RONALDO DELMONTE PIOVEZAN

O RACIOCÍNIO CLÍNICO EM CONTEXTOS DE INCERTEZA: UMA PROPOSTA DE AVALIAÇÃO A PARTIR DE SITUAÇÕES EM GERIATRIA

Tese apresentada à Universidade Federal de São Paulo – Escola Paulista de Medicina, para a obtenção do Título de Mestre em Ciências.

Orientador: Prof. Dr. Nildo Alves Batista
Co-orientadora: Profa. Dra. Maysa Seabra Cendoroglo

SÃO PAULO
2008

PIOVEZAN, Ronaldo Delmonte

O raciocínio clínico em contextos de incerteza: uma proposta de avaliação a partir de situações em geriatria / Ronaldo Delmonte Piovezan. São Paulo, 2008

xiii, 224f.

Tese (Mestrado) -- Universidade Federal de São Paulo. Escola Paulista de Medicina. Centro de Desenvolvimento do Ensino Superior em Saúde.

1. Raciocínio Clínico 2. Teste de Concordância de Scripts

3. Educação Médica.

I. Título

UNIVERSIDADE FEDERAL DE SÃO PAULO

ESCOLA PAULISTA DE MEDICINA

CENTRO DE DESENVOLVIMENTO DO ENSINO SUPERIOR EM SAÚDE

DIRETORA DO CEDESS: Profa. Dra. Maria Cecília Sonzogno

COORDENADOR DO CURSO DE PÓS-GRADUAÇÃO: Prof. Dr. Nildo Alves Batista

RONALDO DELMONTE PIOVEZAN

O RACIOCÍNIO CLÍNICO EM CONTEXTOS DE INCERTEZA: UMA PROPOSTA DE AVALIAÇÃO A PARTIR DE SITUAÇÕES EM GERIATRIA.

Tese apresentada à Universidade Federal de São Paulo – Escola Paulista de Medicina, para a obtenção do Título de Mestre em Ciências.

31 de julho de 2008.

Presidente da Banca: Prof. Dr. Nildo Alves Batista

BANCA EXAMINADORA

Prof. Dr. Joaquim Edson Vieira _____

Prof. Dr. José Antonio Maia de Almeida _____

Prof. Dr. Ricardo Shoiti Komatsu _____

Suplente: Profa. Dra. Maria Cecília Sonzogno _____

Para meu avô,

Por caminhar comigo nos primeiros momentos de vida escolar.

AGRADECIMENTOS

*Aos **professores do Cedess,** pelos ensinamentos e pelo exemplo de dedicação ao ensino.*

*Aos professores **Nildo Alves Batista** e **Maysa Seabra Cendoroglo**, pela imprescindível orientação deste trabalho.*

*Aos **integrantes do Ambulatório de Neuropsiquiatria Geriátrica**, da Disciplina de Geriatria e Gerontologia da UNIFESP, pelo convívio e pela busca de melhores cuidados aos idosos.*

*À **Márcia Menon** e ao **Osvladir Custódio**, pela contribuição e apoio incondicionais a este trabalho.*

*À **Marie**, que soube compreender este nosso momento.*

*À **minha família**, por todos os instantes vividos.*

“Ela não podia comer ou dormir, tornava-se mais magra, tossia, e assim como o médico fez com que eles sentissem, estava em perigo. Eles não podiam pensar em nada, exceto em como ajudá-la. Os médicos vinham vê-la um a um e, em consulta, conversavam em francês, alemão e latim, culpando uns aos outros, e prescrevendo uma grande variedade de medicamentos para todas as doenças conhecidas por eles, mas o simples fato nunca ocorreu a qualquer um deles, de que poderiam não conhecer a doença da qual Natasha estava sofrendo, como nenhuma doença sofrida por um homem pode ser conhecida, pois toda pessoa viva tem sua própria, peculiar, pessoal, insólita, complicada doença, desconhecida para a medicina – não uma doença dos pulmões, fígado, pele, coração, nervos ou mais, mencionada nos livros médicos, mas uma doença consistindo de uma das inumeráveis combinações de injúrias daqueles órgãos. Este simples pensamento não pode ocorrer para os médicos (como não poderia para o mágico que é incapaz de despertar de seus encantamentos), pois o negócio de suas vidas era curar, e eles recebem dinheiro para isto, e tem gasto os melhores anos de suas vidas neste negócio.”

Guerra e Paz, Tolstói.

RESUMO

A avaliação da competência para o raciocínio clínico em situações de incerteza ainda é pouco pesquisada. Os testes escritos mais usados na educação médica são os testes de múltipla escolha. Embora estes sejam capazes de avaliar a capacidade para se lidar com problemas bem definidos, reconhece-se que, na prática, a maioria das situações contém incertezas.

A tomada de decisões nesses contextos é um dos pilares da competência profissional. Compreender como isso se desenvolve pode contribuir com propostas de ensino com ênfase no aprimoramento dessa competência.

Para tanto, a avaliação do raciocínio clínico em contextos de incerteza foi desenvolvida a partir da teoria de scripts, a qual explica o processo de raciocínio clínico, com base na metodologia descrita por Charlin et al. (2000).

Os objetivos dessa pesquisa foram: desenvolver, aplicar e analisar um teste de concordância de scripts (TCS) com temas em geriatria. Um grupo de especialistas formou o painel de referência para a construção do escore do teste. O teste também foi aplicado em um grupo de estudantes de graduação.

Da comparação dos resultados obtidos pelos dois grupos, chegou-se a indícios de validade do instrumento, que foi capaz de diferenciar o raciocínio clínico de acordo com o nível de experiência dos examinandos. As análises de consistência interna e de estudos G forneceram interpretações que se aproximaram da complexidade contida em um escore que busca avaliar a competência profissional. Os coeficientes de alfa de Cronbach e G foram calculados e discutidos segundo as qualidades e as limitações psicométricas dos resultados alcançados.

Com isso, comprovou-se que o teste de concordância de script, com situações em geriatria, desenvolvido em língua portuguesa, em uma instituição de ensino brasileira, pode ser uma alternativa de avaliação do raciocínio clinico em contextos de incerteza.

Palavras-chave: raciocínio clínico, teste de concordância de scripts, educação médica.

ABSTRACT

Assessment of competency for clinical reasoning in contexts of uncertainty is still little studied. The assessment methodologies most applied in medical course are the multiple choice questions. Although these questions are appropriate to evaluate the capacity to solve well defined problems, it is recognized in practice most of situations are under uncertainty.

Decision making in these situations becomes one of the supports of the professional competency. To understand how this process develops can contribute with educational propositions emphasizing the improvement of this competency.

Therefore, a written test for the assessment of the clinical reasoning process in contexts of uncertainty could be helpful. For this purpose, it was developed an instrument of assessment based on the script cognitive theory that explains the clinical reasoning process, according to the description by Charlin et al. (2000).

The purposes of this study were to develop, apply and analyze a script concordance test (SCT) in geriatrics. A group of experts constituted the reference panel for the construction of the test´s score. After that, the instrument was solved by a group of medical students.

Comparing the results for both groups, it was achieved evidences of validation for the approach, which result was able to discriminate the clinical reasoning in agreement with the experience level of the examinees. Internal consistency and G test analyzes gave meanings close to the complexity of a score to measure a professional competency. Cronbach´s alphas and G coefficients were calculated and discussed to support psychometric qualities and limitations of results.

Therewith, it has confirmed the script concordance test with geriatrics situations, developed in Portuguese, at a brazilian educational institution, can be an alternative to the assessment of clinical reasoning in contexts of uncertainty.

Keywords: clinical reasoning, script concordance test, medical education.

SUMÁRIO

1. INTRODUÇÃO..... 1

2. OBJETIVOS 5

3. REFERENCIAL TEÓRICO..... 7

3.1 O Raciocínio, a tomada de decisões e as incertezas em Medicina..... 8

3.2 Construções teóricas sobre o raciocínio clínico..... 9

3.2.1 Teorias normativas..... 11

3.2.2.Teorema de Bayes..... 11

3.2.3 Teorias descritivas..... 12

3.3 A educação médica e os processos de aprendizado do raciocínio clínico..... 14

3.4 A avaliação do raciocínio clínico em contextos de incerteza..... 16

3.5 O teste de concordância de script (TCS)..... 17

3.6 Pesquisas realizadas com o teste de concordância de script (TCS)..... 19

3.7 A geriatria e o seu ensino como elementos de aprendizagem para o desenvolvimento do raciocínio em contextos de Incerteza..... 23

4. METODOLOGIA..... 27

4.1 Tipo de estudo..... 28

4.2 Procedimento ético..... 28

4.3 Princípios do instrumento..... 28

4.4 Desenvolvimento do instrumento..... 29

4.4.1 Princípios de construção..... 29

4.4.2 Casos clínicos..... 30

4.4.3 Escolha do formato de itens..... 31

4.4.4 Outras Considerações sobre a construção dos itens..... 31

4.5 Aplicação em um grupo piloto..... 32
4.6 O Processo de formação dos grupos..... 33
4.7 O Sistema de pontuação..... 33
4.8 Populações integrantes dos grupos..... 34
4.9 Tamanho dos grupos..... 35
4.10 Pesquisa sobre as características dos examinandos..... 36
4.11 Procedimentos para a análise dos dados..... 37

5. RESULTADOS..... 43
5.1 O Processo de construção do instrumento..... 44
5.2 Análise da aplicação em um grupo piloto para ajustes no instrumento..... 46
5.3 Aplicação do teste em um grupo de especialistas..... 47
5.4 Aplicação do teste em um grupo de estudantes..... 48
5.5 Cálculo do escore agregado..... 49
5.6 Análise dos resultados após a formação do escore..... 51
5.6.1 Análises dos resultados após a formação do escore total por itens na diferenciação entre os grupos..... 51
5.6.2 Análise da consistência interna dos resultados do escore por itens..... 52
5.6.3 Comparação dos escores totais por itens entre os grupos após a análise da consistência interna 53
5.6.4 Aplicação do estudo G para os resultados dos itens adequados quanto à consistência interna..... 54
5.6.5 Análise do escore por situações clínicas..... 55

6. DISCUSSÃO..... 57
6.1 Considerações sobre a construção do instrumento..... 58
6.2 Considerações sobre os grupos e a aplicação do instrumento..... 63
6.3 Considerações sobre a validade dos resultados..... 66
6.4 Os resultados em relação às pesquisas anteriores sobre o TCS..... 88
6.5 Considerações sobre a importância de uma avaliação do raciocínio clínico em contextos de incerteza na formação médica atual..... 96
6.6 Perspectivas para novos desenhos de pesquisa sobre o TCS..... 100

7. CONCLUSÕES..... 102

8. REFERÊNCIAS..... 104

9. APÊNDICES..... . 114

10. BIBLIOGRAFIA CONSULTADA..... 219

LISTA DE TABELAS

Tabela 1. Freqüência com que os temas geriátricos foram considerados..... 45

Tabela 2. Número de especialistas do painel de referência, por instituição.. 47

Tabela 3. Frequência das opções de resposta nas questões de 1 a 5..... 49

Tabela 4. Formação do escore nas questões de 1 a 5..48..... 50

Tabela 5. Análise do Escore Agregado do Total de Itens (N=104)..... 51

Tabela 6. Análise do Escore Após Seleção de Itens Viáveis (N=61)..... 53

Tabela 7. Estudo G com desenho P/I para itens..... 54

Tabela 8. Estudo G com desenho P/S para situações..... 56

Tabela 9. Respostas obtidas a partir da aplicação do teste no grupo piloto.. 159

Tabela 10. Respostas obtidas a partir da aplicação do teste no grupo de especialistas..... 161

Tabela 11. Respostas obtidas a partir da aplicação do teste no grupo de estudantes..... 166

Tabela 12. Freqüência das opções de resposta para cada questão, entre os especialistas, para a formação do escore..... 173

Tabela 13. Escores calculados para cada resposta, em cada questão..... 175

Tabela 14. Escores individuais e totais para os especialistas..... 177

Tabela 15. Escores individuais e totais para os estudantes..... 183

Tabela 16. Escore total dos especialistas após seleção de itens por consistência interna..... 192

Tabela 17. Escore total dos estudantes após seleção de itens por consistência interna..... 193

Tabela 18. Escore por situações para os especialistas..... 194

Tabela 19. Escore por situações para os estudantes..... 195

1. INTRODUÇÃO

A educação médica não se atém somente a objetivos técnicos e precisos. A prática profissional exige múltiplas competências. Entre elas, algumas têm seus processos de ensino-aprendizagem pouco compreendidos até o momento. Para a melhor compreensão desse processo, é preciso avaliar o desenvolvimento e o aprimoramento dessas competências.

O desempenho profissional envolve diversos domínios. O conhecimento técnico, a cognição, os aspectos emocionais estão entre eles, sendo que muitos são de difícil mensuração. O reconhecimento desses domínios contribui para a tentativa de definir esse essencial e complexo objeto, ou seja, a competência profissional. Epstein e Hundert (2002, p. 226) assim colocam:

> (...) competência profissional é o uso judicioso e habitual da comunicação, do conhecimento, das habilidades técnicas, do raciocínio clínico, das emoções, dos valores e reflexões na prática diária para o benefício do indivíduo e da comunidade atendida.

Portanto, a competência médica, como elemento que é construído e transformado ao longo de toda vida profissional, envolve aspectos como habilidades clínicas, conhecimento científico, atitudes morais, entre outros. As funções cognitivas participam na formação e desenvolvimento de todos esses aspectos e são úteis como substrato ao processo de aprimoramento dos mesmos.

Essa competência não é estática. Tem como características o desenvolvimento contínuo, não-linear e dependente do contexto. A carreira médica, nesse sentido, contém aspectos evidentes de aquisição continuada do aperfeiçoamento de sua prática.

Quando Morin (2005) propõe um conjunto de sete saberes necessários à formação atual, aprender a enfrentar situações de incerteza está entre eles. Segundo esse autor, o que se ensina normalmente são as certezas. Porém a ciência atual lida invariavelmente com constantes dúvidas. Além disso, as práticas profissionais nem sempre trazem problemas bem definidos.

O raciocínio clínico é uma das principais habilidades médicas, segundo Epstein e Hundert (2002). Porém, não se pode considerar uma competência isoladamente. Todas elas dependem de um contexto. Em medicina, ainda, nem sempre os contextos são bem delineados.

Embora uma parcela da capacidade para o raciocínio profissional recaia sobre o potencial para a aplicação de soluções para problemas bem definidos, reconhece-se que, na prática clínica, muitas situações são mal delimitadas.

Em uma avaliação clínica, nem todos os dados necessários para a tomada de decisões diagnósticas e terapêuticas estão disponíveis. Algumas informações, ainda, podem ser confusas ou imprecisas. Outras também podem estar inconsistentes e sem acurácia.

Essas situações são, portanto, caracterizadas pela incerteza. Mesmo que seja parte do cotidiano profissional, pouca atenção tem sido dada à aprendizagem dessa capacidade para lidar com o impreciso ou imponderável.

Em parte, as dificuldades encontradas no ensino dessa competência podem originar-se na carência de instrumentos capazes de avaliá-la. Seria útil, portanto, a aplicação de um instrumento de avaliação da capacidade de raciocínio clínico em contextos de incerteza.

