

THOMAS EUGENIO PORTES DE ALMEIDA

***NOVOS INSTRUMENTOS DE AVALIAÇÃO NO COMPONENTE
CURRICULAR DE TUTORIA EM CURSO DE MEDICINA POR
METODOLOGIA ATIVA***

Dissertação apresentada ao Programa de Pós-Graduação em Psicologia e Saúde da Faculdade de Medicina de São José do Rio Preto, como requisito para obtenção do Título de Mestre.

Orientadora: Profa. Dra. Maria Cristina de O. S. Miyazaki

SÃO JOSÉ DO RIO PRETO – SP

2018

Almeida, Thomas. E. P.

**Novos instrumentos de avaliação no componente curricular de tutoria em curso de medicina por metodologia ativa. São José do Rio Preto, 2018
71p**

Dissertação de Mestrado – Programa de Mestrado em Psicologia e Saúde – Faculdade de Medicina de São José do Rio Preto.

Orientadora: Profa. Maria Cristina de O. S. Miyazaki

1. Educação de Graduação em Medicina; 2. Avaliação Educacional.

THOMAS EUGENIO PORTES DE ALMEIDA

***NOVOS INSTRUMENTOS DE AVALIAÇÃO NO COMPONENTE
CURRICULAR DE TUTORIA EM CURSO DE MEDICINA POR
METODOLOGIA ATIVA***

BANCA EXAMINADORA

DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE MESTRE

Presidente e Orientadora: Maria Cristina Oliveira Santos Miyazaki

**Instituição: Faculdade de Medicina de São José do Rio Preto
FAMERP**

1ª Examinadora: Patrícia da Silva Fucuta

Instituição: Faculdade de Medicina de São José do Rio Preto

2ª Examinador: Ronaldo Gonçalves da Silva

Instituição: Faculdade de Medicina FACERES

São José do Rio Preto, 22 de Março de 2018

SUMÁRIO

Lista de Anexos.....	VIII
Lista de Tabelas.....	IX
Lista de Figuras.....	XI
Resumo.....	XII
Introdução.....	1
Objetivos	12
Método.....	13
Participantes.....	13
Obtenção dos dados.....	14
Critérios de inclusão e de exclusão da amostra.....	16
Descrição da amostra.....	17
Plano de Análise de Dados.....	20
Aspectos Éticos.....	23
Resultados.....	24
Efeito da Questão de Resposta Curta (QRC) em notas de provas somativas teóricas.....	24
Efeito da QRC nas médias finais.....	27
Agrupamento de resultados por gráficos de Forest Plot.....	29
Impacto da QRC em notas teóricas por estudo de amostras pareadas	34
Impacto da QRC em médias finais por estudo de amostras pareadas	37
Variação das notas em QRC ao longo dos dois semestres de aplicação	40
Avaliação dos efeitos da mudança de Salto Triplo para Duplo.....	40
Discussão	47
Conclusões.....	61
Considerações finais.....	62
Referências.....	64

Anexos.....	66
-------------	----

Dedico este trabalho a Prof. Dra. Patrícia Cury

AGRADECIMENTOS

Agradeço a Deus pela vida que recebi,

À minha família e à minha orientadora pela atenção e carinho

À instituição FAMERP pela estrutura e formação que me brindou

À Faculdade FACERES pela confiança, oportunidade e apoio para realização deste trabalho e da carreira docente.

LISTA DE ANEXOS

Anexo 1 – Parecer do Comitê de Ética	66
Anexo 2 – Relatório Parcial para Comitê de Ética.....	69
Anexo 3 – Parecer do Comitê de Ética sobre Relatório Final.....	72
Anexo 4 – Pirâmide de Miller.....	74

LISTA DE TABELAS

Tabela 1: Comparação entre notas de provas somativas teóricas entre 2º semestre de 2015 (sem QRC) com 1º semestre de 2016 (com QRC).....	24
Tabela 2: : Comparação entre notas de provas somativas teóricas entre 2º semestre de 2015 (sem QRC) com 2º semestre de 2016 (com QRC).	25
Tabela 3: Comparação entre notas de provas somativas teóricas entre 1º semestre de 2016 (com QRC) com 2º semestre de 2016 (com QRC).....	26
Tabela 4 Comparação entre médias finais do 2º semestre de 2015 (sem QRC) e do 1º semestre de 2016 (com QRC).....	27
Tabela 5: Comparação entre médias finais do 2º semestre de 2015 (sem QRC) e do 2º semestre de 2016 (com QRC).....	28
Tabela 6: Comparação entre médias finais do 1º semestre de 2016 (com QRC) e do 2º semestre de 2016 (com QRC).....	29
Tabela 7: Evolução das notas de provas somativas teóricas entre o 2º semestre de 2015 (sem QRC) e o 1º semestre de 2016 (com QRC). Dados pareados por alunos.....	34
Tabela 8: Evolução das notas de provas somativas teóricas entre o 2º semestre de 2015 (sem QRC) e o 2º semestre de 2016 (com QRC). Dados pareados por alunos.....	35
Tabela 9: Evolução das notas de provas somativas teóricas entre o 1º semestre de 2016 (com QRC) e o 2º semestre de 2016 (com QRC). Dados pareados por alunos, separados por turmas.....	36
Tabela 10: Evolução das médias finais entre o 2º semestre de 2015 (sem QRC) e o 1º semestre de 2016 (com QRC). Dados pareados por alunos, separados por turmas.	37

Tabela 11: Evolução das médias finais entre o 2º semestre de 2015 (sem QRC) e o 2º semestre de 2016 (com QRC). Dados pareados por alunos, separados por turmas.38

Tabela 12: Evolução das médias finais entre o 1º semestre de 2016 (com QRC) e o 2º semestre de 2016 (com QRC). Dados pareados por alunos, separados por turmas.39

Tabela 13: Variação da Nota da QRC ao longo de dois semestres seguidos de aplicação, separado por etapas.....40

LISTA DE FIGURAS

Figura 1: Instrumentos de avaliação na disciplina de tutoria da FACERES.....	6
Figura 2: Elementos de avaliação e peso das avaliações em 2014, 2015 e 2016.....	7
Figura 3: Organização do ano letivo, divisão em semestres e módulos.....	15
Figura 4: Descrição da amostra.....	19
Figura 5: Grafico de forest plot demonstrando efeitos observados nas notas de provas somativas teóricas no primeiro e segundo semestre após introdução da QRC.....	32
Figura 6: Grafico de forest plot demonstrando efeitos observados nas medias finais no primeiro e segundo semestre após introdução da QRC.....	33
Figura 7: Distribuição das notas em provas somativas teóricas no segundo semestre de 2015, estratificando pelo resultado obtido na prova Salto Triplo.....	42
Figura 8: Distribuição das notas em provas somativas teóricas no primeiro semestre de 2016, estratificando pelo resultado obtido na prova Salto Triplo.....	43
Figura 9: Distribuição das notas em provas somativas teóricas no segundo semestre de 2016, estratificando pelo resultado obtido na prova Salto Triplo.....	45
Figura 10: Gráfico de aprovações e reprovações na prova Salto Duplo e provas Salto Triplo ao longo dos três semestres de estudo.....	46

RESUMO

Introdução: A formação do médico tem mudado ao longo do tempo, exigindo a expansão das metodologias ativas de ensino, como o aprendizado baseado em problemas. A avaliação dos discentes é uma importante etapa do processo de aprendizagem e representa um desafio para o corpo docente. Neste sentido, duas mudanças foram feitas nos instrumentos de avaliação do componente curricular de tutoria: 1) a aplicação de uma prova dissertativa de respostas curtas aplicada antes da sessão de fechamento da tutoria, e 2) a mudança da avaliação de raciocínio crítico denominada “Salto triplo” para “Salto Duplo”. **Objetivos:** O presente trabalho visa avaliar o impacto destas novas avaliações nas notas dos alunos assim como a fidedignidade e a validade destes instrumentos. **Metodologia:** Pesquisa quantitativa, retrospectiva, estudando o impacto das mudanças efetuadas nos sistemas de avaliação em todas as notas do componente curricular de Tutoria, do segundo semestre de 2015 e de 2016, dos alunos do curso de medicina da Faculdade FACERES de São José do Rio Preto. **Resultados:** 1) De forma global, após a introdução da questão de resposta curta ocorreu aumento nas notas das provas somativas teóricas e médias finais. 2) A prova Salto Duplo manteve a capacidade de discriminar alunos com bom desempenho daqueles com desempenho inferior. **Conclusões:** Os novos instrumentos de avaliação se mostraram eficazes em melhorar o desempenho dos alunos. A prova “Salto Duplo” se mostrou uma alternativa equivalente ao “Salto Triplo”.

Palavras chave DECS-BVS

1. Educação de Graduação em Medicina; 2. Avaliação Educacional; 3. Aprendizagem Baseada em Problemas; 4. Aprendizagem Ativa.

ABSTRACT

Introduction: Medical teaching has changed over time, requiring the expansion of active learning methodologies, such as problem-based learning. Student assessment is an important stage in the learning process and represents a challenge for medical schools. In this direction, two changes were made in the assessment instruments of the tutorial component: 1) the application of short answers questions applied before the closing section of the tutoring meeting, and 2) a change in the critical reasoning evaluation called "Triple Jump" to "Double Jump". **Objectives:** This study aims to evaluate the impact of these new assessments on students' grades as well as the reliability and validity of these instruments. **Methodology:** Quantitative, retrospective research, studying the impact of the changes made in the assessment methods in all the grades obtained in the Tutorial teaching component, by the second semester of 2015 and the complete 2016 year of medical students of the "Faculdade FACERES de São José do Rio Preto". **Results:** 1) Overall, after the introduction of the Short Answer Question, there was an increase in the scores of the summative tests and the final averages scores of students. 2) The Double Jump test maintained the ability to discriminate well-performing students from those with lower performance. **Conclusions:** The new evaluation tools have proved effective in improving students' performance. The Double Jump test proved to be an equivalent to Triple Jump.

Key words DECS-BVS

1. Education, Medical, Undergraduate; 2. Educational Measurement; 3. Problem-Based Learning; 4. Active Learning Methodologies

INTRODUÇÃO

Mudança do ensino tradicional para o PBL

Metodologias ativas de ensino, como o PBL (problem based learning) representam uma mudança no paradigma educacional; outrora focado no professor, hoje é centrado no aluno (Gwee, 2009). Metodologias ativas treinam o estudante para o modo como trabalhará ao longo de sua vida profissional: resolvendo problemas, procurando por respostas e direcionando sua tomada de decisões, individualmente ou em grupo (Wang, Li, Pang, Liang, & Su, 2016).

O ensino médico tradicional é altamente focado nas disciplinas e retarda o momento no qual o aluno enxergará as inter-relações entre as disciplinas, dificultando a percepção de praticabilidade dos conteúdos aprendidos e reduzindo a eficiência do aprendizado (Gwee, 2009). As principais limitações do método tradicional são: 1) a sobrecarga de conteúdo; 2) o desenvolvimento de um hábito frágil de estudo, baseado em memorizar, lembrar e regurgitar (traduzido livremente do inglês: “*memorize, recall, regurgitate*”); 3) o fato de incentivar a dependência do aluno frente ao professor em relação ao quê, quando e como estudar. O método tradicional também está associado à atenção inadequada da instituição perante o desenvolvimento cognitivo de habilidades complexas como pensamento crítico, raciocínio e competência para resolver problemas, assim como habilidades do domínio interpessoal, como comunicação, trabalho em equipe e liderança (Gwee, 2009).

Na década de 60, a Universidade de McMaster, em Hamilton, Ontario, Canadá, mudou um paradigma educacional estabelecido há anos: distanciou-se do método de ensino tradicional em busca de uma forma de fazer o aluno se envolve ativamente no processo de aprendizagem. Mais precisamente, em 1969 o novo currículo foi implantado em McMaster,

introduzindo a metodologia PBL (Gwee, 2009). “O PBL é baseado na crença de que aprender é mais eficiente quando o aluno está ativamente envolvido e aprende no contexto onde o conhecimento está sendo usado” (Boud & Feletti, 1998).

O processo de ensino no PBL se dá por meio da resolução de problemas. No componente curricular de tutoria, apresenta-se um problema em um determinado dia, chamado de “abertura”; normalmente esta sessão utiliza um caso clínico como gatilho para discussão. Os alunos colaboram entre si para analisar o problema à luz de seus conhecimentos prévios e elaborar perguntas que direcionem seu estudo, guiados por um tutor que previamente conhece os objetivos de compreensão desta reunião. O processo se encerra no “fechamento”, uma reunião realizada três dias depois, na qual os alunos, em grupo, discutem os objetivos de compreensão que estudaram individualmente nos dias entre a “abertura” e o “fechamento” (Dolmans, Loyens, Marcq, & Gijbels, 2016; Wood, 2004).

O processo pode ser resumido em sete passos, conforme traduzidos por Santos, Salgado, Barreto, Martins, & Dores (2010).

“1) clarificar termos e conceitos não imediatamente compreensíveis; 2) definir o problema; 3) analisar o problema; 4) elaborar um inventário sistematizado das explicações/hipóteses inferidas a partir do passo 3 e proceder à divisão de tarefas que cada elemento deve desempenhar; 5) formular metas de aprendizagem; 6) compilar informação adicional fora do grupo (estudo individual) e 7) sintetizar e testar a nova informação adquirida”

O ensino é complementado por uma estrutura curricular que utiliza o tema central deste “caso problema” como direcionamento para as demais atividades do calendário do aluno, oferecendo simultaneamente disciplinas de anatomia, patologia, fisiologia, bioquímica, habilidades médicas, etc. (Gwee, 2009).

Processo de Avaliação no PBL

A resolução de problemas, como método de ensino em saúde, tem se mostrado mais eficaz que currículos tradicionais na promoção de um aprendizado de longa duração e na motivação intrínseca dos alunos (Dolmans et al., 2016; Walker & Leary, 2009; Wood, 2004). No entanto, para que este processo se desenvolva adequadamente, é necessário que o aluno participe ativamente das sessões tutoriais e desenvolva um estudo com profundidade e correto direcionamento. Mais do que a simples memorização de conceitos-chave, é preciso que ocorra um processo semelhante à “digestão” do problema. Para tanto, a participação ativa dos alunos deve ser avidamente incentivada pelos docentes (Wood, 2004), atingindo o ideal de “*ensino aos pares e aprendizado aos pares*”, num modo chamado de “*todos ensinam, todos aprendem*”(Gwee, 2009).

O processo de avaliação é um coadjuvante da estratégia de ensino. Segundo Epstein (2007), as provas são ferramentas para aperfeiçoar as capacidades dos aprendizes, atuando como um mecanismo de incentivo e de “feedback” durante o curso de medicina, e neste caso, dentro do componente curricular de tutoria.

O processo de avaliação pode ocorrer de duas formas: a primeira, chamada “somativa”, é realizada no final do curso, período ou unidade de ensino e mostra uma nota ou conceito dados pelo docente frente ao conteúdo ou competência avaliada. Tem como finalidade classificar os alunos ou definir o aproveitamento na unidade para progressão no curso. A segunda, chamada “formativa”, tem como finalidade fornecer dados para aperfeiçoar o processo de ensino-aprendizagem. Permite um “feedback” para os docentes sobre o rendimento de seu trabalho e para os alunos, servindo de recurso de ensino e fonte de motivação (Haydt, 2002). A avaliação formativa também pode ser chamada de “avaliação da aprendizagem” e é um processo contínuo, horizontal, que acompanha o aluno no decorrer do

curso, servindo como um guia para a auto-avaliação das lacunas de competências, ou seja, para tornar ciente a diferença entre o desejado e o encontrado no âmbito acadêmico (Tavakol & Dennick, 2017).

No contexto do PBL as avaliações devem medir: 1) o progresso dos estudantes no conhecimento específico da área curricular abordada, 2) o amadurecimento na compreensão e execução da rotina de resolução de problemas (Savery, 2006). No entanto, uma grande discussão nasce em torno de quais métodos de avaliação devem ser utilizados, qual sua frequência, peso, diversificação, etc. O uso simultâneo de vários métodos é capaz de melhorar a eficácia do processo avaliativo como um todo (Schuwirth & van der Vleuten, 2004), o que reforça a importância de incluir avaliações com questões abertas no componente curricular de tutoria. (Norman, Case, & Swanson, 1996).

Funcionamento do PBL na FACERES / Problematização.

A componente curricular de tutoria da Faculdade de Medicina FACERES de São José do Rio Preto avaliava seus alunos, até final de 2014, por meio de três instrumentos: uma prova somativa de múltipla escolha, uma avaliação dissertativa chamada “Salto Triplo” e um questionário para avaliar a participação dos alunos durante as sessões tutoriais (preenchido pelo docente).

Cada semestre é dividido em três módulos, com temas diferentes e um conjunto de avaliações para cada módulo, sendo as notas de cada módulo independentes entre si.

Devido ao alto grau de subjetividade intrínseca à avaliação por participação, que gerava muita controvérsia entre alunos e tutores, em 2015 foi decidido que este último instrumento não seria mais realizado. Deste modo, em 2015 toda nota de tutoria passou a ser

definida, no final do módulo, pelas provas somativas. O que se observou a seguir foi um desestímulo geral dos alunos para estudar individualmente e participar das sessões de tutoria, levando a um grau elevado de absenteísmo (“como não vale nota, não vou e não estudo”, diriam os alunos) e perda de rendimento não apenas do aluno, mas do grupo tutorial como um todo.

