nota geral do TP realizado no 6° ano em todas as coortes: 2011 (0,61), 2012 (0,62), 2013 (0,65) e piores nos anos de 2010 (0,59) e 2014 (0,45). O estudo dessa correlação com todas as coortes agrupadas mantém a classificação como moderada, contudo elevando o coeficiente para 0,52 nas provas do 6° ano (figuras 40 a 45)

Figura 40 - Correlação do TP do 6º ano com prova de residência na coorte de 2010



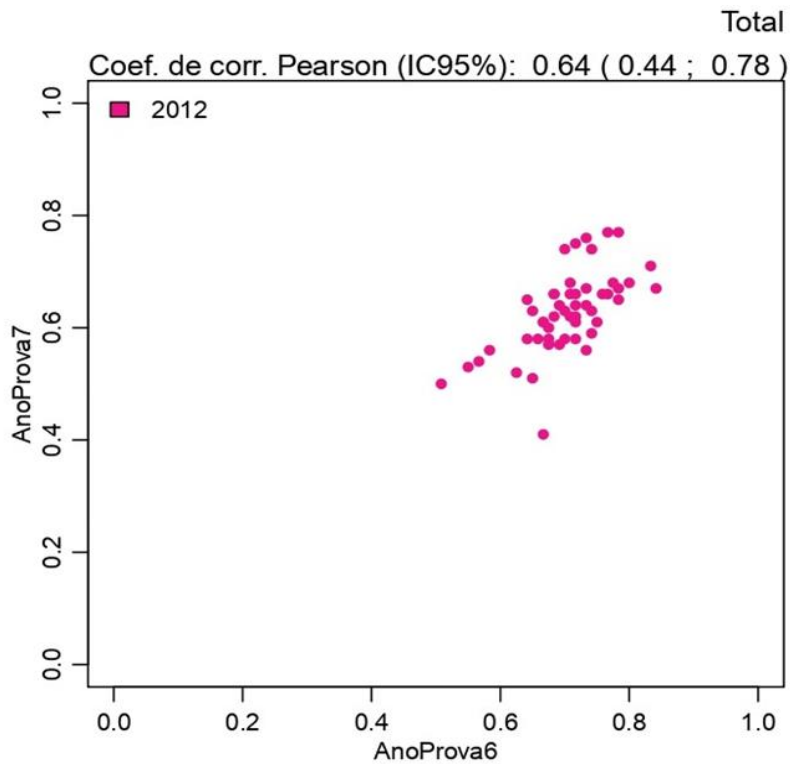
AnoProva7 = prova de residência; AnoProva6 = prova de teste de progresso do 6º ano

Figura 41 - Correlação do TP do 6º ano com prova de residência na coorte de 2011



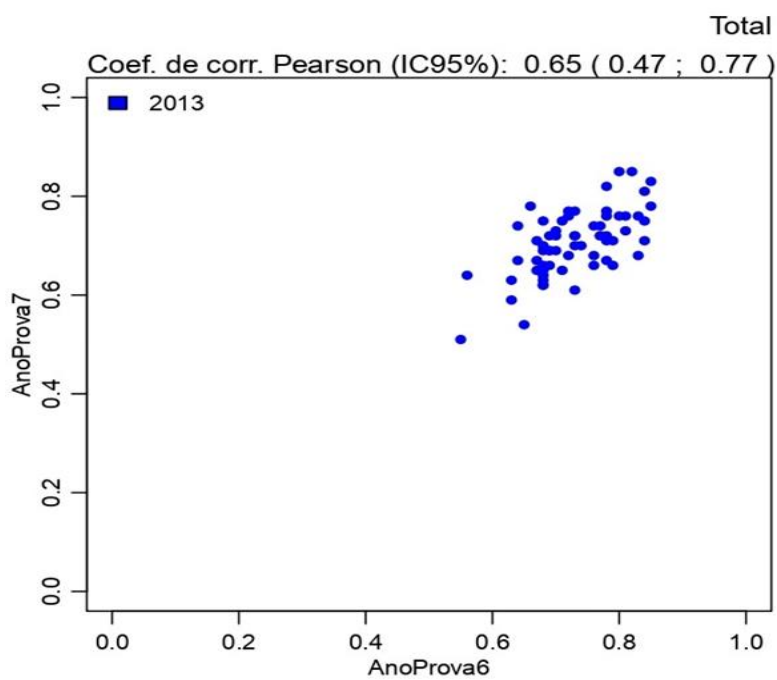
AnoProva7 = prova de residência; AnoProva6 = prova de teste de progresso do 6º ano

Figura 42 - Correlação do TP do 6º ano com prova de residência na coorte de 2012



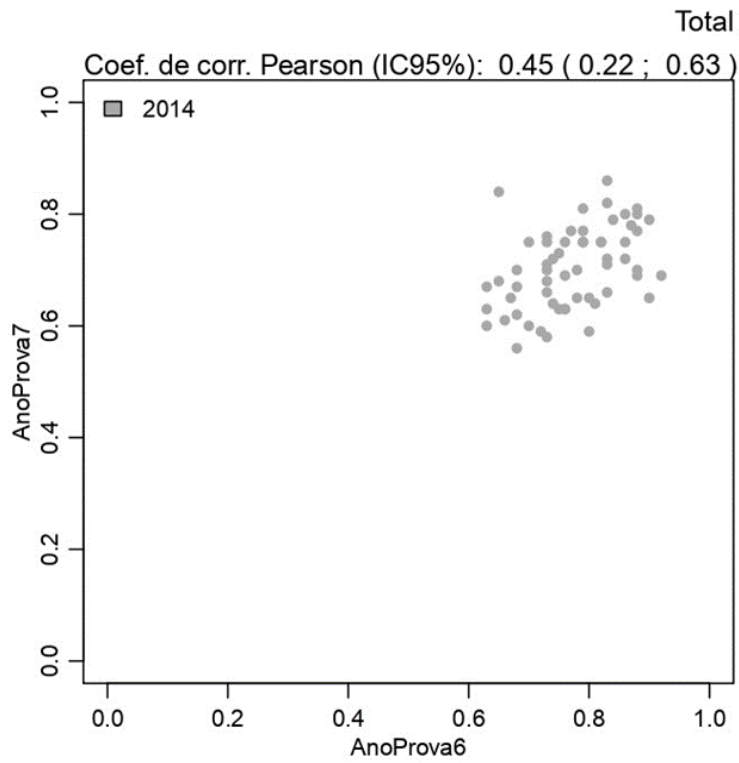
AnoProva7 = prova de residência; AnoProva6 = prova de teste de progresso do 6º ano

Figura 43 - Correlação do TP do 6º ano com prova de residência na coorte de 2013



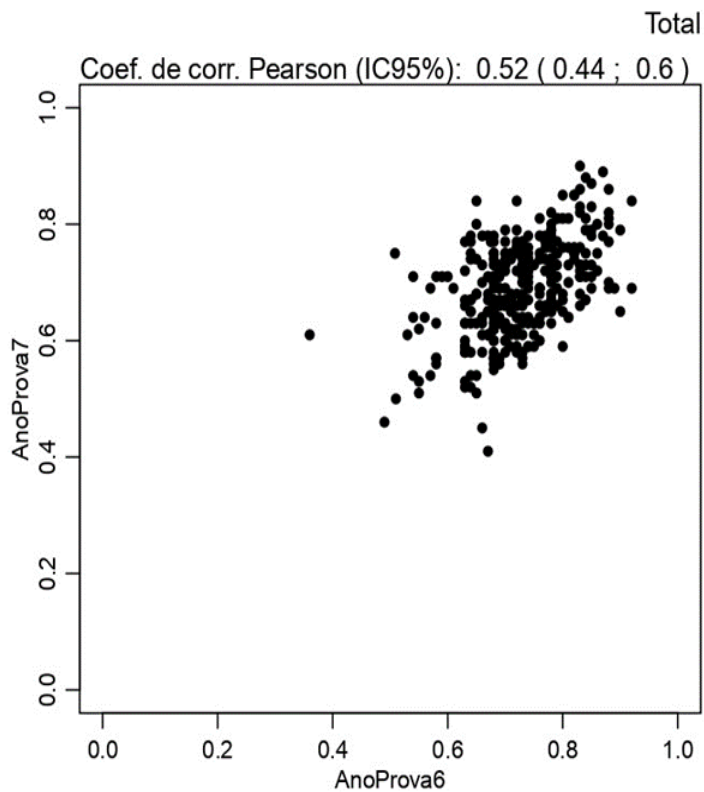
AnoProva7 = prova de residência; AnoProva6 = prova de teste de progresso do 6º ano

Figura 44 - Correlação do TP do 6º ano com prova de residência na coorte de 2014



AnoProva7 = prova de residência; AnoProva6 = prova de teste de progresso do 6º ano

Figura 45 - Correlação do TP do 6º ano com prova de residência de todas as coortes



AnoProva7 = prova de residência; AnoProva6 = prova de teste de progresso do 6º ano

### 4.9.3. Avaliação por quartis

Os alunos foram categorizados em quartis conforme o desempenho na prova da sua própria coorte (tabela 10)

Tabela 10 – Taxas de Acerto e Distribuição por Quartis do TP e da prova de residência

Coorte	Ano Prova	Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo	Desvio Padrão	n
<b>2010</b>	6°	0,38	0,64	0,7	0,70	0,77	0,92	0,09	84
<b>2010</b>	residência	0,58	0,71	0,74	0,74	0,78	0,90	0,06	79
<b>2011</b>	6°	0,36	0,66	0,73	0,71	0,78	0,87	0,09	78
<b>2011</b>	residência	0,45	0,59	0,67	0,65	0,71	0,89	0,08	68
<b>2012</b>	6°	0,51	0,68	0,71	0,70	0,74	0,84	0,06	67
<b>2012</b>	residência	0,41	0,58	0,63	0,62	0,66	0,77	0,07	51
<b>2013</b>	6°	0,43	0,67	0,71	0,71	0,77	0,85	0,08	90
<b>2013</b>	residência	0,51	0,66	0,71	0,70	0,75	0,85	0,06	66
<b>2014</b>	6°	0,5	0,69	0,75	0,75	0,80	0,92	0,08	88
<b>2014</b>	residência	0,56	0,64	0,7	0,70	0,75	0,86	0,07	60

Na tabela é possível perceber que os valores de corte para o 1° quartil são menores que na de residência, com exceção da coorte de 2010. No caso da mediana, os valores da prova da residência são menores que os da prova do 6° ano, salvo nas coortes de 2010 onde é maior e na 2013 onde é igual. Nos valores de corte para o 3° quartil, apenas a Coorte de 2010 apresentou um valor de corte maior na prova do 6° ano.

Tabela 11 - Movimentação de estudantes entre quartis

Movimento	Total	2010	2011	2012	2013	2014
	<i>n=320 (%)</i>	<i>n=79 (%)</i>	<i>n=67 (%)</i>	<i>n=51 (%)</i>	<i>n=64 (%)</i>	<i>n=59 (%)</i>
Igual	138 (43,1)	37 (46,8)	31 (46,3)	23 (45,1)	27 (42,2)	20 (33,9)
Melhorou	80 (25)	18 (22,8)	17 (25,4)	16 (31,4)	14 (21,9)	15 (25,4)
Piorou	102 (31,9)	24 (30,4)	19 (28,4)	12 (23,5)	23 (35,9)	24 (40,7)

Foi possível observar que pelo menos metade dos estudantes mantem-se no mesmo quartil na prova do teste do progresso do 6° ano quando comparado com o quartil baseado na prova de residência. Ainda, que o número de estudantes que melhoram ou pioram seu desempenho é próximo, entre um quarto e um terço da Coorte.

Outro ponto a ser destacado é que na maioria dos grupos, com exceção da coorte de 2014, os estudantes do primeiro ou último quartil tendem a se manter neles, isto é, os estudantes que têm melhores ou pior resultados no TP tendem a ter resultados similares na prova de residência. (Uma tabela mais detalhada pode ser observada no apêndice 5). Apenas no ano de 2014 menos de 50% dos estudantes tiveram esse padrão de movimentação. Nos 2° e 3° quartis há maior variedade de movimentação, com sua permanência na prova seguinte ficando entre 20-40% a depender da Coorte.

Em relação aos estudantes do pior quartil, foi raro que passassem para o melhor quartil na prova de residência. Nos outros quartis, quando há melhora ou piora dos resultados, há tendência de que o estudante suba ou desça 1 quartil, sendo menos comum, ainda que presente, que suba ou desça 2 quartis entre as provas.



## 5. DISCUSSÃO

### 5.1. Explorando a Experiência da FMRP-USP

A revisão da implementação do TP na FMRP-USP mostra que iniciar uma avaliação longitudinal pode ser demorado e oneroso. Apesar de discutida desde os anos 90, a primeira aplicação na FMRP-USP ocorreu no início dos anos 2000. Sendo que sua sedimentação como processo avaliativo aconteceu quase 10 anos após as discussões se iniciarem.

Diversos aspectos podem ser pontuados e relacionados com esse período de ajuste antes do início da aplicação do TP, que se relacionaram com a busca por garantir uma boa organização e entendimento do propósito do teste. Foi necessário a articulação de interesses de diferentes instâncias institucionais da FMRP-USP: departamentos, CG, CoC e Diretoria. Coletivamente, foi necessário alinhar e compartilhar interesses com outras instituições participantes do consórcio, considerando que o mesmo processo interno descrito acima repetiu-se nas demais instituições.