Os testes escritos mais empregados para avaliar o aprendizado dos estudantes na educação médica são os testes de múltipla escolha. Esses são adequados na avaliação de habilidades envolvidas na resolução de situações clínicas bem definidas, nas quais a aplicação de conhecimentos, normas e princípios estão bem estabelecidos.

A capacidade para a resolução de situações mal definidas e duvidosas, porém, não pode ser avaliada por esses testes. A *avaliação padronizada do raciocínio em contextos de incerteza*, descrita por Charlin e Vleuten (2004), e baseada no método de concordância de script, parece ser uma alternativa para analisar a tomada de decisões nessas situações, trazendo novo enfoque na avaliação de competências na área médica.

Esse instrumento tem sido desenvolvido em diversos ambientes educacionais, em diferentes áreas e especialidades de atuação médica, em diversos países e línguas. A estrutura do instrumento se baseia na apresentação escrita de casos clínicos, seguidos de opções de escolha sobre decisões diagnósticas e terapêuticas, em que o formato de resposta é uma escala do tipo Likert (LIKERT, 1932), a qual reflete como uma informação é processada em situações de tomada de decisão, de acordo com a teoria de scripts. A mensuração dos resultados do teste leva em conta a variabilidade do processo de resolução entre especialistas na área de aplicação das questões.

Particularmente no cuidado a idosos, situações duvidosas envolvidas em decisões diagnósticas, terapêuticas e em dilemas éticos são rotineiras. A avaliação geriátrica leva em conta os aspectos globais da saúde, buscando atingir a complexidade e a integralidade necessárias ao cuidado à terceira idade.

Os processos de ensino-aprendizagem e a formação médica, quando recebem a contribuição de tópicos em geriatria, tornam-se importantes quanto ao estudo dos elementos envolvidos no desenvolvimento de competências para a tomada de decisões em circunstâncias de incerteza.

Sendo assim, como médico especializado em geriatria, com atuação no ensino dessa área nos cursos de graduação e pós-graduação em medicina da Universidade Federal de São Paulo (UNIFESP), motivei-me pela curiosidade e pelo desafio em buscar respostas à seguinte questão:

> Como elaborar um instrumento de avaliação do raciocínio clínico em situações de incerteza utilizando-se de questões relacionadas à prática geriátrica, em diferentes momentos de formação médica?

Acredito que a elaboração de um instrumento padronizado e validado a partir dessa pergunta possa aprofundar o conhecimento a respeito do ensino voltado para competências na educação médica.

2. OBJETIVOS

2.1 Objetivo Geral:

Desenvolver, aplicar e analisar um instrumento de avaliação do raciocínio clínico em situações de incerteza, com enfoque em tópicos relacionados à geriatria, em língua portuguesa, baseado no teste de concordância de scripts.

2.2 Objetivos Específicos:

1. Construir um conjunto de testes, baseando-se na teoria de scripts, a partir de um grupo de especialistas em geriatria, para avaliar o raciocínio clínico em situações de incerteza;

2. Investigar as diferenças encontradas nas respostas dadas ao teste em dois momentos de formação, entre especialistas e estudantes em final de graduação, segundo os pressupostos teóricos da teoria de scripts, sobre o desenvolvimento do raciocínio clínico.

3. REFERENCIAL TEÓRICO

3.1 O raciocínio, a tomada de decisões e as incertezas em medicina

A literatura científica referente ao raciocínio clínico desenvolve-se há mais de trinta anos. Porém, as questões sobre esse assunto intrigaram médicos e professores desde os primórdios da medicina. Hipócrates, considerado o pai da medicina, mantém-se atual em alguns de seus princípios. Assim ele se pronuncia em seu primeiro aforismo (p. 45):

> “A vida é curta, e a Arte é longa; a ocasião fugidia; a experiência, enganadora, o julgamento, difícil. “

Mais de dois mil anos depois, outro médico e professor, Willian Osler, considerado um dos maiores médicos da clínica médica moderna, refletiu da seguinte maneira (DANS, 1993):

> “A medicina é a ciência da incerteza e a arte da probabilidade; erros de julgamento têm de ocorrer na prática de uma arte que consiste, largamente, em balancear probabilidades.”

Por isso, a compreensão do processo de raciocínio clínico, que leva à tomada de decisões, torna-se um elemento valioso dentro da pesquisa em educação médica. Entendê-lo faz parte dos esforços para o ensino de competências em situações práticas, as quais frequentemente são carregadas de dúvidas e riscos.

Diz-se na clínica que o médico é tão bom quanto a sua última decisão (GILLIES e SHEEHAN, 2002). Mas como melhor decidir diante de tantos pacientes diferentes, em momentos sempre singulares? As escolhas práticas ultrapassam o domínio técnico e podem percorrer campos abrangentes e complexos, dentro de áreas como a ética, a psicologia e as questões sociais. As situações contêm elementos indeterminados e imprecisos.

O advento atual da medicina baseada em evidências trouxe significativos avanços na tomada de decisões diagnósticas e terapêuticas. Porém, evidências são aplicadas para doenças e não para doentes. Para populações, mas não para indivíduos, propriamente. O indivíduo, considerado em sua completude, é corpo, mente e ser social. As decisões, ao final, são sempre individualizadas. Diante de um indivíduo único, cada situação é inédita.

A singularidade do processo decisório não exclui, porém, a validade do método de interpretação de dados de uma situação clínica. O encontro dos elementos envolvidos no raciocínio clínico e a descoberta da forma como os estudantes e profissionais pensam sobre determinados problemas contribui para a compreensão da aprendizagem para a tomada de decisões.

3.2 Construções teóricas sobre o raciocínio clínico

Segundo Barrows e Tamblyn* (1980), o raciocínio clínico pode ser definido como o processo cognitivo que é necessário para avaliar e manejar o problema médico do paciente (citado por Round, 2001).

Para se pensar clinicamente sobre uma hipótese diagnóstica, por exemplo, seria necessário o conhecimento sobre as doenças que levam àquele quadro. Os dados estatísticos sobre cada doença poderiam auxiliar, ainda, nos testes diagnósticos.

Quando o médico analisa um sintoma ou sinal clínico em um paciente, ele pensa inicialmente nas possíveis causas daquele sintoma ou sinal. A melhor estratégia diagnóstica e terapêutica, em termos de eficácia, segurança e custos, não pode ser empregada sem este primeiro raciocínio.

---

* Barrows HS, Tamblyn RM. Problem Based Learning: an Approach to Medical Education. Springer, New York, 1980 apud Round A. Introduction to clinical reasoning. Journal of Evaluation in Clinical Practice. 2001; 7 (2): p. 109-117.

Seria importante também o conhecimento da freqüência com que cada doença pode levar àquele quadro clínico, pensando-se na probabilidade da mesma para o paciente avaliado. Entretanto, os médicos não parecem pensar rotineiramente dessa maneira. De acordo com Round (2001), a primeira doença que vêm à mente do médico é que vai ser considerada primordialmente.

Diferentes áreas do conhecimento tentam explicar o raciocínio clínico. Entre elas estão a psicologia, a prática clínica e a educação médica. Dentro das proposições que tentam explicar o processo de raciocínio, existem as teorias normativas ou prescritivas e as teorias descritivas. Ambas as linhas levam mais em conta a razão do que a experiência na tomada de decisões.

Todavia, há dúvidas quanto à validade das teorias normativas para a interpretação do processo de tomada de decisões. A experiência pode ser mais relevante do que a razão. Sendo assim, a formulação de hipóteses para a resolução de um problema partiria mais das vivências anteriores dos indivíduos e, portanto, não seriam as mesmas para cada um deles. Grant e Marsden (1987) encontraram variações individuais na capacidade para a resolução de situações clínicas.

Porém, Neufeld et al. (1981) não observaram diferenças no processo de resolução entre estudantes e profissionais. De qualquer forma, o aprendizagem a partir da prática parecia ser relevante.

Hobus et al. (1987) trouxeram uma outra explicação para o processo de raciocínio clínico. Para este autor, tanto a razão quanto a experiência são importantes. A resolução de um problema parece depender do conteúdo e do contexto envolvidos na situação a ser solucionada. Ou seja, o raciocínio clínico parece ser caso-específico.

3.2.1 Teorias normativas

Von Neumann e Morgenstern (1947) propuseram a teoria da decisão. Segundo esses autores, os seres humanos agem para maximizar o valor esperado, seja em termos financeiros, de felicidade, de sucesso, entre outros, em suas decisões.

Esse pressuposto se baseia em proposições racionais da tomada de decisão. Porém se limitam a uma escolha única. E na medicina, raramente as decisões são isoladas. Os médicos lidam com situações seqüenciais, em que múltiplas decisões são feitas para a resolução de um problema. Como o cérebro humano não pode lidar com muitos números ao mesmo tempo, uma alternativa para decisões complexas seria a fragmentação do problema para que fossem realizadas escolhas menores.

Parte da ênfase colocada sobre essa teoria se deve a achados feitos por Elstein et al. (1986), que mostraram que mesmo os especialistas utilizam princípios da mesma em suas decisões. Os clínicos fazem avaliações não tão rigorosas sobre as probabilidades. Apesar disso, costumam ser excessivamente confiantes quanto a seus cálculos, como concluiu Ridderikhoff (1993).

3.2.2. Teorema de Bayes

O teorema de Bayes (BERNARDO & SMITH, 1993) oferece uma proposta normativa e prescritiva para o processamento de informações. Ele considera que o conhecimento pode ser representado como uma hipótese. Cada uma delas é baseada em uma probabilidade subjetiva. Isso acontece por que o conhecimento que fundamenta uma hipótese é variável. Cada médico possui um conhecimento diferente a respeito do mesmo problema.

Dessa maneira, a probabilidade pré-teste (antes da realização de um exame, por exemplo), seria a prevalência conhecida da doença ou a impressão subjetiva sobre a probabilidade da doença, antes de uma nova informação ser adquirida.

Já a probabilidade pós-teste pode ser considerada a possibilidade de existência da doença após uma nova informação. Sob o ponto de vista puramente matemático, essa última probabilidade é função de duas variáveis, provenientes da probabilidade antes de um exame, e de uma evidência, oferecida por um teste. É, portanto, uma razão de probabilidades.

Entretanto, parece evidente que esse teorema tem limitado uso formal na prática. Opiniões informais ainda são as mais comuns na rotina clínica, segundo Wolf e Gruppen (1985).

3.2.3 Teorias descritivas

A partir de observações de atendimentos realizados por médicos ou pela análise dos recordatórios que descrevem o processo de pensamento durante uma entrevista médica, foi possível a elaboração de algumas conclusões a respeito do raciocínio clínico. Segundo Barrows e Bennett (1982), as hipóteses diagnósticas são geradas logo no início da entrevista, ocorrendo uma rápida mudança da coleta de dados para um teste de hipóteses durante a mesma.

A partir dessas interpretações, pode-se dizer que o pensamento humano usa um método de processamento de informações. Mais ainda, o processo de raciocínio poderia ser revelado pela verbalização e que o pensamento mais recente poderia ser acessado e verbalizado.

Porém, essas duas últimas conclusões não foram confirmadas por Ridderikhoff (1993). Parece que, pela limitada memória de trabalho do cérebro humano, os dados devem ser processados seriadamente, segundo um plano, em um contexto que considere os conhecimentos prévios já armazenados.

O médico parece precisar de um modelo simplificado do problema clínico. Após formular esse modelo em sua mente, ambos, razão e experiência, caminharão juntos para a resolução da situação. Neame et al. (1985), consideraram que um processo diagnóstico eficiente recai sobre um pensamento bem organizado e uma boa base de conhecimento prévio.

Elstein et al. (1978), baseando-se na hipótese de estratégia hipotético-dedutiva do raciocínio, descreveram um outro método de processamento de informações. Segundo esse autor, um número de hipóteses é gerado a partir do problema clínico inicial. Essas hipóteses guiam as novas perguntas feitas aos pacientes. As novas informações geradas pelas respostas a essas perguntas vão confirmando ou refutando as hipóteses.

As hipóteses geradas podem ser testadas de diferentes maneiras. As estratégias de teste de hipóteses podem ser de rastreamento ou de focalização. O rastreamento simultâneo ocorre quando se usa uma mesma informação para o teste de várias hipóteses ao mesmo tempo. O rastreamento sucessivo leva em conta que as hipóteses são testadas separadamente e de maneira sucessiva.

O método de focalização conservadora, ao contrário dos métodos de rastreamento, irá construir as hipóteses a partir dos dados, não o inverso, como é feito no rastreamento. Já o foco em jogo ou em aposta é considerado quando se testa um palpite. Por último, o foco negativo trabalha inicialmente com um grande número de hipóteses, mas as vai excluindo sequencialmente.

Parece que o rastreamento simultâneo é a estratégia mais eficiente. O foco em aposta é a pior das estratégias. Neame et al. (1985), encontraram alguns dados que indicam que os estudantes usam mais a focalização conservadora. Nenhum estudante usou a melhor estratégia – o rastreamento simultâneo, pois esta estratégia provavelmente requer uma base de conhecimento mais sólida do que aquela que eles possuíam.

Hoje, há clara evidência de que o raciocínio clínico difere entre novatos e médicos experientes. A quantidade de informação coletada é variável de forma inversa à experiência do indivíduo, conforme dados encontrados por Mcguire (1985). Os especialistas e os novatos parecem processar as informações de maneiras diferentes, segundo Grant e Marsden (1987).

Os especialistas usam com freqüência uma técnica de raciocínio para o reconhecimento de padrões, o que depende de uma base sólida e extensa de conhecimento, concluiu Schmidt et al. (1990). A combinação de informações geralmente não segue o Teorema de Bayes. Eles parecem lembrar especificamente de informações relevantes apenas para o processo diagnóstico, segundo conclusões de Hassebrock et al. (1993).

3.3 A educação médica e os processos de aprendizagem do raciocínio clínico

Tradicionalmente, a aprendizagem inicial nas escolas médicas é orientada, em semiologia, para a coleta sistemática, organizada e exaustiva de informações, dentro dos princípios técnicos da anamnese e do exame físico, para a construção do raciocínio clínico.

Entretanto, como essa estratégia é semelhante ao método de focalização conservadora, citado anteriormente, acredita-se que ele é pouco eficaz. Os clínicos experientes raramente tomam decisões a partir de processos como esse.

A partir das teorias normativas, acredita-se que um bom raciocínio clínico se fundamenta em uma grande base de conhecimentos, além de um uso regular dos princípios do Teorema de Bayes. Porém, como já se discutiu anteriormente, tanto o ensino como a aplicação desses princípios não persiste na prática médica, mesmo entre os especialistas.

Ainda, quando se consideram as teorias descritivas, vê-se que estas têm maior impacto na construção de algumas propostas curriculares. A aprendizagem baseada em problemas (“problem-based learning” - PBL) parte da proposta de raciocínio hipotético-dedutivo. Porém as implicações práticas dessa estratégia de ensino no processo de raciocínio dos estudantes ainda não estão bem definidas, segundo Jayawickramarajah (1996).

Já as teorias de aquisição de habilidades consideram que é difícil ensinar os processos de raciocínio quando os alunos não têm conhecimentos e experiência prática suficientes. Reconhece-se cada vez mais que o número, tipo e diversidade de experiências práticas durante o período curricular são essenciais para o desenvolvimento e aprimoramento do raciocínio para a tomada de decisões.