O PBL enfatiza a integração do conhecimento e das habilidades e não o aprendizado em pequenas partes, peça-a-peça. Para desencadear este fenômeno, a fase de estudo individual tem um papel chave. Cada aluno busca recursos para estudar de forma autônoma e aprende por diferentes ângulos, propiciando riqueza na discussão e desenvolvimento para o coletivo. (Dolmans et al., 2016)

Para o início do ano letivo de 2016 duas mudanças foram implementadas nos sistemas de avaliação da componente curricular de tutoria: a primeira delas foi a introdução de uma nova prova, chamada de Questão de Resposta Curta (QRC), e a segunda foi uma mudança na prova dissertativa, antes chamada de Salto Triplo, que passou a ser nomeada como “Salto Duplo” (a Figura 1 resume os instrumentos de avaliação utilizados em 2016).

A nova modalidade de avaliação, chamada de Questão de Resposta Curta (QRC) foi apelidada informalmente de “Kinder Ovo”, em alusão ao caráter “surpresa” da mesma. Introduzida para contornar o desestímulo ao estudo individual, essa avaliação seria aplicada no dia do “fechamento” do problema, antes de iniciar a discussão, com o objetivo de verificar e estimular o prévio estudo do aluno. A avaliação, aqui denominada como QRC, compreende uma ou duas questões abertas de resposta curta, às quais são atribuídas uma das três notas: certo, meio certo ou errado (considerando que a questão com conceito certo equivale numericamente a 1, a meio certo 0,5 e a errada a zero). A média aritmética das notas destas questões compõe a nota do aluno conforme explicado na *figura 2*.



FIGURA 1:

Instrumentos de avaliação do componente curricular de tutoria em 2016

*QRC = Questão de Resposta Curta

A prova Salto Triplo já era rotineiramente aplicada no componente curricular de tutoria do curso de medicina da FACERES, composta por um exercício em três tempos (daí seu nome).

No primeiro tempo ocorre a apresentação de um problema ou caso clínico ao aluno, que faz a leitura cuidadosa e um “brainstorming” sobre os temas envolvidos no problema; em seguida os alunos elaboram um fluxograma para mostrar possíveis explicações para o caso utilizando seus conhecimentos prévios e redigem duas perguntas que ajudariam a elucidar o problema em questão.

No segundo tempo da prova o aluno vai até a biblioteca da faculdade para fazer pesquisas que o embasem para a resolução do caso clínico, especificamente adquirindo conhecimento

para responder às perguntas que elaborou; após 40 minutos de estudo o aluno retorna à sala de aula onde se iniciará o terceiro tempo de avaliação, onde o discente arguirá com o professor / avaliador que atribuirá um conceito (suficiente ou insuficiente) para tal exercício.

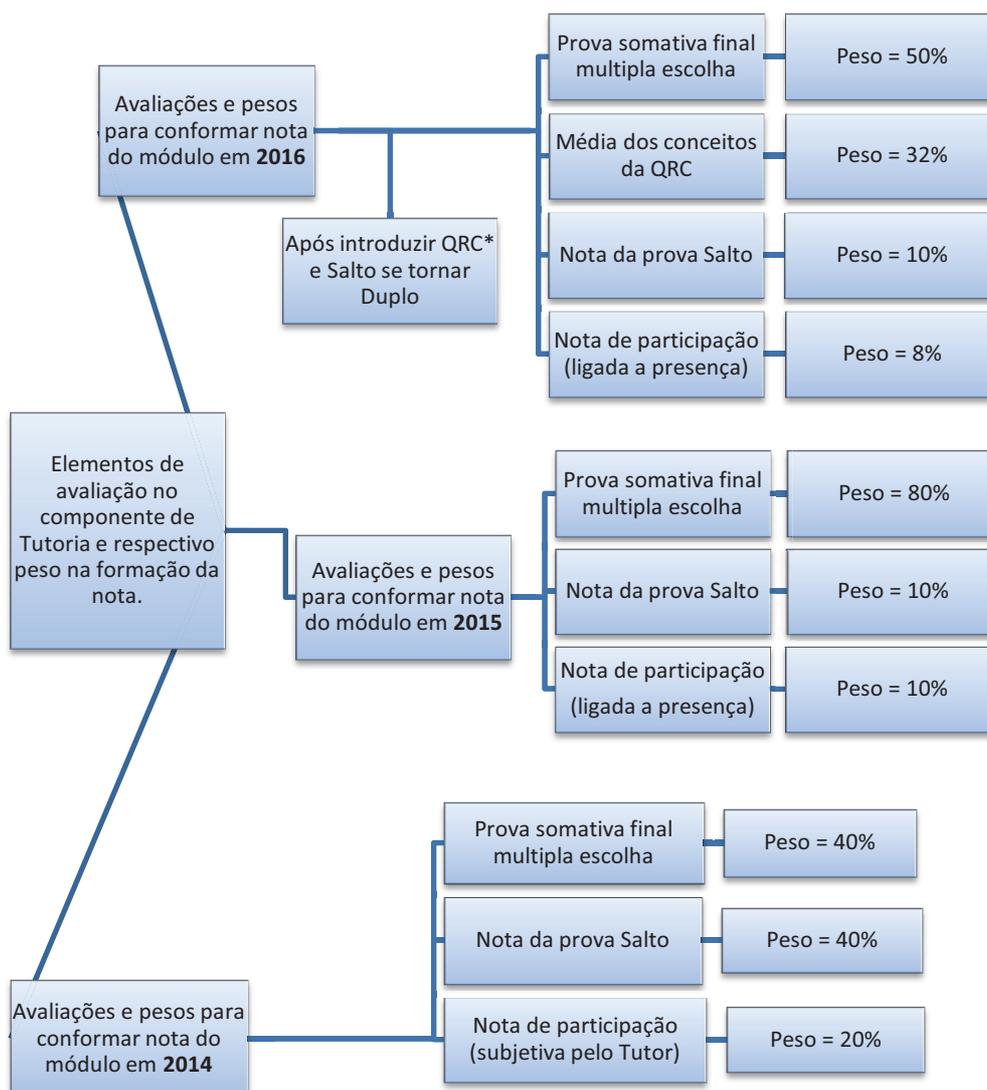


FIGURA 2

Elementos de avaliação e peso das avaliações em 2014, 2015 e 2016 (no componente curricular de tutoria)

*QRC = Questão de Resposta Curta

A prova do tipo Salto Triplo teve sua origem na década de 60 na universidade de McMaster como parte dos primeiros exercícios utilizados no nascimento do PBL (Friedman, Donaldson, & Vantsevich, 2016). Esta avaliação consegue ter um nível elevado na “Miller’s Pyramid of Professional Competence” (Miller, 1990) (Reproduzida no anexo quatro) permitindo uma observação mais profunda quando/ comparada a provas de múltipla escolha, ao estimular análise, pensamento crítico e resolução de problemas (Navazesh, Rich, Chopiuk, & Keim, 2013). Esta avaliação é também conhecida como TJE, sigla de “triple jump examination” ou “three-step examination” na literatura mundial.

A mudança de Salto Triplo para Salto Duplo foi proposta como uma simplificação na estrutura desta avaliação. O objetivo primário foi eliminar a etapa de arguição, sabidamente estressante para os alunos e corpo docente. Neste novo formato são avaliados apenas o fluxograma e as perguntas formuladas pelo aluno. Tal mudança foi inovadora e sem precedentes na literatura corrente (Almeida, Fucuta, Pacca, Ricomini, & Cury, 2017).

Embasamento para uso de QRC.

O uso de questões abertas de resposta curta já foi testado, em outras provas, quanto a sua fidedignidade, sendo que quando bem elaboradas podem ter índices tão bons quanto às questões de múltipla escolha (Rademakers, Ten Cate, & Bär, 2005). A vantagem em relação às questões fechadas está na redução da possibilidade do aluno acertar sua resposta aleatoriamente. Já a principal dificuldade encontrada é a subjetividade para a sua correção, o que, segundo Rademarkers e colaboradores (2005), pode ser reduzida ao se entregar para os corretores modelos para respostas e oferecer instruções claras sobre o que considerar.

O desenvolvimento e a implantação da QRC teve como objetivo estimular o estudo dos alunos entre as sessões de abertura e fechamento, além de poder avaliar o empenho individual do aluno de forma menos subjetiva do que uma avaliação conceitual pela participação durante o fechamento. Esta abordagem, teoricamente, reforçaria a atuação do tutor como “Coach” dos alunos, retirando seu papel de avaliador, o que poderia prejudicar sua relação com os discentes (Allareddy, Havens, Howell, & Karimbux, 2011; Papinczak, 2010; Wang et al., 2016).

A impressão obtida pelos tutores de diversas etapas do curso após a implantação da prova QRC foi de melhora do desempenho dos alunos, uniformização da profundidade de estudo e discussões mais ricas. Resta saber se a prova QRC é fidedigna, se os resultados obtidos pelos alunos nesta avaliação serão proporcionais às notas que obterão na avaliação somativa do módulo, e se as notas e desempenho dos alunos realmente serão impactadas por esta nova modalidade de avaliação.

Cabe ressaltar que nem a introdução da QRC nem as alterações de Salto Triplo para Salto Duplo foram feitas como protocolo experimental, e sim como alterações pedagógicas na estrutura do curso de medicina da faculdade FACERES. Alterações nos instrumentos de avaliação, frequência de aplicação, pesos para ponderamento da nota, entre outras, são permitidas livremente pelas normas correntes do Ministério da Educação, que garante à instituição autonomia na escolha da forma de avaliar os discentes e docentes. Inclusive, estas alterações foram anteriores ao início deste projeto de pesquisa, que será classificado como exclusivamente retrospectivo.

Justificativa implementação da QRC e da prova SALTO.

Não existe uniformidade ou consenso nos métodos de avaliação utilizados pelas escolas médicas que utilizam PBL (Nendaz & Tekian, 1999), cabe a cada uma estabelecer os métodos que melhor se adequem às suas necessidades. Os sistemas de avaliação tem sido explorados pela literatura, buscando estabelecer a utilidade de cada tipo de prova dentro das diversas metodologias ativas, entre elas o PBL. Dentro deste cenário, este estudo visou demonstrar um novo método de avaliação, estudando sua aplicabilidade e impacto nas notas do componente curricular de tutoria. Segundo Tavakol & Dennick (2017) “deve-se enfatizar que a avaliação é uma medida do aprendizado e que compreender os fatores que influenciam na acurácia, fidedignidade e validade deste processo de medição faz-se essencial para a criação de avaliações de qualidade”.

Toda forma de avaliação deve ter três critérios fundamentais: 1) ser válida, ou seja, perguntar o que se deseja saber, sendo que esta pergunta deve refletir o objetivo de ensino do que está sendo avaliado, de forma clara, visando observar raciocínio e compreensão sobre o caso ou problema apresentado; 2) ser fidedigna, ou confiável, o que significa que a prova eficientemente separa o candidato ou aluno que consegue atingir o objetivo exigido de quem não consegue e, finalmente; 3) ser aplicável, onde o custo de elaboração e aplicação são aceitáveis para o objetivo estabelecido (Haydt, 2002; Tabish, 2008).

Dentro do curso de medicina, entre as diferentes disciplinas e etapas, os instrumentos utilizados variam conforme o objetivo da avaliação. Para ter uma psicomетria eficiente do aluno é necessário que o mesmo passe por diversos tipos de avaliação; cabe ao curso oferecer esta pluralidade de opções, permitindo que alcance seu potencial sem ser prejudicado pelo fato de ser avaliado por um método distinto do qual tem mais familiaridade ou segurança, visto que características intrínsecas de cada tipo de prova interferem no seu resultado

(Tabish, 2008). Neste contexto, a existência simultânea de provas QRC e a prova Salto Duplo visam colaborar para a multiplicidade de tipos de provas, em conjunto com a prova final de perguntas múltipla escolha (ver *figura 1*).

A prova QRC é um tipo novo de avaliação no curso de medicina da FACERES, que visa exercer o papel de uma avaliação formativa, ou seja, avaliar, estimular o aluno a estudar para as sessões de tutoria e servir como diagnóstico para os tutores sobre as dificuldades encontradas pelo aluno. A prova Salto Duplo visa proporcionar ao discente uma avaliação dissertativa, na qual possa demonstrar seu raciocínio, capacidade de interpretação e lógica, assim como tenta servir como estímulo ao aguçar a curiosidade e autocrítica.

OBJETIVOS

- Avaliar o efeito da inclusão da prova QRC nas notas das provas somativas finais e médias finais no componente curricular de tutoria;
- Descrever a capacidade de discriminação da prova Salto Duplo, comparando-a com a prova Salto Triplo;
- Verificar a aplicabilidade destes novos métodos de avaliação dentro do componente curricular de tutoria.

MÉTODOS

Delineamento da pesquisa

Caracteriza-se como um estudo descritivo, retrospectivo, com análise quantitativa.

Local

O estudo foi desenvolvido na Faculdade de Medicina FACERES de São José do Rio Preto, SP

Participantes

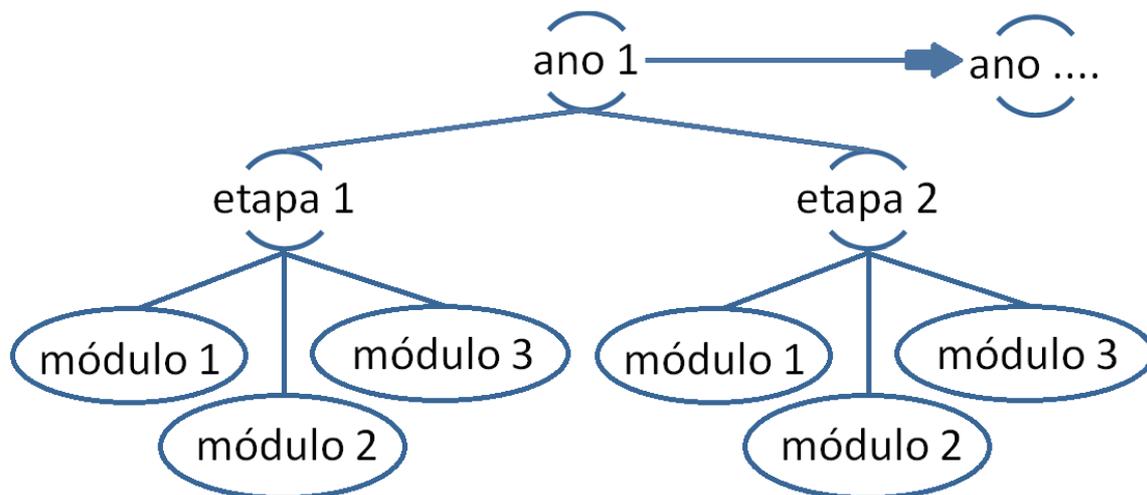
Foram incluídas neste estudo as notas e os índices de presença de 504 alunos da Faculdade de Medicina FACERES, contendo dados referentes aos alunos que cursaram a componente curricular de tutoria no segundo semestre do ano de 2015 e de todo o ano de 2016, portanto os alunos das turmas 1 a 7, permitindo avaliar dados das etapas 1 a 8. (Os índices de presença foram coletados unicamente para definir critérios de inclusão ou exclusão da amostra.)

Os dados foram coletados do sistema de notas da Faculdade FACERES (NOTAMED®), de onde é possível extrair planilhas compatíveis com Excel® contendo um número de identificação do aluno, notas das provas que o aluno fez durante a graduação, notas subjetivas ou recebidas de qualquer outro tipo de avaliação que o aluno tenha feito, índices de presença do aluno em cada uma das disciplinas da faculdade, médias finais nas disciplinas e notas de provas substitutivas ou provas do tipo exame final.

Obtenção de dados

Foram coletados do sistema NOTAMED® as notas de tutoria do segundo semestre de 2015 e de todo o ano letivo de 2016.

As tabelas exportadas do programa NOTAMED® permitiram identificar a qual turma o aluno pertencia, suas notas nas avaliações, divididas conforme disciplina, etapa e módulo. O termo etapa se refere ao semestre que o aluno está cursando, sendo que o curso de medicina é composto por 6 anos, portanto 12 semestres ou, conforme aqui são chamados, 12 etapas. O termo módulo refere-se a uma subdivisão da etapa, cada etapa é composta por três módulos, que não têm relação entre si. Como exemplo, no componente curricular de tutoria, a etapa 2 contém os módulos (1) funções orgânicas, (2) Relação Parasito Hospedeiro e meio ambiente e (3) Práticas de Saúde Pública. (ver *figura 3*)

**FIGURA 3**

Organização do ano letivo, divisão em semestres (etapas) e módulos.

Os dados que possibilitassem identificar o aluno que fez parte da amostra foram excluídos das planilhas assim que foram extraídas do sistema, com a finalidade de evitar a exposição do mesmo no caso perda, extravio, roubo e uso indevido dos dados aqui estudados. Contudo, para permitir análise pareada dos dados, os alunos receberam um código numérico único.

Critérios de inclusão da amostra

Foram analisadas todas as notas e índices de presença dos alunos que cursaram a componente curricular de tutoria durante o segundo semestre de 2015 e todo ano letivo de 2016.

Critério de exclusão da amostra

Foram excluídos da análise:

- Conjunto de notas referente a alunos que faltaram mais que 25% das sessões tutoriais.