Sabe-se da necessidade de desenvolvimento em diversos aspectos para garantir a qualidade do TP (17), que envolvem questões operacionais e logísticas: data e forma de aplicação, sigilo, correção, processamento dos resultados, tipo de devolutiva para os alunos; questões de conteúdo: concepção do teste enquanto temas, matriz de conteúdo, tipo e forma de questão.

Além disso, foi necessária uma preparação orçamentária para a implantação do teste, uma vez que é um processo potencialmente oneroso com diversas fases. Existe um custo implícito relacionado às horas de trabalho docente e administrativo, que foi aplicada à concepção global do projeto, pois houve a necessidade da disposição de tempo para reuniões de colegiados das instituições para discussão da proposta, além de reuniões de representantes das instituições para formação de um consórcio e reuniões de organização final da aplicação.

Como custos externos fez-se necessária a remuneração da consultoria de análise de dados (PRODATA) e de despesas relacionadas à aplicação do TP: impressão, transporte, fiscais de prova, lanches etc. Mas a manutenção do teste como ferramenta avaliativa em múltiplos consórcios ao longo do tempo aponta para a viabilidade de sua implementação e continuação.

Além disso, o desenvolvimento do teste de progresso ao longo do tempo na FMRP-USP foi adquirindo gradualmente características essenciais para aumentar sua qualidade com melhoramentos documentados no campo de construção do TP a fim de aumentar sua qualidade externa. Destaca-se a elaboração de matriz de competências e o *blueprint* da prova, pois essas

estratégias garantem parâmetros tanto para o formato, quanto para o conteúdo da prova que estabelecem maior confiabilidade à prova.

A matriz de competências estratifica a relevância de cada tema de saúde para o médico generalista e cruza essa informação com a frequência epidemiológica e nível de atenção, buscando englobar na prova todos os temas essenciais, alinhados com o perfil do projeto político pedagógico e as DCNs. Isso faz com que a prova não tenha o viés de elaboração de especialistas que tendem a focar em temas relevantes para sua área, porém frequentemente muito específicos, que não atendem a formação desejada.

Outro ponto a ser destacado é a participação no consórcio, que permite uma avaliação externa à instituição. Isso oferece uma melhor perspectiva da qualidade do ensino da escola. Da mesma forma que a matriz de competências, uma avaliação que não depende apenas de um colegiado permite que as questões não adquiram vieses de um projeto curricular específico, aumentando a validade e equivalência do teste. Assim elas tendem a atender as demandas dos cursos de medicina no geral, baseado nas DCNs e permite uma avaliação menos tendenciosa.

A elaboração do TP em consórcio de escolas também permite um ajuste entre o desequilíbrio de forças internamente nas instituições. Cada escola tem particularidades quanto ao perfil epidemiológico de suas regiões, o que reflete na matriz de competências das questões. A presença ou associação a centros de pesquisa, assim como a quantidade de médicos especialistas ou pesquisadores na docência, podem criar tendências curriculares e influenciar no conteúdo ensinado, levando a assimetrias na carga horária entre instituições.

A maturidade do curso e experiência da faculdade também influenciam na qualidade e formato de provas realizadas, tanto positivamente pensando em professores mais treinados para elaborar questões e, sistemas de aplicação seriada já instaurados, quanto negativamente, pois cursos mais bem estabelecidos tendem a ser menos flexíveis para mudanças. Isso pode ser percebido pelo relato de que a associação paulista de escolas médicas não era a favor da implantação do TP nas escolas médicas do estado, especialmente por não concordar com a introdução de uma nova técnica de avaliação.

Ainda, ao conceber diretrizes de elaboração de questões de múltipla escolha (*'blueprint'*) o núcleo de elaboração de questões do consórcio solicitava aos professores não só a questão com a resposta, mas a fonte da informação, assim as questões já eram enviadas em um formato mais homogêneo e confiável, que facilitava a avaliação (14). Além de garantir que as questões não sejam elaboradas a partir de práticas assistenciais regionais ou específicas dos serviços que elaboraram as questões, a busca pelas referências bibliográficas e condutas

embasadas na literatura reduz a ocorrência de contestação do gabarito da prova.

Um efeito dessas práticas em relação a busca de qualidade na elaboração do TP refletiu-se nas recomendações para elaboração de questões em outros processos de avaliação: os testes somativos presentes ao longo do curso de Medicina e na prova de Residência Médica do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCRP) também passaram a ter diretrizes para elaboração de questões. Isso levou a melhora da validade e equivalência também dessas provas, e aponta para o efeito impulsionador e catalítico do TP nos ambientes avaliativos, ao guiar processos de inovação que buscam por qualidade.

Em relação à qualidade interna do teste, também houve registro de melhora dos parâmetros no TP ao longo do tempo, o que pode ser percebido pela melhora na distribuição de dificuldade entre 2011 e 2014 (figura 4), além da redução das questões deficientes ao longo dos anos (figura 5). Quanto à consistência do teste, a análise psicométrica das provas dentro de um mesmo consórcio apresentou variações na dificuldade e discriminação das questões ao longo dos anos, com provas bem mais discriminativas que outras, apesar das medidas tomadas para que ela se mantenha homogênea ao longo do tempo. Sabe-se que questões de múltipla escolha podem ter baixa consistência e uniformidade e itens de teste falhos podem diminuir pontuações, por isso sempre necessitam de construção rigorosa, por isso se entende que a elaboração de questões é sempre um processo que pode ser aprimorado (47).

Apesar disso, destaca-se que nenhuma prova teve menos que metade das questões consideradas boas ou regulares, e pelo menos 65% das questões das provas de todas as coortes se encontravam entre “fácil” e “média”, quando se avaliava dificuldade. Em 2012 foi percebido uma mudança qualitativa importante na prova: as áreas passaram a ter proporções semelhantes de questões com melhor capacidade discriminativa e as proporções de questões também ficou muito parecidas entre as áreas quando se considera a dificuldade e em 2014. Em resumo, ao longo dos 5 anos estudados houve redução da porcentagem de questões classificadas como difíceis até o valor zero e, no sentido contrário, aumentaram as questões com maior capacidade discriminativa de forma proporcional em todas as áreas da prova.

Ao se avaliar os parâmetros de confiabilidade da prova, nota-se que as grandes áreas e por consequência o resultado total, apresentavam resultados entre 0,55-0,65 na sua maioria, considerada uma confiabilidade moderada. Alguns pontos de provas tiveram confiabilidade desprezíveis, mas que não se mantêm ao longo dos anos. Porém, considerando que são dados obtidos dos resultados dos alunos dos sextos anos do consórcio, pode-se considerar que a maioria desses estudantes tinha os conhecimentos necessários para realizar a prova, o que

impacta no quanto se considera esses valores, sendo importante avaliar em conjunto com os outros dados psicométricos.

Ainda, vale apontar que a prova nacional de 2015 foi a que obteve melhores valores nos coeficientes de fidedignidade, reforçando talvez o ponto de que uma prova mais abrangente, com múltiplos atores e com mais elaboradores retire vieses e garanta mais estruturação e qualidade aos testes.

A presença de áreas com questões deficientes e a necessidade de elevar o coeficiente de confiabilidade ao longo do tempo aponta para a importância de múltiplas verificações ao longo do processo de elaboração, e constante revisão da matriz de competências. Além disso, reforça a necessidade de treinamento dos elaboradores de questão e manutenção de um banco de questões de matriz de competências que são regularmente revisados.

As variações nos parâmetros psicométricos da prova são importantes, pois podem justificar mudanças no padrão de acertos de uma turma para mais ou para menos e deve fazer parte do processo de interpretação dos dados. Assim, os dados psicométricos podem também ser fator interessante para contribuir no entendimento do estudante quanto a seus resultados, por contextualizar melhor as características da prova em relação aos anos anteriores. Destaca-se o papel da “dificuldade” das questões para o entendimento dos estudantes quanto a seus resultados: podem avaliar se estão apenas acertando questões fáceis ou errando apenas questões difíceis, ampliando o escopo de interpretação de seus erros e acertos.

Quanto ao formato do teste, as provas do consórcio mantêm o modelo de múltipla escolha com uma única resposta correta (*‘one best answer’*) ao invés de provas com verdadeiro e falso ou resposta menos errada. Entende-se que além de ser um formato familiar ao estudante do consórcio, tem como vantagem evitar adivinhações associativas, por exemplo, no caso do modelo de múltiplos verdadeiros e falsos. Além disso, as provas com questões de múltipla escolha têm a vantagem de ter a correção mais simples e são adequadas para grande quantidade de alunos. Alguns autores defendem a marcação negativa e a presença do “não sei” como resposta, mas esse tipo de ferramenta avaliativa não é comum no Brasil e não mostram melhor efeito, pois os estudantes não estão habituados a ela e há um receio de que gere uso indevido dessa opção (12).

No quesito duração de prova, internacionalmente estipula-se que seriam necessários 75 a 85 segundos para responder um teste. No geral, o tempo de uma prova varia de 2,5 horas a 5 horas, destacando que o objetivo é promover uma avaliação que identifique com precisão se o estudante é capaz de demonstrar o conhecimento (16). Na FMRP-USP o aluno dispõe de 150

segundos para responder cada pergunta, e consideramos que pode ser uma estratégia para tentar familiarizar o estudante ao TP, uma vez que esse é o tempo médio fornecido para questões em provas de vestibular e residência médica no país. Muitos países oferecem mais de um TP por ano, familiarizando mais o estudante com o TP, o que pode facilitar a aplicação de provas mais rápidas e que demandem mais precisão do estudante, com menos tempo para cada questão e as vinhetas clínicas tendem a ser menores (18).

A experiência de 2015, com uma prova nacional, mostrou as potencialidades presentes para elaborações de provas pelo país. A prova apontou a possibilidade de avaliar instituições currículos muito distintos, a possibilidade de modificar o formato para se tornar mais abrangente sem perder a confiabilidade e não houve grande diferença de resultados entre essa prova e as outras nas coortes. Mas essa experiência também apontou as dificuldades em abarcar particularidades de cada região em questão de competências relevantes para epidemiologia local e o peso financeiro de se realizar provas tão abrangentes. Mas dessa experiência surgiram múltiplos núcleos regionais pelo país que passaram a elaborar TP a fim de mitigar as questões que limitavam sua aplicação nacionalmente (30).

E destaca-se, a nível institucional, a transformação do TP em uma disciplina optativa, reforçando sua importância e possibilitando que sua realização seja computada para pontos no currículo em provas de residência, para estágios internos, buscando sua valorização entre os estudantes.

No contexto avaliativo, o TP ocupa o papel singular de atividade formativa para mensuração do aprendizado das bases cognitivas teóricas. Com resultados que independem do desempenho nas provas somativas e complementando a avaliação prática terminal que garante uma avaliação das bases práticas, ambas permitindo uma avaliação seriada, longitudinal e formativa do conhecimento.

Algo a ser pontuado é que apesar do projeto político pedagógico da Medicina da FMRP-USP disponível atualmente referir que as práticas de feedback na instituição foram estabelecidas desde os anos 1990, quando solicitados esses documentos, foi possível perceber que não havia sistematização ou centralização destes dados: cada departamento poderia determinar o modelo de avaliação da disciplina e não necessariamente os resultados eram fornecidos aos coordenadores dos cursos.

Ainda, ao revisar as Atas das reuniões da Comissão de CoC, foi possível notar um empenho ao longo da última década para que as avaliações fossem sistematizadas e um empenho para aumentar a taxa de resposta dos estudantes nas avaliações das disciplinas. Um

processo de centralização e organização do formato foi sedimentado, muito pela organização do CAEG, que é voltado para os processos avaliativos da instituição. Neste contexto, foram exploradas novas práticas e potencialidades e desde 2019 as avaliações são semelhantes entre todos os departamentos e disciplinas.

## **5.2. Resultados das Coortes**

O trabalho apresentou acesso a amplo número de resultados ao longo do tempo e todas as coortes se comportaram de maneira semelhante entre si, com crescimento significativo em todas elas entre o primeiro e sexto ano, como ocorre na abundância de literatura internacional (18). Nesse sentido, destaca-se que o TP foi elaborado inicialmente em e para escolas europeias e da América do Norte, assim o desenvolvimento de trabalhos como esse permite discutir a validade deste tipo de ferramenta avaliativa para outras áreas do mundo, com condições sociais e econômicas distintas.

Este trabalho oferece evidências para o uso do TP fora do eixo do norte global e reforça o TP como ferramenta avaliativa, isto é, o fato de seus resultados serem semelhantes aqui como são em outras localidades com características educacionais muito distintas, reforçam sua consistência e viabilidade como ferramenta avaliativa e efetividade como medidor de desenvolvimento.

No geral, a adesão dos estudantes se mantém alta nos últimos 3 anos do curso, e não chega a cair mais que 20% nos anos iniciais. A adesão é uma preocupação dos elaboradores do teste e gestores de curso, então é interessante saber que ela cresce ao longo do curso, especialmente no contexto de que a participação dos estudantes é voluntária. Apesar de não sabermos a percepção dos estudantes ativamente, o crescimento da adesão pode indicar que eles ao longo do curso percebem a utilidade do TP como ferramenta de avaliação.

Assim, explorar as expectativas dos estudantes em relação ao TP na graduação podem trazer percepções e respostas que aumentem e sustentem a adesão ao longo de todo o curso. Um ponto percebido ao revisar roteiros de disciplinas e atas da CoC foi que muitas vezes o TP era aplicado próximo ao período de prova de disciplinas regulares do ano, podendo levar a maior tensão durante a realização do teste ou que estudantes decidam não comparecer para estudar para uma prova somativa.