Ainda há questionamentos quanto ao papel do método hipotético-dedutivo e da aprendizagem baseada em problemas (PBL) no desenvolvimento do raciocínio clínico. Vernon e Blake, em 1993, demostraram que o PBL pode melhorar o desempenho clínico em relação aos estudantes do grupo controle, porém não demonstrou diferenças na aquisição de conhecimentos. Uma revisão sistemática mais recente (CHOON-HUAT KOH et. al., 2008) concluiu que a competência para lidar com incertezas foi uma das características cognitivas positivamente associadas com o PBL.

Todavia é importante reconhecer, por final, que os resultados em testes que medem desempenho clínico não refletem em capacidade de raciocínio na prática do dia-a-dia. Norman (1985) observou que médicos que alcançam resultados satisfatórios em testes de resolução em pacientes simulados não atingiram os mesmos resultados quando os pacientes foram introduzidos em sua própria prática profissional.

Os instrumentos mais recentes de avaliação da competência clínica também ainda não foram capazes de diferenciar a capacidade para a resolução de problemas entre médicos com diferentes níveis de experiência. Conforme os

achados encontrados por Hodges et al. (1999), o Exame Clínico Objetivo e Estruturado (OSCE), através de seu método de pontuação binário, não foi capaz de avaliar adequadamente as habilidades necessárias para uma correta avaliação clínica e não diferenciou médicos experientes de novatos.

3.4 Avaliação do raciocínio clínico em contextos de incerteza

A capacidade para pensar em situações incertas e para tomar decisões para a resolução de problemas com dados insuficientes é um dos pilares da competência profissional.

A maioria dos métodos de avaliação são medidas de conhecimento ou comportamento. Tanto os testes de múltipla escolha quanto os exames clínicos objetivos e estruturados ("Objective Structured Clinical Exams" - OSCEs) podem avaliar esses elementos da competência médica.

A mensuração da capacidade cognitiva nas profissões não pode ser feita pelos métodos de avaliação tradicionais. Mais do que aferir os resultados desse processo cognitivo, é preciso considerar a maneira como ele se desenvolve.

A adaptação da teoria de script, da psicologia cognitiva, realizada por Charlin et al. (2000) e Schmidt et al. (1990), às características do raciocínio nas profissões, permitiu a construção de um instrumento de avaliação em situações de incerteza.

Segundo Nelson* (1986), a definição de script consiste em uma estrutura de conhecimento direcionada a um objetivo, adaptada para a resolução de tarefas de forma eficiente (citado por Charlin, 2004).

---

* Nelson K, Event knowledge: structure and function in development. Hillsdale, NJ: Lawrence Erlbraum apud Charlin B, Vleuten CVD. Standardized Assessment of Reasoning in Contexts of Uncertainty: The Script Concordance Approach. Eval Heal Prof. 2004; 27 (3): p. 304-319.

Os scripts começam a aparecer no raciocínio clínico quando os estudantes de medicina começam a ser desafiados por tarefas reais, de acordo com Schmidt et al. (1990). Esses scripts podem ser desenvolvidos e aprimorados ao longo de toda a vida profissional.

A teoria dos scripts coloca que, quando um profissional é exposto a uma situação, ele ativa os scripts necessários para a resolução dessa situação. Essa geração de processos cognitivos é útil na formulação de hipóteses diagnósticas e na escolha terapêutica, por exemplo.

Ainda com relação a essa teoria, o raciocínio é construído através de julgamentos qualitativos. Esses julgamentos poderiam ser mensurados e comparados a um painel de especialistas, o qual seria uma referência. Esse método de avaliação foi denominado teste de concordância de scripts, conforme citação de Charlin et al (2000).

3.5 O teste de concordância de scripts

Um teste educacional engloba algumas características essenciais que devem ser analisadas. Segundo Norman et al (1996), primeiramente é preciso saber qual a tarefa exigida pelo teste. Em segundo, como as respostas são obtidas e, por último, como é definida uma pontuação e uma escala de valores para o teste.

As tarefas da avaliação precisam ser desafiadoras para qualquer nível de formação profissional. No caso do raciocínio clínico em situações de incerteza, o teste é elaborado a partir de uma situação prática, através da descrição por escrito de um caso clínico.

Considera-se que o teste representa um desafio, mesmo para os especialistas, pois o caso clínico, como uma situação que representa a realidade, não deve apresentar todos os dados necessários para a solução do problema,

como uma situação diagnóstica ou de tratamento, ou porque todas as atitudes ou decisões apresentadas podem ter prós e contras, como em um teste envolvendo dilemas éticos, por exemplo.

O formato de respostas está de acordo com as teorias sobre o processo de raciocínio clínico, segundo colocação de Grant e Marsden (1988). Uma escala do tipo Likert é empregada para a pontuação dos julgamentos realizados, conforme descrição de Charlin et al. (2000).

O método de pontuação empregado leva em conta a variação de respostas obtidas por um grupo de jurados, como uma espécie de ranqueamento agregado, conforme descrição feita por Norman (1985).

De acordo com este último autor, a técnica de escore agregado é um método criado para avaliar resolução de problemas, inicialmente em um encontro com um paciente simulado. Esse método já fora analisado anteriormente quanto à sua confiabilidade e à sua validade.

Como uma estratégia objetiva de avaliação de resolução de situações clínicas, esta forma de pontuação demonstrou-se adequada. As soluções em diagnósticos, investigações ou estratégias terapêuticas não se enquadram simplesmente em diâmetros absolutamente opostos, como certo ou errado.

Contrariamente, o processo de resolução de problemas clínicos abrange uma série de alternativas de diagnóstico ou intervenção, que podem ser mais ou menos apropriadas. O escore agregado lida de forma objetiva com essa complexa variabilidade.

O formato do item pode variar de acordo com o objetivo de avaliação, ou seja, em caso de avaliação de competências para o diagnóstico, investigação, tratamento ou atitudes. Primeiramente há uma descrição de um caso clínico.

Cada item do teste é composto de três partes, sendo que a primeira é formada por uma hipótese diagnóstica, uma ação de investigação ou uma opção de tratamento (CHARLIN et al. 2004).

A segunda parte apresenta uma nova informação, como um sinal clínico, uma condição, um exame de imagem, um teste de laboratório, que sejam relevantes para a resolução das situações clínicas.

A terceira parte é uma escala do tipo Likert, que pontua as opções de resposta assinaladas. O método de escore agregado demonstra a variabilidade de respostas oferecidas por especialistas no painel de referência. O crédito para cada resposta é igual ao número de membros do painel de referência que ofereceu aquela resposta, dividido pelo valor modal para cada item. Dessa forma, a pontuação máxima em cada item será igual a um, e a mínima, igual à zero. A pontuação total para o teste resulta da soma de créditos obtidos em todos os itens. Esta pontuação pode ser dividida pelo número de itens e multiplicada por 100, obtendo-se, assim, a pontuação em forma de porcentagem de aproveitamento.

3.6 Pesquisas realizadas com o teste de concordância de scripts (TCS)

Essa forma de avaliação tem sido testada de diversas maneiras, na busca de dados de acurácia ou precisão, validade, aplicabilidade e viabilidade de seus resultados.

Inicialmente, buscou-se encontrar o poder discriminativo do teste nos diversos níveis de experiência profissional. Esse dado é importante para a demonstração da aplicabilidade e da validade do instrumento, já que comprovaria que esse método pode eliminar o denominado efeito intermediário.

É chamado de efeito intermediário de uma avaliação o fato de clínicos experientes obterem resultados semelhantes ou até inferiores em testes de múltipla escolha, por exemplo, quando comparados com residentes em final de treinamento. Ou seja, esse resultado demonstra que este tipo de exame não é capaz de avaliar o grau de experiência do profissional.

Um estudo publicado por Charlin e colaboradores em 1998 demonstrou a validade do teste de concordância de scripts (TCS) em diferenciar profissionais de acordo com seus níveis de experiência clínica.

A validade do processo de ranqueamento do teste foi visto em um estudo que aplicou a avaliação na área de ginecologia, usando os resultados obtidos por 150 estudantes e 14 especialistas. Com isso, analisou-se o valor do método de pontuação agregada, proposto por Norman (1985).

Para se alcançar o efeito de variabilidade de respostas entre especialistas, duas formas de obtenção de respostas foram propostas. Primeiramente, os especialistas responderam o teste individualmente. Um ano depois, as respostas ao mesmo teste foram obtidas por uma resolução em grupo, ou seja, através de um consenso entre os especialistas.

A primeira conclusão desse estudo é que os especialistas respondem o teste de forma diferente, de acordo com o contexto, já que 95% das respostas consensuais foram diferentes das individuais. A interação entre colegas, portanto, pode modificar o raciocínio clínico.

Como resultado dessa diferença, foi possível criar dois ranqueamentos entre especialistas, sendo um o resultado do conjunto de respostas individuais, ou seja, pelo método agregado, diferente do escore consensual. Entretanto, somente o ranqueamento agregado foi capaz de mostrar a diferença de resultado entre os testes respondidos por estudantes e especialistas. Portanto, essa estratégia de ranqueamento foi a única validada.

A validade preditiva do TCS foi demonstrada através de um exame realizado em médicos residentes em medicina de família no Canadá. Brailovsky et al. (2001), fizeram uma comparação entre os resultados de testes aplicados para avaliar o raciocínio clínico em dois momentos diferentes da formação dos 24 residentes. Os resultados demonstraram que a capacidade para organizar e aplicar conhecimentos para o raciocínio clínico se manteve ao longo da residência médica.

A estabilidade do teste em dois ambientes de aprendizado diferentes foi analisada em uma pesquisa que comparou o desempenho entre estudantes e residentes e especialistas em um exame que seguia a metodologia do TCS, usando assuntos da área urológica.

Esse exame foi aplicado inicialmente em língua inglesa, no Canadá. Em seguida, foi traduzido e validado também, segundo um painel de referência francês, em língua francesa, para ser aplicado em uma universidade na França. As diferenças encontradas entre os grupos se correlacionaram nos dois países onde o teste foi aplicado, comprovando sua validade através de diferentes ambientes de aprendizado.

A avaliação de como se lida com dilemas éticos na profissão médica é difícil de ser feita através de testes objetivos e padronizados. O julgamento ético consiste em uma dimensão complexa da competência profissional. Recente pesquisa, entretanto, tentou aplicar os recursos do TCS para avaliar as respostas a assuntos controversos em decisões terapêuticas.

Llorca* (2003) construiu um instrumento através da apresentação de um caso clínico, seguido de seis atitudes que poderiam ser tomadas. As respostas eram realizadas em uma escala tipo Likert. Além dos estudantes e residentes examinados, quatro painéis de referência foram obtidos, sendo o primeiro formado de membros do comitê de ética de um hospital universitário, o segundo por professores com interesses em ética, o terceiro por professores em geral e, o último, por médicos de família (citado por Charlin et al., 2004).

Os painéis de referência apresentaram resultados semelhantes. Porém as respostas do painel formado por médicos de família foram mais semelhantes às respostas oferecidas pelos residentes, a qual foi mais técnica. Quando se obteve uma pontuação a partir das respostas dos painéis de referência, os residentes demonstraram um desempenho superior àquele dos estudantes.

A partir desse estudo, uma nova opção foi proposta na busca de alternativas válidas e viáveis para a avaliação de itens tão complexos dentro das dimensões de competência, como a capacidade de julgamento em dilemas éticos.

---

* Llorca G, Evaluation de résolution de problème mal definis en éthique clinique: variation des scores selon les methods de correction et selon les caracteristiques des juris. Pédagogie Médicale, 2003; 4: p. 80-88 apud Charlin B, Vleuten CVD. Standardized Assessment of Reasoning in Contexts of Uncertainty: The Script Concordance Approach. Eval Heal Prof. 2004; 27 (3): p. 304-319.

Outra etapa de validação de construção do teste, sob o ponto de vista cognitivo, foi alcançada por Gagnon et al. (2006), através de uma pesquisa que avaliou a diferença de tempo de processamento de novas informações, de acordo com o fato de elas serem típicas ou atípicas, e compatíveis ou incompatíveis com a hipótese diagnóstica proposta.

Esse objetivo adveio do pressuposto teórico de que os scripts, quando ativados, contêm expectativas em características que são correlacionadas com cada doença e com uma variabilidade de valores que podem ser típicos, atípicos ou incompatíveis para aquela doença.

Sendo assim, os autores documentaram, nessa pesquisa, que o processamento de informações tem velocidades diferentes de acordo com o fato de uma nova informação ser típica, atípica ou incompatível com o script ativado. Para isso, foram recrutados dois grupos de participantes (trinta estudantes e trinta geriatras), que receberam 64 casos clínicos com temas em geriatria acompanhados de hipóteses diagnósticas geradoras de scripts.

A partir de cada hipótese, em cada item, uma nova informação foi apresentada. Os participantes, então, teriam de decidir se aquela nova informação aumentava, diminuía ou não afetava a probabilidade para a hipótese diagnóstica. A velocidade de processamento de cada item foi aferida através da aplicação do teste através de um programa de computador.

Como resultado, constatou-se que as informações típicas são processadas mais rapidamente do que as atípicas, sendo que as incompatíveis estiveram em uma velocidade intermediária. Com isso, foi possível predizer qual tipo de informação é processada mais rapidamente em termos de compatibilidade com as hipóteses formuladas. Esse resultado demonstrou que o teste de concordância de scripts está em acordo com a teoria cognitiva que o fundamenta.

3.7 A geriatria e o seu ensino como elementos de aprendizagem para o raciocínio em contextos de incerteza

O envelhecimento traz desafios e perspectivas para a prática médica. A população idosa dobrou nos últimos 50 anos, chegando a 9,1% da população. Desse modo, as principais causas de morte passaram a ser as doenças crônico-degenerativas, que acompanham os idosos em média por 20 anos de suas vidas (IBGE, 1999).

O debate que recai sobre as atividades pedagógicas traz uma série de preocupações a respeito do ensino médico nas escolas brasileiras. Segundo o Relatório Geral do CINAEM (1997), a falta de interações que favoreçam a aprendizagem, a fragmentação do conhecimento, a dicotomia teoria-prática, a memorização de dados, entre outros, são elementos persistentes nas escolas médicas (citado por Batista e Silva, em 1998).

Entretanto, reconhece-se cada vez mais a necessidade de mudanças. As conferências mundiais de educação médica em Edimburgo em 1988 e 1993, e as conferências internacionais de promoção de saúde (Otawa, 1986; Adelaide, 1988: Sundswall, 1991, Jacarta, 1997; México, 2000) trazem desafios para a construção de uma nova proposta de formação médica, buscando o modelo da integralidade na assistência (LAMPERT, 2003).

Assim, o processo de ensino-aprendizagem em geriatria conjuga-se a um novo perfil de competências médicas. A preeminência da função, a natureza multidisciplinar do cuidado, o atendimento fora de ambientes hospitalares e a participação de pacientes e familiares nas decisões diagnósticas e terapêuticas são alguns dos princípios geriátricos que ultrapassam o domínio atual dessa especialidade, fazendo parte do acompanhamento médico em geral (GILL, 2002).

Em contrapartida, segundo Morin (2005, p.18):

> ”Nossa formação escolar e, mais ainda, a universitária nos ensina a separar os objetos do contexto, as disciplinas umas das outras, para não ter de relacioná-las.”

Sem contextualização, há um prejuízo na reconstrução dos saberes, o que traz dificuldades para a aprendizagem significativa. Recursos necessários à prática profissional, como a capacidade de compreender o ser humano e de lidar com incertezas também são difíceis de serem ensinados.

