

UNIVERSIDADE JOSÉ DO ROSÁRIO VELLANO – UNIFENAS
Galileu Bonifácio da Costa Filho

IMPACTO DA REFLEXÃO ESTRUTURADA NA CALIBRAGEM DIAGNÓSTICA
DE ALUNOS DO 6º ANO DO CURSO DE MEDICINA DA UNIFENAS BH

Belo Horizonte

2017

Galileu Bonifácio da Costa Filho

**IMPACTO DA REFLEXÃO ESTRUTURADA NA CALIBRAGEM DIAGNÓSTICA
DE ALUNOS DO 6º ANO DO CURSO DE MEDICINA DA UNIFENAS BH**

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

Orientador: Prof. Dr. Alexandre Sampaio Moura.

Belo Horizonte

2017

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

611.77

C837i

Costa Filho, Galileu Bonifácio da.

Impacto da reflexão estruturada na calibragem diagnóstica de alunos do 6º ano do curso de medicina da Unifenas BH [manuscrito] / Galileu Bonifácio da Costa Filho. -- 2017.

52f. : il.

Dissertação (Mestrado) - Universidade José do Rosário Vellano, Programa de Pós-Graduação Stricto Sensu em Ensino em Saúde, 2017.

Orientador : Prof. Dr. Alexandre Sampaio Moura.

1. Calibragem. 2. Acurácia. 3. Reflexão Estruturada. 4. Dermatologia. 5. Confiança. I. Moura, Alexandre Sampaio de. II. Título.

Bibliotecária responsável: Kely A. Alves CRB6/2401

Certificado de Aprovação

"IMPACTO DA REFLEXÃO ESTRUTURADA NA CALIBRAGEM DIAGNÓSTICA DE ALUNOS DO 6º ANO DO CURSO DE MEDICINA DA UNIFENAS BH".


AUTOR: Galileu Bonifácio da Costa Filho

ORIENTADOR: Prof. Dr. Alexandre Sampaio Moura

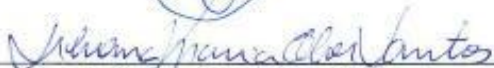
Aprovado como parte das exigências para obtenção do Título de Mestre Profissional em Ensino em Saúde pela Comissão Examinadora.



Prof. Dr. Alexandre Sampaio Moura




Prof. Dr. Josemar de Almeida Moura



Profa. Dra. Silvana Maria Eloi Santos

Belo Horizonte, 16 de dezembro de 2017.



Prof. Dr. Antônio Carlos de Castro Toledo JR.
Coordenador do Mestrado Profissional
Em Ensino em Saúde
UNIFENAS

LISTA DE ILUSTRAÇÕES

Figura 1	- Análise entre <i>insight</i> /desempenho.....	10
Gráfico 1	- Comparação da confiança e acurácia global entre os grupos reflexão e controle.....	22
Gráfico 2	- Comparação da confiança e acurácia em lesões melanocíticas e não melanocíticas entre os grupos reflexão e controle.....	23
Gráfico 3	- Análise comparativa entre os 2 grupos de estudos quanto à acurácia, no global e por síndrome dermatológica.....	24
Gráfico 4	- Análise comparativa entre os 2 grupos de estudos quanto à confiança, no global e por síndrome dermatológica.....	24
Gráfico 5	- Análise comparativa entre os 2 grupos de estudos quanto à calibragem do teste, no global e por síndrome dermatológica.....	25
Gráfico 6	- Comparação da confiança e acurácia em lesões fáceis e difíceis entre os grupos reflexão e controle.....	28
Gráfico 7	- Comparação da calibragem em lesões fáceis e difíceis entre os grupos reflexão e controle.....	29

LISTA DE TABELAS

Tabela 1	- Análise comparativa entre os 2 grupos de estudo quanto ao conhecimento e à experiência prévios em lesões dermatológicas.....	19
Tabela 2	- Comparação entre os grupos, controle e reflexão quanto ao acerto de cada um dos 12 casos estudados.....	20
Tabela 3	- Análise comparativa entre os 2 grupos de estudos quanto à confiança do aluno no diagnóstico em cada um dos 12 casos estudados.....	21
Tabela 4	- Análise comparativa entre os 2 grupos de estudos quanto à acurácia, à confiança e à calibragem do teste, no global e por síndrome dermatológica	22
Tabela 5	- Análise comparativa entre as síndromes quanto à acurácia, à confiança e à calibragem do teste, por grupo.....	26
Tabela 6	- Análise de Correlação entre o percentual de acerto previsto pelo aluno e as medidas de acurácia, confiança e calibragem do teste no global e por grupo.....	26
Tabela 7	- Análise comparativa entre os 2 grupos de estudos quanto à acurácia, à confiança e à calibragem do teste, no global e pela dificuldade da resolução da questão.....	29

RESUMO

Introdução: A calibragem diagnóstica pode ser definida como a relação entre acurácia e confiança do profissional em relação a esta acurácia. A prática reflexiva é uma potencial estratégia para melhorar a calibragem diagnóstica, pois pode aumentar o *insight* sobre a performance. O uso da reflexão estruturada para melhoria da calibragem diagnóstica ainda não foi avaliado. **Objetivo:** Avaliar o impacto da reflexão estruturada na calibragem diagnóstica, acurácia e confiança na resolução de casos dermatológicos. **Métodos:** estudo realizado entre alunos do 6º ano do curso de medicina da universidade José do Rosário Vellano (UNIFENAS-BH). Os alunos foram divididos em dois grupos (grupo reflexão (GR) e grupo controle (GC)). Os dois grupos avaliaram as mesmas 12 lesões dermatológicas, apontando diagnóstico inicial mais provável e a confiança diagnóstica. Em seguida, os alunos foram orientados a refletirem de maneira estruturada sobre o diagnóstico (GR) ou engajarem em atividade distratora (GC) e então fornecerem o diagnóstico final e o grau de confiança. Foram comparadas acurácia diagnóstica, confiança e calibragem diagnóstica globalmente, por síndrome dermatológica e por grau de dificuldade. **Resultados:** Foram incluídos 61 alunos, alocados 33 para GR e 28 para GC. Houve predominância de mulheres no GR, mas não houve diferença significativa de idade ou de auto-percepção do conhecimento/experiência prévios entre os grupos. Observou-se uma maior acurácia diagnóstica dos alunos do GR em relação ao GC tanto para síndromes melanocíticas ($p=0,025$) quanto para não-melanocíticas ($p=0,022$). Em relação à calibragem diagnóstica, tanto o GC ($p=0,017$) quanto o GR ($p=0,001$) apresentaram melhor calibragem para síndromes melanocíticas, sendo que para síndromes não-melanocíticas o aluno apresenta maior excesso de confiança. A calibragem foi pior para os casos difíceis com os alunos apresentando excesso de confiança em relação ao seu real desempenho. A reflexão não impactou de maneira significativa em relação à confiança e calibragem de maneira global. Nos casos fáceis, houve diferença de calibragem entre GR e GC, observando-se no GR maior acurácia diagnóstica, mas com confiança semelhante à observada do GC, ocorrendo inversão do sinal da calibragem com os alunos tendo um desempenho levemente superior à sua confiança. **Conclusão:** A atividade reflexiva aumentou a acurácia diagnóstica de maneira global para as lesões dermatológicas testadas, mas não alterou globalmente a confiança e calibragem diagnóstica. A calibragem piorou nos casos mais difíceis e a reflexão não foi capaz de melhorar esta situação.

Palavras chave: Calibragem. Acurácia. Reflexão Estruturada. Dermatologia. Confiança

ABSTRACT

Introduction: Diagnostic calibration can be defined as the relationship between diagnostic accuracy and the confidence in that accuracy. Reflective practice seems to be an interesting strategy to improve diagnostic calibration by improving performance insights and has been proposed to foster clinical reasoning by improving diagnostic accuracy. However, its effect on diagnostic calibration has yet to be assessed. **Aim:** This study seeks to evaluate the impact of structured reflection on diagnostic calibration, diagnostic accuracy, and the confidence in that accuracy while solving dermatology case vignettes. **Methods:** This study recruited sixth year medical students from Universidad José do Rosário Vellano, who were systematically allocated into one of two groups: reflection group (RG) or control group (CG). Students from both groups were presented the same 12 dermatology cases and were initially asked to point out the probable diagnosis and their confidence in that diagnosis. The students allocated to the RG were then told to reflect on the case using a structured tool. Those allocated to the CG engaged in a time-filler activity before providing their final diagnosis for each case and their final confidence in that diagnosis. Both groups were compared as to their diagnostic accuracy, confidence, and diagnostic calibration overall as well as stratified by dermatological syndrome and case difficulty. **Results:** Sixty-one students were included (33 in the RG and 28 in the CG). Age and self-assessed perception of knowledge/previous experience were similar in both groups. Gender was significantly different among the groups. Students from the RG showed a better diagnostic accuracy for both melanocytic and non-melanocytic lesions ($p=0.025$ and $p=0.022$, respectively). Regarding diagnostic calibration, both RG and CG showed better calibration for melanocytic lesions than for non-melanocytic lesions. Calibration for non-melanocytic lesions was more distant from zero, in a positive direction (i.e. greater overconfidence). It was observed that calibration was worse among more difficult cases, with students showing greater overconfidence when solving such cases. Reflection did not impact significantly on confidence nor on calibration ($p=0.228$ and $p=0.197$, respectively). It is worth noting that a significant difference was observed in the calibration in both groups when solving easier cases; students from the RG had a greater diagnostic accuracy, but their confidence did not differ from those in the CG, resulting in an opposite direction of calibration (i.e. among students in the RG, diagnostic accuracy was slightly greater than their confidence). **Conclusion:** Reflective practice increased the overall diagnostic accuracy, but it did not significantly change overall confidence and calibration.

Calibration was worse when solving more difficult cases, and reflection was not enough to improve it.

Keywords: Calibration. Accuracy. Structured Reflection. Dermatology. Confidence

SUMÁRIO

1	INTRODUÇÃO	08
2	OBJETIVOS	13
2.1	Objetivo geral.....	13
2.2	Objetivos específicos.....	13
3	MATERIAL E MÉTODOS	14
3.1	Local do estudo	14
3.2	População	14
3.3	Critérios de inclusão.....	14
3.4	Critérios de exclusão	14
3.5	Procedimentos	14
3.6	Aspectos éticos	16
3.7	Análise de dados	16
4	RESULTADOS	19
5	DISCUSSÃO	31
6	CONCLUSÃO.....	35
	REFERÊNCIAS	36
	APÊNDICES	38

1 INTRODUÇÃO

Erros médicos são importantes causas de morbimortalidade no mundo e são frequentemente associados ao mau desempenho profissional (MAKARY; MICHAEL, 2016). Uma revisão recente da literatura estimou entre 210.000 e 400.000 o número de mortes anuais associadas a erros médicos em pacientes internados, colocando o erro médico como a terceira causa mais comum de morte nos Estados Unidos, atrás apenas das mortes por doenças cardíacas e câncer (MAKARY; MICHAEL, 2016).