A adesão dos estudantes é alvo de discussão e preocupação dos docentes: desde o início da elaboração do teste e nos períodos avaliados, foi aventada a possibilidade de tornar o teste obrigatório, pensando que os alunos faltantes seriam os estudantes cujo acompanhamento

longitudinal seria essencial, mas essa ideia não se manteve ao longo dos anos. Em outras instituições a nota do TP entra na nota do currículo, para estimular que o teste seja levado a sério pelos estudantes. Contudo, como se esperam notas baixas nos primeiros anos do teste, pode haver desconforto entre os estudantes.

Uma estratégia para contornar o efeito desestimulante entre os alunos dos anos iniciais é atribuir a nota com base na comparação com os colegas do mesmo ano. Assim, espera-se que as notas dos alunos dos anos iniciais sejam baixas, mas estejam próximas entre si e frações da nota mais alta observada não seriam desestimulantes. Nesse sentido, talvez a abordagem utilizada atualmente na FMRP-USP, de um registro formativo estar presente no currículo, seja associada a estratégias de inclusão e melhora do feedback, além de alocação estratégica na grade curricular, pode trazer poucas repercussões negativas entre os estudantes e ainda assim ofereça um estímulo ao teste.

O TP confirma o ganho cumulativo de conhecimento médico ao longo do curso. Em todas as turmas houve crescimento positivo significativo até o fim do curso, indicando que a prova se manteve como uma boa ferramenta avaliativa no decorrer dos anos. Também se percebe rapidamente entre provas quando houve queda ou nenhum crescimento. Isso mostra a potência para se fazer um diagnóstico rápido, ainda que não seja possível identificar imediatamente uma causa, permitiria iniciar discussões sobre as demandas daquela turma de estudantes.

Quanto ao contexto curricular, a introdução de disciplinas de uma determinada grande área na prova implica em aumento de acertos a partir daí. De forma geral em todas as turmas o crescimento do acerto de pediatria e cirurgia se iniciam entre o 2º e 3º ano, mas tem maior inflexão da curva entre o 3º e o 4º anos (figura 11, figura 12). Para GO e CM o crescimento mais consistente e a inflexão ocorrem entre o terceiro e quarto ano (figura 10, figura 25). Por outro lado, saúde coletiva e área básica apresentam notas médias iniciais mais altas já no primeiro ano, pois fazem parte do conteúdo inicial do currículo. Para estas duas últimas, não se observa um momento de deflexão acentuada na curva de acúmulo de conhecimento, indicando uma progressão/acúmulo mais gradativa e constante.

Pensando na relação da curva de crescimento cognitivo e a introdução de disciplinas, poderia se esperar que o conhecimento relacionado às áreas básicas deveria ser maior em momentos próximos a quando essas disciplinas são ministradas, mas o que se observa é que ao longo do curso o aluno passa a dominar melhor esse conteúdo (Figuras 8, 14, 20, 32). Uma explicação para isso pode ser a solicitação de recordação do conteúdo de fisiologia antes das

aulas das disciplinas clínicas, ou a repetição desse conteúdo com associação a outras disciplinas básicas, como farmacologia, nas disciplinas clínicas, e anatomia nas disciplinas cirúrgicas e de imagem.

Por outro lado, cargas horárias maiores não necessariamente indicam melhores desempenhos na prova, reforçando a tendência atual que aponta que não é necessário grandes volumes de aulas teóricas para que o conhecimento seja fixado. Quanto se avaliou a organização curricular, notou-se que a introdução de disciplinas ao longo das turmas, assim como mudança de carga horária feitas ao longo dos anos não resultaram em alterações no desempenho das turmas subsequentes no TP, indicando que a forma de elaboração da prova permite que haja menor influência de um único currículo sobre o conteúdo e, conseqüentemente, resultados nas provas.

Um exemplo importante foram as reduções de carga horária da área básica, com redução de carga horária das disciplinas em cerca de 300 horas, sem impactos significativos no desempenho. Por outro lado, Ginecologia e Obstetrícia manteve a carga horária praticamente a mesma durante esse período (Apêndice 1), e houve inflexões nos desempenhos, implicando que outros motivos influenciam mais nos resultados do teste.

Através de documentos da CoC, foi relatado o envio de e-mails e relatórios de desenvolvimento para os departamentos da unidade, para discutir tanto elaboração de prova quanto desempenho, porém não havia registro desses e-mails que pudesse ser acessado atualmente. Buscou-se obter registros do NATEPROM e CAEG, mas também não havia registro desses dados. Também por atas de reuniões da CoC, viu-se uma preocupação em se atentar a erros recorrentes de questões-âncora, mas não obtivemos mais informações de quais foram as estratégias utilizadas para evitar tal evento e quais as questões em cada prova são consideradas âncora.

### **5.3 Comparação com a Prova de Residência**

Em relação às pontuações médias da prova de residência, a ausência de diferença significativa entre as provas de 2011-2012 e 2013-2014 podem indicar provas ou perfis de estudantes mais parecidos que culminaram em resultados aproximados.

Quanto às correlações dos resultados do TP com a prova de residência, é interessante que os resultados apresentem correlações moderadas, especialmente ao se comparar os resultados gerais e do 6º ano, mas com resultados inconsistentes entre as coortes, com algumas turmas tendo correlações tão altas quanto 0,65 e tão baixas quanto 0,4. A correlação do 6º ano



ser mais forte que a do 5º ano também é interessante, pois indica que os estudantes estão mesmo mais preparados ao fim do curso.

De toda forma, não se esperavam correlações fortes pois, apesar de ocorrerem próximas uma da outra, com conteúdo similar dividido nas mesmas grandes áreas, elas possuem diferentes formatos de questões, e principalmente, tem diferentes objetivos: uma formativa para acompanhamento longitudinal durante a graduação e outra somativa com objetivo classificatório para ingresso na residência. Assim, espera-se que o fator humano, isto é, a reação dos estudantes frente a prova - seja bastante distinto, podendo justificar a correlação moderada e não forte entre as provas, mas não descartando seu potencial dentro de um contexto de planejamento a longo prazo para o estudante.

Nas divisões por quartis, destaca-se a tendência dos estudantes com melhores e piores desempenhos (1º e 4º quartis) a manterem-se nos quartis em que finalizam o curso na prova de residência, mas quando mudam de quartil não há uma tendência bem definida nas turmas de estudantes. Esse resultado é interessante, pois indica uma tendência de que os estudantes dos extremos mantenham seus desempenhos, o que pode guiar de alguma maneira seu planejamento. Também mostra que é possível mudar os resultados obtidos previamente, podendo ser uma informação adicional interessante para processos de remediação ativa de desempenho do aluno.

As comparações entre a prova do TP e a prova teórica de residência, assim como as comparações entre as provas de residência de cada coorte, neste estudo não puderam ser consideradas equivalentes, então não se pode comparar diretamente resultados do TP com a prova de residência, mas sim refinar e explorar os resultados do TP como ferramenta preditiva em um contexto que permita também outros tipos de avaliação. Estudos recentes brasileiros apontaram potencialidades ao se compararem os resultados do TP com a da prova de residência. Tanto com relação ao crescimento geral do 1º ao 6º ano quanto ao desempenho ao fim do curso (43, 44), indicando que o teste pode consolidar-se como uma ferramenta útil e complementar para guiar os estudantes quanto ao seu uso para autoavaliação ao longo do curso. Pode, também, ser usado para planejamento de estudos e planos futuros em relação à residência médica.

## 5.4 Exemplos de Uso Prático do Teste do Progresso

### 5.4.1. Observação de questões curriculares baseada nos resultados

Como exemplo de possível intervenção curricular usando os resultados do TP, tem-se que na coorte de 2011, entre o 4º e o 5º ano, houve uma queda do desempenho da turma os resultados gerais e se pode observar - inclusive visualmente- alterações na progressão esperada de ganho de conhecimento (Figura 7). Percebe-se que houve queda desempenho da turma nas áreas de ginecologia e obstetrícia (Figura 19), além de ausência de crescimento na área de cirurgia (Figura 17).

Nessa coorte, não houve mudança na carga horária em nenhuma dessas grandes áreas anteriormente a essas provas. Quanto aos dados da prova realizada em 2015, foi uma prova com boa confiabilidade, destacando-se que foi a prova de realização nacional, com a mudança de panorama sendo uma possível explicação para a maior dificuldade de crescimento em acertos.

Nesse caso, se os coordenadores achassem necessário, seria possível investigar se os resultados se reproduziram nas provas somativas, se houve dificuldade nas matérias e buscar um feedback ativo dos estudantes sobre a prova, conseguindo uma visão global sobre um possível problema, ou descartando-o como uma ocorrência que não necessitava de remediação.

### 5.4.2. Relatos de Caso de Progressão de Resultados de Estudantes

A apresentação de exemplos de casos de progressão de estudantes aponta também a facilidade de verificação da evolução do estudante, com a possibilidade de planejamento a partir dos resultados. Os gráficos mostrados são incidentais, isto é, as recuperações e pioras apresentadas pelos estudantes não ocorram mediante intervenção institucional, mas a ocorrência ou não de recuperação de alguns estudantes leva à reflexão sobre a necessidade de criar caminhos para que esse tipo de intervenção exista e possa gerar resultados positivos no desempenho dos estudantes.

A figura 38, com exemplo de apresentação gráfico de linha e áreas, ao invés de aspecto de histograma, apresenta-se de maneira mais intuitiva e é de maneira similar à apresentação dos dados aos estudantes que é feita no consórcio das escolas holandesas (41), permite uma visualização clara de onde o estudante se encontra em relação a seus pares, facilitando o entendimento de seus resultados.

Ainda, permite observações visuais sobre planos de mudanças no percurso de aprendizagem tanto para estudantes quanto para gestores. A longo prazo, a denominação das áreas (nesse trabalho divididos em quartis) podem mudar para se adequar a possíveis novos objetivos do teste, como no caso das escolas holandesas, com áreas com áreas denominadas “adequadas” ou “inadequadas” para o período específico do aprendizado.

A FMRP-USP atualmente tem passado por uma reforma curricular, que além de buscar um currículo baseado em competências, tem um plano de tutoria transversal ao curso, em que o estudante teria um docente-tutor que o auxiliaria a acompanhar o desempenho e planejar remediações intencionais caso sejam necessárias. Esse seria talvez um novo ambiente ideal para uma utilização mais ativa dos resultados do TP e implementar novas estratégias de visualização dos dados. Reforçando a ideia que as novas estratégias curriculares, p.e. construtivistas ou baseados em competências, devem também refletir nos espaços de avaliação e feedback (50).

### **5.5. Devolutiva e Feedback dos Resultados aos Estudantes**

A devolutiva que os estudantes recebem é um processo muito independente do restante do currículo. Não há espaço na grade para discussão de resultados que fossem considerados insuficientes para o estudante ou para os coordenadores do curso e a correção das questões ocorre apenas a partir do interesse do estudante. Isso pode ser um fator que leve a pouco uso pelos estudantes do TP como ferramenta de gerenciamento do próprio aprendizado. A fim de assegurar adesão de avaliações formativas, sugere-se que a apresentação dos dados estatísticos, comparativos e contextuais pode melhorar o entendimento dos resultados e como ele pode ser utilizado para autogestão de desempenho pelos estudantes.

Desde que esses resultados e análises sejam devidamente discutidos e apresentados aos estudantes, mais dados sobre o TP poderiam ser oferecidos, como desvio padrão das comparações, para apontar ao aluno se a diferença entre discente de diferentes consórcios é ou não significativa, além de distribuição de quartis/percentis, permitindo uma visão mais ampla ao estudante de sua colocação na prova e o que ela significa num panorama geral e se ela se adequa a suas metas individuais. Além de ampliar o panorama de percepção de aprendizado em relação aos pares, seria interessante que os estudantes pudessem também ter acesso aos dados da prova, se foi mais fácil ou mais difícil que a anterior, quais as questões difíceis da prova que realizaram.

Caso se mantenha e aprimore os dados relacionados ao padrão preditivo do TP no consórcio, tanto pensando nas metas para a expertise médica quanto para as provas de

residência, o fornecimento deste tipo de comparação também poderia estimular os estudantes a utilizarem o TP como uma ferramenta de autoeficácia e autoavaliação, para que possam fazer ajustes nos estudos de acordo com os objetivos traçados.

A potencialidade do teste de progresso como ferramenta de autoavaliação embasada depende também de tempo e estratégias para se debruçar sobre eles e compará-los com outras ferramentas avaliativas presentes no currículo. Tal constatação apontaria para necessidade de tempo, na estrutura curricular, para que o estudante desenvolva atividades relacionadas a portfólio, gestão de aprendizado e tutoria guiada, nos anos estudados. Sabe-se que quando cursos de medicina integram as avaliações formativas e os espaços de devolutiva no currículo em espaços protegidos, as avaliações tendem a ser mais recebidos pelos estudantes e o feedback mais bem utilizado por eles (49). Na FMRP-USP, observa-se que, apesar de ter havido uma diminuição na carga horária total do curso ao longo do tempo, não houve mudança na estrutura da grade curricular que permitisse a exploração de estratégias de crescimento profissional e pessoal.