Com exceção das áreas ligadas à pediatria e à obstetrícia, os profissionais da área de saúde lidam cada vez mais com a figura do idoso. Seu número absoluto deve dobrar nos próximos vinte e cinco anos.

Com essas transformações demográficas e epidemiológicas, alguns tópicos relacionados ao cuidado ganham ainda mais relevância, como a cronicidade das doenças e dos doentes, incapacidade funcional, perda de qualidade de vida, iatrogenias, fragilidade, medidas paliativas, acompanhamento extra-hospitalar e participação de familiares em decisões diagnósticas e terapêuticas.

Todas essas situações criam uma outra perspectiva de acompanhamento aos pacientes, baseada na busca da integralidade da assistência. A funcionalidade do doente, a multidisciplinaridade do cuidado, entre outros princípios da atividade geriátrica, ultrapassam os domínios dessa área e passam a contribuir para a superação das fronteiras de nossos padrões de atendimento em qualquer área de atuação.

Portanto, o que define a medicina geriátrica não é a faixa etária dos pacientes atendidos, mas sim o intensivo foco do cuidado para a preservação e restauração da funcionalidade. Raciocinar clinicamente, sob o ponto de vista geriátrico, é uma outra maneira de olhar o paciente com incapacidades e doenças crônicas, buscando independência e qualidade de vida.

Embora baseada em uma população de outro país, uma pesquisa realizada por Freudenheim (1996), sugere que somente um quarto dos pacientes com doenças crônicas são idosos. Entre aqueles que precisam de auxílio pessoal ou assistência domiciliar, 40% estão abaixo dos 65 anos de idade (citado por Gill, 2002).

Além disso, é importante considerar o quanto a cronicidade de doenças está ligada à piora da funcionalidade. Entre as nove doenças crônicas mais comuns, oito delas levaram a perdas funcionais em diversos domínios (físico, social e mental), segundo Stewart et al. (1989).

Sendo assim, o processo de aprendizagem a partir da geriatria, com novos olhares sobre a assistência aos processos de saúde e doença em condições crônicas, com foco especial à funcionalidade, ultrapassa os limites de uma determinada faixa etária.

Pelo que foi exposto e pelo caráter integral do cuidado oferecido pela geriatria, muitas decisões nessa área consideram situações com alto nível de complexidade. Como necessidade, portanto, a interdisciplinaridade é um pressuposto a ser sublinhado.

Importante se torna, ainda, a participação de pacientes e familiares nas decisões diagnósticas e terapêuticas, que podem não levar à risco de morte, mais que na maioria das vezes leva a conseqüências persistentes ou permanentes sobre a vida dos enfermos ou de seus cuidadores.

As escolhas quanto a realizações de exames diagnósticos ou tratamentos, muitas vezes mais objetivas e técnicas em indivíduos jovens ou hígidos, podem necessitar de outras considerações quando o paciente possui múltiplas comorbidades ou é idoso. O prognóstico e os resultados dessas intervenções podem ser mais incertos ou imprevisíveis em pacientes crônicos ou naqueles com idades mais avançadas.

Mais difíceis ainda podem ser as decisões geriátricas que envolvem aspectos éticos. Situações de vida ou de morte, decisões de prolongamento artificial de condições clínicas irreversíveis ou fora dos recursos de cura são cada vez mais comuns em diversas áreas médicas, de uma maneira geral, e são uma rotina para o geriatra, em particular. Para tais decisões, os aspectos técnicos são apenas uma parte do problema. A complexidade que envolve a conjunção de elementos emocionais, afetivos, culturais, entre outros, precisa ser considerada para uma boa tomada de decisões.

Sendo assim, decidir em situações de incerteza faz parte do cotidiano das atividades em geriatria, em seus diversos ambientes, como os hospitais, ambulatórios, instituições de longa permanência ou atividades assistenciais domiciliares. Os processos de ensino-aprendizagem nessa área, portanto, criam oportunidades diversas para o desenvolvimento do raciocínio clínico.

4. METODOLOGIA

4.1 Tipo de estudo

Este estudo apresentou um desenho de coleta de dados transversal e análises de naturezas quantitativas.

4.2 Procedimento ético

Alguns cuidados foram adotados para o adequado desenvolvimento desta pesquisa e devido respeito aos sujeitos envolvidos:

- O projeto foi submetido ao Comitê de Ética em Pesquisa (CEP), na Universidade Federal de São Paulo, em protocolo número 0406/07 (apêndice 7);

- O pesquisador esclareceu os motivos da pesquisa e os sujeitos tiveram o anonimato preservado, bem como o sigilo das respostas;

- Todos os participantes, em todas as etapas, ou seja, desenvolvimento, piloto e aplicação, da pesquisa concordaram em assinar o Termo de Consentimento Livre e Esclarecido (apêndice 6).

4.3 Princípios do instrumento

O teste foi baseado na teoria de concordância de script, que, segundo Charlin et al. (2000), consiste em apresentar aos examinandos uma série de problemas, na forma de descrições de casos clínicos breves, e, então, questioná-los quanto a elementos de diagnósticos, investigações ou decisões terapêuticas, após a apresentação de uma nova informação em cada item ou questão.

O sistema de pontuação do teste foi elaborado com o objetivo de aferir a concordância entre as respostas dos estudantes e os dos especialistas, sendo que estes últimos formaram o painel de referência.

4.4Desenvolvimento do instrumento

4.4.1 Princípios de construção

Gagnon et al. (2006) e Charlin et al. (2000), recomendaram que o número de especialistas para a colaboração no desenvolvimento dos testes seja pequeno, em geral em torno de duas ou três pessoas. Por isso, foram considerados três geriatras vinculados à UNIFESP, que estão diretamente ligados à proposta de ensino dessa disciplina na graduação, com especialização ou residência médica em geriatria e com título de especialista nesta área, para compor o grupo responsável pela elaboração do instrumento. Este grupo foi coordenado pelo pesquisador.

Primeiramente, em uma entrevista, foram expostos a estes especialistas os princípios e a estrutura do teste de concordância de scripts. Em seguida, foi solicitada aos mesmos a presença em uma seqüência de encontros, quando deveriam trazer, individualmente, a elaboração de uma proposta por escrito que incluísse tópicos em geriatria e que fossem relevantes para a construção de situações clínicas problemáticas, mal definidas ou duvidosas.

Para cada uma dessas situações, foram especificados os seguintes tópicos: hipóteses diagnósticas consideradas, estratégias de investigação e opções de tratamento. Além disso, cada um deles deveria especificar as perguntas que eles fariam para solucionar cada problema, o exame físico que realizariam e o que esperariam para cada hipótese, além dos exames que solicitariam. Por final, deveriam especificar também as informações clínicas, positivas ou negativas, que eles buscariam em cada caso.

A partir do material recolhido nessa fase, foram construídos os itens do teste. De acordo com os coeficientes de alfa de Cronbach encontrados em estudos prévios, realizados por Charlin et al. (1998), e Brailovsky et al. (2000), o número de itens desenvolvidos deveria ser maior do que cinqüenta. Considerou-se exceder o número mínimo de itens necessários, pois alguns poderiam estar inadequados ou mal construídos e, portanto, deveriam ser descartados ao longo do processo de desenvolvimento e de aplicação do teste.

4.4.2 Casos clínicos

Os itens que formaram o teste sempre estavam relacionados a um problema do paciente, que foi redigido na forma de relato de caso clínico. A apresentação da situação não deveria conter todos os dados necessários para a sua resolução, como acontece em situações reais.

4.4.3 Escolha do formato dos itens

A forma dos itens dependeu dos objetivos a serem alcançados. Assim os itens foram agrupados de acordo com uma meta de avaliação que, em geral, foi de investigação, diagnóstico ou tratamento. Houve também testes relacionados a prognóstico e dilemas éticos.

Portanto, para um mesmo caso clínico, poderia haver mais de um grupo de itens. Cada item apresentou três partes. A primeira apresentou uma hipótese diagnóstica, uma ação para investigação ou uma opção terapêutica. A segunda trouxe uma nova informação clínica, como um sinal, sintoma, condição, exame de imagem, resultado de testes de laboratório etc., que eram relevantes para a resolução do problema proposto. A última parte foi formada por uma escala tipo Likert, com cinco pontos, com os valores de 1 a 5, que representava o espectro de possibilidades diante de cada nova informação.

4.4.4 Outras considerações sobre a construção dos itens

Cada item foi construído com a intenção de levantar as dúvidas para a resolução da situação proposta. Dessa maneira, haveria a necessidade de uma reflexão para respondê-lo. Os itens eram independentes uns dos outros. O novo dado clínico foi diferente e independente para cada questão.

O número de testes para cada grupo de objetivos foi determinado pelo grupo responsável pela elaboração do instrumento, de acordo com o número de situações relevantes colocadas. Para grupos de hipóteses diagnósticas, o número mínimo de itens necessários para um teste de hipóteses foi quatro.

Já o número máximo de hipóteses não deveria exceder cinco. Esse limite está de acordo com as conclusões encontradas por Elstein et al, em 1978, sobre o raciocínio clínico. Segundo esse autor, o cérebro humano não consegue trabalhar ao mesmo tempo com mais de cinco hipóteses. Quando esse número foi ultrapassado, as hipóteses eram reformuladas e limitadas em número.

Seguem os apêndices de um a três com formatos teóricos para a construção de itens de acordo com o objeto de avaliação e o apêndice quatro com um exemplo de caso clínico seguido de um grupo de itens em geriatria.

4.5 Aplicação em um grupo piloto

Em uma etapa imediatamente anterior à aplicação final do teste, este foi resolvido por cinco geriatras que não participaram da elaboração do mesmo e que eram vinculados à UNIFESP, mas que também não iriam compor o painel de referência. Além de receberem explicações sobre como resolver o teste, esses geriatras puderam opinar sobre os itens, que foram reformulados ou descartados de acordo com esta primeira opinião e análise.

Com isso, a intenção foi de treinamento e aperfeiçoamento da instrução para a resolução do teste, realizada por escrito (apêndice 8) e pelo próprio pesquisador. Esse material foi desenvolvido para que os examinandos compreendessem claramente o teste antes de respondê-lo.

Esperava-se também, nessa primeira oportunidade, colher informações sobre a adequação do instrumento. Uma análise inicial de dados foi realizada antes da aplicação do instrumento no painel de referência e nos estudantes, buscando o aperfeiçoamento da metodologia dessa análise e a discussão de resultados preliminares.

4.6 O processo de formação dos grupos

O instrumento de avaliação foi analisado segundo a aplicação do mesmo primeiramente em um grupo de especialistas para a formação de um painel de referência para a pontuação dos examinandos.

Após a resolução dos testes, os especialistas foram requisitados, através de um questionário (apêndice 5), a responder perguntas sobre dados demográficos e sobre a opinião dos mesmos sobre o grau de dificuldade e de desafio proporcionados pelo teste.

Como etapa de validação do instrumento, ele foi aplicado em um grupo de estudantes de internato. Da comparação entre as respostas oferecidas pelos especialistas e os estudantes, foi possível concluir se o instrumento seguia um padrão de respostas de acordo com a teoria do teste de concordância de scripts.

4.7 O sistema de pontuação

Previa-se que um grupo de especialistas com no mínimo quinze integrantes seria necessário para expressar a variabilidade de respostas possíveis em cada item, já que esse era baseado em situações de incerteza. Portanto, não se considerou uma resposta correta para cada item, empregando-se o método de escore agregado.

Por esse método, qualquer resposta dada por um especialista possui um valor próprio, unitário e intrínseco, mesmo que os outros especialistas não concordem com ele. O escore de cada item é feito segundo as freqüências dadas para cada ponto da escala Likert.

Como exemplo, se para um determinado item, oito especialistas entre dez respondessem a opção de resposta três da escala Likert, um respondesse a opção dois e outro a opção quatro, o valor intrínseco da opção três seria 8/8 (1,0), das opções dois e quatro, 1/8 (0,125) e das opções um e cinco, 0/8 (0). O escore máximo para esse item, portanto, seria alcançado ao se assinalar a resposta três, obtendo-se a pontuação de 1,0.

Com isso, o escore máximo para cada item foi sempre igual a um, de acordo com a freqüência de respostas oferecidas pelos especialistas que formaram o painel de referência. Essa escala de valor está de acordo com a teoria de script e reflete o padrão de variabilidade de respostas oferecidas por especialistas diante de situações de incerteza.

A pontuação total do teste foi obtida através da somatória do escore obtido em cada item. O valor máximo para essa pontuação foi obtido pelo maior escore para cada item (um ponto em cada item), o que seria considerado como cem por cento de aproveitamento na resolução do teste.

4.8 Populações integrantes dos grupos

A pesquisa foi realizada através da elaboração, aplicação e análise do instrumento. Ele foi aplicado em um grupo de especialistas em geriatria e em um grupo de estudantes de graduação em medicina.

O grupo de especialistas foi convidado segundo uma lista de geriatras de referência, vinculados a instituições de ensino, pesquisa e assistência, envolvidos com a área de geriatria nas mesmas. Foram considerados, para tanto geriatras atuantes no estado de São Paulo, sugeridos pela direção da Disciplina de Geriatria e Gerontologia da UNIFESP.

Para a análise do desempenho dos estudantes de graduação, foram considerados os estudantes de medicina da UNIFESP. Como particularidade, esses estudantes aprendem sobre tópicos em geriatria do 1º ao 5º anos do curso médico. A Disciplina de Geriatria e Gerontologia da UNIFESP é diretamente responsável por esse ensino que, sob ponto de vista curricular, diferencia-se daquele que é presente na maioria das escolas médicas brasileiras.

Os especialistas em geriatria deveriam apresentar mais de dez anos de prática médica e pelo menos cinco anos de experiência nesta área. E os estudantes que integraram a pesquisa estavam no período de internato, especificamente no quinto ano do curso médico. Este grupo deveria ter desenvolvido a formação completa em geriatria proposta pelo currículo formal da UNIFESP.

4.9 Tamanho dos grupos

Segundo dados publicados por Gagnon et al. (2005), um número por volta de vinte especialistas é associado com uma boa acurácia para a formação do painel de referência. Nosso painel de especialistas foi composto por 21 integrantes.

Para a formação do grupo de estudantes, foram considerados os alunos do quinto ano do curso de medicina da UNIFESP nos anos de 2007 e de 2008. Foram convidados os estudantes que cursaram o módulo de geriatria no período de setembro de 2007 a janeiro de 2008. Destes, foi obtida uma amostra de conveniência com 41 integrantes.

4.10 Pesquisa sobre as características dos examinados

Após a resolução do teste, os especialistas preenchiam o questionário sobre características demográficas, formação educacional e práticas docentes e assistenciais. As questões englobavam informações sobre idade, sexo, ano de graduação, função docente, ano de obtenção de título de especialista e os locais de prática clínica atual.

Esperava-se, com essas informações, conhecer algumas características desse grupo que poderiam influenciar suas respostas ao teste. Além disso, algumas questões abertas ao final permitiram a esses profissionais colocarem suas opiniões sobre a utilidade e o grau de dificuldade envolvidos na resolução do teste. O mesmo não foi feito com o grupo de estudantes, já que este grupo possuía uma homogeneidade inerente à situação de estarem no mesmo momento de formação e freqüentarem o mesmo ambiente educacional.