O erro humano é inevitável e embora não possamos eliminá-lo, podemos desenvolver estratégias para reduzi-lo. A maneira como os médicos desenvolvem sua expertise e adquirem novos conhecimentos está relacionada com a habilidade de refletirem sobre seu desempenho. De acordo com Berlinger e Wu (2005), erros médicos podem ser parcialmente atribuídos à falta de reflexão do profissional em sua prática clínica. Muitos médicos com desempenho clínico abaixo do esperado não têm consciência disso por não refletirem sobre suas práticas sugerindo que a capacidade de transformação do médico requer *insight* que serviria como uma motivação para a mudança.

Hayset al. (2002) sugerem que *insight* é ter consciência e deve ser compreendido como a combinação de três construtos: a consciência de seu próprio desempenho, consciência da performance dos outros e a capacidade de refletir sobre ambos e fazer seus julgamentos (BECKETT; HAGER, 2002). A literatura sobre *insight* é vasta e complexa. Na teoria psicológica, *insight* era um termo primeiro descrito por Kohler para mostrar o comportamento de macacos em resolver enigmas (GOULD; REAN, 1994). Os macacos progrediam em seus saltos (*insights*) na tentativa de capturar objetos, para, estrategicamente, localizar em alimentos. O termo *insight* foi generalizado para outras formas como “aha!”, para resolver enigmas ou outros problemas mais complexos, onde a resolução parece surgir do nada.

Insights, em relação ao desempenho e a capacidade de mudar este desempenho, são nitidamente relevantes para o desenvolvimento profissional dos médicos, particularmente para aqueles cujo desempenho é fraco. Por exemplo, o médico que tem consciência de sua deficiência em conduzir um diabetes, pode estudar mais o tema ou encaminhar para outros colegas que conduzem melhor que ele.

Como o desenvolvimento de *insight* é fundamental, o uso de estratégias que fomentem a reflexão pode ser potencialmente útil. O uso do *feedback* adequado pode aprimorar o raciocínio clínico por tornar visíveis os erros diagnósticos, identificar pontos fracos e fortes, levando à melhora, também, na calibração. Entretanto, intervenções utilizando o simples uso de *feedback* podem não ser suficientes para corrigir o excesso de confiança (SIECK; ARCKS, 2005).

Dentro das estratégias com foco no indivíduo, a prática reflexiva é uma abordagem definida como habilidade dos médicos em analisar de maneira crítica suas condutas e decisões durante as atividades profissionais. Esta estratégia incorpora os princípios da metacognição dentre os quais destacam-se: (a) capacidade de pensar e explorar hipóteses alternativas quando os casos forem pouco familiares ou complexos, (b) testar as hipóteses frente aos fatos, (c) abertura para reflexão e habilidade para lidar com incertezas (BERNER; GRABER, 2008). Desta forma, a prática reflexiva parece constituir uma interessante estratégia para melhorar a calibragem diagnóstica.

A calibragem diagnóstica, objeto de nosso estudo, pode ser definida como a relação entre a acurácia diagnóstica e a confiança do profissional em relação a esta acurácia. Há diferentes níveis de percepção entre os indivíduos sobre seu desempenho. Hayset al. (2002) ilustraram diferentes categorias de calibragem diagnóstica, definida como a relação entre acurácia diagnóstica e confiança (FIG. 1). O médico classificado como “inconsciente incompetente” (letra G) além de apresentar uma baixa performance, não tem a percepção de que está falhando em seus diagnósticos. É um grupo de difícil abordagem, pois este profissional, provavelmente, não identificará a necessidade de aperfeiçoar sua prática. O médico “inconsciente competente” (letra I) também não configura o ideal, pois a insegurança em relação a sua prática pode levar à ansiedade excessiva ou insegurança que podem vir a impactar no seu desempenho, bem como resultar em encaminhamentos desnecessários ou solicitação excessiva de exames complementares.

Figura 1 - Análise entre *insight*/desempenho

↑ <i>Insight</i>		
Alto <i>Insight</i> Baixo Desempenho A	Alto <i>Insight</i> Desempenho Padrão B	Alto <i>Insight</i> Alto Desempenho C
<i>Insight</i> Padrão Baixo Desempenho D	<i>Insight</i> Padrão Desempenho Padrão E	<i>Insight</i> Padrão Alto Desempenho F
Baixo <i>Insight</i> Baixo Desempenho G	Baixo <i>Insight</i> Desempenho Padrão H	Baixo <i>Insight</i> Alto Desempenho I
Desempenho →		

Fonte: Hayset al. (2002)

Estratégias para melhorar a calibragem entre acurácia e confiança são claramente relevantes na educação permanente dos médicos, particularmente para aqueles com performance abaixo do esperado. Sem *insight* para mudar, as atividades educacionais se tornam pouco relevantes e dificultam o aprendizado significativo e a mudança da prática.

O excesso de confiança, frequentemente, ocorre durante uma sequência de ações e deve ser avaliado no contexto do julgamento e decisão. Parece ser influenciado por um número de fatores relacionados ao indivíduo e ao tipo de atividade desempenhada. Nas decisões médicas, frequentemente, o excesso de confiança resulta em atraso e erros diagnósticos, com efeitos prejudiciais (BERNER; GRABER, 2008).

Há uma variedade de discussões do porquê o médico sozinho demonstra excesso de confiança no seu julgamento. Isto é percebido como um viés cognitivo comum e pode ser propagado como componente de padrão comportamental dentro da cultura da medicina.

Numerosos estudos tentam corrigir falhas no raciocínio clínico e nas tomadas de decisões (CROSKERRY, 2007). A falta de consciência e de *insight* nas decisões teóricas pode ser corrigida introduzindo treinamentos específicos na graduação enfatizando treinamentos para

tomada de decisão levando em conta o contexto, assim como ressaltando as vulnerabilidades e potenciais vieses das diferentes maneiras de se tomar uma decisão.

Vieses cognitivos e afetivos podem ser trabalhados com treinamentos específicos em diferentes níveis da graduação utilizando exemplos que ilustrem esses vieses com sugestão de estratégias de correção. A falta de pensamento crítico pode ser corrigida introduzindo capacitações, precocemente no currículo, que abordem os princípios básicos do pensamento crítico. A falsa sensação de segurança pode ser combatida com treinamentos visando superar barreiras pessoais e culturais contra a admissão da incerteza e o reconhecimento de que a certeza não é sempre possível. Estes treinamentos permitem que o aluno aceite que o “diagnóstico não foi fechado ainda”.

Podem ainda ser adotadas estratégias cognitivas que levem o aluno a considerar hipóteses concorrentes e evidências conflitantes

Nenhuma dessas estratégias têm demonstrado superioridade de uma estratégia e a redução do excesso de confiança possivelmente requer a combinação de atividades (KORLAT; LICHTENSTEIN; FISCHOFF, 1980).

Schon (1987) expandiu a teoria da reflexão e desenvolveu o conceito da prática reflexiva. Desde então, a reflexão passou a ser incorporada na pós-graduação, na graduação e no ensino médio. Para Rego, Billett e Morgan (2009) os benefícios da prática reflexiva para os praticantes incluem a redefinição dos conhecimentos profissionais, a expansão do autoconhecimento ou conhecimento pessoal e a reavaliação das ações apropriadamente.

A ideia de reflexão como uma ferramenta de aprendizagem foi primeira colocada por John Dewey, um filósofo educacional. Dewey achou que a reflexão começa com uma dificuldade inesperada, um problema ou um desafio (CUNHA, 2009). Isto então promove no pensador um exame de suas fontes e verificação das evidências para defendê-las, além de considerar as expectativas situacionais, para só assim, resolver a dúvida.

O uso da reflexão estruturada no desenvolvimento do raciocínio clínico já foi bem avaliado e, comprovadamente, acelera o ganho desta habilidade pelos alunos (MAMEDE et al., 2012). A associação de casos clínicos com a reflexão estruturada também mostrou melhores resultados

em estudos anteriores (MAMEDE et al., 2014). Em um estudo com 42 residentes que resolveram 16 casos clínicos, a reflexão estruturada mostrou melhores resultados em casos difíceis (MAMEDE; SCHMIDT; PENAFORTE, 2008).

Além de contribuir para a aprendizagem do raciocínio clínico, a prática reflexiva, ao levar o aluno a considerar hipóteses concorrentes e evidências conflitantes, pode constituir uma interessante estratégia para reduzir o excesso de confiança e melhorar a calibragem diagnóstica.

2 OBJETIVOS

2.1 Objetivo geral

Avaliar o impacto da reflexão estruturada na calibragem diagnóstica de alunos do 6ºano do curso de medicina da Unifenas-BH, no diagnóstico de lesões clínicas dermatológicas.

2.2 Objetivos específicos

- Avaliar o impacto da reflexão estruturada na acurácia diagnóstica de lesões dermatológicas por alunos de medicina.
- Avaliar o impacto da reflexão estruturada na confiança dos alunos de medicina na resolução de lesões dermatológicas.
- Analisar a interação entre o tipo de lesão dermatológica e o efeito da reflexão estruturada na calibragem diagnóstica
- Analisar a interação entre dificuldade diagnóstica da lesão dermatológica e o efeito da reflexão estruturada na calibragem diagnóstica.

3 MATERIAL E MÉTODOS

3.1 Local do estudo

O estudo foi realizado entre alunos do 6º ano, no Curso de Medicina do Campus da Unidade de Belo Horizonte da Universidade José do Rosário Vellano- Unifenas. Os alunos do curso de Medicina da Unifenas-BH, atualmente, têm o primeiro contato formal no curriculum com a dermatologia no sétimo período, cujo módulo de cinco semanas nesta área clínica e novamente no último ano no internato de atenção integral à saúde é cursado ,com atividades práticas no ambulatório de especialidades da Universidade (Centro de Estudos e Atenção à Saúde da Comunidade Professor Edson Antônio Vellano-Ceasc). A Unifenas-BH tem um currículo voltado para a formação de um médico generalista, centrado na Aprendizagem Baseada em Problemas (ABP).

3.2 População

Alunos do último ano do curso de Medicina da Unifenas-BH.

3.3 Critérios de inclusão

Foram incluídos alunos regularmente matriculados no sexto ano do curso de Medicina no ano de 2017 que concordaram em participar da pesquisa.

3.4 Critérios de Exclusão

Foram excluídos os alunos que não compareceram ao dia do recrutamento ou não concluíram a atividade proposta

3.5 Procedimentos

Trata-se de estudo experimental controlado realizado em uma única fase. Antes do início do estudo, foi avaliada a percepção da experiência e do conhecimento prévios do aluno com lesões dermatológicas. Uma com 20 diagnósticos dermatológicos (12 diagnósticos utilizados no estudo e 8 distratores) foi entregue aos alunos e, para cada diagnóstico listado, o estudante

registrava sua auto-percepção em relação ao grau de conhecimento prévio e experiência clínica com cada uma delas (APÊNDICE A) .

Doze imagens de lesões dermatológicas foram utilizadas no estudo. Os diagnósticos eram de Psoríase, Impetigo, Dermatite de contato, Pênfigo, Urticária, Hanseníase, tinha, Carcinoma Basocelular, Ceratose Seborréica pigmentada, Nevo Melanocítico, Nevo Congênito e Melanoma (APÊNDICE B). Previamente ao estudo, foi feito um piloto com residentes de clínica médica e alunos do oitavo período do curso de medicina para avaliar o nível de complexidade diagnóstica das lesões. Uma lesão de Líquen plano foi excluída porque nenhum aluno acertou o diagnóstico e uma lesão de tinha (aqui é tinha ou tínea?) inguinal porque todos acertaram, sendo substituída por uma lesão de dermatite de contato e uma de tínea corporis, respectivamente.

