



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE CIÊNCIAS MÉDICAS

SHEYLA RIBEIRO ROCHA

TRADUÇÃO, ADAPTAÇÃO TRANSCULTURAL E EVIDÊNCIAS DE VALIDADE DE
DUAS ESCALAS DE AVALIAÇÃO DE HABILIDADES DE COMUNICAÇÃO EM
EDUCAÇÃO MÉDICA

*BRAZILIAN-PORTUGUESE TRANSLATION, CULTURAL ADAPTATION AND
VALIDITY EVIDENCE OF TWO SCALES TO ASSESS COMMUNICATION SKILLS
IN MEDICAL EDUCATION*

CAMPINAS

2018

SHEYLA RIBEIRO ROCHA

TRADUÇÃO, ADAPTAÇÃO TRANSCULTURAL E EVIDÊNCIAS DE VALIDADE DE
DUAS ESCALAS DE AVALIAÇÃO DE HABILIDADES DE COMUNICAÇÃO EM
EDUCAÇÃO MÉDICA

*BRAZILIAN-PORTUGUESE TRANSLATIO, CULTURAL ADAPTATION AND
VALIDITY EVIDENCE OF TWO SCALES TO ASSESS COMMUNICATION SKILLS
IN MEDICAL EDUCATION*

Tese apresentada à Faculdade de Ciências Médicas da
Universidade Estadual de Campinas como parte dos
requisitos exigidos para a obtenção do título de Doutora
em Ciências na área de concentração Ensino em Saúde

*Thesis presented to the School of Medical Sciences of the
State University of Campinas as part of the requirements
required to obtain the Doctor of Science degree*

ORIENTADORA: ELIANA MARTORANO AMARAL
COORIENTADOR: CARLOS FERNANDO COLLARES

ESTE EXEMPLAR CORRESPONDE À VERSÃO
FINAL DA TESE DEFENDIDA PELA
ALUNA SHEYLA RIBEIRO ROCHA E ORIENTADA PELA
PROF^A. DRA. ELIANA MARTORANO AMARAL.

CAMPINAS

2018

Agência(s) de fomento e nº(s) de processo(s): CAPES, 2036/2010
ORCID: <https://orcid.org/0000-0003-1926-4439>

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Ciências Médicas
Maristella Soares dos Santos - CRB 8/8402

R582t Rocha, Sheyla Ribeiro, 1972-
Tradução, adaptação transcultural e evidências de validade de duas escalas de avaliação de habilidades de comunicação em educação médica / Sheyla Ribeiro Rocha. – Campinas, SP : [s.n.], 2018.

Orientador: Eliana Martorano Amaral.
Coorientador: Carlos Fernando Collares.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Ciências Médicas.

1. Comparação transcultural. 2. Psicometria. 3. Comunicação em saúde. 4. Avaliação educacional. 5. Educação médica. I. Amaral, Eliana Martorano, 1960-. II. Collares, Carlos Fernando. III. Universidade Estadual de Campinas. Faculdade de Ciências Médicas. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Brazilian-Portuguese translation, cultural adaptation and validity evidence of two scales to assess communication skills in medical education

Palavras-chave em inglês:

Cross-cultural comparison

Psychometrics

Health communication

Educational measurement

Medical education

Área de concentração: Ensino em Saúde

Titulação: Doutora em Ciências

Banca examinadora:

Eliana Martorano Amaral [Orientador]

Suely Grosseman

Joana Frões Bragança Bastos

Luiz Emersto de Almeida Troncon

Erich Vinicius de Paula

Data de defesa: 23-11-2018

Programa de Pós-Graduação: Clínica Médica

BANCA EXAMINADORA DA DEFESA DE DOUTORADO

SHEYLA RIBEIRO ROCHA

ORIENTADOR: ELIANA MARTORANO AMARAL

COORIENTADOR: CARLOS FERNANDO COLLARES

MEMBROS:

1. PROF^A. DRA. ELIANA MARTORANO AMARAL

2. PROF^A. DRA. SUELY GROSSEMAN

3. PROF^A. DRA. JOANA FRÓES BRAGANÇA BASTOS

4. PROF. DR. LUIZ ERNESTO DE ALMEIDA TRONCON

5. PROF. DR. ERICH VINICIUS DE PAULA

Programa de Pós-Graduação em Clínica Médica da Faculdade de Ciências Médicas da Universidade Estadual de Campinas.

A ata de qualificação com as respectivas assinaturas dos membros da banca examinadora encontra-se no processo de vida acadêmica do aluno.

Data: 23/11/2018

*“Posso saber pedagogia, biologia como astronomia,
posso cuidar da terra como posso navegar.
Sou gente.
Sei que ignoro e sei que sei.
Por isso tanto posso saber o que ainda não sei
como posso saber melhor o que já sei.
E saberei tão melhor e mais autenticamente
quanto mais eficazmente construa minha autonomia
em respeito à dos outros.”*

Paulo Freire, 1996

AGRADECIMENTOS

Gratidão...

A Deus, presença constante, por não me deixar esquecer quão maravilhosa é a oportunidade de aprender em qualquer fase da vida!

Aos meus amados pais, Elvécio e Graça, e aos meus irmãos, por respeitarem minhas escolhas e me apoiarem em todos os momentos.

Ao meu amado Gustavo, colaborador desse projeto desde a sua concepção, por me apoiar incondicionalmente, compartilhando comigo sonhos e vida.

À minha orientadora Profa. Eliana Amaral, por me orientar com firmeza, respeitando minha autonomia e me permitindo ir além dos meus próprios limites.

Ao meu coorientador Dr. Carlos Collares, por aceitar o desafio de me conduzir pelo desconhecido mundo da psicometria e não me deixar desistir.

Aos professores Dra. Angélica Bicudo, Dr. Paulo Velho, Dra. Giuliane Jesus Lajos e Dra. Adriana Gomes Luz, pelo apoio institucional ao projeto, colaborando para que ele se tornasse uma realidade.

À amiga Dra. Maria Silvia Vellutini, por sua qualificada colaboração em momentos importantes de execução do projeto.

À minha irmã Sandra Rocha, por suas valiosas contribuições com os intrincados caminhos da língua portuguesa.

À comunidade acadêmica do Curso de Medicina da Faculdade de Ciências Médicas da Unicamp, por apoiar este estudo em diferentes etapas, e em especial ao NAPEM, pelo suporte logístico ao projeto.

À comunidade Faimer Brasil, que direta ou indiretamente esteve envolvida neste trabalho, deixando contribuições preciosas.

Aos amigos Solange Sousa, Danilo Olzon, Vera Marinzeck e Jorge Ferreira (*in memoriam*), pela amizade e solidariedade durante esta jornada.

RESUMO

Este estudo teve como objetivo traduzir, adaptar e obter evidências de validade de dois instrumentos de avaliação da comunicação de médicos em ambiente simulado. O primeiro, o *Communication Assessment Tool* é um questionário de 14 itens desenvolvido por Makoul e colaboradores para avaliar a comunicação médico-paciente pelo paciente. Os itens são pontuados em uma escala de cinco pontos (1 = ruim a 5 = excelente) e somados para obter uma pontuação total (variando de 14 a 70). O segundo, *Analytic Global OSCE Rating Form*, desenvolvido por Hodges e McIlroy avalia quatro domínios da comunicação médica (empatia, coerência na entrevista, expressão verbal e não verbal), utilizando uma escala de cinco pontos e preenchida pelo avaliador médico. A primeira fase do estudo consistiu na tradução, adaptação transcultural e análise de conteúdo realizada por um comitê de especialistas em comunicação. A versão adaptada dos instrumentos foi aplicada em um estudo-piloto durante o Curso de Urgências Obstétricas do Programa de Residência em Tocoginecologia da Faculdade de Ciências Médicas da Unicamp, em duas estações de simulação focadas em comunicação. Participaram dessa etapa quatro pacientes simuladas, quatro avaliadores médicos e 28 residentes. A segunda fase consistiu no estudo das propriedades psicométricas da versão final da *Analytic Global OSCE Rating Form*. A escala foi aplicada durante a avaliação de habilidades clínicas a 249 estudantes do quarto ao sexto ano do Curso de Medicina da Unicamp por 40 avaliadores. Para o *Communication Assessment Tool*, treze dos 14 itens da versão em português do Brasil foram considerados “equivalentes” por pelo menos 70% dos juízes; ajuste foi necessário em um item. Quanto à validade de conteúdo, todos os itens foram considerados relevantes por pelo menos 70% dos juízes. O Índice de Validade do Conteúdo do Item variou de 0,9 a 1 e o Índice de Validade do Conteúdo da Escala foi de 0,99. A versão brasileira do *Communication Assessment Tool* mostrou boa consistência interna (alfa de Cronbach = 0,93 e 0,95). As pacientes simuladas consideraram o instrumento fácil de entender e completar. Os resultados indicam que o *Communication Assessment Tool* é uma opção válida para avaliar a comunicação dos estudantes de medicina. Para o *Analytic Global OSCE Rating Form*, três dos quatro itens foram considerados “equivalentes” na análise linguística e apenas um necessitou ajustes; na análise de conteúdo, nove de 10 especialistas classificaram todos os itens como relevantes. A análise pelo modelo multifacetado de Rasch e a análise dos componentes principais indicaram a unidimensionalidade da escala. A consistência interna foi > 0,80, indicando elevada confiabilidade e reprodutibilidade dos dados. O desempenho em comunicação teve correlação forte e significativa com a habilidade dos estudantes ($r = 0,883$), mas correlações moderadas com os escores do *checklist* para as estações de medicina interna e ginecologia ($r = 0,376$ e $0,476$, respectivamente). A escala foi utilizada com precisão por examinadores não especialistas em comunicação. O *Analytic Global OSCE Rating Form* demonstrou ser uma escolha confiável para a avaliação da comunicação clínica de estudantes de medicina no Brasil.

Palavras-chave: comparação transcultural; psicometria; avaliação educacional; comunicação em saúde; educação médica.

ABSTRACT

This study aims to provide validity evidence for using two scales to assess medical students' communication skills in a simulated environment. The first, the Communication Assessment Tool is a 14-item questionnaire developed by Makoul et al. to assess doctor–patient communication from the patient's perspective. Items are scored according to a five-point scale (1=poor to 5=excellent) and summed to obtain a total score (ranging from 14 to 70). The second, the Analytic Global OSCE Rating Form developed by Hodges and Mcllroy to assess four domains of medical communication (empathy, interview coherence, verbal and nonverbal expression) using a likert scale of 5 points completed by the examiner. The authors translated and adapted the original scale into Brazilian Portuguese using translation, transcultural adaptation, back-translation and content analysis. The adapted versions were applied in a pilot study during the Obstetric Emergency Course of the Ob-Gyn Residency Program at the Medical Sciences School of Unicamp, in two OSCE stations focused on communication. Four simulated patients, four medical examiners, and 28 residents participated in this phase. The second phase consisted of the study of the psychometric properties of the Brazilian-Portuguese version of the Analytic Global OSCE Rating Form. The scale was applied during the clinical skills assessment to 249 students from the fourth to the sixth year of the medical school of Unicamp by 40 examiners. For the Communication Assessment Tool, thirteen of the 14 items of the Brazilian-Portuguese version were considered as “equivalent” by at least 70% of the judges; adjustment was necessary in one item. Regarding content validity, all items were considered relevant by at least 70% of the judges. The Item Content Validity Index ranged from 0.9 to 1 and Scale Content Validity Index was 0.99. The Brazilian-Portuguese version of the Communication Assessment Tool showed good internal consistency (Cronbach's alpha=0.93 and 0.95). Simulated patients considered the tool easy to understand and complete. Results indicate that the Communication Assessment Tool is a valid option for assessing communication skills of medical students. For the Analytic Global OSCE Rating Form, In the content analysis, 9 of 10 experts rated all items as relevant for assessing communication. Forty faculty members then used the scale to assess 249 medical students during a two-station OSCE. The MFRM and PCA demonstrated good model fit and suggested a one-factor model for the scale. The internal consistency reliability (MFRM reliability of separation) was >0.80 for all items and examinees. The Analytic Global OSCE Rating Form had a strong and significant correlation with the estimated examinees' measure ($r=0.883$) but moderate correlations with the checklist scores for the internal medicine and Ob-Gyn stations ($r=0.376$ and 0.476 , respectively). The Brazilian version of the Analytic Global OSCE Rating Form demonstrated good validity. Furthermore, the scale was accurately completed by OSCE examiners, rather than communication specialists. Altogether, the scale may be a useful and robust assessment of communication during OSCEs in medical education.

Key words: cross-cultural comparison; psychometrics; health communication; educational measurement; medical education.

SUMÁRIO

1. INTRODUÇÃO	10
2. OBJETIVOS	22
2.1. Objetivo geral	22
2.2. Objetivos específicos	22
3. MÉTODO	23
3.1. Tipo de estudo.....	23
3.2. Cenário de estudo	23
3.3. População e amostra	24
3.4. Procedimentos para tradução, adaptação e validade de conteúdo do CAT e AG-OSCE-R	25
3.5. Procedimentos para estudo das propriedades psicométricas da versão final da AG-OSCE-R	32
3.6. Processamento e análise dos dados.....	32
3.6.1. Fase 1: tradução, adaptação transcultural, validade de conteúdo e confiabilidade do CAT e AG-OSCE-R	32
3.6.2. Fase 2: estudo das propriedades psicométricos do AG-OSCE-R.....	34
3.7. Aspectos éticos	36
4. RESULTADOS	37
4.1. Artigo 1: Brazilian-Portuguese translation and cultural adaptation of the Communication Assessment Tool: Additional validity evidence for use in OSCE	38
4.2. Artigo 2: Assessing the Communication Skills of Medical Students: A Validation Study of the Brazilian Portuguese Adaptation of the Analytic Global OSCE Rating	57
5. DISCUSSÃO	86
5.1. Communication Assessment Tool – CAT	87
5.2. Analytic Global OSCE Rating – AG-OSCE-R	90
6. CONCLUSÃO	97
7. REFERÊNCIAS	98
APÊNDICES	109
APÊNDICE A – Artigo de Revisão.....	109
APÊNDICE B: AG-OSCE-R – Orientações de preenchimento.....	129
ANEXOS	130

1. INTRODUÇÃO

Nas últimas décadas, a sociedade vem demonstrando mudanças em suas expectativas quanto ao desempenho do médico, solicitando mais cuidado e atenção desses profissionais. Sensíveis a esses anseios, os profissionais reconhecem a necessidade de transição do paradigma atual da formação médica em busca de uma abordagem mais integral no cuidado à saúde(1). Influenciado por esse movimento, o contexto da educação médica vivencia uma mudança de foco, do processo educacional para o perfil do profissional a ser formado(2, 3). Assim, a formação passa a enfatizar a aquisição de novas competências e incorpora conceitos, por exemplo, de medicina centrada na pessoa, trabalho em equipe, capacidade de liderança, habilidades de relação interpessoal e de comunicação, gestão dos serviços de saúde, prática baseada na comunidade e nos sistemas de saúde(4). Essa mudança tem influência direta sobre o desenho curricular, e o desenvolvimento dessas competências torna-se essencial na formação médica(5-7).

As habilidades de comunicação e relação interpessoal têm sido objeto de diversas pesquisas. Na assistência à saúde, a comunicação eficaz entre médicos, paciente e seus familiares é reconhecida como condição indispensável para a qualidade dos serviços prestados(8). Estudos demonstram que a melhora na qualidade da comunicação durante a entrevista clínica está diretamente relacionada a aumento na qualidade da atenção em saúde(9), maior satisfação de médicos e pacientes, melhor prognóstico físico e psicológico do paciente, melhor adesão ao tratamento(10) e redução de reclamações por má-prática(11), aumentando a segurança do paciente e reduzindo erros(12).

Apesar da crescente valorização das habilidades de comunicação na formação médica, ainda não há consenso sobre o que seria uma comunicação médico-paciente adequada(13, 14). O Consenso de Kalamazoo(15), baseando-se em cinco modelos teóricos de comunicação médico-paciente, determinou sete componentes-chaves que caracterizam uma comunicação adequada no contexto clínico: 1) construção de relacionamento; 2) abertura da consulta; 3)

coleta de informações; 4) inclusão da perspectiva do paciente; 5) compartilhamento de decisões; 6) formulação de um acordo entre médico e paciente e 7) encerramento da consulta.

As habilidades interpessoais estão diretamente relacionadas à qualidade da comunicação durante a consulta e dependem do desenvolvimento de habilidades, como a empatia, cordialidade e tonalidade da voz. As habilidades interpessoais podem ser percebidas pelo efeito da comunicação no interlocutor. O uso das habilidades de comunicação e interpessoais de forma combinada cria e sustenta a relação médico-paciente(16). Dessa forma, podemos definir as HC, de modo global, como sendo a capacidade de o profissional se relacionar de forma empática, efetiva e eficaz, capaz de evitar e resolver conflitos, utilizando comunicação verbal e não verbal no processo de comunicação com os pacientes, suas famílias e outros profissionais(17, 18). Espera-se, portanto, que os médicos que apresentem essas habilidades bem desenvolvidas sejam capazes de: 1) criar e manter uma relação empática, terapêutica e ética sólida com os pacientes; 2) utilizar a escuta ativa para obter as informações necessárias, incluindo habilidades verbais (explicativas, questionadoras, orientadoras), não verbais e de linguagem escrita e 3) trabalhar harmoniosamente com outros profissionais, tanto como membro ou líder de uma equipe de saúde(17, 18).

Sabe-se que a habilidade em se comunicar adequadamente é uma competência que não pode ser aprendida apenas por observação e tende a declinar ao longo do curso de medicina(19). Recomenda-se que seu ensino se estenda de forma coerente a todos os níveis de formação, incluindo os internatos e programas de residência(9, 19). Vários artigos, guias e consensos publicados em diversos países reafirmam a importância das HC na formação médica e na qualidade da assistência prestada(8, 9, 20-22). Assim, a qualificação da comunicação e o desenvolvimento de competência social pelos médicos assumem papel de destaque na formação profissional, e o ensino de HC passa a ser incluído nos currículos em muitos países ao redor do mundo(19, 23-27). Sua importância na prática médica é também reconhecida por organizações de classe e organismos regulatórios dos serviços de saúde(19).

Uma das propostas mais abrangentes de currículo para ensino das HC na formação médica é o Consenso Britânico. Publicado em 2008, a proposta inglesa apresenta uma linha de base para o desenvolvimento e integração do ensino, aprendizagem e avaliação da comunicação clínica. Este consenso de especialistas não é organizado por competências por defender uma flexibilidade na construção curricular, acreditando que os estudantes devem ser expostos a todas as áreas descritas em diferentes momentos da formação e de acordo com o currículo de cada escola médica. O este modelo propõe que o ensino da comunicação clínica esteja apoiado em quatro princípios fundamentais que governam não apenas a comunicação, mas todas as áreas da medicina: a prática reflexiva, o profissionalismo, a prática baseada em evidências e princípios éticos e legais. Dentre os domínios propostos para a construção curricular, o ponto central é o “respeito pelos outros”. A partir dele, os demais domínios da comunicação clínica são organizados em forma de esferas que se movem concentricamente para orientar o delineamento curricular, conforme pode ser observado na figura 1. Esta proposta permite a ligação de domínios-chave e incentiva explicitamente um processo reiterativo em que os estudantes revejam, refinem e desenvolvam habilidades existentes, ao mesmo tempo em que adquirem novas habilidades e progridem em complexidade(25).

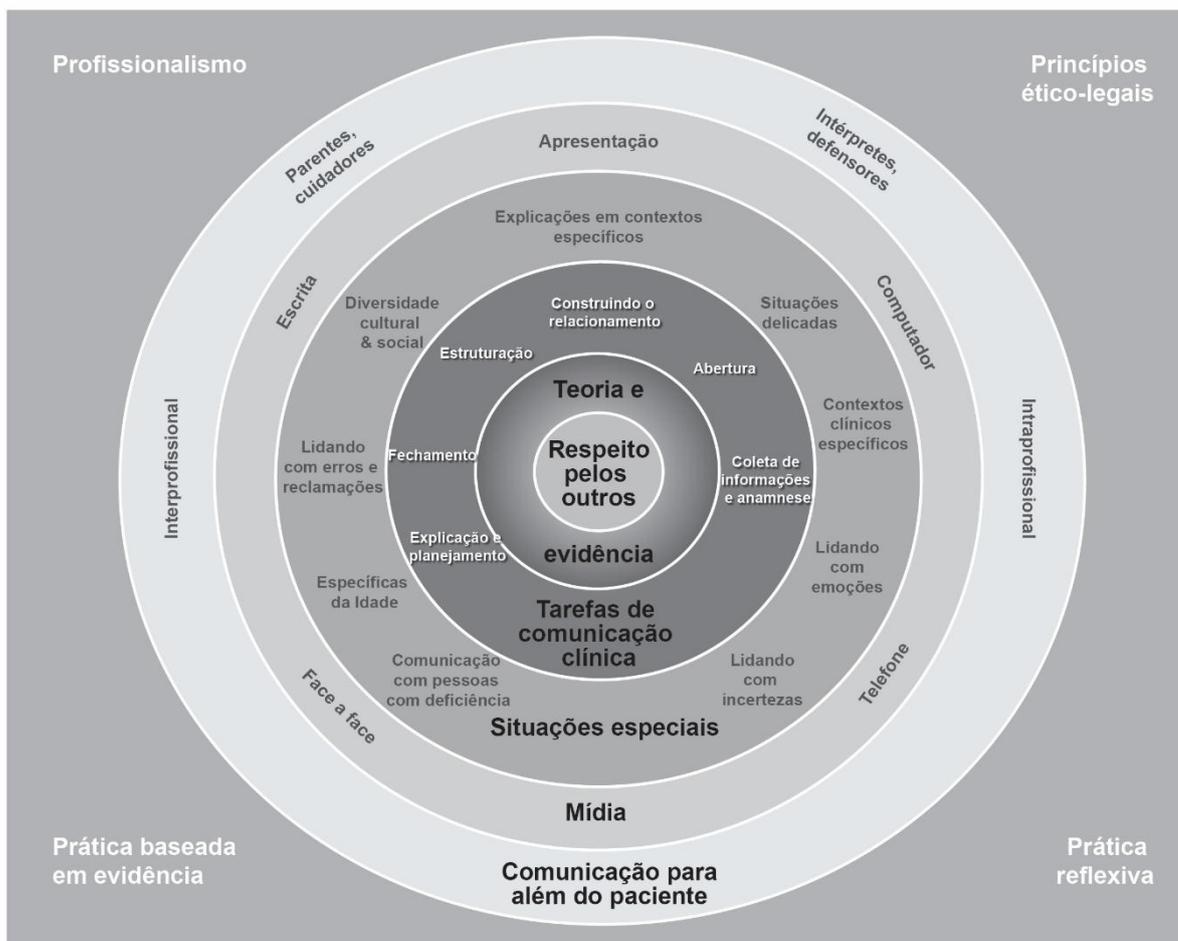


Figura 1: Representação gráfica do Consenso Britânico sobre currículo em comunicação clínica. Adaptado de von Frangstein et al, 2008.

Em 2013, foi publicado um Consenso Latino-americano, Português e Espanhol (LAPE) sobre o currículo essencial em comunicação para cursos de medicina. Este consenso foi formado por especialistas de países de língua espanhola e portuguesa, incluindo representantes brasileiros. Esse consenso pretende ser uma proposta para a concepção e o desenvolvimento de currículos em comunicação clínica para faculdades de medicina no contexto cultural da LAPE. A proposta reúne 136 objetivos educacionais agrupados em 6 domínios de competência representados na figura 2. Os objetivos educacionais foram formulados de acordo com taxonomias de Bloom e definidos como habilidades demonstráveis durante a prática médica. Um diferencial deste consenso é um grande número de objetivos educacionais relacionados à comunicação com as famílias. Segundo os autores, a maioria dos objetivos educacionais propostos são mais atitudinais e comportamentais do que cognitivos e incorporá-los ao

currículo das escolas médicas implicará em uma série de desafios para essas instituições.(28)

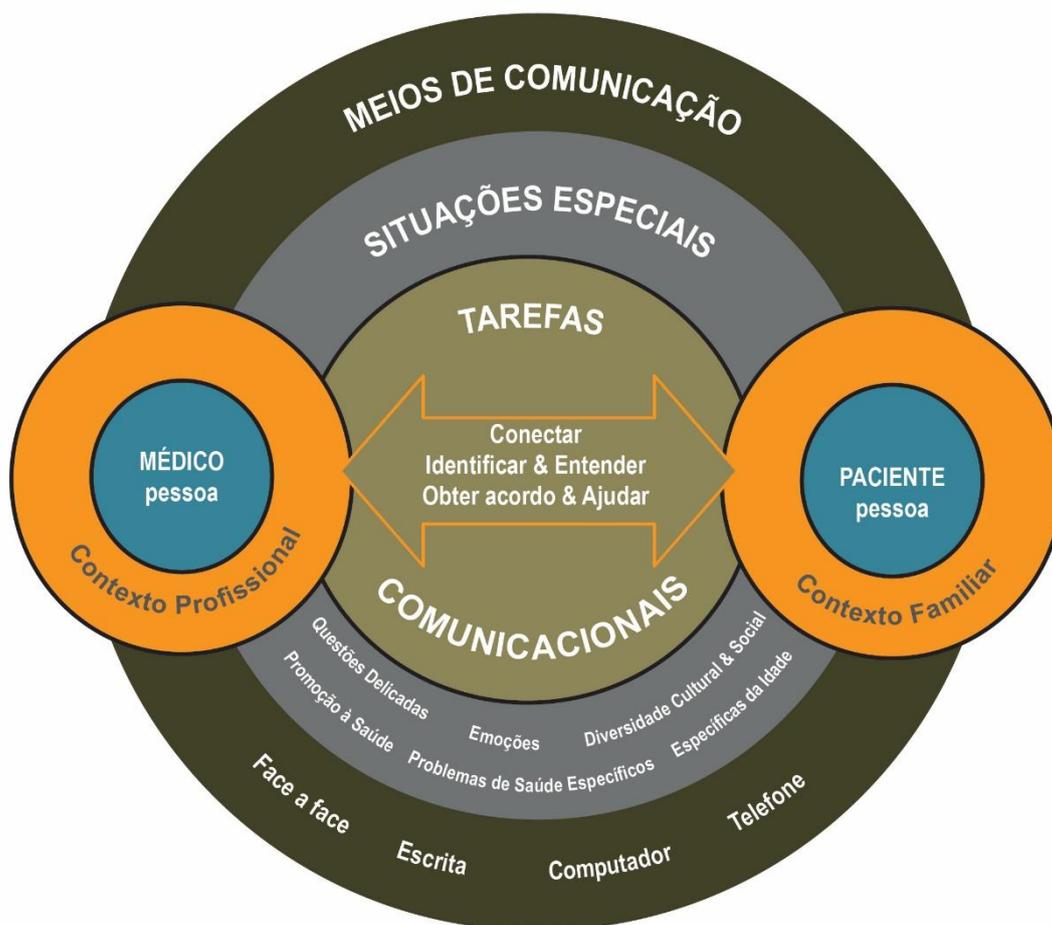


Figura 2: Representação gráfica do Consenso Latino-americano, Português e Espanhol britânico sobre currículo essencial em comunicação para cursos de medicina. Adaptado de García de Leonardo et al, 2016.

No Brasil, desde 2001, as Diretrizes Curriculares Nacionais (DCN) para os cursos de graduação na área da saúde incluem a comunicação como uma das competências a serem alcançadas pelos egressos(29). Em 2014, as novas DCN para os cursos de graduação em medicina detalharam as expectativas em relação à HC. As atuais diretrizes definem que o futuro médico deve ser capaz de se comunicar por meio de linguagem verbal e não verbal, com empatia, sensibilidade e interesse. Estabelece, ainda, que o egresso deverá estar preparado para o cuidado centrado na pessoa, desenvolvendo relação horizontal e compartilhada – condição intrinsecamente ligada à habilidade para se comunicar(30). Assim, os currículos para formação dos futuros médicos no

Brasil devem contemplar conteúdo específico para o ensino de habilidades de comunicação e relação interpessoal. No entanto, ainda não temos um consenso brasileiro que estabeleça quais habilidades devem ser ensinadas e aprendidas durante a formação médica.

Rossi e Batista (2006), analisando o processo de ensino da comunicação nos cursos de medicina brasileiros, identificaram uma diversidade de concepções e uma tendência em considerá-la uma habilidade instrumental. A principal oportunidade de aprendizado é baseada em modelos durante o internato. Os autores apontaram para a necessidade de haver uma nova abordagem do tema nos currículos médicos(31). Num outro estudo nacional, Liberali e colaboradores (2018) constataram que os currículos brasileiros abordam menos frequentemente o ensino formal das HC quando comparados com currículos de outros países, constando uma baixa adesão ao recomendado nas DCN. Os conteúdos mais trabalhados são: componentes verbais e não verbais da comunicação, escuta ativa, atendimento centrado na pessoa e técnicas de acolhimento. Identificou-se também que modalidades eficazes de ensino, como narrativas e pacientes padronizados, são subutilizadas(32).

Lampert(33) (2014) adverte que é necessário haver bons sistemas de avaliação dos estudantes, de modo que incluam a análise das habilidades de comunicação. Ele também reforça a importância na qualificação dos recursos humanos, para facilitar e orientar o desenvolvimento dessas habilidades durante a formação médica. Essa discussão já tem ocorrido na literatura internacional há alguns anos. Durante Oficina sobre o Ensino de Habilidades de Comunicação na Área da Saúde no Brasil, conduzida pela Associação Brasileira de Educação Médica (Abem) no 52º Congresso Brasileiro de Educação Médica (Cobem), em outubro de 2014, que reuniu docentes e gestores de escolas médicas de todo o país, foram determinados cinco fatores primordiais para qualificar o processo formativo em comunicação:

“...definir melhor o que é comunicação (definições e referencial teórico); estabelecer um consenso sobre HC; sistematizar o ensino das HC não só em espiral, mas também com ênfase nos laboratórios de comunicação, desde a primeira até a sexta série do curso; rever o processo de inserção curricular das HC e o papel do professor; disseminar a importância das HC nas instituições de origem, em um movimento coordenado docente/discente

para sensibilizar outros atores da escola a fim de deslanchar o ensino de HC.”(34)

A obrigatoriedade do ensino das HC constante nas DCN para os cursos de graduação em medicina, apesar de significar um avanço importante para a educação médica brasileira, ainda não é seguida de maneira sistemática pelas escolas médicas(31, 32, 35). Para que os futuros médicos brasileiros possam ser formados em consonância com as recomendações das DCN, faz-se imperativo que as atividades educacionais curriculares contemplem especificamente o ensino e a avaliação de habilidades comunicacionais essenciais para a entrevista clínica, construção de vínculo na relação médico-paciente e prática clínica de qualidade(15, 20, 23, 36). Nesse contexto, programas de desenvolvimento docente são fundamentais para que estratégias eficazes de ensino e avaliação sejam implementadas e permitam ao futuro médico a aquisição de habilidades essenciais ao ético exercício profissional(34).

Apesar das evidências quanto à importância das HC, aprendizes e educadores vivenciam problemas na integração do ensino dessa competência com os demais conhecimentos clínicos(36, 37). Um dos fatores que contribui com esse problema é o aparente conflito entre os dois modelos de entrevista clínica: o modelo construído em bases comunicacionais, que se estrutura no processo da entrevista, e o modelo tradicional de coleta da história clínica, que se estrutura no conteúdo da entrevista(38). Isso resulta em confusão para os aprendizes e interfere na aplicação das HC na prática médica cotidiana(39). No sentido de fortalecer a formação médica na aquisição dessa competência, recomenda-se que seja instituída uma matriz de avaliação das HC que permita repetidas oportunidades de avaliação e *feedback*, reforçando o uso das habilidades mais básicas da anamnese até as mais complexas, por exemplo, dar más notícias(9).