Entre os vinte e um especialistas que aceitaram fazer parte do painel de referência, cinco (23,8%) eram do sexo feminino e dezesseis (76,2%) eram do sexo masculino. A média de idade desse grupo foi de 45,6 anos, sendo a menor idade de 33 anos e a maior de 58 anos. A média de anos de exercício da medicina foi de 22,5 anos, com o mínimo de 11 anos e o máximo de 33 anos. Quanto ao tempo de especialização em geriatria, a média foi de 15,4 anos, com o mínimo de seis anos e o máximo de 22 anos.

Entre os estudantes que aceitaram participar desta pesquisa, dezessete (41,5%) eram do sexo feminino e 24 (58,5%) eram do sexo masculino. A média de idade desse grupo foi de 26,7 anos.

4.11 Procedimentos para a análise dos dados

Os dados foram armazenados e analisados nos programas estatísticos “SPSS for Windows” (Statistical Package for Social Sciences, Version 16.0. Chicago, Illinois: SPSS; 2008.) e “EduG 3.04 for Windows” (English Version. Neuchatel, Switzerland: IRDP; 2006). Este último programa foi utilizado exclusivamente na aplicação de estudos G para a análise dos dados.

Levando-se em conta que a escala de medida do item é ordinal, foram utilizadas medidas de tendência central e de dispersão compatíveis com tal escala. Sendo assim, a descrição dos itens foi feita por meio da mediana e da amplitude interquartil (Q3 –Q1) para cada grupo analisado. Foram comparadas as médias dos dois grupos, ou seja, do grupo de especialistas e do grupo de estudantes.

Na análise dos escores dos itens, foi calculado o coeficiente de correlação parcial-total, que correlaciona cada um dos escores de cada item com o escore total, incluindo todos os participantes da pesquisa. Buscava-se, com isso, analisar a consistência interna dos resultados da aplicação do instrumento, o que se traduz em sua confiabilidade, precisão e potencial de replicação (DOWNING, 2004). Os itens que apresentaram correlação baixa foram retirados para maximizar essa qualidade psicométrica do resultado.

A consistência interna dos resultados poderia ser prejudicada por diversos motivos. Entre eles, estariam a presença de itens muito fáceis, ou de itens muito difíceis, ou ainda aqueles sem poder discriminatório entre os indivíduos que responderam o teste. Pela integração dos dados dos coeficientes de correlação parcial-total individuais, chegamos ao grau de acurácia dos resultados para novas aplicações do instrumento nos examinandos que fizeram parte do estudo, desconsiderando o efeito de memória sobre os resultados de repetidas aplicações do mesmo teste ou a perda ou ganho de competência ao longo do tempo. Esse grau de capacidade de replicação é expresso numericamente por meio do coeficiente de alfa de Cronbach (CRONBACH, 1951).

A normalidade da distribuição do escore total por itens foi analisada por meio de teste formal (Teste de Shapiro-Wilk). Quando houve normalidade na distribuição dos dados dentro dos grupos, foi realizado o teste t de Student. Em caso de rejeição da normalidade, foram utilizados os métodos não-paramétricos (teste de Mann-Whitney).

Ainda, para garantir a validade dos métodos de comparação das variáveis, foi utilizado o teste de homogeneidade de variâncias. Dessa forma, a hipótese de homogeneidade foi avaliada por meio do teste de Levene.

Por final, buscamos aprimorar a análise das fontes de erro do escores obtidos através de medidas a partir da Teoria G, que pode também fornecer indícios a respeito da potencial validade do instrumento. Com relação a esse conceito, Cronbach et al. (1972, p. 15) assim se pronunciaram:

> O escore (em um teste ou em outra medida) a partir do qual a decisão se baseia é somente um dos muitos escores que poderiam servir ao mesmo propósito. Quem deve tomar a decisão, quase sempre não está interessado na resposta dada a uma questão, a um objeto ou a um estímulo em particular, no examinando em particular, ou no momento particular do teste. Algumas, no mínimo, destas condições de medida poderiam ser alteradas sem fazer o escore menos aceitável para quem vai tomar a decisão. O dado ideal em que a decisão deve se basear seria alguma coisa como o escore médio da pessoa em todas as observações aceitáveis e possíveis.

A teoria de testes Clássica, da qual derivam os conceitos de consistência interna e de coeficiente de alfa de Cronbach, pode estimar somente uma fonte de erro por vez, como por exemplo, em uma avaliação de confiabilidade no formato teste e re-teste (SHAVELSON e WEBB, 1991).

Uma das vantagens da teoria Moderna, ou teoria G, é que as múltiplas fontes de erro em uma medida podem ser separadas e analisadas isoladamente, por meio de uma única aplicação do teste. Com isso, há a possibilidade de se determinar quantas ocasiões, qual o número de itens e quantos avaliadores são necessários para se obter resultados válidos, o que se denominou de Estudo D (SHAVELSON e WEBB, 1991).

A teoria G também permite a mensuração de um coeficiente, denominado coeficiente G, que reflete a proporção do escore observado que se refere ao escore universal, o qual seria obtido a partir da grande média dos resultados dos examinandos, em um número “infinito” de itens que poderiam ser resolvidos em um teste com os mesmo formatos e objetivos ao que foi aplicado. Este coeficiente é semelhante ao coeficiente de alfa de Cronbach em seu cálculo e significado.

Outra limitação da teoria Clássica é que ela permite somente uma interpretação dos resultados, a partir das posições relativas entre os examinandos, chamado de escore relativo. A teoria G permite também avaliações absolutas de performance, ou de escore absoluto, segundo “notas de corte”, por exemplo, importantes em alguns tipos de avaliação, como quando o objetivo é a licença para o exercício profissional ou a aprovação em um curso (“ou passou ou reprovou”). Os escores absolutos também são úteis em avaliações formativas e em avaliações somativas.

Entre os modelos possíveis de desenho de um estudo G, esta análise considerou uma única fonte de variabilidade, ou seja, os itens (*i*), que são denominados de faceta de instrumentação. Os examinandos (*p*) são considerados como faceta de diferenciação, ou seja, são aqueles que vão ser diferenciados pelo teste.

Em suma, obtivemos, com esse desenho, quatro fontes de variabilidade para os resultados dos escores obtidos: a diferença entre as pessoas avaliadas quanto ao raciocínio clínico (denominado de componente *p*), a diferença entre o grau de dificuldade dos itens (componente *i* da variância para o escore observado), a interação entre examinandos e itens (“*p* x *i* “) e os eventos não identificados ou ocorridos ao acaso.

Essas duas últimas fontes de variância para o escore obtido não podem ser separadas em um modelo com somente uma faceta (*i*) no universo observável. Sendo assim elas são calculadas conjuntamente e não podem ser separadas pela análise. O componente “*p* x *i*” da variância da média do escore obtido surge a partir de histórias de experiências educacionais ou profissionais que os examinandos trazem para a resolução do teste. Se, por exemplo, um especialista já vivenciou um caso clínico semelhante ao apresentado no teste, uma maior parte do escore obtido a partir desse caso será devido à aprendizagem obtida por essa vivência. Uma grande variação de desempenho ao longo dos itens para um mesmo examinando também é fonte de variabilidade para a interação entre os examinandos e os itens.

Porém os eventos não identificados ou ocorridos ao acaso não podem ser analisados separadamente, a não ser que fossem considerados como uma fonte isolada de variância, o que aumentaria muito a complexidade da análise. São eventos como desatenções dos examinandos durante a aplicação do teste, resolução em locais diferentes e em momentos diferentes para cada participante ou grupo de participantes, escolha de alternativas ao acaso (o chamado “chute” ou “aposta” de opção de resposta), entre outras fontes de variabilidade de um resultado, que o tornam não generalizável e impreciso.

Sendo assim, obtivemos as proporções das três fontes de variabilidade (examinandos, itens e "*p* x *i*") dos escores de itens para os itens com melhor consistência interna para os especialistas e para os estudantes. Também foram calculados os coeficientes G relativos e absolutos para ambos os grupos.

A existência de dois tipos de coeficiente G se deve às possíveis decisões que podem ser tomadas a partir do escore obtido. Decisões relativas são aquelas que classificam os resultados através de um ranqueamento, ou seja, uma classificação onde um examinando é sempre visto em relação aos outros. A validação desse processo de análise dos resultados pode ser obtida pelo valor do coeficiente G relativo.

As decisões absolutas são consideradas mais rigorosas e são obtidas pelo coeficiente G absoluto. Nesse caso os resultados individuais são vistos através de "notas de corte", ou do tipo bom ou ruim, ou habilitado ou insuficiente, em relação a uma padronização prévia. Avaliações como essas necessitam de decisões mais complexas e dependem do grau de dificuldade dos itens, que são considerados uma fonte de erro no escore. Os itens mais difíceis alteram isoladamente o julgamento dos resultados dos examinandos, e podem levar a aprovação ou reprovação de um examinando, independentemente do seu nível de competência para o atributo a ser avaliado. Nas decisões relativas, o grau de dificuldade dos itens não altera a posição relativa entre os examinandos.

Além disso, durante a construção, a aplicação e a análise do instrumento, foi levantada a hipótese de interação entre os itens de uma mesma situação ou caso clínico. Como os itens estavam dispostos sequencialmente em quadros, parecia difícil, tanto durante a construção, como na resolução dos mesmos, o isolamento do raciocínio entre itens relacionados a uma mesma situação clínica. Portanto, a interação item-item pode ser uma fonte significativa de variabilidade do escore observado e estaria contida no componente "*p* x *i*" da análise da variância dos resultados.

Sendo assim, esta hipótese foi estudada através da análise estatística do escore separado por situações. Com isso, obteve-se a consistência interna dos resultados para a somatória dos escores para cada descrição de situação clínica isoladamente, bem como a capacidade que cada uma dessas situações tem para diferenciar os grupos estudados.

O estudo G para os escores por situação clínica também foi realizado, com o objetivo principal de se analisar a diferença entre os componentes de variância "*p* x *i*" dos escores por itens e por situações, buscando fundamentação para a presença e a importância da interação item-item em uma mesma situação clínica nesse tipo de teste.

Com relação à perda de valores durante a aplicação do teste, estas foram minimizadas e desprezíveis diante da análise dos resultados finais.

5. RESULTADOS

5.1 O processo de construção do instrumento

De acordo com as descrições anteriores, cada um dos integrantes dessa etapa da pesquisa desenvolveu uma série de descrições de casos clínicos por escrito, envolvendo situações geriátricas, para serem discutidas pela equipe, que deveria decidir sobre aquelas que seriam mais adequadas para a composição do teste.

Sugeriu-se englobar as situações mais freqüentes em geriatria (demências, depressão, delirium, quedas, tonturas, síncope, incontinências, iatrogenias, entre outras) nesses casos.

Por fim, foram selecionados doze casos clínicos, sendo que o último apresentou uma segunda parte, que descrevia a evolução clínica de um paciente. Todos os casos foram revisados pela equipe para que fossem concisos, curtos e para que não apresentassem todas as informações necessárias para a sua resolução. Segue abaixo, na próxima página, a freqüência com que cada tema em geriatria foi considerado de forma direta ou indireta ao longo das situações clínicas:

Tabela 1 – Freqüência com que os temas geriátricos foram considerados

Tema	*Freqüência dos temas*
1. Demências	4
2. Iatrogenias	4
3. Síndromes Depressivas	3
4. Delirium	2
5. Obstipação	2
6. Quedas	2
7. Instabilidade Postural	2
8. Perda Involuntária de Peso	2
9. Tontura	1
10. Síncope	1
11. Cuidados Paliativos/ Dilemas Éticos	1
12. Incontinências	0

Entre as características dos pacientes descritos, a média de idade dos mesmos foi de 79,5 anos, com o desvio-padrão de 6,695 anos, idade mínima de 72 anos e máxima de 92 anos, com curtose de – 0,612. Portanto, a faixa etária dos pacientes descritos seguiu uma distribuição normal.

Entre os ambientes de assistência em que os pacientes se encontravam, cinco estavam em situações de atendimento hospitalar, cinco foram atendidos em ambiente ambulatorial, um deles estava em domicílio e um outro em instituição de longa permanência.

A partir dos casos selecionados, desenvolveram-se alguns itens relevantes à resolução dos mesmos, envolvendo raciocínio para hipóteses diagnósticas, investigação, tratamentos e decisões éticas. Todos os casos continham um quadro de itens envolvendo raciocínio sobre hipóteses diagnósticas. Nem todos apresentavam quadros com itens relacionados à investigação e a tratamento. Somente o último caso apresentava itens envolvendo dilemas éticos.

Ressalta-se que a construção dos itens foi discutida diversas vezes em equipe. Os itens foram modificados ou excluídos até o alcance de um consenso entre os seus membros. Finalmente obtivemos 115 itens para a aplicação do instrumento em um grupo piloto.

5.2 Análise da aplicação em um grupo piloto para ajustes no instrumento

Cinco especialistas responderam aos 115 itens construídos inicialmente. Foram registradas as freqüências de resposta em cada item, para cada um dos integrantes deste grupo. Essas respostas estão representadas no apêndice 9 (Tabela 9).

De acordo com Cicchitti et al. (1997), foram realizadas análises de concordância entre os especialistas com relação às respostas para cada questão ou item. Foram consideradas as concordâncias denominadas de

justas ou boas, cujos valores calculados se situavam entre 0,70 e 0,85. Consideraram-se impróprias as concordâncias abaixo ou acima destes valores.

Os níveis de concordância abaixo de 0,70 (baixa concordância) foram considerados inadequados por serem interpretados como uma falta de tendência nas respostas dos especialistas. Esses itens podem ter sido mal interpretados durante a resolução. Em contrapartida, os itens com respostas com concordâncias muito elevadas (acima de 0,85) podem não ter gerado incerteza e, portanto, também foram excluídos ou modificados.

Os itens com respostas com nível de concordância considerado baixo ou muito elevado foram reapresentados ao grupo responsável pela construção do teste. Cada um desses itens foi reconstruído buscando melhores níveis de compreensão e de incerteza. Alguns itens foram considerados completamente inadequados e foram excluídos definitivamente. Por final, o teste modificado e em sua versão completa apresentou 104 itens, e está representado no item 5.7 deste capítulo.

5.3 Aplicação do teste em um grupo de especialistas

Durante o período de coleta dos dados, dentro da uma lista de trinta especialistas selecionados para fazer parte do painel de referência, conseguiu-se que vinte e um deles respondessem ao teste, de acordo com disponibilidade dos participantes para essa etapa. Seguem abaixo os números de geriatras do painel de referência de acordo com o vínculo institucional que apresentavam:

Tabela 2 – Número de especialistas do painel de referência, por instituição

Instituição	*N participantes*
Universidade Federal de São Paulo - UNIFESP	7
Universidade de São Paulo - FMUSP	6
Universidade Estadual Paulista – UNESP	2
FCM da Santa Casa de São Paulo	2
Faculdade de Medicina de Marília - FAMEMA	1
Pontifícia Universidade Católica – PUC – São Paulo	1
Hospital Israelita Albert Einstein	1
Fleury Medicina e Saúde	1

Todos os especialistas responderam ao teste individualmente. As explicações sobre a resolução do teste e sobre a teoria de script foram feitas pelo pesquisador antes do início da resolução. Havia também um capítulo introdutório, por escrito, na primeira e na segunda página, com instruções sobre a sua resolução do teste (apêndice 8). Neste capítulo também existia um modelo estrutural dos itens e um caso clínico como exemplo, para treinamento e adaptação antes da resolução definitiva do instrumento.