O objetivo desta exclusão recai na necessidade de padronização dos dados dos alunos estudados. Como o intuito da pesquisa é observar o efeito da QRC, foi entendido que alunos que reprovaram por faltas forneceriam dados que não contribuiriam para a compreensão do desfecho relacionado a implementação da QRC, além de incluir um viés relacionado à baixa aderência às técnicas pedagógicas implementadas pela instituição, inclusive a aqui estudada.

- Notas referentes a provas tabuladas com o valor zero.

Este valor poderia representar provas que foram entregues em branco ou que tenham sido anuladas por fraude de qualquer tipo. A base de dados avaliada não contém marcações específicas que definam o motivo pelo qual a nota foi igual a zero, considerando que podem representar qualquer situação de desclassificação.

- Notas de provas substitutivas ou de notas de “exame final” não foram utilizadas

O motivo da exclusão destas informações recai sobre a impossibilidade de comparar os resultados destas ferramentas alternativas de avaliação, visto que as provas substitutivas

ou de “exame final” são preparadas exclusivamente para o aluno ou população de alunos específica, sem repetição da aplicação, o que dificulta a avaliação psicométrica do instrumento em questão.

Obs: o termo “exame final” se refere à uma prova aplicada ao aluno que obteve conceito inferior a 7 na média final da disciplina.

Seleção da amostra

Todas as notas de alunos que se enquadraram nos critérios de inclusão e que não apresentaram fatores impeditivos discriminados nos critérios de exclusão foram incluídas nas análises deste trabalho.

Descrição da amostra

Separando apenas o período do segundo semestre de 2015 até o segundo semestre de 2016 encontramos 3505 módulos (divisão da etapa ilustrada na *figura 3*), dos quais participaram 504 alunos. Estes dados se distribuem nos semestres avaliados da seguinte forma:

2º semestre de 2015 = 1029 módulos, dos quais participaram 344 alunos, nas etapas 1 a 7.

1º semestre de 2016 = 1252 módulos, dos quais participaram 419 alunos, nas etapas 1 a 8.

2º semestre de 2016 = 1224 módulos, dos quais participaram 419 alunos, nas etapas 1 a 8.

Prosseguindo com a análise, foram retirados os módulos nos quais os alunos apresentam índice de presença inferior a 75% ($<75\%$):

2º semestre de 2015 = retirados 46 módulos, resultando na supressão de 10 alunos.

1º semestre de 2016 = retirados 63 módulos, resultando na supressão de 7 alunos.

2º semestre de 2016 = retirados 35 módulos, resultando na supressão de 13 alunos.

Em sequência foram retirados os dados de alunos com nota igual a zero:

2º semestre de 2015 = retirados 2 módulos, o que não suprimiu nenhum aluno.

1º semestre de 2016 = retirados 3 módulos, o que não suprimiu nenhum aluno.

2º semestre de 2016 = retirados 3 módulos, o que não suprimiu nenhum aluno.

Retirando os dados descritos acima restaram para análise:

2º semestre de 2015 = 981 módulos, dados de 334 alunos.

1º semestre de 2016 = 1186 módulos, dados de 412 alunos.

2º semestre de 2016 = 1185 módulos, dados de 406 alunos.

Esta descrição da amostra e aplicações de critérios de exclusão está esquematizada na figura 4.

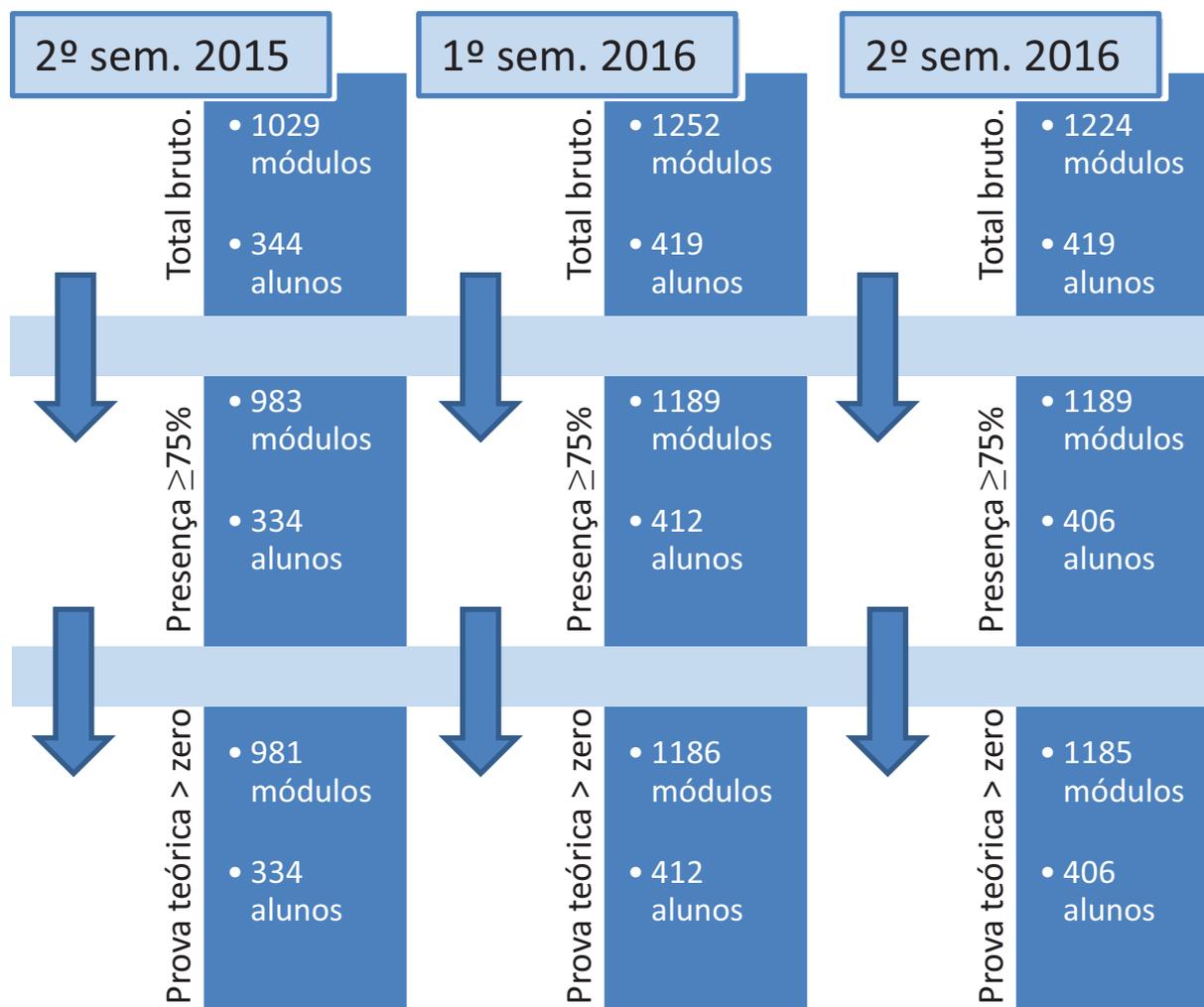


FIGURA 4

Descrição da amostra

PLANO DE ANÁLISE DOS DADOS

Metodologia para análise estatística dos dados

Foi realizada análise quantitativa das notas dos alunos antes e após a implementação da QRC, período que coincide com a mudança da prova salto de Triplo para Duplo. Os dados obtidos foram transferidos para planilhas digitais (Excel® Microsoft Corporation) e submetidos a análise estatística utilizando o software SPSS ® IBM.

Em todas as análises, excetuando os estudos por Gráficos de Forest Plot, os dados foram inicialmente observados por estatística descritiva da amostra; caso as amostras apresentassem dados normais (o que foi verificado pelos testes de kolmogorov-Smirnov) seria realizada comparação entre as médias utilizando teste T de Student para amostras não pareadas, ou teste T de Student para amostras pareadas quando pertinente. Nas amostras não normais foram realizados testes para amostras não paramétricas (testes de Wilcoxon e de Mann-Titney), conforme definido no texto.

A seguir estão enumeradas as etapas de análise realizadas; a numeração está pareada com os resultados, demonstrados a partir da página: 24

1 - Estudo do efeito da QRC nas notas de provas somativas teóricas e médias finais do aluno:

1.1 Estudando as **notas de provas somativas teóricas** de cada etapa do curso, comparado o segundo semestre de 2015 (sem QRC) e os dois semestres de 2016 (com QRC).

1.2 Estudando as **médias finais** de cada etapa do curso, comparado o segundo semestre de 2015 (sem QRC) e os dois semestres de 2016 (com QRC).

Para compreender como a média é formada ver a figura 2.

2 - Agrupamento dos resultados observados nas análises das etapas isoladas para cálculo do desfecho global, através do gráfico de Forest Plot

2.1- Agrupamento dos resultados obtidos em 1.1, relativo às mudanças observadas nas **notas de provas somativas teóricas** de cada etapa do curso, comparado o segundo semestre de 2015 (sem QRC) e os dois semestres de 2016 (com QRC).

2.2- Agrupamento dos resultados obtidos em 1.2, relativo às mudanças observadas nas **médias finais** de cada etapa do curso, comparado o segundo semestre de 2015 (sem QRC) e os dois semestres de 2016 (com QRC).

3 - Avaliação de dados pareados da evolução das notas de cada aluno

Estudo das notas dos alunos, de forma **pareada**, antes e após a implementação da prova QRC, sendo feita uma análise de como evoluiu a nota de cada aluno (agrupados conforme sua turma) ao longo dos 3 semestres analisados. (quando o aluno não cursou algum deles poderá ser incluído também, desde que tenha cursado ao menos dois semestre dentro do período da análise). Dois conjuntos de dados foram avaliados, sendo eles:

3.1 Evolução das notas de provas somativas teóricas após inclusão da QRC. Análise de amostras pareadas por alunos, agrupados por turmas.

3.2 Evolução das médias finais após inclusão da QRC. Análise de amostras pareadas por alunos, agrupados por turmas.

4 - Evolução das notas obtidas nas avaliações QRC, por etapas, ao longo do dois semestres de 2016

5 - Estudo do impacto da mudança da prova Salto Triplo para Duplo

Para avaliar se mesmo após a simplificação proposta na prova Salto, de Triplo para Duplo foi mantido seu poder de discriminação, foram estudadas as correlações entre notas de provas teóricas somativa e a nota da prova salto.

Considerando que a prova Salto tem resultado categórico dicotômico (aprovado ou reprovado) e as notas da prova teórica somativa são variáveis numéricas contínuas, análises estatísticas de correlação, como correlação de Pearson ou Spearman não puderam ser realizadas. A estratégia utilizada para avaliar o poder de discriminação foi separar as notas dos alunos que foram aprovados no Salto dos que foram reprovados, formando dois grupos. Foram aplicados testes para comparar médias (como o teste T para variáveis não pareadas) com intuito de descobrir se existe diferença significativa entre as notas destes dois grupos, e se o grupo com notas maiores se correlacionam com os aprovados na prova Salto.

Esta análise foi realizada para três semestres, sendo que no segundo semestre de 2015 os alunos responderam a prova Salto Triplo, e nos dois semestres de 2016 responderam a prova Salto Duplo.

O intuito deste trabalho, quanto à prova Salto, não foi comparar o efeito da mudança de Salto Triplo para Salto Duplo nas notas dos alunos, mas sim observar se ao longo destas três aplicações as provas mantiveram seu poder de discriminação.

ASPECTOS ÉTICOS

Projeto (CAAE: 65687417.4.0000.5415) aprovado pelo Comitê de Ética em pesquisa da FAMERP (Anexo - 1) (Parecer de nº 2.074.864, no dia 22 de maio de 2017).

Foi enviado Relatório Final (Anexo - 2) para Plataforma Brasil, datado no dia 21 de janeiro de 2018, Sendo avaliado pelo Comitê de Ética em pesquisa da FAMERP e aceito pelo parecer nº 2.495.367 no dia 15 de fevereiro de 2018. (Anexo - 3)

RESULTADOS

1. ESTUDO DO EFEITO DA QRC NAS NOTAS DE PROVAS SOMATIVAS TEÓRICAS E MÉDIAS FINAIS

1.1 IMPACTO DA QRC NAS NOTAS DE PROVAS SOMATIVAS TEÓRICAS POR ETAPAS

A- Comparando 2º semestre de 2015 (sem QRC) com 1º semestre de 2016 (com QRC)

Aumento das notas da prova somativa teórica em três etapas: 1, 4 e 7. Observamos redução destas notas na etapa 6 e as mudanças nas etapas 2, 3 e 5 não foram significativas. Conforme mostrado na *Tabela 1*.

TABELA 1:

Comparação entre notas de provas somativas teóricas entre 2º semestre de 2015 (sem QRC) com 1º semestre de 2016 (com QRC).

etapa	(a) notas	(b) notas	resultado	p	Z	mediana	mediana	N	N
	2ºsem2015	1ºsem2016				(a)	(b)	(a)	(b)
1	6,84	7,28	Aumento*	0,003	-2,921	7	7,5	166	187
2	7,63	7,83	Não sig *	0,052	-1,939	7,525	8	164	150
3	7,67	7,42	Não sig *	0,123	-1,543	7,85	7,5	91	169
4	7,98	7,6	Aumento*	0,001	-3,284	7,15	7,5	112	111
5	6,96	7,24	Não sig *	0,112	-1,588	7,15	7,5	109	113
6	7,49	6,83	Redução *	<0,001	-4,19	7,425	7	126	118
7	6,56	8,18	Aumento*	<0,001	-11,216	6,5	8,5	213	125

* Teste de Mann-Whitney U.

Nota:QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos.

B- Comparando 2º semestre de 2015 (sem QRC) com 2º semestre de 2016 (com QRC)

Aumento das notas da prova somativa teórica em três etapas: 4, 5 e 7. Observamos redução destas notas nas etapas 2 e 3 e as mudanças nas etapas 1 e 6 não foram significativas. Conforme mostrado na *Tabela 2*.

TABELA 2:

Comparação entre notas de provas somativas teóricas entre 2º semestre de 2015 (sem QRC) com 2º semestre de 2016 (com QRC).

etapa	(a) notas teóricas 2ºsem2015	(b) notas teóricas 2ºsem2016	resultado	p	Z	Mediana (a)	Mediana (b)	N (a)	N (b)
1	6,84	6,88	Não sig *	0,677	-0,417	7	7	166	192
2	7,63	7,12	Redução *	<0,001	-3,895	7,525	7	164	190
3	7,67	7,08	Redução *	<0,001	-3,733	7,85	7	91	156
4	7,98	7,56	Aumento*	<0,001	-4,036	7,5	7,66	112	170
5	6,96	7,78	Aumento*	<0,001	-4,774	7,5	8	109	114
6	7,49	7,41	Não sig *	0,602	-0,521	7	7,34	126	117
7	6,56	7,81	Aumento*	<0,001	-8,186	8,5	7,66	213	123

* Teste de Mann-Whitney U.

Nota:QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos.

C- Comparando 1º semestre de 2016 (com QRC) com 2º semestre de 2016 (com QRC)

Aumento das notas da prova somativa teórica em três etapas: 5, 6 e 8. Observamos redução destas notas nas etapas 1, 2, 3 e 7 e as mudanças na etapa 4 não foram significativas.

Conforme mostrado na *Tabela 3*.

TABELA 3:

Comparação entre notas de provas somativas teóricas entre 1º semestre de 2016 (com QRC) com 2º semestre de 2016 (com QRC).

etapa	(a) notas	(b) notas	resultado	p	Z	mediana		N (a)	N (b)
	1ºsem2016	2ºsem2016				(a)	(b)		
1	7,28	6,88	Redução*	<0,001	-3,902	7,5	7	187	192
2	7,83	7,12	Redução*	<0,001	-5,212	8	7	150	190
3	7,42	7,08	Redução*	0,009	-2,627	7,5	7	169	156
4	7,6	7,56	Não sig *	0,945	-0,069	7,5	7,66	111	170
5	7,24	7,78	Aumento*	0,001	-3,303	7,5	8	113	114
6	6,83	7,41	Aumento*	0,001	-3,423	7	7,34	118	117
7	8,18	7,81	Redução*	0,008	-2,635	8,5	7,66	125	123
8	5,96	6,88	Aumento*	<0,001	-5,514	6	7	214	123

* Teste de Mann-Whitney U.

Nota:QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos.

1.2 IMPACTO DA QRC NAS MÉDIAS FINAIS DOS MÓDULOS POR ETAPA

A- Comparando 2º semestre de 2015 (sem QRC) com 1º semestre de 2016 (com QRC)

Aumento das médias finais de quatro etapas: 1, 4, 5 e 7, com redução destas médias nas etapa 3 e 6 e sem alterações significativas na etapa 2. Conforme mostrado na *Tabela 4*.

TABELA 4:

Comparação entre médias finais do 2º semestre de 2015 (sem QRC) e do 1º semestre de 2016 (com QRC).

etapa	(a) médias 2ºsem2015	(b) médias 1ºsem2016	resultado	p	Z	mediana (a)	mediana (b)	N (a)	N (b)	DP (a)	DP (b)
1	7,21	7,53	Aumento*	0,026	-2,229	7,375	7,61	166	187	1,31	1,18
2	7,8	7,8	não sig**	0,976		7,8	7,95	160	150	1,02	1,17
3	7,8	7,32	Redução**	0,001		8	7,42	91	169	1,02	1,07
4	7,03	7,61	Aumento*	<0,001	-3,695	7,01	7,61	112	111	1,14	1,1
5	7,15	7,97	Aumento*	<0,001	-4,853	7,3	8,1	109	117	1,31	0,91
6	7,48	7,16	Redução**	0,019		7,5	7,19	126	118	1,17	0,93
7	6,85	8,43	Aumento*	<0,001	-11,8	7	8,61	213	125	1,14	0,68

* Teste de Mann-Whitney U. ** Teste t de student

Nota:QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos; DP=desvio padrão.