Processos adequados de avaliação são importantes tanto para qualificar a formação quanto para dar garantias à população que será beneficiada pelos serviços prestados pelos profissionais(5). Em educação médica, também está bem estabelecido o conceito de que a avaliação desempenha o importante papel de direcionar e estimular o aprendizado – os alunos se sentem motivados

a estudar os temas sobre os quais serão testados(5, 40). Recomenda-se que os processos avaliativos contemplem diversos métodos, combinando testes somativos e formativos, em diferentes momentos do processo educacional(2, 5). Considerando-se as HC uma habilidade compulsória nos currículos das escolas médicas, fazem-se necessários métodos e instrumentos avaliativos com sólido embasamento teórico e propriedades psicométricas reconhecidas, adequados ao uso em ambiente simulado e/ou ambiente da prática real(4, 41, 42).

A necessidade de instrumentos de avaliação das HC confiáveis e com evidências de validade vem sendo apontada em vários estudos ao longo das últimas décadas(9, 13, 41, 43-48). Uma revisão sistemática recente identificou instrumentos de avaliação da comunicação médico-paciente publicados em bases de dados internacionais e testados em suas propriedades psicométricas. Ao investigar a qualidade metodológica dos estudos, recomendou-se fortemente que sejam realizadas novas pesquisas que testem as propriedades psicométricas dessas escalas. Os autores sugerem que, durante os novos estudos, utilizem-se critérios de qualidade metodológica, como o *checklist* do **CO**nsensus-based **S**tandards for the Selection of Health **M**easurement **I**Nstruments (COSMIN), que avalia o rigor estatístico da pesquisa, permitindo a comparação de seus resultados(46).

O Objective Structured Clinical Examination (OSCE), traduzido livremente como Exame Clínico Objetivo Estruturado, é um método bastante recomendado para a avaliação de habilidades clínicas e HC em ambiente simulado(13, 48, 49). Entende-se que as HC podem ser avaliadas em estações de simulação construídas especificamente para esse fim ou em estações que combinem a avaliação de habilidades clínicas com a de comunicação (p.ex., coleta de história clínica e do exame físico)(50). Revisões sistemáticas sobre a validade e confiabilidade dos instrumentos de avaliação utilizados durante o Osce demonstraram uma clara ausência de consenso entre os pesquisadores com relação às suas interpretações e definições das competências em HC a serem avaliadas, bem como a fragilidade das propriedades psicométricas desses instrumentos(13, 48). Outra reconhecida ferramenta para avaliar as HC dos médicos é a inclusão da percepção do paciente na avaliação da qualidade do

atendimento prestado(49). Apesar de sua reconhecida importância, também há uma carência de estudos de validade desses instrumentos(51).

O Communication Assessment Tool (CAT) é um questionário de 15 itens desenvolvido e validado por Makoul e colaboradores, para ser preenchido pelo paciente na avaliação das HC do médico durante a consulta clínica. Esse instrumento foi inicialmente desenvolvido a partir de outros instrumentos descritos na literatura e em grupos focais com usuários dos serviços de saúde norte-americanos. Na etapa seguinte, foi conduzida uma pesquisa nacional para verificar o grau de importância que a população dava aos itens propostos. Foram mantidos os itens classificados como “muito importantes” por 70% dos entrevistados. A versão final de 15 itens também passou por um estudo de análise léxica, para verificar a sua compreensão pelos respondentes, sendo classificado como acessível a pessoas com baixo nível de leitura, o que aumenta a probabilidade de a escala ser compreendida e utilizada, seja autoadministrada ou aplicada por um entrevistador. O CAT é considerado uma ferramenta simples, direta e prática, acessível a pacientes com pouca escolaridade(52).

O CAT avalia elementos básicos da comunicação, tendo 14 itens focados no médico e um item voltado para a equipe de saúde. Cada item é pontuado numa escala de 5 pontos (1=ruim, 5=excelente) (anexo 1). A nota total obtida pelo profissional pode variar de 14 (pior desempenho) a 70 (melhor desempenho). A seleção da escala de resposta foi definida em teste-piloto, em que a escala de cinco pontos forneceu mais informações que a escala dicotomizada “sim-não”. No entanto, quando se agruparam as quatro categorias mais baixas (ruim, aceitável, bom e muito bom), comparando-as à categoria mais alta (excelente), a escala se comportou quase exatamente como a escala dicotômica “não-sim”. A escala também foi subdividida de outras maneiras, sugerindo que uma classificação de “excelente” é semelhante a “sim”, enquanto que mesmo a escolha de “muito bom” está mais próximo de “não” do que de “sim”. Diante desse resultado, o desenvolvedor manteve a escala de cinco pontos, mas recomenda que a avaliação do profissional seja obtida pela porcentagem de itens classificados como “excelentes”, por guardar maior correlação com o “sim”(52).

No estudo conduzido por Makoul et al.(52), o CAT mostrou-se unidimensional, no qual todos os itens estão avaliando um único constructo. Os resultados também indicam que o CAT de 15 itens tem elevada consistência interna e alta confiabilidade (coeficiente de Cronbach alfa=0,98), o que significa dizer que as pontuações obtidas no CAT se devem à habilidade do examinado, e não ao erro. Também se observou, por análises de funcionamento diferencial do item (FDI), que as respostas a esses itens não são influenciadas por outras variáveis, como especialidades médicas e características do paciente (p. ex., sexo, raça/etnia, nível de escolaridade, estado de saúde e visitas prévias ao médico)(52).

O CAT tem uma versão de 14 itens recomendada pelo Conselho de Acreditação de Pós-Graduação em Educação Médica dos Estados Unidos (ACGME), para avaliação de médicos-residentes(53). Nessa versão, não consta o item que avalia a equipe de saúde. Estudos demonstram que a versão de 14 itens é eficaz na avaliação de residentes tanto em cenários reais(51, 54-56) quanto em ambientes simulados(57). Recentemente, o CAT também foi adaptado para uso no contexto italiano(58). A versão de 14 itens foi escolhida para ser adaptada por ser a recomendada para avaliação de residentes em ambiente simulado.

O Analytic Global OSCE Rating Form (AG-OSCE-R) desenvolvido por Hodges e McIlroy(44) é uma escala de avaliação global desenvolvida para avaliar HC em ambiente simulado. A motivação para o desenvolvimento dessa escala global foi a evidência de que as listas de verificação (*checklist*) não eram capazes de discriminar estudantes em níveis diferentes de *expertise*(59). A construção do instrumento se baseou em quatro domínios e comportamentos observáveis: 1) resposta a sentimentos e necessidades do paciente (empatia); 2) grau de coerência na entrevista; 3) expressão verbal e 4) expressão não verbal. O AG-OSCE-R é composto por quatro itens (um para cada domínio) e um conceito global. Para cada item, as notas variam de 1 (pior desempenho) a 5 (melhor desempenho), sendo detalhados os comportamentos/desempenhos esperados para as categorias 1, 3 e 5 da escala (anexo 2). A nota total é obtida pela soma da pontuação em cada item, variando de 4 (pior desempenho) a 20 (melhor desempenho). O quinto item é uma Avaliação Global baseada na

observação de todo o desempenho do estudante durante a realização da tarefa. A Avaliação Global utiliza três níveis de desempenho observáveis, A=incompetente, B=limítrofe e C=competente (anexo 2), e para cada nível está descrito o desempenho esperado. É considerado competente o estudante que demonstra, de forma precisa, as habilidades necessárias para se realizar um atendimento médico de qualidade, integrando consistentemente todos os aspectos da comunicação médico-paciente em uma consulta clínica(44).

Os estudos de validade foram desenvolvidos na Universidade de Toronto, com 50 estudantes de medicina em um Osce de dez estações, elaborado especificamente para esse fim. As estações abrangiam problemas clínicos mais frequentes nas áreas de clínica médica, cirurgia, psiquiatria, pediatria e obstetrícia e ginecologia. Todos os cenários foram estabelecidos no contexto da atenção primária e envolviam a coleta de história clínica e/ou exame físico. Cada estudante foi avaliado em dez estações de dez minutos por examinadores médicos treinados. Cada examinador preencheu dois instrumentos de avaliação: uma lista de verificação marcada com respostas dicotômicas “fez” e “não fez” e o AG-OSCE-R. Os resultados indicaram que a confiabilidade entre as estações foi maior para o AG-OSCE-R (coeficiente alfa=0,7) do que para o *checklist* (coeficiente alfa=0,54). As pontuações do AG-OSCE-R foram significativamente maiores para estudantes dos anos mais avançados, tanto na somatória das notas quanto nas subescalas que avaliavam coerência da entrevista e comunicação verbal. Dessa forma, evidenciou-se a validade de construto da AG-OSCE-R, com capacidade da escala para discriminar estudantes em diferentes níveis de treinamento. Os resultados corroboram a superioridade da escala global em relação ao *checklist*(44).

Considerando ser a comunicação efetiva entre médico e paciente um conhecimento clínico compulsório, faz-se necessário o desenvolvimento de métodos de avaliação com sólido embasamento teórico. A literatura sugere que, na existência de um instrumento validado em outra língua, a adaptação transcultural é uma boa alternativa, por ser mais rápido do que criar um instrumento novo e permitir que os resultados sejam comparados com outros estudos(60). O processo de adaptação transcultural é muito mais do que

simplesmente traduzir um instrumento e deve seguir um protocolo bem estabelecido, para que seja obtida uma equivalência entre o instrumento original e o adaptado(60).

Em 2015, não foi encontrado nenhum instrumento desenvolvido ou adaptado para avaliar HC de médicos brasileiros. Diante disso, decidiu-se por adaptar dois instrumentos em língua inglesa desenvolvidos para avaliar HC durante consultas médicas em ambiente simulado. Também buscou-se investigar se as versões adaptadas desses dois instrumentos de avaliação de HC atendem às exigências de confiabilidade e validade praticadas no Brasil.

Os resultados deste trabalho serão importantes para o avanço na estruturação da avaliação de HC no país, com benefícios potenciais para os profissionais em formação e para a população por eles atendida. Considera-se que este estudo contribuirá com os processos de avaliação de HC em cursos da área da saúde, uma vez que não existem trabalhos com esse enfoque publicados no Brasil. Acredita-se que pesquisas conduzidas nessa área trarão uma importante contribuição para o ensino médico.

2. OBJETIVOS

2.1. Objetivo geral

Estudar as propriedades psicométricas dos instrumentos AG-OSCE-R, de Hodges e McIlroy (2003), e CAT, de Makoul e colaboradores (2007), para avaliar HC de estudantes de medicina em ambiente simulado no Brasil.

2.2. Objetivos específicos

2.2.1. Traduzir e adaptar os instrumentos para uso no Brasil, verificando a equivalência linguística dos itens adaptados por meio da análise das traduções realizadas por um comitê de juízes especialistas.

2.2.2. Obter evidências de validade baseada no conteúdo da escala, por meio da análise da pertinência dos itens adaptados por um comitê de juízes especialistas.

2.2.3. Obter evidências de validade baseada na estrutura interna dos instrumentos, por meio do coeficiente alfa de Cronbach, da análise dos componentes principais e da análise de Rasch (validade de constructo).

2.2.4. Obter evidências de validade baseada em outras variáveis por meio da comparação dos desempenhos dos estudantes na avaliação de HC utilizando o AG-OSCE-R e o *checklist* de habilidades clínicas (validade concorrente).

3. MÉTODO

3.1. Tipo de estudo

Tradução, adaptação e validação de dois instrumentos de avaliação de HC de estudantes de medicina no contexto brasileiro.

3.2. Cenário de estudo

O estudo foi realizado no Curso de Medicina da Faculdade de Ciências Médicas da Universidade Estadual de Campinas (FCM-Unicamp). A FCM-Unicamp é responsável pelos cursos de Medicina e Fonoaudiologia, somando, aproximadamente, 860 alunos, sendo 720 (840%) em Medicina. A Faculdade também disponibiliza 79 programas de residência médica, credenciados pelo Ministério da Educação e Cultura (MEC), com cerca de 550 residentes.

O curso de Medicina da FCM-Unicamp, inaugurado em 1963, tem por objetivo formar profissionais com conhecimentos, habilidades e atitudes para o exercício da medicina numa postura ética, numa visão humanística, com senso de responsabilidade social e compromisso com a cidadania. O Curso espera também formar profissionais; com capacidade de compreensão, integração e aplicação dos conhecimentos básicos na prática profissional; resolvendo com qualidade os problemas prevalentes de saúde; com competência para o primeiro atendimento das urgências e emergências; para comunicar-se e lidar com os múltiplos aspectos da relação médico-paciente; para aprendizagem contínua durante toda a vida profissional e auditoria do próprio desempenho e para atuação e eventual liderança em equipe de saúde. O curso conta ainda com o Núcleo de Avaliação e Pesquisa em Educação Médica (Napem) que, vinculado à Comissão de Ensino de Graduação em Medicina, tem entre suas finalidades desenvolver estudos de avaliação da aprendizagem, avaliação discente e projetos de pesquisas em Educação em Saúde.

Em vista disso, com apoio do Napem, este estudo foi desenvolvido durante o Exame de Avaliação de Habilidades Clínicas que é realizado anualmente como parte do processo de avaliação formativa dos alunos dos dois últimos anos do Curso de Medicina da FCM-Unicamp. Essa avaliação é conduzida

pela Coordenação do Curso e o Napem, e o apoio à pesquisa decorreu do interesse institucional em incorporar instrumentos com evidências de validade em seu programa de avaliação discente.

3.3. População e amostra

Por se tratar de estudo composto de duas fases distintas, as definições da população e do tamanho amostral levaram em consideração os objetivos de cada uma delas, a saber:

- 1) A primeira fase deste estudo compreendeu as etapas de adaptação transcultural, validade de conteúdo e pré-teste dos instrumentos. A população estudada foi diversificada e variou de acordo com a etapa do estudo:
 - a. **Etapa de adaptação transcultural e validade de conteúdo:** contou com a participação de dez especialistas que compuseram um Comitê de Juízes. Os critérios de inclusão foram: ser docente em curso de medicina e ter experiência no ensino de habilidades de comunicação. Critério de exclusão: não ter conhecimento em língua inglesa para leitura de textos científicos.
 - b. **Etapa do pré-teste:** ocorreu durante o curso Urgências e Emergências Obstétricas do Programa de Residência Médica em Tocoginecologia da FCM-Unicamp. Participaram 28 médicos residentes e oito avaliadores, sendo quatro médicos ginecologistas e quatro pacientes simuladas. Os critérios de inclusão foram: a) residente: estar regularmente matriculado no programa de residência em Tocoginecologia da FCM da Unicamp, b) avaliador: ser docente ou paciente simulado nas estações de simulação que compunham o curso Urgências e Emergências Obstétricas da FCM da Unicamp. Critério de exclusão (para ambos): não participar do curso. Todos leram e preencheram o termo de consentimento livre e esclarecido (TCLE). O cálculo amostral foi baseado no protocolo de adaptação transcultural de instrumentos de medida que recomenda que a versão adaptada seja aplicada a uma amostra de 30 a 40

sujeitos(60).

2) A segunda fase compreendeu a aplicação da versão final do instrumento para o estudo das propriedades psicométricas. Apenas o AG-OSCE-R foi aplicado a uma população de 249 discentes em duas estações de simulação. Essa amostra corresponde a todos os estudantes que participaram do Exame de Avaliação de Habilidades Clínicas em setembro de 2016. Por se tratar de estudo de validação de instrumento utilizado durante avaliação regular dos estudantes de medicina, foi solicitada e concedida a dispensa de TCLE para essa fase do estudo. O cálculo amostral da segunda fase foi baseado na recomendação específica para estudos envolvendo validação de instrumentos, que recomenda um mínimo de dez observações para cada variável estudada. Esse tamanho amostral permite a generalização dos dados com uma margem de erro 0,03 e alfa de 0,05(61). Considerando que o instrumento de AG-OSCE-R tem quatro itens, o tamanho amostral mínimo seria 40 sujeitos. No entanto, para utilizar a análise dos componentes principais (ACP), recomenda-se um mínimo de 100 sujeitos. Por questões éticas e logísticas, optou-se por aplicar o instrumento a todos os participantes. Foi considerado o fato de que uma amostra maior permitiria análises estatísticas mais robustas(61).

3.4. Procedimentos para tradução, adaptação e validade de conteúdo do CAT e AG-OSCE-R

A primeira fase da pesquisa foi desenvolvida entre agosto e novembro de 2016. Antes do início da pesquisa, foi solicitada a autorização dos desenvolvedores dos dois instrumentos (anexos 3 e 4). Todo o protocolo foi aprovado pelo Comitê de Ética em Pesquisa (CEP) da Unicamp (CAAE: 56935716.5.0000.5404) e seguiu as recomendações da Convenção de Helsinki (anexo 5). O protocolo utilizado baseou-se na proposição de Beaton et al.(60), com algumas modificações apresentadas a seguir (figura 1).

Etapa 1: Tradução Inicial

O instrumento foi traduzido para o português por dois tradutores juramentados, bilíngues e independentes, cuja língua nativa é o português brasileiro.

Etapa 2: Síntese da Tradução

Os pesquisadores realizaram análise criteriosa e síntese das duas traduções para produzir a primeira versão dos instrumentos em português. Palavras e expressões idiomáticas inadequadas foram modificadas após consenso entre os pesquisadores. Todo o processo foi documentado.

Etapa 3: Retrotradução (*back-translation*)

A versão-síntese produzida na etapa 2 foi retrotraduzida para o idioma original por dois tradutores profissionais, independentes, fluentes em português (cuja língua nativa é o inglês) e sem acesso ao instrumento original.

Etapa 4: Equivalência Linguística

Constituiu-se um comitê externo de dez docentes de cursos de medicina, bilíngues, especialistas em educação médica, com experiência no ensino de habilidades de comunicação e de diversas regiões do país (Quadro 1). Após assinarem o TCLE, todos receberam orientações sobre os objetivos da pesquisa, o instrumento original, a versão-síntese da tradução, as duas retrotraduções e os relatórios produzidos durante a etapa 2.

Os especialistas avaliaram a versão traduzida em quatro aspectos linguísticos: (1) equivalência semântica, para verificar se o significado das palavras era o mesmo em ambas as línguas; (2) equivalência idiomática, para identificar os itens de difícil tradução, sugerindo expressão equivalente mais apropriada; (3) equivalência cultural, para verificar a coerência dos termos com a realidade brasileira e (4) equivalência conceitual, para verificar se os termos ou expressões utilizados são experimentados da mesma forma pelos brasileiros. Para cada equivalência, os juízes atribuíram pontuação de acordo com a seguinte escala: (-1) “não equivale”, (0) “indeciso” e (+1) “equivale” sendo orientados a propor modificações para os itens

considerados inadequados. Cada item foi considerado adequado quando classificado como “equivalente” por pelo menos sete dos dez juízes(62) nos quatro parâmetros acima descritos. Itens considerados inadequados foram reformulados de acordo com as sugestões dos especialistas(62).

Etapa 5: Validade de Conteúdo

Essa avaliação foi realizada pelo mesmo Comitê de Especialistas e consistiu em verificar se os itens adaptados representavam o domínio-alvo (habilidades interpessoais e de comunicação) na língua portuguesa(62). Cada item foi avaliado em três critérios: (1) clareza e compreensão; (2) associação com habilidades interpessoais e de comunicação esperadas para médicos; (3) grau de relevância na escala.

Para avaliar os critérios 1 e 2, foi utilizada uma escala com resposta dicotômica “sim” e “não”. Cada item foi considerado adequado quando classificado positivamente por, pelo menos, sete dos dez juízes(62). Itens considerados inadequados foram reformulados de acordo com as sugestões do comitê(62).

Para avaliar o critério 3, foi utilizada uma escala Likert de quatro pontos: (1) = irrelevante; (2) = algo relevante; (3) = muito relevante e (4) = altamente relevante. Para cada item, o grau de concordância entre os juízes foi estimado pelo Índice de Validade de Conteúdo do Item (IVC-I), que corresponde à proporção de conceitos “3” ou “4” em relação ao total de conceitos atribuídos. Foram considerados adequados os itens com IVC-I maior ou igual a 0,78(63, 64).

Para a escala como um todo, o grau de concordância entre os juízes foi estimado pelo Índice de Validade de Conteúdo da Escala (IVC-E), que corresponde à média do IVC-I de todos os itens na escala dividida pelo número de itens. Para IVC-E maior ou igual a 0,9, a escala foi considerada adequada(63).

Etapa 6: Pré-teste da Versão dos Instrumentos CAT e AG-OSCE-R

O pré-teste dos dois instrumentos adaptados ocorreu durante o curso Urgências e Emergências Obstétricas do Programa de Residência Médica

em Ginecologia e Obstetrícia da FCM-Unicamp e consistiu na aplicação do CAT e do AG-OSCE-R a 28 médicos residentes em duas estações simuladas cujo foco era habilidades de comunicação. A versão adaptada do CAT (anexo 6) foi aplicada por quatro pacientes simuladas, atrizes profissionais com experiência em provas de habilidades clínicas e treinadas para preencher o instrumento e dar *feedback*. A versão adaptada do AG-OSCE-R foi aplicada por quatro médicos especialistas em ginecologia e obstetrícia, que receberam treinamento para preencher o instrumento. Durante os treinamentos, os itens do instrumento foram lidos e esclarecidos um a um, até a resolução de todas as dúvidas. Todos preencheram o TCLE.

Participaram 11 residentes do 1º ano (R1), nove do 2º ano (R2) e oito do 3º ano (R3); 78,5% eram do sexo feminino; 71,4% já haviam participado anteriormente de treinamentos em habilidades de comunicação e 92,8% já haviam sido submetidos a provas de habilidades clínicas em ambiente simulado.

Cada residente foi avaliado em duas estações de simulação focadas na comunicação médico-paciente em situações desafiadoras em obstetrícia (uma paciente na primeira consulta de pré-natal com VDRL positivo e uma gestante com diagnóstico inesperado de pré-eclâmpsia e alto risco de perda fetal). Cada estação teve duração de 10 minutos e foi observado o intervalo recomendado de um minuto para preenchimento do CAT(52) e do AG-OSCE-R(44). Ao final da atividade, foi realizada a devolutiva aos residentes sobre seu desempenho a partir dos instrumentos preenchidos.

Etapa 7: Equivalência Operacional para a Versão do CAT

Reichenheim e Moraes(65) recomendam avaliar a equivalência operacional quando o instrumento adaptado for utilizado de forma diferente da versão original, a fim de garantir que a eficácia seja mantida. Considerando que o CAT foi originalmente desenvolvido para uso em ambientes de prática real(52), mesmo já tendo sido utilizado em ambiente simulado(57), decidiu-se por verificar a operacionalização da versão adaptada durante a prova de

habilidades clínicas (simulação), avaliando os aspectos de utilização nesse cenário no contexto brasileiro.

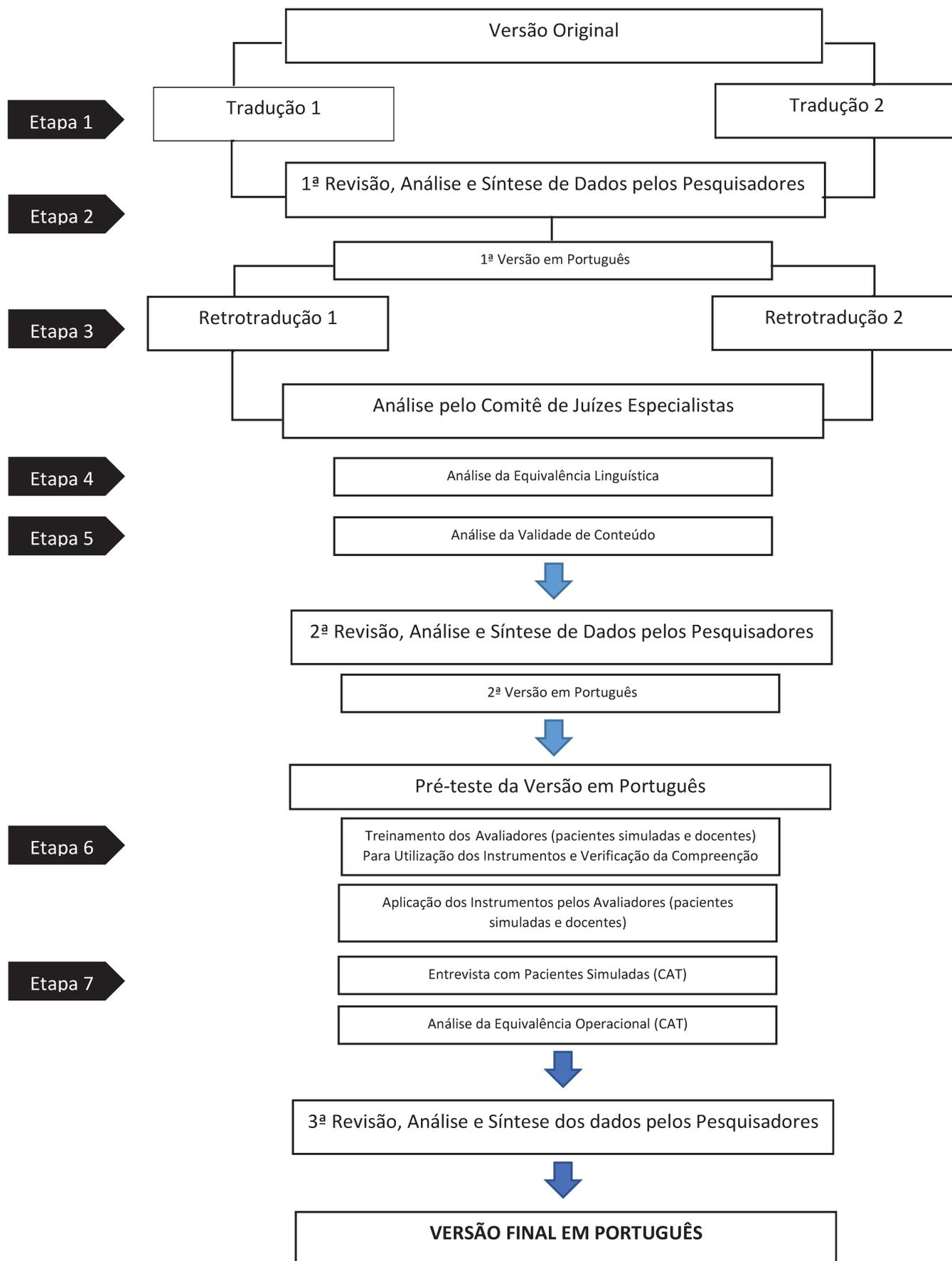
Durante o pré-teste, os pesquisadores registraram dados sobre o tempo de preenchimento, a logística e a aplicabilidade do instrumento em ambiente simulado. Após o pré-teste, as pacientes simuladas foram entrevistadas para verificar a sua percepção quanto ao formato das questões, às instruções e ao processo de resposta utilizado no preenchimento do instrumento(66). Todas essas informações foram utilizadas para avaliar a pertinência e adequação do CAT para uso em ambiente simulado(65).

Quadro 1: Características Demográficas do Comitê de Especialistas

Característica	n	%	Média (anos)
Sexo			
Feminino	6	60	
Masculino	4	40	
Titulação			
Doutor	6	60	
Mestre	4	40	
Graduação			
Medicina	9	90	
Biomedicina	1	10	
Região do Brasil			
Nordeste	1	10	
Sul	2	20	
Sudeste	7	70	
Tempo de docência em curso de medicina		25	
2-9 anos	2	20	
10-20 anos	1	10	
21-36 anos	7	70	
Tempo de ensino de habilidades de comunicação		14	
2-9 anos	4	40	
10-20 anos	4	40	
21-36 anos	2	20	
Experiência em adaptação de instrumentos			
Sim	4	40	
Não	6	60	

Fonte: A autora

Figura 1: Desenho Esquemático do Processo de Adaptação Transcultural



3.5. Procedimentos para estudo das propriedades psicométricas da versão final da AG-OSCE-R

O estudo das propriedades psicométricas do AG-OSCE-R em português (anexo 7) ocorreu durante o Exame de Avaliação de Habilidades Clínicas da FCM-Unicamp. A participação foi voluntária e contou com adesão de grande parte dos estudantes, principalmente daqueles dos dois últimos anos. O exame constou de cinco estações de OSCE desenvolvidas para avaliar habilidades clínicas nas áreas de clínica médica, pediatria, ginecologia e obstetrícia (GO), cirurgia e saúde coletiva. As estações foram elaboradas exclusivamente para essa avaliação por uma comissão de docentes do curso de medicina e têm duração de dez minutos, sendo oito minutos para a realização da tarefa e dois minutos para a devolutiva do avaliador.

Por questões logísticas, a versão final da AG-OSCE-R foi aplicada em duas estações (GO e clínica médica) por 40 avaliadores médicos a 249 discentes, sendo 59 do 4º ano, 92 do 5º ano e 98 do 6º ano. Essas estações avaliavam habilidades clínicas e de comunicação simultaneamente. Todos os avaliadores receberam orientações para preencher o instrumento no momento da avaliação (Apêndice 8).

3.6. Processamento e análise dos dados

3.6.1. Fase 1: tradução, adaptação transcultural, validade de conteúdo e confiabilidade do CAT e AG-OSCE-R

Os dados foram coletados em formulários de papel e digitados num banco de dados Excel. A estatística descritiva foi utilizada para a caracterização da amostra.

Na etapa de equivalência semântica e pertinência dos itens, o grau de concordância entre os especialistas foi calculado por porcentagem simples, utilizando a fórmula(67):

$$\% \text{ concordância} = \frac{\text{n}^\circ \text{ de participantes que concordaram}}{\text{n}^\circ \text{ total de participantes}}$$

A utilização de dez especialistas permitiu considerar como adequado o item

classificado como equivalente por pelo menos sete dos dez especialistas(62, 63). Itens que apresentaram índice de concordância abaixo desse valor foram reformulados de acordo com as sugestões do comitê(62). Os dados coletados nessa fase direcionaram as decisões críticas na produção da versão final do CAT em português.

Na etapa de validade do conteúdo, o grau de concordância entre os juízes foi estimado a partir do Índice de Validade de Conteúdo para cada Item (IVC-I) e para toda a escala (IVC-E). O IVC-I calcula a proporção de especialistas que classificou o item com nota “3” ou “4”, utilizando a fórmula(63, 67):

$$\text{IVC-I} = \frac{\text{n}^\circ \text{ de respostas "3" ou "4"}}{\text{n}^\circ \text{ total de respostas}}$$

Os itens que recebem nota “1” ou “2” foram revisados ou eliminados. É considerado adequado o item com IVC-I maior ou igual a 0,78(63).

O IVC-E calcula a média do IVC-I de todos os itens na escala, utilizando a fórmula:

$$\text{IVC-E} = \frac{\sum \text{IVC-I}}{n}$$

É considerada adequada a escala com IVC-E maior ou igual a 0,9(63). Dessa forma, pode-se avaliar a validade do conteúdo de cada item e do instrumento como um todo.

A confiabilidade foi testada por meio da análise da consistência interna, utilizando os dados da aplicação dos instrumentos após o pré-teste (etapa 6). A consistência interna é o cálculo da proporção em que os itens em um questionário estão correlacionados (homogêneos), medindo assim o mesmo constructo/conceito. São considerados adequados valores acima de 0,7(68). Todos os resultados foram comparados com os dados disponíveis na literatura.

Para o CAT, também foi verificada a validade baseada no processo de resposta das pacientes simuladas que foram entrevistadas após aplicação do instrumento no pré-teste(66).

3.6.2. Fase 2: estudo das propriedades psicométricos do AG-OSCE-R

Os dados foram coletados em formulários de papel e digitados num banco de dados Excel. Para a estatística analítica, foi utilizada a análise de Rasch, modelo baseado na teoria de resposta ao item (TRI). Essa análise utiliza um modelo matemático que estabelece que a probabilidade de resposta a um item é modelada como função da habilidade que a pessoa possui (traço latente) e de parâmetros que expressam propriedades dos itens. Quanto maior a habilidade da pessoa, maior a probabilidade dela acertar o item(69).