O estudo foi realizado em sala de aula, com as seis turmas do internato. As turmas, com número semelhante de alunos, foram alocadas em bloco, de maneira alternada sequencial, ao grupo reflexão (GR) ou ao grupo controle (GC).

As imagens das lesões dermatológicas foram apresentadas aos alunos em um formato de bloco de notas, onde cada página continha a imagem de uma lesão. Todos os alunos receberam as imagens em uma mesma sequência, juntamente com as instruções das tarefas a serem realizadas.

Para cada imagem, os alunos de ambos os grupos eram orientados a prover, em 30 segundos, uma hipótese diagnóstica inicial e assinalar seu grau de confiança em relação àquele diagnóstico, utilizando uma escala de 1 a 10 (1 equivalia a “não tenho confiança” e 10a “eu tenho total confiança”) .A seguir, o aluno realizava uma segunda tarefa, com duração de 3 minutos e 30 segundos, que diferia dependendo da condição experimental para a qual o aluno tivesse sido alocado. Os alunos do GR foram orientados a se engajar em uma atividade reflexiva, por meio do preenchimento de um quadro (FIG. 1) no qual deveriam listar os aspectos visuais que favoreciam e aqueles que falavam contra a hipótese diagnóstica, apontar dois diagnósticos diferenciais e listar tais aspectos para cada um deles. Após a reflexão, o aluno era solicitado então a apontar a hipótese diagnóstica final e o grau de confiança neste diagnóstico (na mesma escala utilizada para avaliação da hipótese diagnóstica inicial). Para os alunos do GC, a segunda tarefa era uma atividade distratora que consistia em encontrar em

um caça-palavras nove termos médicos (não relacionados às lesões utilizadas no estudo), e também finalizava a atividade listando a hipótese diagnóstica final e o grau de confiança neste diagnóstico (APÊNDICE C). Ao final da atividade com as doze lesões, os alunos de ambos os grupos responderam a uma última pergunta referente ao número de lesões do estudo que julgava ter acertado o diagnóstico principal.

3.6 Aspectos éticos

Este estudo foi realizado em consonância com a Resolução nº. 466/2012 do Conselho Nacional de Saúde. Todos os estudantes participaram voluntariamente e o resultado individual dos testes foi mantido em sigilo. O autor e seus colaboradores não puderam identificar o nome dos estudantes e o desempenho individual de cada participante, pois os questionários foram codificados.

Foi fornecido termo de consentimento livre e esclarecido, em duas vias, a todos os alunos participantes, ficando uma com o aluno e outra com o pesquisador, e este deveria ser lido e assinado.

Os testes foram realizados em horário favorável à atividade, compartilhado pelos membros envolvidos. A participação em pesquisa também criou um ambiente favorável para estimular a pesquisa na Universidade. Os alunos receberam feedback 1, através de e-mail, com o resumo dos resultados do trabalho. As mudanças de cenários no último ano do curso dificultaram reunir os alunos para um feedback presencial do desempenho.

Os riscos decorrentes da participação foram ansiedade, redução do tempo para lazer e atividades de estudo em curso naquele dia. Alguns alunos ficaram cansados devido à atividade cognitiva realizada.

3.7 Análise dos dados

A análise descritiva das características da população elegível foi realizada a partir da distribuição de frequência das variáveis categóricas e por meio das medidas de tendência central e de dispersão das variáveis contínuas. A acurácia diagnóstica foi computada para cada lesão dermatológica. Para a avaliação dos diagnósticos foi utilizado o sistema binário (0

e 1) para pontuação das respostas dadas no diagnóstico principal. Um ponto foi dado quando o diagnóstico correto e 0 ponto quando o diagnóstico estava incorreto. A pontuação foi dada por P.R.B, especialista em dermatologia, e por G.B.C. especialista em cirurgia geral, de maneira conjunta e consensual, sem conhecimento acerca do grupo ao qual o estudante tinha sido alocado.

Para a confiança, foi analisada a pontuação da confiança dada pelo aluno ao final de cada caso. Os resultados tanto para acurácia quanto para confiança foram apresentados como percentual do total de acertos ou do total de confiança possíveis.

Foi avaliada como a acurácia de cada diagnóstico estava alinhada com a confiança para aquele caso específico, variando de 0 (melhor alinhamento possível) a 1 (pior alinhamento possível). Foi utilizado o índice “over-under” (O-U) que indicou a direção e magnitude do erro de calibragem, variando de -1 (maior nível possível de confiança subestimada) a +1 (maior nível possível de confiança superestimada) (MEYER et al., 2013). Foi feita uma análise também da calibragem global, utilizando-se o escore global de acurácia (média dos acertos no diagnóstico final de cada caso) e a percepção global da porcentagem de acertos nos casos, respondida ao final do estudo.

Para subanálise por síndrome dermatológica, foram agrupados os escores (tanto de acurácia quanto de confiança) de cada aluno para todas as cinco lesões pigmentadas (compondo o escore para síndromes melanocíticas) e aqueles para as outras sete lesões (compondo o escore para as síndromes não-melanocíticas). A dificuldade dos casos, utilizada em outra subanálise, foi definida a partir da análise do desempenho dos alunos no próprio estudo. As seis lesões nas quais os alunos tiveram melhor desempenho foram classificadas como “fáceis” e as demais “difíceis”.

Como não foi observado um efeito global da reflexão na calibragem diagnóstica, foi realizada uma subanálise *post-hoc* para melhor compreensão do fenômeno, com as lesões agrupadas por dificuldade dos casos (fáceis x difíceis), utilizando A NOVA de medidas repetidas. A dificuldade dos casos foi definida a partir do desempenho dos alunos no próprio estudo. As seis lesões nas quais os alunos tiveram melhor desempenho foram classificadas como “fáceis” e as demais como “difíceis”.

Utilizou-se o teste de χ^2 de Pearson, ou teste exato de Fisher, para análises univariadas de variáveis categóricas e o teste T de Student para as variáveis contínuas. A comparação da calibragem quanto à direção do erro foi avaliada utilizando o teste T de Student.

O coeficiente de correlação de Pearson foi utilizado para analisar a correlação entre a estimativa do total de acertos (0 a 12) com a acurácia, confiança e a calibragem (índice O-U).

4 RESULTADOS

Participaram do estudo, 61 alunos do 6º ano do curso de Medicina da Unifenas - Belo Horizonte, divididos em grupo reflexão (n=33) e grupo controle (n=28). Houve predominância de mulheres no grupo reflexão (72,7% vs 42,9%; p=0,035). Não houve diferença em relação à idade entre os grupos controle e reflexão (27,1 vs 27,7 anos, respectivamente; p=0,62). Não houve perda amostral ou aluno excluído.

Os alunos dos dois grupos de estudos (grupo controle – GC e grupo reflexão – GR) não diferiram, significativamente, entre si quanto aos graus de autopercepção do conhecimento e de experiências prévias relacionados a doenças dermatológicas, conforme mostrados na TAB. 1.

Tabela 1 - Análise comparativa entre os 2 grupos de estudo quanto à percepção do conhecimento e a experiência prévias em lesões dermatológicas.

Variável	Grupo		p
	Controle (GC)	Reflexão (GR)	Conclusão
Percepção do conhecimento	3,5 ± 0,8	3,4 ± 0,6	0,449
Percepção da experiência	2,9 ± 0,8	2,7 ± 0,7	0,277

Base de dados: 61 alunos → Grupo Controle (28) e Grupo Reflexão (33)

Nota: 1ª linha → Média ± d. pp → probabilidade de significância (p) do teste *t de Student p/ amostras independentes*

A porcentagem de acertos de cada lesão dermatológica do estudo está apresentada na TAB.2. Comparando-se o desempenho dos alunos dos diferentes grupos, foi observada diferença estatisticamente significativa quanto à porcentagem de acerto somente em relação ao diagnóstico de melanoma, sendo o desempenho superior observado para o GR quando comparado ao GC (87,9% versus 64,3%, respectivamente; p= 0,029). Independentemente do grupo de estudo, nenhum acertou o diagnóstico de *Tineacorporis*

Tabela 2 - Comparação entre os grupos controle e reflexão quanto ao acerto de cada uma das 12 lesões dermatológicas estudadas

Síndromes	Grupo		p (conclusão)
	Controle (GC)	Reflexão (GR)	
Não melanocíticos			
Psoríase	15 (53,6%)	25 (75,8%)	0,069*
Urticária	21 (75,0%)	26 (78,8%)	0,726*
Hanseníase Dimorfo	1 (3,6%)	5 (15,2%)	0,205**
Dermatite de contato	2 (7,1%)	1 (3,0%)	0,589**
TineaCorporis	0 (0,0%)	0 (0,0%)	----
Pênfigo Bolhoso	16 (57,1%)	21 (63,6%)	0,605*
Impetigo	14 (50,0%)	24 (72,7%)	0,068*
Melanocíticos			
Melanoma	18 (64,3%)	29 (87,9%)	0,029*
Nevo congênito	17 (60,7%)	27 (81,8%)	0,067*
Carcinoma basocelular	9 (32,1%)	11 (33,3%)	0,921*
Ceratose seborreica pigmentada	4 (14,3%)	6 (18,2%)	0,741**
Nevo melanocítico	12 (42,9%)	22 (66,7%)	0,062*

Base de dados: 61 alunos → Grupo Controle (28) e Grupo Reflexão (33)

Nota:* → probabilidade de significância (p) do teste *qui-quadrado de Pearson*

Nota:** → probabilidade de significância (p) do teste *Exato de Fisher*

Analisando-se a confiança do aluno em relação ao seu próprio desempenho, observou-se diferença estatisticamente significativa entre os dois grupos estudados quanto à confiança no diagnóstico no que diz respeito ao carcinoma basocelular, no qual, os alunos do GR apresentaram grau de confiança, em média, significativamente maior do que os alunos do GC (média 6,6 versus 5,2; $p=0,046$). Para os demais casos avaliados, nenhuma diferença estatisticamente significativa entre os dois grupos de alunos foi encontrada quanto ao grau de confiança no diagnóstico, demonstrado na TAB. 3.

Tabela 3 - Análise comparativa entre os 2 grupos de estudo quanto à confiança do aluno no diagnóstico em cada um dos 12 casos estudados

Síndromes	Grupo		p conclusão
	Controle (GC)	Reflexão (GR)	
Não melanocíticos			
Psoríase	6,3 ± 2,6	6,8 ± 2,3	0,452
Urticária	7,2 ± 2,8	7,3 ± 2,1	0,882
Hanseníase dimorfo	5,0 ± 3,1	6,0 ± 2,1	0,139
Dermatite de contato	5,6 ± 3,0	5,9 ± 2,3	0,662
TíneaCorporis	5,9 ± 3,0	6,0 ± 2,6	0,877
Pênfigo Bolhoso	5,9 ± 3,0	6,3 ± 2,0	0,536
Impetigo	7,7 ± 3,3	7,5 ± 2,2	0,785
Melanocíticos			
Melanoma	6,4 ± 3,2	7,1 ± 2,0	0,345
Nevo congênito	6,0 ± 2,7	6,6 ± 2,0	0,361
Carcinoma basocelular	5,2 ± 3,2	6,6 ± 1,7	0,046
Ceratose seborréica pigmentada	4,6 ± 2,9	5,6 ± 1,8	0,133
Nevo melanocítico	4,9 ± 2,9	5,6 ± 1,9	0,294

Base de dados: 61 alunos → Grupo Controle (28) e Grupo Reflexão (33)

Nota: 1ª linha → Média ± d,pp → probabilidade de significância (p) do teste *t de Student p/ amostras independentes*

Quando os diagnósticos são agrupados por síndrome (melanocítica versus não-melanocíticas), observou-se maior acurácia diagnóstica dos alunos do GR em relação ao GC para ambas as síndromes (TAB. 4). Entretanto, não se observou diferença estatisticamente significativa entre os grupos em relação à Confiança e à Calibragem. Os GRAF. de 1 a 5 permitem melhor visualização destes resultados.