B- Comparando 2º semestre de 2015 (sem QRC) com 2º semestre de 2016 (com QRC)

Aumento das médias finais de 5 etapas: 1, 4, 5, 6 e 7, com redução destas medias nas etapas 2 e 3. Conforme mostrado na *Tabela 5*.

TABELA 5:

Comparação entre médias finais do 2º semestre de 2015 (sem QRC) e do 2º semestre de 2016 (com QRC).

etapa	(a) médias	(b) médias	Resultado	p	Z	mediana	mediana	N	N	DP	DP
	2ºsem2015	2ºsem2016				(a)	(b)	(a)	(b)	(a)	(b)
1	7,21	7,78	Aumento*	<0,001	-4,053	7,375	7,87	166	192	1,31	0,86
2	7,8	7,56	Redução**	0,021		7,8	7,56	160	190	1,02	0,90
3	7,8	7,19	Redução**	<0,001		8	7,23	91	156	1,02	0,95
4	7,03	7,51	Aumento*	<0,001	-3,568	7,01	7,53	112	170	1,14	0,93
5	7,15	8,11	Aumento*	<0,001	-5,639	7,3	8,27	109	114	1,31	0,94
6	7,48	7,57	Aumento**	0,501		7,5	7,67	126	117	1,17	1,04
7	6,85	8,12	Aumento*	<0,001	-9,834	7	8,18	213	123	1,14	0,91

* Teste de Mann-Whitney U. ** Teste t de student

Nota:QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos; DP=desvio padrão.

C- Comparando 1º semestre de 2016 (com QRC) com 2º semestre de 2016 (com QRC)

Aumento das médias finais de 3 etapas: 1, 6 e 8, com redução destas médias nas etapas 2 e 7. As variações nas médias das etapas 3, 4 e 5 não foram significativas. Conforme mostrado na *Tabela 6*.

TABELA 6:

Comparação entre médias finais do 1º semestre de 2016 (com QRC) e do 2º semestre de 2016 (com QRC).

etapa	(a) médias (b) médias		resultado	p	Z	mediana	mediana	N	N	DP	DP
	1ºsem2016	2ºsem2016				(a)	(b)	(a)	(b)	(a)	(b)
1	7,53	7,78	Aumento*	0,054	-1,924	7,61	7,87	187	192	1,18	0,86
2	7,8	7,56	Redução**	0,033		7,95	7,56	150	190	1,17	0,90
3	7,32	7,19	Não sig*	0,168	-1,377	7,42	7,23	169	156	1,07	0,95
4	7,61	7,51	Não sig*	0,312	-1,011	7,61	7,53	111	170	1,10	0,93
5	7,97	8,11	Não sig*	0,217	-1,235	8,1	8,27	117	114	0,91	0,94
6	7,16	7,57	Aumento**	0,001		7,19	7,67	118	117	0,93	1,04
7	8,43	8,12	Redução*	0,001	-3,427	8,61	8,18	125	123	0,68	0,81
8	6,77	7,82	Aumento*	<0,001	-8,503	6,75	7,87	214	123	1,03	1,00

* Teste de Mann-Whitney U. ** Teste t de student

Nota: QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos; DP=desvio padrão.

2- AGRUPAMENTO DE RESULTADOS OBTIDOS EM 1.1 E 1.2 / GRÁFICOS DE FOREST PLOT E CÁLCULO ESTATÍSTICO POR VARIANCIA INVERTIDA

O objetivo de entender o impacto da inclusão da QRC no componente curricular de tutoria não pode ser claramente elucidado com as análises prévias, que mostraram o efeito separado por etapas, traduzindo, de forma truncada, o comportamento que as notas obtiveram. Como a observação foi heterogênea, com algumas etapas sendo beneficiadas e

outras sendo prejudicadas, foi necessário agrupar o resultado para permitir observar o desfecho final da intervenção. Para isto foi proposta uma estratégia de análise inovadora em estudos educacionais. (Almeida et al., 2017)

O gráfico de Forest Plot, também chamado de Blobograma, é uma forma de exibição visual dos resultados de diferentes estudos, expondo uma estimativa do efeito global das intervenções. Seu uso é comum em meta-análises, onde diferentes trabalhos são agrupados com a finalidade de determinar a consequência de alguma intervenção.

A seguinte análise foi realizada utilizando o programa Review Manager (RevMan) 5.3, um software gratuito distribuído pela “The Nordic Cochrane Centre”. Todos os dados foram estudados considerando um intervalo de confiança de 95%, erro alfa de 5% e foram considerados significativos quando $p < 0,05$. Todos os cálculos para desenho do gráfico Forest Plot utilizaram o método estatístico “Inverted Variance”.

2.1 – AGRUPAMENTO DE RESULTADOS REFERENTES A VARIAÇÃO DA NOTA DA PROVA SOMATIVA TEÓRICA

O gráfico de forest plot abaixo (*figura 5*) mostra, no primeiro conjunto de dados, o comportamento das notas de provas somativas teóricas durante a primeira inclusão da QRC, portanto mostra as variações que cada etapa obteve entre o segundo semestre de 2015 e o primeiro semestre de 2016. Cada etapa é apresentada no gráfico como um ponto, que representa a variação da média das notas das provas somativas teóricas entre os períodos analisados. A linha horizontal que acompanha o ponto representa o intervalo de confiança desta variação média, sendo proporcional à heterogenicidade dos dados, do desvio padrão e do número de amostras dentro de cada análise. Se a barra horizontal cruzar a linha média

vertical significa que as mudanças não são estatisticamente significativas. Abaixo do conjunto de dados está um ponto em formato de diamante, este ponto representa o desfecho global deste conjunto de dados, se estiver para direita da linha vertical, sem tocá-la, conclui-se que as notas, após inclusão da QRC, foram maiores que as notas anteriores.

O desfecho global da primeira inclusão da prova QRC, portanto, comparando a nota da prova somativa teórica do segundo semestre de 2015 com o primeiro semestre de 2016 foi favorável ao uso da QRC, com aumento de 0,45 pontos na nota da prova somativa teórica (numa escala de 0-10) significativo com $p < 0.00001$. Também conclui que os dados são muito heterogêneos, com $I^2 = 97\%$.

O segundo conjunto de dados respeita as mesmas regras explicadas acima, e representa a variação da média das notas teóricas entre o segundo semestre de 2015 e o segundo semestre de 2016. O defecho global calculado foi favoravel ao uso da QRC, com aumento de 0,21 pontos na nota teórica (numa escala de 0-10) significativo com $p < 0.00001$. Também conclui que os dados são muito heterogêneos, com $I^2 = 97\%$.

O último ponto em formato de diamante do gráfico mostra o desfecho global dos dois conjuntos de dados, portanto, agrupa 14 grupos de intervenções independentes. O ponto está a direita da linha vertical, e isso representa um aumento das notas das provas teóricas após a introdução da QRC. A diferença média observada foi de 0,33 pontos (numa escala de 0-10) significativo com $p < 0.00001$. Também conclui que os dados são muito heterogêneos, com $I^2 = 96\%$.

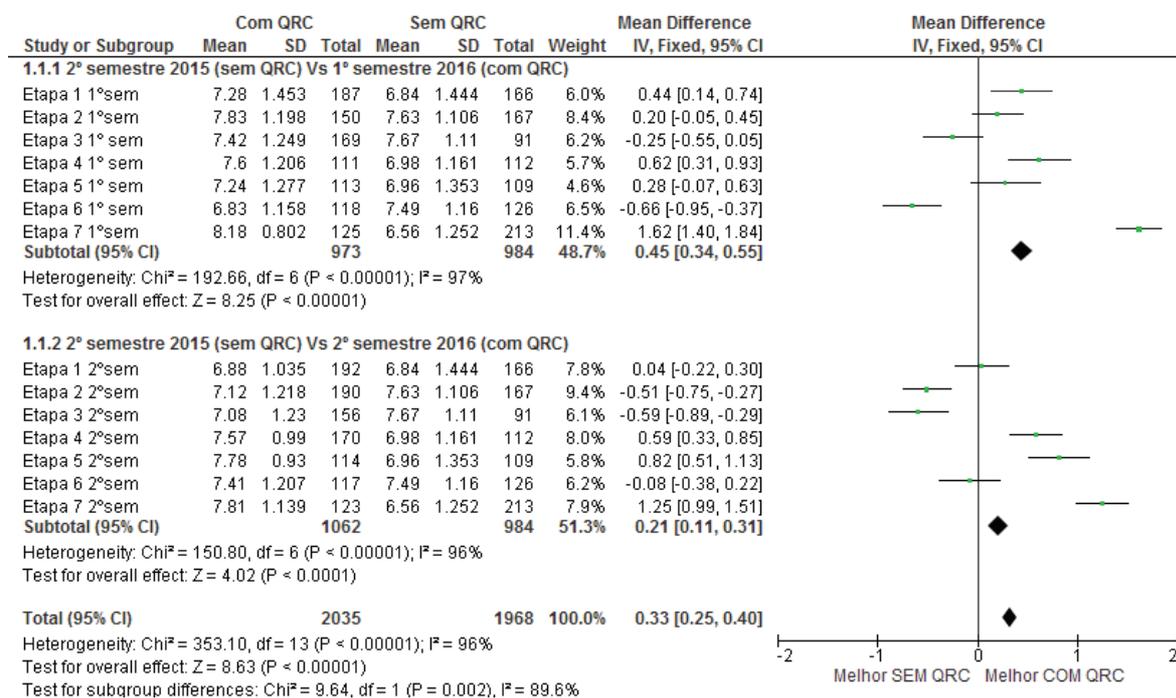


FIGURA 5:

Grafico de forest plot demonstrando efeitos observados nas notas de provas somativas teóricas no primeiro e segundo semestre após introdução da QRC.

2.2 AGRUPAMENTO DE RESULTADOS REFERENTES A VARIAÇÃO DA MÉDIA FINAL

O mesmo método explicado anteriormente foi utilizado para explorar o comportamento das médias finais dos alunos.

O gráfico de Forest Plot está representando na [figura 6](#). O primeiro conjunto de dados representa a variação das médias finais e o desfecho global da primeira inclusão da prova QRC, portanto, comparando o segundo semestre de 2015 e o primeiro semestre de 2016 foi favorável ao uso da QRC, com aumento de 0,23 pontos na média final (numa escala de 0-10)

significativo com $p < 0.00001$. Também conclui que os dados são muito heterogêneos, com $I^2 = 93\%$.

O segundo conjunto de dados representa a variação das médias finais entre o segundo semestre de 2015 e o segundo semestre de 2016. O defecho global calculado foi favorável ao uso da QRC, com aumento de 0,36 pontos na nota teórica (numa escala de 0-10) significativo com $p < 0.00001$. Também conclui que os dados são muito heterogêneos, com $I^2 = 97\%$.

O último ponto em formato de diamante do gráfico mostra o desfecho global dos dois conjuntos de dados, portanto, agrupa os 14 grupos de intervenções independentes. O ponto está a direita da linha vertical, e isso representa um aumento das médias finais após a introdução da QRC. A diferença média observada foi de 0,30 pontos (numa escala de 0-10) significativo com $p < 0.00001$. Também conclui que os dados são muito heterogêneos, com $I^2 = 95\%$.

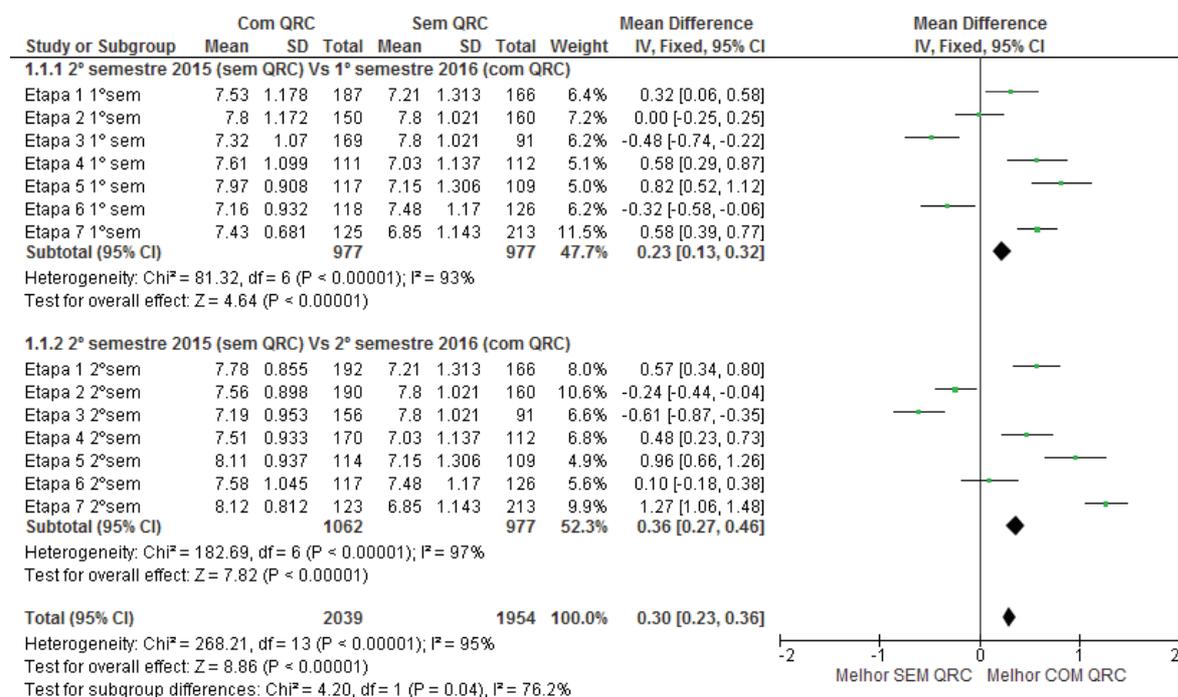


FIGURA 6:

Grafico de forest plot demonstrando efeitos observados nas medias finais no primeiro e segundo semestre após introdução da QRC.

3.1- IMPACTO DA QRC NAS NOTAS DE PROVAS SOMATIVAS TEÓRICAS, ANÁLISE PAREADA POR ALUNOS

A- Comparando 2º semestre de 2015 (sem QRC) com 1º semestre de 2016 (com QRC)

Alunos das turmas 2, 4 e 7 apresentaram aumento significativo de suas notas de provas somativas teóricas, alunos da turma 1 sofreram redução de suas notas e as demais turmas não apresentaram mudanças significativas. Conforme mostrado na *Tabela 7*.

TABELA 7:

Evolução das notas de provas somativas teóricas entre o 2º semestre de 2015 (sem QRC) e o 1º semestre de 2016 (com QRC).

turma	etapas	(a) Notas	(b) Notas	resultado	p	DP (a)	DP (b)	N	Mediana (a)/(b)
		teóricas 2ºsem2015	teóricas 1ºsem2016						
7	1 2	7,12	7,99	Aumento*	<0,001	0,66954	0,74514	35	
6	2 3	7,73	7,6	Não sig*	0,2	0,77665	0,69675	45	
5	3 4	7,64	6,69	Não sig**	0,767	0,75682	0,5945	25	7,65/7,67
4	4 5	7,01	7,4	Aumento*	0,01	0,77154	0,50421	32	
3	5 6	7,07	6,89	Não sig*	0,185	0,57521	0,87027	32	
2	6 7	7,55	8,19	Aumento*	<0,001	0,85966	0,52637	38	
1	7 8	6,58	6,02	Redução*	<0,001	0,87949	1,00341	66	

Nota: Dados pareados por alunos, separados por turmas.

* Teste t de student para amostras pareadas; ** Teste de Wilcoxon para amostras pareadas

QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos; DP=desvio padrão.

Comparando 2º semestre de 2015 (sem QRC) com 2º semestre de 2016 (com QRC)

Alunos das turmas 3, 4, 5 e 7 apresentaram aumento significativo de suas notas de provas somativas teóricas, a turma 2 sofreu redução destas notas e a turma 6 não apresentou mudanças significativas, conforme mostrado na *Tabela 8*.

Nesta análise a turma 1 não foi incluída por não estar mais cursando a componente curricular de tutoria no segundo semestre de 2016.

TABELA 8:

Evolução das notas de provas somativas teóricas entre o 2º semestre de 2015 (sem QRC) e o 2º semestre de 2016 (com QRC).

turma	etapas	(a) notas	(b) notas	resultado	p	DP (a)	DP (b)	N	Mediana (a)/(b)
		teóricas 2ºsem2015	teóricas 2ºsem2016						
7	13	6,93	7,23	Aumento*	0,011	0,80401	0,73179	35	
6	24	7,64	7,71	Não sig*	0,483	0,72705	0,69002	43	
5	35	7,62	7,94	Aumento**	0,046	0,74998	0,56339	24	7,65/8,92
4	46	6,97	7,53	Aumento*	<0,001	0,79366	0,6318	33	
3	57	7,07	7,91	Aumento*	<0,001	0,58451	0,71016	31	
2	68	7,54	6,9	Redução**	<0,001	0,88998	0,80102	38	

Nota: Dados pareados por alunos, separados por turmas.