Para utilizar a análise de Rasch, duas premissas importantes devem ser satisfeitas: 1) os dados obtidos empiricamente devem se ajustar ao modelo, ou seja, os sujeitos devem responder conforme a expectativa do modelo proposto (a discrepância entre os dados observados e esperados devem estar em um nível razoável) e 2) os itens devem representar uma única característica (conceito de unidimensionalidade)(69). A análise de Rasch possibilita a transformação dos escores ordinais obtidos durante a avaliação em medidas lineares, as quais são expressas na mesma escala linear (*logits*), permitindo inferências corretas e comparações de dados(69). Dessa forma, esse tipo de análise estatística permite verificar se o instrumento é adequado para mensurar o traço latente, que, no caso do AG-OSCE-R, é a habilidade de comunicação de estudantes de medicina durante a consulta clínica.

Para a análise de Rasch, foi utilizado o Modelo Multi-Facetado de Rasch (MMFR) por meio do software Facets® - versão 3.71.4 (Oregon: Winsteps, 2013). O nível de significância estatística foi estabelecido em 0,05. Dentre os parâmetros disponibilizados pela análise de Rasch, estão:

- A estatística média quadrática *infit* (MNSQ-*infit*) e a média quadrática *outfit* (MNSQ-*outfit*) são utilizadas para avaliar o ajuste ou desajuste dos dados ao modelo. Esses índices de cada uma das facetadas estudadas (examinandos,

avaliadores, estações e itens) revelam se há uma maior ou menor variação entre o padrão observado de respostas da faceta e o padrão esperado de respostas. O MNSQ-*infit* é sensível à variação do escore de habilidade dos avaliados. O MNSQ-*outfit* reflete a ocorrência de dados inesperados. Valores entre 0,5 e 1,5 são considerados aceitáveis(70).

- A unidimensionalidade do instrumento é obtida por meio da análise de resíduos dos componentes principais na busca por indicações de estruturas secundárias. Para verificar a dimensionalidade, analisa-se a diferença entre o que o modelo prediz e o que é de fato, sendo essa diferença denominada resíduo. O instrumento é considerado unidimensional se a maioria da variância pode ser explicada pelo componente principal. Dessa forma, um valor adequado para a variância do componente principal seria de 60% ou mais. Além disso, a variância não explicada dos contrastes deve ter o valor menor que duas unidades de autovalores, do inglês *eigenvalue*(71).

- A confiabilidade é avaliada por meio dos índices Separation Reliability e Separation Index, que fornecem informações sobre a reprodutibilidade dos dados. Esse índice é interpretado de forma semelhante aos índices de confiabilidade mais tradicionais como o alfa de Cronbach, e valores maiores que 0,80 são adequados(72). O Separation Index acima de 2,0 é considerado apropriado(70).

Para examinar a dimensionalidade da versão brasileira do AG-OSCE-R, realizou-se inicialmente uma análise de componentes principais – ACP utilizando o software SPSS. O número de componentes retidos foi baseado na combinação dos seguintes critérios: (a) autovalor de Kaiser maior que 1.0 (critério K1); (b) teste scree de Cattell; (c) teste de esfericidade de Bartlett e (d) análise paralela(73, 74). Posteriormente foi realizada uma Análise de Componentes Principais dos Resíduos pelo modelo Rasch (PCAR), utilizando o *software* Winsteps® - versão 4.0 (Linacre, Chicago, 2017) para identificar a existência de um segundo fator que pudesse representar uma ameaça ao conceito de unidimensionalidade(75, 76).

3.7. Aspectos éticos

Todos os procedimentos dessa pesquisa foram aprovados pelo Comitê de Ética em Pesquisa da Unicamp (CAAE: 56935716.5.0000.5404) e seguiu as recomendações da Convenção de Helsinki (anexo 5).

Na primeira fase desse estudo, a população estudada foi diversificada e variou de acordo com a etapa do estudo: a) Etapa do Comitê de Juízes: composta por 10 docentes de cursos da área da saúde; b) Etapa do pré-teste do instrumento: composta por 28 residentes do Programa de Residência em Ginecologia e Obstetrícia da FCM da Unicamp, 4 médicos avaliadores e 4 pacientes simuladas que participaram da avaliação de habilidades clínicas. Nessa etapa, todos os sujeitos foram convidados a participar e assinaram o TCLE. As informações coletadas foram identificáveis, porém sigilosas, de forma que os dados pudessem ser recuperados. Os instrumentos de coleta de dados foram identificados por um código numerado e pelas iniciais do sujeito. Também lhes foi assegurado o não comprometimento de suas atividades educacionais caso não concordassem em participar da pesquisa. Ficaram assim garantidas a liberdade de participação, a integridade dos participantes da pesquisa e a preservação dos dados que pudessem identificá-los, garantindo especialmente a privacidade, o sigilo e a confidencialidade. A todos os participantes foi concedido o direito de, a qualquer momento, encerrar sua participação ante algum constrangimento ou desconforto emocional. No entanto, todos os participantes convidados permaneceram em todas as etapas do estudo.

A segunda fase envolveu os estudos de evidências de validade da versão adaptada do instrumento AG-OSCE-R. Os sujeitos foram todos os discentes do curso de medicina da FCM da Unicamp que participaram do Exame de Avaliação de Habilidades Clínicas no dia 26 de setembro de 2016. Por se tratar de pesquisa realizada em ambiente educacional sobre a validade de um instrumento utilizado em uma avaliação regular dos estudantes de medicina da FCM, foram feitas as seguintes considerações ao CEP: 1) os instrumentos estudados avaliam habilidades interpessoais e de comunicação – informar aos estudantes que estaria sendo conduzido um estudo sobre esse aspecto poderia deixá-los constrangidos, trazer riscos emocionais desnecessários e

prejudicá-los em seu desempenho no teste; 2) nenhum dos estudantes seria individualmente avaliado, pois o estudo não visa ao desempenho dos alunos, mas às propriedades psicométricas do instrumento na busca de evidências de validade para sua utilização no Brasil; 3) o presente estudo não oferece riscos adicionais aos participantes, o que permitiu ao CEP Unicamp autorizar a dispensa de TCLE para essa fase do estudo (anexo 5).

4. RESULTADOS

Os resultados serão apresentados na forma de dois artigos produzidos ao longo do doutorado:

1) Artigo 1: Brazilian-Portuguese translation and cultural adaptation of the Communication Assessment Tool: Additional validity evidence for use in OSCE (tradução, adaptação transcultural e validade de conteúdo do CAT).

2) Artigo 2: Assessing Communication Skills of Medical Students: Psychometric properties of the Brazilian-Portuguese adaptation of the Analytic Global OSCE Rating (tradução, adaptação transcultural e estudo de validação do AG-OSCE-R).

Durante a construção do referencial teórico para esta tese foi produzido um artigo de revisão narrativa que está apresentado no Apêndice A deste manuscrito (pág. 109).

4.1. Artigo 1: Brazilian-Portuguese translation and cultural adaptation of the Communication Assessment Tool: Additional validity evidence for use in OSCE

Brazilian-Portuguese translation and cultural adaptation of the Communication Assessment Tool: Additional validity evidence for use in OSCE

1º) Sheyla Ribeiro Rocha: Assistant Professor, MD, PhD, Department of Medicine, Center for Biological and Health Sciences, Federal University of São Carlos, São Carlos, Brazil. Post Graduate Program in Internal Medicine, Faculty of Medical Sciences, State University of Campinas, Campinas, Brazil.

2º) Gustavo Salata Romão: Full-Professor, MD, PhD, Medicine School, University of Ribeirão Preto, Ribeirão Preto, Brazil.

3º) Maria Sílvia Velutini Setúbal: PhD, post-doc at Obstetric Department at University of São Paulo School of Medicine, São Paulo, Brazil

4º) Giuliane Jesus Lajos: Assistant Professor, MD, PhD, Obstetrics & Gynecology Department, Medicine School, Faculty of Medical Sciences, State University of Campinas, Campinas, Brazil

5º) Adriana Gomes Luz: Assistant Professor, MD, PhD, Obstetrics & Gynecology Department, Medicine School, Faculty of Medical Sciences, State University of Campinas, Campinas, Brazil

6º) Carlos Fernando Collares: Assistant Professor, MD, PhD, Department of Educational Development and Research, School of Health Professions Education, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands.

7º) Eliana Amaral: Full-Professor, MD, PhD, Obstetrics & Gynecology Department,
Medicine School, Faculty of Medical Sciences, State University of Campinas,
Campinas, Brazil

Correspondence author: Sheyla Ribeiro Rocha

Address: Departamento de Medicina, Rodovia Washington Luis, km 235 - São Carlos -
SP - Brasil - CEP: 13565-905

e-mail: sheyla.ribeiro.rocha@gmail.com

phone: +5516 99766-4550 - +55 (16)3351-8340

Running title: Communication Assessment Tool: Validation in OSCE

Brazilian-Portuguese translation and cross-cultural adaptation of the Communication Assessment Tool: Additional validity evidence for use in OSCE

ABSTRACT

Construct: The Communication Assessment Tool is a 14-item instrument developed to assess residents' interpersonal communication skills from the patients' perspective. **Background:** Using validated instruments and the simulated patient's perception is considered good practice to assess doctor-patient communication. The Communication Assessment Tool was designed for use in real practice and has not yet applied in the assessment of Ob-Gyn residents in Objective Structured Clinical Examination stations. This study aims to provide validity evidence for using the Communication Assessment Tool to assess residents' interpersonal communication skills in a simulated environment.

Approach: The adaptation process included translations into Portuguese, synthesis of translations, back translation, linguistic equivalence, content validity, and operational equivalence. The Brazilian-Portuguese version was applied by four standardized patients, to 28 Ob-Gyn residents in two 10-minute Objective Structured Clinical Examination stations focused on communication skills.

Findings: Thirteen of the 14 items of the Brazilian-Portuguese version were considered as "equivalent" by at least 70% of the judges; adjustment was necessary in one item. Regarding content validity, all items were considered relevant by at least 70% of the judges. The Item Content Validity Index ranged from 0.9 to 1 and Scale Content Validity Index was 0.99. The Brazilian-Portuguese version of the Communication Assessment Tool showed good internal consistency in both scenarios (Cronbach's alpha=0.93 and 0.95, respectively). Simulated patients considered the Communication Assessment Tool easy to understand and complete. All reported using their emotions and feelings to assess residents' performance. **Conclusion:** This study provides additional evidence regarding the reliability and validity of the Communication Assessment Tool in a simulated environment. Results of the cross-cultural adaptation, validity evidence based on test content, internal consistency, and simulated patient interviews indicate that the Communication Assessment Tool is

a valid option for assessing interpersonal communication skills of medical residents in Objective Structured Clinical Examination stations.

Introduction

Since 2001, the Brazilian Curriculum Guidelines for Medical Education include communication skills as an educational objective¹. The effective communication among health professionals, patients and their families is recognized as a core condition for the quality of health care². Studies have shown that the improvement in doctor–patient communication is directly related to better outcomes³, greater adherence to treatment⁴, increased user satisfaction with reduction of complaints due to malpractice⁵, and improved patient safety with reduction of errors⁶. There is global consensus that the teaching of communication skills (CS) must be extended to all levels of medical training^{7; 8}. To strengthen the importance of the acquisition of CS, repeated opportunities of assessment and feedback are recommended throughout vocational training⁸. The assessment should include the perspective of patients; where they experience a doctor–patient relationship more realistically, they provide meaningful feedback to the professional⁹.

The Objective Structured Clinical Examination (OSCE) is a recommended approach to assess CS using interactions with standardized patients (SP)¹⁰. Assessment of CS is a complex task because the relationship between the student performance and the assessment criteria is often less clear than in knowledge assessment¹¹. In this context, the quality of instrument used to rate a performance is crucial for providing valid inferences from an assessment.

The Communication Assessment Tool (CAT) is a 15-item questionnaire developed by Makoul et al. (2007) to assess doctor–patient communication from the patient's perspective¹². Items are scored according to a five-point scale (1=poor to

5=excellent) and summed to obtain a total score (ranging from 14 to 70). The professional score can still be obtained by the percentage of items classified as excellent. The 14-item CAT is recommended to assess medical residents by the Accreditation Council for Graduate Medical Education¹³.

In Brazil, there is no validated instrument to assess CS from the patient's perspective in a real or simulated environment. The literature suggests that is preferable to perform a cultural adaptation of a validated instrument developed in another language than to create a new one, considering that cross-cultural adaptation is faster and the results are comparable to other studies¹⁴. This study aims to translate and adapt the CAT to the Brazilian context, and collect validity evidence to use it as a formative assessment of medical residents in a simulated environment.

Method

Our research project was approved by Institutional Human Ethical Committee of State University of Campinas (Unicamp) (CAAE 56935716.5.0000.5404) and was carried in a multi-step process (see Figure 1). We obtained the developer's permission to use the instrument.

[Insert Figure 1 about here.]

Translation and Cross-Cultural Adaptation

The research protocol was based on previous studies and guidelines for the cross-cultural adaptation process^{14; 15; 16}. Initially, two independent official translators, who were fluent in English and native speakers in Portuguese, translated two versions. The researchers compared these two translations and carried out the synthesis version of the CAT in Portuguese. Inappropriate words and expressions were modified after consensus. The process was all carefully documented. The synthesis version was back-

translated by two independent, professional, who were fluent in Portuguese and native speakers in English. None of the translators had access to the original instrument.

Throughout this process, 10 external and independent faculties were invited to form a committee. They were selected based on their experiences teaching CS in different schools and regions in Brazil (see Table 1). The CS Experts (CSEs) assessed the compatibility between the translated versions and the original instrument in four different areas: (1) semantic equivalence, to assess if the meaning of the words was the same in both languages and to check for grammatical errors in the translation; (2) idiomatic equivalence, to assess difficult-to-translate items and suggest equivalent expressions that do not change the cultural meaning of the items; (3) experiential equivalence, to assess if items are known in the new culture; and (4) conceptual equivalence, to assess whether a given term or expression, even if adequately translated, rates the same aspect in Brazilian culture¹⁴. In this step, the CSE could propose modifications and new translations that were better suited to the CAT Brazilian-Portuguese version. They assigned ratings based on three documents—the original CAT, synthesis version, and two back-translations—and produced reports of the previous steps. Full written documentation of the CSEs' contribution was analyzed by the research group to rationalize critical decisions in the adapted items.

Evidence of Content Validity

The CSEs also evaluated, reviewed, and rated all the CAT items concerning test content. Essentially, their task was to determine if the adapted items fully and sufficiently represented the targeted domain in the CAT Brazilian-Portuguese version, according the degree of item relevance on the scale. Each item was considered appropriate when rated positively by at least 70% of the judges¹⁷.

[Insert Table 1 about here.]

Instrument Application

An OSCE was organized during the Obstetric Emergencies Course of Ob-Gyn Medical Residency in Unicamp Medical School (FCM-Unicamp). We invited all residents and informed them about the study. They performed two 10-minute-long OSCE stations focused on CS (women at their first prenatal visit with a serologic diagnosis of syphilis, and pregnant women with an unexpected diagnosis of preeclampsia and high risk of fetal loss). Four simulated patients (SPs), professional actresses, with experience in clinical skills assessment were trained to use the instrument by an experienced SP trainer. At the end, the residents received feedback based on the instrument. Participants provided written informed consent, prior to data collection. After the OSCE, the SPs were interviewed to verify their perception regarding the format of the items, instructions, and the response process used to complete the CAT¹⁸. These data were used to evaluate the adequacy of the CAT for use in a simulated scenario¹⁶.

Data Analysis

Descriptive statistics were used for sample characterization. Validity evidence based on test content came from the CSE judgments. We applied the Content Validity Index (CVI) at item-level (I-CVI) and scale-level (S-CVI/Ave)¹⁹. The I-CVI was computed as the relevance agreement proportion (number of CSEs rating 3 or 4 for each item, divided by the number of CSEs)¹⁹. We calculated the CVI/Ave, computing the I-CVI for each item, then calculated the average I-CVI across all items. The instrument was judged as having excellent content validity if items showed an I-CVI of 0.78 or higher and an S-CVI/Ave of 0.90 or higher¹⁹. Internal consistency was estimated with Cronbach's alpha coefficients, and a value ≥ 0.7 was considered good reliability²⁰.

Results

Translation and Cross-Cultural Adaptation

In the adapted version, nine items were considered “equivalent” in all criteria by at least 70% of the CSEs; four items were considered “equivalent” by 100% of the CSEs in all criteria (items 2, 6, 9, and 10). Only item 3, “*Showed interest in my ideas about my health,*” obtained a lower agreement (60%) in the cultural equivalence due to the meaning of the word “ideas” in the Brazilian context. This item was rewritten according to the CSEs’ suggestions and the term “*ideia*” was replaced by “*opinions.*” No other adaptations were necessary.

Evidence of Content Validity

All items were associated to the construct and considered “relevant” or “very relevant” with 100% of CSE agreement (n=10). The IVC-I ranged from 0.9 to 1. The CVI/AVE was 0.99. Based on this data, all items were maintained in the CAT Brazilian-Portuguese version.

Interviews with SPs

The SPs evaluated the adapted instrument before and after the application. At the first meeting, they received guidance on completing the CAT, and the meaning of each item. Their understanding of each item also was verified. After the application of the instrument, SPs were individually interviewed. None had doubt regarding the items’ content and instrument’s form. All considered the CAT easy to understand and fill out. The items “*Involved me in decisions as much as I wanted*” and “*Discussed next steps, including follow-up plans*” were considered as the most difficult to assess. All reported using their emotions and feelings during the simulation to assess residents’ performance. They completed the instrument in one minute but one suggested that a

little more time would be more “*comfortable*.” No adaptation was required after this step. The final version of the CAT Brazilian-Portuguese version is presented in Table 2.

[Insert Table 2 about here.]

Evaluation of Residents’ Scores in the CAT Brazilian-Portuguese Version

Participants in the OSCE included 28 Ob-Gyn residents (11 PGY-1, 9 PGY-2, and 8 PGY-3; 78.5% female); 71.4% had previously training in CS and 92.8% had already be participated in OSCE assessment.

The Cronbach’s alpha coefficients were calculated at each station, including all 28 residents. Cronbach’s alpha was 0.93 for station 1 and 0.95 for station 2. for the entire instrument.

The residents’ scores ranged from 17 to 64 (mean=36.1, SD=11.11). Scores did not show significantly variation by resident gender (mean male=38.3, mean female=35.5, $p=0.19$). Concerning items rating, only 21.4% (6/28) of residents received at least one item rated as “Excellent”. Among those, the proportion of “Excellent” ratings ranged from 4 to 32% of items. The highest scored items were “*Treated me with respect*” (mean=2.96), “*Paid attention to me (looked at me, listened carefully)*” (mean=2.96), and “*Greeted me in a way that made me feel comfortable*” (meane=2.89). The lowest scored items were “*Encouraged me to ask questions*” (mean=2.14) and “*Involved me in decisions as much as I wanted*” (mean=2.05). The average percentage of items with “Excellent” performance was 4% (31/779). Mean score, and percentage of excellent rating by items are showed in Table 3.

[Insert Table 3 about here.]

Discussion

The use of validated instruments and the inclusion of simulated patients’ perceptions are recommended as good practice to assess doctor–patient

communication^{9; 10}. With this objective, we choose to translate and adapt the CAT to be used in Brazilian medical education. According to our results, the CAT is an appropriate choice for formative assessment of CS in medical residents in a simulated environment.

The process of cross-cultural adaptation was conducted according to the standards recommended by the international literature^{14; 15; 16}. In addition to the semantic, idiomatic, cultural, and conceptual equivalences, content validity and operational equivalence to use the CAT in a simulated environment were also verified. All items in the original tool were maintained in the CAT Brazilian-Portuguese version, observing the high degree of agreement among the CSEs. Only one item was adapted to the Brazilian context, according to the CSEs' reports.

There is no consensus about the ideal strategy for transcultural adaptation of instruments, and methodological choice has been pointed out as a matter of preference and logistics^{16; 21}. The protocols used for similar purposes are quite different in their approach (technical translation, focus group, expert committee, etc.). Most recommend checking the content validity while verifying cross-cultural adaptation. However, they rarely define the steps to achieve this objective²¹. In our protocol, the content validity was verified by 10 CSEs to ensure that the CAT Brazilian-Portuguese version contained only relevant items. The role of the CSEs was fundamental, since there is no Brazilian consensus for CS curriculum in medical schools. For this reason, the members were selected carefully; all were experienced teachers and CS Experts from different Brazilian regions. Thus, the methodology used to verify the content validity was described in detail and the experts acted in two distinct ways (assessing linguistic equivalence and test content), playing a vital role in critical decisions to finalize the CAT Brazilian-Portuguese version.

Reicheneim and Moraes²⁸ recommend assessing operational equivalence, if the adapted instrument is used differently from the original version, to ensure effectiveness. This step was included because the CAT was originally developed for use in real settings¹⁷ and this research was conducted during an OSCE. Although the CAT had already been used in a simulated environment²³, we consider it fundamental to verify its operationalization in a simulated environment by Brazilian SPs. The SPs evaluated the adapted instrument before and after the application. At the first meeting, they received guidance on completing the CAT and their understanding of each item was verified. At the second meeting, after the application of the instrument, individual interviews were conducted. The interviewees were encouraged to express their difficulties, opinions, and suggestions on possible changes. No problems were identified, and no further adaptations were required. The time for completing the instrument was considered adequate, as recommended by the instrument developers (1–2 minutes)¹². This phase confirmed the conceptual and operational equivalence of the adapted instrument. Based on the SPs' perspective, the CAT Brazilian-Portuguese version is a simple and accessible tool to be used in OSCEs.

The internal consistency of the CAT was adequate in both stations (Cronbach's $\alpha=0.93$ and 0.95) and these results are similar to other studies using CAT with medical residents^{22; 23; 24}. However, the residents' scores showed values below those described in other studies using real patients^{22; 23} but similar to those in a simulated environment²⁴. We suppose that the lower ratings were due to the training of SPs to rating the residents' performance as reported by Trickey et al.²⁴ However, studies showed that the CS assessment by SPs has a high reliability and higher correlation with the assessment of medical students in real environments¹⁰. Considering this, another important question arises: the transposition of CS knowledge to performance during the

OSCE. In our study, residents had no prior knowledge of the CAT items and did not know exactly which tasks would be assessed. However, 71.4% of them had received training in CS and 92.8% had participated in assessments in a simulated environment. Thus, it is clear that the residents experienced difficulty in translating their knowledge into a clinical practice perceived as “*excellent*” by SPs during the simulation. After these results, the Obstetrics & Gynecology Department of FCM-Unicamp has organized an annual formative assessment of residents to improve their skills in doctor-patient relationship. We believe that the absence of a Brazilian consensus on CS made the definition of learning goals and assessment standards during medical training difficult²; 25; 26 .

Our study has some limitations. Firstly, it is the first evaluation of validity evidence of a Brazilian-Portuguese version of the CAT. Independent replication of these results is required in other institutions and contexts. Secondly, we used a small sample size and only two OSCE stations, due restrictions of time and financial resources. Finally, as we did not have another validated scale for CS assessment in Brazilian medical education at the time of this research, we could not design a convergent study.

We conclude that the CAT Brazilian-Portuguese version is a valid and reliable option to assess medical residents’ CS in Brazil. The CAT proved to be a simple and feasible instrument when used in simulated environment by trained SPs. Studies on larger samples and in different scenarios are needed to confirm the results of this study.

"This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001"

References

- ¹ BRASIL. Ministério da Educação. Conselho Nacional de Educação. Câmara de Educação Superior. Resolução CNE/CES Nº 4, de 7 de novembro de 2001.

Institui diretrizes curriculares nacionais do curso de graduação em Medicina. Diário Oficial da União. Brasília, 09 nov. 2001; Seção 1, p.38

- ² LAIDLAW, A.; HART, J. Communication skills: an essential component of medical curricula. Part I: Assessment of clinical communication: AMEE Guide No. 51. *Med Teach*, v. 33, n. 1, p. 6-8, 2011. ISSN 0142-159x.
- ³ STEWART, M. A. Effective physician-patient communication and health outcomes: a review. *CMAJ: Canadian Medical Association Journal*, v. 152, n. 9, p. 1423-1433, 1995. ISSN 0820-3946/1488-2329.
- ⁴ HASKARD ZOLNIEREK, K. B.; DIMATTEO, M. R. Physician Communication and Patient Adherence to Treatment: A Meta-analysis. *Medical care*, v. 47, n. 8, p. 826-834, 2009. ISSN 0025-7079/1537-1948.
- ⁵ LEVINSON, W. et al. Physician-patient communication: The relationship with malpractice claims among primary care physicians and surgeons. *JAMA*, v. 277, n. 7, p. 553-559, 1997. ISSN 0098-7484.
- ⁶ SANDARS, J. et al. Educating undergraduate medical students about patient safety: Priority areas for curriculum development. *Medical Teacher*, v. 29, n. 1, p. 60-61, 2007/01/01 2007. ISSN 0142-159X.
- ⁷ KIESSLING, C. et al. Communication and social competencies in medical education in German-speaking countries: the Basel consensus statement. Results of a Delphi survey. *Patient Educ Couns*, v. 81, n. 2, p. 259-66, Nov 2010. ISSN 1873-5134.
- ⁸ VON FRAGSTEIN, M. et al. UK consensus statement on the content of communication curricula in undergraduate medical education. *Med Educ*, v. 42, n. 11, p. 1100-7, Nov 2008. ISSN 0308-0110.

- 9 NORCINI, J. et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach*, v. 33, n. 3, p. 206-14, 2011. ISSN 1466-187X.
- 10 SWANSON, D. B.; VAN DER VLEUTEN, C. P. Assessment of clinical skills with standardized patients: state of the art revisited. *Teach Learn Med*, v. 25 Suppl 1, p. S17-25, 2013. ISSN 1040-1334.
- 11 HULSMAN, R. L. The art of assessment of medical communication skills. *Patient Educ Couns*, v. 83, n. 2, p. 143-4, May 2011. ISSN 0738-3991.
- 12 MAKOUL, G.; KRUPAT, E.; CHANG, C. H. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. *Patient Educ Couns*, v. 67, n. 3, p. 333-42, Aug 2007. ISSN 0738-3991.
- 13 SWING, S. R. Assessing the ACGME General Competencies: General Considerations and Assessment Methods. *Academic Emergency Medicine*, v. 9, n. 11, p. 1278-1288, 2002. ISSN 1553-2712.
- 14 BEATON, D. E. et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*, v. 25, n. 24, p. 3186-91, Dec 2000. ISSN 0362-2436.
- 15 GJERSING, L.; CAPLEHORN, J. R.; CLAUSEN, T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC Med Res Methodol*, v. 10, p. 13, Feb 2010. ISSN 1471-2288.
- 16 REICHENHEIM, M. E.; MORAES, C. L. [Operationalizing the cross-cultural adaptation of epidemiological measurement instruments]. *Rev Saude Publica*, v. 41, n. 4, p. 665-73, Aug 2007. ISSN 0034-8910.
- 17 SIRECI, S.; FAULKNER-BOND, M. Validity evidence based on test content. *Psicothema*, v. 26, n. 1, p. 100-7, 2014. ISSN 1886-144X.

- 18 PADILLA, J. L.; BENITEZ, I. Validity evidence based on response processes. *Psicothema*, v. 26, n. 1, p. 136-44, 2014. ISSN 0214-9915.
- 19 POLIT, D. F.; BECK, C. T.; OWEN, S. V. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*, v. 30, n. 4, p. 459-67, Aug 2007. ISSN 0160-6891.
- 20 TAVAKOL, M.; DENNICK, R. Making sense of Cronbach's alpha. *Int J Med Educ*, v. 2, p. 53-55, Jun 27, 2011. ISSN 2042-6372.
- 21 EPSTEIN, J.; SANTO, R. M.; GUILLEMIN, F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol*, v. 68, n. 4, p. 435-41, Apr 2015. ISSN 1878-5921.
- 22 STAUSMIRE, J. M. et al. Measuring general surgery residents' communication skills from the patient's perspective using the Communication Assessment Tool (CAT). *J Surg Educ*, v. 72, n. 1, p. 108-16, Jan-Feb 2015. ISSN 1878-7452.
- 23 MYERHOLTZ, L. Assessing Family Medicine Residents' Communication Skills from the Patient's Perspective: Evaluating the Communication Assessment Tool. *J Grad Med Educ*, v. 6, n. 3, p. 495-500, Sep 2014. ISSN 1949-8349 (Print) 1949-8357.
- 24 TRICKEY, A. W. et al. Assessment of Surgery Residents' Interpersonal Communication Skills: Validation Evidence for the Communication Assessment Tool in a Simulation Environment. *J Surg Educ*, May 20 2016. ISSN 1878-7452.
- 25 KIESSLING, C. et al. Communication and social competencies in medical education in German-speaking countries: the Basel consensus statement. Results of a Delphi survey. *Patient Educ Couns*, v. 81, n. 2, p. 259-66, Nov 2010. ISSN 0738-3991.

- ²⁶ KURTZ, S. et al. Marrying content and process in clinical method teaching: enhancing the Calgary-Cambridge guides. *Acad Med*, v. 78, n. 8, p. 802-9, Aug 2003. ISSN 1040-2446 (Print) 1040-2446.

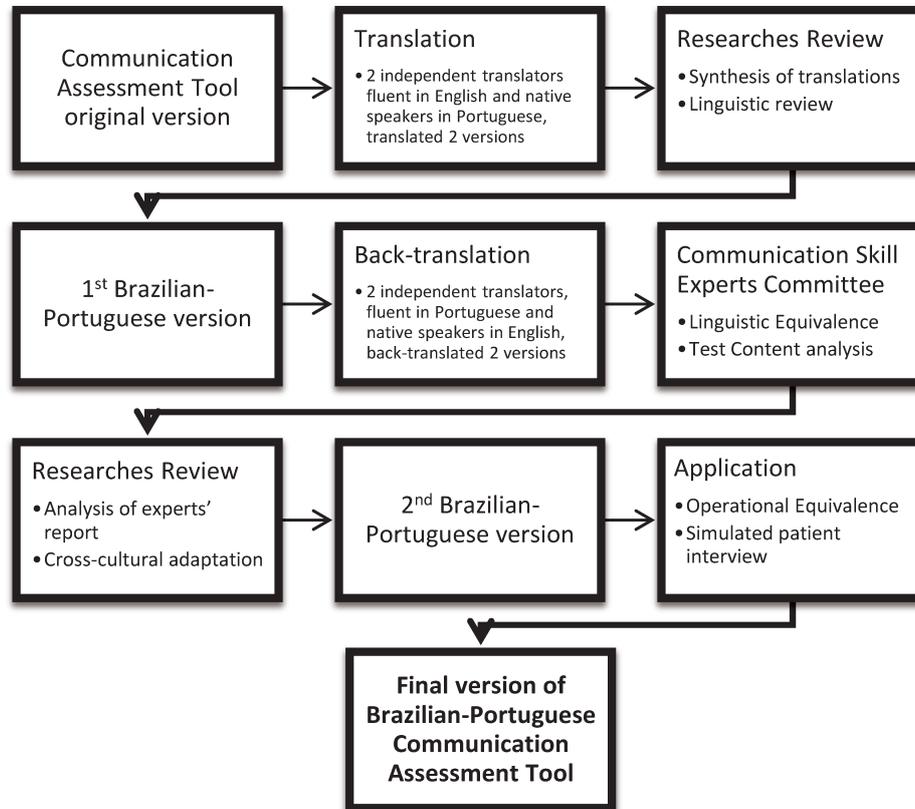


Figure 1 - Step sequence for the translation and cross-cultural adaptation process of The Brazilian-Portuguese Communication Assessment Tool.