Tabela 4 - Análise comparativa entre os 2 grupos de estudo quanto à acurácia, à confiança e à calibragem do teste, no global e por síndrome dermatológica

Síndromes	Grupo		p Conclusão
	Controle (GC)	Reflexão (GR)	
Global			
Acurácia (%)	38,4 ± 14,6	49,7 ± 12,1	0,002
Confiança (%)	58,9 ± 20,1	64,3 ± 13,2	0,228
Calibragem (-1 a +1)	0,20 ± 0,19	0,15 ± 0,16	0,197
Não melanocíticos			
Acurácia (%)	35,2 ± 17,2	44,2 ± 10,9	0,022
Confiança (%)	62,1 ± 21,3	65,3 ± 14,8	0,507
Calibragem (-1 a +1)	0,27 ± 0,22	0,21 ± 0,17	0,270
Melanocíticos			
Acurácia (%)	42,9 ± 25,9	57,6 ± 23,3	0,025
Confiança (%)	54,4 ± 23,1	62,9 ± 13,2	0,091
Calibragem (-1 a +1)	0,12 ± 0,28	0,05 ± 0,24	0,360

Base de dados: 61 alunos → Grupo Controle (28) e Grupo Reflexão (33)

Nota: 1ª linha → Média ± d.pp → probabilidade de significância (p) do teste *t de Student p/ amostras independentes*

Gráfico 1 - Comparação da confiança e acurácia global entre os grupos reflexão e controle

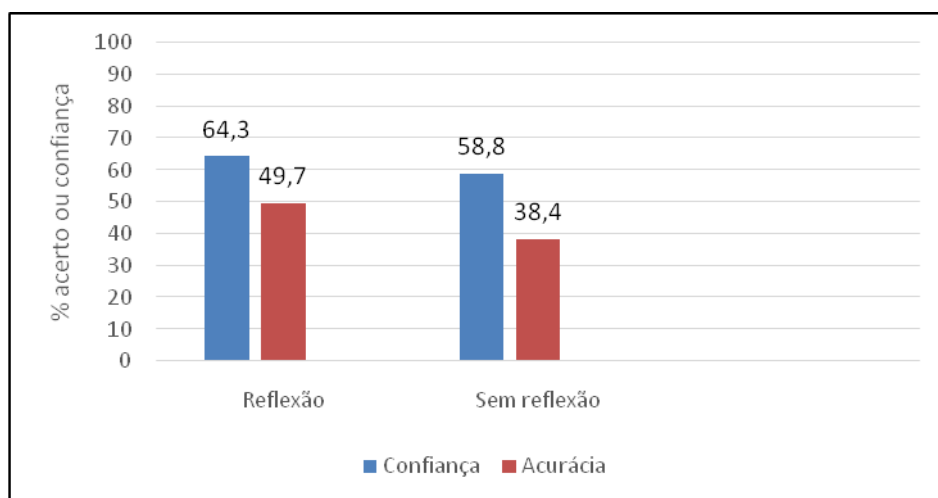


Gráfico 2 - Comparação da confiança e acurácia em lesões melanocíticas e não melanocíticas entre os grupos reflexão e controle

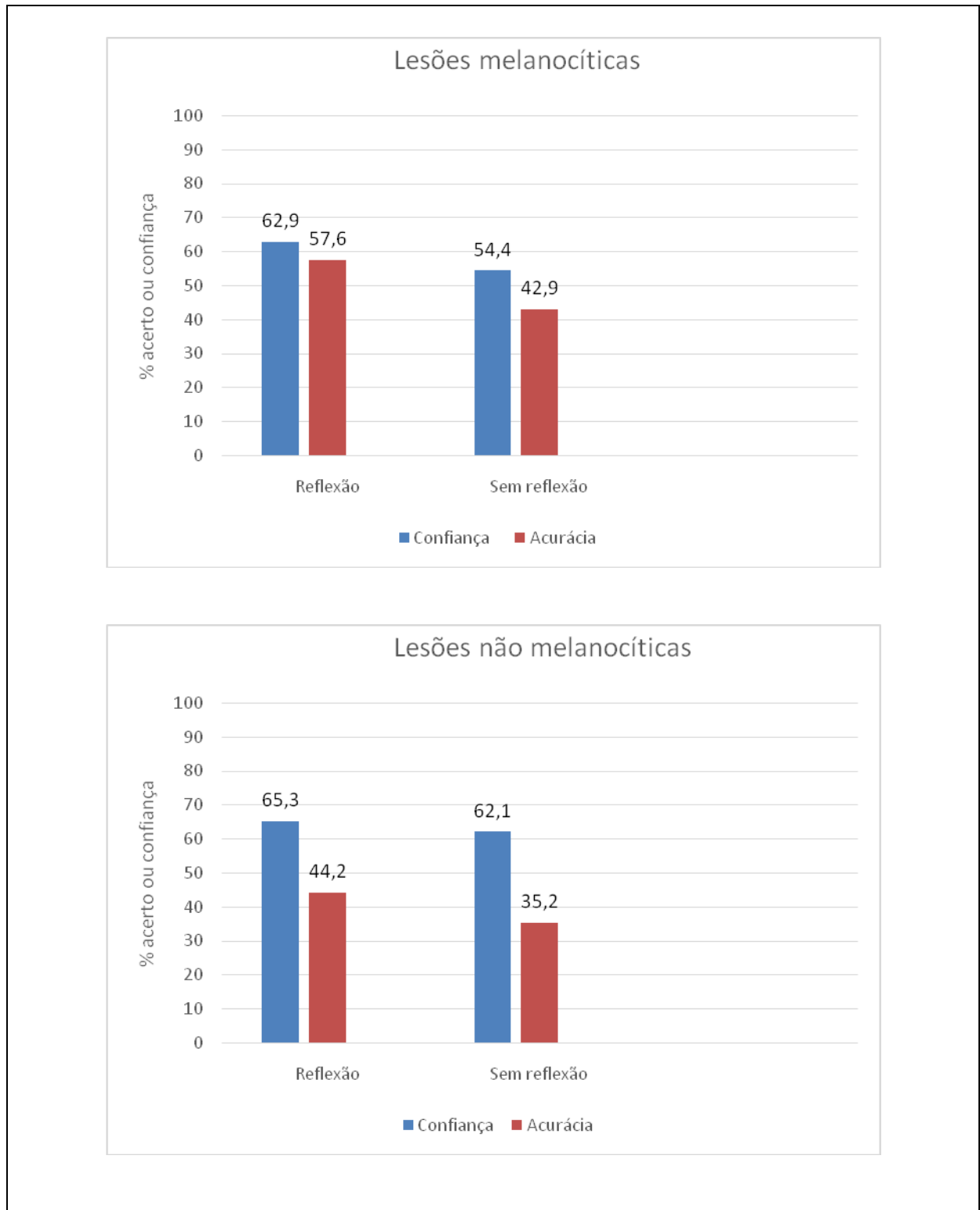
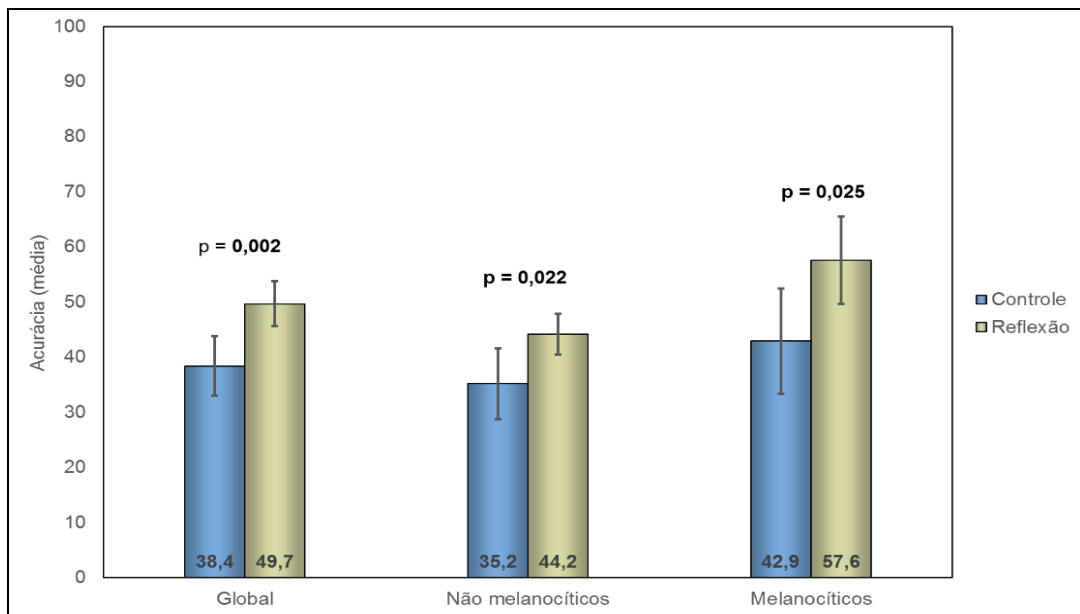


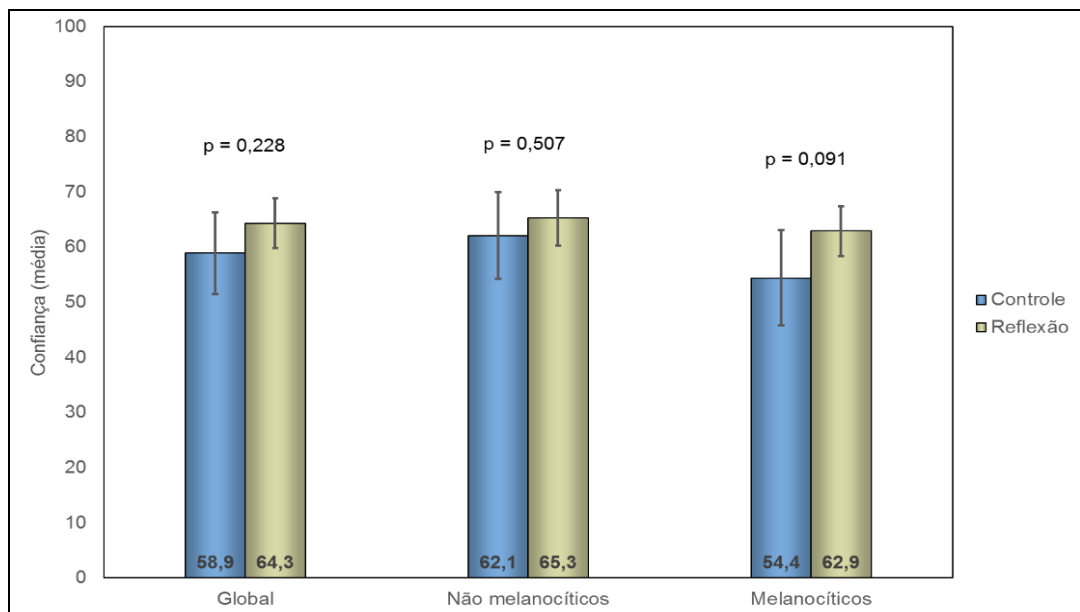
Gráfico 3 - Análise comparativa entre os 2 grupos de estudo quanto à acurácia, no global e por síndrome dermatológica