* Teste t de student para amostras pareadas; ** Teste de Wilcoxon para amostras pareadas
QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos; DP=desvio padrão.

B- Comparando 1º semestre de 2016 (com QRC) com 2º semestre de 2016 (com QRC)

Alunos das turmas 3 e 5 apresentaram aumento significativo de suas notas de provas somativas teóricas, a turma 2 e 7 sofreram redução destas notas e as turmas 4 e 6 não apresentaram mudanças significativas, conforme mostrado na *Tabela 9*..

Nesta análise a turma 1 não foi incluída por não estar mais cursando a componente curricular de tutoria no segundo semestre de 2016.

TABELA 9:

Evolução das notas de provas somativas teóricas entre o 1º semestre de 2016 (com QRC) e o 2º semestre de 2016 (com QRC).

turma	etapas	(a) notas	(b) notas	resultado	p	DP (a)	DP (b)	N	Mediana (a)/(b)
		1ºsem2016	2ºsem2016						
7	2 3	8,01	7,31	Redução*	<0,001	0,73612	0,73179	30	
6	3 4	7,57	7,71	Não sig*	0,101	0,69857	0,69711	42	
5	4 5	7,72	7,95	Aumento**	0,033	0,54918	0,56339	23	7,67/8,11
4	5 6	7,4	7,54	Não sig*	0,242	0,50421	0,6384	32	
3	6 7	6,89	7,91	Aumento*	<0,001	0,88459	0,71016	31	
2	7 8	8,2	6,87	Redução*	<0,001	0,53264	0,79047	37	

Nota: Dados pareados por alunos, separados por turmas.

* Teste t de student para amostras pareadas; ** Teste de Wilcoxon para amostras pareadas
QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos; DP=desvio padrão.

3.2 IMPACTO DA QRC NAS MÉDIAS FINAIS, ANÁLISE PAREADA POR ALUNOS

A- Comparando 2º semestre de 2015 (sem QRC) com 1º semestre de 2016 (com QRC)

Alunos das turmas 2, 4 e 7 apresentaram aumento significativo de suas médias finais, sendo que as médias da turma 6 sofreram redução e os demais dados não revelaram mudanças significativas. Conforme mostrado na *Tabela 10*.

TABELA 10:

Evolução das médias finais entre o 2º semestre de 2015 (sem QRC) e o 1º semestre de 2016 (com QRC).

turma	etapas	(a) médias	(b) médias	resultado	p	DP (a)	DP (b)	N
		2ºsem2015	1ºsem2016					
7	1 2	7,44	8,1	Aumento*	<0,001	0,60195	0,68574	35
6	2 3	7,91	7,54	Redução*	<0,001	0,70503	0,70081	45
5	3 4	7,81	7,76	Não sig*	0,694	0,72843	0,68665	25
4	4 5	7,06	8,16	Aumento*	<0,001	0,76986	0,43527	32
3	5 6	7,24	7,21	Não sig*	0,793	0,57892	0,71422	32
2	6 7	7,52	8,47	Aumento*	<0,001	0,79562	0,4745	38
1	7 8	6,87	6,8	Não sig*	0,416	0,81314	0,75998	66

Nota: Dados pareados por alunos, separados por turmas.

* Teste t de student para amostras pareadas

QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos; DP=desvio padrão.

B- Comparando 2º semestre de 2015 (sem QRC) com 2º semestre de 2016 (com QRC)

Alunos das turmas 2, 3, 4 e 5 apresentaram aumento significativo de suas médias finais, e as turmas 6 e 7 não apresentaram mudanças significativas, conforme mostrado na *Tabela 11*.

Nesta análise a turma 1 não foi incluída por não estar mais cursando a componente curricular de tutoria no segundo semestre de 2016.

TABELA 11:

Evolução das médias finais entre o 2º semestre de 2015 (sem QRC) e o 2º semestre de 2016 (com QRC).

turma	etapas	(a) médias	(a) médias	resultado	p	DP (a)	DP (b)	N
		2ºsem2015	2ºsem2016					
7	13	7,28	7,43	Não sig*	0,146	0,71118	0,67945	35
6	24	7,83	7,67	Não sig*	0,067	0,66623	0,72903	43
5	35	7,78	8,28	Aumento*	0,001	0,71026	0,62348	24
4	46	7,01	7,73	Aumento*	<0,001	0,80575	0,65273	33
3	57	7,24	8,25	Aumento*	<0,001	0,58744	0,51243	31
2	68	7,51	7,85	Aumento*	<0,001	0,81886	0,57256	38

Nota: Dados pareados por alunos, separados por turmas.

* Teste t de student para amostras pareadas

QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos; DP=desvio padrão.

C- Comparando 1º semestre de 2016 (com QRC) com 2º semestre de 2016 (com QRC)

Alunos das turmas 3 e 5 apresentaram aumento significativo de suas médias finais; as turmas 2, 4, 6 e 7 sofreram redução destas médias, conforme mostrado na *Tabela 12*.

Nesta análise a turma 1 não foi incluída por não estar mais cursando a componente curricular de tutoria no segundo semestre de 2016.

TABELA 12:

Evolução das médias finais entre o 1º semestre de 2016 (com QRC) e o 2º semestre de 2016 (com QRC).

turma	etapas	(a) médias	(a) médias	resultado	p	DP (a)	DP (b)	N
		1ºsem2016	2ºsem2016					
7	2 3	8,1	7,55	Redução*	<0,001	0,68323	0,57703	30
6	3 4	7,49	7,69	Redução*	0,002	0,69604	0,72757	42
5	4 5	7,77	8,28	Aumento*	<0,001	0,57352	0,63623	23
4	5 6	8,16	7,75	Redução*	<0,001	0,43527	0,65115	32
3	6 7	7,21	8,25	Aumento*	<0,001	0,72589	0,51243	31
2	7 8	8,48	7,85	redução	<0,001	0,47208	0,57946	37

Nota: Dados pareados por alunos, separados por turmas.

* Teste t de student para amostras pareadas

Nota: QRC=Questão de Resposta Curta; sem=semestre; sig=significativo; N=número de alunos; DP=desvio padrão.

4 – VARIAÇÃO DA NOTA DA QRC AO LONGO DE DOIS SEMESTRES SEGUIDOS DE APLICAÇÃO, POR ETAPAS

A comparação das notas obtidas nas provas QRC entre o 1º semestre e o 2º semestre de 2016 mostrou que as notas obtidas nas etapas 1, 2 e 8 apresentaram significativa melhora; já as notas obtidas na etapa 4 sofreram redução. Nas etapas 3, 5, 6 e 7 não apresentaram variações significativas. Estes resultados estão expressos na *Tabela 13*.

TABELA 13:

Variação da Nota da QRC ao longo de dois semestres seguidos de aplicação.

etapa	(a)	(b)	p	Desv. Padrão	Desv. Padrão	N	N
	1ºsem2016	2ºsem2016		(a)	(b)	(a)	(b)
1	7,779	8,58	<0,001*	1,737	1,264	187	192
2	7,91	8,55	0,001*	1,825	1,399	150	190
3	7,48	7,56	0,611*	1,636	1,533	169	156
4	7,79	7,38	0,032*	1,645	1,483	111	170
5	8,46	8,58	0,494*	1,295	1,334	113	114
6	7,601	8,044	0,22*	1,483	1,46	118	117
7	8,82	8,52	0,057*	1,171	1,294	125	123
8	8,137	8,87	<0,001*	1,618	1,274	214	123

* Teste t de student

Nota: QRC=Questão de Resposta Curta; sem=semestre; N=número de alunos; Desv.Padrão=desvio padrão.

5- AVALIAÇÃO QUANTITATIVA DOS EFEITOS DA MUDANÇA DE SALTO TRIPLO PARA DUPLO

O presente trabalho visava constatar se a mudança de Salto Triplo para Duplo interferiu na validade deste instrumento de avaliação. Estudou-se a capacidade das provas Salto Triplo e Duplo discriminar os alunos aprovados dos reprovados, utilizando notas das avaliações teóricas e médias finais como “padrão ouro” para verificar se aqueles alunos que

foram aprovados na prova salto coincidiam com os alunos que obtiveram melhores notas, e se esta relação se manteve após a mudança da prova Salto Triplo para Duplo.

5.1 - AVALIAÇÃO DA CAPACIDADE DE DISCRIMINAÇÃO DO SALTO TRIPLO

A primeira análise considerou as notas dos alunos no segundo semestre de 2015, último semestre a utilizar a avaliação Salto Triplo no componente curricular de tutoria do curso de medicina da Faceres.

Teste T mostrou que as notas de provas teóricas dos alunos que foram aprovados (nota numérica igual a 1) na prova Salto Triplo diferiram significativamente ($p < 0,001$) daqueles que foram reprovados (nota igual a 0) na prova Salto Triplo, sendo que os aprovados obtiveram notas maiores ($7,23 \pm 1,27$ $n=639$) que os reprovados ($6,86 \pm 1,35$ $n=342$) concluindo que neste semestre, a prova Salto Triplo foi eficiente em discriminar o rendimento dos alunos, tendo por base a nota da prova Teórica.

Estes resultados foram expressos graficamente na figura 7.

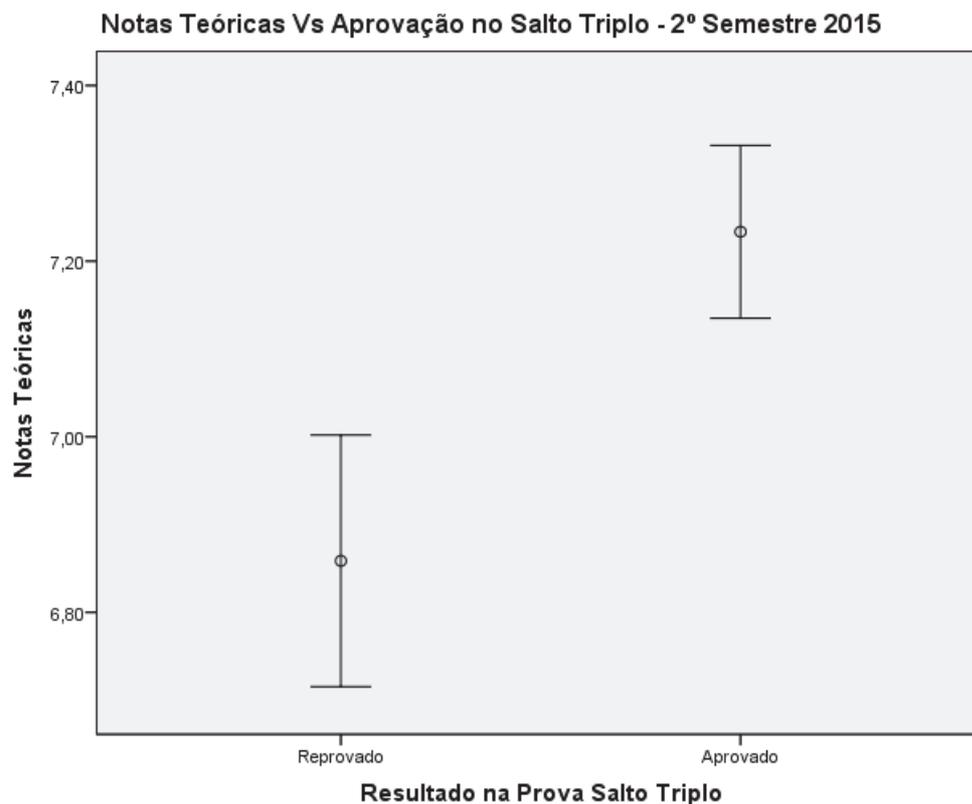


FIGURA 7:

Distribuição das notas em provas somativas teóricas no segundo semestre de 2015, estratificando pelo resultado obtido na prova Salto Triplo. Diferença significativa ($p < 0,001$) entre a média de notas teóricas obtidas por alunos com conceito “aprovado” ($7,23 \pm 1,27$ $n=639$) e aqueles com conceito “reprovado” ($6,86 \pm 1,35$ $n=342$).

5.2 - AVALIAÇÃO DA CAPACIDADE DE DISCRIMINAÇÃO DO SALTO DUPLO NO PRIMEIRO SEMESTRE DE APLICAÇÃO

A segunda análise considerou as notas dos alunos no primeiro semestre de 2016, primeiro semestre a utilizar a avaliação Salto Duplo no componente curricular de tutoria do curso de medicina da FACERES.

Teste T mostrou que as notas de provas somativas teóricas dos alunos que obtiveram conceito aprovado na prova Salto Duplo diferiram significativamente ($p < 0,001$) daqueles considerados reprovados na prova Salto Duplo, sendo que os aprovados tem notas maiores ($7,59 \pm 1,35$ $n=491$) que os reprovados ($6,94 \pm 1,46$ $n=696$) concluindo que neste semestre, a prova Salto Duplo foi eficiente em discriminar o rendimento dos alunos, tendo por base a nota da prova Teórica.

Estes resultados estão expressos graficamente na *figura 8*.

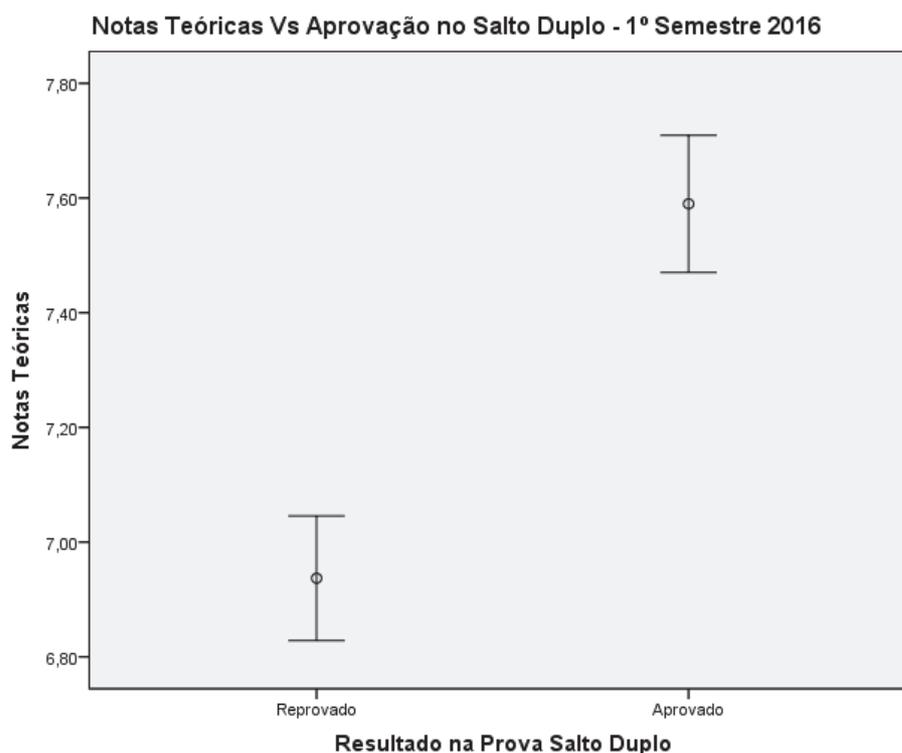


FIGURA 8:

Distribuição das notas em provas somativas teóricas no primeiro semestre de 2016, estratificando pelo resultado obtido na prova Salto Duplo. Diferença significativa ($p < 0,001$) entre a média de notas teóricas obtidas por alunos com conceito “aprovado” ($7,59 \pm 1,35$ $n=491$) e aqueles com conceito “reprovado” ($6,94 \pm 1,46$ $n=696$).

5.3 - AVALIAÇÃO DA CAPACIDADE DE DISCRIMINAÇÃO DO SALTO DUPLO NO SEGUNDO SEMESTRE DE APLICAÇÃO

A terceira análise considerou as notas dos alunos no segundo semestre de 2016, segundo semestre a utilizar a avaliação Salto Duplo no componente curricular de tutoria do curso de medicina da FACERES.

Teste T mostrou que as notas de provas somativas teóricas dos alunos que foram aprovados na prova Salto Duplo foram significativamente ($p=0,048$) diferentes daqueles que foram reprovados na prova Salto Duplo, sendo que os aprovados obtiveram notas maiores ($7,36 \pm 1,17$ $n=498$) que os reprovados ($7,22 \pm 1,20$ $n=687$) concluindo que neste semestre, a prova Salto Duplo, mais uma vez, se mostrou eficiente em discriminar o rendimento dos alunos, tendo por base a nota da prova Teórica.

Estes resultados estão expressos graficamente na *figura 9*.