Table 1 – Demographic characteristics of Communication Skill Experts Committee

Feature	N	%	Mean (Y)
Sex			
Female	6	60	
Degree			
Doctorate	6	60	
Master	4	40	
Teaching in Medical Education			25
4-20 years	3	30	
21-36 years	7	70	
Teaching Communication Skills			14
2-20 years	7	70	
21-36 years	2	20	
Experience in scale adaptation			
Yes	4	40	
Brazilian Region			
Northeast	1	10	
South	2	20	
Southeast	7	70	

Table 2 – Communication Assessment Tool scores

CAT item	Mean score (SD)	Range	% of items rated as “Excellent”
1. Greeted me in a way that made me feel comfortable	2.89 (1.15)	1–5	8.9
2. Treated me with respect	2.96 (0.92)	1–5	5.5
3. Showed interest in my ideas about my health	2.45 (1.04)	1–5	1.8
4. Understood my main health concerns	2.32 (0.99)	1–5	1.8
5. Paid attention to me (looked at me, listened carefully)	2.96 (1.06)	1–5	7.1
6. Let me talk without interruption	2.79 (0.99)	1–5	1.8
7. Gave me as much information as I wanted	2.68 (0.96)	1–5	3.6
8. Talked in terms I could understand	2.68 (1.10)	1–5	5.4
9. Checked to be sure I understood everything	2.50 (1.06)	1–5	1.8
10. Encouraged me to ask questions	2.14 (1.05)	1–4	0
11. Involved me in decisions as much as I wanted	2.05 (0.87)	1–4	0
12. Discussed next steps, including any follow-up plans	2.71 (0.97)	1–5	7.1
13. Showed care and concern	2.61 (1.00)	1–5	7.1
14. Spent the right amount of time with me	2.58 (1.06)	1–5	3.8
Total	36.1 (11.11)	17–64	4

4.2. Artigo 2: Assessing the Communication Skills of Medical Students: A Validation Study of the Brazilian Portuguese Adaptation of the Analytic Global OSCE Rating

Assessing the Communication Skills of Medical Students: A Validation Study of the Brazilian Portuguese Adaptation of the Analytic Global OSCE Rating

Sheyla Ribeiro Rocha, MD, MS, Gustavo Salata Romão, MD, PhD, Carlos Fernando Collares, MD, PhD, Angélica Maria Bicudo, MD, PhD, and Eliana Amaral, MD, PhD

Sheyla Ribeiro Rocha is an assistant professor, Department of Medicine, Center for Biological and Health Sciences, Federal University of São Carlos, São Carlos, Brazil. She completed the post graduate program in Internal Medicine, Faculty of Medical Sciences, State University of Campinas, Campinas, Brazil.

Gustavo Salata Romão is a professor, School of Medicine, University of Ribeirão Preto, Ribeirão Preto, Brazil.

Carlos Fernando Collares is an assistant professor, Department of Educational Development and Research, School of Health Professions Education, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, The Netherlands.

Angélica Maria Bicudo is an associate professor, Pediatrics Department, Medicine School, Faculty of Medical Sciences, State University of Campinas, Campinas, Brazil

Eliana Amaral is a professor, Obstetrics & Gynecology Department, Medicine School, Faculty of Medical Sciences, State University of Campinas, Campinas, Brazil

Correspondence should be addressed to Sheyla Ribeiro Rocha, Department of Medicine, Center for Biological and Health Sciences, Federal University of São Carlos, Rodovia Washington Luís, km 235, São Carlos, São Paulo CEP: 13565-905, Brazil; telephone: +55(16)99766-4550 - +55(16)3351-83404; e-mail: sheyla.ribeiro.rocha@gmail.com

Abstract

Purpose This authors examined the validity of the Analytic Global Objective Structured Clinical Examination Rating (AG-OSCE-R) for assessment of communications skills (CS) in medical education.

Method The authors translated and adapted the original scale into Brazilian Portuguese using translation, transcultural adaptation, back-translation, and pretesting to produce the Brazilian version of the AG-OSCE-R. This scale was then used by faculty members to assess medical students participating in their final-year formative OSCE in October 2016, and evaluated its content validity using a committee of experts in CS; its internal structure using the many-facet Rasch model (MFRM) and principal component analysis (PCA); and convergent validity using Spearman correlations.

Results In the content analysis, 9 of 10 experts rated all items as relevant for assessing CS. Forty faculty members then used the scale to assess 249 medical students during a two-station OSCE. The MFRM and PCA demonstrated good model fit and suggested a one-factor model for the AG-OSCE-R. The internal consistency reliability (MFRM reliability of separation) was >0.80 for all items and examinees. The AG-OSCE-R had a strong and significant correlation with the estimated examinees measure ($r=0.883$) but moderate correlations with the checklist scores for the internal medicine and obstetrics and gynecology stations ($r=0.376$ and 0.476 , respectively).

Conclusions The Brazilian version of the AG-OSCE-R demonstrated good validity. Furthermore, the scale was accurately completed by OSCE examiners, rather than CS specialists. Altogether, the scale may be a useful and robust formative assessment of CS during OSCEs in medical education.

Introduction

Physicians' communication skills (CS) can strongly influence patient care: it can improve patients' physical and psychological health outcomes, adherence to treatment, and satisfaction, as well as improve health care quality, reduce malpractice claims, and increase patient safety and error reduction.^{1; 2; 3; 4} Accordingly, CS is a core clinical skill that medical students must be effectively taught.^{5; 6; 7; 8; 9}

The objective structured clinical examination (OSCE) is the current recommended approach for assessing CS. OSCEs rely on interactions with standardized patients (SPs).¹⁰ OSCE stations may be designed to assess CS in an integrated fashion (such that CS are evaluated during clinical tasks such as management planning or clinical examination) or separate from other aspects of care.¹¹ CS assessment is a complex task because the relationship between students' performance and assessment criteria is often less clear than in knowledge assessment.¹² Good assessment criteria are thus needed to provide a solid basis for performance judgments.¹³

Several rating scales have been developed to assess medical students' CS in OSCE settings. However, recent systematic reviews have identified several research gaps concerning the measurement of physician–patient communication and the methodological quality of CS instruments.^{14; 15; 16} Particularly, the factorial structure of OSCE rating scales has rarely been reported, and the content validity not at all reported.¹⁵ Thus, further psychometric evaluation studies is necessary to guarantee valid CS assessment.

In 2015, a literature search of Medline/PubMed and LILACS revealed no valid scale for assessing CS during OSCEs in Brazilian medical education. To correct this, we adapted the Analytic Global OSCE Rating (AG-OSCE-R) developed by Hodges and McIlroy,¹⁷

into the Brazilian context. We chose this scale because it (1) was developed specifically for use in OSCEs, (2) can be completed quickly (1 min), (3) can differentiate between students with different levels of training, and (4) can be used by trained OSCE examiners rather than CS experts.¹⁸ Following its adaptation, we sought to collect evidence of its validity.

We first performed the cross-cultural adaptation process (CCAP), which was followed by an investigation of the measurement properties. We were guided by the following hypotheses: (1) The content of the adapted scale supports formative assessment of CS in Brazilian medical education; (2) the rating scale categories are used sequentially by examiners (i.e., improved performance generates higher scores); (3) the scale is sensitive to students' training level (i.e., more advanced students have higher scores); (4) scores are reliable across repeated administrations of the scale; and (5) there is a correlation between the AG-OSCE-R score and scores on checklists of clinical skills.

Method

This research project was approved by the Institutional Human Ethical Committee of State University of Campinas (Unicamp) (CAAE 56935716.5.0000.5404) and was carried out in two phases.

Phase 1: Cross-cultural adaptation process and content validity

Procedure

Initially, we adapted the AG-OSCE-R into a Brazilian context and evaluated its content validity. To translate the scale, we used the following steps based on CCAP guidelines¹⁹: First, two independent official translators fluent in English and who were

native Portuguese speakers created two translated versions. Second, we compared the translations, assessing their linguistic and contextual differences to create a single, summarized version. Third, two independent English native speakers back-translated the summarized version into English (neither had access to the original version nor had been informed of the scale's concepts). Fourth, experts in CS (ECS) analyzed the adapted scale for semantic, idiomatic, conceptual, and experiential equivalence with the original and its content validity²⁰ using a 4-point scale. The ECS based their ratings on the original AG-OSCE-R, the summarized translation, the two back-translations, and a written report explaining the CCAP rationale. The ECS's reports were used to adjust the Brazilian version of the scale. Finally, the prefinal version was pretested in a formative OSCE for medical residents in obstetrics and gynecology (OB-GYN).

The Analytic Global OSCE Rating (AG-OSCE-R)

The AG-OSCE-R comprises four single-item domains (hereafter, items): (1) response to the patient's feelings and needs (empathy); (2) degree of coherence in the interview (e.g., organization, demonstrating command of cohesive devices); (3) verbal expression (e.g., fluency, volume and modulation of voice); and (4) nonverbal expression (e.g., facial expressions, use of silence). Each item is assessed on a 5-point Likert scale, ranging from 1=poor to 5=excellent. Scores of 1, 3 and 5 are behaviorally anchored to a description of the expected performance (Annex 1). The item scores are summed to generate a global rating ranging from 4 to 20. At the end of each OSCE station, examiners assess students' overall performance in terms of three behaviorally anchored categories: (A) incompetent, (B) borderline, and (C) competent.¹⁷

Pilot OSCE Organization

In 2016, a pilot OSCE was organized during the Obstetric Emergencies Course of the OB-GYN Medical Residency Program at the State University of Campinas Medical School (FCM-Unicamp). Four examiners (medical school staff) were invited to receive a 1-hour AG-OSCE-R training session and provide feedback on the scale. All 28 residents were informed of the pilot and invited to participate. Residents completed two OSCE stations on CS with trained SP (a woman making her first prenatal visit with a positive venereal disease research laboratory and a pregnant woman with unexpected preeclampsia and high risk of fetal loss). All participants provided written informed consent before data collection.

Participants

The ECS included 10 faculty members (60% female) with a median of 13.5 years of teaching CS (80% had a PhD and 20% had just an MD) from different regions of Brazil. Nine were medical doctors. Pilot study participants were 28 OB-GYN residents (78.5% female), including 11 (39.3%) in postgraduate year one, 9 (32.2%) in year 2, and 8 (28.5%) in year 3, and four medical examiners (two females). The sample size according to the CCAP guidelines agrees with Beaton et al.'s guidelines¹⁹—namely, about 30 pilot study participants and 10 experts.

Phase 2: Validation Process

In Phase 2, we examined the adapted scale's internal structure and relations with other variables during the Clerkship OSCE for undergraduate medical students of FCM-Unicamp.

The undergraduate medical curriculum at Unicamp Medical School comprises four preclinical years and two years of clinical clerkship in the major areas of medicine. In the second year, students are taught history-taking skills in seminars on interviewing

techniques. In the next four years, they practice their CS in primary and outpatient care and in the hospital wards in different medical disciplines. In the sixth year, simulations with SPs are scheduled as a curricular activity, with a focus on empathetic communication (including feedback by the SPs), during their emergency rotation.²¹

Clerkship OSCE Organization

The Clerkship OSCE is a routine formative assessment offered to medical students in their 4th to 6th years. The Nucleus for Assessment and Research in Medical Education of the Medical Science School coordinates the organization and administration of this OSCE. Students' participation is completely facultative. All students performed the same five stations (involving history taking, clinical reasoning, and/or physical examination) covering five disciplines (internal medicine, surgery, pediatrics, OB-GYN, and public health). At each station, students spent 7 minutes performing the task and 3 minutes receiving formative feedback. Due to limited resources, only one examiner rated students' performance at each station. Examiners were faculty members or physicians recruited from the medical school staff. Academic community members were employed and trained as SP. The OSCE runs in 20 parallel circuits. In 2016, we administered the adapted AG-OSCE-R to assess CS in the internal medicine and OB-GYN stations (along with the standard clinical skills checklist). We chose integrated stations because running stations exclusively for assessing CS exceeded our personal and financial resources. All examiners received instructions on administering the scale on the OSCE day.

Participants

Clerkship OSCE participants in 2016 were 249 (75% of the 330 medical students enrolled, 60% female) undergraduate medical students, including 59, 92, and 98 from

years 4, 5, and 6, respectively. Furthermore, 40 medical examiners participated (55% female). The Institutional Human Ethical Committee did not require us to attain written informed consent for this phase because the data collection and analysis focused on regular educational practice and added minimal risk for participants.

Data analysis

Descriptive statistics were used for sample characterization. We used item-level and scale-level content validity index (I-CVI and S-CVI/Ave) to assess content validity.²² The I-CVI was computed as the number of experts giving a rating of 3 or 4 divided by the total number of experts (i.e., the proportion who agree on item relevance).²² We calculated the S-CVI/Ave as the average I-CVI across all items. The scale had excellent content validity if all I-CVIs were 0.78 or higher and the S-CVI/Ave was 0.90 or higher.²²

The construct validity and internal consistency were estimated with the many-facet Rasch measurement model (MFRM) using the Facets® Version N° 3.71.4 (Oregon: Winsteps.com, 2013.). We chose the partial credit model because the AG-OSCE-R contains items with five response options and behavioral anchors specific to each domain (empathy, coherence, and verbal and nonverbal expression). The MFRM enables conversion of ordinal scores into linear measures expressed on the same linear scale (logits), allowing for correct inferences and data comparisons²³. Two important assumptions must be satisfied to use MFRM: local independence (i.e., residuals must be uncorrelated) and unidimensionality (i.e., the data must represent a single construct). The infit and outfit mean square values are relevant for evaluating the data quality. They also offer additional evidence of the internal structural validity. The infit measure is sensitive to variations in personal ability scores, while the outfit measure reflects the occurrence of unexpected data (particularly on the extremes of the latent variable

spectrum). The MFRM also enables reports of item difficulty and discrimination. We adopted an acceptable range for the infit and outfit values is 0.5–1.5.²⁴ The MFRM also provides reliability and separation indices. The reliability index is interpreted similarly to the Kuder–Richardson Formula 20 and Cronbach’s alpha; values of >0.80 indicate a more internally consistent measure.²⁴ Correlations were assessed using Spearman’s rank correlation coefficient.

A conventional principal component analysis (PCA) was also conducted to examine the factor structure of the Brazilian AG-OSCE-R using SPSS Statistics version 22. The number of retained components was based on the following criteria: (a) Kaiser’s eigenvalue greater than 1.0 (K1 criterion), (b) Cattell’s scree test, (c) Bartlett’s chi-square test, and (d) parallel analysis.^{25; 26} A PCA of the Rasch model residuals (PCAR) still was conducted using Winsteps® 4.0 (Linacre, Chicago, 2017) to test for the existence of a second factor that can threaten the unidimensionality concept. We adopted an acceptable range for the first contrast of residual values of ≤ 2 eigenvalue as an indicative of unidimensionality^{24; 27}

Results

Content Validity

During the CCAP, we had difficulty in translating the word “engage” in two instances: “Fails to engage, frustrates and/or antagonizes the patient” and “Exhibits enough control of nonverbal expression to engage a patient willing to overlook deficiencies such as passivity, self-consciousness or inappropriate aggressiveness.” We therefore contacted the scale’s developer to better understand these items, and conceived the final

versions with the support of the ECS reports. All other items were classified as “equivalent” by at least 70% of the experts. The ECS’s suggestions were used to improve the final version.

All items were rated as highly or quite relevant by at least 90% of experts (see Figure 1). The I-CVI of items 1, 2, 3, and 4 were 1, 0.9, 1, and 0.9, respectively, and the S-CVI/Ave was 0.95. During the pilot study, the examiners reported clear understanding of all items in the Brazilian AG-OSCE-R and no further adjustments were required. The final version is available as Supplemental Digital Content.

Internal Structure

Rasch Analysis

We used 1906 responses to estimate the model parameters. Only 20 responses were extreme scores (1.04%). The mean fit indices of the four facets were all 0.5–1.5 (see Table 1). A more detailed analysis revealed that all items had adequate discrimination, infit, and outfit indexes (see Table 2), altogether indicating that the MFRM had a satisfactory fit to the data.²⁴

The mean item difficulty is 0.00 (SD=0.46), with item 3 being the easiest and item 2 the most difficult. The mean examinee ability was 1.81 logits (SD=1.94), suggesting that the items measured higher-level CS in this sample. The mean severity imposed by the examiners was fixed at 0.00 logits (SD=1.38), with a separation ratio of 4.70 ($p=0.00$).

This finding suggests that the severity ranged considerably among examiners.²³

However, the range in examiner severity was narrower than was that for examinee ability (–3.89 to 9.16), supporting the assumption that examiners generally followed the requirements of the scoring process.²⁸ The two stations had similar difficulty levels ($M=0.00$, $SD=0.01$). These results are evident in the Wright Map (see Figure 2), which

depicts how examinees, examiners, items, and stations locations are plotted in terms of CS ability. The items and stations are distributed linearly in order of difficulty (from the easiest to most difficult), the examinees in order of ability (from the least to most skilled), and examiners in order of severity (from the least to most severe).²³

The examinee separation analysis (see Table 1) indicated that examinees could be divided into three levels of ability, with a reliability of separation index of 0.87 (separation ratio=2.6, $p=0.00$)²³. Thus, the scale can statistically distinguish between high, medium, and low performers.^{23; 24} The item separation analysis revealed that the items were distributed into five difficulty levels with a reliability of separation index of 0.97 (separation ratio=5.51, $p=0.00$); thus, the sample was large enough to confirm the item difficulty hierarchy. Overall, these results support the construct validity and reproducibility of the scale.²⁴

Additional evidence of the validity comes from the increase in the observed average measure of each item response category at both stations (see Figure 3). In other words, the threshold parameters of the five response categories were ordered and reflected the intended order for all items. The frequency of observations of each category indicates that they were used for all items at both stations (save for category 1). Taken together, the results suggested that the examiners used the AG-OSCE-R in line with the developers' intentions.²⁴

Principal Components Analysis (PCA)

The conventional PCA was conducted on 240 participants who responded to all items. The Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy (KMO=0.829); KMO values for the individual items were all greater than 0.8 (save item

three). Bartlett's sphericity test ($\chi^2(6)=1058.058$, $p=0.001$) indicated that the items were sufficiently correlated for PCA.

The initial analysis revealed only one component that observed the K1 criterion; this single component explained 74.626% of the variance (eigenvalue=2.985). The scree plot similarly showed only one component positioned before inflection (see Figure 4). The loadings were high for all items. A parallel analysis similarly recommended a one-factor solution, as expected. Taking all these results together, the instrument had only one component.

In the PCAR between examinees and examiners facets, Rasch measures explained 56.8% of the variance, with the first residual contrast explaining only 2.5% of the variance (eigenvalue=1.999). This again suggests a unidimensional solution. However, when examinees and stations were analyzed, even though the Rasch measures explained 52.2% of the variance, the first residual contrast explained 22.9% of the variance (eigenvalue=3.830). This finding suggests some content/task-specificity between the two stations.

Relationship to Other Variables

The AG-OSCE-R Score was very strongly correlated ($r=0.883$, $p=0.000$) with the examinees ability and moderately correlated with the internal medicine and OB-GYN station checklist scores ($r=0.376$ and $r=0.476$, respectively; $p=0.000$).

Discussion

We evaluated the psychometric properties of the Brazilian AG-OSCE-R for formative assessment of medical students in OSCEs. We collected evidence of its content validity, validity based on internal structure, and convergent validity.

The researchers conducted this study during the regular final year Clerkship OSCE of Unicamp Medical School, a formative assessment of clinical skills proposed for the last three years of medical undergraduate. We used the AG-OSCE-R in two stations simultaneously assessing clinical skills and CS because running stations exclusively to assess CS exceeded personal and financial resources. For the same reason, we had only one examiner at each station.

The developers of the AG-OSCE-R provided no clear definition of CS, and previous studies on this measure have not evaluated its content validity.^{17; 18; 29} We corrected these gaps by having ten ECS analyze the content of the AG-OSCE-R in terms of comprehension, domain representation, and relevance to assessing CS in medical education. The translated version required only minor corrections in item 4 (based on experts' reports), and all items were rated as relevant. The pilot study also revealed no discrepancies in meaning or terminology, and no modifications were required.

Therefore, we assume that the AG-OSCE-R contains only relevant items for formative assessments of CS in medical education.

Previous studies have not reported on the internal structure of the AG-OSCE-R or its dimensionality.^{17; 18; 29} Because all items in the AG-OSCE-R are summed to form a total score, it is tacitly assumed that the scale is unidimensional¹⁷. To confirm this, we performed a PCA using multiple criteria and a PCAR. The existence of a substantive secondary residual dimension requires at least 5% variance explained, or an eigenvalue higher than 2.0.³⁰ However, simulation studies have shown that an eigenvalue of >1.4

indicates some structure in the residuals, and recommend determining the criterion eigenvalue directly from the relevant simulations.³¹ Therefore, while the PCA clearly pointed out a one-factor solution, the most rigorous PCAR did not suggest absolute unidimensionality; rather, it reflected an “essentially unidimensional” solution. This concept of “essential unidimensionality” refers to when the dominant latent trait coexists with one or more minor latent traits but is not affected by their presence.³² However, it should be applied to educational measures with caution.²⁶ Researchers should use multiple criteria when deciding on the appropriate number of factors to retain in order to discern essential unidimensionality from genuine multi-dimensionality.²⁶ However, there is no universally accepted rule for this.^{25; 26} Based on the statistical and content analysis, the AG-OSCE-R appears to be essentially unidimensional. It means that the items measure predominantly CS with some possible secondary minor latent variables but the major latent trait (CS) is not affected by the presence of the secondary one. Content test analysis support this statement.

The Rasch analysis demonstrated that the Brazilian AG-OSCE-R showed good item discrimination with highly positive correlations. Furthermore, it could roughly divide examinees into three skills levels, which might be considered suitable. Furthermore, the high reliability (based on a separation index of 0.87, $p=0.00$) suggests that we can reproduce these findings in similar samples elsewhere. In previous studies of the AG-OSCE-R, item reliability was estimated by the Cronbach’s alpha, and it varied from 0.52²⁹ to 0.70.¹⁷ The Cronbach's alpha and the separation index reported in a Rasch analysis are both estimates of the ratio of “true measure variance” to “observed measure variance.”³³ We did not observe ceiling or floor effects. The high person and item separation indexes for the AG-OSCE-R further support the internal structure of the scale.

Importantly, the Rasch analysis provides a reliability index for examiners, but this is not interrater reliability. The high reliability index for examiners (0.96) in this study indicated that although examiners' severity differed somewhat, their ratings are likely to remain stable on subsequent administrations of the scale. Variability in examiners' severity is a common problem among clinical assessments.³⁴ Increasing use of performance-based assessments in medical education has highlighted variability in observed examiner judgements, and there is debate about whether and how this variability can (or if it needs to) be overcome^{35; 36; 37; 38}.

The high correlation between the AG-OSCE-R total score and the examinees ability provide further evidence of validity.³⁹ We also found significant and moderate positive correlation between the AG-OSCE-R total score and checklist scores for both stations (0.376 and 0.476 for internal medicine and OB-GYN stations, respectively; $p=0.000$). These results suggest that better clinical performance scores are positively correlated with good clinical communication, thus providing evidence of the convergent validity. At the time of this study, there were no validated scales for CS assessment in Brazilian medical education; therefore, we could not design a criterion validity study.

Our study has several limitations. First, a greater sample size, more stations, and two examiners per station would have improved the robustness of the results. Second, the examiners were volunteer faculty members running routine assessments, all of whom were of different backgrounds and were not specialists in CS, which may have contributed to high variability of examiners' severity. However, the examiners all had experience in OSCEs and had received instructions on using the scale on the OSCE day. Assessor training is recommended for improving the reliability of assessments, but effective training is not a perfect solution.^{40; 41} Examiner decision making during assessments is a complex task involving hypothesis generation, high-level inferences

about the competencies rated, and a balance between objectivity and subjectivity, which is influenced by the examiner's initial impressions of the examinee.^{35; 36; 37; 42} Because we had only one rater at each station, we could not compare interrater agreement between expert and non-expert examiners. Furthermore, none of the examiners were blinded to students because all were faculty members or preceptors at the medical school.

This was the first examination of the psychometric properties of a Brazilian adaptation of the AG-OSCE-R. Independent replication of these results is necessary in other institutions and contexts. Our results point to the necessity of further investigations regarding the examiners' cognitive processes during CS assessment as well as other sources of examiner heterogeneity, such as professional background and experience. Furthermore, it is necessary to determine how much more structured training improves the reliability of examiners. Given the current findings suggesting that objectivity and reliability differ and that more subjective, less structured measures might be more reliable, studies on the development and impact of a more efficient rater-based assessment system that incorporates more qualitative data in the assessment of clinical performances would be a welcome development.^{35; 36; 37; 43; 44}

Conclusions

The Brazilian version of the AG-OSCE-R demonstrated good content, structural, and convergent validity. The AG-OSCE-R also showed high reliability, despite its few items and behaviorally anchored rating scale. The AG-OSCE-R can be completely and accurately completed by OSCE examiners of various backgrounds, none of whom are CS specialists. No erratic response patterns were observed and higher performance led

to higher CS scores. Thus, the Brazilian AG-OSCE-R may be a useful and robust choice for assessing CS during OSCE in medical education. Further research examining the response processes of the AG-OSCE-R could highlighted examiners' judgment process during assessment of clinical communication.

"This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001"

References

- ¹ LAIDLAW, A.; HART, J. Communication skills: an essential component of medical curricula. Part I: Assessment of clinical communication: AMEE Guide No. 51. **Med Teach**, v. 33, n. 1, p. 6-8, 2011. ISSN 0142-159x.
- ² HASKARD ZOLNIEREK, K. B.; DIMATTEO, M. R. Physician Communication and Patient Adherence to Treatment: A Meta-analysis. **Medical care**, v. 47, n. 8, p. 826-834, 2009. ISSN 0025-7079/1537-1948.
- ³ LEVINSON, W. et al. Physician-patient communication: The relationship with malpractice claims among primary care physicians and surgeons. **JAMA**, v. 277, n. 7, p. 553-559, 1997. ISSN 0098-7484.
- ⁴ SANDARS, J. et al. Educating undergraduate medical students about patient safety: Priority areas for curriculum development. **Medical Teacher**, v. 29, n. 1, p. 60-61, 2007/01/01 2007. ISSN 0142-159X.
- ⁵ MAKOUL, G. The SEGUE Framework for teaching and assessing communication skills. **Patient Educ Couns**, v. 45, n. 1, p. 23-34, Oct 2001. ISSN 0738-3991.
- ⁶ VON FRAGSTEIN, M. et al. UK consensus statement on the content of communication curricula in undergraduate medical education. **Med Educ**, v. 42, n. 11, p. 1100-7, Nov 2008. ISSN 0308-0110.
- ⁷ SWING, S. R. Assessing the ACGME General Competencies: General Considerations and Assessment Methods. **Academic Emergency Medicine**, v. 9, n. 11, p. 1278-1288, 2002. ISSN 1553-2712.
- ⁸ KIESSLING, C. et al. Communication and social competencies in medical education in German-speaking countries: the Basel consensus statement. Results of a Delphi survey. **Patient Educ Couns**, v. 81, n. 2, p. 259-66, Nov 2010. ISSN 0738-3991.

- 9 LIBERALI, R. et al. Communication skills teaching in Brazilian medical schools: What lessons can be learned? **Patient Educ Couns**, v. 101, n. 8, p. 1496-1499, Aug 2018. ISSN 0738-3991.
- 10 SWANSON, D. B.; VAN DER VLEUTEN, C. P. Assessment of clinical skills with standardized patients: state of the art revisited. **Teach Learn Med**, v. 25 Suppl 1, p. S17-25, 2013. ISSN 1040-1334.
- 11 KHAN, K. Z. et al. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. **Med Teach**, v. 35, n. 9, p. e1447-63, Sep 2013. ISSN 0142-159x.
- 12 HULSMAN, R. L. The art of assessment of medical communication skills. **Patient Educ Couns**, v. 83, n. 2, p. 143-4, May 2011. ISSN 0738-3991.
- 13 NORCINI, J. et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. **Med Teach**, v. 33, n. 3, p. 206-14, 2011. ISSN 1466-187X.
- 14 ZILL, J. M. et al. Measurement of physician-patient communication--a systematic review. **PLoS One**, v. 9, n. 12, p. e112637, 2014. ISSN 1932-6203.
- 15 COMERT, M. et al. Assessing Communication Skills of Medical Students in Objective Structured Clinical Examinations (OSCE)--A Systematic Review of Rating Scales. **PLoS One**, v. 11, n. 3, p. e0152717, 2016. ISSN 1932-6203.
- 16 SETYONUGROHO, W.; KENNEDY, K. M.; KROPMANS, T. J. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. **Patient Educ Couns**, Jun 27 2015. ISSN 0738-3991.
- 17 HODGES, B.; MCILROY, J. H. Analytic global OSCE ratings are sensitive to level of training. **Med Educ**, v. 37, n. 11, p. 1012-6, Nov 2003. ISSN 0308-0110.
- 18 SCHEFFER, S. et al. Assessing students' communication skills: validation of a global rating. **Adv Health Sci Educ Theory Pract**, v. 13, n. 5, p. 583-92, Dec 2008. ISSN 1382-4996.
- 19 BEATON, D. E. et al. Guidelines for the process of cross-cultural adaptation of self-report measures. **Spine (Phila Pa 1976)**, v. 25, n. 24, p. 3186-91, Dec 2000. ISSN 0362-2436..
- 20 SIRECI, S.; FAULKNER-BOND, M. Validity evidence based on test content. **Psicothema**, v. 26, n. 1, p. 100-7, 2014. ISSN 1886-144X.

- 21 SCHWELLER, M. et al. The impact of simulated medical consultations on the empathy levels of students at one medical school. **Acad Med**, v. 89, n. 4, p. 632-7, Apr 2014. ISSN 1040-2446.
- 22 POLIT, D. F.; BECK, C. T.; OWEN, S. V. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. **Res Nurs Health**, v. 30, n. 4, p. 459-67, Aug 2007. ISSN 0160-6891.
- 23 BOONE, W. J.; STAVER, J. R.; YALE, M. S. **Rasch analysis in the human sciences**. Springer, 2013. ISBN 9400768575.
- 24 LINACRE, J. M. Facets computer program for many-facet Rasch measurement, version 3.71.4. **Beaverton, Oregon: Winsteps.com**, 2013.
- 25 ZWICK, W. R.; VELICER, W. F. Comparison of five rules for determining the number of components to retain. **Psychological bulletin**, v. 99, n. 3, p. 432, 1986. ISSN 1939-1455.
- 26 SLOCUM-GORI, S. L.; ZUMBO, B. D. Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. **Social Indicators Research**, v. 102, n. 3, p. 443-461, 2011. ISSN 0303-8300.
- 27 TENNANT, A.; PALLANT, J. F. Unidimensionality matters!(A tale of two Smiths?). **Rasch measurement transactions**, v. 20, n. 1, p. 1048-1051, 2006.
- 28 TOFFOLI, S. F. L.; DE ANDRADE, D. F.; BORNIA, A. C. Evaluation of open items using the many-facet Rasch model. **Journal of Applied Statistics**, v. 43, n. 2, p. 299-316, 2016/01/25 2016. ISSN 0266-4763.
- 29 MORTSIEFER, A. et al. Summative assessment of undergraduates' communication competence in challenging doctor-patient encounters. Evaluation of the Düsseldorf CoMeD-OSCE. **Patient Education and Counseling**, v. 95, n. 3, p. 348-355, ISSN 0738-3991.
- 30 LINACRE, J. M. Local independence and residual covariance: A study of Olympic figure skating ratings. **Journal of applied measurement**, v. 10, n. 2, p. 157-169, 2009.
- 31 RAÏCHE, G. Critical eigenvalue sizes in standardized residual principal components analysis. **Rasch measurement transactions**, v. 19, n. 1, p. 1012, 2005.
- 32 SLOCUM-GORI, S. L. et al. A Note on the Dimensionality of Quality of Life Scales: An Illustration with the Satisfaction with Life Scale (SWLS). **Social Indicators Research**, v. 92, n. 3, p. 489-496, 2009/07/01 2009. ISSN 1573-0921.