Apesar de mais de um momento se discutir em reuniões da CoC, medidas para adesão e sugestões de participações mais ativas dos estudantes ao longo do processo de elaboração do TP, a participação do estudante no processo é mínima, sendo um gargalo importante para a expansão dos usos e importância do TP na graduação. Sabe-se que os estudantes devem ocupar cada vez mais espaços educacionais relacionados a avaliação e feedback, para que se torne o mais autêntico e efetivo possível (50). No momento não há, na FMRP-USP, um sistema de feedback que permita saber como os estudantes têm entendido sua participação no TP, com que seriedade realizam a prova e interpretam seus resultados, assim como a instituição está utilizando seus dados.

Consideramos que esse processo de avaliação, apesar de bem estabelecido, é recente, e ainda é preciso explorar e desenvolver maturidade da instituição e do corpo discente para sedimentar a importância da prova como ferramenta para todos, e expandir seus usos. Nesse contexto, associado a boa adesão ao teste que indica, no mínimo, um interesse do estudante pela prova, surge a demanda e a oportunidade de se desenvolver discussões sobre as causas do bom ou mau desempenho dos alunos na avaliação, considerando componentes como: empenho do aluno, modelo curricular, desenvolvimento das disciplinas, contexto universitário, validade do TP, dentre outros.

No mesmo sentido, apesar de haver a recomendação de que haja um feedback intencional para o processo de aprendizado no TP, ele acaba não acontecendo como iniciativa

regular dos departamentos ou das disciplinas do curso. Restando aos alunos apenas a checagem da pontuação e conferência do gabarito como mecanismo formativo. Nas reuniões da CoC, apesar de ter sido comentada a importância de participação dos estudantes, não foram elaboradas atividades para esse fim.

Dessa forma, apesar do grande empenho para elaborar a prova, vê-se pouca iniciativa para o uso efetivo do TP como atividade formativa. Observa-se que estudantes de medicina quando perguntados sobre avaliações no contexto de TP, apesar de considerar a prova útil, tendem a apontar que não consideram o TP uma prova justa e preferem fazer provas para as quais conseguem estudar (48). Somando-se isso há a percepção de que estudantes podem não compreender os objetivos do TP e ficam frustrados com o teste, pois não sentem que houve explicação para tal (28).

Assim, reforça-se a importância de estruturar um feedback e uma explicação prévia robusta para a realização do TP. Isso pode tanto auxiliar na adesão nos primeiros anos quanto na existência de dados mais confiáveis de resultados, pois os alunos poderiam ter maior comprometimento com o teste, bem como disponibilizar valores preditivos melhores para a avaliação do curso e do estudante. Assim forma-se um ciclo virtuoso, em que a melhor participação intencional e melhor devolutiva aos alunos pode levar a melhores resultados do TP, auxiliando no manejo de curso e autoavaliação.

## **5.6. Limitações do Estudo**

Por ser um estudo retrospectivo não foi possível a avaliação de percepção dos participantes quanto à realização do teste e uso de seus resultados à época. Assim, não se sabe como os estudantes da FMRP-USP se relacionaram com o TP, tanto em relação a sua realização quanto à utilização e entendimento de seus resultados.

Também não se tem estudos sobre as expectativas dos estudantes quanto a seu desempenho, sendo necessárias abordagens nesse sentido a fim de avaliar quais intervenções podem ser úteis para a adesão ao TP e para que os resultados sejam bem utilizados pelos estudantes. Da mesma forma, não foram avaliadas as percepções de gestores e docentes quanto ao uso do TP. Outros estudos são necessários para explorar como esses são interpretados e quais potencialidades ainda podem ser exploradas.

Na revisão documental, alguns documentos referentes à implementação do TP na FMRP-USP não estavam disponíveis pois eram antigos e sem cópias on-line e apesar de

inicialmente ter sido nossa intenção contatar alguns dos principais pioneiros na implementação do teste na faculdade, não foi possível contatar algumas dessas pessoas.

Apesar de correlações interessantes com os resultados das provas teóricas de residência, não foi possível correlacionar os resultados no TP, com o desfecho final na prova (aprovação no concurso em si) e mesmo considerando apenas a prova teórica, ainda são necessárias análises mais longas, com amostras maiores de estudantes para determinar como o valor preditivo do TP deve ser considerado para processos de autoavaliação e planejamento.

### **5.7. Considerações Finais**

O trabalho apresentou acesso a amplo número de resultados ao longo do tempo e todas as coortes se comportaram de maneira semelhante entre si, com o teste sendo capaz de determinar aumento nos acertos entre o primeiro e último ano, como também descrito nos relatos de literatura internacional.

Nesse sentido, vale destacar que o TP foi elaborado em e para escolas europeias e da América do Norte, e assim o desenvolvimento de trabalhos como esse permite discutir a validade deste tipo de ferramenta avaliativa para outras áreas do mundo, com condições sociais e econômicas distintas. Ademais, respalda sua efetividade como ferramenta avaliativa para estudantes brasileiros, nos permite iniciar novos questionamentos sobre suas aplicações no Brasil e outros países do hemisfério sul, além de abrir espaço para explorar novos campos metodológicos e perguntas de pesquisa referentes ao TP.

Além de registrar o uso do TP na FMRP-USP, este trabalho ofereceu novas evidências para destacar a potencialidade do uso do TP como versátil ferramenta avaliativa.

## 6. CONCLUSÕES

Esse trabalho conseguiu retomar e registrar a história do TP na FMRP-USP, além de registrar os resultados de seus estudantes ao longo do tempo. Com esse trabalho é possível apontar o TP como uma ferramenta útil para avaliar a progressão do conhecimento dos estudantes, permite a comparação com pares e uma avaliação mais imparcial, também em contextos de escolas de medicina brasileiras. A participação em consórcios e a elaboração externa tendem a melhorar a qualidade do teste e permite uma melhor confiabilidade do teste.

É potencialmente uma ferramenta preditiva para orientar a autoavaliação e o próprio planejamento dos alunos para atingir seus objetivos tanto na própria faculdade quanto para a residência.

Este trabalho também permitiu explorar as potencialidades do TP na FMRP-USP e apontar para a necessidade de aprimorar os processos de feedback e a inclusão dos alunos na avaliação das questões. Há espaço e necessidade de se explorar as percepções de estudantes e professores sobre o TP, além de buscar estratégia para registro do uso contemporâneo dos dados do teste.

## BIBLIOGRAFIA

1. MINISTÉRIO DA EDUCAÇÃO. Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina. RESOLUÇÃO Nº 3, DE 20 DE JUNHO DE 2014 Brasil; 2014.
2. dos Reis FJC. Uma Reflexão sobre os Desafios para a Graduação na FMRP-USP. *Jornal Eletrônico FMRP*. 2012 Sep 28;
3. Piccinato CE, Figueiredo JF de C, Troncon LE de A, Peres LC, Cianflone ARL, Colares M de FA, et al. Análise do desempenho dos formandos em relação a objetivos educacionais da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, em duas estruturas curriculares distintas. *Rev Assoc Med Bras*. 2004;50(1):68–73.
4. Norcini J, Anderson BASC, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 conference. In: *Medical Teacher*. 2011. p. 206–14.
5. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001 Mar 24;357(9260):945–9.
6. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990 Sep;65(9 Suppl):S63-7.
7. Cruess RL, Cruess SR, Steinert Y. Amending Miller’s Pyramid to Include Professional Identity Formation. *Academic Medicine*. 2016 Feb;91(2):180–5.
8. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programs. *Med Educ*. 2005 Mar;39(3):309–17.
9. Handy H, Prasad K, Anderson MB, Scherpbie A, Williams R, Zwierstra R, Cuddihy H. BEME systematic review: Predictive values of measurements obtained in medical schools and future performance in medical practice. *Medical Teacher*. 2006. 28:2, 103-116
10. Sordi MRL. Alternativas propostas no campo da avaliação: por que não? In 2009.
11. Perrenoud P. Avaliação da excelência à regulação das aprendizagens. São Paulo: Artmed; 1999.
12. Moraes DAF de. Avaliação Formativa: re-significando a prova no cotidiano escolar. Universidade Estadual de Londrina; 2008.
13. Hadji C. A avaliação, regras do jogo. Das intenções aos instrumentos. Porto Editora; 1994.
14. Black P, Wiliam D. Assessment and Classroom Learning. *Assess Educ*. 1998 Mar 28;5(1):7–74.
15. Epstein RM. Assessment in Medical Education. *New England Journal of Medicine*. 2007 Jan 25;356(4):387–96.
16. Shumway JM, Harden RM. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach*. 2003 Nov 3;25(6):569–84.
17. Wrigley W, van der Vleuten CPM, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints, and issues: AMEE Guide No. 71. *Med Teach*. 2012;34(9):683–97.
18. Freeman A, Van Der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Med Teach*. 2010 Jan;32(6):451–5.
19. Vleuten CPM Van Der, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach*. 1996 Jan 3;18(2):103–9.
20. Arnold L, Willoughby TL. The quarterly profile examination. *Academic Medicine*. 1990 Aug;65(8):515–6.
21. Verhoeven H, Verwijnen GM, Basc AJ. An analysis of progress test results of PBL and non-PBL students. *Med Teach*. 1998 Jan 3;20(4):310–6.
22. Karay Y, Reiss BASC, Schaubert SK. Progress testing anytime and anywhere – Does a mobile-learning approach enhance the utility of a large-scale formative assessment tool? *Med Teach*. 2020 Oct 2;42(10):1154–62.



23. Swanson DB, Holtzman KZ, Butler A, Langer MM, Nelson M V., Chow JWM, et al. Collaboration across the pond: The multi-school progress testing project. *Med Teach*. 2010;32(6):480–5.
24. Portanova R, Adelman M, Jollick JD, Schuler S, Modrzakowski M, Soper E, et al. Student assessment in the Ohio University College of Osteopathic Medicine CORE system: Progress testing and objective structured clinical examinations. *J Am Osteopath Assoc*. 2000 Nov 1;100:707–12.
25. Dijksterhuis MGK, Schuwirth LWT, Braat DDM, Scheele F. An exploratory study into the impact and acceptability of formatively used progress testing in postgraduate obstetrics and gynecology. *Perspect Med Educ*. 2013 Jun 21;2(3):126–41.
26. Aarts R, Steidel K, Manuel BAF, Driessen EW. Progress testing in resource-poor countries: a case from Mozambique. *Med Teach*. 2010;32(6):461–3.
27. Lillis S, Yelder J, Mogol V, O'Connor BASC, Bacal K, Booth R, et al. Progress Testing for Medical Students at the University of Auckland: Results from the First Year of Assessments. *J Med Educ Curric Dev*. 2014 Jan;1:JMECD.S20094.
28. Matsuyama Y, Muijtjens AMM, Kikukawa M, Stalmeijer R, Murakami R, Ishikawa S, et al. A first report of East Asian students' perception of progress testing: a focus group study. *BMC Med Educ*. 2016 Dec 22;16(1):245.
29. Cecilio-Fernandes D, Bicudo AM, Filho PTH. Progress testing as a pattern of excellence for the assessment of medical students' knowledge - Concepts, history, and perspective. Vol. 54, *Medicina (Brazil)*. Faculdade de Medicina de Ribeirão Preto - U.S.P.; 2021.
30. Bicudo AM, Hamamoto Filho PT, Abbade JF, Hafner M de LMB, Maffei CML. Teste de Progresso em Consórcios para Todas as Escolas Médicas do Brasil. *Rev Bras Educ Med*. 2019 Dec;43(4):151–6.
31. Tomic ER, Martins MA, Lotufo PA, Benseñor IM. Progress testing: evaluation of four years of application in the School of Medicine, University of São Paulo. *Clinics*. 2005 Oct;60(5):389–96.
32. Sakai MH, Ferreira Filho OF, Almeida MJ de, Mashima DA, Marchese M de C. Teste de progresso e avaliação do curso: dez anos de experiência da medicina da Universidade Estadual de Londrina. *Rev Bras Educ Med*. 2008 Jun;32(2):254–63.
33. de Sá MFS, Romão GS, Fernandes CE, Filho AL da S. The Individual Progress Test of Gynecology and Obstetrics Residents (TPI-GO): The Brazilian Experience by FEBRASGO. *Revista Brasileira de Ginecologia e Obstetrícia / RBGO Gynecology and Obstetrics*. 2021 Jun 27;43(06):425–8.
34. Blake JM, Norman GR, Keane DR, Mueller CB, Cunningham J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum. *Academic Medicine*. 1996 Sep;71(9):1002–7.
35. Cecilio-Fernandes D, Kerdijk W, Jaarsma AD (Debbie) C, Tio RA. Development of cognitive processing and judgments of knowledge in medical students: Analysis of progress test results. *Med Teach*. 2016 Nov 1;38(11):1125–9.
36. Van Berkel HJM, Nuy HJP, Geerligs T. The influence of progress tests and block tests on student behavior. Vol. 22, *Instructional Science*. Kluwer Academic Publishers; 1995.
37. Chen Y, Henning M, Yelder J, Jones R, Wearn A, Weller J. Progress testing in the medical curriculum: students' approaches to learning and perceived stress. *BMC Med Educ*. 2015 Dec 11;15(1):147.
38. Muijtjens AMM, Timmermans I, Donkers J, Peperkamp R, Medema H, Cohen-Schotanus J, et al. Flexible electronic feedback using the virtues of progress testing. *Med Teach*. 2010 Jan;32(6):491–5.
39. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programs. *Med Educ*. 2005 Mar;39(3):309–17.