Seguindo-se às instruções iniciais, as dúvidas sobre o teste eram discutidas entre o pesquisador e o especialista. Foi sugerida ao especialista a resolução individual do teste, preferencialmente sem a presença do pesquisador no mesmo local. Somente dois especialistas resolveram o teste com o acompanhamento do pesquisador. As respostas obtidas a partir da aplicação do teste no grupo de especialistas estão representadas no apêndice 9, Tabela 10.

Entre os 104 itens respondidos pelos 21 especialistas (2184 respostas), somente quatro não foram assinaladas. Esse número de perdas foi considerado desprezível para fins de análise dos dados.

5.4 Aplicação do teste em um grupo de estudantes

Entre setembro de 2007 e janeiro de 2008, foi realizada a aplicação do teste nos estudantes do quinto ano médico da UNIFESP, no final do estágio de geriatria. Essa coleta foi realizada quinzenalmente, em grupos de cinco a seis estudantes, de acordo com o rodízio dos mesmos no ambulatório de Neuropsiquiatria Geriátrica da Disciplina de Geriatria e Gerontologia da UNIFESP, local onde o pesquisador realizava suas atividades de ensino junto a esses estudantes.

Durante esse período, 41 estudantes aceitaram participar da pesquisa, formando o grupo a ser comparado com o painel de referência. As mesmas instruções apresentadas aos especialistas foram também realizadas com esse grupo. Entre as 4264 respostas que poderiam ser obtidas neste grupo, houve somente 4 respostas em branco, o que também foi considerada uma perda desprezível. As respostas obtidas nessa fase da pesquisa estão representadas no apêndice 9, Tabela 11.

5.5 Cálculo do escore agregado

Para a obtenção do valor do escore de cada opção de resposta em cada item ou questão, foram utilizados os princípios metodológicos descritos no item 5.5 do capítulo sobre metodologia.

Desse modo, foram computadas as frequências de respostas dos especialistas em cada uma das questões (exemplo na Tabela 3, abaixo). Em seguida, para cada uma das questões, foi obtido o valor modal, ou seja, o número de vezes que a opção de resposta mais respondida foi assinalada. Esse valor foi utilizado como denominador para o cálculo do escore agregado.

Tabela 3 - Frequência das opções de resposta nas questões de 1 a 5

Resposta	***Questão 1***	***Questão 2***	***Questão 3***	***Questão 4***	***Questão 5***
1	0	0	0	0	0
2	0	0	2	4	1
3	9	14	3	11	3
4	12	7	15	6	15
5	0	0	1	0	2

* freqüência de todas as questões está no apêndice 9, Tabela 10.

Como exemplo, consideraremos o cálculo para a questão 3, na tabela acima. Nessa questão, a opção de resposta 4 apresentou a maior freqüência de escolhas. Essa freqüência de respostas foi então considerada o valor modal para essa questão, que no caso, foi igual a 15. Sendo assim, o escore para cada resposta da questão 3 foi obtido dividindo-se o número de vezes que essa resposta foi assinalada pelo valor modal.

Portanto, o escore da opção de resposta 1, para a questão 3, é 0/15; para a opção de resposta 2, 2/15; para a opção de resposta 3, 3/15; para a opção de resposta 4, 15/15; e para a opção de resposta 5, 1/15. Com isso, foram formados os valores de escores para as respostas da questão 3, conforme pode ser visto na tabela 4, abaixo:

Tabela 4 - Formação do escore nas questões de 1 a 5

Resposta	*Questão 1*	*Questão 2*	*Questão 3*	*Questão 4*	*Questão 5*
1	0	0	0	0	0
2	0	0	0,13	0,36	0,07
3	0,75	1	0,2	1	0,2
4	1	0,5	1	0,55	1
5	0	0	0,07	0	0,13

* escores de todas as questões está no apêndice 9, Tabela 13.

Essa pontuação foi utilizada para ranquear cada resposta de cada integrante de ambos os grupos que participaram da pesquisa, como pode ser visto na tabela 13, do apêndice 9. Essas pontuações estão representadas nas tabelas 14 e 15 deste mesmo apêndice. A partir da omatória dos escores de cada questão, para cada participante, obtivemos os escores totais de cada especialista ou estudante, que também podem ser nestas últimas tabelas. Realizamos ainda o cálculo da porcentagem de aproveitamento com relação ao escore máximo possível para cada um dos integrantes da pesquisa.

5.6 Análises dos resultados após a formação do escore do teste

5.6.1 Análise dos resultados dos escores totais por itens na diferenciação entre os grupos

Por meio da análise estatística descritiva, foram obtidas as médias, as medianas, os desvios-padrões (D. P.), os valores mínimos (Min.) e máximos (Max.), entre outros, dos escores dos grupos que resolveram a versão final do teste. Esses dados podem se vistos na tabela 5, logo abaixo. A variação dos escores foi maior no grupo de estudantes (32,96) do que no grupo de especialistas (21,82).

Tabela 5 - Análise do Escore Agregado do Total de Itens (N=104)

grupos	*N*	*Média*	*Mediana*	*D.P.*	*Mín.*	*Máx.*	*Variação*
Especialistas	21	81,41	80,91	5,46	70,27	92,09	21,82
Estudantes	41	70,71	72,17	7,28	51,16	84,12	32,96

* Estudantes vs. Especialistas: p<0.001 (teste t de Student)

A hipótese de normalidade das distribuições foi testada pelo teste de Shapiro-Wilk (SW). Não foi observada violação desta hipótese na distribuição dos escores nos grupos especialistas (SW= 0,953; p=0,384) e estudantes (SW=0,978; p=0,616).

Para se testar a homogeneidade das variâncias de distribuição dos escores dos grupos, foi realizado o Teste de Levene. Não houve violação desta hipótese (*F*=3,002, *p*=0,088), corroborando que as duas variâncias eram iguais.

Seguindo-se a hipótese de normalidade das amostras, o teste t de Student foi empregado na comparação entre as médias de escores totais dos grupos de especialistas e de estudantes. A média do escores do grupo de especialistas (81,41) foi superior à média dos escores do grupo de estudantes (70,71), com *p*<0,001.

5.6.2 Análise da consistência interna dos resultados dos escores por itens

O coeficiente de alfa de Cronbach, considerando-se todos os participantes, para o total de itens, foi de 0.842. A eliminação do especialista com o menor escore total aumentou a média do grupo e diminuiu seu desvio-padrão, mas não alterou a acurácia do teste. Portanto, decidiu-se manter este especialista durante toda a análise.

O coeficiente correlação parcial-total foi utilizado para selecionar os itens que melhor poderiam distinguir os indivíduos que obtêm pontuações altas e os que obtêm pontuações baixas. Para ser mantido, um item deveria apresentar este coeficiente com valor de 0,19 ou maior. Quarenta e três itens apresentavam baixa correlação e foram excluídos. Com os 61 itens restantes (58,7% do total), o coeficiente alfa de Cronbach elevou-se para 0,882.

5.6.3 Comparação dos escores totais por itens entre os grupos após a análise da consistência interna

Na análise subseqüente, consideraram-se apenas os itens com boa correlação parcial-total, assegurando-se a consistência interna, ou seja, a capacidade de discriminância entre os indivíduos que responderam ao teste e a capacidade de replicar os resultados obtidos. Os itens descartados estão assinalados com grifo e itálico no sub-capítulo 5.7. Sendo assim, 61 itens foram novamente analisados, isoladamente, de forma descritiva, conforme demonstrado na tabela 6 (ver abaixo). Os escores totais o especialistas e dos estudantes após essa seleção de itens estão no apêndice 9, tabelas 16 e 17.

Tabela 6 - Análise do Escore Após Seleção de Itens Viáveis (N=61)

grupos	***N***	***Média***	***Mediana***	***D.P.***	***Mín.***	***Máx.***	***Variação***	***SW***
Especialistas	21	49,01	50,06	4,57	36,43	56,65	20,22	0,021
Estudantes	41	38,87	39,59	6,43	22,82	50,70	27,88	0,272

* Estudantes vs. Especialistas: $p<0.001$ (Teste de Mann-Whitney)

Novamente, foi utilizado o teste de Shapiro-Wilk (SW), que neste caso não comprovou a normalidade de distribuição dos resultados do grupo de especialistas (p=0,021). Nessa nova distribuição, o especialista com menor escore colocou-se abaixo do primeiro intervalo inter-quartil. Entretanto, sua retirada do grupo não foi significante para a mudança de resultados. Na reaplicação do teste de Levene, observou-se a homogeneidade das variâncias entre os grupos (F=3,116, p=0,083).

Devido à ausência de normalidade na distribuição do grupo de especialistas para o escore total calculado a partir dos 61 itens selecionados, a comparação entre as médias desses escores entre os grupos de especialistas e de estudantes, neste caso, foi realizada através da aplicação do teste da Mann-Whitney. Os resultados demonstraram que os grupos não foram iguais quanto à resolução do teste, após a seleção dos itens com maior consistência interna (p<0,001). Nessa situação, a média do escore total do grupo de especialistas foi de 49,01, superior à média obtida pelo grupo de estudantes, que foi de 38,87.

5.6.4 Aplicação do Estudo G para os resultados dos itens adequados quanto à consistência interna

Como não foi objetivo desta pesquisa generalizar os resultados do teste para todos os grupos possíveis, de acordo com o grau de competência na área de geriatria, os grupos de especialistas e estudantes foram considerados como facetas fixas (não generalizáveis) e, portanto, analisados separadamente. Seguem abaixo os resultados do estudo G segundo os escores totais por itens, considerando o teste final com 61 itens selecionados pela análise de consistência interna realizada anteriormente:

Tabela 7 – Estudo G com desenho P/I para itens

Fonte de Variabilidade no Escore (%)	*Especialistas*	*Estudantes*
Items (i)	1,7	10,7
Resíduo (*p* x *i*)	98,3	89,3

Esse desenho de estudo foi obtido a partir de cálculos com três fontes de variância: os participantes (*p*), os itens (*i*) e a interação *p* x *i*, a partir de um desenho com facetas cruzadas. As porcentagens de variâncias para esses componentes foram, respectivamente 4,4% (*p*), 1,6% (*i*) e 94% (*p* x *i*) para o grupo de especialistas, e 6,1% (*p*), 10% (*i*) e 83,9% (*p* x *i*) para o grupo de estudantes.

Os coeficientes G relativo e absoluto para o grupo de especialistas foram respectivamente 0,741 e 0,738. Para os estudantes, os coeficientes G relativo e absoluto foram 0,816 e 0,798.

5.6.5 Análise do escore por situações clínicas

Como o último caso clínico contém a descrição de duas situações, optou-se pela análise da soma dos escores de 13 situações como uma opção para a consideração da interação item-item para cada situação clínica. Desse modo, foram obtidos treze escores que correspondiam à somatória dos escores de cada item dentro de cada situação.

Porém, como cada situação apresentava um número diferente de itens (entre 4 e 15 itens), a variação de escore máximo entre as situações era igualmente diferente (entre 4 e 15 pontos), tornando a análise comparativa entre os escores das situações inviável. Optou-se, portanto, por dividir esse escore pelo número de itens que o formou, ou seja, o escore de uma situação com 4 itens foi dividido por 4 e o escore de uma situação com 15 itens foi dividido por 15, por exemplo.

Com isso, formou-se hipoteticamente um escore médio, como se cada item dentro de uma mesma situação clínica apresenta-se o mesmo resultado. Esse valor foi denominado de escore por situações obtido pela média dos valores de escore de cada item para cada situação e estão representados nas tabelas 18 e 19, no apêndice 9. Este artifício tornou a variação entre os escores mínimos e máximos iguais e comparáveis entre si, ou seja, entre 0 e 1.

Considerando-se esse novo resultado, por situação clínica, observou-se que, por meio do teste de Mann-Whitney, as situações 8, 10, 11 e 12 não foram capazes de diferenciar especialistas de estudantes. Na análise da consistência interna dos resultados, a situação 11 foi a única que apresentou correlação parcial total abaixo de +0,19 (-0,12). O coeficiente de alfa de Cronbach para as 13 situações foi de 0,78. Descartando-se a situação 11, alcançou-se o coeficiente de alfa de Cronbach de 0,80. Seguem abaixo os resultados do estudo G para as 13 situações.

Tabela 8 - Estudo G com desenho P/S para situações

Fonte de Variabilidade do Escore (%)	***Especialistas***	***Estudantes***
Situações (*s*)	8,4	17,8
Resíduo (*p* x *s*)	91,6	82,2

Novamente, esse desenho de estudo foi obtido a partir de cálculos com três fontes de variância: os participantes (*p*), situações (*s*) e a interação "*p* x *s*", a partir de um desenho com facetas cruzadas. As porcentagens de variância para esses componentes foram, respectivamente 11,3% (*p*), 7,5% (*s*) e 81,2% ("*p* x *s*") para o grupo de especialistas , e 12,5% (*p*), 15,5% (*s*) e 71,9% ("*p* x *s*") para o grupo de estudantes. Analisando-se os dois grupos conjuntamente, as porcentagens de variâncias para esse componentes foram respectivamente 19,4% (*p*), 9,6% (*i*) e 71,1% ("*p* x *s*").

Levando-se em conta esse novo escore, por situações, os coeficientes G relativo e absoluto para o grupo de especialistas foram, respectivamente 0,644 e 0,624. Os resultados por situações no grupo de estudantes apresentaram os coeficientes G relativo e absoluto de 0,694 e 0,650, respectivamente. Ainda, em análise conjunta englobando especialistas e estudantes, sem diferenciação por grupos, obteve-se o coeficiente G relativo de 0,779 e o coeficiente G absoluto de 0,757.

6. DISCUSSÃO

6.1 Considerações sobre a construção do instrumento

A construção do teste foi baseada em descrições de casos clínicos, envolvendo diversas situações sugeridas pelo grupo responsável por sua elaboração. Esse grupo, embora pudesse ser considerando homogêneo, formado por três geriatras na faixa etária dos 35 anos de idade, com formação em geriatria na UNIFESP, não se tratava de um grupo de experts nessa área de atuação. Experiências anteriores, que discutem a formação do expertise, colocam que seria necessário cerca de 10 anos para a formação de um expert (CROSKERRY, 2005).

Essa opção não invalidou de maneira alguma os princípios de construção propostos anteriormente para a elaboração do instrumento. Porém, segundo Fournier et al., em 2008, poderia ser útil a supervisão e revisão do trabalho deste grupo por um expert na área de desenvolvimento do instrumento.

Isso não deveria ser limitado somente à área de geriatria, como no caso deste estudo, mas também um profissional experiente em instrumentos de avaliação, psicometria e até mesmo com experiências anteriores na elaboração de testes de concordância de scripts (TCS). Como nossa elaboração foi de natureza experimental, não eram esperados problemas desse tipo em seu planejamento. Vale lembrar que, quando foi desenvolvido o projeto desta pesquisa, não dispunhamos ainda de “guidelines” de recomendação sobre a construção deste tipo de teste, como o que foi recentemente publicado, em maio de 2008, por Fournier et al..

Um item ou uma questão do teste não deve ser considerado isoladamente, tanto em sua construção, como em sua resolução. Um item deve ser sempre ele mesmo e o caso clínico ao qual ele se refere. Pode ainda ser influenciado por itens presentes ao seu redor. Sendo assim, qualquer construção isolada ou distanciada de itens em relação a seus respectivos casos clínicos, poderia prejudicar a compreensão e a resolução do teste.