- 33 CLAUSER, B.; LINACRE, J. M. Relating Cronbach and Rasch reliabilities. **Rasch Measurement Transactions**, v. 13, n. 2, p. 696, 1999.
- 34 MCGILL, D. A.; VAN DER VLEUTEN, C. P. M.; CLARKE, M. J. Supervisor assessment of clinical and professional competence of medical trainees: a reliability study using workplace data and a focused analytical literature review. **Advances in health sciences education**, v. 16, n. 3, p. 405-425, 2011. ISSN 1382-4996.
- 35 GINGERICH, A.; REGEHR, G.; EVA, K. W. Rater-based assessments as social judgments: rethinking the etiology of rater errors. **Academic Medicine**, v. 86, n. 10, p. S1-S7, 2011. ISSN 1040-2446.
- 36 GAUTHIER, G.; ST-ONGE, C.; TAVARES, W. Rater cognition: review and integration of research findings. **Medical Education**, v. 50, n. 5, p. 511-522, 2016/05/01 2016. ISSN 0308-0110.
- 37 ST-ONGE, C. et al. Expectations, observations, and the cognitive processes that bind them: expert assessment of examinee performance. **Adv Health Sci Educ Theory Pract**, v. 21, n. 3, p. 627-42, Aug 2016. ISSN 1382-4996.
- 38 WASS, V. et al. Assessment in clinical competence. **Lancet**, v. 357, 2001// 2001.
- 39 AMERICAN EDUCATIONAL RESEARCH, A.; AMERICAN PSYCHOLOGICAL, A.; NATIONAL COUNCIL ON MEASUREMENT IN, E. **Standards for educational and psychological testing**. Washington, DC: American Educational Research Association, 2014. ISBN 0935302255.
- 40 DOWNING, S. M. Threats to the validity of clinical teaching assessments: what about rater error? **Med Educ**, v. 39, n. 4, p. 353-5, Apr 2005. ISSN 0308-0110 (Print)0308-0110.
- 41 COOK, D. A. et al. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. **J Gen Intern Med**, v. 24, n. 1, p. 74-9, Jan 2009. ISSN 0884-8734.
- 42 YAPHE, J.; STREET, S. How do examiners decide?: a qualitative study of the process of decision making in the oral examination component of the MRCGP examination. **Medical education**, v. 37, n. 9, p. 764-771, 2003. ISSN 0308-0110.
- 43 WHITEHEAD, C. R. et al. Conceptual and practical challenges in the assessment of physician competencies. **Med Teach**, v. 37, n. 3, p. 245-51, Mar 2015. ISSN 0142-159x.
- 44 HODGES, B. Assessment in the post-psychometric era: learning to love the subjective and collective. **Med Teach**, v. 35, n. 7, p. 564-8, Jul 2013. ISSN 0142-159x.

Supplemental Digital Content - Brazilian adapted version of the Analytic Global OSCE Rating (AG-OSCE-R)

Estação nº – Instrumento de Avaliação da comunicação médico-paciente durante o OSCE

Avaliador: _____

Aluno: _____

Resposta aos sentimentos e às necessidades do paciente (empatia)				
1	2	3	4	5
Não responde a sinais (verbais e não verbais) óbvios apresentados pelo paciente e/ou responde de maneira inapropriada.		Responde às necessidades e aos sinais apresentados pelo paciente, porém nem sempre de forma efetiva.		Responde consistentemente às necessidades e sinais apresentados pelo paciente, de maneira perspicaz e autêntica.
Grau de coerência da entrevista				
1	2	3	4	5
Não foi identificado nenhum plano durante a interação; o plano utilizado não demonstra coesão ou o paciente precisa determinar o direcionamento da entrevista.		A abordagem quanto à organização da entrevista é padronizada, pouco flexível e/ou o controle da entrevista é inconsistente.		A organização da entrevista é excelente, demonstrando domínio de estratégias coesas, flexibilidade e controle consistente da entrevista.
Expressão verbal				
1	2	3	4	5
Comunica-se de modo a interferir na e/ou impedir a compreensão pelo paciente, ou se comunica inadequadamente com o paciente.		Demonstra controle suficiente da expressão verbal de modo a ser compreendido por um ouvinte ativo e participativo (paciente).		Demonstra domínio da expressão verbal (fluência, dicção, gramática, vocabulário, tom, volume e modulação de voz, ritmo da fala, cadência e pronúncia).
Expressão não verbal				
1	2	3	4	5
Não consegue envolver, frustra e/ou entra em conflito com o paciente.		Demonstra controle suficiente de expressão não-verbal para envolver o paciente, e não se perturba diante de certos comportamentos por ele apresentados como passividade, insegurança ou agressividade inadequadas.		Demonstra elegância e domínio da expressão não verbal (contato visual, gestos, postura, uso do silêncio etc.)
Avaliação geral do conhecimento e das habilidades demonstradas na entrevista				
A = Incompetente		B = Limítrofe		C = Competente
Responde de forma inadequada e inefetiva à tarefa, indicando falta de conhecimento e/ou habilidades interpessoais e para entrevistar mal desenvolvidas.		Responde de forma efetiva a alguns componentes da tarefa, apresenta algum desenvolvimento de habilidades interpessoais e para entrevistar.		Responde de forma precisa e com perspicácia à tarefa, integrando todos os componentes de forma consistente.

Table 1. Summary fit statistics using Rasch analysis for each facet (examinees, examiners, and stations) and items from the adapted version of the Analytic Global OSCE Rating (AG-OSCE-R)

	Items (n=4)	Examinees (n=249)	Assessors (n=40)	Stations (n=2)
<i>Measure</i>				
Mean (SD)	0.00 (0.46)	1.81 (1.94)	0.00 (1.38)	0.00 (0.01)
<i>Infit mean square (infit MnSq)</i>				
Mean (SD)	1.04 (0.05)	1.04 (0.71)	1.06 (0.40)	1.03 (0.01)
<i>Outfit mean square (outfit MnSq)</i>				
Mean	1.19 (0.22)	1.07 (1.00)	1.24 (1.01)	1.20 (0.15)
<i>Other measures</i>				
Separation ratio	5.51	2.60	4.70	0.00
Reliability of separation	0.97*	0.87*	0.96*	0.00

***p=0.00**

Table 2. Fit statistics using Rasch analysis for item difficulty and discrimination from the adapted version of the Analytic Global OSCE Rating (AG-OSCE-R)

Item	Name	Difficulty	SE	Discrimination	Point-Biserial Correlation (r)	MNSQ	
						Infit	Outfit
1	Response to patient's feelings and needs (empathy)	-0.09	0.08	0.93	0.74	1.07	1.06
2	Degree of coherence in the interview	0.54	0.08	0.85	0.70	1.11	1.29
3	Verbal expression	-0.69	0.08	1.06	0.73	0.96	0.92
4	Nonverbal expression	0.24	0.08	0.96	0.76	1.01	1.50

SE= Standard Error

* Adequate infit and outfit mean squares (MNSQ) are defined as a range from 0.5 to 1.5

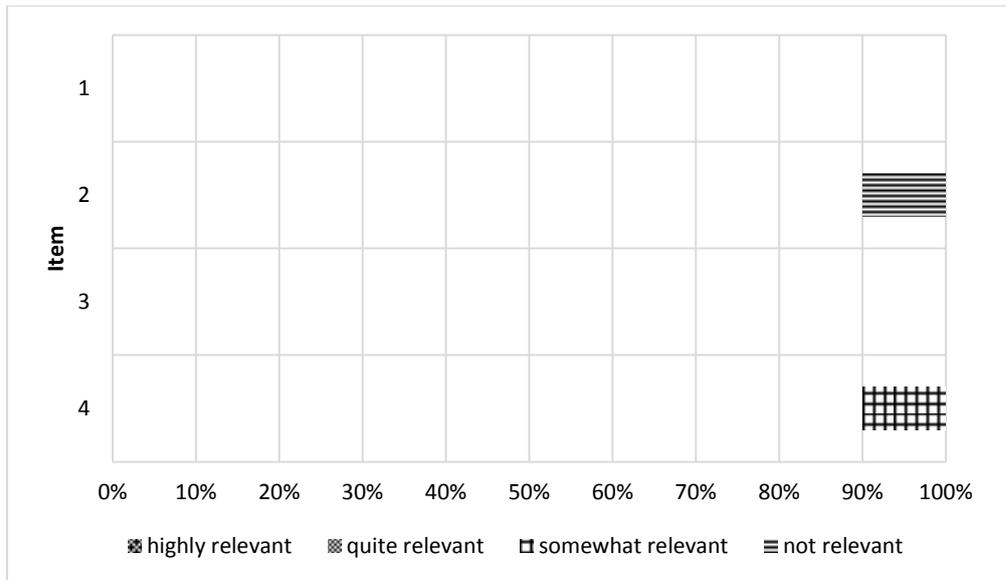


Figure 1: (A) Experts' agreement about item relevance to assess communication skills

Measr	+examinees	-Stations	-Assessors	-Items	S.1	S.2	S.3	S.4	S.5	S.6	S.7	S.8
7	+ *.	+	+	+	+	(5)	+	(5)	+	(5)	+	(5)
6	+ .	+	+	+	+	+	+	+	+	+	+	+
5	+ *.	+	+	+	+	+	+	+	+	+	+	+
4	+ *.	+	+	+	+	+	---	+	+	---	+	---
3	+ ***.	+	+	+	+	4	---	+	+	4	+	4
2	+ ****.	+	+	+	+	+	4	+	4	+	+	+
1	+ *****.	+	+	+	+	+	---	+	+	---	+	---
0	+ *****.	CM GO	+	+	+	+	3	---	---	+	3	3
-1	+ **.	+	+	+	+	+	+	+	+	+	+	+
-2	+ .	+	+	+	+	+	3	---	3	3	---	---
-3	+ .	+	+	+	+	2	+	+	+	2	+	2
-4	+ .	+	+	+	+	(1)	+	(2)	+	(1)	+	(1)
Measr	* = 2	-Stations	* = 1	-Items	S.1	S.2	S.3	S.4	S.5	S.6	S.7	S.8

Figure 2. Wright item-person maps related to the communication skills (CS) ability of the adapted version of the Analytic Global OSCE Rating (AG-OSCE-R)

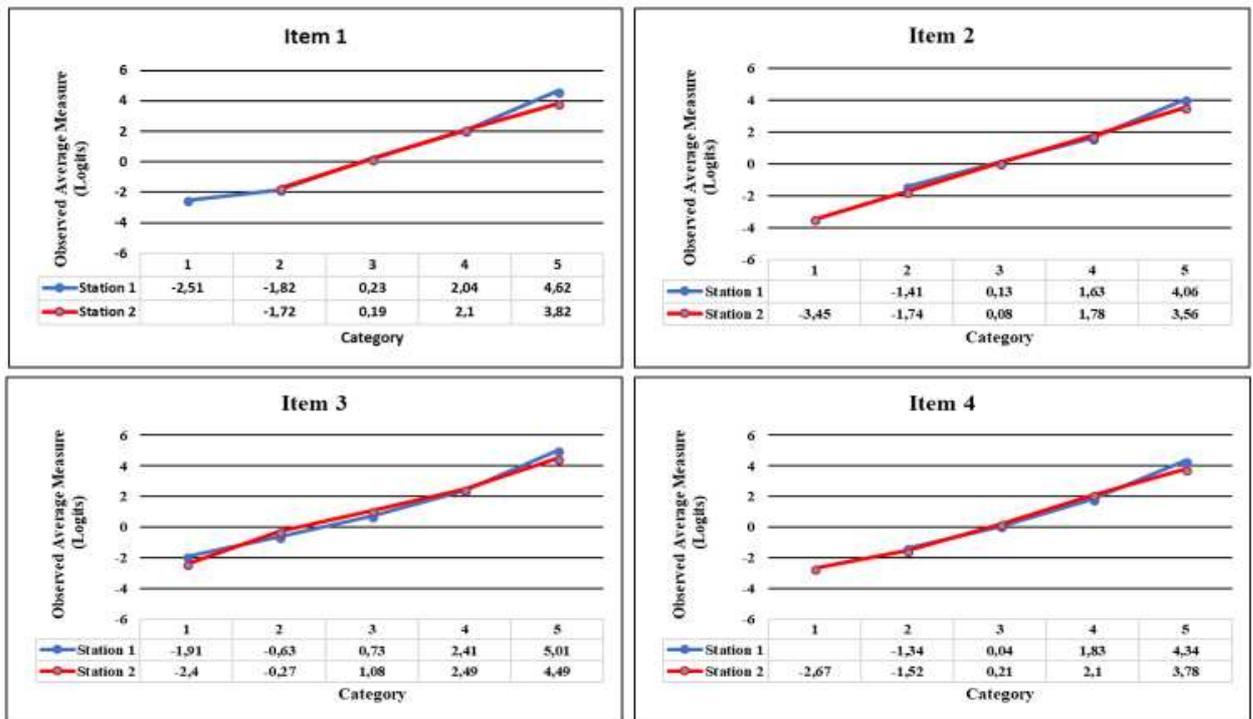


Figure 3. Fit statistics using Rasch analysis of the item response category locations according to the observed average measure from the adapted version of the Analytic Global OSCE Rating (AG-OSCE-R) at each station. The results indicate that examiners used the rating scale to assess examinees in line with the AG-OSCE-R developers' intentions (i.e., student with higher communication skill (CS) ability received higher category scores).

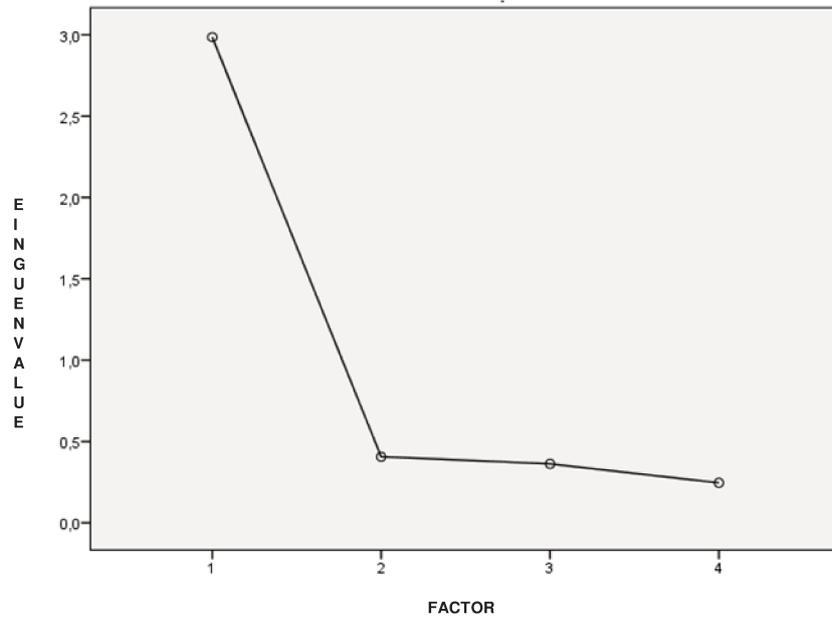


Figure 4. The PCA Scree Plot of the AG-OSCE-R

5. DISCUSSÃO

A adaptação transcultural é muito mais do que simplesmente traduzir um instrumento. Utilizar um protocolo bem estabelecido é fundamental para obter equivalência entre o instrumento original e o adaptado. Estudos mostram não haver consenso na literatura sobre a estratégia ideal para adaptação transcultural de instrumentos, e a escolha metodológica tem sido apontada como uma questão de preferência e logística(65, 77). Os protocolos utilizados para finalidade similar são bastante diferentes. A maioria recomenda verificar a validade de conteúdo, ao mesmo tempo em que verificam a adaptação transcultural – no entanto, raramente definem os passos para alcançar esse objetivo(77). Em nosso estudo, após a revisão de vários protocolos(60, 65, 67), adotamos o proposto por Beaton et al.(60), acrescido do estudo de validade de conteúdo realizado pelo comitê de especialistas(62) e da avaliação da equivalência operacional proposta por Reicheneim e Moraes(65). Essas duas etapas foram acrescentadas com o objetivo de se obter um protocolo mais completo, para que os objetivos propostos pudessem ser alcançados.

No protocolo utilizado neste estudo, a validade de conteúdo foi verificada pelo comitê de especialistas, para garantir que as versões em português contivessem apenas itens relevantes para a avaliação das habilidades de comunicação dos médicos brasileiros. Diante da diversidade cultural brasileira, foram convidados juízes de diferentes partes do país, com representantes das regiões sul, sudeste e nordeste. Entretanto, não encontramos juízes que se preenchem os critérios de inclusão nas demais regiões brasileiras. A opinião dos especialistas foi fundamental, uma vez que não existe consenso nacional sobre as HC a serem ensinadas nos cursos de medicina no Brasil. Por essa razão, os membros foram criteriosamente selecionados, sendo todos docentes experientes, especialistas em educação médica e no ensino de HC em cursos de medicina de diversas regiões do país (tabela 1). A metodologia utilizada para verificar a validade de conteúdo foi detalhadamente descrita, e o comitê de especialistas atuou de forma distinta em duas etapas: na avaliação das equivalências linguísticas e no estudo de validade de conteúdo, desempenhando papel vital nas decisões críticas para a concepção da versão final dos instrumentos.

5.1. Communication Assessment Tool – CAT

Neste estudo, optou-se por adaptar o CAT, pois não havia na literatura nenhum instrumento validado, desenvolvido ou adaptado para avaliar habilidades de comunicação de médicos brasileiros a partir das percepções do paciente. Em seu processo de adaptação transcultural, utilizamos o protocolo proposto por Beaton et al. e incluímos a etapa de equivalência operacional proposta por Reicheneim e Moraes(65) para o CAT. Os autores recomendam avaliar a equivalência operacional quando o instrumento adaptado for utilizado de modo diferente da versão original, a fim de garantir que a eficácia seja mantida. Essa etapa foi incluída porque o CAT, apesar de já ter sido utilizado em um estudo anterior em ambiente simulado(57), foi originalmente desenvolvido para uso em ambientes de prática real(52). Sendo assim, consideramos fundamental verificar a sua operacionalização durante a prova de habilidades clínicas, avaliando os aspectos de utilização nesse cenário.

O processo de adaptação transcultural do CAT foi conduzido dentro dos melhores padrões internacionais e, além das equivalências semântica, idiomática, cultural e conceitual, verificou-se também a validade de conteúdo e a equivalência operacional para o uso do CAT em ambiente simulado. Todos os itens da escala original foram mantidos na versão adaptada, observando o elevado grau de concordância entre os especialistas. Os elevados índices de validade de conteúdo corroboraram a pertinência dos itens para avaliar as habilidades interpessoais e de comunicação dos médicos brasileiros. Dessa forma, a utilização de uma metodologia criteriosa foi essencial para a produção de uma versão do CAT apropriada para uso no contexto brasileiro.

Foram realizados dois encontros com as pacientes simuladas. No primeiro deles, antes do pré-teste, elas receberam treinamento sobre o preenchimento e foi verificada a compreensão de cada item. No segundo, após o pré-teste, elas foram entrevistadas e estimuladas a fazer comentários, expressar dificuldades, emitir opiniões e sugestões sobre possíveis mudanças. O tempo preestabelecido para o preenchimento foi considerado adequado, conforme recomendado pelo desenvolvedor do instrumento (1-2 minutos)(52). Não foram identificados problemas e não foram necessárias novas adaptações. Essa fase

constatou a equivalência conceitual e a operacional da versão do CAT em ambiente simulado.

O alfa de Cronbach da versão em português do CAT, detectado nas duas estações, apresentou valores elevados e semelhantes ao valor encontrado no estudo original(52), em outros estudos americanos(54, 56) e na versão adaptada para o italiano(58). Sabe-se que os valores desse coeficiente podem ser influenciados não apenas pela consistência interna dos itens, mas também pelo número e pela forma de redação deles(68). Os valores do alfa de Cronbach aqui encontrados corroboram a confiabilidade do instrumento adaptado para o grupo de residentes que participou do estudo.

As notas dos médicos residentes mostraram valores abaixo daqueles descritos em outros estudos conduzidos tanto em ambiente de prática real(51, 54, 55) quanto simulado(57). Um aspecto importante a ser considerado é a forma de composição da pontuação no CAT. Ao utilizar apenas os itens marcados como excelente para compor a pontuação do profissional, o instrumento perde a capacidade de discriminar os examinandos na extremidade inferior do traço latente (efeito chão), ou seja, aqueles de menor habilidade de comunicação. Como a maioria dos residentes teve pontuação baixa, poderíamos considerar que os itens do CAT são muito difíceis para essa amostra de examinandos e questionar se as tarefas avaliadas são indicativas de uma boa comunicação durante a consulta médica no contexto brasileiro. A opinião dos especialistas nos permite afirmar que sim, pois todos os itens do CAT foram considerados associados às habilidades de comunicação dos médicos, e a sua presença na escala tida como relevante ou muito relevante.

Considerando ainda a baixa pontuação dos residentes na amostra estudada, outro questionamento importante refere-se à equivalência do desempenho dos médicos entre os ambientes real e simulado. Estudos demonstram que a avaliação de habilidades de comunicação por pacientes simulados tem elevada confiabilidade e maior correlação com a avaliação de estudantes de medicina e residentes em ambientes reais(49). No presente estudo, os residentes não tinham conhecimento prévio dos itens do CAT e não sabiam exatamente em que tarefas seriam avaliados, no entanto, 71,4% deles já tinham recebido treinamento em habilidades de comunicação e 92,8% participado de avaliações

de habilidades clínicas em ambiente simulado. Diante disso, pode-se considerar a dificuldade dos residentes em transpor o conhecimento para a prática real ou simulada, e desempenhar os comportamentos e atitudes percebidos como “excelentes” pelos pacientes durante o atendimento médico. Apesar da inclusão compulsória das habilidades de comunicação na graduação em medicina, seu ensino ainda não ocorre de maneira sistemática(78), o que dificulta a transposição de tais habilidades para a prática clínica(8, 9, 37). Acreditamos que a ausência de um consenso nacional sobre as competências em comunicação a serem alcançadas durante a formação médica dificulta a compreensão e construção de padrões de ensino e avaliação.

Ainda no que concerne às notas obtidas pelos residentes, considerou-se que a baixa proporção de itens classificados como “excelente” pode ter relação com o significado dessa palavra na cultura brasileira. No Brasil, o termo “excelente” é frequentemente associado a um elevado padrão de desempenho, considerado difícil de ser alcançado por profissionais ainda em formação. Considerar as notas de um a quatro como tendo um mesmo significado e somente a nota cinco ser valorada como desempenho satisfatório pode não ser adequado quando estamos utilizando o instrumento para avaliar profissionais ainda em formação e em outra cultura. Dessa forma, esse aspecto pode ser visto como uma limitação dos estudos de adaptação transcultural.

Em termos de generalização dos resultados, a etapa do pré-teste apresenta limitações, devido a um número restrito de sujeitos (28 médicos residentes), o que não possibilita avaliações psicométricas mais apuradas como a análise fatorial para verificação apropriada da dimensionalidade do instrumento adaptado. Sendo assim, serão necessários estudos com amostras maiores para confirmar esses resultados(68).

De modo geral, em nosso estudo, o CAT mostrou-se um instrumento simples e viável para avaliar as HC dos residentes a partir da percepção do paciente. Suas características permitem que seja utilizado para fornecer *feedback* sistematizado e objetivo, em benefício do profissional em treinamento, podendo auxiliar os residentes a refletirem sobre suas atitudes durante a consulta médica e identificarem áreas que necessitam de aprimoramento. Considerando-se que a avaliação de habilidades de comunicação seja bastante

complexa e que a associação entre o desempenho e os critérios utilizados apresentam maior grau de subjetividade em relação às avaliações cognitivas(42), a possibilidade de prover *feedback* estruturado, a partir de avaliações objetivas realizadas por meio de um instrumento confiável, torna ainda mais potente o uso do CAT na educação médica.

5.2. Analytic Global OSCE Rating – AG-OSCE-R

No presente estudo, optou-se por adaptar o AG-OSCE-R, pois não encontramos na literatura nenhum instrumento validado, desenvolvido ou adaptado para avaliar habilidades de comunicação de médicos brasileiros durante estações de Osce. Ele também foi escolhido por ser um instrumento simples e fácil de ser preenchido pelo avaliador durante um Osce(79). O processo de adaptação transcultural do AG-OSCE-R foi conduzido dentro dos melhores padrões internacionais. Todos os itens da escala original foram mantidos na versão adaptada, observando o elevado grau de concordância entre os especialistas. Os elevados índices de validade de conteúdo corroboraram a pertinência dos itens para avaliar as habilidades interpessoais e de comunicação dos médicos brasileiros. Sendo assim, a utilização de uma metodologia criteriosa foi essencial para a produção de uma versão do AG-OSCE-R apropriada para uso no contexto brasileiro.

Durante o pré-teste, não foram identificados problemas na utilização do AG-OSCE-R. O tempo preestabelecido para o preenchimento do instrumento (1 minuto) foi considerado adequado e de acordo com o descrito na literatura(79). Não foram necessárias novas adaptações no instrumento. Os dados do pré-teste foram utilizados para o cálculo do alfa de Cronbach, e valores elevados indicam que existe correlação entre os itens, sendo aceitáveis valores que variam de 0,7 a 0,95(68). Os valores de alfa obtidos nessa amostra (Estação 1= 0,91 e Estação 2=0,87) já indicavam que o instrumento apresentava uma elevada consistência interna, sugerindo estarmos diante de um instrumento homogêneo, ou seja, unidimensional. Esse valor de alfa é semelhante ao relatado no estudo original (alfa=0,7)(44). Essa fase constatou a equivalência conceitual e a aplicabilidade da versão do AG-OSCE-R no contexto brasileiro, garantindo a execução da fase seguinte, que compreendeu a aplicação da

versão final do instrumento a uma população maior para o estudo das propriedades psicométricas.

Para verificar as propriedades psicométricas do AG-OSCE-R, o instrumento foi aplicado durante o Osce para avaliar habilidades clínicas de 249 discentes dos três últimos anos do Curso de Medicina da FCM-Unicamp. A maioria das escolas de medicina utiliza o Osce para avaliar as HC, em que os estudantes interagem com PP e são avaliados por um observador (docente/preceptor ou pelos próprios PP)(50). Existem diferentes tipos de Osce para avaliar as HC, que variam de acordo com a finalidade. Alguns abordam a avaliação das HC de forma integrada a outras tarefas clínicas (por exemplo, coleta de história clínica, exame físico), enquanto outros se concentram exclusivamente na avaliação de HC(50). Nesse estudo, utilizaram-se duas estações em que as HC foram avaliadas de forma integrada a outras tarefas, com a utilização de dois instrumentos distintos: um *checklist* para as tarefas clínicas e o AG-OSCE-R para as HC. A escolha dessa opção considerou a logística e as possibilidades de recursos financeiros e humanos disponíveis para a execução do Osce.

Estudos demonstraram a importância de uma avaliação válida das habilidades de comunicação dos estudantes de medicina. Sabe-se que existem diferentes tipos de instrumento de avaliação de HC desenvolvidos para avaliar o desempenho dos estudantes de medicina durante estações de Osce(13, 48). No entanto, a validade das notas de desempenho está relacionada à qualidade do instrumento de avaliação utilizado durante a prova(47).

Para a análise das propriedades psicométricas da versão final do instrumento AG-OSCE-R aplicado durante as estações de Osce, utilizou-se o Modelo Multifacetado de Rasch (MMFR) por dois motivos: 1) por ser uma técnica mais moderna para estudos de propriedades psicométricas e recomendada para estudos de validação de instrumentos, com o objetivo de melhorar a qualidade das interpretações dos resultados obtidos(80); e 2) pelo fato do objeto de estudo tratar-se de um instrumento de avaliação com itens pontuados em uma escala Likert (itens politômicos) preenchidos por diferentes avaliadores (81, 82).

O modelo MMFR é uma extensão do modelo da teoria da resposta ao item

(TRI) que incorpora mais parâmetros (facetos), além das duas variáveis tradicionalmente utilizadas nas avaliações (traço latente e dificuldade do item). O MMFR estima o desempenho dos examinados considerando o grau de exigência do examinador, o grau de dificuldade da tarefa ou qualquer outro parâmetro (faceta) que afete o desempenho dos examinados(83).

As avaliações de Osce estão sujeitas a variadas fontes de erros que irão influenciar o resultado do desempenho dos examinados. As fontes de erro mais frequentes são conteúdo inapropriado do caso, desempenho impreciso dos PP e inconsistências no desempenho do avaliador(82). A análise de Rasch utiliza métodos mais complexos para produzir resultados que não apenas identificam fontes de erro de medição e falta de confiabilidade, mas também identificam a forma como a dificuldade do item interage com a capacidade do aluno.

No presente estudo, em que o instrumento de avaliação foi aplicado por avaliadores em estações do Osce, o uso do MMFR permitiu que fossem incorporadas nas análises, além da habilidade do examinando (traço latente) e da dificuldade das tarefas (itens e estações), outras variáveis como a severidade dos avaliadores e a estrutura da escala de avaliação. Dessa forma, a análise Rasch também ajuda a interpretar, de forma mais sofisticada, a relação entre o instrumento e os construtos (variáveis) que se deseja medir (validade de construto)(69).

Um instrumento é considerado unidimensional se seus itens medem um único construto ou traço latente. No entanto, a consistência interna, apesar de ser uma condição necessária, não é suficiente para verificar a homogeneidade ou unidimensionalidade de um instrumento. A literatura recomenda que, para confirmar os resultados, deve-se processar a análise fatorial dos dados(68, 84). Esse passo é fundamental para confirmar a homogeneidade do instrumento, pois o conceito de confiabilidade assume que a unidimensionalidade existe na amostra de itens que compõem o instrumento e, se essa suposição for violada, ela causa uma subestimação da confiabilidade da medida(68). Quando não existe uma hipótese prévia quanto à dimensionalidade do instrumento, como no caso da AG-OSCE-R, recomenda-se a utilização da análise de componentes principais (ACP) ou análise fatorial exploratória (AFE). Somente quando já

existe uma hipótese clara quanto à estrutura dos fatores, seja devido a um modelo teórico existente ou porque a estrutura já foi determinada anteriormente, a análise fatorial confirmatória (AFC) deve ser usada(85).

Nos estudos anteriores utilizando a AG-OSCE-R, não foi verificada a dimensionalidade do instrumento. Em nosso estudo, utilizamos uma metodologia combinada para verificar a dimensionalidade da escala. Optou-se por esse caminho porque a ACP convencional utiliza dados brutos em seu modelo matemático, e a ACP dos resíduos pelo modelo Rasch utiliza os escores lineares (reais), sendo considerada uma avaliação mais robusta por investigar também as dimensões/fatores presentes na porcentagem da variância não explicada pelo modelo(86). Os dados da ACP convencional e da análise de Rasch constataram a unidimensionalidade essencial da AG-OSCE-R, fornecendo evidências de validade baseada na estrutura interna de escala (validade de constructo). Dessa forma, esses resultados corroboram os achados da validade de conteúdo e nos permitem afirmar que a AG-OSCE-R é uma escala essencialmente unidimensional e, portanto, com elevada consistência interna, apresentando boas definições de construto com itens pertinentes e bem elaborados.

A análise de Rasch demonstrou que a versão brasileira do AG-OSCE-R apresentou boa discriminação de itens com alta correlação positiva e dividiu os estudantes em três níveis de habilidades, o que pode ser considerado adequado. A elevada confiabilidade do índice de separação dos examinandos (0,87, $p=0,00$) indica a possibilidade de reprodutibilidade desses achados em outras amostras semelhantes. Em estudos prévios do AG-OSCE-R, a confiabilidade do item foi estimada pelo alfa de Cronbach e variou de 0,52(87) a 0,70(44). Tanto o alfa de Cronbach quanto o índice de separação de confiabilidade relatado na análise de Rasch são estimativas da razão entre "variância verdadeira" e "variância observada" das medidas(72). Uma diferença importante é que a análise de Rasch usa medidas lineares, enquanto o alfa de Cronbach usa escores brutos(69). É importante ressaltar que "confiabilidade" não é um índice de qualidade, mas de reprodutibilidade relativa(88). Em outras palavras, quanto maior o coeficiente de confiabilidade, mais os itens estão correlacionados e provavelmente medem o mesmo construto(88). Nossos

resultados mostraram elevada reprodutibilidade em nossa amostra, utilizando uma moderna metodologia estatística e revelando elevada consistência interna de escala. Eles também mostraram elevada confiabilidade dos itens (0,97, $p=0,00$), indicando que tais itens avaliaram diferentes níveis de traço latente e distinguiram indivíduos com alta e baixa habilidade de comunicação. Não observamos efeitos teto ou piso nos resultados das avaliações. Estes valores elevados de confiabilidade para o AG-OSCE-R fornecem mais duas evidências de validade baseadas na estrutura interna de escala.