40. De Champlain AF, Cuddy MM, Scoles P V., Brown M, Swanson DB, Holtzman K, et al. Progress testing in clinical science education: Results of a pilot project between the National Board of Medical Examiners and a US Medical School. *Med Teach*. 2010 Jan;32(6):503–8.
41. Tio RA, Schutte B, Meiboom AA, Greidanus J, Dubois EA, Bremers AJ, & Dutch Working Group of the Interuniversity Progress Test of Medicine (2016). The progress test of medicine: the Dutch experience. *Perspectives on medical education*, 5(1), 51–55.
42. Schuwirth LWT, van der Vleuten CPM. Programmatic assessment and Kane’s validity perspective. *Med Educ*. 2012 Jan;46(1):38–48
43. Hamamoto Filho PT, de Arruda Lourenção PLT, do Valle AP, Abbade JF, Bicudo AM. The Correlation Between Students’ Progress Testing Scores and Their Performance in a Residency Selection Process. *Med Sci Educ*. 2019 Dec 11;29(4):1071–5.
44. Guaraná CVP dos S, Duque TB, Lima TR de M et al. Desempenho no teste do progresso e coeficiente de rendimento final de curso são preditores de aprovação na residência médica? *Medicina (Ribeirão Preto)*. 2022 Jul 6;55(2).
45. Rosa MI da, Isoppo CC, Cattaneo HD, Madeira K, Adami F, Ferreira Filho OF. O Teste de Progresso como Indicador para Melhorias em Curso de Graduação em Medicina. *Rev Bras Educ Med*. 2017 Jan;41(1):58–68.
46. Schuwirth LWT, Van der Vleuten CPM. The use of progress testing. *Perspect Med Educ*. 2012 Mar 10;1(1):24–30.
47. Ali SH, Ruit KG. The Impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. *Perspect Med Educ*. 2015 Jul, 4(5), 244–251.
48. Ali K, Cockerill J, Zahra D, Tredwin C, Ferguson C. Impact of Progress testing on the learning experiences of students in medicine, dentistry, and dental therapy. *BMC Med Educ*. 2018 Dec 9;18(1):253
49. Schüttpelz-Brauns K, Karay Y, Arias J, Gehlhar K, Zupanic M. Comparison of the evaluation of formative assessment at two medical faculties with different conditions of undergraduate training, assessment, and feedback. *GMS J Med Educ*. 2020; 37(4):Doc41
50. Harrison CJ, Könings KD, Dannefer EF et al. Factors influencing students’ receptivity to formative feedback emerging from different assessment cultures. *Perspect Med Educ* 2016. 5, 276–284.

## APÊNDICES

Apêndice 1 - Distribuição de Disciplinas ao longo do curso médico - Número de disciplinas (horas)

Coorte Ano	Ano do curso	Área Básica	Clínica Médica	Cirurgia	Pediatria	Saúde Coletiva	Ginecologia e Obstetrícia
2010	1°	8 (815)	1 (45)	-	-	1 (120)	-
	2°	8 (1047)	-	-	-	1 (112)	-
	3°	5 (589)	3 (256)	4 (199)	1 (90)	2 (85)	-
	4°	-	16 (879)	4 (157)	1 (80)	1 (45)	2 (255)
	5°	-	3 (780)	2 (435)	2 (555)	1 (450)	1 (420)
	6°	-	5 (617)	3 (456)	2 (323)	1 (240)	1 (360)
	Total	21 (2451)	28 (2577)	13 (1247)	6 (1048)	7 (1052)	4 (1035)
2011	1°	8 (905)	1 (45)	-	-	1 (120)	-
	2°	10(1362)	-	-	-	1 (112)	-
	3°	5 (342)	3 (237)	4 (229)	1 (90)	1 (45)	-
	4°	-	18 (975)	4 (157)	1 (135)	1 (45)	2 (255)
	5°	-	3 (780)	2 (435)	2 (555)	1 (450)	1 (420)
	6°	-	4 (548)	3 (456)	2(323)	1 (240)	1 (360)
	Total	23(2609)	29 (2584)	13 (1276)	6 (1102)	6 (1012)	4 (1035)
2012	1°	8 (905)	1 (45)	-	-	1 (120)	-
	2°	10 (1362)	-	-	-	1 (64)	-
	3°	5 (342)	3 (253)	4 (229)	1 (90)	1 (45)	-
	4°	-	18 (935)	4 (157)	1 (135)	1 (40)	2 (255)
	5°	-	3 (780)	2 (435)	2 (555)	1 (360)	1 (420)
	6°	-	5 (617)	3 (456)	2 (323)	1 (330)	1 (360)
	Total	23 (2609)	30 (2630)	13 (1276)	6 (1102)	6 (959)	4 (1035)
2013	1°	7 (758)	1 (45)	-	-	1 (90)	-
	2°	10(1050)	-	-	-	1 (64)	-
	3°	8 (550)	3 (250)	4 (207)	1 (90)	1 (45)	-
	4°	-	17 (908)	4 (157)	1 (135)	1 (40)	2 (255)
	5°	-	2 (555)	2 (435)	2 (555)	1 (240)	1 (420)
	6°	-	5 (617)	3 (456)	2 (323)	1 (330)	1 (360)
	Total	25(2358)	28 (2555)	13 (1254)	5 (1102)	6 (809)	4 (1035)
2014	1°	7(759)	1 (45)	-	-	1 (90)	-
	2°	10 (1079)	-	-	-	1 (64)	-
	3°	8 (700)	3 (250)	4 (186)	1 (90)	1 (45)	-
	4°	-	17 (910)	4 (157)	1 (80)	1 (40)	2 (255)
	5°	-	2 (425)	2 (435)	2 (555)	1 (240)	1 (420)
	6°	-	6 (755)	3 (456)	2 (323)	1 (330)	1 (360)
	Total	25 (2538)	29 (2383)	13 (1234)	5 (1048)	6 (809)	4 (1035)

## Apêndice 2 -Exploratória média de resultados gerais por Coorte por ano

Coorte	Ano	Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo	Desvio Padrão	n
2010	1	0,16	0,25	0,28	<b>0,29</b>	0,31	0,46	<b>0,05</b>	82
2010	2	0,15	0,36	0,41	<b>0,40</b>	0,44	0,54	<b>0,07</b>	70
2010	3	0,23	0,35	0,4	<b>0,40</b>	0,44	0,59	<b>0,07</b>	69
2010	4	0,24	0,49	0,55	<b>0,53</b>	0,58	0,69	<b>0,08</b>	67
2010	5	0,38	0,56	0,62	<b>0,61</b>	0,68	0,81	<b>0,08</b>	78
2010	6	0,38	0,64	0,7	<b>0,70</b>	0,76	0,91	<b>0,09</b>	84
2011	1	0,27	0,34	0,38	<b>0,38</b>	0,42	0,52	<b>0,05</b>	57
2011	2	0,20	0,32	0,37	<b>0,36</b>	0,41	0,55	<b>0,07</b>	60
2011	3	0,2	0,38	0,45	<b>0,44</b>	0,52	0,65	<b>0,09</b>	59
2011	4	0,35	0,51	0,57	<b>0,57</b>	0,65	0,76	<b>0,09</b>	65
2011	5	0,25	0,50	0,57	<b>0,56</b>	0,63	0,80	<b>0,11</b>	73
2011	6	0,35	0,65	0,72	<b>0,71</b>	0,77	0,86	<b>0,09</b>	78
2011	7	0,45	0,59	0,67	<b>0,65</b>	0,71	0,89	<b>0,08</b>	68
2012	1	0,21	0,26	0,3	<b>0,30</b>	0,32	0,41	<b>0,04</b>	60
2012	2	0,13	0,32	0,38	<b>0,38</b>	0,43	0,53	<b>0,07</b>	60
2012	3	0,25	0,42	0,47	<b>0,46</b>	0,52	0,64	<b>0,07</b>	65
2012	4	0,36	0,45	0,51	<b>0,50</b>	0,56	0,66	<b>0,08</b>	57
2012	5	0,21	0,6	0,66	<b>0,64</b>	0,70	0,81	<b>0,10</b>	65
2012	6	0,50	0,675	0,70	<b>0,70</b>	0,74	0,84	<b>0,06</b>	67
2013	1	0,2	0,305	0,35	<b>0,34</b>	0,38	0,45	<b>0,05</b>	83
2013	2	0,21	0,35	0,41	<b>0,41</b>	0,48	0,58	<b>0,08</b>	77
2013	3	0,24	0,35	0,4	<b>0,41</b>	0,48	0,61	<b>0,08</b>	80
2013	4	0,33	0,505	0,55	<b>0,55</b>	0,63	0,77	<b>0,10</b>	83
2013	5	0,28	0,58	0,64	<b>0,63</b>	0,70	0,84	<b>0,09</b>	88
2013	6	0,43	0,67	0,71	<b>0,71</b>	0,77	0,85	<b>0,08</b>	90
2014	1	0,24	0,33	0,37	<b>0,37</b>	0,42	0,53	<b>0,05</b>	85
2014	2	0,23	0,3	0,33	<b>0,33</b>	0,38	0,49	<b>0,05</b>	86
2014	3	0,28	0,44	0,49	<b>0,48</b>	0,53	0,65	<b>0,08</b>	86
2014	4	0,20	0,48	0,56	<b>0,52</b>	0,60	0,70	<b>0,10</b>	84
2014	5	0,43	0,58	0,63	<b>0,62</b>	0,68	0,78	<b>0,06</b>	87
2014	6	0,50	0,69	0,75	<b>0,75</b>	0,80	0,92	<b>0,08</b>	88

## Apêndice 3 - Tabelas de detalhamento das comparações por área por Coorte

Tabela 12 – Comparação de resultados da área básica na Coorte de 2010

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
BASC	2010	1	2	-0,1197	0,0001	-0,1549	-0,08446
BASC	2010	2	3	-0,1213	0,0001	-0,1579	-0,08476
BASC	2010	3	4	0,08183	0,0001	0,04464	0,119
BASC	2010	4	5	-0,1224	0,0001	-0,1587	-0,08616
BASC	2010	1	5	-0,2816	0,0001	-0,3159	-0,2474

Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 13 - Comparação de resultados da área de saúde coletiva na Coorte de 2010

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
SC	2010	1	2	-0,1537	0,0001	-0,1897	-0,1177
SC	2010	1	6	-0,2782	0,0001	-0,3124	-0,244
SC	2010	2	3	0,04389	0,0215	0,006471	0,08131
SC	2010	3	4	0,007816	0,6863	-0,03013	0,04576
SC	2010	4	5	-0,1197	0,0001	-0,1566	-0,08275
SC	2010	5	6	-0,05659	0,0014	-0,09127	-0,0219

Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 14 - Comparação de resultados da área de Clínica Médica na Coorte de 2010

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
CM	2010	1	2	-0,0973	0,0001	-0,1318	-0,0628
CM	2010	2	3	0,0409	0,0255	0,0050	0,0768
CM	2010	3	4	-0,2317	0,0001	-0,2681	-0,1953
CM	2010	4	5	-0,0627	0,0005	-0,0981	-0,0273
CM	2010	5	6	-0,1647	0,0001	-0,1980	-0,1315
CM	2010	1	6	-0,5155	0,0001	-0,5483	-0,4827

Legenda: CM: área clínica médica da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 15 - Comparação de resultados da área de Cirurgia na Coorte de 2010

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
-------------	--------	------------	--	--------------------	---------	---	--

		AnoProva				Limite Inferior	Limite Superior
<b>CX</b>	2010	1	2	-0,1014	0,0001	-0,1379	-0,0649
<b>CX</b>	2010	1	6	-0,4882	0,0001	-0,5229	-0,4534
<b>CX</b>	2010	2	3	-0,0008	0,9685	-0,0388	0,0373
<b>CX</b>	2010	3	4	-0,2487	0,0001	-0,2873	-0,2102
<b>CX</b>	2010	4	5	0,0148	0,4394	-0,0227	0,0523
<b>CX</b>	2010	5	6	-0,1521	0,0001	-0,1873	-0,1168

Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 16 - Comparação de resultados da área de Pediatria na Coorte de 2010

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>P</b>	2010	1	2	-0,1004	0,0001	-0,1357	-0,0652
<b>P</b>	2010	1	6	-0,4853	0,0001	-0,5189	-0,4518
<b>P</b>	2010	2	3	0,0101	0,5906	-0,0266	0,0468
<b>P</b>	2010	3	4	-0,2269	0,0001	-0,2641	-0,1897
<b>P</b>	2010	4	5	-0,0327	0,0762	-0,0689	0,0034
<b>P</b>	2010	5	6	-0,1353	0,0001	-0,1693	-0,1013

Legenda: P: área de Pediatria da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 17 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2010

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>GO</b>	2010	1	2	-0,1021	0,0001	-0,1370	-0,0672
<b>GO</b>	2010	1	6	-0,4044	0,0001	-0,4376	-0,3711
<b>GO</b>	2010	2	3	0,0583	0,0017	0,0220	0,0947
<b>GO</b>	2010	3	4	-0,2178	0,0001	-0,2547	-0,1810
<b>GO</b>	2010	4	5	-0,1346	0,0001	-0,1704	-0,0988
<b>GO</b>	2010	5	6	-0,0082	0,6337	-0,0419	0,0255

Legenda: GO: área de ginecologia e obstetrícia prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 18- Comparação de resultados da área básica na Coorte de 2011

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>BASC</b>	2011	1	2	-0,2196	0,0001	-0,2599	-0,1793
<b>BASC</b>	2011	1	6	-0,3291	0,0001	-0,3671	-0,291
<b>BASC</b>	2011	2	3	0,1081	0,0001	0,06816	0,1481
<b>BASC</b>	2011	3	4	-0,1827	0,0001	-0,2216	-0,1438