Nesse sentido, como exemplo, os especialistas que não concordassem com a hipótese diagnóstica, a investigação ou a terapêutica que foi sugerida pelo

item, poderiam cognitivamente desmembrá-lo da situação clínica a qual ele se refere, prejudicando a acurácia dos resultados obtidos.

Portanto, uma revisão do instrumento, após a sua construção, por um expert na área de atuação à qual ele se refere, poderia contribuir com a coerência das questões. Outra estratégia de construção que poderia melhorar a correlação de itens com as situações clínicas é a redução no número de itens por caso clínico (GAGNON et al., 2008).

O cenário descrito no caso clínico deve estar claro e preciso. Ainda, é necessário que ele descreva uma situação típica, mas desafiadora para os especialistas da área à qual o teste se refere, por motivo de contexto duvidoso ou por carência de informações suficientes (FOURNIER et al., 2008). No teste desenvolvido nesta pesquisa, todas as situações envolviam práticas profissionais comuns à área de geriatria.

Analisando algumas características dos casos desenvolvidos, nota-se que a média de idade dos pacientes apresentados era de 79,5 anos, adequada, portanto, à faixa etária de pacientes atendidos por essa especialidade.

Quanto à diversidade e representatividade dos problemas geriátricos mais comuns, percebe-se que a maioria deles está representado no teste, contribuindo para a sua validade de conteúdo. Somente o assunto “incontinências” não esteve envolvido nas situações propostas, provavelmente pelo seu menor potencial de geração de incertezas diagnósticas.

Ainda, é importante considerar que as situações precisavam gerar incertezas. Sendo assim, muitos problemas clínicos comuns para a população idosa não foram considerados por apresentarem menor potencial de gerar dilemas práticos ou por não serem grandes representantes do domínio da prática geriátrica, como hipertensão, diabetes, doenças cardiovasculares em geral, entre outros.

Porém, reconhece-se que, considerando-se a especificidade do raciocínio clínico (NORMAN, 1985), mais situações diferentes poderiam ter sido descritas. Um teste envolvendo vinte casos, sessenta questões, para uma hora de resolução, pode apresentar uma boa acurácia (METERRISIAN, 2006). Portanto, três itens por caso ou situação clínica seriam suficientes para garantir uma boa

representatividade de situações para o raciocínio e diminuiriam o risco de uma resolução de questões sem a consideração do contexto a que elas estão envolvidas.

Segundo Crocker e Algina (1986), uma seqüência de muitos itens após um caso clínico, relacionados a diferentes objetivos educacionais (raciocínio clínico para diagnóstico, investigação, tratamento etc.) prejudica a independência entre os itens ou questões, violando um princípio de construção do teste. No momento em que essa pesquisa foi planejada, os estudos sobre o teste de concordância de scripts não restringiam os tipos ou o número de itens por caso clínico.

Mais recentemente, entretanto, Gagnon et al. (2008), publicou conclusões baseadas em análises com estudos G e estudos D que indicam que um ótimo número para um TCS com boa precisão seriam 2 a 4 itens por situação clínica, em um teste com 15 a 25 situações.

Por isso, o desenvolvimento de algumas situações clínicas com números excessivos de itens (maiores que cinco) e com mais de um quadro de itens (casos com até três quadros de itens) podem ter prejudicado a interpretação e a validade de nossos resultados. As análises psicométricas contribuíram para a compreensão desse fenômeno.

Uma das principais hipóteses de interpretação sobre o instrumento, a capacidade de diferenciar indivíduos, principalmente quanto a seus níveis de experiência, precisa ser estudada por meio de experimentação e fundamentação analítica. Para isso, a construção de itens poderia ter seguido como parâmetro exclusivamente os níveis médios de dificuldade. Os itens muito fáceis ou muito difíceis contribuem menos para discriminar as competências individuais.

Entretanto, o grau de dificuldade das questões pode ser diferente de acordo com os grupos estudados. A intenção de se construir um teste com um determinado nível de dificuldade é uma tarefa imprecisa quando não se tem experiência ou se dispõe de um banco de dados previamente, com itens que possam ser escolhidos para a aplicação. Uma alternativa seria a análise do grau de dificuldade dos itens por meio de testes psicométricos dos resultados em

grupos pilotos semelhantes aos grupos estudados, antes da aplicação definitiva do instrumento.

Além dos contextos, formados a partir das descrições de situações clínicas, outra fonte de incerteza do instrumento é o conteúdo dos itens. O grau de incerteza que um item é capaz de gerar pode ser medido pela distribuição de respostas dos especialistas ao longo da escala Likert. Por esse motivo realizamos um teste piloto com especialistas, cujo análise buscou obter o grau de concordância das respostas em cada item.

Caso essa análise demonstrasse itens com respostas consensuais, poder-se-ia concluir que os mesmos geraram pouca incerteza, assemelhando-se a testes de múltipla escolha. O contrário também prejudica a qualidade do item, já que dispersões muito grandes (alto grau de discordância) podem indicar erros de interpretação ou de construção do item. Portanto, itens com qualidade para gerar maiores graus de incerteza devem ser aqueles que apresentam perfis de respostas com níveis de concordância bom ou justos entre os especialistas .

Todavia, pesquisadores mais experientes com a metodologia de construção de TCS (FOURNIER et al., 2008), recomendam a existência de itens com respostas consensuais entre os especialistas ao longo do teste. Isso provavelmente porque itens com essa característica poderiam aumentar a freqüência de respostas 1 e 5 (extremos na escala Likert) entre os especialistas.

Caso as respostas 1 e 5 fossem respondidas tão frequentemente quanto as outras, uma das fontes de erro de mensuração em testes no formato TCS, a desproporcionalidade entre as frequências de respostas do painel de referência ao longo da escala Likert, poderia ser diminuída. Estudantes ou outros grupos que respondam ao acaso as opções de resposta 2, 3 e 4 têm maior chance de obter maiores escores.

Uma das grandes dificuldades na construção do TCS é a ausência de respostas certas. Parece improvável que se consiga construí-lo de forma a garantir uma proporção semelhante de respostas com o valor de escore mais alto (igual a um), entre todas a opções da escala Likert, já que um dos problemas com esse tipo de escala é a tendência de centralidade em suas respostas. Nesse sentido, a presença de itens com respostas consensuais entre os especialistas

nos extremos da escala poderia amenizar, mas não extinguir, as conseqüências desse problema.

Há uma clara tendência entre os especialistas de responder as opções centrais na escala de respostas. Nesse estudo, foram obtidas as seguintes frequências de respostas, entre os especialistas, aos 104 itens: 91 (4,2%) respostas 1, 394 (18%) respostas 2, 759 (34,8%) respostas 3, 788 (36,1%) respostas 4 e 151 (6,9%) respostas 5.

Mais interessante ainda é observar esses resultados diante das frequências de respostas ao longo da escala entre os estudantes. Nessa amostra, obtiveram-se os seguintes resultados: 278 (6,6%) respostas 1, 781 (18,49%) respostas 2, 1193 (28,24%) respostas 3, 1569 (37,14%) respostas 4 e 403 (9,54%) respostas 5.

Esses dados corroboram resultados anteriores (Charlin, 2006) que indicam que é uma característica do TCS a maior proporção de respostas 3 entre os especialistas, a qual significa a ausência de impacto positivo ou negativo da informação sobre a hipótese diagnóstica, à investigação ou a terapêutica sugeridas nos itens. A opção de resposta 3 parece ser menos freqüentemente escolhida entre os novatos, que podem não ter uma opinião clara para decidir em situações de dúvida, o que dificulta decisões do tipo “nem mais, nem menos provável”. Essa opção de resposta tem um significado diferente de uma resposta com significado equivalente à “não sei” em uma escala. Optar pela resposta 3 não é uma escolha neutra.

Entretanto, nota-se que as respostas nos extremos da escala são pouco freqüentes, principalmente a opção com significados negativos, ou seja a resposta 1, principalmente entre os especialistas. Além disso, as respostas 4 e 5 conjuntamente foram mais assinaladas do que as respostas 1 e 2, confirmando a desproporcionalidade nas respostas entre os dois lados da escala e a preferência por respostas com correlações positivas.

Esse resultado pode ter dois significados. Primeiramente pode ser uma conseqüência de uma tendência, durante a construção do teste, de formar itens com associações positivas entre a proposição e a dedução. Em segundo lugar, pode haver entre os especialistas uma maior dificuldade, comparando-se ao

grupo de estudantes, em assumir posições mais precisas, discordantes ou de descarte, diante de incertezas, ou seja, optar pelas respostas 1 e 2. Nesse ponto, algumas teorias sobre tomada de decisões podem corroborar estes achados (VON NEUMAN e MORGENSTERN, 1947). De qualquer maneira, em geral, a tendenciosidade central nas respostas é uma fonte de erro sistemático em testes com escala Likert.

Portanto, as respostas aos extremos da escala Likert parecem ser ainda menos freqüentes entre os indivíduos mais experientes. Se comprovarmos que os novatos respondem significativamente menos a resposta 3 (central), esse fenômeno por si só poderia justificar parte da capacidade deste instrumento em diferenciar especialistas de estudantes. Essa desproporção entre as escolhas de opções de respostas ao longo do teste, no painel de referência, também pode prejudicar algumas etapas de validação dos resultados obtidos.

A carência de uma análise sobre a concordância das respostas entre os membros do painel de referência não permite prosseguir ainda mais nesse tópico de discussão. Um trabalho estatístico semelhante ao empregado no teste piloto necessitaria de recursos de análise que não dispunhamos para grupos maiores, como o grupo de especialistas.

6.2 Considerações sobre os grupos e a aplicação do instrumento

O desenho dessa pesquisa escolheu pela comparação entre dois grupos, segundo diferentes momentos de formação profissional. Primeiramente, diante da metodologia de avaliação educacional proposta, foi realizada a aplicação do teste para a formação de um painel de referência, formado por especialistas na área a que se refere o instrumento.

Segundo, o painel de referência deveria ser formado de acordo com os objetivos de aplicação do teste. Nesse sentido, um painel pode ser adequado para a comparação com alguns grupos, mas não com outros. Dependendo

também do grau de exigência da avaliação e da seriedade com que os resultados da avaliação serão levados em conta, esse painel pode ser modificado.

Como exemplo, poderíamos colocar a situação de uma avaliação para a obtenção de título de especialista em uma determinada área de atuação. Nessa ocasião, estaríamos avaliando pessoas potencialmente especialistas naquela área. Ainda, o objetivo desta avaliação, ou seja, determinar se um indivíduo é competente naquela área, é da mais alta responsabilidade.

Sendo assim, nesse particular cenário, o painel de referência, que não deixa de representar o julgamento sobre o potencial desempenho no teste, precisa ser formado por um grupo de excelência na área, capaz de oferecer as melhores soluções para as situações colocadas pelo teste.

Não há descrições ou discussões sobre o TCS referindo-se à questão da representatividade sócio-demográfica do painel de referência com relação aos especialistas em determinada área de atuação. Ou seja, a proporção de mulheres e homens, a faixa etária, entre outras características, não deve representar necessariamente o que se encontra dentro dessa especialidade em determinado momento ou local.

Por isso, o painel de referência precisa simplesmente ser um padrão, de modo relativo, em relação aos grupos comparados pelo teste. Não deve, necessariamente, representar de forma fidedigna o perfil da área de atuação a que se refere e não precisa conter somente ou obrigatoriamente os melhores especialistas desta área.

Outra questão a se detalhar seria o número de especialistas a formar esse painel. Gagnon et al. (2005) não coloca um número preciso ou ideal. Considerando-se as características psicométricas dos resultados encontrados por esse autor, um número mínimo de dez seria necessário para garantir a consistência interna dos resultados. Os números maiores que vinte contribuíram pouco para a melhora da acurácia do instrumento, mas de maneira alguma prejudicaram a análise.

Ainda com relação a essa questão, é preciso considerar que a maioria dos estudos sobre o TCS publicados formaram painéis de referência com especialistas de um único centro ou de poucos centros. Esta decisão pode ter

influenciado os resultados obtidos, pois garante uma maior homogeneidade entre os especialistas e limita o número de integrantes nos grupos formados. Outra explicação para essa limitação seria a dificuldade metodológica que se tem em aplicar o instrumento de forma padronizada em diferentes centros.

Além disso, esta pesquisa foi realizada com geriatras do estado de São Paulo. Mesmo considerando que suas características sócio-demográficas eram heterogeneas, principalmente a idade e o tempo de formação profissional, háque se dar relevância ainda maior ao fato que suas formações e exercícios profissionais têm um histórico bastante distinto. A geriatria em nosso meio não é uma área de saber técnico, seu ensino não é padronizado, os centros de formação podem divergir em suas prioridades e ênfases, e a sua prática profissional se abre para múltiplas atuações.

Portanto, esse estudo buscou elaborar uma proposta multi-cêntrica para a formação do painel de referência. Poderíamos ter alcançado mais do que oito centros participantes, com um número maior e, ainda assim, adequado de especialistas. Mas as dificuldades metodológicas impediram essa ampliação.

De qualquer maneira, as características do painel de referência formado não invalidam os resultados e as conclusões desta pesquisa. O número de vinte e um integrantes foi mais do que suficiente para demonstrar a variabilidade de opiniões e decisões dentro das questões elaboradas. A variabilidade do tempo de formação e experiência entre os especialistas poderia comprometer os resultados em um teste para título de especialidade, como exemplificado anteriormente. Mas demonstrou ser válido na comparação com os estudantes de graduação, como se vê nas evidências provenientes de nossos resultados.

O grupo de estudantes apresentou mais homens (58,5%) do que mulheres, tendência semelhante à do grupo de especialistas (76,2% de homens). Entretanto, o cálculo do qui-quadrado não mostrou diferença estatística entre homens e mulheres quanto ao desempenho obtido. Além disso, não há estudos sobre a influência desse fator sobre os resultados do TCS. Sendo assim, também não planejamos previamente uma determinada proporcionalidade entre os sexos, nem intra, nem inter-grupos.

Reconhece-se que não há estudos que diferenciam o raciocínio clínico entre homens e mulheres. Portanto, tentativas de comparação ou diferenciação quanto a essa característica não foram objetivos deste estudo. Contudo, pode-se questionar se alguma diferença nestas proporcionalidades poderia influenciar os resultados do teste.

Os estudos publicados sobre o TCS formam grupos de estudantes com números bastante diversos, entre 15 e 150 integrantes. O número da amostra de conveniência que conseguimos formar (41) está dentro desses limites. No entanto, números tão diversos permitem aferir que o TCS pode ser aplicado para avaliar grupos de estudantes de diversos tamanhos. Os números maiores podem contribuir para algumas características de validade, capacidade de replicação e acurácia dos resultados obtidos em pesquisas que buscam generalizar seus resultados.

Nesse sentido, amostras maiores de estudantes poderiam trazer outras interpretações aos resultados. Porém uma série de limitações metodológicas e cronológicas impediu a formação de um grupo maior. Todavia, nossas análises dos resultados demonstram que muito pode se compreender sobre as características desse grupo diante do TCS com base na amostra obtida. Nosso principal objetivo de avaliação nesta pesquisa, diferenciar especialistas de estudantes, também foi claramente alcançado.