Para avaliar as evidências de validade baseadas em outras variáveis, realizamos um estudo de validade concorrente comparando a nota do AG-OSCE-R, a habilidade dos examinandos e a nota obtida no checklist de habilidades clínicas desenvolvido para cada estação (relação teste-critério)(89). Encontramos uma elevada correlação significativa entre a nota do AG-OSCE-R e a medida de habilidade do examinando (0,883, $p=0,000$), confirmando que os alunos com maiores habilidades em HC tiveram escores mais altos na avaliação. Ainda encontramos uma correlação positiva significativa entre a nota no AG-OSCE-R e a nota obtida no *checklist* de habilidades clínicas em ambas as estações (0,376 e 0,476, nas estações de Clínica Médica e GO, respectivamente; $p = 0,000$). Nossos resultados indicam que melhores escores de desempenho clínico estão correlacionados positivamente com boa comunicação clínica. Considerando que, na época em que realizamos o estudo, não havia nenhuma outra escala validada para avaliação de HC publicada em português, não foi possível desenvolver um estudo de validade de critério com um instrumento padrão-ouro nesta pesquisa.

A análise de Rasch também fornece um índice de separação e confiabilidade dos avaliadores. No entanto, esses dados não podem ser interpretados como estatísticas convencionais de confiabilidade interobservadores (por ex., o Kappa de Cohen). Nesta análise de Rasch, os valores de confiabilidade para avaliadores próximos de zero significam que todos os avaliadores atuaram com o mesmo grau de severidade, e os valores elevados indicam que os avaliadores atuaram de forma independente. No presente estudo, o elevado índice de confiabilidade dos avaliadores (0,96) demonstra que eles atuaram com diferentes níveis de rigor durante as observações. (os avaliadores atuam

de forma independente). No entanto, o intervalo da variação da habilidade dos avaliadores (-3,0 a 2,0) é mais restrito do que o intervalo de habilidade dos examinandos (-3,89 a 9,16), e esses dados apoiam a suposição de que os avaliadores seguiram as orientações recebidas para a atribuição de pontuação durante o processo de avaliação(81). Os resultados encontrados são semelhantes aos relatados em estudos que demonstram baixa confiabilidade entre avaliadores, mesmo quando diferentes avaliadores observam o mesmo desempenho(90-92). A baixa confiabilidade entre avaliadores é considerada um problema em relação à validade das avaliações de desempenho clínico(93). O aumento da utilização da avaliação de desempenho clínico utilizando observadores traz à tona o importante papel do observador e a variabilidade presente nos julgamentos. Vários pesquisadores questionam se essa variabilidade pode (ou precisa) ser superada(92, 94-96). O treinamento de examinadores é a recomendação para melhorar a confiabilidade da avaliação, mas o treinamento eficaz não é uma solução pacífica para esse problema(90, 97). Estudos sobre a tomada de decisão do examinador durante o processo de avaliação mostraram que essa é uma tarefa complexa, influenciada pelas impressões iniciais do examinador sobre o examinando, envolvendo a geração de hipóteses e inferências complexas sobre as competências a serem avaliadas(92, 94, 95, 98, 99). Um constante jogo de equilíbrio entre objetividade e subjetividade é observado no processo de avaliação de desempenho clínico e mais estudos são necessários para orientar a construção de um sistema de avaliação mais eficiente, que incorpore coerentemente os aspectos subjetivos da avaliação prática de desempenho clínico(92) (4, 94, 95, 100). As estações de Osce utilizadas neste estudo contavam com um avaliador e por isso não foi possível apresentar evidências adicionais comparando a concordância entre avaliadores especialistas e não especialistas em HC.

Esses resultados permitem concluir que estamos diante de dois instrumentos válidos e confiáveis para a avaliação de estudantes de medicina em ambiente simulado. Devido ao crescente interesse no tema “comunicação em saúde” e aos desafios relacionados ao ensino dessas habilidades aos profissionais brasileiros, acreditamos que ferramentas como essas podem auxiliar na promoção da prática reflexiva e nas mudanças de comportamento

profissional. Considerando-se que a avaliação de habilidades de comunicação seja bastante complexa e apresente um elevado grau de subjetividade, a utilização de instrumentos de avaliação confiáveis, que incorporem tanto a perspectiva do paciente quanto a do docente/preceptor, é de grande valia e torna ainda mais potente o seu uso na educação médica. Estudos em amostras maiores e diferentes cenários são necessários para confirmar os resultados desta pesquisa. Esperamos que este trabalho contribua para a formação profissional e a avaliação da qualidade do cuidado ofertado aos usuários dos serviços de saúde.

6. CONCLUSÃO

As versões brasileiras do Communication Assessment Tool (CAT) e do Analytic Global OSCE Rating (AG-OSCE-R) demonstraram adequada adaptação transcultural com elevado grau de concordância entre os especialistas.

As versões brasileiras do CAT e AG-OSCE-R demonstraram ter itens adequados e uteis para avaliar habilidades de comunicação (validade de conteúdo).

A versão brasileira do CAT demonstrou evidências de validade baseada na estrutura interna quando aplicado a essa amostra de residentes (validade de constructo), e isso dá suporte à introdução desse instrumento na educação médica brasileira.

A versão brasileira do AG-OSCE-R demonstrou evidências robustas de validade baseada na estrutura interna sugerindo a confiabilidade desses resultados em outros contextos (validade de constructo) sendo recomendado seu uso na avaliação de médicos brasileiros em ambiente simulado.

A versão brasileira do AG-OSCE-R demonstrou evidências de validade baseada na relação com outras variáveis (checklist de habilidades clínicas) sugerindo que melhores pontuações no desempenho clínico estão correlacionadas positivamente com boa comunicação clínica.

7. REFERÊNCIAS

1. Jadete Barbosa L, Rinaldo Henrique A-d-S, Gianna Lepre P, Regina Celes de Rosa S, Ively Guimarães A, Nilce Maria da Silva Campos C. Projeto de avaliação de tendências de mudanças no curso de graduação nas escolas médicas brasileiras. Project for the evaluation of change trends in the undergraduate course of brazilian medical schools. *Rev bras educ med*.5-18.
2. van der Vleuten CP, Schuwirth LW, Driessen EW, Govaerts MJ, Heeneman S. 12 Tips for programmatic assessment. *Med Teach*. 2014;1-6.
3. Wass V, Van der Vleuten CPM, Shatzer J, Jones R. Assessment in clinical competence. *Lancet*. 2001;357.
4. Whitehead CR, Kuper A, Hodges B, Ellaway R. Conceptual and practical challenges in the assessment of physician competencies. *Med Teach*. 2015;37(3):245-51.
5. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach*. 2011;33(3):206-14.
6. Ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Med Teach*. 2015;37(11):983-1002.
7. Pruskil S, Deis N, Druener S, Kiessling C, Philipp S, Rockenbauch K. Implementation of "social and communicative competencies" in medical education. The importance of curriculum, organisational and human resource development. *GMS Z Med Ausbild*. 2015;32(5):Doc50.
8. Laidlaw A, Hart J. Communication skills: an essential component of medical curricula. Part I: Assessment of clinical communication: AMEE Guide No. 51. *Med Teach*. 2011;33(1):6-8.

9. von Fragstein M, Silverman J, Cushing A, Quilligan S, Salisbury H, Wiskin C. UK consensus statement on the content of communication curricula in undergraduate medical education. *Med Educ*. 2008;42(11):1100-7.
10. Haskard Zolnierok KB, DiMatteo MR. Physician Communication and Patient Adherence to Treatment: A Meta-analysis. *Medical care*. 2009;47(8):826-34.
11. Levinson W, Roter DL, Mullooly JP, Dull VT, Frankel RM. Physician-patient communication: The relationship with malpractice claims among primary care physicians and surgeons. *JAMA*. 1997;277(7):553-9.
12. Sandars J, Bax N, Mayer D, Wass V, Vickers R. Educating undergraduate medical students about patient safety: Priority areas for curriculum development. *Medical Teacher*. 2007;29(1):60-1.
13. Comert M, Zill JM, Christalle E, Dirmaier J, Harter M, Scholl I. Assessing Communication Skills of Medical Students in Objective Structured Clinical Examinations (OSCE)--A Systematic Review of Rating Scales. *PLoS One*. 2016;11(3):e0152717.
14. Gillis AE, Morris MC, Ridgway PF. Communication skills assessment in the final postgraduate years to established practice: a systematic review. *Postgraduate Medical Journal*. 2014.
15. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Acad Med*. 2001;76(4):390-3.
16. Duffy FD, Gordon GH, Whelan G, Cole-Kelly K, Frankel R, Buffone N, et al. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. *Acad Med*. 2004;79(6):495-507.
17. Gontijo ED, Alvim C, Megale L, Melo JRC, Lima M. Matriz de competências essenciais para a formação e avaliação de desempenho de estudantes de Medicina. *Rev Bras Educ Médica*. 2013;37(4):526-39.

18. Accreditation Council for Graduate Medical E. Advancing education in interpersonal and communication skills: an educational resource from the ACGME Outcome Project. Accreditation Council for Graduate Medical Education: Chicago, IL, USA. 2005.
19. Rider EA, Hinrichs MM, Lown BA. A model for communication skills assessment across the undergraduate curriculum. *Med Teach*. 2006;28(5):e127-34.
20. Kiessling C, Dieterich A, Fabry G, Holzer H, Langewitz W, Muhlinghaus I, et al. Communication and social competencies in medical education in German-speaking countries: the Basel consensus statement. Results of a Delphi survey. *Patient Educ Couns*. 2010;81(2):259-66.
21. LeBlanc TW. Communication skills training in the twenty-first century. *AMA J Ethics*. 2015;17(2):140-3.
22. Drazen JM, Shields HM, Loscalzo J. A division of medical communications in an academic medical center's department of medicine. *Acad Med*. 2014;89(12):1623-9.
23. Kurtz SM, Silverman JD. The Calgary-Cambridge Referenced Observation Guides: an aid to defining the curriculum and organizing the teaching in communication training programmes. *Med Educ*. 1996;30(2):83-9.
24. Makoul G. The SEGUE Framework for teaching and assessing communication skills. *Patient Education and Counseling*. 2001;45(1):23-34.
25. von Fragstein M, Silverman J, Cushing A, Quilligan S, Salisbury H, Wiskin C, et al. UK consensus statement on the content of communication curricula in undergraduate medical education. *Medical Education*. 2008;42(11):1100-7.
26. Kiessling C, Dieterich A, Fabry G, Hölzer H, Langewitz W, Mühlinghaus I, et al. Communication and social competencies in medical education in German-speaking countries: the Basel consensus statement. Results of a Delphi survey. *Patient Educ Couns*. 2010;81(2):259-66.

27. Bachmann C, Abramovitch H, Barbu CG, Cavaco AM, Elorza RD, Haak R, et al. A European consensus on learning objectives for a core communication curriculum in health care professions. *Patient Education and Counseling*. 2013;93(1):18-26.
28. Garcia de Leonardo C, Ruiz-Moral R, Caballero F, Cavaco A, Moore P, Dupuy LP, et al. A Latin American, Portuguese and Spanish consensus on a core communication curriculum for undergraduate medical education. *BMC Med Educ*. 2016;16:99.
29. Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina., Resolução CNE/CES Nº 4, de 7/11/2001. (2001.).
30. Brasil. **Diretrizes Curriculares para o Curso de Graduação em Medicina**. In: Educação. Md, editor. Brasília, DF: Senado Federal; 2014.
31. Rossi PS, Batista NA. O ensino da comunicação na graduação em medicina: uma abordagem. *Interface - Comunicação, Saúde, Educação*. 2006;10:93-102.
32. Liberali R, Novack D, Duke P, Grosseman S. Communication skills teaching in Brazilian medical schools: What lessons can be learned? *Patient Educ Couns*. 2018;101(8):1496-9.
33. Lampert JB. Habilidades de Comunicação e Qualidade de Vida/Saúde. *Cadernos da ABEM*. 2014;10:77.
34. Grosseman S, Loures LFM, Mariussi AP, Grossman E, Muraguchi EMe, Ogatta. Projeto Ensino de Habilidades de Comunicação na Área da Saúde: uma Trajetória Inicial. *Cadernos da ABEM*. 2014;10:6.
35. Dohms M, Tesser CD, Grosseman S. Potencialidades no ensino-aprendizagem da comunicação médico-paciente em três escolas brasileira, espanhola e holandesa. *Revista Brasileira de Educação Médica*. 2013;37:311-9.

36. Kidd J, Patel V, Peile E, Carter Y. Clinical and communication skills: Need to be learnt side by side. *BMJ : British Medical Journal*. 2005;330(7488):374-5.
37. Stewart MA. Effective physician-patient communication and health outcomes: a review. *CMAJ: Canadian Medical Association Journal*. 1995;152(9):1423-33.
38. Kurtz S, Silverman J, Benson J, Draper J. Marrying content and process in clinical method teaching: enhancing the Calgary-Cambridge guides. *Acad Med*. 2003;78(8):802-9.
39. Nunes SOV, Vargas HO, Liboni M, Martins Neto D, Vargas LHM, Turini B. O ensino de psiquiatria, habilidades de comunicação e atitudes no currículo integrado do curso de Medicina da Universidade Estadual de Londrina. *Revista Brasileira de Educação Médica*. 2008;32:210-6.
40. Van Der Vleuten CP. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ Theory Pract*. 1996;1(1):41-67.
41. Schirmer JM, Mauksch L, Lang F, Marvel MK, Zoppi K, Epstein RM, et al. Assessing communication competence: a review of current tools. *Fam Med*. 2005;37(3):184-92.
42. Hulsman RL. The art of assessment of medical communication skills. *Patient Educ Couns*. 2011;83(2):143-4.
43. Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G. Evaluating communication skills in the OSCE format: reliability and generalizability. *Med Educ*. 1996;30(1):38-43.
44. Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. *Med Educ*. 2003;37(11):1012-6.

45. Troncon LEdA. Significance of experts' overall ratings for medical student competence in relation to history-taking. Sao Paulo Medical Journal. 2006;124:101-4.
46. Zill JM, Christalle E, Müller E, Härter M, Dirmaier J, Scholl I. Measurement of physician-patient communication--a systematic review. PLoS One. 2014;9(12):e112637.
47. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. Med Teach. 2010;32(10):802-11.
48. Setyonugroho W, Kennedy KM, Kropmans TJ. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. Patient Educ Couns. 2015.
49. Swanson DB, van der Vleuten CP. Assessment of clinical skills with standardized patients: state of the art revisited. Teach Learn Med. 2013;25 Suppl 1:S17-25.
50. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. Med Teach. 2013;35(9):e1447-63.
51. Myerholtz L, Simons L, Felix S, Nguyen T, Brennan J, Rivera-Tovar A, et al. Using the communication assessment tool in family medicine residency programs. Fam Med. 2010;42(8):567-73.
52. Makoul G, Krupat E, Chang CH. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. Patient Educ Couns. 2007;67(3):333-42.
53. Swing SR, Clyman SG, Holmboe ES, Williams RG. Advancing resident assessment in graduate medical education. J Grad Med Educ. 2009;1(2):278-86.

54. Stausmire JM, Cashen CP, Myerholtz L, Buderer N. Measuring general surgery residents' communication skills from the patient's perspective using the Communication Assessment Tool (CAT). *J Surg Educ.* 2015;72(1):108-16.
55. Myerholtz L. Assessing Family Medicine Residents' Communication Skills From the Patient's Perspective: Evaluating the Communication Assessment Tool. *J Grad Med Educ.* 2014;6(3):495-500.
56. Ferranti DE, Makoul G, Forth VE, Rauworth J, Lee J, Williams MV. Assessing patient perceptions of hospitalist communication skills using the Communication Assessment Tool (CAT). *J Hosp Med.* 2010;5(9):522-7.
57. Trickey AW, Newcomb AB, Porrey M, Wright J, Bayless J, Piscitani F, et al. Assessment of Surgery Residents' Interpersonal Communication Skills: Validation Evidence for the Communication Assessment Tool in a Simulation Environment. *J Surg Educ.* 2016.
58. Scala D, Menditto E, Armellino MF, Manguso F, Monetti VM, Orlando V, et al. Italian translation and cultural adaptation of the communication assessment tool in an outpatient surgical clinic. *BMC Health Serv Res.* 2016;16:163.
59. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med.* 1999;74(10):1129-34.
60. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25(24):3186-91.
61. Barlett JE, Kotrlik JW, Higgins CC. Organizational research: Determining appropriate sample size in survey research. *Information technology, learning, and performance journal.* 2001;19(1):43.
62. Sireci S, Faulkner-Bond M. Validity evidence based on test content. *Psicothema.* 2014;26(1):100-7.

63. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007;30(4):459-67.
64. Lynn MR. Determination and quantification of content validity. *Nurs Res*. 1986;35(6):382-5.
65. Reichenheim ME, Moraes CL. [Operationalizing the cross-cultural adaptation of epidemiological measurement instruments]. *Rev Saude Publica*. 2007;41(4):665-73.
66. Padilla JL, Benitez I. Validity evidence based on response processes. *Psicothema*. 2014;26(1):136-44.
67. Alexandre NMC, Coluci MZO. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Ciência & Saúde Coletiva*. 2011;16:3061-8.
68. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53-5.
69. Boone WJ, Staver JR, Yale MS. *Rasch analysis in the human sciences*: Springer; 2013.
70. Linacre JM. Facets computer program for many-facet Rasch measurement, version 3.71. 4. Beaverton, Oregon: Winsteps com. 2013.
71. Raïche G. Critical eigenvalue sizes in standardized residual principal components analysis. *Rasch measurement transactions*. 2005;19(1):1012.
72. Clauser B, Linacre JM. Relating Cronbach and Rasch reliabilities. *Rasch Measurement Transactions*. 1999;13(2):696.
73. Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. *Psychological bulletin*. 1986;99(3):432.

74. Slocum-Gori SL, Zumbo BD. Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research*. 2011;102(3):443-61.
75. Linacre JM. Winsteps (Version 4.0.0)[Computer software]. Chicago, IL: winsteps com. 2017.
76. Tennant A, Pallant JF. Unidimensionality matters!(A tale of two Smiths?). *Rasch measurement transactions*. 2006;20(1):1048-51.
77. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol*. 2015;68(4):435-41.
78. Marcela D, Charles Dalcanale T, Suely G. Potencialidades no ensino-aprendizagem da comunicação médico-paciente em três escolas brasileira, espanhola e holandesa
Potentialities in the teaching-learning of physician-patient communication in brazilian, spanish and dutch medical schools. *Rev bras educ med*.311-9.
79. Scheffer S, Muehlinghaus I, Froehmel A, Ortwein H. Assessing students' communication skills: validation of a global rating. *Adv Health Sci Educ Theory Pract*. 2008;13(5):583-92.
80. Boone WJ. Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sci Educ*. 2016;15(4).
81. Toffoli SFL, de Andrade DF, Bornia AC. Evaluation of open items using the many-facet Rasch model. *Journal of Applied Statistics*. 2016;43(2):299-316.
82. Iramaneerat C, Yudkowsky R, Myford CM, Downing SM. Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. *Adv Health Sci Educ Theory Pract*. 2008;13(4):479-93.
83. Tavakol M, Dennick R. Psychometric evaluation of a knowledge based examination using Rasch analysis: an illustrative guide: AMEE guide no. 72. *Med Teach*. 2013;35(1):e838-48.

84. Reichenheim ME, Hökerberg YH, Moraes CL. Assessing construct structural validity of epidemiological measurement tools: a seven-step roadmap. *Cad Saude Publica*. 2014;30(5):927-39.
85. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
86. Green KE. Dimensional analyses of complex data. *Structural Equation Modeling: A Multidisciplinary Journal*. 1996;3(1):50-61.
87. Mortsiefer A, Immecke J, Rotthoff T, Karger A, Schmelzer R, Raski B, et al. Summative assessment of undergraduates' communication competence in challenging doctor-patient encounters. Evaluation of the Düsseldorf CoMeD-OSCE. *Patient Education and Counseling*. 95(3):348-55.
88. Rios J, Wells C. Validity evidence based on internal structure. *Psicothema*. 2014;26(1):108-16.
89. American Educational Research A, American Psychological A, National Council on Measurement in E. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 2014.
90. Downing SM. Threats to the validity of clinical teaching assessments: what about rater error? *Med Educ*. 2005;39(4):353-5.
91. Beckman TJ, Cook DA, Mandrekar JN. What is the Validity Evidence for Assessments of Clinical Teaching? *Journal of General Internal Medicine*. 2005;20(12):1159-64.
92. Gingerich A, Regehr G, Eva KW. Rater-based assessments as social judgments: rethinking the etiology of rater errors. *Academic Medicine*. 2011;86(10):S1-S7.
93. McGill DA, Van der Vleuten CPM, Clarke MJ. Supervisor assessment of clinical and professional competence of medical trainees: a reliability study

using workplace data and a focused analytical literature review. *Advances in health sciences education*. 2011;16(3):405-25.

94. Gauthier G, St-Onge C, Tavares W. Rater cognition: review and integration of research findings. *Medical Education*. 2016;50(5):511-22.
95. St-Onge C, Chamberland M, Levesque A, Varpio L. Expectations, observations, and the cognitive processes that bind them: expert assessment of examinee performance. *Adv Health Sci Educ Theory Pract*. 2016;21(3):627-42.
96. Avan BI, Raza SA, Kirkwood BR. An epidemiological study of urban and rural children in Pakistan: examining the relationship between delayed psychomotor development, low birth weight and postnatal growth failure. *Trans R Soc Trop Med Hyg*. 2015;109(3):189-96.
97. Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med*. 2009;24(1):74-9.
98. St-Onge C, Chamberland M, Levesque A, Varpio L. The role of the assessor: exploring the clinical supervisor's skill set. *Clin Teach*. 2014;11(3):209-13.
99. Yaphe J, Street S. How do examiners decide?: a qualitative study of the process of decision making in the oral examination component of the MRCGP examination. *Medical education*. 2003;37(9):764-71.
100. Hodges B. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach*. 2013;35(7):564-8.

APÊNDICES

APÊNDICE A – Artigo de Revisão

Avaliação de habilidades de comunicação em ambiente simulado na formação médica: conceitos, desafios e possibilidades

Autores:

1º) Sheyla Ribeiro Rocha: Professora Assistente, Departamento de Medicina, Centro de Ciências Biológicas e da Saúde, Universidade Federal de São Carlos. Programa de Pós-Graduação em Clínica Médica, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brasil.

2º) Gustavo Salata Romão: Professor Titular, Curso de Medicina, Universidade de Ribeirão Preto, Brasil.

3º) Maria Sílvia Velutini Setúbal: Pesquisadora post-doc do Departamento de Ginecologia e Obstetrícia, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brasil.

4º) Carlos Fernando Collares: Professor Assistente, Departamento de Desenvolvimento e Pesquisa Educacional, Escola de Educação de Profissões da Saúde, Faculdade de Saúde, Medicina e Ciências da Vida, Universidade de Maastricht, Maastricht, Holanda.

5º) Eliana Amaral: Professora Titular, Departamento de Tocoginecologia, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brasil.

Resumo

Nas últimas décadas, a sociedade vem demonstrando mudanças em suas expectativas quanto ao desempenho do médico, solicitando mais cuidado e atenção desses profissionais – o que tem influenciado o processo de formação. Sensíveis a esses anseios, os profissionais reconhecem a necessidade de transição do paradigma atual da formação médica em busca de uma abordagem mais integral no cuidado à saúde. No Brasil, as diretrizes curriculares para cursos de graduação na área da saúde incluem a comunicação como uma das competências a serem alcançadas. As Diretrizes

Curriculares Nacionais para Graduação em Medicina estabelecem que o egresso deve ser preparado para se comunicar por meio de linguagem verbal e não verbal, com empatia, sensibilidade e interesse, promovendo o cuidado centrado na pessoa, e uma relação horizontal e compartilhada com o paciente. O desenvolvimento dessa competência exige que o processo de formação dissemine novos conceitos e promova a aquisição de habilidades, incluindo trabalho em equipe, capacidade de liderança, habilidades de relação interpessoal e de comunicação, gestão dos serviços de saúde, prática baseada na comunidade e nos sistemas de saúde. Essas novas demandas se refletem no desenho curricular, visto que o desenvolvimento dessas competências é requerido durante a graduação em medicina. Este artigo discute o papel das habilidades de comunicação na formação médica e aspectos práticos para a implementação de uma avaliação estruturada em ambiente simulado.

Introdução:

Desde 2001 as Diretrizes Curriculares Nacionais para Graduação em Medicina (DCNGM) definem que as habilidades de comunicação devem ser ensinadas aos médicos brasileiros para que estes sejam capazes de se comunicar efetivamente com a equipe de trabalho, os pacientes, as famílias e a comunidade¹. Em 2014, as DCNGM foram revisadas e definiram que o médico deverá ser preparado para o cuidado centrado na pessoa, desenvolvendo relação horizontal e compartilhada, condição intrinsecamente ligada à habilidade para se comunicar. Definiram, ainda, que o estudante deverá ser capaz de se comunicar por meio de linguagem verbal e não verbal, com empatia, sensibilidade e interesse². Em 2017, o Conselho Nacional de Saúde (CNS) aprovou a Resolução 569/2017 que expressa diretrizes comuns para os cursos de graduação em saúde considerando as competências necessárias para a formação profissional. Dentre outros aspectos, destacou a importância das competências em comunicação e relações interpessoais para um cuidado em saúde seguro e de qualidade.

Na comunidade acadêmica internacional, a comunicação eficaz entre profissionais de saúde, pacientes e seus familiares é reconhecida como condição indispensável para a qualidade dos cuidados em saúde³. Estudos revelam que a melhora na qualidade da comunicação médico-paciente está diretamente relacionada a melhores desfechos⁴, maior adesão ao tratamento⁵, maior satisfação dos usuários com redução de reclamações por má-prática⁶ e maior segurança do paciente com redução de erros⁷.

Diante disso, percebe-se a importância de que haja a intencionalidade em promover o desenvolvimento desta competência durante a formação médica⁸. Consensos mundiais recomendam que o ensino das habilidades de comunicação se estenda a todos os níveis da

formação, passando pelo internato e programas de residência^{9 10}. Recomendam ainda que sejam instituídos processos que permitam repetidas oportunidades de avaliação e *feedback*.^{3; 10}.

Este artigo trata-se de uma revisão narrativa sobre as habilidades de comunicação (HC) na formação médica e avaliação em ambiente simulado, apresentando suas características em termos de validade, confiabilidade, impacto educacional, viabilidade e aceitabilidade. Também aborda aspectos práticos para a implementação de uma avaliação estruturada, baseada nas melhores evidências.

1. A habilidade interpessoal e de comunicação em educação médica

As habilidades de comunicação em educação médica podem ser definidas como tarefas e/ou comportamentos observáveis durante a entrevista: coleta de dados, explicação ao paciente sobre seu diagnóstico, prognóstico e proposta terapêutica, discussão do plano de cuidados e oferta das informações necessárias sobre procedimentos, além da obtenção do consentimento para a realização deles¹¹. Estudos demonstram que a melhora na qualidade da comunicação durante a entrevista clínica está diretamente relacionada a aumento na qualidade da atenção em saúde, maior satisfação de médicos e pacientes, melhor prognóstico físico e psicológico do paciente, melhor adesão ao tratamento, redução de reclamações por má-prática, maior segurança do paciente e redução de erros¹².

Apesar da crescente valorização, ainda não há discordância sobre o que seria uma comunicação médico-paciente adequada^{13; 14}. O Consenso de Kalamazoo¹¹, baseando-se em cinco modelos teóricos de comunicação médico-paciente, determinou sete componentes-chaves que caracterizam uma comunicação adequada no contexto clínico: 1) construção de relacionamento; 2) abertura da consulta; 3) coleta de informações; 4) inclusão da perspectiva do paciente; 5) compartilhamento de decisões; 6) formulação de um acordo entre médico e paciente e 7) encerramento da consulta.

As habilidades interpessoais estão diretamente relacionadas à qualidade da comunicação realizada durante a consulta e dependem do desenvolvimento de habilidades, como a empatia, cordialidade, linguagem não verbal e tonalidade da voz. As habilidades interpessoais podem ser percebidas pelo efeito da comunicação no interlocutor e, juntamente às habilidades de comunicação, criam e sustentam, de forma combinada, a relação médico-paciente¹⁵.

Dessa forma, podem-se definir as HC como sendo a capacidade de o profissional se relacionar com os pacientes, suas famílias e outros profissionais, o que resulta em uma comunicação médico-paciente efetiva e eficaz. Espera-se, portanto, que os médicos que apresentem essas habilidades bem desenvolvidas sejam capazes de: 1) criar e manter uma relação terapêutica e ética sólida com os pacientes; 2) utilizar a escuta ativa para obter informações necessárias, incluindo habilidades verbais (explicativas, questionadoras, orientadoras) e não verbais, além de registrar por escrito as informações coletadas (linguagem escrita) e 3) trabalhar harmoniosamente com outros profissionais, seja como membro ou líder de uma equipe de saúde¹⁶.

2. **Qualidade de avaliação do estudante**

Os processos de avaliação em educação médica desenvolveram-se muito nas últimas décadas, com foco na avaliação de habilidades clínicas e no profissionalismo¹⁷. Os novos métodos incluem avaliação de desempenho com atividades desenvolvidas em ambientes simulados, com a participação de pacientes padronizados (PP) e em cenários reais da prática profissional. Em ambos os casos, o *feedback* ofertado ao estudante ao final da atividade é entendido como importante ferramenta catalizadora do aprendizado¹⁷.

A escolha do método de avaliação deve considerar critérios de qualidade e eficiência. Segundo Norcini & McKinley (2007), deve-se levar em consideração cinco fatores: validade, confiabilidade, efeitos educacionais do teste, viabilidade e aceitação pelo público-alvo. Se o teste não preencher um dos cinco quesitos, ainda que seja válido e confiável, não deve ser considerado um método adequado para uso naquela realidade. Recomenda-se que as atividades de avaliação sejam programáticas, isto é, sejam parte do programa educacional ou do currículo^{17; 18; 19}. Deve haver uma mudança de foco do processo de avaliação, sendo desejável que se concentre na avaliação “para o aprendizado”, e não somente na avaliação “do aprendizado”¹⁹. A avaliação realizada em diferentes momentos, por diferentes métodos e múltiplos avaliadores, permite maior segurança nas decisões de progressão ou retenção dos estudantes¹⁹.

Uma estratégia importante para potencializar o aprendizado é o uso do *feedback*. No meio médico-educacional, *feedback* pode ser definido como sendo a apreciação objetiva, informada e não avaliativa do desempenho de um profissional, com o propósito de aperfeiçoar suas habilidades clínicas²⁰. Evidências mostram que a observação do estudante enquanto realiza uma atividade prática associada a um

feedback apreciativo pode afetar positivamente o seu desenvolvimento e a qualidade do atendimento prestado ao paciente²¹. Nesse contexto, o momento do *feedback* é uma oportunidade para corrigir erros e deficiências, evitando que falhas não percebidas se consolidem como maus hábitos. Além disso, é importante para reforçar comportamentos positivos, evitando que a não valorização favoreça a descontinuidade de comportamentos desejáveis.

Na avaliação de habilidades de comunicação (HC), especialistas recomendam que seja instituída uma matriz que permita repetidas oportunidades de avaliação e *feedback*, reforçando o uso das habilidades mais básicas de entrevista até as mais complexas, como a comunicação de más notícias^{11; 22}. Dessa forma, para uma avaliação eficiente, são necessários métodos e instrumentos com sólido embasamento teórico e propriedades psicométricas reconhecidas^{9; 10; 22; 23}. Assim, as avaliações em ambiente simulado com PP têm sido amplamente utilizadas para HC^{13; 24; 25; 26; 27; 28}.