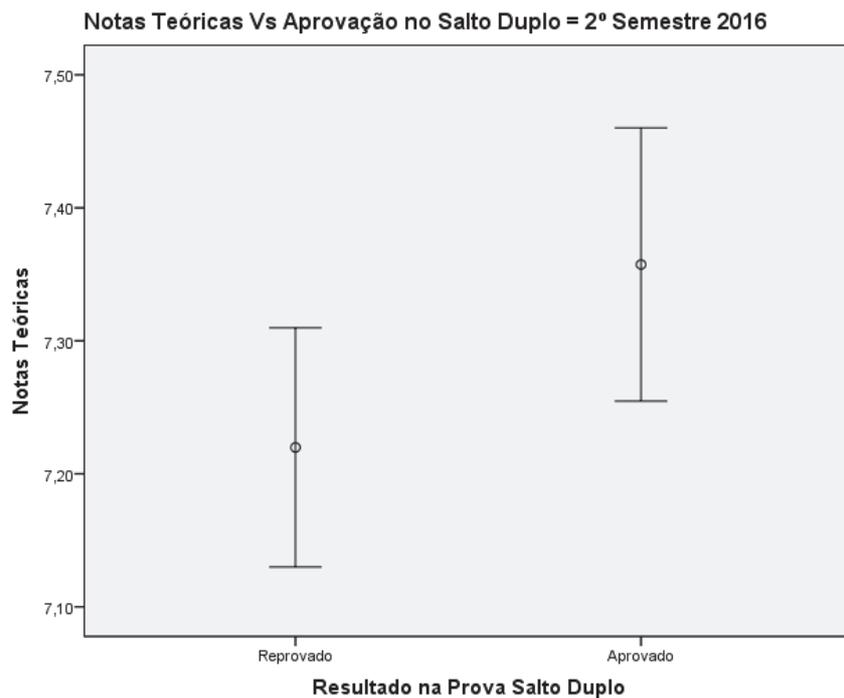


FIGURA 9:

Distribuição das notas em provas somativas teóricas no segundo semestre de 2016, estratificando pelo resultado obtido na prova Salto Duplo. Diferença significativa ($p=0,048$) entre a média de notas teóricas obtidas por alunos com conceito “aprovado” ($7,36 \pm 1,17$ $n=498$) e aqueles com conceito “reprovado” ($7,22 \pm 1,20$ $n=687$).

5.4 CONSIDERAÇÕES SOBRE A VARIAÇÃO DOS ÍNDICES DE APROVAÇÃO ENTRE PROVAS SALTO DUPLO E SALTO TRIPLO

Foi evidente que o número de alunos reprovados na prova Salto aumentou após a mudança de Salto Triplo para Salto Duplo (342 vs 696), fato que se manteve também no segundo semestre de aplicação (687 reprovações). (*figura 10*)

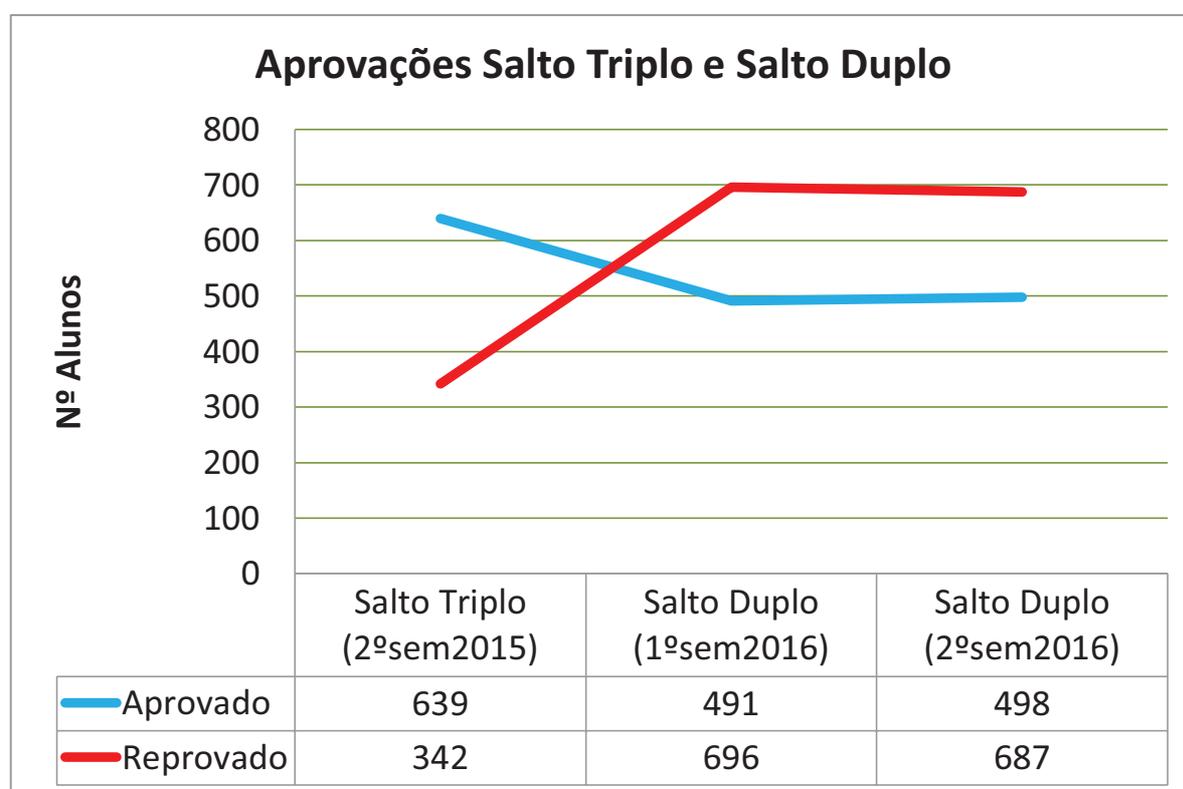


FIGURA 10:

Gráfico de aprovações e reprovações na prova Salto Triplo e provas Salto Duplo ao longo dos três semestres dest4 estudo.

DISCUSSÃO

Considerações sobre o impacto da QRC nas notas de provas somativas teóricas

Com intuito de avaliar se a introdução da QRC foi capaz de melhorar o rendimento dos alunos, o presente estudo avaliou quantitativamente as variações que as notas de provas somativas teóricas e médias finais sofreram após o uso de QRC. Inicialmente foram comparados os resultados de provas teóricas somativas finais, pensando nesta avaliação como um “padrão ouro” para o rendimento dos alunos.

Nesta análise, tanto no primeiro como no segundo semestre onde foram utilizadas as QRC (1º semestre de 2016 e 2º semestre de 2016) ficou constatado que ocorreu aumento, estatisticamente significativo das notas. No caso de 1º semestre de 2016 uma etapa sofreu redução de suas notas (*tabela 1*) e no 2º semestre de 2016 duas etapas sofreram reduções de suas notas (*tabela 2*). Esta primeira análise já colabora com a hipótese de melhora do rendimento dos alunos nas provas somativas teóricas após introdução do uso de QRC.

Para confirmar se realmente houve aumento nas notas, de forma global, foi utilizado uma ferramenta de meta-análise, chamada de forest plot com estatística por variância invertida. Esta estratégia mostrou melhora das notas no primeiro semestre de inclusão (*figura 5*), com aumento de 0,45 pontos (0,34-0,55 $p < 0,00001$), e também confirmou aumento das notas no segundo semestre de inclusão, com aumento de 0,21 pontos (0,11-0,31 $p < 0,00001$). Esta análise confirma a hipótese de efeito positivo do uso da QRC nas notas de provas teóricas somativas, contudo, também é claro que, apesar de significativo, este aumento é pequeno.

Também com objetivo de verificar se houve melhora do rendimento dos alunos nas provas teóricas foram feitos teste pareados, onde a nota do mesmo aluno é acompanhada ao longo dos três semestres desta análise (um antes da QRC e dois depois). Neste ensaio estatístico os alunos foram agrupados por turmas e então foi verificada como evoluíram suas notas. A *tabela 7* descreve como variaram as notas quando se compara o semestre sem QRC (2º semestre 2015) com o que o procedeu imediatamente (1º semestre 2016). Neste grupo de dados ficou evidente que alunos de três turmas melhoraram seu rendimento (2, 4 e 7) e a apenas uma turma reduziu seu rendimento (turma 1), colaborando para conclusão geral de melhora nas notas. Comparando o semestre sem QRC (2º semestre 2015) com o segundo semestre de avaliações com QRC (2º semestre 2016, exibido na *tabela 8*) ficou mais clara a melhora das notas, sendo que quatro turmas tiveram melhora nas suas notas (3, 4, 5 e 7) e apenas uma piorou (turma 2).

Este estudo contraria um pensamento inicial sobre a QRC. Na sua idealização, foi inferido que esta avaliação tentaria ajudar na motivação dos alunos e, portanto, como os alunos mais “velhos”, de turmas mais avançadas, já estariam intrinsecamente motivados, a QRC teria menos influência sobre eles. Nesta análise, não ficou claro um padrão de melhora no rendimento que predominasse nas etapas mais avançadas em comparação às mais jovens. Talvez, o tópico da questão elaborada, explorando mais fisiopatologia nas etapas iniciais, e tratamento nas etapas finais, poderia explicar esta discordância de rendimento, mas esta análise (qualitativa sobre a questão) não foi o enfoque deste estudo.

Considerações sobre o impacto da QRC nas médias finais

O padrão de alterações relatado nas notas teóricas também se repete nas médias finais. Analisando o mesmo período descrito anteriormente, é possível verificar que a maioria das

etapas obteve médias melhores após o QRC do que tinha antes (*tabelas 4 e 5*). No entanto, também é possível constatar grande heterogeneidade destas variações, colaborando para a hipótese de falta de padronização da dificuldade dos instrumentos aplicados nas diferentes etapas.

Outro fator de confusão ao analisar mudanças que ocorreram nas médias dos alunos é que, após a inclusão da QRC, o método de calcular a média final foi modificado, incluindo mudanças nos pesos que são atribuídos a cada avaliação. Estas modificações interferem de forma substancial no uso da média como parâmetro para definir se houve melhora no rendimento do aluno ou se traduz apenas a mudança que ocorreu numericamente na nota.

Na análise da variação de médias finais, por estudo de dados pareados, semelhante ao que foi observado no estudo das notas de provas teóricas somativas, não houve diferenças nítidas entre as variações das médias das turmas mais novas comparando com a mais velhas. Ficando evidente a melhora das notas, na primeira inclusão da QRC, nas turmas 2, 4 e 7, com redução da nota da turma 6. Na segunda inclusão da QRC observamos melhora nas turmas 2, 3, 4 e 5, sem redução da nota de nenhuma turma. A avaliação de variação de médias finais, mais uma vez, repete o viés de que esta média sofreu mudança na sua forma de cálculo, com pesos diferentes após a inclusão da QRC, e a interferência da nota deste próprio instrumento (QRC) na conformação da média.

Considerações sobre o impacto da QRC no rendimento dos alunos, de forma geral

Pode ser argumentado que existem interferências nas análises realizadas, principalmente ao se considerar que não houve uma rígida padronização tanto da dificuldade das provas somativas teóricas aplicadas como também do método de correção das QRC.

Cada etapa pode ter utilizado variações qualitativas nos instrumentos de avaliação, podendo, a diferença constatada nas notas teóricas ser apenas reflexo de mudanças na dificuldade ou validade da prova. Contudo, o grande número de indivíduos e diferentes testes realizados neste estudo visam contornar este tipo de viés.

A grande heterogeneidade de variações das notas entre as diferentes etapas colabora com a hipótese de que exista outro fator de confusão: a falta de padronização no desenvolvimento e correção da QRC. A dificuldade em padronizar a redação, a profundidade e a fidedignidade da QRC pode acarretar efeitos diferentes dos desejados e não colaborar para a melhora da qualidade do estudo do aluno.

Estes problemas foram identificados nestes primeiros semestres e após isto a faculdade Faceres passou a normatizar as características da QRC, instituindo uma comissão destinada a pré-avaliar as questões de todas as etapas e equilibrar as questões, tanto na dificuldade como na pertinência e relevância do conteúdo questionado. O corpo docente foi instruído a elaborar dois itens em cada avaliação, sendo que estas questões devem ser claras e objetivas, focadas centralmente nos desempenhos de compreensão do problema estudado. O padrão de resposta esperado é discutido entre os tutores com intuito de padronizar a correção e esta última fase não deve ser feita individualmente, por um único docente, e sim numa banca de tutores do módulo.

Variações nas notas dentro de dois semestres consecutivos de aplicação de QRC

Parte deste trabalho avaliou como as notas se comportaram dentro dos dois semestres que utilizaram QRC e Salto Duplo, portanto, avalia grupos que não sofreram nenhuma mudança curricular entre si. Não houve mudança de peso de avaliações ou mudança no tipo e

número de avaliações que realizaram. Entre estes grupos, que cursaram o 1º e o 2º semestres de 2016, somente se pode esperar que tivessem ocorrido mudanças na qualidade dos instrumentos aplicadas, assim como na forma de corrigir o mesmo. Estes conjunto de estudos, dentro dos dois semestres que utilizaram QRC, visam observar se ocorreram mudanças estatísticas nas notas, que somente poderão representar mudanças intrínsecas nos instrumento, incluindo melhora da habilidade do professor e do aluno no uso dos novos instrumento de avaliação (QRC e Salto Duplo).

O estudo das variações das notas de provas somativas teóricas e de médias finais dentro dos dois semestres que utilizaram o QRC foi feito utilizando duas metodologias distintas. A primeira avaliou a variação das notas dentro de cada etapa; nesta metodologia não foi visto variações com padrão constante para permitir uma conclusão clara. As notas de provas teóricas somativas sofreram redução em quatro etapas e aumento em três (*tabela 3*). Já as médias finais sofreram aumento em três etapas e redução em duas (*tabela 6*).

A outra metodologia utilizada foi avaliar dados pareados, estudando o comportamento de cada aluno, agrupado em sua turma. Esta análise também se mostra inconclusiva. Nas notas de provas somativas teóricas foi constatado aumento da nota em duas turmas e redução também em duas (*tabela 9*). O mesmo tipo de análise nas médias finais dos alunos mostrou tendência para piora, com redução das médias de quatro turmas, frente o aumento na média de duas turmas (*tabela 12*).

Ainda estudando o comportamento de notas dentro de dois semestres que envolveram o uso de QRC, foi comparado a variação que a nota do próprio QRC sofreu (*tabela 13*), constatando melhora das notas da QRC em três etapas frente à piora em apenas uma. Estes dados colaboram para hipótese de que a elaboração da pergunta pode ter se aprimorado ao longo dos semestres, na medida em que os tutores se tornaram mais hábeis na elaboração da

QRC. O mesmo poderá ser concluído sobre o aluno, que teve melhores resultados nas QRC devido à familiarização com o método de avaliação. É possível considerar que após um semestre praticando o hábito de ser avaliado por QRC, o aluno tenha também criado uma disciplina de estudo mais regrada, permitindo melhora no despenho, o que é compatível também com a melhora das notas de provas teóricas, salvo os vieses anteriormente discutidos.

Considerações sobre a mudança de Salto Triplo para Salto Duplo

A mudança de Salto Triplo para Duplo teve como objetivo simplificar sua execução e reduzir o estresse da terceira fase (discussão oral). O receio do corpo docente em relação a esta simplificação era perder a capacidade deste instrumento discriminar os alunos, ou seja, a capacidade de apontar como reprovado aqueles alunos que de fato tiveram rendimento (de forma geral) aquém daqueles que foram definidos como aprovados.

Para verificar se o Salto Duplo continuava a ter boa capacidade de discriminação foi feito um teste estatístico para analisar a diferença na nota da prova somativa teórica dos alunos aprovados na prova salto daqueles reprovados. Foi verificado que tanto na prova Salto Triplo (*figura 7*), como nos dois semestres onde foram aplicados provas Salto Duplo (*figuras 8 e 9*), aqueles alunos definidos como aprovados na prova Salto coincidiram com os que obtiveram melhor rendimento nas provas somativas teóricas e, portanto, foi estatisticamente significativa a capacidade de discriminação da prova Salto Duplo.

Este teste de discriminação aplicado ao salto foi uma maneira de verificar a fidedignidade do instrumento. Intrinsecamente, o teste estatístico verificou se nas diversas reaplicações da prova Salto Duplo (para diferentes alunos) a avaliação continuou “medindo”

o aluno de forma eficaz. No entanto, um questionamento pode ser feito em relação à natureza da habilidade que está sendo “medida”. Provas do estilo Salto têm intuito de verificar raciocínio e aplicação de conhecimento, que pode ou não coincidir com a habilidade que a prova somativa múltipla-escolha está verificando, que muitas vezes é mais pontual e objetiva, cobrando conhecimentos técnicos isolados.

De qualquer forma, a correlação entre ser aprovado na prova Salto e ter bons resultados na prova somativa final existiu e foi estatisticamente significativa, podendo concluir que ambas as provas estão alinhadas na habilidade que está sendo mensurada, mesmo sendo muito diferentes entre si, o que garante ao aluno instrumentos diferentes para avaliar suas habilidades e, com isto, torna o processo de avaliação mais justo e fidedigno.

Considerando que os cursos de medicina vivem o desafio de formar alunos que aprendem ao longo de toda vida (tradução para “life-long learners”) e que se tornarão experts versáteis em sua área de atuação (Dolmans et al., 2016) métodos de avaliação que motivem o estudo, especialmente no início do curso, tem papel importante no desenvolvimento da habilidade de “life-long learners”. Neste ponto, a prova Salto Duplo reforça o exercício de pesquisa e auto aprendizado, complementando o treino em resolução de problemas e pensamento crítico exigido pela sessão de tutoria.

Considerações sobre os índices de aprovação nas provas Salto Duplo

Os índices de aprovação nas provas Salto variaram de forma significativa após a mudança de Salto Triplo para Duplo, caindo o número de aprovações na prova Salto Duplo em relação ao índice que existia no Salto Triplo (*figura 10*). Um possível motivo para esta observação é a retirada do terceiro passo do Salto Triplo; sem a arguição entre tutor e aluno o

processo de avaliação pode ter se tornado mais rigoroso, tanto ao cobrar do aluno mais habilidade de escrita quanto por eliminar a discussão e arguição verbal do aluno sobre o fluxograma desenvolvido.