<b>BASC</b>	2011	4	6	-0,03483	0,0612	-0,07131	0,001641
-------------	------	---	---	----------	--------	----------	----------

Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 19 - Comparação de resultados da área de Saúde Coletiva na Coorte de 2011

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>SC</b>	2011	1	2	0,1115	0,0001	0,07048	0,1525
<b>SC</b>	2011	2	3	-0,02553	0,2186	-0,06621	0,01516
<b>SC</b>	2011	3	4	-0,1191	0,0001	-0,1588	-0,07934
<b>SC</b>	2011	4	5	-0,02283	0,2359	-0,0606	0,01494
<b>SC</b>	2011	5	6	-0,1404	0,0001	-0,1763	-0,1045
<b>SC</b>	2011	1	6	-0,1963	0,0001	-0,235	-0,1576

Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 20 - Comparação de resultados da área de Clínica Médica na Coorte de 2011

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>CM</b>	2011	1	2	0,0305	0,1287	-0,0089	0,0699
<b>CM</b>	2011	2	3	-0,1531	0,0001	-0,1921	-0,1141
<b>CM</b>	2011	3	4	-0,1343	0,0001	-0,1725	-0,0962
<b>CM</b>	2011	4	5	-0,0806	0,0001	-0,1169	-0,0444
<b>CM</b>	2011	5	6	-0,0504	0,0042	-0,0849	-0,0159
<b>CM</b>	2011	1	6	-0,3880	0,0001	-0,4251	-0,3509

Legenda: CM: área clínica médica da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 21 - Comparação de resultados da área de Cirurgia na Coorte de 2011

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>CX</b>	2011	1	2	0,0830	0,0001	0,0413	0,1247
<b>CX</b>	2011	2	3	-0,2567	0,0001	-0,2980	-0,2153
<b>CX</b>	2011	3	4	0,0298	0,1478	-0,0106	0,0702
<b>CX</b>	2011	4	5	-0,0284	0,147	-0,0668	0,0100
<b>CX</b>	2011	5	6	-0,1940	0,0001	-0,2305	-0,1575
<b>CX</b>	2011	1	6	-0,3662	0,0001	-0,4055	-0,3269

Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 22 - Comparação de resultados da área de Pediatria na Coorte de 2011

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>P</b>	2011	1	2	-0,0105	0,6098	-0,0507	0,0298
<b>P</b>	2011	2	3	-0,1619	0,0001	-0,2018	-0,1221
<b>P</b>	2011	3	4	-0,0322	0,1049	-0,0712	0,0067
<b>P</b>	2011	4	5	-0,0899	0,0001	-0,1269	-0,0529
<b>P</b>	2011	5	6	-0,1346	0,0001	-0,1698	-0,0993
<b>P</b>	2011	1	6	-0,4291	0,0001	-0,4670	-0,3912

Legenda: P: área de Pediatria da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 23 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2011

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>GO</b>	2011	1	2	0,0349	0,0852	-0,0049	0,0747
<b>GO</b>	2011	2	3	-0,0266	0,1869	-0,0660	0,0129
<b>GO</b>	2011	3	4	-0,2657	0,0001	-0,3043	-0,2271
<b>GO</b>	2011	4	5	0,0713	0,0001	0,0346	0,1079
<b>GO</b>	2011	5	6	-0,2009	0,0001	-0,2358	-0,1660
<b>GO</b>	2011	1	6	-0,3870	0,0001	-0,4245	-0,3494

Legenda: GO: área de ginecologia e obstetrícia prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 24 - Comparação de resultados da área básica na Coorte de 2012

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>BASC</b>	2012	1	2	-0,07862	0,0001	-0,1183	-0,03896
<b>BASC</b>	2012	1	6	-0,3136	0,0001	-0,3524	-0,2749
<b>BASC</b>	2012	2	3	-0,09866	0,0001	-0,1374	-0,05992
<b>BASC</b>	2012	3	5	-0,1005	0,0001	-0,1386	-0,06246
<b>BASC</b>	2012	5	6	-0,03581	0,0615	-0,07335	0,001723

Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 25 - Comparação de resultados da área de Saúde Coletiva na Coorte de 2012

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>SC</b>	2012	1	2	-0,04657	0,0242	-0,08705	-0,00608
<b>SC</b>	2012	2	3	-0,1423	0,0001	-0,1819	-0,1027



SC	2012	3	4	0,02404	0,2405	-0,01611	0,06419
SC	2012	4	5	-0,1522	0,0001	-0,1924	-0,112
SC	2012	5	6	-0,05375	0,0061	-0,09215	-0,01534
SC	2012	1	6	-0,3708	0,0001	-0,4103	-0,3313

Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 26 - Comparação de resultados da área de Clínica Médica na Coorte de 2012

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
CM	2012	1	2	-0,1344	0,0001	-0,1732	-0,0955
CM	2012	2	3	-0,0620	0,0014	-0,1000	-0,0241
CM	2012	3	4	-0,0835	0,0001	-0,1220	-0,0449
CM	2012	4	5	-0,0796	0,0001	-0,1181	-0,0410
CM	2012	5	6	-0,1749	0,0001	-0,2118	-0,1381
CM	2012	1	6	-0,5343	0,0001	-0,5722	-0,4964

Legenda: CM: área clínica médica da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 27 - Comparação de resultados da área de Cirurgia na Coorte de 2012

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
CX	2012	1	2	-0,1413	0,0001	-0,1824	-0,1002
CX	2012	2	3	-0,0063	0,759	-0,0465	0,0339
CX	2012	3	4	-0,0957	0,0001	-0,1365	-0,0550
CX	2012	4	5	-0,1571	0,0001	-0,1979	-0,1163
CX	2012	5	6	-0,0945	0,0001	-0,1335	-0,0555
CX	2012	1	6	-0,4950	0,0001	-0,5351	-0,4548

Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 28 - Comparação de resultados da área de Pediatria na Coorte de 2012

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
P	2012	1	2	-0,0957	0,0001	-0,1354	-0,0561
P	2012	2	3	-0,0792	0,0001	-0,1180	-0,0404
P	2012	3	4	-0,0931	0,0001	-0,1324	-0,0537
P	2012	4	5	-0,1415	0,0001	-0,1809	-0,1021
P	2012	5	6	0,0167	0,3857	-0,0210	0,0543
P	2012	1	6	-0,3928	0,0001	-0,4315	-0,3541

Legenda: P: área de Pediatria da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 29 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2012

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>GO</b>	2012	1	2	-0,0311	0,1204	-0,0704	0,0082
<b>GO</b>	2012	2	3	-0,0796	0,0001	-0,1181	-0,0412
<b>GO</b>	2012	3	4	-0,0592	0,0029	-0,0981	-0,0202
<b>GO</b>	2012	4	5	-0,1439	0,0001	-0,1829	-0,1049
<b>GO</b>	2012	5	6	-0,0301	0,1139	-0,0674	0,0072
<b>GO</b>	2012	1	6	-0,3439	0,0001	-0,3822	-0,3056

Legenda: GO: área de ginecologia e obstetrícia prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 30 - Comparação de resultados da área básica na Coorte de 2013

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>BASC</b>	2013	1	2	-0,1307	0,0001	-0,1649	-0,09643
<b>BASC</b>	2013	1	6	-0,2002	0,0001	-0,233	-0,1675
<b>BASC</b>	2013	2	3	-0,1386	0,2439	-0,3718	0,09461
<b>BASC</b>	2013	3	4	0,06466	0,5858	-0,168	0,2973
<b>BASC</b>	2013	4	5	-0,02935	0,08	-0,06221	0,003514
<b>BASC</b>	2013	5	6	0,03373	0,0404	0,001475	0,06599

Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 31- Comparação de resultados da área de saúde coletiva na Coorte de 2013

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>SC</b>	2013	1	2	-0,1315	0,0001	-0,1663	-0,09659
<b>SC</b>	2013	2	3	-0,03817	0,0337	-0,0734	-0,00294
<b>SC</b>	2013	3	4	-0,08428	0,0001	-0,1187	-0,04981
<b>SC</b>	2013	4	5	-0,03594	0,0363	-0,06958	-0,0023
<b>SC</b>	2013	5	6	-0,1079	0,0001	-0,141	-0,0749
<b>SC</b>	2013	1	6	-0,3978	0,0001	-0,4313	-0,3643

Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 32 - Comparação de resultados da área de Clínica Médica na Coorte de 2013

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior

		AnoProva				Limite Inferior	Limite Superior
<b>CM</b>	2013	1	2	-0,0896	0,0001	-0,1231	-0,0562
<b>CM</b>	2013	2	3	-0,0284	0,1004	-0,0623	0,0055
<b>CM</b>	2013	3	4	-0,1068	0,0001	-0,1400	-0,0736
<b>CM</b>	2013	4	5	-0,2466	0,0001	-0,2789	-0,2143
<b>CM</b>	2013	5	6	0,0320	0,0475	0,0004	0,0637
<b>CM</b>	2013	1	6	-0,4394	0,0001	-0,4716	-0,4073

Legenda: CM: área clínica médica da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 33 - Comparação de resultados da área de Cirurgia na Coorte de 2013

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>CX</b>	2013	1	2	0,0174	0,339	-0,0183	0,0531
<b>CX</b>	2013	2	3	-0,1404	0,0001	-0,1764	-0,1043
<b>CX</b>	2013	3	4	-0,1180	0,0001	-0,1530	-0,0830
<b>CX</b>	2013	4	5	-0,0954	0,0001	-0,1295	-0,0612
<b>CX</b>	2013	5	6	-0,0104	0,5446	-0,0439	0,0232
<b>CX</b>	2013	1	6	-0,3467	0,0001	-0,3807	-0,3126

Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 34 - Comparação de resultados da área de Pediatria na Coorte de 2013

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>P</b>	2013	1	2	-0,0326	0,0618	-0,0668	0,0016
<b>P</b>	2013	2	3	-0,0401	0,023	-0,0746	-0,0055
<b>P</b>	2013	3	4	-0,1554	0,0001	-0,1892	-0,1216
<b>P</b>	2013	4	5	0,0039	0,8169	-0,0291	0,0369
<b>P</b>	2013	5	6	-0,1732	0,0001	-0,2056	-0,1408
<b>P</b>	2013	1	6	-0,3973	0,0001	-0,4302	-0,3645

Legenda: P: área de Pediatria da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 35 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2013

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>GO</b>	2013	1	2	-0,0370	0,0333	-0,0711	-0,0029
<b>GO</b>	2013	2	3	0,0384	0,0296	0,0038	0,0729

<b>GO</b>	2013	3	4	-0,1593	0,0001	-0,1929	-0,1258
<b>GO</b>	2013	4	5	-0,0514	0,0021	-0,0840	-0,0187
<b>GO</b>	2013	5	6	-0,2089	0,0001	-0,2410	-0,1768
<b>GO</b>	2013	1	6	-0,4183	0,0001	-0,4508	-0,3858

Legenda: GO: área de ginecologia e obstetrícia prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 36- Comparação de resultados da área básica na Coorte de 2014

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>BASC</b>	2014	1	3	-0,1437	0,0001	-0,1766	-0,1108
<b>BASC</b>	2014	1	6	-0,2163	0,0001	-0,2491	-0,1836
<b>BASC</b>	2014	3	4	0,002721	0,8715	-0,03027	0,03571
<b>BASC</b>	2014	4	5	-0,05538	0,001	-0,08828	-0,02249
<b>BASC</b>	2014	5	6	-0,01995	0,2275	-0,05237	0,01247

Legenda: BASC: área básica da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 37 - Comparação de resultados da área de saúde coletiva na Coorte de 2014

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>SC</b>	2014	1	2	0,02336	0,1742	-0,01034	0,05706
<b>SC</b>	2014	2	3	-0,1326	0,0001	-0,1662	-0,09898
<b>SC</b>	2014	3	4	0,04191	0,0151	0,008122	0,07569
<b>SC</b>	2014	4	5	-0,1171	0,0001	-0,1507	-0,08341
<b>SC</b>	2014	5	6	-0,1446	0,0001	-0,1778	-0,1114
<b>SC</b>	2014	1	6	-0,329	0,0001	-0,3625	-0,2954

Legenda: SC: área de saúde coletiva da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 38 - Comparação de resultados da área de Clínica Médica na Coorte de 2014

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>CM</b>	2014	1	2	0,0136	0,4102	-0,0188	0,0459
<b>CM</b>	2014	2	3	-0,1100	0,0001	-0,1422	-0,0778
<b>CM</b>	2014	3	4	-0,2214	0,0001	-0,2538	-0,1890
<b>CM</b>	2014	4	5	-0,0015	0,9278	-0,0338	0,0308
<b>CM</b>	2014	5	6	-0,0761	0,0001	-0,1080	-0,0443
<b>CM</b>	2014	1	6	-0,3954	0,0001	-0,4275	-0,3632

Legenda: CM: área clínica médica da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 39 - Comparação de resultados da área de Cirurgia na Coorte de 2014

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>CX</b>	2014	1	2	-0,0012	0,9477	-0,0355	0,0332
<b>CX</b>	2014	2	3	-0,1854	0,0001	-0,2195	-0,1513
<b>CX</b>	2014	3	4	-0,0394	0,0244	-0,0737	-0,0051
<b>CX</b>	2014	4	5	-0,0682	0,0001	-0,1024	-0,0340
<b>CX</b>	2014	5	6	-0,1625	0,0001	-0,1962	-0,1287
<b>CX</b>	2014	1	6	-0,4566	0,0001	-0,4908	-0,4224