Base de dados: 61 alunos → Grupo Controle (28) e Grupo Reflexão (33)

Nota: p → probabilidade de significância (p) do teste *t de Student p/ amostras Independentes*

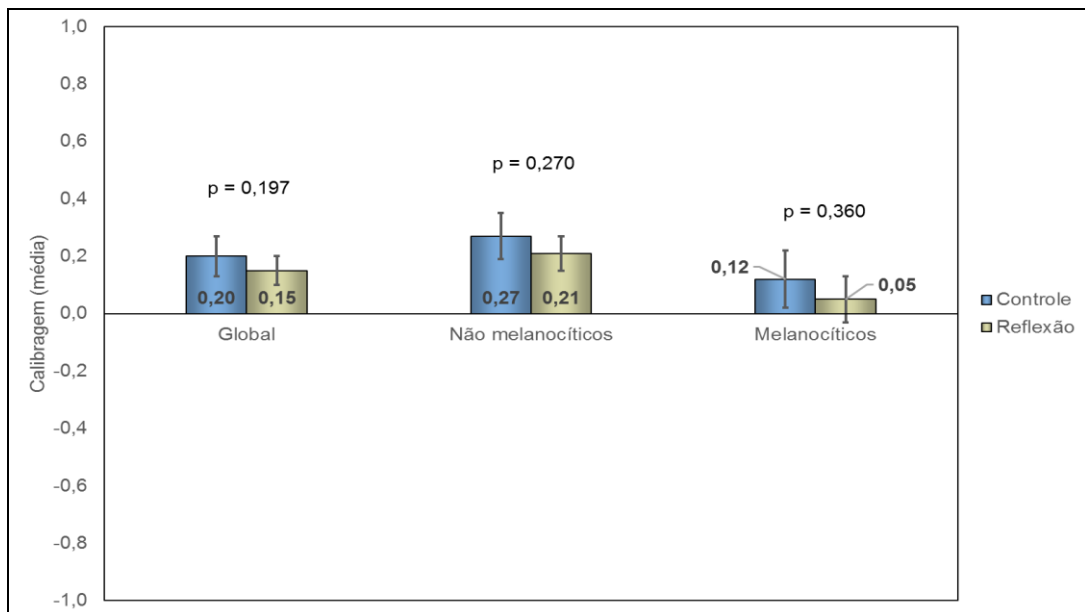
Gráfico 4 - Análise comparativa entre os 2 grupos de estudo quanto à confiança, no global e por síndrome dermatológica



Base de dados: 61 alunos → Grupo Controle (28) e Grupo Reflexão (33)

Nota: p → probabilidade de significância (p) do teste *t de Student p/ amostras independentes*

Gráfico 5 - Análise comparativa entre os 2 grupos de estudo quanto à calibragem do teste, no global e por síndrome dermatológica



Base de dados: 61 alunos → Grupo Controle (28) e Grupo Reflexão (33)

Nota: p → probabilidade de significância (p) do teste *t de Student p/ amostras independentes*

Em relação à acurácia diagnóstica, observa-se no GR um melhor desempenho para diagnóstico das doenças melanocíticas em relação às não-melanocíticas ($p=0,004$); no GC, apesar do desempenho também ter sido superior para doenças melanocíticas, esta diferença não alcançou significância estatística ($P=0,208$).

Na TAB. 5, comparando-se as diferentes síndromes dermatológicas, observa-se que no GC os alunos, em média, mostraram-se significativamente mais confiantes no diagnóstico dos casos de síndromes não-melanocíticas do que no diagnóstico de síndrome melanocíticas ($p=0,036$), o que não foi observado para o GR.

Em relação à calibragem, tanto o GC quanto o GR apresentaram melhor calibragem para síndromes melanocíticas ($p=0,017$ no GC e $p=0,001$ no GR), sendo que para síndromes não-melanocíticas a calibragem está mais distante de zero na direção positiva, ou seja, o aluno apresenta um maior excesso de confiança.

Tabela 5 - Análise comparativa entre as síndromes quanto à acurácia, à confiança e à calibragem do teste, por grupo

Grupo	Síndromes		p conclusão
	Não melanocíticos (SNM)	Melanocíticos (SM)	
Controle			
Acurácia	35,2 ± 17,2	42,9 ± 25,9	0,208
Confiança	62,1 ± 21,3	54,4 ± 23,1	0,036
Calibragem	0,27 ± 0,22	0,12 ± 0,28	0,017
Reflexão			
Acurácia	44,2 ± 10,9	57,6 ± 23,3	0,004
Confiança	65,3 ± 14,8	62,9 ± 13,2	0,195
Calibragem	0,21 ± 0,17	0,05 ± 0,24	0,001

Base de dados: 61 alunos → Grupo Controle (28) e Grupo Reflexão (33)

Nota: 1ª linha → Média ± d.p.p → probabilidade de significância (p) do teste *t de Student p/ amostras pareadas*

A TAB. 6 mostra que existe correlação direta estatisticamente significativa entre o percentual de acerto previsto pelo aluno ao final da resolução de todos os casos e a Confiança e Calibragem em ambos os grupos (GR e GC) ($p < 0,001$). Ressalta-se que não existe correlação estatisticamente significativa entre o percentual de acerto previsto pelo aluno e a acurácia do diagnóstico em nenhum dos grupos sugerindo que os alunos não conseguiram prever bem o seu desempenho no diagnóstico dos casos estudados ($p = 0,567$).

Tabela 6 - Análise de Correlação entre o percentual de acerto previsto pelo aluno e as medidas de acurácia, confiança e calibragem do teste no global e por grupo

(Continua)				
Grupo	Correlação	Acurácia	Confiança	Calibragem
Global	<i>r</i>	0,08	0,54	0,47
	<i>p</i>	0,567	< 0,001	< 0,001

Tabela 6 - Análise de Correlação entre o percentual de acerto previsto pelo aluno e as medidas de acurácia, confiança e calibragem do teste no global e por grupo

Grupo	Correlação	Acurácia	Confiança	(conclusão)
				Calibragem
Controle	<i>r</i>	0,285	0,604	0,420
	<i>p</i>	0,142	0,001	0,026
Reflexão	<i>r</i>	-0,05	0,530	0,501
	<i>p</i>	0,775	0,002	0,003

Base de dados: 60 alunos → Grupo Controle (28) e Grupo Reflexão (32)

Nota: *p* → probabilidade de significância da Análise de *Correlação de Pearson* → Coeficiente de Correlação de Pearson

Ao agruparmos os casos fáceis e difíceis (TAB. 7), observamos que a acurácia foi maior no grupo reflexão em ambos os grupos de casos, mas alcançou significância estatística apenas nos casos fáceis ($p=0,004$). Em relação à calibragem diagnóstica, observou-se que de maneira geral a calibragem é pior para os casos difíceis com os alunos apresentando excesso de confiança em relação ao seu real desempenho. Comparando-se os resultados do GR com o GC em relação aos casos fáceis, observa-se que o GR apresenta maior acurácia diagnóstica, ($p=0,004$) mas com confiança semelhante à observada do GC ($p=0,473$) e, desta forma, o sinal da calibragem se inverte (os alunos do GR têm um desempenho levemente superior ao da sua confiança) (GRAF. 6 e 7).

Gráfico 6 - Comparação da confiança e acurácia em lesões fáceis e difíceis entre os grupos reflexão e controle

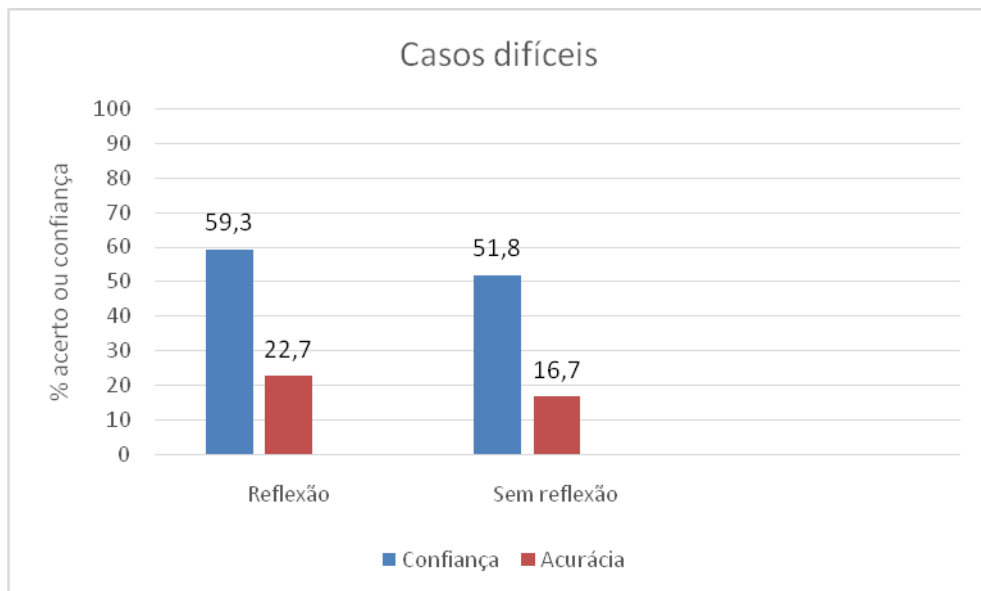
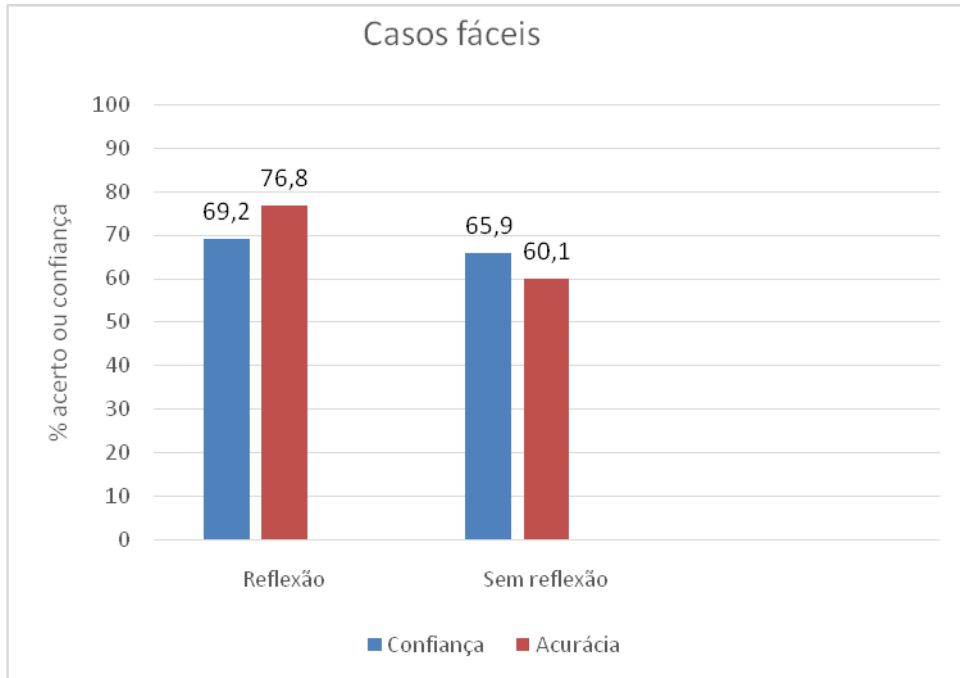