3. Avaliação de HC em ambiente simulado do tipo Osce

Na década de 70, Harden e colaboradores²⁹ desenvolveram um método de avaliação de habilidades médicas com o objetivo de padronizar e reduzir o número de variáveis que poderiam afetar o desempenho dos estudantes durante as provas práticas²⁹. Esse método ficou mundialmente conhecido como “OSCE”, sigla em inglês para *Objective Structured Clinical Examination* (Exame Clínico Objetivo Estruturado). E assim também se popularizou no Brasil, sendo amplamente utilizado na literatura científica nacional^{30; 31; 32; 33; 34; 35}. Dessa forma, utilizaremos o anglicismo Osce para nos referirmos a esse tipo de avaliação.

Osce é um exame baseado no desempenho, que permite a avaliação estruturada de habilidades clínicas, incluindo as HC, tanto em nível de graduação quanto pós-graduação^{13; 26}, bastante utilizado inclusive no Brasil^{30; 31; 33; 36; 37; 38}. Durante o exame, os estudantes realizam tarefas em um cenário simulado, denominado estação, e são avaliados por examinadores treinados. Tipicamente, os estudantes percorrem um determinado número de estações interagindo com pacientes padronizados (PP). Manequins de baixa e alta fidelidade também podem ser utilizados, bem como estações híbridas que combinam PP com manequins. Essas últimas são úteis para avaliar o desempenho em procedimentos clínico-cirúrgicos ou situações de emergência¹⁷. A avaliação de desempenho clínico inclui, entre outras ações, a coleta de história clínica,

realização de exame físico, comunicação, relação interpessoal e realização de procedimentos.

O processo de elaboração de um Osce deve seguir as recomendações gerais para esse tipo de avaliação^{19; 39; 40; 41; 42}. As tarefas escolhidas devem se basear nos objetivos de aprendizagem do curso, no nível de aprendizagem e no desempenho esperado dos estudantes. Pode ter caráter formativo ou somativo.

Considerando que o desenvolvimento e a realização de um exame tipo Osce é um empreendimento laborioso e de custo elevado, seu uso deve ser criterioso e reservado para avaliar habilidades que não podem ser verificadas por outros tipos de teste. Uma estação de Osce se limita a uma tarefa realizada entre 5 e 10 minutos, utilizando-se de 5 a 20 estações por teste⁴³. Podem ser criadas estações para avaliação exclusiva das HC ou para avaliar outras habilidades clínicas simultaneamente. Nas estações em que forem avaliadas as habilidades de comunicação, deve-se definir com clareza a situação que poderá desencadear emoções complexas e exigirá maior domínio da comunicação por parte do estudante^{44; 45}.

Membros do corpo docente ou preceptores nos serviços de saúde podem atuar como avaliadores no Osce⁴⁰. Recomendam-se cursos de qualificação para esses examinadores. O objetivo é capacitá-los quanto aos princípios do Osce, o papel do examinador, a padronização na atribuição de notas ao estudante e as regras de *feedback* e de confidencialidade do exame⁴⁰. Inconsistências entre os avaliadores reduzem a confiabilidade do exame.⁴⁶

A seleção e capacitação dos PP constituem etapa fundamental para a padronização do Osce. Podem-se utilizar voluntários ou atores profissionais, e o tipo de paciente para cada estação dependerá das especificações do caso clínico e da tarefa. Em estações de avaliação da HC, recomenda-se que os PP sejam atores profissionais treinados tanto para desempenhar de forma consistente o papel proposto quanto para atuar como avaliadores e ofertar *feedback* ao final das estações^{27; 40; 42; 44; 45; 47}. No quadro 3, apresentamos exemplos de estações de Osce para avaliação de HC.

Uma estação bem estruturada de Osce, projetada para avaliar uma tarefa específica e baseada nos conteúdos curriculares, possui um elevado nível de validade e alto grau de confiabilidade^{39; 40}. Os conceitos de validade e confiabilidade, impacto educacional, viabilidade e aceitação da avaliação são detalhados a seguir.

3.1. Validade e Confiabilidade

Verificar a validade e a confiabilidade de um Osce tem por objetivo garantir que padrões mínimos de qualidade sejam alcançados, legitimando o conteúdo do exame, a pontuação dos alunos e a interpretação dos resultados⁴⁸. Devido ao seu impacto na vida acadêmica, são desejáveis índices elevados de confiabilidade e validade no Osce de alto-impacto (que embasa decisões de aprovação/reprovação). Nos exames de caráter formativo, são aceitáveis níveis mais baixos de confiabilidade^{39; 40; 49; 50}.

Na atualidade, a “validade” é compreendida como um conceito único que se refere ao grau em que as evidências dão suporte à interpretação dos resultados de um teste para o uso proposto. Nesse conceito, diferentes tipos de evidências podem ser utilizados na verificação da validade de um teste, e a confiabilidade passa a ser vista como parte do processo de obtenção de validade⁵⁰.

A validade de conteúdo é fundamental na elaboração do Osce⁴⁹, pois identifica se as questões do teste são relevantes e representativas dos conteúdos (constructos) a serem avaliados. Isso requer que um grupo de especialistas analise (ou crie) as estações, a fim de assegurar que elas refletem adequada e completamente o domínio que se pretende avaliar. A elaboração de uma matriz de avaliação é indispensável para se definir os parâmetros do teste e orientar os especialistas nessa análise⁵¹. A validade de um teste, portanto, depende diretamente dos padrões de qualidade em que é elaborado^{39; 40}.

A confiabilidade ou precisão refere-se à consistência dos resultados ao longo da aplicação do teste e pode ser estimada de diversas formas: coeficiente de confiabilidade, coeficiente de generalizabilidade, teoria de resposta aos itens (TRI), entre outros⁵⁰. No Osce, a confiabilidade está fortemente relacionada ao número de estações, ao tempo total de duração da prova e à diversidade de tipos de estações utilizadas²⁷, além da padronização do desempenho dos pacientes simulados, dos avaliadores e dos critérios de avaliação³⁹. Um número maior de estações permite uma melhor abordagem do conteúdo a ser avaliado, aumentando, portanto, a validade e confiabilidade⁴¹. A atribuição adequada e realista do tempo para a execução das tarefas, em cada uma das estações, aumenta a validade do teste³⁹. Portanto, para se generalizar os resultados da avaliação e considerar que o estudante tem desempenho satisfatório, são necessárias repetidas avaliações em contextos padronizados e diferentes⁴⁰. Dessa forma, a validade e confiabilidade do Osce serão influenciadas pelos seguintes fatores^{27; 39; 52}:

a) Matriz de Avaliação – O primeiro passo é desenvolver uma matriz do conteúdo curricular e, a seguir, definir peso e relevância de cada conteúdo diante dos objetivos de aprendizagem. O ideal é que essa etapa seja desenvolvida por um grupo de

docentes envolvidos na gestão do curso⁵¹. O próximo passo é definir a matriz de avaliação com as habilidades a serem avaliadas e o número de estações para cada conteúdo, a fim de garantir que o exame terá uma amostra representativa dos conteúdos curriculares a serem avaliados⁴⁹. A matriz de avaliação pode ser construída em um modelo bidimensional no qual um eixo representa as habilidades gerais avaliadas (por exemplo, comunicação, exame físico ou procedimentos) e o outro representa o contexto em que elas serão demonstradas. As estações do Osce serão criadas a partir da definição dessa matriz⁴⁰. Idealmente, a matriz de avaliação deve ser desenvolvida em conjunto por todos os envolvidos na elaboração do Osce⁴⁹. Um exemplo de matriz de avaliação é apresentado no quadro 1.

b) Número de estações e tempo de duração da prova – Esse é um quesito extremamente importante, pois a confiabilidade e a validade do exame serão influenciadas pelo número de estações e pela duração total da prova. Ao implementar um Osce, esses parâmetros devem estar definidos. Há relatos de Osce compostos de 5 até 20 estações, com média de duração de 5 a 10 minutos cada⁴³. Acredita-se que uma confiabilidade adequada poderia ser alcançada com 14-18 estações, em uma prova com duração total de 3 a 4 horas⁴⁰. De forma prática, é necessário equilibrar a qualidade do exame com viabilidade e recursos.

c) Instrumentos de avaliação – Durante as estações de Osce, vários examinadores (geralmente um por estação) avaliam o desempenho dos estudantes por meio de listas de verificação (*checklist*) e/ou escalas de avaliação (*rating scale*) ou uma combinação de ambas²⁶. Recomenda-se que uma lista de verificação (*checklist*) tenha de 8 a 25 itens⁵³ objetivamente relacionados à tarefa proposta. Cada item deve conter apenas um conceito, evitando a avaliação de tarefas combinadas no mesmo item⁵³. A pontuação deve ser decidida por consenso entre especialistas, podendo cada item ser valorado de forma igual ou diferente⁵³. São geralmente classificadas de forma dicotômica, como “fez” ou “não fez”⁵³. Os itens na lista visam identificar se a ação ocorreu e parecem ser mais eficientes ao avaliar estudantes em fases iniciais de formação ou para habilidades procedimentais^{24; 26}. As listas de verificação podem ter um efeito negativo no aprendizado na medida em que valorizam ações predefinidas. Sugerem que os estudantes podem memorizar as ações que serão pontuadas e ter um desempenho mecanizado e desconectado com o desempenho em ambientes da prática real²⁷.

As escalas de avaliação global são formuladas a partir de itens mais gerais e requerem um julgamento do avaliador em relação à realização da tarefa¹³. O avaliador deve decidir se a tarefa foi ou não realizada e com que grau de qualidade, exigindo, portanto, mais que uma simples observação do desempenho⁴⁰. As escalas exigem mais tempo para serem preenchidas e refletem melhor o desempenho do candidato quando preenchidas ao final da estação. O número de itens, portanto, deve ser reduzido, em relação às listas de verificação, podendo variar de um a nove⁵³. Estas devem utilizar escalas do tipo Likert, em que cada item possui pontuação variando de um a cinco pontos. As escalas tendem a valorizar mais a integração entre o conhecimento e o desempenho e são consideradas mais adequadas à avaliação de estudantes com níveis mais elevados de habilidades, residentes e profissionais mais experientes^{13; 24; 25}.

Estudos têm avaliado o uso das listas de verificação em contraponto com as escalas de avaliação, não se chegando a um consenso sobre qual seria a melhor escolha^{13; 26; 27; 28}. Escolher o melhor instrumento depende do propósito da avaliação. Em situações nas quais é exigido um maior domínio de habilidade técnica, a lista de verificação parece ser mais apropriada, tanto nos níveis iniciais (manobras do exame físico) quanto nos níveis mais avançados (manobras de ressuscitação e intubação)²⁷. As escalas de avaliação global parecem avaliar melhor os estudantes durante tarefas que exigem a integração de conhecimentos de maior complexidade, visando a uma abordagem mais holística do paciente (raciocínio clínico, habilidade de comunicação, empatia, profissionalismo)^{27; 40}. Pesquisas sugerem que as escalas de avaliação apresentam maior confiabilidade quando comparadas a listas de verificação e que usar uma combinação de escala de avaliação com a lista de verificação melhora a validade do teste²⁶.

No caso de utilização de instrumentos com a combinação dos tipos (lista e escala), a pontuação deverá ser calculada de acordo com o peso dado a cada item. Para avaliação de HC, sugere-se que seja estabelecido um instrumento padrão a ser utilizado em todas as estações, permitindo que o desempenho dos estudantes possa ser comparado ao longo de todo teste⁵⁴. Recomenda-se que sejam utilizados instrumentos submetidos a estudos de evidências de validade^{13; 26; 27; 28}. Deve-se ter cuidado ao elaborar os instrumentos de avaliação, para não estimular a memorização dos itens, levando à banalização e fragmentação do aprendizado, nem a artificialização do desempenho durante o teste²⁷.

d) Quantidade, tipo e treinamento de examinadores

As estações de Osce geralmente contam com a presença de, pelo menos, um examinador por estação. A presença de diversos examinadores reduz o viés de avaliador e aumenta a confiabilidade do teste³⁹. Estudos sugerem que a presença de dois avaliadores por estação parece aumentar a confiabilidade⁵⁵ – no entanto, a elevação do custo pode inviabilizar essa opção.

Estações de Osce podem utilizar avaliador que seja médico, profissional de outra área da saúde e não especialista⁴⁰. Na literatura, são encontradas três categorias de examinadores não especialistas: PP que avaliam o desempenho do estudante após finalizada a estação, examinadores leigos e estudantes de medicina treinados²⁷. A maioria dos estudos revela que todos esses avaliadores apresentam confiabilidade semelhante a examinadores que são médicos²⁷. No entanto, avaliadores leigos tendem a ser mais indulgentes quando comparados aos que são médicos²⁷. O tipo de avaliador guarda relação com o tipo de habilidade avaliada. Médicos especialistas estão mais indicados para a avaliação de habilidades em procedimentos, manobras do exame físico e organização da entrevista clínica⁴⁰. PP são considerados bons examinadores para HC e domínio do idioma, mas podem ter dificuldades para avaliar a organização da estrutura da entrevista, mesmo que recebam treinamento adequado²⁷. Com relação aos instrumentos de avaliação, avaliadores que são médicos têm melhor desempenho quando utilizam escalas de avaliação durante a observação de desempenho clínico. Avaliadores leigos podem utilizar qualquer um dos tipos de instrumento para avaliar os estudantes²⁷.

A presença de examinadores treinados e com experiência relevante tem papel fundamental na confiabilidade do Osce, garantindo que todos os estudantes sejam avaliados com os mesmos critérios³⁹. Portanto, o treinamento adequado dos examinadores tende a reduzir a variação entre eles e melhorar o desempenho, elevando a confiabilidade (precisão) dos resultados do teste⁴⁰. Por outro lado, inconsistências entre os avaliadores reduzem a confiabilidade do Osce⁴⁶. Baseando-se nessas considerações, os organizadores da prova podem decidir que tipos de avaliador são mais efetivos para quais tipos de estação/habilidades.

e) Padronização dos pacientes simulados

A padronização da atuação dos pacientes simulados é fundamental para garantir a padronização da estação, especialmente quando o exame é conduzido em mais de um

circuito simultaneamente (mais de um ator desempenhando o mesmo papel). A padronização consiste em treinar os PP para representar sintomas, respostas emocionais e atitudes semelhantes⁴⁷. O treinamento garante que todos os alunos enfrentarão a mesma situação durante o exame. Pacientes mal padronizados, cujo desempenho varia de um candidato para outro, podem reduzir a confiabilidade do Osce³⁹.

3.2. Viabilidade

O Osce, quando comparado a outros métodos de avaliação, exige maior investimento de recursos financeiros e logísticos, tanto na sua elaboração quanto na sua execução. Lidar com a viabilidade desse tipo de exame é um desafio para todos que trabalham com educação médica¹⁷. Por essa razão, o Osce deve ser reservado para avaliação de habilidades práticas, e não para avaliação de conhecimento teórico. As estações que passaram por todo um processo sistematizado de elaboração apresentam qualidade assegurada. Considerando o esforço necessário para se organizar um Osce, manter um banco de estações criado a partir de estações de qualidade contribui significativamente para uma maior confiabilidade e validade dos resultados do exame⁴⁰. Os relatos de experiências exitosas no cenário nacional, tanto em cursos de medicina^{31; 32; 36; 37; 38; 56} quanto em outras áreas da saúde^{30; 57; 58}, apontam estratégias locais para lidar com dificuldades logísticas e financeiras. Esses estudos demonstram a viabilidade da utilização do Osce nos cursos da área da saúde no Brasil. As vantagens e desvantagens do Osce estão listadas no quadro 2.

3.3. Impacto educacional

O impacto na aprendizagem dos alunos, resultante de um processo de avaliação, é uma fonte de validade do teste, denominada “evidência de validade baseada nas consequências do teste”^{50; 46}. Sabe-se que a avaliação direciona o aprendizado, uma vez que os alunos se concentram mais em ter bom desempenho nos testes do que em seu próprio aprendizado^{27; 59}. Essa dedicação também é influenciada pelo peso que a avaliação tem na composição da nota final⁶⁰. O Osce, ao avaliar habilidades clínicas, pode ter um impacto educacional positivo ao direcionar o aprendizado para a aquisição das competências clínicas desejadas⁴⁶. Esse impacto depende do grau de realismo das estações do Osce ao reproduzir situações semelhantes às vivenciadas nos cenários de prática real³⁹. Mas o impacto educacional pode diminuir quando a avaliação das habilidades clínicas é reduzida a cumprir um determinado número de tarefas

predefinidas e está relacionada ao uso das listas de verificação como instrumento de avaliação^{39; 42}. O uso de escalas de avaliação tendem a estimular os estudantes a aprender e praticar a habilidade de uma forma mais integrada ao contexto clínico⁴⁶.

3.4. Aceitação da avaliação

A aceitação de um método de avaliação significa que a comunidade acadêmica o reconheceu como sendo pertinente, adequado e justo⁶¹. O Osce é considerado um exame de elevada aceitabilidade tanto entre educadores quanto estudantes¹⁷. Seu impacto educacional positivo e as evidências de validade de seus resultados levaram à incorporação do método ao processo de avaliação de estudantes de medicina em todo o mundo, tanto em avaliações que embasam decisões de alto impacto (aprovação/reprovação, certificação profissional) quanto formativas^{13; 14; 26; 39; 40; 41; 48; 55}. A cultura institucional, os altos custos e o elevado consumo de tempo em sua elaboração e execução são desafios a serem vencidos³⁹. Na fase de implantação de um novo método de avaliação, recomenda-se que os envolvidos participem dela ativamente, visando assegurar sustentabilidade à mudança⁶². Ter o apoio institucional é fundamental para o sucesso do projeto. No Brasil, estudos experimentais apontam tanto para a satisfação dos estudantes quanto dos docentes envolvidos no Osce^{30; 36; 56}.

4. Conclusões

A avaliação de habilidades clínicas é fundamental para garantir a qualidade do profissional na formação médica, devendo fazer parte de um programa ou sistema de avaliação mais amplo, que inclua outros métodos, como avaliação cognitiva e avaliação da prática profissional em ambiente real. Sabe-se que nenhum método isoladamente pode fornecer todas as informações necessárias para se definir a competência profissional.

O Osce avalia habilidades clínicas de forma integrada em ambiente simulado. Por ser padronizado e usar múltiplos examinadores, é considerado uma avaliação mais justa. Para que tenha níveis razoáveis de validade e confiabilidade, exige elevados padrões de qualidade na sua elaboração e execução. Avaliações de alto impacto (aprovação/reprovação, certificação) exigem que os quesitos “validade” e “confiabilidade” cumpram padrões mais rigorosos. Executar um Osce requer muita base logística e treinamento para todos os envolvidos, com custos mais elevados.

As HC são habilidades clínicas complexas, fundamentais no exercício da medicina. Os métodos de avaliação precisam abranger aspectos distintos, tais como empatia, organização da entrevista, entonação da voz e linguagem não verbal. O Osce pode ser utilizado para a avaliação das HC, tanto em nível de graduação quanto pós-graduação, em caráter formativo ou somativo. A avaliação em ambiente simulado permite que o estudante demonstre essas habilidades de forma integrada, simulando a prática. No entanto, não há garantia da transposição de conhecimento. A necessária avaliação do desempenho dos estudantes no contexto real da prática profissional exige a utilização de outros métodos dirigidos a essa finalidade específica. Como exemplo, podemos citar o Mini-Cex⁴ – desenvolvido para avaliar competências clínicas no ambiente de trabalho, pode ser adaptado ao objetivo da avaliação, permitindo *feedback* imediato ao aluno no final da observação.

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

Referências

- ¹ BRASIL. Ministério da Educação. Conselho Nacional de Educação. Câmara de Educação Superior. Resolução CNE/CES Nº 4, de 7 de novembro de 2001. **Institui diretrizes curriculares nacionais do curso de graduação em Medicina**. Diário Oficial da União. Brasília, 09 nov. 2001; Seção 1, p.38
- ² BRASIL. Ministério da Educação. Conselho Nacional de Educação. Câmara de Educação Superior. Resolução CNE/CES Nº 3 de 20 de junho de 2014. **Institui diretrizes curriculares nacionais do curso de graduação em Medicina e dá outras providências**. Diário Oficial da União. Brasília, 23 jun. 2014; Seção 1, p.8-11.
- ³ LAIDLAW, A.; HART, J. Communication skills: an essential component of medical curricula. Part I: Assessment of clinical communication: AMEE Guide No. 51. **Med Teach**, v. 33, n. 1, p. 6-8, 2011. ISSN 0142-159x.
- ⁴ STEWART, M. A. Effective physician-patient communication and health outcomes: a review. **CMAJ: Canadian Medical Association Journal**, v. 152, n. 9, p. 1423-1433, 1995. ISSN 0820-3946 1488-2329.
- ⁵ HASKARD ZOLNIEREK, K. B.; DIMATTEO, M. R. Physician Communication and Patient Adherence to Treatment: A Meta-analysis. **Medical care**, v. 47, n. 8, p. 826-834, 2009. ISSN 0025-7079 1537-1948.

- 6 LEVINSON, W. et al. Physician-patient communication: The relationship with malpractice claims among primary care physicians and surgeons. **JAMA**, v. 277, n. 7, p. 553-559, 1997. ISSN 0098-7484.
- 7 SANDARS, J. et al. Educating undergraduate medical students about patient safety: Priority areas for curriculum development. **Medical Teacher**, v. 29, n. 1, p. 60-61, 2007/01/01 2007. ISSN 0142-159X.
- 8 HAWKINS, R. E. et al. Implementation of competency-based medical education: are we addressing the concerns and challenges? **Med Educ**, v. 49, n. 11, p. 1086-102, Nov 2015. ISSN 0308-0110.
- 9 KIESSLING, C. et al. Communication and social competencies in medical education in German-speaking countries: the Basel consensus statement. Results of a Delphi survey. **Patient Educ Couns**, v. 81, n. 2, p. 259-66, Nov 2010. ISSN 0738-3991.
- 10 VON FRAGSTEIN, M. et al. UK consensus statement on the content of communication curricula in undergraduate medical education. **Med Educ**, v. 42, n. 11, p. 1100-7, Nov 2008. ISSN 0308-0110.
- 11 MAKOUL, G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. **Acad Med**, v. 76, n. 4, p. 390-3, Apr 2001. ISSN 1040-2446.
- 12 _____. Improving communication with all patients. **Med Educ**, v. 42, n. 11, p. 1050-2, Nov 2008. ISSN 0308-0110.
- 13 COMERT, M. et al. Assessing Communication Skills of Medical Students in Objective Structured Clinical Examinations (OSCE)--A Systematic Review of Rating Scales. **PLoS One**, v. 11, n. 3, p. e0152717, 2016. ISSN 1932-6203.
- 14 GILLIS, A. E.; MORRIS, M. C.; RIDGWAY, P. F. Communication skills assessment in the final postgraduate years to established practice: a systematic review. **Postgraduate Medical Journal**, 2014.
- 15 DUFFY, F. D. et al. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. **Acad Med**, v. 79, n. 6, p. 495-507, Jun 2004. ISSN 1040-2446 (Print) 1040-2446.
- 16 ACCREDITATION COUNCIL FOR GRADUATE MEDICAL, E. Advancing education in interpersonal and communication skills: an educational resource from the ACGME Outcome Project. **Accreditation Council for Graduate Medical Education: Chicago, IL, USA**, 2005.

- 17 NORCINI, J. J.; MCKINLEY, D. W. Assessment methods in medical education. **Teaching and Teacher Education**, v. 23, n. 3, p. 239-250, 4// 2007. ISSN 0742-051X.
- 18 NORCINI, J. et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. **Med Teach**, v. 33, n. 3, p. 206-14, 2011. ISSN 1466-187X.
- 19 VAN DER VLEUTEN, C. P. et al. 12 Tips for programmatic assessment. **Med Teach**, p. 1-6, Nov 20 2014. ISSN 0142-159x.
- 20 ENDE, J. Feedback in clinical medical education. **Jama**, v. 250, n. 6, p. 777-81, Aug 12 1983. ISSN 0098-7484 (Print) 0098-7484.
- 21 KILMINSTER, S. et al. AMEE Guide No. 27: Effective educational and clinical supervision. **Med Teach**, v. 29, n. 1, p. 2-19, Feb 2007. ISSN 0142-159x.
- 22 RIDER, E. A.; HINRICHS, M. M.; LOWN, B. A. A model for communication skills assessment across the undergraduate curriculum. **Med Teach**, v. 28, n. 5, p. e127-34, Aug 2006. ISSN 0142-159x.
- 23 MAKOUL, G.; KRUPAT, E.; CHANG, C. H. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. **Patient Educ Couns**, v. 67, n. 3, p. 333-42, Aug 2007. ISSN 0738-3991.
- 24 HODGES, B. et al. OSCE checklists do not capture increasing levels of expertise. **Acad Med**, v. 74, n. 10, p. 1129-34, Oct 1999. ISSN 1040-2446 (Print) 1040-2446.
- 25 HODGES, B.; MCILROY, J. H. Analytic global OSCE ratings are sensitive to level of training. **Med Educ**, v. 37, n. 11, p. 1012-6, Nov 2003. ISSN 0308-0110.
- 26 SETYONUGROHO, W.; KENNEDY, K. M.; KROPMANS, T. J. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. **Patient Educ Couns**, Jun 27 2015. ISSN 0738-3991.
- 27 SWANSON, D. B.; VAN DER VLEUTEN, C. P. Assessment of clinical skills with standardized patients: state of the art revisited. **Teach Learn Med**, v. 25 Suppl 1, p. S17-25, 2013. ISSN 1040-1334.
- 28 ZILL, J. M. et al. Measurement of physician-patient communication--a systematic review. **PLoS One**, v. 9, n. 12, p. e112637, 2014. ISSN 1932-6203.
- 29 HARDEN, R. M. et al. Assessment of clinical competence using objective structured examination. **British Medical Journal**, v. 1, n. 5955, p. 447, 1975.

- 30 CACHO, R. D. O. et al. Metodologias Ativas em Fisioterapia: Estudo de Confiabilidade Interexaminador do Método Osce. **Revista Brasileira de Educação Médica**, v. 40, p. 128-137, 2016. ISSN 0100-5502.
- 31 NUNES, S. O. V. et al. O ensino de psiquiatria, habilidades de comunicação e atitudes no currículo integrado do curso de Medicina da Universidade Estadual de Londrina. **Revista Brasileira de Educação Médica**, v. 32, p. 210-216, 2008. ISSN 0100-5502.
- 32 NUNES, A. A. et al. Desenvolvimento de competências para o uso da ultrassonografia point-of-care em Nefrologia. **Jornal Brasileiro de Nefrologia**, v. 38, p. 209-214, 2016. ISSN 0101-2800.
- 33 SANDOVAL, G. E. et al. Análise de um sistema de avaliação de aprendizagem para internato em pediatria baseado em exame clínico objetivo estruturado, observação de prática clínica e exame escrito. **Jornal de Pediatria**, v. 86, p. 131-136, 2010. ISSN 0021-7557.
- 34 TRONCON, L. E. D. A. Significance of experts' overall ratings for medical student competence in relation to history-taking. **Sao Paulo Medical Journal**, v. 124, p. 101-104, 2006. ISSN 1516-3180.
- 35 AMARAL, E.; DOMINGUES, R. C. L.; BICUDO-ZEFERINO, A. M. Avaliando competência clínica: o método de avaliação estruturada observacional. **Revista Brasileira de Educação Médica**, v. 31, p. 287-290, 2007. ISSN 0100-5502.
- 36 DAUD-GALLOTTI, R. M. et al. A new method for the assessment of patient safety competencies during a medical school clerkship using an objective structured clinical examination. **Clinics**, v. 66, p. 1209-1215, 2011. ISSN 1807-5932.
- 37 TRONCON, L. E. D. A. Clinical skills assessment: limitations to the introduction of an "OSCE" (Objective Structured Clinical Examination) in a traditional Brazilian medical school. **Sao Paulo Medical Journal**, v. 122, p. 12-17, 2004. ISSN 1516-3180.
- 38 LUCCHETTI, G.; EZEQUIEL, O. S.; LUCCHETTI, A. L. An OSCE with very limited resources: Is it possible? **Med Teach**, v. 39, n. 2, p. 227, Feb 2017. ISSN 0142-159x.
- 39 KHAN, K. Z. et al. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. **Med Teach**, v. 35, n. 9, p. e1437-46, Sep 2013. ISSN 0142-159x.
- 40 _____. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. **Med Teach**, v. 35, n. 9, p. e1447-63, Sep 2013. ISSN 0142-159x.

- 41 PELL, G. et al. How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. **Med Teach**, v. 32, n. 10, p. 802-11, 2010. ISSN 0142-159x.
- 42 BOURSICOT, K.; ROBERTS, T. How to set up an OSCE. **The Clinical Teacher**, v. 2, n. 1, p. 16-20, 2005. ISSN 1743-498X.
- 43 HASTIE, M. J. et al. Designing and implementing the objective structured clinical examination in anesthesiology. **Anesthesiology**, v. 120, n. 1, p. 196-203, Jan 2014. ISSN 0003-3022.
- 44 AMIEL, G. E. et al. Ability of primary care physician's to break bad news: a performance based assessment of an educational intervention. **Patient Educ Couns**, v. 60, n. 1, p. 10-5, Jan 2006. ISSN 0738-3991 (Print) 0738-3991.
- 45 MORTSIEFER, A. et al. Summative assessment of undergraduates' communication competence in challenging doctor-patient encounters. Evaluation of the Delft CoMeD-OSCE. **Patient Education and Counseling**, v. 95, n. 3, p. 348-355, ISSN 0738-3991.
- 46 BOURSICOT, K. A. M. Structured assessments of clinical competence. **British Journal of Hospital Medicine**, v. 71, n. 6, p. 342-344, 2010/06/01 2010. ISSN 1750-8460.
- 47 CLELAND, J. A.; ABE, K.; RETHANS, J.-J. The use of simulated patients in medical education: AMEE Guide No 42. **Medical Teacher**, v. 31, n. 6, p. 477-486, 2009/01/01 2009. ISSN 0142-159X.
- 48 TAVAKOL, M.; DENNICK, R. Post-examination interpretation of objective test data: monitoring and improving the quality of high-stakes examinations--a commentary on two AMEE Guides. **Med Teach**, v. 34, n. 3, p. 245-8, 2012. ISSN 0142-159x.
- 49 NEWBLE, D. Techniques for measuring clinical competence: objective structured clinical examinations. **Med Educ**, v. 38, n. 2, p. 199-203, Feb 2004. ISSN 0308-0110 (Print) 0308-0110.
- 50 AMERICAN EDUCATIONAL RESEARCH, A.; AMERICAN PSYCHOLOGICAL, A.; NATIONAL COUNCIL ON MEASUREMENT IN, E. **Standards for educational and psychological testing**. Washington, DC: American Educational Research Association, 2014. ISBN 0935302255.
- 51 CODERRE, S.; WOLOSCHUK, W.; MCLAUGHLIN, K. Twelve tips for blueprinting. **Medical Teacher**, v. 31, n. 4, p. 322-324, 2009/01/01 2009. ISSN 0142-159X.
- 52 VAN DER VLEUTEN, C. P. The assessment of professional competence: Developments, research and practical implications. **Adv Health Sci Educ Theory Pract**, v. 1, n. 1, p. 41-67, Jan 1996. ISSN 1382-4996 (Print) 1382-4996.