Legenda: CX: área de cirurgia da prova; AnoProva: ano do curso em que a prova foi realizada

Tabela 40 - Comparação de resultados da área de Pediatria na Coorte de 2014

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>P</b>	2014	1	2	-0,0143	0,3975	-0,0473	0,0188
<b>P</b>	2014	2	3	-0,1431	0,0001	-0,1760	-0,1101
<b>P</b>	2014	3	4	0,0271	0,1087	-0,0060	0,0603
<b>P</b>	2014	4	5	-0,1363	0,0001	-0,1693	-0,1033
<b>P</b>	2014	5	6	-0,1668	0,0001	-0,1993	-0,1342
<b>P</b>	2014	1	6	-0,4333	0,0001	-0,4662	-0,4004

Legenda: P: área de Pediatria da prova; AnoProva: ano do curso em que a prova foi realizada;

Tabela 41 - Comparação de resultados da área de Ginecologia e Obstetrícia na Coorte de 2014

Grande Área	Coorte	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
		AnoProva	AnoProva			Limite Inferior	Limite Superior
<b>GO</b>	2014	1	2	0,0501	0,0027	0,0174	0,0829
<b>GO</b>	2014	2	3	-0,0003	0,9853	-0,0329	0,0323
<b>GO</b>	2014	3	4	-0,1058	0,0001	-0,1386	-0,0729
<b>GO</b>	2014	4	5	-0,2205	0,0001	-0,2532	-0,1878
<b>GO</b>	2014	5	6	-0,1582	0,0001	-0,1905	-0,1260
<b>GO</b>	2014	1	6	-0,4347	0,0001	-0,4672	-0,4021

Legenda: GO: área de ginecologia e obstetrícia prova; AnoProva: ano do curso em que a prova foi realizada;

## Apêndice 4 - Comparação entre resultados da prova interCoorte, por ano

Ano Prova	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
	Coorte	Coorte			Limite Inferior	Limite Superior
1	2010	2011	-0,07213	0,0001	-0,09869	-0,0456
1	2010	2012	-0,00763	0,5715	-0,03406	0,0188
1	2010	2013	-0,04457	0,0004	-0,06903	-0,0201
1	2010	2014	-0,08176	0,0001	-0,1062	-0,0574
1	2011	2013	0,02756	0,0405	0,001191	0,0539
1	2011	2014	-0,00963	0,4731	-0,03594	0,0167
1	2011	2012	0,0645	0,0001	0,0363	0,0927
1	2013	2014	-0,03719	0,0026	-0,06139	-0,013
1	2012	2013	-0,03694	0,0058	-0,06319	-0,0107
1	2012	2014	-0,07413	0,0001	-0,1003	-0,0479
2	2010	2011	0,04974	0,0003	0,02282	0,0767
2	2010	2012	0,0247	0,0739	-0,00239	0,0518
2	2010	2013	-0,00185	0,8865	-0,02731	0,0236
2	2010	2014	0,06579	0,0001	0,04074	0,0908
2	2011	2012	-0,02504	0,0791	-0,053	0,00291
2	2011	2013	-0,05159	0,0001	-0,07797	-0,0252
2	2011	2014	0,01605	0,2259	-0,00994	0,042
2	2012	2013	-0,02655	0,0499	-0,0531	-6,9E-06
2	2012	2014	0,04109	0,0021	0,01493	0,0673
2	2013	2014	0,06764	0,0001	0,04318	0,0921
3	2010	2011	-0,03197	0,0207	-0,05906	-0,00488
3	2010	2012	-0,07021	0,0001	-0,09698	-0,0434
3	2010	2013	-0,01447	0,2637	-0,03984	0,0109
3	2010	2014	-0,08137	0,0001	-0,1065	-0,0563
3	2011	2012	-0,03824	0,0068	-0,06592	-0,0106
3	2011	2013	0,0175	0,1925	-0,00883	0,0438
3	2011	2014	-0,0494	0,0002	-0,0755	-0,0233
3	2012	2013	0,05574	0,0001	0,02974	0,0817
3	2012	2014	-0,01117	0,3954	-0,03693	0,0146
3	2013	2014	-0,06691	0,0001	-0,09122	-0,0426
4	2010	2011	-0,03616	0,008	-0,06289	-0,00943
4	2010	2012	0,03153	0,0249	0,003981	0,0591
4	2010	2013	-0,03169	0,0144	-0,05705	-0,00632
4	2010	2014	-0,00051	0,9685	-0,02585	0,0248
4	2011	2012	0,06769	0,0001	0,03983	0,0956
4	2011	2013	0,004475	0,7328	-0,02123	0,0302
4	2011	2014	0,03565	0,0065	0,009967	0,0613

Ano Prova	Comparação		Diferença Estimada	Valor p	Intervalo de Confiança 95% da diferença	
	Coorte	Coorte			Limite Inferior	Limite Superior
4	2012	2013	-0,06322	0,0001	-0,08977	-0,0367
4	2012	2014	-0,03204	0,018	-0,05857	-0,00551
4	2013	2014	0,03118	0,0118	0,006923	0,0554
5	2011	2012	-0,07688	0,0001	-0,1035	-0,0502
5	2011	2013	-0,06987	0,0001	-0,09476	-0,045
5	2011	2014	-0,06322	0,0001	-0,08818	-0,0383
5	2010	2011	0,04698	0,0003	0,02146	0,0725
5	2010	2012	-0,0299	0,0255	-0,05613	-0,00366
5	2010	2013	-0,02289	0,0663	-0,04732	0,00154
5	2010	2014	-0,01624	0,194	-0,04074	0,00827
5	2012	2013	0,007008	0,5918	-0,01862	0,0326
5	2012	2014	0,01366	0,2973	-0,01204	0,0394
5	2013	2014	0,006652	0,5845	-0,0172	0,0305
6	2010	2011	-0,0059	0,6424	-0,03082	0,019
6	2010	2012	0,000864	0,9477	-0,02494	0,0267
6	2010	2013	-0,00388	0,7516	-0,0279	0,0201
6	2010	2014	-0,0453	0,0002	-0,06945	-0,0212
6	2011	2012	0,006765	0,6133	-0,01948	0,033
6	2011	2013	0,002023	0,8714	-0,02247	0,0265
6	2011	2014	-0,03939	0,0017	-0,06402	-0,0148
6	2012	2013	-0,00474	0,7143	-0,03014	0,0207
6	2012	2014	-0,04616	0,0004	-0,07168	-0,0206
6	2013	2014	-0,04142	0,0006	-0,06513	-0,0177

Legenda: Ano Prova: ano do curso em que a prova foi realizada

#### Apêndice 5 – Detalhamento de movimentação de estudantes entre quartis

Quartil Original	Coorte Quartil Final	2010	2011	2012	2013	2014
		n (%)*	n (%)	n (%)	n (%)	n (%)
Q1	Q1	13 (72,2)	10 (58,9)	11 (57,9)	6 (50,0)	5 (45,5)
	Q2	1 (5,5)	3 (17,6)	5 (26,3)	3 (25,0)	5 (45,5)
	Q3	4 (22,2)	3 (17,6)	3 (15,8)	1 (8,3)	0
	Q4	0 (0)	1 (5,9)	0 (0)	2 (16,7)	1 (9,0)

Quartil Original	Coorte		2010	2011	2012	2013	2014
	Quartil Final		n (%) <sup>*</sup>	n (%)	n (%)	n (%)	n (%)
<b>Q2</b>	Q1		8 (34,8)	5 (27,8)	2 (25,0)	8 (44,4)	5 (35,7)
	Q2		5 (21,7)	9 (50,0)	2 (25,0)	5 (27,8)	3 (21,4)
	Q3		8 (34,8)	2 (11,1)	2 (25,0)	5 (27,8)	5 (35,7)
	Q4		2 (8,7)	2 (11,1)	2 (25,0)	0 (0)	1 (7,2)
<b>Q3</b>	Q1		2 (11,1)	2 (11,1)	2 (15,4)	2 (13,3)	3 (23,1)
	Q2		6 (33,3)	6 (33,3)	4 (30,8)	4 (26,7)	4 (30,7)
	Q3		7 (38,9)	7 (38,9)	3 (23,0)	6 (40,0)	3 (23,1)
	Q4		3 (16,7)	3 (16,7)	4 (30,8)	3 (20,0)	3 (23,1)
<b>Q4</b>	Q1		2 (10,0)	0	0 (0)	1 (5,3)	1 (4,8)
	Q2		5 (25,0)	2 (15,4)	1 (9,1)	5 (26,3)	5 (23,8)
	Q3		1 (5)	3 (23,1)	3 (27,3)	3 (15,8)	6 (28,6)
	Q4		12 (60)	8 (61,5)	7 (63,6)	10 (52,6)	9 (42,8)

(\*) – porcentagem calculada baseada no n do quartil inicial



## ANEXOS

## Anexo 1 – Autorização da Comissão de Graduação – FMRP/USP



UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

Ofic. N° 005 SVGRAD/FMRP/26012022

MAH/ps

Prezado Professor,

A Comissão de Graduação **aprovou no mérito**, em sua 892ª reunião ordinária, realizada em 25 de janeiro de 2022, o projeto de pesquisa intitulado "**Avaliação de Estratégias em Educação Médica: Ensino Interprofissional e Desempenho Longitudinal do Estudante**", que será desenvolvido pela mestranda Mariana Passos Souza, do Programa de Pós-graduação em Ciências da Saúde Aplicadas ao Aparelho Locomotor da FMRP-USP, sob sua orientação.

A referida aprovação ocorreu com base no parecer do relator pela Comissão de Graduação, sendo que a aprovação ainda foi condicionada a que:

- Os proponentes da pesquisa se atentem para a importância da atenção a Lei Geral de Proteção de Dados sensíveis em vigor no país.
- O orientador e seus participantes, ao convidarem estudantes de graduação e docentes da FMRP-USP para participarem do Projeto de Pesquisa, ratifiquem que a participação dos mesmos será voluntária;
- O início das atividades somente ocorra após a aprovação pelo Comitê de Ética e Pesquisa (CEP);
- Os horários e locais designados para a realização da coleta de dados não prejudiquem as atividades acadêmicas curriculares;

Atenciosamente,

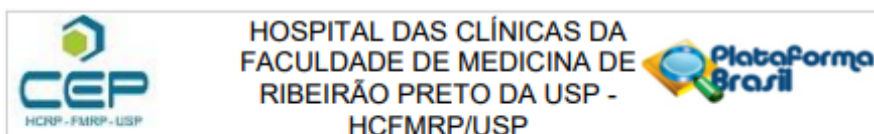
**Prof. Dr. Miguel Angelo Hyppolito**  
Presidente da Comissão de Graduação

Ilmo. Sr.

**PROF. DR. MARCELO RIBERTO**

Pesquisadora responsável

## Anexo 2 - Aprovação do Comitê de Ética – HCFMRP/USP

**PARECER CONSUBSTANCIADO DO CEP****DADOS DA EMENDA**

**Título da Pesquisa:** Avaliação de Estratégias em Educação Médica: Ensino Interprofissional e Desempenho Longitudinal do Estudante

**Pesquisador:** Mariana Passos de Souza

**Área Temática:**

**Versão:** 3

**CAAE:** 52905121.0.0000.5440

**Instituição Proponente:** Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo

**Patrocinador Principal:** Financiamento Próprio

**DADOS DO PARECER**

**Número do Parecer:** 6.017.157

**Apresentação do Projeto:**

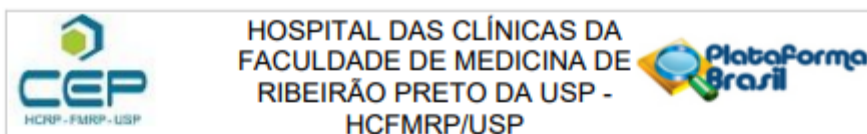
As diretrizes curriculares nacionais para os cursos de saúde (DCNs) determinam que os cursos de saúde estabeleçam diversas estratégias de ensino, com destaque recente para ferramentas de avaliação longitudinal e dos espaços de educação interprofissional. Tal abordagem possibilita a formação de profissionais mais plurais, capacitados e autônomos. O objetivo deste estudo é verificar de que forma experiências demandadas nas DCNs são ofertadas a estudantes de graduação da FMRP. Assim o estudo tem dois focos: 1- Avaliar as oportunidades interprofissionais disponibilizadas aos alunos dos cursos da FMRP-USP através de análise da grade curricular e de questionários semiestruturados; e 2- Avaliar cinco turmas de estudantes de medicina (2010 a 2014) quanto ao desempenho no teste de progresso. Com os resultados, espera-se discutir a presença e uso de práticas e estratégias de ensino atuais, a fim de buscar constantes melhorias na formação dos discentes.

**Objetivo da Pesquisa:**

**Objetivo Primário:**

Considerando que são duas temáticas importantes nas DCNs de 2014 e 2017, o objetivo deste trabalho é avaliar de que forma as experiências de Ensino Interprofissional e Desempenho Longitudinal do Estudante ocorreram nos cenários escolhidos. Isto é, identificar como a questão

**Endereço:** Hospital das Clínicas de Ribeirão Preto da Universidade de São Paulo/Campus Universitário, Prédio Central-  
**Bairro:** MONTE ALEGRE **CEP:** 14.048-900  
**UF:** SP **Município:** RIBEIRÃO PRETO  
**Telefone:** (16)3602-2228 **Fax:** (16)3633-1144 **E-mail:** cep@hcrp.usp.br



Continuação do Parecer: 6.017.157

do trabalho em equipe é abordada na formação de profissionais da saúde no contexto de graduação da FMRP-USP e como a aplicação de uma avaliação longitudinal tem sido explorada no curso de Medicina da FMRP-USP.