O fato de não existir uma padronização rígida entre as diferentes provas elaboradas entre as etapas e também ao decorrer dos semestres, permitindo que existam provas com dificuldades diferentes, pode ser vista como causa da variação no índice de aprovação da prova Salto. Semelhante ao discutido em relação à QRC.

Ao longo do ano seguinte (2017), a faculdade realizou sessões de treinamento para os tutores quanto à elaboração dos problemas para provas Salto, assim como orientações para o desenvolvimento de comandos que norteiem o aluno na construção de sua resposta. Durante o treinamento foram feitas simulações nas quais os tutores assumiram o papel de aluno, respondendo a provas Salto elaboradas com diferentes graus de “qualidade”, com objetivo de conscientizar sobre a importância da construção adequada dos instrumentos de avaliação.

Considerações sobre uso de diferentes (variados) métodos de avaliação

Segundo Nendaz & Tekian (1999), “as avaliações tem que ser desenhadas respeitando os princípios do PBL, sendo confiáveis e válidas, buscando não ter efeito negativo no aprendizado do aluno” e além de procurar avaliar os conteúdos cognitivos, deve também identificar traços latentes nos domínios afetivos e de atitudes (Gwee, 2009). Este processo é postulado como formativo e muitas vezes realizado após a sessão tutorial, onde um “feedback” sobre toda a reunião é feito entre os alunos e tutores. A avaliação do domínio das atitudes pode ser realizada por meio de notas subjetivas para itens específicos, como comunicação e trabalho em equipe (Gwee, 2009).

As avaliações formativas podem ser realizadas na forma de “feedback” onde o tutor, subjetivamente, informa aos alunos seus respectivos rendimentos durante a sessão de tutoria e estabelece uma nota para diferentes domínios, como comunicação, trabalho em equipe e raciocínio. Esse tipo de avaliação carece de fidedignidade (tradução para reliability) e de validade, reforçando a necessidade de desenvolver avaliações rigorosamente mais efetivas e confiáveis, sem fugir dos princípios do PBL (Gwee, 2009).

É sabido que a abordagem por métodos variados melhora a eficácia da avaliação (Schuwirth & van der Vleuten, 2004) e, portanto, reforça a importância de provas com questões abertas (Norman et al., 1996), como foram as QRC e a questão Salto Duplo. Contudo, postular que estas avaliações são instrumentos válidos e fidedignos é uma tarefa essencial para consolidar sua utilidade. Visto que provas mal elaboradas, que buscam por conhecimentos pontuais intensificam e estimulam o desenvolvimento de hábitos viciosos e pouco construtivos de estudo, baseados em memorizar, lembrar e regurgitar (traduzido livremente do inglês: “memorize, recall, regurgitate” chamado de MRR como acrônimo deste termo) (Gwee, 2009) .

Diversos trabalhos tentam dissociar a avaliação do desenvolvimento de habilidades relacionadas à “resolução de problemas” da avaliação de “conhecimentos específicos” do contexto do problema (Elstein, Schwartz, & Schwarz, 2002; Nendaz & Tekian, 1999), contudo, isto ainda não se mostrou possível. Consideramos que a QRC e a prova Salto têm um papel importante na avaliação dos conhecimentos específicos do aluno, reforçando seu papel no processo de avaliação.

A QRC falha ao cobrar, eventualmente, fatos pontuais e isolados, sendo que no PBL as avaliações deveriam focar predominantemente nos processos e contextos. Segundo Nendaz e Tekian (1999), “o fato dos alunos saberem uma informação em específico não

significa que saibam quando e como usá-la” e isto reforça a importância de desenvolver cuidadosamente a QRC, para que estimule o aluno a estudar e aprender, mais do que simplesmente decorar e memorizar. A avaliação Salto Duplo avalia o processo e contexto, associado à interpretação e juízo.

O motivo de escolher Questões de Resposta Curta (abertas) ou dissertativas como o Salto Duplo, envolve várias vantagens em psicometria, incluindo o fato de evitar o acerto por acaso ou aleatório (“chute”), mantendo o mesmo poder de fidedignidade e validade de um teste for bem elaborado. Entretanto, as desvantagens incluem a subjetividade ao corrigir o instrumento, a falta de precisão por não saber exatamente como os alunos interpretarão as questões, o tempo gasto para a correção (Nendaz & Tekian, 1999) e a necessidade de treinamento para elaboração de questões com alto nível de clareza, visando evitar margem de dúvidas interpretativas.

Ao longo da implantação da QRC foi nítida a dificuldade em estabelecer parâmetros para correção. Gradativamente, para contornar este empecilho, os professores foram tornando as questões mais pontuais, com respostas mais objetivas, o que facilitou a correção e reduziu o viés da subjetividade. No entanto, este direcionamento muitas vezes tem o custo de tornar as questões mais exatas e fechadas, cobrando fatos isolados e restringindo a possibilidade de avaliar o processo e contexto de uma forma ampla como a inicialmente sugerida.

Questões de resposta curta tem amplo uso na literatura, como nos testes de progresso descritos por Rademarkers e colaboradores (2005), que citam este tipo de questão como confiável, válida e fidedigna, desde que seja muito bem elaborada. Neste sentido, o autor relata que o processo de elaboração de cada QRC durou por volta de 4,7 horas. O mesmo autor afirma que as QRC foram usadas em quatro testes do progresso sucessivamente, sempre com boa fidedignidade, refletindo que seu uso deve ser estimulado, reiterando que o fato

destas questões serem pouco estudadas estatisticamente provoca a falsa ideia de inferioridade perante as questões de múltipla escolha (Rademakers et al., 2005).

Assim como neste trabalho, a publicação de Rademarkers e colaboradores (2005) observou uma melhora das notas de QRC nas aplicações sucessivas deste tipo de avaliação na instituição (*tabela 13*). A progressão das notas pode ser interpretada sob a luz de vários focos, um deles é a familiarização do aluno com o tipo de teste, outro é o ganho de expertise por parte dos examinadores e corretores, refletindo em ganhos em rendimento.

Quanto à alta frequência com que a prova QRC é aplicada (semanalmente), o estudo de seu efeito pode tomar diferentes rumos. O uso muito frequente pode levar a um efeito negativo no aluno, onde ele deixa de estudar os verdadeiros objetivos de aprendizado e passa a decorar apenas os objetivos que ele supõe que serão cobrados na prova (Swanson, Case, & Van-der-Vleuten, 1997). Os autores Nendaz e Tekian (1999) recomendam que avaliações somativas sejam espaçadas em no mínimo um mês. Entretanto, se pensarmos na avaliação QRC como instrumento formativo, o fato de ser aplicado repetidamente faria parte de sua essência como processo incentivador de melhora de aprendizado e qualidade de estudo.

A validade do instrumento de avaliação é determinada por vários fatores, como pela sua habilidade de “investigar” o que realmente se deseja saber (Haydt, 2002), indagar conceitos abrangidos pela disciplina ministrada e questionar conceitos referentes à interpretação e raciocínio sobre a matéria. Neste último critério, onde se busca acessar o conhecimento profundo (traduzido de “deep learning”), existe uma carência de evidências sobre quais os melhores métodos a se usar (Dolmans et al., 2016). Enxergamos a prova Salto Duplo como uma alternativa para esta finalidade.

O efeito provocado pela QRC e o Salto Duplo sobre a profundidade do estudo e da compreensão do aluno ainda não é bem conhecido. Segundo diversos estudos (Dolmans et

al., 2016; Marton & Säljö, 1976), o aluno pode estudar num nível superficial, apenas memorizando o texto, decorando (traduzido de “rote learning”), buscando conseguir reproduzir o que leu quando questionado. Outros alunos podem estudar de maneira mais profunda, com objetivo de compreender o texto. Neste caso, o aluno tenta extrair significado do que está lendo, relaciona com conhecimento prévio, estrutura as ideias num contexto com sentido e avalia criticamente os conceitos e conclusões apresentados pelo texto.

Devido à frequência e importância que a QRC assumiu dentro da componente curricular de tutoria, o direcionamento que faz sobre o estudo do aluno é extremamente importante. Com base nos dados observados podemos inferir que o aluno tenha conseguido estudar mais, tendo em vista a melhora das notas. Se considerarmos essas provas como válidas e confiáveis para se medir conhecimento, raciocínio e pensamento crítico, estaremos confirmando que o QRC propiciou aos alunos um estudo mais “profundo”. No entanto, esses dados precisam de mais análise para poder confirmar o direcionamento (do inglês *steering*) que a QRC fez no estudo do aluno.

Ainda refletindo sobre a influência da QRC no estudo do aluno, ressaltamos que a qualidade da pergunta seja rigorosamente conferida, cabendo às QRC questionar os conceitos chave para o problema em questão, de forma a avaliar raciocínio, compreensão e interpretação, evitando perguntas pontuais, superficiais ou que meramente questionem fatos isolados ou periféricos. Entendemos que este tipo de avaliação, especialmente devido à periodicidade com que o QRC é aplicado, possa desenvolver hábitos viciosos de estudo, direcionando para o estudo superficial, visando apenas decorar para acertar a questão, o que prejudicaria totalmente o objetivo desta avaliação. Por outro lado, segundo alguns autores como Baeten, Kyndt, Struyven, & Dochy (2010), o que direciona ou não o aluno para um determinado tipo de aprendizado não é um único fator isolado, mas múltiplos fatores

interligados, sendo que estas inter-relações ainda são desconhecidas nos ambientes de ensino centrado nos alunos.

Considerações sobre avaliação formativa

A avaliação formativa é um dos principais pilares para o PBL, onde o aluno aprende a se auto avaliar e o professor a receber um “feed-back” do processo de ensino. Este tipo de avaliação tem que ser contínua ao longo do curso e integrada à metodologia de ensino e aprendizagem (Barrows & Tamblyn, 1980). Uma revisão de literatura realizada por Nendaz e Tekian (1999) registrou uma grande proporção de artigos explicando os processos de avaliação somativa de suas escolas, contudo, algumas já incluíam rotineiramente avaliações formativas, em especial a Universidade de McMaster, berço do PBL, onde o aluno é continuamente avaliado por método formativo, sem exame final ou provas somativas intermediárias. Estes autores concluíram que a falta de evidências sobre o uso de avaliações formativas não significa que estas não sejam importantes, mas reflete a dificuldade de instalá-las como processo formal de ensino.

A avaliação formativa pode ser uma atividade complexa, como uma prova com fidedignidade diagnóstica e validade para conferir se o aluno está atingindo os objetivos propostos, mas pode ser também uma observação subjetiva, como ocorre na sessão de tutoria, ou durante o acompanhamento de um aluno em um ambulatório ou laboratório de habilidades, em que o professor consegue perceber a evolução do aluno, o ganho de habilidades, a aplicação dos conhecimentos prévios, gerando uma avaliação subjetiva sobre a adequação do método de ensino empregado, assim como do aproveitamento do aluno.

Tanto a avaliação QRC como a Salto Duplo são propriamente avaliações somativas, sua nota é definida com base na adequação ao conteúdo cobrado; o QRC é mais pontual e objetiva, contudo, devido ao fato de ser aplicada continuamente ao longo da componente curricular de tutoria, permite que o tutor e o aluno enxerguem horizontalmente a adequação do desempenho do aluno, atuando como um termômetro de seu estudo, permitindo encontrar falhas e modificá-las antes do final do módulo, propriamente como uma avaliação formativa o faria. Já o Salto Duplo reserva seu papel como instrumento somativo, avaliando tanto o conteúdo teórico quanto aspectos de raciocínio, interpretação, análise e síntese.

CONCLUSÕES

I. Os efeitos da inclusão da QRC no componente curricular de tutoria foram heterogêneos. Tanto as notas de provas somativas finais como a média final tiveram tendência para melhora na maioria das etapas do curso.

II. O agrupamento dos resultados permitiu definir, como desfecho final, que houve aumento das notas somativas finais e médias finais, contudo esse aumento foi pequeno.

III. A Avaliação da capacidade de discriminação da prova Salto Duplo mostrou equivalência à prova Salto Triplo. No entanto, ocorreu um menor índice de aprovação nas provas Salto Duplo do que o observado na prova Salto Triplo.

IV. A mudança de Salto Triplo para Duplo se mostrou aplicável, ao manter sua fidedignidade mesmo com a eliminação da dispendiosa terceira fase de prova.

V. A implementação da QRC no componente curricular de tutoria se mostrou eficaz e aplicável, dentro da perspectiva da metodologia ativa. Otimizou o estudo dos alunos e discussões em sala de modo perceptível ao favorecer as notas somativas finais e médias finais. No entanto, ainda precisa de ajustes.

CONSIDERAÇÕES FINAIS

Este trabalho suscitou discussões no corpo docente da faculdade que culminaram em importantes alterações na metodologia de avaliação do componente curricular de tutoria do curso de medicina da Faculdade FACERES. As principais observações deste trabalho que levaram a mudanças foram:

- 1) O pequeno impacto provocado pela QRC nas notas, aliado à percepção de grande carga de estresse gerada nos alunos.
- 2) O efeito possivelmente provocado na profundidade do estudo dos discentes ao forçar uma mudança na forma de estudar, priorizando acertar a questão pontual e objetiva que era aplicada, reduzindo capacidade de raciocínio.
- 3) A constatação de grande heterogeneidade entre os resultados nas diferentes turmas e etapas.

A partir destas observações (mas não limitada a elas), a faculdade considerou alguns ajustes na QRC e no funcionamento do sistema de avaliações do componente curricular de tutoria, sendo as principais delas:

- a) Reuniões frequentes entre a coordenação do componente curricular de tutoria com os tutores, a fim de padronizar a elaboração e correção dos dispositivos de avaliação, incluindo a QRC.
- b) Manutenção do uso da prova de raciocínio Salto Duplo.
- c) Redução no peso da avaliação QRC na formação da média final (de 30% para 22%).
- d) Orientação aos tutores para sempre preparar Questões de Resposta Curta que abordem raciocínio e compreensão, evitando-se questões que cobrem unicamente fatos pontuais, isolados ou de baixo nível de cognição.

- e) Definiu-se que será permitido ao aluno utilizar como material de consulta, durante a resolução da QRC, um mapa conceitual (ou resumo), elaborado previamente pelo próprio aluno, à mão, e em uma página de folha A4. Portanto o QRC passou a ser uma avaliação com consulta. Esta estratégia teve como objetivo reduzir o estresse aliado a QRC e incentivar o poder de síntese do aluno.

REFERÊNCIAS

- Allareddy, V., Havens, A. M., Howell, T. H., & Karimbux, N. Y. (2011). Evaluation of a new assessment tool in problem-based learning tutorials in dental education. *J Dent Educ*, 75(5), 665-671.
- Almeida, T. E. P., Fucuta, P. S., Pacca, F. C., Ricomini, T. E. P. A., & Cury, P. M. (2017). *Uso do gráfico de Forest Plot como estratégia estatística inovadora para estudos educacionais*. Paper presented at the Congresso Brasileiro de Educação Médica COBEM, Porto Alegre, RS.
- Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review*, 5(3), 243-260. doi:<http://dx.doi.org/10.1016/j.edurev.2010.06.001>
- Barrows, H., & Tamblyn, R. (1980). Problem-Based Learning: Rationale and Definition. In *Problem-Based Learning / An Approach to Medical Education* (pp. 1-18). New York: Springer Publishing Company.
- Boud, D., & Feletti, G. (1998). Changing problem-based learning. Introduction to the second edition. In *The Challenge of Problem-based Learning* (2^o edition ed., pp. 1-14): Kogan Page.
- Dolmans, D. H. J. M., Loyens, S. M. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: a review of the literature. *Adv Health Sci Educ Theory Pract*, 21(5), 1087-1112. doi:10.1007/s10459-015-9645-6
- Elstein, A. S., Schwartz, A., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*, 324(7339), 729-732.
- Epstein, R. M. (2007). Assessment in medical education. *N Engl J Med*, 356(4), 387-396. doi:10.1056/NEJMra054784
- Friedman, C. P., Donaldson, K. M., & Vantsevich, A. V. (2016). Educating medical students in the era of ubiquitous information. *Med Teach*, 38(5), 504-509. doi:10.3109/0142159X.2016.1150990
- Gwee, M. C. (2009). Problem-based learning: a strategic learning system design for the education of healthcare professionals in the 21st century. *Kaohsiung J Med Sci*, 25(5), 231-239. doi:10.1016/s1607-551x(09)70067-1
- Haydt, R. (2002). Funções, modalidades e propósitos da avaliação. In *Avaliação do Processo de Ensino-Aprendizagem* (6^a edição ed., pp. 17-18). São Paulo: Ática.
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I-outcome and process. *British journal of educational psychology*, 46(1), 4-11. doi:10.1111/j.2044-8279.1976.tb02980.x
- Mehay, R., & Burns, R. (2009). Miller's Prism Of Clinical Competence. In. Internet site: http://www.gp-training.net/training/educational_theory/adult_learning/miller.htm acessado 13/09/2017.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Acad Med*, 65(9 Suppl), S63-67.
- Navazesh, M., Rich, S. K., Chopiuk, N. B., & Keim, R. G. (2013). Triple jump examinations for dental student assessment. *J Dent Educ*, 77(10), 1315-1320.
- Nendaz, M., & Tekian, A. (1999). Assessment in Problem-Based Learning Medical Schools: A Literature Review. *Teaching and Learning in Medicine*, 11(4), 232-243. doi:10.1207/S15328015TLM110408

- Norman, G. R., Case, S. M., & Swanson, D. B. (1996). Conceptual and Methodological Issues in Studies Comparing Assessment Formats. *Teaching and Learning in Medicine*, 8(4), 208-216.
- Papinczak, T. (2010). assessment: An exploration of perceptions of tutor evaluation in problem-based learning tutorials. *Medical Education*, 44(9), 892-899. doi:10.1111/j.1365-2923.2010.03749.x
- Rademakers, J., Ten Cate, T. J., & Bär, P. R. (2005). Progress testing with short answer questions. *Med Teach*, 27(7), 578-582. doi:10.1080/01421590500062749
- Santos, A. M., Salgado, A., Barreto, J. F., Martins, H., & Dores, A. r. (2010). Problem-Based Learning e suas implicações: Breve revisão teórica. *ESTSP-IPP*, 1, 1-8. Retrieved from [http://recipp.ipp.pt/bitstream/10400.22/1404/1/COM AndreiaSantos_2010.pdf](http://recipp.ipp.pt/bitstream/10400.22/1404/1/COM_AndreiaSantos_2010.pdf)
- Savery, J. R. (2006). Overview of Problem-based Learning: Definitions and Distinctions. *Interdisciplinary Journal of Problem-Based Learning*, 1(1), 9-20.
- Schuwirth, L. W., & van der Vleuten, C. P. (2004). Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ*, 38(9), 974-979. doi:10.1111/j.1365-2929.2004.01916.x
- Swanson, D., Case, S., & Van-der-Vleuten, C. (1997). Strategies for student assessment. In D. Boud & G. Feletti (Eds.), *The Challenge of Problem-Based Learning* (pp. 269-282). London: Kogan Page.
- Tabish, S. (2008). Assessment methods in medical education. *Int J Health Sci (Qassim)*, 2(2), 3-7.
- Tavakol, M., & Dennick, R. (2017). The foundations of measurement and assessment in medical education. *Med Teach*, 1-6. doi:10.1080/0142159X.2017.1359521
- Walker, A., & Leary, H. (2009). A Problem Based Learning Meta Analysis: Differences Across Problem Types, Implementation Types, Disciplines, and Assessment Levels. *Interdisciplinary Journal of Problem-Based Learning*, 3(1). doi:10.7771/1541-5015.1061
- Wang, Q., Li, H., Pang, W., Liang, S., & Su, Y. (2016). Developing an integrated framework of problem-based learning and coaching psychology for medical education: a participatory research. *BMC Med Educ*, 16(1), 2. doi:10.1186/s12909-015-0516-x
- Wood, E. J. (2004). Problem-based learning. *Acta Biochimica Polonica*, 51(2), 21-26.