- 53 PUGH, D.; SMEE, S. **Guidelines for the Development of Objective Structured Clinical Examination (OSCE) Cases**. Ottawa: Medical Council of Canada, 2013.
- 54 SETYONUGROHO, W. et al. Calibration of communication skills items in OSCE checklists according to the MAAS-Global. **Patient Education and Counseling**, v. 99, n. 1, p. 139-146, ISSN 0738-3991.
- 55 BRANNICK, M. T.; EROL-KORKMAZ, H. T.; PREWETT, M. A systematic review of the reliability of objective structured clinical examination scores. **Med Educ**, v. 45, n. 12, p. 1181-9, Dec 2011. ISSN 0308-0110.
- 56 FRANCO, C. A. G. D. S. et al. OSCE para Competências de Comunicação Clínica e Profissionalismo: Relato de Experiência e Meta-Avaliação. **Revista Brasileira de Educação Médica**, v. 39, p. 433-441, 2015. ISSN 0100-5502.
- 57 SILVA, C. C. B. M. et al. Objective structured clinical evaluation as an assessment method for undergraduate chest physical therapy students: a cross-sectional study. **Brazilian Journal of Physical Therapy**, v. 15, p. 481-486, 2011. ISSN 1413-3555.
- 58 GALATO, D. et al. Exame clínico objetivo estruturado (ECOEF): uma experiência de ensino por meio de simulação do atendimento farmacêutico. **Interface - Comunicação, Saúde, Educação**, v. 15, p. 309-320, 2011. ISSN 1414-3283.
- 59 SCHUWIRTH, L.; VAN DER VLEUTEN, C. Merging views on assessment. In: (Ed.). **Med Educ**. England, v.38, 2004. p.1208-10. ISBN 0308-0110 (Print) 0308-0110 (Linking).
- 60 WORMALD, B. W. et al. Assessment drives learning: an unavoidable truth? **Anat Sci Educ**, v. 2, n. 5, p. 199-204, Oct 2009. ISSN 1935-9772.
- 61 TRONCON, L. E. D. A. Estruturação de Sistemas para Avaliação Programática do Estudante de Medicina. **Revista Brasileira de Educação Médica**, v. 40, p. 30-42, 2016. ISSN 0100-5502.
- 62 VAN DER VLEUTEN, C. P.; SCHUWIRTH, L. W. Assessing professional competence: from methods to programmes. **Med Educ**, v. 39, n. 3, p. 309-17, Mar 2005. ISSN 0308-0110 (Print) 0308-0110.

Quadro 1: Exemplo de uma Matriz de Avaliação do Osce

Tópico/Área	Habilidade em exame físico	Habilidade em procedimentos	Habilidade de comunicação	Nº total de questões
Obstetrícia	Estação 1	Estação 6		2
Pediatria	Estação 2		Estação 9	2
Clínica Médica	Estação 3		Estação 10	2
Cirurgia		Estação 7		1
Emergência		Estação 8		1
Atenção Primária	Estação 4 Estação 5		Estação 11 Estação 12	4
Nº total de questões	5	3	4	12

Adaptado: Khan, K. Z. et al., 2018

Quadro 2: Osce - Vantagens e desvantagens

Vantagens	Limitações
<ul style="list-style-type: none"> • Observação direta de habilidades clínicas • Adaptável aos objetivos educacionais • Avaliação mais justa por ser baseada em padrões bem definidos • Avaliação de diversas habilidades em um período de tempo relativamente curto • Minimização do viés do avaliador devido ao uso de múltiplos examinadores • Elevada confiabilidade quando implementado dentro dos padrões de qualidade 	<ul style="list-style-type: none"> • Custo elevado • Consumo elevado de tempo • Limitado escopo do que se pode avaliar

Adaptado: Khan, K. Z. et al., 2018

Quadro 3: Exemplos de estações de Osce - Comunicação em situações desafiadoras

Tópico de comunicação	Descrição do caso clínico e postura do paciente	Tarefa/Comunicação em situação desafiadora	Cenário
1. Orientação e educação do paciente: compartilhamento de decisão	Mulher soropositiva de 25 anos, encaminhada para orientação sobre contracepção e que manifesta resistência ao uso de preservativo.	Lidar com a resistência e aconselhar com base nas recomendações específicas para o caso.	Atenção Primária – Ginecologia
2. Resultado desfavorável de exames	Mulher casada, 30 anos, grávida do primeiro filho com exame de VDRL positivo e que questiona como adquiriu a doença. Está tensa e ansiosa.	Lidar com as dúvidas e ansiedade da paciente diante do diagnóstico inesperado.	Atenção Primária – Obstetrícia
3. Incerteza do diagnóstico ainda não confirmado	Mulher, 45 anos, apresenta nódulo palpável em mama durante atendimento em consulta de rotina.	Preparar o paciente para lidar com uma possível situação desfavorável, mas que ainda precisa ser confirmada.	Atenção Primária – Ginecologia
4. Orientação e educação do paciente: comunicação de risco	Homem, 50 anos, sobrepeso, hipertenso e sem sintomas. Demonstra resistência em acreditar que sua condição seja algum problema de saúde. Não aceita tomar medicamentos.	Orientar sobre o risco cardiovascular e definir plano compartilhado de tratamento.	Atenção Primária – Clínica Médica
5. Comunicação de más notícias: comunicação de diagnóstico grave	Homem, 60 anos, retorna com resultado de biópsia gástrica positiva para carcinoma. No atendimento anterior, você considerou tratar-se de uma gastrite.	Lidar com a comunicação do diagnóstico grave recente.	Atenção Primária – Clínica Médica
6. Erro médico: gerenciamento de conflitos	Menino, 4 anos, com dor nas pernas, tratada como dor do crescimento, mas posteriormente recebe o diagnóstico de leucemia. Pais revoltados.	Lidar com a agressividade dos familiares e dilemas éticos.	Atenção Primária – Pediatria
7. Abuso sexual infantil	Menina, 10 anos, apresentando lesões genitais compatíveis com Condiloma (DST). Acompanhada da mãe.	Comunicar o diagnóstico e lidar com a necessidade de notificar o caso às autoridades competentes.	Atenção Primária – Pediatria
8. Violência doméstica	Mulher, 19 anos, múltiplas feridas e contusões. Paciente sozinha e chorando muito.	Lidar com casos de violência doméstica.	Serviço de Urgência – Cirurgia
9. Diversidade de gênero	Transexual masculino, 23 anos, solicita orientação sobre a cirurgia para mudança de sexo.	Lidar com a diversidade de gênero.	Atenção Primária – Cirurgia
10. Consentimento para realizar procedimento cirúrgico	Adolescente, 12 anos, apendicite aguda acompanhada dos pais. Pais inseguros diante dos riscos cirúrgicos e verbalizam o medo da morte da filha.	Comunicar diagnóstico, orientar sobre o procedimento e solicitar consentimento ao responsável.	Serviço de urgência – Cirurgia.

Fonte: Autores

APÊNDICE B: AG-OSCE-R – Orientações de preenchimento

Orientações aos avaliadores para o preenchimento da escala de avaliação de habilidades de comunicação durante a prova de habilidades clínicas do Curso de Medicina da FCM-Unicamp

Prezados avaliadores, a capacidade dos alunos em colher a história clínica/fazer anamnese e se comunicar com o paciente será avaliada por uma escala de desempenho composta por 5 itens: 4 itens analíticos e 1 conceito global.

Os 4 itens analíticos são: empatia, coerência durante a consulta, linguagem verbal e linguagem não verbal. A nota em cada item vai de 1 a 5, onde 1 é a pior nota e 5 a melhor. As notas nos 4 itens serão somadas, sendo a maior nota possível 20 pontos e a menor 4 pontos. O item conceito global é o momento em que o docente faz uma avaliação geral do desempenho do aluno. Esse item permite 3 conceitos: incompetente, limítrofe e competente e deve guardar relação direta com as notas dadas nos 4 itens anteriores.

Por exemplo, um aluno com notas entre 1 e 2 nos 4 itens anteriores (somando de 4 a 8 pontos), provavelmente não deverá ser considerado competente, assim como aquele com notas entre 4 e 5 (somando de 16 a 20 pontos) não será considerado incompetente. A utilização do conceito global se torna especialmente útil para aqueles alunos cujas notas variaram muito entre os 4 itens anteriores (somando entre 9 e 15 pontos). Nesse momento o docente tem a oportunidade de fazer uma avaliação global de desempenho do aluno que pode ser melhor ou pior que a simples soma das notas dos itens anteriores, mas sobretudo coerente com o desempenho observado durante toda a estação.

Por favor, pedimos que leiam atentamente cada item. Sugerimos que essa escala seja preenchida ao final da estação.

Muito obrigada!

ANEXOS

ANEXO 1: Communication Assessment Tool (CAT) – Instrumento original

Communication with patients is a very important part of quality medical care. We would like to know how you feel about the way your doctor communicated with you. Your answers are completely confidential, so please be as open and honest as you can. Thank you very much.

1	2	3	4	5
poor	fair	good	very good	excellent

Please use this scale to rate the resident physician's communication with you. Circle your answer for each item below.

<u>The doctor</u>	<u>poor</u>					<u>excellent</u>
1. Greeted me in a way that made me feel comfortable	1	2	3	4	5	
2. Treated me with respect	1	2	3	4	5	
3. Showed interest in my ideas about my health	1	2	3	4	5	
4. Understood my main health concerns	1	2	3	4	5	
5. Paid attention to me (looked at me, listened carefully)	1	2	3	4	5	
6. Let me talk without interruptions	1	2	3	4	5	
7. Gave me as much information as I wanted	1	2	3	4	5	
8. Talked in terms I could understand	1	2	3	4	5	
9. Checked to be sure I understood everything	1	2	3	4	5	
10. Encouraged me to ask questions	1	2	3	4	5	
11. Involved me in decisions as much as I wanted	1	2	3	4	5	
12. Discussed next steps, including any follow-up plans	1	2	3	4	5	
13. Showed care and concern	1	2	3	4	5	
14. Spent the right amount of time with me	1	2	3	4	5	

ANEXO 2: Analytic Global OSCE Rating (AG-OSCE-R) – Instrumento original

Response to patient's feelings and needs (empathy)				
1	2	3	4	5
Does not respond to obvious patient cues (verbal and non-verbal) and / or responds inappropriately		Responds to patient's needs and cues, but not always effectively		Responds consistently in a perceptive and genuine manner to the patient's needs and cues
Degree of coherence in the interview				
1	2	3	4	5
No recognizable plan to the interaction; the plan does not demonstrate cohesion or the patient must determine the direction of the interview		Organizational approach is formulaic and minimally flexible and/or control of the interview is inconsistent		Superior organization, demonstrating command of cohesive devises, flexibility, and consistent control of the interview
Verbal expression				
1	2	3	4	5
Communicates in manner that interferes with and/or prevents understanding by patient, or communicates inappropriately with the patient		Exhibits sufficient control of expression to be understood by an active, engaged listener (patient)		Exhibits command of expression (fluency, diction, grammar, vocabulary, tone, volume and modulation of voice, rate of speech, pace and pronunciation)
Non-verbal expression				
1	2	3	4	5
Fails to engage, frustrates and /or antagonises the patient		Exhibits enough control of non-verbal expression to engage a patient willing to overlook deficiencies such as passivity, self-consciousness or inappropriate aggressiveness		Exhibits finesse and command of non-verbal expression (eye-contact, gesture, posture, use of silence, etc.)
Overall assessment of the knowledge and skills demonstrated in the interview				
A = Incompetent		B = Borderline		C = Competent
Responds inappropriately and ineffectively to the task, indicating a lack of knowledge and/or undeveloped interpersonal and interviewing skills		Responds effectively to some components of the task, some development of interpersonal and interviewing skills		Responds precisely and perceptively to the task, consistently integrating all components

ANEXO 3: Autorização para adaptação do Analytic Global OSCE Rating (AG-OSCE-R)



Sheyla Ribeiro Rocha <sheyla.ribeiro.rocha@gmail.com>

validating Hodges&McIlroy communication tool in Portuguese

Jodi Herold <Jodi.Herold@utoronto.ca>

26 de abril de 2016 14:25

Para: Eliana Amaral <amaraleli@gmail.com>

Cc: "brian.hodges@utoronto.ca" <brian.hodges@utoronto.ca>, Sheyla Ribeiro Rocha <sheyla.ribeiro.rocha@gmail.com>, "Hodges, Brian" <Brian.Hodges@uhn.ca>, Carlos Fernando Collares <c.collares@maastrichtuniversity.nl>

Dear Eliana

Apologies for the delay in responding

Thanks for sending the English version of the proposal

The proposal and support that you have assembled are both very strong,

Of course you have my (our) permission to validate a Portuguese version.

Best of luck with the project,

We look forward to seeing your results

Please let us know if there is anything else we can do to support this endeavor

(I'm sure I speak for Dr. Hodges as well in making this offer)

Best

Jodi Herold (previously McIlroy)

From: Eliana Amaral [mailto:amaraleli@gmail.com]

Sent: Saturday, April 23, 2016 5:12 AM

To: Jodi Herold

Cc: brian.hodges@utoronto.ca; Sheyla Ribeiro Rocha; Hodges, Brian; Carlos Fernando Collares

Subject: Re: validating Hodges&McIlroy communication tool in Portuguese

[Texto das mensagens anteriores oculto]

ANEXO 4: Autorização para adaptação do Communication Assessment Tool (CAT)

Fwd: RE: collaboration on CAT from Brazil

Gregory Makoul <gmakoul@patientwisdom.com>

23 de abril de 2016 18:18

Para: Eliana Amaral <amaraleli@gmail.com>

Cc: Sheyla Ribeiro Rocha <sheyla.ribeiro.rocha@gmail.com>, gmakoul@uchc.edu

Hi. There is no cost or payment. I simply want to make it clear that I hold the copyright even if the work in Brazil leads to refining wording. Thank you.
-gm

Gregory Makoul, PhD MS
via iphone

On Apr 23, 2016, at 4:19 PM, Eliana Amaral <amaraleli@gmail.com> wrote:

Hi Gregory (may I use only first name?),

Thanks for your kind answer.

We intend to use only P7 of CAT. Since you mentioned copyright, I am not sure if this means it is necessary to pay even if it is part of a scientific project for a PhD student. Or could we use CAT - P7 as it is, already translated, partnering with you? At this point, we may use in a simulated scenario, but we could explore other possibilities depending on the possibilities for scientific partnership. Could you please clarify what possible collaboration we could have for the use of CAT? If you prefer, we could even Skype another time during the week.

Best regards,

Eliana
UNICAMP

Prof. Dra. Eliana Amaral
Professora-Titular de Obstetrícia & Chefe do Departamento de Tocoginecologia, FCM/UNICAMP
Presidente da Comissão de Corpo Docente - FCM/UNICAMP & Membro da CIDD/UNICAMP
Membro da Comissão de Ensino e Avaliação da Febrasgo & Co-diretora Faimer Brasil
<http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4787083P3>

2016-04-23 10:31 GMT-03:00 Gregory Makoul <gmakoul@patientwisdom.com>:

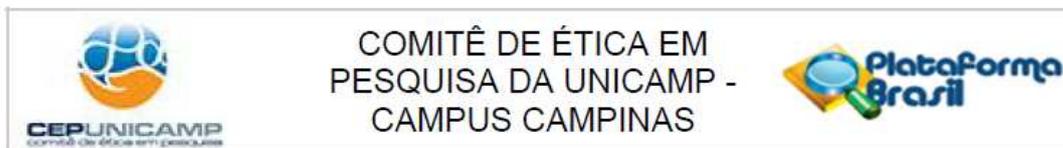
Hi Eliana. Very sorry for the delay in getting back to you – I missed the original message. We do have a version of the CAT in Portuguese (attached). You are welcome to use – and refine – the Brazilian CAT for academic purposes. Please note that I retain copyright in all languages. The only thing I ask is that you and Sheyla keep me posted on your progress and share de-identified data for our growing CAT database.

As you may know, I have been working closely with colleagues from many countries on translation and cultural adaptation of the CAT. I would be honored to work with you, and am very happy to help along the way. With thanks, and all good wishes,

-gm

Gregory Makoul, PhD MS
mobile 860.216.7010

ANEXO 5: Aprovação do Comitê de Ética em Pesquisa da Unicamp



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Validade de instrumentos de avaliação de habilidades interpessoais e de comunicação em educação médica.

Pesquisador: SHEYLA RIBEIRO ROCHA

Área Temática:

Versão: 1

CAAE: 56935716.5.0000.5404

Instituição Proponente: Faculdade de Ciências Médicas - UNICAMP

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 1.613.689

Apresentação do Projeto:

Reconhecendo a habilidade de comunicação como competência essencial na formação médica, multiplicam-se as estratégias educacionais para o seu ensino. Em 2014, as Diretrizes Curriculares Nacionais (DCN) brasileiras para os Cursos de Medicina foram revistas e incluíram novos aspectos da comunicação obrigatórios nos currículos das escolas médicas. Esse trabalho tem como objetivo validar dois diferentes instrumentos de avaliação de habilidades interpessoais e de comunicação na relação médico-paciente a partir da análise de suas propriedades psicométricas. O primeiro, desenvolvido por Hodges e McIlroy (2003) para ser aplicado pelo docente-avaliador, avalia quatro elementos-chave da comunicação médica: a empatia (resposta a sentimentos e necessidades dos pacientes), a coerência na entrevista (flexibilidade e domínio da entrevista), a expressão verbal (ritmo e modulação da voz) e a expressão não-verbal (contato visual e gestos). Cada item é avaliado por uma escala de 5 pontos e ao final, é formado um conceito global para o desempenho do estudante. O segundo, desenvolvido por Makoul, Krupat e Chang (2007) para ser aplicado pelo paciente, avalia a percepção do usuário quanto a qualidade da comunicação durante a entrevista clínica e é composto por 15 itens, sendo cada um pontuado de acordo com uma escala de 5 pontos. As escalas serão traduzidas para o Português Brasileiro de acordo com protocolo específico para adaptação transcultural de instrumentos e serão incorporados ao Exame Clínico

Endereço: Rua Tessália Vieira de Camargo, 126
Bairro: Barão Geraldo **CEP:** 13.083-887
UF: SP **Município:** CAMPINAS
Telefone: (19)3521-8936 **Fax:** (19)3521-7187 **E-mail:** cep@fcm.unicamp.br



Continuação do Parecer: 1.613.689

Objetivo Estruturado (OSCE) que avalia habilidades clínicas de estudantes de medicina da Faculdade de Ciências Médicas da Unicamp. Esse OSCE formativo é regularmente aplicado a estudantes ao final do 5º e 6º ano do curso de medicina, nas áreas de pediatria, ginecologia e obstetrícia, clínica médica, saúde coletiva e cirurgia. A validade de conteúdo e a pertinência dos construtos será avaliada por um comitê de juizes experts em comunicação médica. A validade do construto será investigada usando análise fatorial exploratória. A confiabilidade entre avaliadores será medida pelo coeficiente de correlação intraclasse. A validade preditiva será testada com o desempenho nas provas de admissão para residência médica, bem como a pontuação acadêmica regular (RA) no último ano para estudantes do sexto ano.

Objetivo da Pesquisa:

Objetivo Geral:

Estudar evidências de validade dos instrumentos Analytic Global Rating Form (AGRF) de Hodges e Mclroy (2003) e Communication Assessment Tool (CAT) de Makoul, Krupat e Chang (2007) para avaliar habilidades de comunicação e relação interpessoal em estudantes de medicina no Brasil.

Objetivos específicos

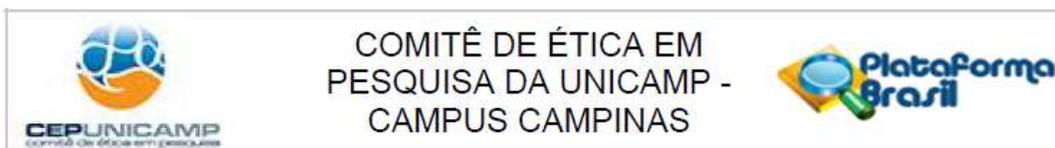
1. Traduzir e adaptar os instrumentos para uso no Brasil;
2. Analisar a equivalência semântica dos instrumentos adaptados, por meio da análise do significado geral e referencial;
3. Analisar a pertinência dos itens dos instrumentos adaptados aos construtos pretendidos por meio da análise semântica com a população-alvo, e por meio da análise de juizes;
4. Obter evidências de validade com base na estrutura interna por meio da análise de eixos principais (AEP) e análise por teoria de resposta ao item (TRI);
5. Obter evidências de validade com base no processo de resposta por meio de entrevista cognitiva com os juizes/avaliadores;
6. Estimar a confiabilidade (precisão) dos instrumentos adaptados por meio dos coeficientes de Cronbach e, por meio da TRI, a descrição da confiabilidade condicional (precisão local) e das funções de informação de construtos e itens.

Avaliação dos Riscos e Benefícios:

Segundo a pesquisadora este projeto não apresenta riscos previsíveis para os participantes. Se prevê, inclusive dito no TCLE, que caso haja qualquer desconforto ou constrangimento para o sujeito, ele pode desistir do mesmo sem prejuízo nenhum.

Os benefícios virão por apresentar novas propostas de comunicação para o profissional médico, favorecendo uma formação da classe em solo brasileiro, assim como melhorando o atendimento à

Endereço: Rua Tessália Vieira de Camargo, 126			
Bairro: Barão Geraldo	CEP: 13.083-887		
UF: SP	Município: CAMPINAS		
Telefone: (19)3521-8936	Fax: (19)3521-7187	E-mail: cep@fcm.unicamp.br	



Continuação do Parecer: 1.613.689

população.

Comentários e Considerações sobre a Pesquisa:

Trata-se de um doutorado da FCM/Unicamp. Projeto muito bem estruturado. Quanto mais se melhora a comunicação médico-paciente, melhor para a Saúde Brasileira.

Essa pesquisa será desenvolvida na Faculdade de Ciências Médicas (FCM) da Unicamp durante Exame de Avaliação de Habilidades Clínicas que acontece anualmente como parte do processo de avaliação conduzida pela Coordenação do Curso e o Núcleo de Avaliação e Pesquisa em Ensino Médico (NAPEM), previsto em 24 de setembro de 2016.

A etapa do pré-teste que precede a fase de aplicação da versão final do instrumento será realizada durante a avaliação de habilidades clínicas dos residentes do Programa de Residência em Ginecologia e Obstetrícia da Unicamp. Essa atividade é conduzida pela Coordenação do Programa de Residência em Ginecologia e Obstetrícia, e está prevista para 06 de agosto de 2016.

Ao todo serão envolvidos 276 sujeitos: residentes e ou alunos em internato (34) -Curso de Medicina ou no Programa de Residência em Ginecologia e Obstetrícia da FCM da Unicamp/ avaliadores (8)/ juizes (10) / alunos do 5° e 6° anos (grupo maior - 220) / pacientes simulados (4) /

Considerações sobre os Termos de apresentação obrigatória:

A pesquisadora atende em seu projeto a todas as exigências da resolução 466/12: Folha de Rosto com autorização / Critérios de inclusão e exclusão em todas as fases / Cronograma / Financiamento / TCLEs de acordo / Assim como esta bem justificado o pedido de dispensa do TCLE para alunos do 5° e 6° anos - na segunda fase da pesquisa.

Conclusões ou Pendências e Lista de Inadequações:

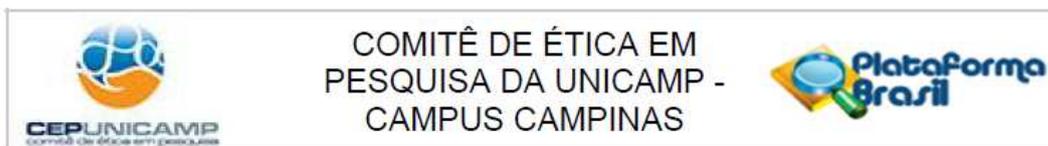
Nenhuma.

Considerações Finais a critério do CEP:

- O sujeito de pesquisa deve receber uma via do Termo de Consentimento Livre e Esclarecido, na íntegra, por ele assinado (quando aplicável).

- O sujeito da pesquisa tem a liberdade de recusar-se a participar ou de retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo ao seu cuidado (quando aplicável).

Endereço: Rua Tessália Vieira de Camargo, 126
 Bairro: Barão Geraldo CEP: 13.083-887
 UF: SP Município: CAMPINAS
 Telefone: (19)3521-8936 Fax: (19)3521-7187 E-mail: cep@fcm.unicamp.br



Continuação do Parecer: 1.613.689

- O pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado. Se o pesquisador considerar a descontinuação do estudo, esta deve ser justificada e somente ser realizada após análise das razões da descontinuidade pelo CEP que o aprovou. O pesquisador deve aguardar o parecer do CEP quanto à descontinuação, exceto quando perceber risco ou dano não previsto ao sujeito participante ou quando constatar a superioridade de uma estratégia diagnóstica ou terapêutica oferecida a um dos grupos da pesquisa, isto é, somente em caso de necessidade de ação imediata com intuito de proteger os participantes.

- O CEP deve ser informado de todos os efeitos adversos ou fatos relevantes que alterem o curso normal do estudo. É papel do pesquisador assegurar medidas imediatas adequadas frente a evento adverso grave ocorrido (mesmo que tenha sido em outro centro) e enviar notificação ao CEP e à Agência Nacional de Vigilância Sanitária – ANVISA – junto com seu posicionamento.

- Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas e aguardando a aprovação do CEP para continuidade da pesquisa. Em caso de projetos do Grupo I ou II apresentados anteriormente à ANVISA, o pesquisador ou patrocinador deve enviá-las também à mesma, junto com o parecer aprovatório do CEP, para serem juntadas ao protocolo inicial.

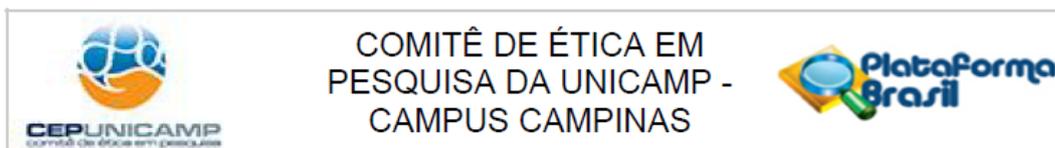
- Relatórios parciais e final devem ser apresentados ao CEP, inicialmente seis meses após a data deste parecer de aprovação e ao término do estudo.

- Lembramos que segundo a Resolução 466/2012, item XI.2 letra e, "cabe ao pesquisador apresentar dados solicitados pelo CEP ou pela CONEP a qualquer momento".

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_708879.pdf	10/06/2016 15:51:36		Aceito
Outros	Carteir_Estudantil_UNICAMP.pdf	10/06/2016 15:41:46	SHEYLA RIBEIRO ROCHA	Aceito

Endereço: Rua Tessália Vieira de Camargo, 126
 Bairro: Barão Geraldo CEP: 13.083-887
 UF: SP Município: CAMPINAS
 Telefone: (19)3521-8936 Fax: (19)3521-7187 E-mail: cep@fcm.unicamp.br



Continuação do Parecer: 1.613.689

Outros	Atestado_Matricula_Doutorado_2016.pdf	10/06/2016 15:40:13	SHEYLA RIBEIRO ROCHA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLEs.pdf	08/06/2016 22:17:36	SHEYLA RIBEIRO ROCHA	Aceito
Projeto Detalhado / Brochura Investigador	Proj_Plata_Brasil_08062016.pdf	08/06/2016 22:16:09	SHEYLA RIBEIRO ROCHA	Aceito
Outros	Parecer_Com_Pes_CAISM.pdf	08/06/2016 22:08:46	SHEYLA RIBEIRO ROCHA	Aceito
Folha de Rosto	Folha_Rosto.pdf	08/06/2016 22:06:21	SHEYLA RIBEIRO ROCHA	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

CAMPINAS, 29 de Junho de 2016

Assinado por:

Renata Maria dos Santos Celeghini
(Coordenador)

Endereço: Rua Tessália Vieira de Camargo, 126
 Bairro: Barão Geraldo CEP: 13.083-887
 UF: SP Município: CAMPINAS
 Telefone: (19)3521-8936 Fax: (19)3521-7187 E-mail: cep@fcm.unicamp.br

ANEXO 6: Versão adaptada do Communication Assessment Tool (CAT)

Estação n° ____ – Instrumento de Avaliação da Comunicação pelo Paciente Simulado

Avaliador: _____

Residente: _____

A comunicação com os pacientes é uma parte muito importante da consulta médica de qualidade. Gostaríamos de saber como você se sentiu sobre a maneira como seu/a médico (a) se comunicou com você. Suas respostas são sigilosas, portanto, seja o mais sincero possível. Muito obrigado.

1	2	3	4	5
Ruim	Aceitável	Boa	Muito boa	Excelente

Por favor, use esta escala para classificar como o médico se comunicou com você. Circule sua resposta para cada item abaixo.

<u>O Médico</u>	Ruim					Excelente
1. Me cumprimentou de modo a me deixar à vontade	1	2	3	4	5	
2. Me tratou com respeito	1	2	3	4	5	
3. Demonstrou interesse por minhas opiniões sobre minha saúde	1	2	3	4	5	
4. Compreendeu minhas principais preocupações com a saúde	1	2	3	4	5	
5. Prestou atenção em mim (olhou para mim, me ouviu com atenção)	1	2	3	4	5	
6. Deixou que eu falasse sem me interromper	1	2	3	4	5	
7. Me deu todas as informações que eu pedi	1	2	3	4	5	
8. Falou usando palavras que eu consegui entender	1	2	3	4	5	
9. Verificou se eu entendi tudo	1	2	3	4	5	
10. Me incentivou a fazer perguntas	1	2	3	4	5	
11. Me envolveu nas decisões tanto quanto eu queria	1	2	3	4	5	
12. Discutiu os próximos passos, incluindo planos de acompanhamento	1	2	3	4	5	
13. Demonstrou cuidado e interesse	1	2	3	4	5	
14. Passou comigo a quantidade de tempo necessária	1	2	3	4	5	

ANEXO 7: Versão adaptada do Analytic Global OSCE Rating (AG-OSCE-R)

Estação nº – Instrumento de Avaliação da comunicação médico-paciente durante o OSCE

Avaliador: _____

Aluno: _____

Resposta aos sentimentos e às necessidades do paciente (empatia)				
1	2	3	4	5
Não responde a sinais (verbais e não verbais) óbvios apresentados pelo paciente e/ou responde de maneira inapropriada.		Responde às necessidades e aos sinais apresentados pelo paciente, porém nem sempre de forma efetiva.		Responde consistentemente às necessidades e sinais apresentados pelo paciente, de maneira perspicaz e autêntica.
Grau de coerência da entrevista				
1	2	3	4	5
Não foi identificado nenhum plano durante a interação; o plano utilizado não demonstra coesão ou o paciente precisa determinar o direcionamento da entrevista.		A abordagem quanto à organização da entrevista é padronizada, pouco flexível e/ou o controle da entrevista é inconsistente.		A organização da entrevista é excelente, demonstrando domínio de estratégias coesas, flexibilidade e controle consistente da entrevista.
Expressão verbal				
1	2	3	4	5
Comunica-se de modo a interferir na e/ou impedir a compreensão pelo paciente, ou se comunica inadequadamente com o paciente.		Demonstra controle suficiente da expressão verbal de modo a ser compreendido por um ouvinte ativo e participativo (paciente).		Demonstra domínio da expressão verbal (fluência, dicção, gramática, vocabulário, tom, volume e modulação de voz, ritmo da fala, cadência e pronúncia).
Expressão não verbal				
1	2	3	4	5
Não consegue envolver, frustra e/ou entra em conflito com o paciente.		Demonstra controle suficiente de expressão não-verbal para envolver o paciente, e não se perturba diante de certos comportamentos por ele apresentados como passividade, insegurança ou agressividade inadequadas.		Demonstra elegância e domínio da expressão não verbal (contato visual, gestos, postura, uso do silêncio etc.)
Avaliação geral do conhecimento e das habilidades demonstradas na entrevista				
A = Incompetente		B = Limítrofe		C = Competente
Responde de forma inadequada e inefetiva à tarefa, indicando falta de conhecimento e/ou habilidades interpessoais e para entrevistar mal desenvolvidas.		Responde de forma efetiva a alguns componentes da tarefa, apresenta algum desenvolvimento de habilidades interpessoais e para entrevistar.		Responde de forma precisa e com perspicácia à tarefa, integrando todos os componentes de forma consistente.