**Objetivo Secundário:**

Identificar através de abordagem aos documentos dos cursos de graduação da FMRP (ementas e roteiros de atividades didáticas disponíveis no sistema Júpiter) a presença dos temas "Multiprofissional/multidisciplinar" e/ou "Interprofissional/interdisciplinar" no respectivos currículos, seja em disciplinas teóricas, práticas ou de prática profissional (práticas supervisionadas, estágios, etc.); Identificar a percepção dos discentes, aprimorando/residentes e docentes da FMRP-USP sobre a presença de atividades teóricas e/ou práticas envolvendo os temas Multiprofissional/multidisciplinar" e/ou Interprofissional/interdisciplinar" e o impacto em sua formação; Descrever o processo de formação do consórcio na qual a FMRP-USP está inserido, descrever o modelo da prova e como ela se organiza e avaliar resultados do teste de Progresso aplicado na FMRP-USP, considerando o progresso das coortes por áreas temática ao longo dos anos.

**Avaliação dos Riscos e Benefícios:**

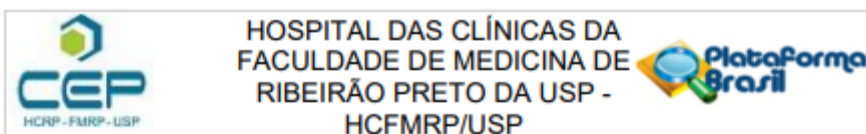
**Riscos:**

Entendendo que qualquer pesquisa com seres humanos pode gerar riscos ou desconfortos, neste caso pela relação de dependência em relação a instituição, nos comprometemos a deixar claro que a participação é livre, que os procedimentos para responder ao questionário através do Google Docs garantem a confidencialidade e a não identificação do respondente; sendo que igualmente para os procedimentos de entrevista a confidencialidade é garantida. Além disso, salientamos que os procedimentos são simples (responder a um questionário, após receber orientação), que o tempo necessário para participar da pesquisa é aproximadamente de 20 minutos, não oferecendo outros riscos ou prejuízos no desempenho de suas funções acadêmicas. Assumimos a responsabilidade e o compromisso ético de garantir que os resultados da pesquisa possam ser utilizados em benefício dos envolvidos.

**Benefícios:**

Os resultados obtidos podem contribuir para a melhor compreensão sobre a presença de

**Endereço:** Hospital das Clínicas de Ribeirão Preto da Universidade de São Paulo/Campus Universitário, Prédio Central-  
**Bairro:** MONTE ALEGRE **CEP:** 14.048-900  
**UF:** SP **Município:** RIBEIRÃO PRETO  
**Telefone:** (16)3602-2228 **Fax:** (16)3633-1144 **E-mail:** cep@hcrp.usp.br



Continuação do Parecer: 6.017.157

abordagens/atividades multi/interdisciplinares ao longo da formação na graduação, identificando lacunas, o que pode permitir melhorar o planejamento educacional no sentido de investir na formação para atuar em equipes. Da mesma forma, a descrição do Teste de Progresso como ferramenta avaliativa pode fornecer novos insights sobre seu uso e discussão ao longo da graduação dos cursos de saúde.

#### Comentários e Considerações sobre a Pesquisa:

Trata-se de emenda 1 ao projeto de pesquisa conforme PB\_INFORMAÇÕES\_BÁSICAS\_2102799\_E1.pdf de 16/04/2023.

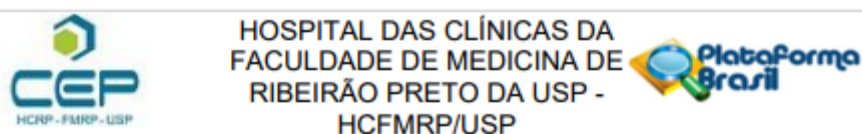
#### Justificativa da Emenda:

A intenção da resubmissão do projeto é sumariamente notificar o comitê para a expansão do uso de dados referente à porção do Teste de Progresso desse projeto. Inicialmente foram analisados os dados das turmas de estudantes ingressantes em medicina de 2010, 2011 e 2012 do primeiro ao sexto ano. Gostaríamos de expandir as análises para as turmas de alunos ingressantes de medicina de 2013 e 2014, também do primeiro ao sexto ano, avaliando também seu desempenho. O objetivo seria obter dados os dados mais atuais possíveis anteriores ao período pandêmico, a fim de possuir mais dados para comparação e expandir as percepções sobre a aplicação e resultado do Teste de Progresso na FMRP/USP. Nesse caso, nenhum processo metodológico seria alterado além da expansão dos dados, nessa porção do estudo os dados se manteriam agregados, sem possibilidade de identificação dos estudantes, da mesma forma que as demais turmas e não haveria contato com os estudantes das coortes expandidas, mantendo-se a dispensa do TCLE para essa porção do projeto. Assim, na nova Brochura de projeto submetida, foi alterada apenas a quantidade de grupos (destacada ao longo do texto com sublinhado) para análise e a versão do projeto no rodapé. E novo cronograma para o ano de 2023 foi adicionado.

#### Documentos:

- E\_Brochura\_v21.docx
- E\_Carta\_de\_Ementa.docx
- E\_Relatorio\_de\_Atividades\_Realizadas.docx
- E\_Cronograma\_v3.docx

**Endereço:** Hospital das Clínicas de Ribeirão Preto da Universidade de São Paulo/Campus Universitário, Prédio Central-  
**Bairro:** MONTE ALEGRE **CEP:** 14.048-900  
**UF:** SP **Município:** RIBEIRAO PRETO  
**Telefone:** (16)3602-2228 **Fax:** (16)3633-1144 **E-mail:** cep@hcrp.usp.br



Continuação do Parecer: 6.017.157

**Considerações sobre os Termos de apresentação obrigatória:**

Todos os termos da emenda 1 foram devidamente apresentados ao CEP.

**Conclusões ou Pendências e Lista de Inadequações:**

Diante do exposto e à luz da Resolução CNS 466/2012, a emenda 1 submetida em 16/04/2023, assim como o Projeto de pesquisa versão 2.1 de 16 de abril de 2023, podem ser enquadradas na categoria APROVADO. O CEP tomou ciência do Relatório parcial de 12/03/2023.

**Considerações Finais a critério do CEP:**

Emenda ao Projeto Aprovado: Tendo em vista a legislação vigente, devem ser encaminhados ao CEP, relatórios parciais anuais referentes ao andamento da pesquisa e relatório final ao término do trabalho. Qualquer modificação do projeto original deve ser apresentada a este CEP em nova versão, de forma objetiva e com justificativas, para nova apreciação.

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_2102799_É1.pdf	16/04/2023 13:58:12		Aceito
Projeto Detalhado / Brochura Investigador	E_Brochura_v21.docx	16/04/2023 13:58:00	Mariana Passos de Souza	Aceito
Outros	E_Carta_de_Ementa.docx	12/03/2023 23:53:48	Mariana Passos de Souza	Aceito
Outros	E_Relatorio_de_Atividades_Realizadas.docx	12/03/2023 23:52:54	Mariana Passos de Souza	Aceito
Cronograma	E_Cronograma_v3.docx	12/03/2023 23:37:10	Mariana Passos de Souza	Aceito
Outros	AprovacaoCG_DCN.pdf	28/02/2022 23:43:45	Mariana Passos de Souza	Aceito
Outros	Carta_de_resposta_questionamentosdo CEP_DCN.pdf	28/02/2022 23:41:21	Mariana Passos de Souza	Aceito
Cronograma	Cronograma_DNC_v2.pdf	28/02/2022 22:58:23	Mariana Passos de Souza	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Dispensa_TCLE_DCN_v2.pdf	28/02/2022 22:11:48	Mariana Passos de Souza	Aceito
Outros	Orcamento_UPC.pdf	28/10/2021 23:27:22	Mariana Passos de Souza	Aceito
Folha de Rosto	Folha_de_rosto.pdf	28/10/2021	Mariana Passos de Souza	Aceito

**Endereço:** Hospital das Clínicas de Ribeirão Preto da Universidade de São Paulo/Campus Universitário, Prédio Central-  
**Bairro:** MONTE ALEGRE **CEP:** 14.048-900  
**UF:** SP **Município:** RIBEIRÃO PRETO  
**Telefone:** (16)3602-2228 **Fax:** (16)3633-1144 **E-mail:** cep@hcrp.usp.br



HOSPITAL DAS CLÍNICAS DA  
FACULDADE DE MEDICINA DE  
RIBEIRÃO PRETO DA USP -  
HCFMRP/USP



Continuação do Parecer: 6.017.157

Folha de Rosto	Folha_de_rosto.pdf	23:26:25	Souza	Aceito
Outros	Questionarios_de_Pesquisa.docx	24/10/2021 23:01:08	Mariana Passos de Souza	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Dispensa_TCLE_DCN.docx	24/10/2021 22:57:36	Mariana Passos de Souza	Aceito
Orçamento	Orcamento_DCN.docx	24/10/2021 22:56:32	Mariana Passos de Souza	Aceito
Cronograma	Cronograma_DCN.docx	24/10/2021 22:56:18	Mariana Passos de Souza	Aceito

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

RIBEIRAO PRETO, 24 de Abril de 2023

Assinado por:

**MARCIA GUIMARÃES VILLANOVA**  
(Coordenador(a))

**Endereço:** Hospital das Clínicas de Ribeirão Preto da Universidade de São Paulo/Campus Universitário, Prédio Central-  
**Bairro:** MONTE ALEGRE **CEP:** 14.048-900  
**UF:** SP **Município:** RIBEIRAO PRETO  
**Telefone:** (16)3602-2228 **Fax:** (16)3633-1144 **E-mail:** cep@hcrp.usp.br



## Anexo 3 – Carta de Aceite de trabalho na 2023 AMEE Conference como apresentação oral



Mariana Passos de Souza  
<mariana.passos.souza@alumni.usp.br>

---

**Your research paper has been accepted for AMEE 2023 Conference.**

---

<AMEE2023@smartsubmissions.io>

qui., 2 de fev. 07:58

Responder para: <conferences@amee.org>

Para: <mariana.passos.souza@alumni.usp.br>



INCLUSIVE LEARNING ENVIRONMENTS TO TRANSFORM THE FUTURE

Dear Mariana Passos,

**Abstract ID:** 0482

**Abstract Title:** Progress Testing as an planning and self-assessment tool:  
student results and curriculum intervention

We are pleased to inform you that your abstract has been accepted as a Research Paper for AMEE 2023. Congratulations!

Please log in at <https://app.smartsubmissions.io/AMEE/AMEE2023/MySubmissions> using your [mariana.passos.souza@alumni.usp.br](mailto:mariana.passos.souza@alumni.usp.br) email address to confirm your acceptance. You will be required to read and accept the terms and conditions of acceptance, accept the presentation awarded and confirm the presenters registration ID. You will also have the opportunity to review and update your author information and add a presenter contact number. You are kindly requested to do this within 14 days of receipt of this email.

As part of our multi-step review process for Research Papers, all **419** submitted abstracts were first reviewed by medical education scientists from the AMEE Research Committee. Of these, **191** abstracts were sent out for further review. All of these abstracts were reviewed by at least two medical education scientists with expertise related to the methodology and/or content area of your abstract. Your abstract was one of only **48** that were selected for presentation as Research Papers at AMEE 2023.

Please find below the individual feedback received from the reviewers about your abstract, which you may want to keep in mind as you prepare your presentation.

**Feedback:****0482 Progress Testing as an planning and self-assessment tool: student results and curriculum intervention**

- This is a well-executed study and clearly written.
- This research paper investigated the knowledge growth from progress test to progress test. While research on progress tests and how students engage in these is clearly relevant, the approach of this study is lacking theoretical background and is not well embedded in current literature. The references mentioned in the reference list do not come back in the abstract itself, this would be a point for attention in the future. Further, I am wondering whether the conclusions the authors made can be drawn from the results, as the data are the progress test results 'only'. It is a pity that the authors did not include qualitative data to get more insight into the factors that influence the improvement or lack of improvement from progress test to progress test. Without these insights, it remains some speculation what caused the irregularities in progress test results.
- Please include a short description of the length of the undergraduate medical program at your institution. In the methods section it is unclear how the data from different cohorts was analyzed. Also please describe the residency test in more detail, including information on its timing in residency, format, and subjects covered. In the results section, please include any relevant statistics and confidence intervals when stating significant and non-significant results.

Research Paper presentations will (as in previous years) be 15 minutes long, followed by 5 minutes for questions/discussion. These presentations are all expected to include a strong focus on theory/methodology/method, and this will also be the focus of discussion led by the Session Moderators.

On behalf of the AMEE Research Committee, thank you again for your interest and we look forward to meeting you in Lyon and to hearing your presentation! All research paper presenters must register and pay the in-person attendance registration fee for AMEE 2023 no later than the 9th June. If you submitted the abstract but are not the named presenter of the accepted abstract, we would be grateful if you could ask the presenter to register and pay by the 9th June deadline. Abstracts with no named presenter registered will be withdrawn from the programme without further notification as agreed in the terms and conditions.

Registrations can be completed using the link - [Home - AMEE 2023 \(cvent.com\)](https://www.cvent.com)

Thank you

The AMEE 2023 Conference Team.