Gráfico 7 - Comparação da calibragem em lesões fáceis e difíceis entre os grupos reflexão e controle

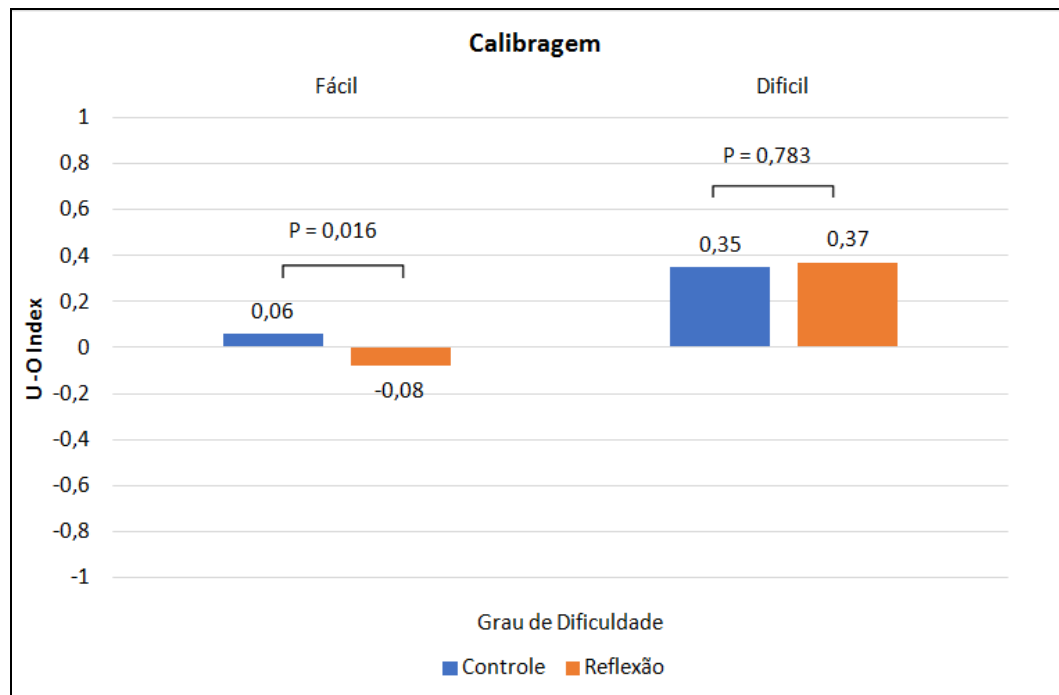


Tabela 7 - Análise comparativa entre os 2 grupos de estudo quanto à acurácia, à confiança e à calibragem do teste, no global e pela dificuldade da resolução da questão

(Continua)

Síndromes	Grupo		p Conclusão
	Controle (GC)	Reflexão (GR)	
Global			
Acurácia (%)	38,4 ± 14,6	49,7 ± 12,1	0,002
Confiança (%)	58,9 ± 20,1	64,3 ± 13,2	0,228
Calibragem (-1 a +1)	0,20 ± 0,19	0,15 ± 0,16	0,197
Resolução fácil			
Acurácia (%)	60,1 ± 23,3	76,8 ± 18,6	0,004
Confiança (%)	65,9 ± 21	69,2 ± 13,7	0,473
Calibragem (-1 a +1)	0,06 ± 0,22	-0,08 ± 0,19	0,016

Tabela 7 - Análise comparativa entre os 2 grupos de estudo quanto à acurácia, à confiança e à calibragem do teste, no global e pela dificuldade da resolução da questão

Síndromes	Grupo		(conclusão)
	Controle (GC)	Reflexão (GR)	p Conclusão
Resolução Difícil			
Acurácia (%)	16,7 ± 12,8	22,7 ± 13,7	0,080
Confiança (%)	51,8 ± 22,1	59,3 ± 14,5	0,131
Calibragem (-1 a +1)	0,35 ± 0,20	0,37 ± 0,20	0,783

Base de dados: 61 alunos → Grupo Controle (28) e Grupo Reflexão (33)

Nota: 1ª linha → Média ± d.pp → probabilidade de significância (p) do teste *t de Student p/ amostras independente*

5 DISCUSSÃO

O presente estudo mostrou que a reflexão estruturada aumentou a acurácia diagnóstica de casos dermatológicos de maneira global. Achados semelhantes com o uso da reflexão estruturada melhorando a acurácia diagnóstica foram demonstrados em estudos anteriores com casos clínicos (MAMEDE; SCHIMIDT; PENAFORTE, 2008) (MAMEDE et al., 2012), mas ainda não tinha sido demonstrada no diagnóstico de lesões dermatológicas. Apesar de Aldridge, Maxwell e Ress (2012) ponderar sobre a natureza essencialmente visual do diagnóstico dermatológico, os nossos resultados sugerem que refletir sobre estes aspectos visuais resulta em melhor desempenho diagnóstico.

Analisando individualmente cada diagnóstico, a reflexão estruturada aumentou a acurácia diagnóstica de maneira significativa apenas em relação ao melanoma. Este achado é importante pelo melanoma ser uma das lesões dermatológicas mais importantes de ser diagnosticada corretamente (GLAZER et al., 2017). Chamou a atenção o baixo desempenho de ambos os grupos no diagnóstico da dermatofitose *Tineacorporis*, doença altamente prevalente. Este baixo índice de acerto desta lesão pode ser devido à escolha inadequada da imagem desta lesão, sem as características mais importantes dessas erupções cutâneas, com o formato de anéis ou círculos com bordas levemente elevadas e pele central da lesão com aparência saudável para o diagnóstico.

Em relação à calibragem diagnóstica, observou-se que esta foi pior nos casos mais difíceis, sendo observada menor acurácia diagnóstica para estes casos, mas com os alunos mantendo confiança semelhante àquela observada para casos fáceis. Nos casos difíceis, a confiança foi apenas um pouco menor, enquanto se observou acentuada redução na acurácia. Já nos casos fáceis, os alunos ficaram melhor calibrados, mostrando que a acurácia e confiança estavam mais alinhadas.

Quando se analisa a calibragem entre casos fáceis e difíceis, os achados foram semelhantes aos observados por Meyer e Singh (2016) que também observaram que tanto a acurácia quanto a confiança estavam menos alinhadas quando os médicos lidavam com casos mais difíceis. O excesso de confiança em relação à acurácia também aumentou com a dificuldade dos casos,

contrário à hipótese inicial desses autores de que ambos, acurácia e confiança, iriam diminuir à medida que os médicos se deparassem com casos difíceis. Além disso, os autores observaram que a confiança aumentava quando mais dados do caso clínico eram disponibilizados, mas a acurácia não aumentava proporcionalmente, resultando em piora da calibragem, com excesso de confiança em relação à acurácia. Isto era também contrária à hipótese inicial dos autores de que os médicos seriam melhor calibrados quanto mais informações se desse no processo diagnóstico (MEYER et al., 2013).

O presente estudo ainda mostrou que tanto no GC quanto no GR houve melhor calibragem para lesões melanocíticas, quando comparadas às lesões não-melanocíticas, nas quais os alunos mostraram excesso de confiança. Mais uma vez a confiança se manteve relativamente semelhante nas duas síndromes, mas a piora curácia nas lesões não-melanocíticas levou à pior calibragem nesta última.

A falta de correlação entre confiança e acurácia entre estudantes de medicina e médicos residentes já foi demonstrada por outros autores. Morgane Cleave-Hogg (2002) analisaram alunos em um módulo de treinamento de habilidades anestésicas e mostraram que a confiança se correlacionou bem com a experiência clínica do aluno, mas observaram uma baixa correlação entre confiança e desempenho em um ambiente de simulação realística.

Leopold et al. (2008) avaliaram médicos durante treinamento de procedimentos cirúrgicos simples e mostraram que o treinamento resultava em ganho na performance, mas os participantes superestimavam suas habilidades tanto antes e quanto depois do treinamento oferecido.

Barnsley et al. (2004) avaliando a competência para desempenhar habilidades procedurais habituais de médicos no primeiro ano de residência demonstrou não haver relação entre o nível de confiança autodeclarado pelo residente e o desempenho formalmente avaliado por seus tutores. Achado esse semelhante ao nosso estudo em que o aluno não conseguiu prever seu desempenho, nem refletindo sobre sua prática.

Cavalcanti e Sibbald (2014) analisaram a correlação entre confiança e desempenho em residentes de clínica médica no diagnóstico de achados ao exame físico em ambiente de simulação cardiológica com manequim de alta fidelidade. Estes autores encontraram apenas

correlação moderada entre acurácia diagnóstica e confiança em condições experimentais e esta correlação piorava em situações que existia inconsistência dos dados fornecidos aos residentes. Isto os levou a concluir que a correlação entre confiança e acurácia diagnóstica é dependente do contexto e o uso da confiança como correlato de acurácia diagnóstica não é confiável.

Cavalcanti e Sibbald (2014) apontam que a identificação de condições clínicas nas quais os alunos apresentam excesso de confiança pode ser uma boa oportunidade para os alunos refletirem sobre lacunas no conhecimento, na obtenção de informações clínicas ou nas estratégias de raciocínio clínico utilizadas. Estimular os alunos a explorarem os casos onde ocorreram excesso de confiança pode levá-los a reverem quais as manifestações clínicas estão sendo levadas em consideração por eles para se chegar a determinados diagnósticos, tornando mais precisos o que os autores chamam de “links diagnósticos”.

Para tentar melhorar esta relação entre confiança e acurácia diagnóstica, exploramos o uso da reflexão estruturada que permitiria ao aluno perceber com mais clareza suas lacunas de aprendizagem e se tornar mais consciente da dificuldade diagnóstica em lesões mais difíceis. Entretanto, em nosso estudo não observamos diferença significativa entre os grupos em relação à confiança e calibragem, ou seja, a reflexão não modifica a confiança e não melhora a calibragem. Nos casos considerados mais fáceis, foi observada inversão do sinal da calibragem, com os alunos do grupo reflexão melhorando sua acurácia diagnóstica, sem ganho proporcional na confiança, ou seja, os alunos passaram a ter uma confiança abaixo do seu real desempenho. De uma maneira geral, os nossos achados sugerem uma refratariedade da avaliação de confiança à reflexão.