ANEXOS

Anexo 1 – Parecer do comitê de ética



FACULDADE DE MEDICINA DE
SÃO JOSÉ DO RIO PRETO-
FAMERP - SP



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Papel da Avaliação Somativa Entre as Seções Tutoriais no Ensino Baseado em Problemas

Pesquisador: THOMAS EUGENIO PORTES DE ALMEIDA

Área Temática:

Versão: 2

CAAE: 65687417.4.0000.5415

Instituição Proponente: Faculdade de Medicina de São José do Rio Preto- FAMERP - SP

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 2.074.864

Apresentação do Projeto:

Trata-se de um estudo transversal, observacional e descritivo acerca de uma nova modalidade de avaliação dos alunos do curso de Medicina da FACERES. Serão coletadas, do sistema informatizado de notas da Faculdade de Medicina FACERES, as notas das avaliações dissertativas de respostas curtas e das provas somativas da disciplina de tutoria dos alunos da primeira turma (Ingressantes no primeiro semestre de 2012) a nona turma (ingressantes no segundo semestre de 2016) da faculdade FACERES, também serão coletados dados relacionados aos índices de assiduidade dos alunos nas reuniões de tutoria destas mesmas turmas. Os dados avaliados serão referentes a um semestre antes e dois semestres depois da implantação da prova de resposta curta (também chamada de prova kinder ovo, ou abreviada como QRC). Estudos estatísticos avaliarão o impacto desta nova prova no desempenho dos alunos.

Objetivo da Pesquisa:

Estudar uma nova modalidade de avaliação, chamada de prova "Kinder Ovo" no desempenho dos alunos na disciplina de tutoria.

Avaliação dos Riscos e Benefícios:

Os riscos para os participantes serão mínimos e os benefícios estão diretamente associados aos possíveis achados da pesquisa.

Endereço: BRIGADEIRO FARIA LIMA, 5416

Bairro: VILA SAO PEDRO

CEP: 15.090-000

UF: SP

Município: SAO JOSE DO RIO PRETO

Telefone: (17)3201-5813

Fax: (17)3201-5813

E-mail: cepfamerp@famerp.br



FACULDADE DE MEDICINA DE
SÃO JOSÉ DO RIO PRETO-
FAMERP - SP



Continuação do Parecer: 2.074.864

Comentários e Considerações sobre a Pesquisa:

Trata-se de uma pesquisa factível, com elevado potencial para produção de novos conhecimentos na área e sem maiores implicações à Ética ou à segurança dos participantes.

Considerações sobre os Termos de apresentação obrigatória:

Presentes e adequados. O estudo propões dispensa do TCLE, uma vez que, segundo os proponentes, serão coletados dados do sistema digital de notas da faculdade, apenas serão estudados dados retrospectivos referentes a uma proposta de avaliação que já foi aplicada, as planilhas utilizadas para análise não contem informações que permitam identificar diretamente os alunos.

Recomendações:

Recomenda-se a aprovação do presente estudo e o deferimento da dispensa do TCLE.

Conclusões ou Pendências e Lista de Inadequações:

Não há.

Considerações Finais a critério do CEP:

Projeto Aprovado.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_868835.pdf	16/05/2017 16:26:51		Aceito
Projeto Detalhado / Brochura Investigador	projeto famerp.docx	16/05/2017 16:26:33	THOMAS EUGENIO PORTES DE ALMEIDA	Aceito
Declaração de Instituição e Infraestrutura	declaracaoassinada.pdf	16/05/2017 16:24:47	THOMAS EUGENIO PORTES DE ALMEIDA	Aceito
Folha de Rosto	folha_de_rostoassinada.pdf	16/05/2017 16:23:08	THOMAS EUGENIO PORTES DE ALMEIDA	Aceito
Outros	declaracaoDEResponsabilidade.pdf	13/03/2017 20:00:00	THOMAS EUGENIO PORTES DE ALMEIDA	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Endereço: BRIGADEIRO FARIA LIMA, 5416

Bairro: VILA SAO PEDRO

CEP: 15.090-000

UF: SP

Município: SAO JOSE DO RIO PRETO

Telefone: (17)3201-5813

Fax: (17)3201-5813

E-mail: cepfamerp@famerp.br



FACULDADE DE MEDICINA DE
SÃO JOSE DO RIO PRETO-
FAMERP - SP



Continuação do Parecer: 2.074.864

Não

SAO JOSE DO RIO PRETO, 22 de Maio de 2017

Assinado por:
GERARDO MARIA DE ARAUJO FILHO
(Coordenador)

Endereço: BRIGADEIRO FARIA LIMA, 5416

Bairro: VILA SAO PEDRO

CEP: 15.090-000

UF: SP

Município: SAO JOSE DO RIO PRETO

Telefone: (17)3201-5813

Fax: (17)3201-5813

E-mail: cepfamerp@famerp.br

Anexo 2 – Relatório Parcial / Final CEP

 <p>Comitê de Ética em Pesquisa em Seres Humanos CEP/FAMERP</p>	<p>Em 31 de Janeiro de 2018</p>
<p>PREZADO(A) SENHOR (A):</p>	
<p>Pelo presente, envio o relatório parcial/final para informar sobre o andamento do projeto abaixo relacionado:</p>	
<p>N.º / CAAE do projeto: 65687417.4.0000.5415</p>	
<p>Título: Papel Da Avaliação Somativa Entre As Seções Tutoriais No Ensino Baseado Em Problemas</p>	
<p>Pesquisador: Thomas Eugenio Portes de Almeida</p>	
<p>Relatório N.º: 1</p>	
<p>Estudo Encerrado? Sim Terminou em 27/11/2017</p>	
<p>Caráter do Estudo: MESTRADO</p>	
<p>Houve Emendas ao Projeto Inicial? NÃO</p>	
<p>Houve necessidade de alteração na estrutura do projeto? Título? Material? Método? Objetivos? Se sim, detalhar e justificar alterações. SIM</p>	
<p>O título do projeto foi alterado para “<u>Novos Instrumentos de Avaliação no Componente Curricular de Tutoria em Curso de Medicina por Metodologia Ativa</u>”</p>	
<p>A Justificativa para esta mudança foi a seguinte: Inicialmente, este projeto, visava estudar o impacto, exclusivamente, da “Questão de Resposta Curta (QRC)” retratada no título anterior como “avaliação somativa realizada entre as seções tutoriais”, no entanto, ao longo da execução da análise de dados, foi observado que outra mudança nos métodos de avaliação ocorreu no mesmo período, especificamente, a mudança da prova “salto triplo” para “salto duplo”. Ao longo da análise quantitativa dos dados ficou claro que essa mudança influenciaria de forma tão importante quanto à avaliação QRC, que inicialmente foi proposta.</p>	
<p>Av. Brigadeiro Faria Lima, - 5416 - Vila São Pedro CEP: 15090-000 - São José do Rio Preto - SP / (17) 3201-5813</p>	
	



Comitê de Ética em
Pesquisa em Seres Humanos
CEP/FAMERP

Portanto, o objetivo do trabalho passou a incluir estudar o impacto e efetividade tanto da implementação da QRC no componente curricular de Tutoria como também da prova "Salto Duplo". Deixo claro que em nenhum momento esta nova avaliação obrigou qualquer tipo de coleta adicional de dados, ou mudança nos aspectos éticos do trabalho, tendo em vista que foi apenas uma exploração adicional dos resultados obtidos na análise estatística dos dados já obtidos inicialmente.

Foi apresentado em Congresso () reunião Qualificação de Mestrado

Foi publicado? **Não**

Data da Inclusão 1ª Participante: **Não foram alocados participantes diretamente, tendo em vista que o projeto avaliou exclusivamente base de dados.**

Número de Participantes Inclusos na Pesquisa até o momento: **A base de dados estudada inclui dados válidos de 504 alunos, no momento da coleta.**

O estudo apresentou alguma intercorrência? **Não** Algum participante deixou de assinar Termo de Consentimento Livre e Esclarecido? **Não se aplica** Houve algum evento adverso durante a execução? **Não se aplica**

Por quanto tempo mais o estudo se estenderá? Caso haja alteração, anexar novo cronograma. **Estudo encerrado**

Resultados parciais ou totais já publicados: Foi publicado, na forma de Poster, no congresso mundial de educação médica (AMEE, Helsinque, Finlândia) 26-30 de agosto de 2017, sob título: **"Double jump" assessment as an alternative way to improve clinical reasoning in undergraduate medical students.** Traduzindo livremente para o português seria: **Avaliação "Salto Duplo" como uma forma alternativa de melhorar raciocínio clínico em graduandos do curso de medicina.**

Thomas



Comitê de Ética em
Pesquisa em Seres Humanos
CEP/FAMERP

O pôster envolvia tanto aspectos institucionais sobre a elaboração e aplicação da prova "Salto Duplo", que fogem do escopo deste trabalho, por se tratar de mudanças curriculares que ocorreram antes mesmo deste estudo. Também incluiu resultados quantitativos que foram derivados deste trabalho, sendo eles: A demonstração de equivalência, com $p < 0.0001$ entre a efetividade do "Salto Triplo" com o "Salto Duplo". As provas "salto" separam os alunos de forma dicotômica, tendo como nota "suficiente" ou "insuficiente" sendo que os alunos com notas "suficiente" tiveram médias significativamente maiores nos exames finais do que as médias em exames finais dos alunos com nota "insuficiente" nas provas Salto.

Outras informações Relevantes: Os resultados de todo o projeto foram completamente apresentados durante a qualificação para mestrado do autor deste estudo, no dia 27 de novembro de 2017. No evento não foi solicitado nenhuma correção metodológica.

Sem mais, coloco-me a disposição para demais esclarecimentos.

Assinatura do pesquisador



FACULDADE DE MEDICINA DE
SÃO JOSÉ DO RIO PRETO-
FAMERP - SP



Continuação do Parecer: 2.495.357

Conclusões ou Pendências e Lista de Inadequações:

Sem pendências ou inadequações.

Considerações Finais a critério do CEP:

O Comitê de Ética em Pesquisa em Seres Humanos da Faculdade de Medicina de São José do Rio Preto está ciente do Relatório Final datado de 21 de janeiro de 2018, referente ao projeto CAAE: 65687417.4.0000.5415 sob responsabilidade de Thomas Eugenio Portes de Almeida com o título "Papel da Avaliação Somativa Entre as Seções Tutoriais no Ensino Baseado em Problemas".

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Envio de Relatório Parcial	relatorioparcial.pdf	12/02/2018 11:30:12	THOMAS EUGENIO PORTES DE ALMEIDA	Postado

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

SAO JOSE DO RIO PRETO, 15 de Fevereiro de 2018

Assinado por:

GERARDO MARIA DE ARAUJO FILHO
(Coordenador)

Endereço: BRIGADEIRO FARIA LIMA, 5416

Bairro: VILA SAO PEDRO

CEP: 15.090-000

UF: SP

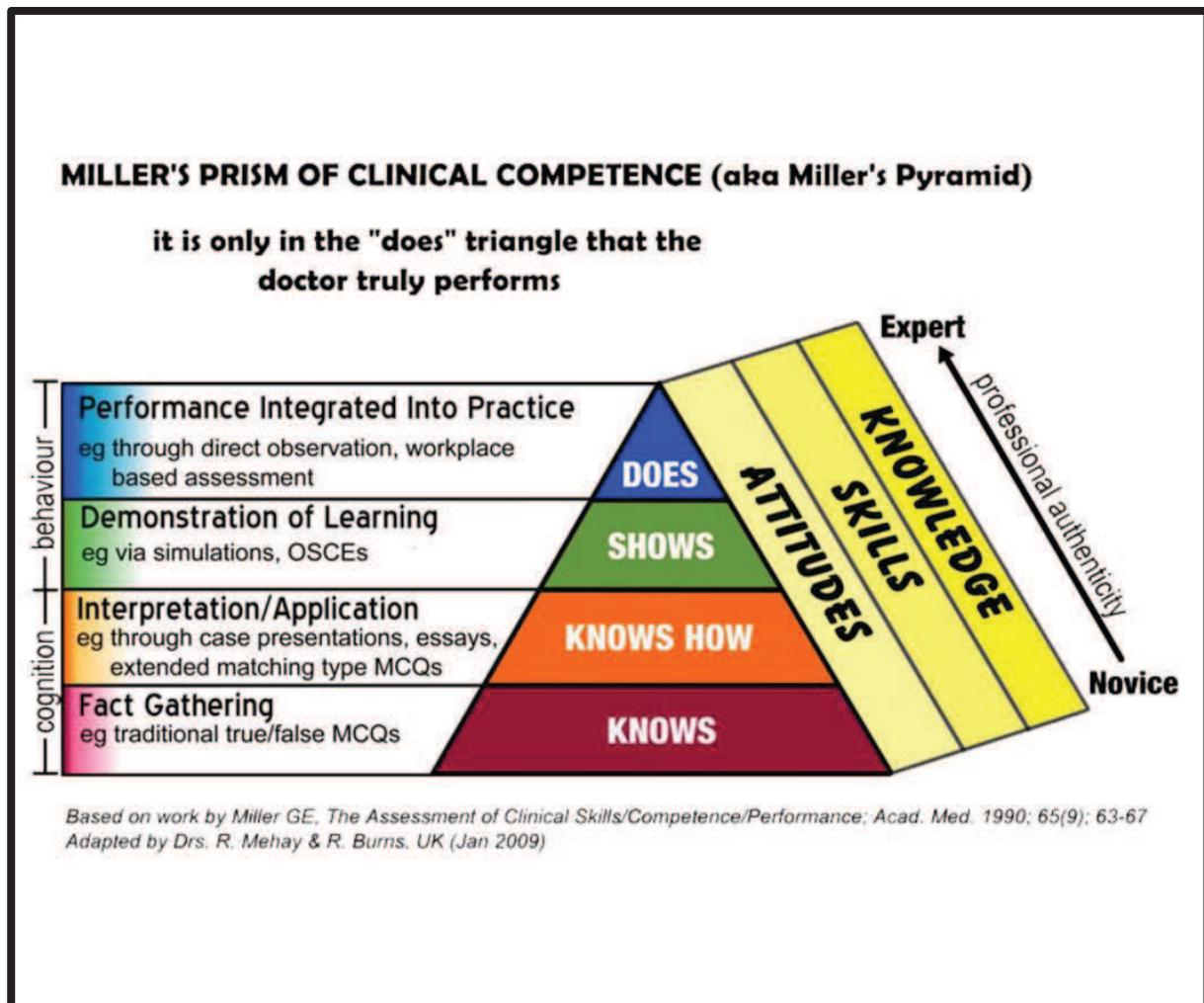
Município: SAO JOSE DO RIO PRETO

Telefone: (17)3201-5813

Fax: (17)3201-5813

E-mail: cepfamerp@famerp.br

Anexo 4 – Pirâmide de Miller



(Mehay & Burns, 2009)