O presente estudo tem algumas limitações. O momento da intervenção foi único, o que pode ter limitado o efeito da reflexão sobre a calibragem. Além disso, um efeito pequeno da reflexão dificilmente seria detectado pelo pequeno tamanho amostral do estudo. As subanálises por dificuldade de caso também têm que ser avaliadas com cautela, pois a dificuldade não foi definida a priori, e sim a partir do próprio desempenho dos alunos no estudo. Por fim, observou-se predominância do sexo feminino no grupo reflexão. Existem poucos estudos analisando a questão de gênero em relação à confiança entre estudantes de Medicina. Blanchet al. (2008) mostraram que durante a realização de avaliação clínica estruturada (OSCE) os estudantes de Medicina do sexo feminino, geralmente, têm menor grau

de autoconfiança do que seus colegas do sexo masculino. Como o nosso estudo mostrou excesso de confiança em relação à acurácia tanto globalmente quanto nas subanálises, a menor confiança entre mulheres favoreceria o grupo reflexão em relação à calibragem e isto não foi observado. Desta forma consideramos que a diferente distribuição de sexo entre os grupos não interferiu nos nossos principais resultados.

Nosso estudo tem implicações importantes para a educação médica. Este foi o primeiro estudo a demonstrar que o uso da reflexão estruturada melhora a acurácia diagnóstica também para lesões dermatológicas. Em relação à calibragem diagnóstica, além de reforçar a extensa literatura que mostra a falta de correlação entre confiança e acurácia, nossos resultados mostram que o uso isolado e em um momento único da reflexão estruturada não é suficiente para melhorar esta correlação, principalmente, por seu pouco efeito na confiança do aluno em relação à sua real capacidade de dar o diagnóstico correto.

6 CONCLUSÃO

A calibragem não alterou com a atividade reflexiva globalmente, por sua falta de efeito na confiança, principalmente.

A atividade reflexiva aumentou a acurácia diagnóstica de maneira global para as lesões dermatológicas testadas.

Esperávamos que a melhoria na acurácia observada no grupo reflexão fosse acompanhada de uma melhoria na confiança, melhorando a calibragem diagnóstica. Os resultados mostraram que a calibragem piorou com o aumento da dificuldade dos casos sugerindo que a acurácia e confiança estão menos alinhadas nos casos mais difíceis e a reflexão não foi capaz de melhorar esta situação.

O presente estudo ainda mostrou que tanto no GC quanto no GR houve pior calibragem para lesões não-melanocíticas, mostrando um maior excesso de confiança. A confiança se manteve relativamente semelhante nas duas síndromes dermatológicas, mas a piora da acurácia nas lesões não-melanocíticas levou a uma pior calibragem nesta última.

Estudos adicionais utilizando outros instrumentos para melhoria do *insight* que pudessem ser agregadas à reflexão são necessários para se tentar obter um efeito favorável na calibragem diagnóstica.

REFERÊNCIAS

- ALDRIGE, R.B.; MAXWELL, S. S.; RESS, J.L. Dermatology undergraduate skin cancer training: a disconnect between recommendations, clinical exposure and competence. **BMC Medical Education**, Londres, v. 12, n. 27, p. 1-9, 2012.
- BARNESLEY, L. et al. Clinical skills in junior medical officers: a comparison of self-reported confidence and observed competence. **Medical education**, Oxford, v. 38, n. 4, p. 358-367, Apr. 2004.
- BECKETT, D.; HAGER P. J. **Life, Work and Learning**: practice in post-modernity. London: Routledge International Studies in Education, 2002.
- BERLINGER, N.; WU, A. W. Subtracting insult from injury: addressing cultural expectations in the disclosure of medical error. **Journal of medical ethics**, Londres, v. 31, p. 106-108, 2005.
- BERNER, E. S.; GRABER, M. L. Overconfidence as a cause of diagnostic error in medicine. **The American journal of medicine**, New York, v. 121, n. 5, p. S2-S23, 2008.
- BLANCH, D. C. et al. Medical student gender and issues of confidence. **Patient education and counseling**, Princeton, v. 72, n. 3, p. 374-381, Sep. 2008.
- CAVALCANTI, R. B.; SIBBALD, M. Am I right When I Am Sure? Data consistency influences the relationship between diagnostic accuracy and certainty. **Academic Medicine**, Philadelphia, v. 89, n. 1, p. 107-113, Jan. 2014.
- GLAZER, A. M. et al. Clinical Diagnosis of skin cancer: enhancing inspection and early recognition. **Dermatologic clinics**, Philadelphia, v. 35, n. 4, p. 409-416, Oct. 2017.
- CROSSKERRY, P. Timely recognition and diagnosis of illness. In: MAC-KINNON N, N. T. (Ed.). **Safe and Effective**: the eight essential elements of an Optimal Medication-use System. Ottawa, Ontario: Canadian Pharmacists Association, 2007.
- CUNHA, M. V. Uma reflexão com John Dewey sobre a educação contemporânea. **Revista USP**, São Paulo, n. 82, p. 215-217, jun./ago. 2009. Disponível em: <<https://www.revistas.usp.br/revusp/article/view/13761>>. Acesso em: 20 ago. 2017.
- GOULD, D.; REAM, E. Nurses' views of infection control: an interview study. **Journal of advanced nursing**, Oxford, v. 19, n. 6, p. 1121-1131, Jun. 1994.
- HAYS, R. B. et al. Is insight important? Measuring capacity to change performance. **Medical education**, Oxford, v. 36, n. 10, p. 965-971, 2002.
- KINSELLA, E. A. Professional knowledge and the epistemology of reflective practice. **Nursing philosophy**, Oxford, v. 11, n. 1, p. 3-14, Jan. 2010.

KORIAT, A.; LICHTENSTEIN, S.; FISCHOFF, B. Reasons for confidence. **Journal of Experimental Psychology Human Learning and Memory**, [S.l.], v. 6, n. 2, p. 107-118, 1980.

LEOPOLD, S. S. et al. Impact of Educational Intervention of confidence and competence in the performance of a simple surgical task. **The Journal of bone and joint surgery**, Boston, v. 87, n. 5, p. 1031- 1037, May 2005.

MAMEDE, S. et al. Reflection as a strategy to foster medical students' acquisition of diagnostic competence. **Medical education**, Oxford, v. 46, n. 5, p. 464-472, 2012.

MAMEDE, S. et al. How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. **Academic Medicine**, Philadelphia, v. 89, n. 1, p. 121-127, 2014.

MAMEDE, S.; SCHMIDT, H. G.; PENAFORTE, J. C. Effects of reflective practice on the accuracy of medical diagnoses. **Medical Education**, Oxford, v. 42, n. 5, p. 468-475, May 2008.

MAKARY, M. A.; MICHAEL, D. Medical error—the third leading cause of death in the US, **BMJ**, London, v. 353, p. 2139. 2016.

MEYER, A. N. D. et al. Physicians diagnostic accuracy, confidence, and resource requests. **JAMA Intern Med.**, Chicago, v. 173, n. 21, p. 1952-1958, Aug. 2013.

MEYER, A. N. D.; SINGH, H. Calibrating how doctors think and seek information to minimize errors in diagnosis. **BMJ Quality e Safety**, London, v. 0, p. 1-3, 2016. Disponível em: <<http://qualitysafety.bmj.com/content/qhc/early/2016/09/26/bmjqs-2016-006071.full.pdf>>. Acesso em: 20 Aug. 2017.

MORGAN, P. J.; CLEAVE-HOGG, D. Comparison between medical students experience, confidence and competence medical education 2002. **Medical Education**, Oxford, v. 36, n. 6, p. 534-539, Jun. 2002.

REGO, L. L.; BILLETT, M. T.; MORGAN, N. A. Consumer-based brand equity and firm risk. **Journal of Marketing**, [S.l.], v. 73, p. 47-60, Nov. 2009.

SCHON, D. A. **Educating the reflective practitioner**: toward a new design for teaching and learning in the professions. San Francisco: [s.n], 1987.

SIECK, W. R.; ARKES, H. R. The recalcitrance of overconfidence and its contribution to decision aid neglect. **Journal of Behavioral Decision Making**, [S.l.], v. 18, n. 1, p. 29-53, 2005.

APÊNDICE B – Lesões dermatológicas de acordo com o grau de dificuldade e pigmentação

Grau de dificuldade	Diagnóstico	Pigmentação
Fácil	Psoríase	Não Melanocítico
Fácil	Urticária	Não Melanocítico
Fácil	Pênfigo	Não Melanocítico
Fácil	Impetigo	Não Melanocítico
Fácil	Melanoma	Melanocítico
Fácil	Nevo congênito	Melanocítico
Difícil	Hanseníase dimorfa	Não Melanocítico
Difícil	Dermatite de contato	Não Melanocítico
Difícil	Tineas	Não Melanocítico
Difícil	Ceratose pigmentada	Melanocítico
Difícil	Carcinoma basocelular	Melanocítico
Difícil	Nevo melanocítico	Melanocítico

APÊNDICE C - Instrumento utilizado pelos alunos para realizar a reflexão estruturada

HIPÓTESE DIAGNÓSTICA INICIAL	NO ASPECTO VISUAL, O QUE FAVORECE A SUA HIPÓTESE	NO ASPECTO VISUAL, O QUE DESFAVORECE A SUA HIPÓTESE.

DOIS DIAGNÓSTICOS DIFERENCIAIS	NO ASPECTO VISUAL, O QUE FAVORECE A SUA HIPÓTESE	NO ASPECTO VISUAL, O QUE DESFAVORECE A SUA HIPÓTESE.

Agora, você terá 20 segundos para responder a mesma pergunta da página anterior.

Qual o diagnóstico final do caso?

Atenção:

Você recebeu um bloco de fotos enumeradas de 1 a 12. O número registrado na foto corresponde ao número da questão.

Questão 1

Figura 1 do seu bloco.



A) Qual o diagnóstico mais provável para este caso?

B) Assinale sua confiança em seu diagnóstico?

1	2	3	4	5	6	7	8	9	10
----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------

APÊNDICE D – Termo de consentimento livre esclarecido (TCLE)Termo de Consentimento Livre e Esclarecido(TCLE)

Você está sendo convidado para participar, como voluntário, de uma pesquisa proposta pela UNIVERSIDADE JOSÉ DO ROSÁRIO VELLANO que está descrita em detalhes abaixo. Para decidir se você deve concordar ou não em participar desta pesquisa, leia atentamente todos os itens a seguir que irão informá-lo e esclarecê-lo de todos os procedimentos, riscos e benefícios pelos quais você passará, segundo as exigências da Resolução 196/96 do Conselho Nacional de Saúde do Ministério da Saúde.

1. Identificação do(a) voluntário(a) da pesquisa: Nome: GALILEU BONIFACIO DA COSTA FILHO

2. Dados da pesquisa:

Título do Projeto: Impacto da reflexão estruturada na calibragem diagnóstica de alunos sextanistas do curso de Medicina da UNIFENAS - BH

a. Universidade José do Rosário Vellano, Faculdade de Medicina, Campus Itapoã, Belo Horizonte.

b. Projeto:

(X) Unicêntrico () Multicêntrico

c. Instituição:

d. Universidade Jose Do Rosário Vellano Campus Itapoã, Belo Horizonte.

e. Coparticipante: não se aplica

f. Patrocinador: não se aplica

g. Professor Orientador: Alexandre Moura Sampaio.

h. Pesquisador Responsável:

(X) Estudante de Pós-graduação (...) Professor Orientador

3. Objetivo da pesquisa:

Avaliar o impacto da reflexão estruturada na calibragem diagnóstica de alunos do sexto ano do curso de medicina na resolução de casos clínicos dermatológicos.

4. Justificativa da pesquisa:

Erros médicos são importantes causas de morbimortalidade no mundo e estão, frequentemente, associados à falta de acurácia diagnóstica. Infelizmente, muitos médicos que apresentam desempenhos insuficientes não têm consciência disso e, desta forma, não buscam aprimorar sua prática. Uma adequada calibragem dos alunos sextanistas, futuros médicos, ou seja, uma boa correlação entre a confiança em desempenhar uma tarefa e o desempenho real nesta mesma tarefa é fundamental para uma boa prática e para o adequado desenvolvimento profissional.

5. Descrição detalhada e explicação dos procedimentos realizados:

Para investigar a calibragem diagnóstica de alunos sextanistas, estamos conduzindo este estudo, no qual você deverá resolver 12 casos clínicos dermatológicos, registrando seus diagnósticos principais e diagnósticos diferenciais, além de avaliarmos quão confiantes vocês estão em relação a esses diagnósticos. Antes de iniciar a resolução dos casos, você será apresentado a uma lista contendo 20 diagnósticos, para avaliação da confiança pré-teste. Os alunos serão alocados em dois grupos, sendo que cada grupo receberá uma orientação distinta em relação à forma de resolução dos casos. Esta tarefa será executada em um único encontro com duração de aproximadamente 60 minutos. Os casos clínicos foram selecionados por um *expert*, todos baseados em pacientes reais, com um diagnóstico confirmado, e acompanhados de pranchas com fotos das lesões, juntamente com as instruções sobre como você deve proceder para resolver cada caso. Após será dado um *feedback* aos participantes e sorteio de jaleco. Embora seja muito importante para a nossa pesquisa a sua participação, você tem liberdade de escolher participar ou não, sem que isso lhe traga qualquer prejuízo. Estes dados serão utilizados, apenas, para a pesquisa proposta. Não é necessária a sua identificação e seus resultados não serão divulgados individualmente, sendo do conhecimento apenas dos pesquisadores.

6. Descrição dos desconfortos e riscos da pesquisa:

(X) Risco Mínimo () Risco Baixo () Risco Médio () Risco Alto

Os riscos decorrentes da participação são ansiedade, redução do tempo para lazer e atividades de estudo em curso naquele dia. Alguns alunos poderão ficar cansados devido à atividade cognitiva realizada.

7. Descrição dos benefícios da pesquisa:

Os alunos receberão *feedback* sobre os casos ao final da aplicação dos questionários, mesmo que em dia diferente ao da aplicação devido ao sigilo dos casos entre os grupos e serão sorteados jalecos entre os participantes.

8. Despesas, compensações e indenizações:

- a. Você não terá despesa pessoal nessa pesquisa
- b. Você não terá compensação financeira relacionada à sua participação nessa pesquisa.

9. Direito de confidencialidade:

- a. Você tem assegurado que todas as suas informações pessoais obtidas durante a pesquisa serão consideradas estritamente confidenciais e os registros estarão disponíveis apenas para os pesquisadores envolvidos no estudo.
- b. Os resultados obtidos nessa pesquisa poderão ser publicados com fins científicos, mas sua identidade será mantida em sigilo.
- c. Imagens ou fotografias que possam ser realizadas se forem publicadas, não permitirão sua identificação.

10. Acesso aos resultados da pesquisa:

Você tem direito de acesso atualizado aos resultados da pesquisa, ainda que os mesmos possam afetar sua vontade em continuar participando da mesma.

11. Liberdade de retirada do consentimento:

Você tem direito de retirar seu consentimento, a qualquer momento, deixando de participar da pesquisa, sem qualquer prejuízo à continuidade de seu cuidado e tratamento na instituição.

12. Acesso aos profissionais responsáveis pela pesquisa:

Você tem garantido o acesso, em qualquer etapa da pesquisa, aos profissionais responsáveis pela mesma, para esclarecimento de eventuais dúvidas acerca de procedimentos, riscos, benefícios, entre outros, através dos contatos abaixo:

APÊNDICE E – Caso modelo e questão 1 (um) do grupo reflexão

Instruções para a primeira página (modelo)

Você vai visualizar a foto, escrever seu diagnóstico mais provável e assinalar seu grau de confiança, baseado realmente na sua primeira impressão para o caso. O tempo é limitado nesta fase (30 segundos).

Exemplo:*Figura do caso modelo*



Figura do caso modelo

A) Qual o diagnóstico mais provável para este caso?

R: Verruga vulgar

B) Assinale o grau de confiança no seu diagnóstico. (1 para **menor** confiança e 10 para **maior**).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	--------------	---	----

Instruções para segunda página

Na segunda página de cada caso, você terá 3 minutos e 30 segundos para realizar uma reflexão estruturada.

Na primeira coluna do quadro abaixo liste a hipótese diagnóstica mencionada na página anterior e duas hipóteses alternativas. A seguir, para cada uma das hipóteses aponte os aspectos visuais que favorecem ou não seu diagnóstico.

HIPÓTESE DIAGNÓSTICA INICIAL	NO ASPECTO VISUAL, O QUE FAVORECE A SUA HIPÓTESE	NO ASPECTO VISUAL, O QUE DESFAVORECE A SUA HIPÓTESE.
Verruga vulgar	Lesões múltiplas elevadas, circulares, descamativas com ponto negro em superfície.	Número excessivo de lesões

DOIS DIAGNÓSTICOS DIFERENCIAIS	NO ASPECTO VISUAL, O QUE FAVORECE A SUA HIPÓTESE	NO ASPECTO VISUAL, O QUE DESFAVORECE A SUA HIPÓTESE.
Ceratose actínica	Lesões elevadas e descamativas.	Múltiplas lesões no mesmo local.
Corno cutâneo	Lesão hiperkeratótica	Múltiplas lesões no mesmo local

Agora você terá 20 segundos para responder a mesma pergunta da página anterior.

A) Qual o diagnóstico final do caso?

R: Verruga vulgar

B) Assinale sua confiança em seu diagnóstico?

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---------------

Atenção:

Você recebeu um bloco de fotos enumeradas de 1 a 12. O número registrado na foto corresponde ao número da questão.

Questão 1

Figura 1 do seu bloco.



A) Qual o diagnóstico mais provável para este caso?

B) Assinale sua confiança em seu diagnóstico?

1	2	3	4	5	6	7	8	9	10
----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------

APÊNDICE F – Caso modelo e questão 1 (um) grupo controle

Instruções para a primeira página

Você vai visualizar a foto, escrever seu diagnóstico mais provável e assinalar seu grau de confiança, baseado realmente na sua primeira impressão para o caso. O tempo é limitado nesta fase (30 segundos).

Exemplo: *Figura do caso modelo*



Figura do caso modelo

A) Qual o diagnóstico mais provável para este caso?

R: *Verruga vulgar*

B) Assinale o grau de confiança no seu diagnóstico. (1 para **menor** confiança e 10 para **maior**).

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---------------

Instruções para a segunda página (modelo)

Na segunda página de cada caso você terá 3 minutos e 30 segundos e vai encontrar um anagrama (caça-palavras) que contém termos relevantes do ponto de vista clínico. Marque 9 (nove) palavras no anagrama.

As palavras podem estar em qualquer posição no anagrama (diagonal, horizontal e vertical). Sua tarefa é encontrar essas palavras, circulando-as, no restante de tempo que você tiver. Só as palavras relevantes do ponto de vista médico interessam. Essa pode parecer uma tarefa irrelevante para você, mas não é assim, porque há indicações de estudos em outros campos de que tarefas como estas podem afetar a resolução de problemas. Por isso, por favor, faça um esforço para localizar todos os termos médicos no anagrama.

g	o	u	r	t	e	d	e	r	t	c	a	s	e	d	r	e	t	c	e
l	i	j	e	r	e	c	a	s	a	u	t	r	v	i	i	n	o	r	t
g	h	f	s	v	a	n	a	m	n	e	s	e	l	s	t	r	i	t	x
p	n	e	m	p	y	e	b	m	b	q	w	o	r	t	a	y	m	i	o
e	e	o	p	a	s	s	d	s	p	a	l	p	a	ç	ã	o	a	x	e
r	r	c	o	o	i	g	o	o	m	e	h	m	ã	n	b	e	r	i	w
c	e	e	r	t	o	g	m	e	n	o	e	o	c	i	i	a	p	a	l
u	l	l	a	e	k	h	e	n	c	o	n	s	t	i	p	a	ç	ã	o
s	l	u	e	m	p	y	e	m	g	o	d	i	a	r	r	e	i	a	t
s	i	l	l	f	x	i	a	r	e	s	t	e	s	i	a	p	n	b	s
ã	r	i	k	h	f	e	t	t	c	i	a	n	o	s	e	u	c	c	u
o	b	t	m	i	l	f	s	a	a	y	c	t	f	a	c	w	u	r	d
a	i	e	q	s	w	d	o	r	ç	w	o	s	i	b	i	l	o	n	o
m	f	o	e	g	o	i	w	a	a	ã	r	i	g	e	l	o	a	r	r
t	m	m	n	e	b	s	r	v	d	e	o	a	i	s	i	a	t	e	e
n	o	a	j	z	d	p	r	a	n	d	d	i	g	i	t	a	i	o	s
i	e	s	p	s	i	a	r	l	i	j	k	t	f	l	r	w	e	n	e
a	k	s	s	q	w	r	o	a	u	s	c	u	l	t	a	o	d	o	d
p	a	l	p	e	r	e	a	g	r	i	v	r	e	j	s	t	i	k	o
r	r	e	t	e	h	u	a	c	n	b	h	e	i	n	m	d	k	t	x
k	e	r	a	e	s	n	p	o	p	i	m	k	m	e	r	i	i	u	p
f	e	b	r	e	s	i	a	t	i	l	i	c	ã	n	p	m	e	v	s
s	c	h	e	s	d	a	e	f	u	o	o	c	o	l	o	t	b	a	n
l	a	a	t	a	b	s	c	e	s	s	o	t	i	v	i	n	a	l	l

Agora você tem 20 segundos para responder qual o diagnóstico final do caso e seu grau de confiança

A) Qual o diagnóstico mais provável para este caso?

R: Verruga vulgar

B) Assinale o grau de confiança no seu diagnóstico dado, logo acima?

Atenção:

Você recebeu um bloco de fotos enumeradas de 1 a 12. O número registrado na foto corresponde ao número da questão.

Questão 1

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	--------------	----

Figura 1 do seu bloco.



A) Qual o diagnóstico mais provável para este caso?

B) Assinale sua confiança em seu diagnóstico?

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

APÊNDICE G – Pergunta final sobre confiança após avaliação dos 12 (doze) casos nos 2 grupos

Dos 12 (doze) casos que você avaliou, quantos acha que acertou ?

0	1	2	3	4	5	6	7	8	9	10	11	12
----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	-----------	-----------