6.3 Considerações sobre a validade dos resultados

Todas as avaliações em educação médica precisam de evidências de validade para ser interpretadas significativamente (DOWNING, 2003). Porém, a validação é um processo contínuo. Mais ainda, fornecendo valor, qualidade e fundamentos à interpretação dos resultados, é um processo que jamais termina (CRONBACH e MEEHL, 1955).

O que se valida não é o teste em si, mas uma interpretação que se faz dos resultados de sua aplicação. Por isso, a validação é um teste de hipóteses.

Qualquer tentativa de análise ou de interpretação de um instrumento de avaliação educacional é, inevitavelmente, uma proposta de validação, que terá determinadas limitações ou direcionamentos específicos, de acordo com as metodologias e teorias utilizadas. Os dados serão válidos para uma proposta em particular, um significado ou uma interpretação específicos, em uma determinada situação, em um momento único, para uma população o mais bem delimitada possível.

Por isso, embora não seja um objetivo explícito desta pesquisa, validar os resultados da aplicação do teste é o caminho natural e necessário no percurso da análise a ser empregada. Elementos como o desenho do estudo, os grupos envolvidos, o tamanho e a especificidade do teste desenvolvido, as ocasiões de aplicação e os recursos e técnicas de análise contribuem e ao mesmo tempo trazem limitações para o potencial de compreensão e de conclusão a respeito dos resultados obtidos.

O próprio percurso de análise a ser empregado é apenas uma alternativa entre tantas outras opções já conhecidas ou ainda à espera de serem elaboradas ou descobertas para compreensão do fenômeno. Portanto, o processo de validação, como recurso que fundamenta a interpretação de um teste, sempre estará inacabado.

Conceitualmente, validade é um significado unitário que procura por múltiplas fontes de evidência (KANE et al., 1999). Atualmente, todas as validações são consideradas validações de constructo. No passado, a validade era classificada em três diferentes tipos: de conteúdo, de critério e de constructo. A validade de critério era ainda subdividida em validades concorrente e preditiva, de acordo com os momentos de coleta dos dados.

A justificativa para que todos os tipos de validade sejam, hoje, considerados validade de constructo provém de conceitos atuais da filosofia da ciência. Em qualquer circunstância, haveria muitas teias ou conexões de inferências inter-relacionadas, associadas com os conteúdos de uma amostragem ou de uma aplicação de um teste. Isto seria suficiente para levar a conclusões significativas e viáveis para um domínio, área de conhecimento ou população de maior de interesse. Ou seja, todas as avaliações em educação médica lidariam

com constructos, que são conceitos abstratos ou princípios inferidos a partir de comportamentos e explicados por teorias psicológicas ou educacionais.

Um constructo é um atributo postulado das pessoas, refletido em um desempenho em um teste (CRONBACH e MEEHL, 1955). Por isso, os resultados do processo de ensino-aprendizagem, ou da educação propriamente dita, são constructos ou abstrações que podem ser inferidas a partir do desempenho que consideramos em instrumentos de avaliação elaborados em suas mais diversas maneiras. A aptidão em uma determinada área de atuação é um constructo ainda mais abstrato, difícil de ser definido e mensurado tanto por testes padronizados e objetivos, quanto por metodologias que consideram múltiplos avaliadores e opiniões subjetivas.

Os testes educacionais não são válidos ou inválidos. O que pode ter mais ou menos evidência para ser defendido é o escore obtido pela aplicação do teste, de acordo com os objetivos de avaliação. Em nosso caso, o objetivo foi o de interpretar as diferenças de desempenho entre os indivíduos, segundo suas competências para o raciocínio clínico em situações de incerteza em geriatria, de acordo com as situações propostas.

A validade percorre um caminho longo e com múltiplos obstáculos. É um conceito sempre buscado, mas nunca assumido, complexo. Envolve a geração de hipóteses, a coleta e a análise dos dados, uma avaliação crítica e uma dedução lógica. O argumento de validade correlaciona teorias, conceitos pré-determinados e evidência empírica, de modo a sugerir que algumas interpretações podem ser mais coerentes que outras para a proposta de aplicação do teste. Segundo a publicação da American Educational Research Association (1999, p. 9):

> “A validade se refere ao grau com que a evidência e a teoria fornecem suporte às interpretações oferecidas aos escores, relacionados com uma proposta de utilidade para o teste.”

É preciso, portanto, aprimorar o desenvolvimento do conteúdo, do processo de resposta, da análise da estrutura interna, do significado dos

resultados em comparação com outras medidas anteriormente obtidas em testes com objetivos e formatos semelhantes e também em testes com propostas diferentes, e discutir as justificativas e as conseqüências do desenvolvimento e da aplicação do instrumento. Buscar as diferenças entre os especialistas e os estudantes e fundamentar alguns sentidos para essas diferenças é etapa primária e essencial desse processo. Esse foi nosso objetivo principal de aplicação do TCS.

Algumas fontes de validade, dependendo do propósito e da interpretação do instrumento, poderiam ser: a representação de conteúdo, a reprodutibilidade e a generalização dos escores, as características estatísticas das questões ou das formas de resposta, as relações estatísticas entre outras medidas do mesmo construto ou de constructos semelhantes, as evidências do impacto do escore sobre os estudantes e a consistência dos critérios de aprovação ou reprovação segundo os escores obtidos (DOWNING, 2003).

Se um teste diferencia indivíduos por meio de seu método de pontuação e este teste está de acordo com os níveis de experiência e o raciocínio clínico que os tornam diferentes, logo, portanto, esse instrumento pode estar detectando, entre tantos outros elementos, a capacidade de raciocínio desses profissionais. Porém, o instrumento e seus resultados não são o fenômeno propriamente dito e devem ser compreendidos diante de suas limitações e imperfeições para a detecção e a interpretação de determinado nível de competência.

Claramente, todavia, conforme desenvolvido na introdução, a competência para o raciocínio clínico em situação de incerteza não é uma característica isolada. Pode ser considerada como um elemento interdependente de múltiplos outros fenômenos que podem ser aferidos ou detectados, mas nem sempre diferenciados, isolados ou quantificados, por um teste. Os conhecimento prévios, as experiências pessoais, educacionais e profissionais anteriores, os aspectos emocionais, entre tantos outros, são elementos que se confundem com o processo de raciocínio e de tomada de decisões. Portanto, aferir de forma precisa e isolada essa competência seria uma proposta inverossímil.

Mais ainda, é preciso considerar que uma das maiores ameaças à medida confiável em um exame de performance é a especificidade do caso (DOWNING, 2004). As avaliações complexas de desempenho podem necessitar de modelos de análise ainda mais elaborados e detalhados para se obter o grau de confiança de seus resultados. O número de situações ou casos clínicos dispostos ao longo do exame será sempre inferior à multiplicidade de cenários que a realidade para aquela situação pode envolver.

Então, afinal, o TCS é uma medida escrita de cognição ou é uma medida de desempenho clínico? Segundo Schuwirth e Vleuten (2004), nem uma coisa, nem outra. A proposta do TCS é única, na medida em que assume o desafio de ter um formato híbrido. Ele está no limite entre três características.

Primeiramente, ele se aproxima de um teste com um formato objetivo de respostas, ou seja, um teste escrito de múltipla escolha. Em segundo, seu método de julgamento, por meio de um escore numérico, formado a partir da organização e quantificação sistemática de decisões qualitativas e subjetivas de um grupo de especialistas, permite a comparação com decisões de um painel de referência, que representa os avaliadores ou juízes do desempenho clínico. Todos os exames que pretendem dimensionar performance devem ter seu escore baseado na opinião de especialistas (CLAUSER, 2000).

E o formato de estímulo ou desafio para a sua resolução, ou seja, casos clínicos genuínos e envolvendo contextos de incerteza, é representativo de sugestões mais elaboradas de medida de performance profissional. Quanto maior a diversidade de situações, maior a chance dos resultados refletirem o universo de possibilidades para esses elementos.

Esta singularidade na miscigenação de características provenientes de diferentes metodologias de avaliação faz com que a análise psicométrica do TCS necessite de métodos mais abrangentes, amplos e diversos para a garantia da acurácia e da consistência de seus resultados. Os testes de múltipla escolha, por exemplo, precisam somente de medidas de consistência interna na análise da precisão de seus resultados. Porém, as avaliações de desempenho profissional não podem abdicar de estudos que considerem a complexidade de suas mensurações, como as análises provenientes da teoria G. A interface do TCS

com essas múltiplas formas de avaliação educacional não pode se abster de análises mais pormenorizadas, que considerem as múltiplas propostas que são base para a criação do instrumento.

Quanto mais complexa a forma de avaliação e quanto mais abrangente o seu objetivo, como, por exemplo, os exames que buscam avaliar competências práticas, mais diversas devem ser a propostas para validá-la. É difícil desenvolver um teste objetivo para avaliar competências profissionais, já que essas são intangíveis e complexas. Além disso, a performance profissional varia de local para local, de caso para caso e de um momento para o outro (ERICSSON, 2007).

A validade de um dado depende de sua acurácia e de sua precisão. A acurácia é o grau de veracidade de uma medida, enquanto a precisão é o seu grau de reprodutibilidade. A acurácia se verifica através de respostas a perguntas como por exemplo: este teste realmente mede o raciocínio clínico em contextos de incerteza? As respostas a uma questão como essa envolvem várias etapas de validação por meio de desenhos de pesquisa que possam comprovar que os resultados do teste estão de acordo com as teorias cognitivas e as teorias sobre o raciocínio clínico que o fundamentam. Nossos resultados comprovam que os especialistas apresentaram desempenho superior ao dos estudantes, de acordo com a teoria de scripts, as teorias sobre o raciocínio hipotético-dedutivo e as conclusões sobre o reconhecimento de padrões na resolução de situações práticas na atividade médica. Com isso, agregamos validade e acurácia ao instrumento proposto.

Reproduzir os resultados indefinidamente reflete diretamente na capacidade do teste para diferenciar os examinandos de maneira adequada. Sendo a precisão o grau de reprodutibilidade dos dados obtidos, o cálculo da consistência interna, por meio da correlação parcial-total entre os itens ou entre as situações, e do coeficiente de alfa de Cronbach, é capaz de analisar a confiança que podemos ter nos resultados por meio de escores. Essa qualidade dos dados é uma das principais fontes de evidência para a validação de um instrumento de avaliação em educação médica. Os níveis baixos de consistência interna indicam que grandes variações podem ser esperadas em uma reaplicação

do teste nos mesmos grupos estudados, o que inviabilizaria sua aplicabilidade e interpretação.

Em contrapartida, um valor de consistência interna elevado indica que os escores seriam os mesmos se o teste fosse reaplicado em um outro momento. Além disso, os erros ao acaso, como aqueles que resultam de cansaço do examinando, desatenção, diferenças de performance para o mesmo indivíduo ao longo do tempo, escolha de respostas por aposta, entre outros, serão suficientemente baixos para garantir a validade do escore obtido, ou seja, a sua capacidade de ser reproduzido.

Alcançar uma boa consistência interna em uma avaliação profissional é um desafio por dois motivos: o papel profissional é composto de comportamentos complexos, e o comportamento profissional depende da natureza e dos detalhes do problema a ser resolvido, o que é chamado de caso-especificidade (DOWNING, 2004).

A análise da correlação parcial-total entre os itens do escore total para os 104 itens resolvidos considerou somente os itens com correlação igual ou superior a +0,19, de acordo com Bisquerra et al. (2004). Com isso, garantimos que um bom número de itens (61), para serem resolvidos em cerca de um hora de teste, pudesse ser considerado adequado para a uma possível reaplicação do instrumento.

Embora o coeficiente de alfa de Cronbach para o teste inteiro (104 itens) fosse de 0.842, seria adequada a busca por coeficientes próximos ao valor de 0.90. Vários autores e escritores de livros-texto discutem com uma variedade de opiniões sobre o melhor valor de alfa (DOWNING, 2004). Muitos profissionais da área de avaliação educacional sugerem alfas de no mínimo 0.90 para exames “high stakes”, ou seja aqueles que devem aprovar os examinandos para ocupações ou obrigações de grande responsabilidade, como podem ser considerados o exercício da medicina, a obtenção de um título de especialista ou a aprovação em um concurso para admissão profissional.

O alfa do teste com 61 itens selecionados pela correlação parcial-total foi de 0.882, mais próximo de 0.90. Para exames com menor grau de exigência

ou responsabilidade, como avaliações somativas de final de estágio ou de final de um curso, por exemplo, os valores maiores ou iguais a 0.80 são suficientes.

Foram também obtidos os valores denominados de escores por situação, por meio da média dos valores de escore de cada item, após a somatória dos escores por itens em cada situação. Além de permitir a exploração de fontes de erro ainda desconsideradas em publicações anteriores sobre o TCS, como a imprecisão do escore provocada pela interação item-item, esse artifício permitiu uma visão crítica mais fundamentada a respeito das limitações do coeficiente de alfa de Cronbach para a análise do escore por itens. Sobre isso, Cronbach e Shavelson (2004, p. 403) assim se pronunciaram:

> “A fórmula de alfa não é estritamente apropriada para muitos testes construídos de acordo com um plano que aloca algumas frações de itens a um particular tópico ou processo.”

Sendo assim, as análises anteriores sobre o TCS não consideraram que a reunião de itens em torno de um caso ou uma situação clínica pudesse comprometer a veracidade do cálculo do alfa por itens. Nesse formato de teste, a interação entre os itens na mesma situação é maior do que a interação entre todos os itens do teste entre si e fornece uma fonte de erro ao escore verdadeiro. As instruções aos responsáveis pela construção do teste e ao examinandos sobre a independência de um item em relação aos outros pode ser incapaz de eliminar completamente essa fonte de erro.

Como foi discutido anteriormente. Um item do teste não tem significado isoladamente. Ele só existe em relação à situação clínica a que ele se refere. Portanto, itens relacionados à mesma situação clínica apresentam mais componentes em comum do que itens de situações clínicas diferentes. Além disso, a proximidade física dos itens de uma mesma situação durante a resolução do teste dificulta o “isolamento cognitivo” necessário para que eles não sejam considerados como sendo relacionados entre si.

O coeficiente de alfa de Cronbach calculado para as treze situações foi de 0,78. Descartando-se a situação onze, que apresentou uma correlação parcial-total inferior à + 0,19, alcançou-se o coeficiente de alfa de Cronbach de 0,80. Esse coeficiente foi inferior àquele obtido pelo escore para 61 itens, que foi de 0,882.

Essa metodologia de análise dos resultados da aplicação do TCS trouxe indagações diferentes sobre o escore obtido. Primeiramente, surge a seguinte pergunta: deve-se considerar o escore por itens ou o escore por situações? Em segundo lugar: qual é o grau de importância da interação item-item no desempenho obtido pelos examinandos?

Considerando-se uma das principais teorias em que se baseia o teste, ou seja, a teoria de scripts, o escore por itens ainda parece ser adequado para a medida dos resultados do TCS. Vale lembrar que cada item colabora com a ativação específica de um script. Mas os scripts podem estar de alguma forma relacionados ou interligados entre si, embora isso não impeça que um item seja um componente único, que poderia ser detectado ou aferido. Nesse sentido, considerar apenas os resultados de escores por situações também pode limitar a compreensão sobre o processo de raciocínio clínico que o teste busca avaliar.

Porém, há uma complexidade inerente a esse processo de raciocínio que impede uma análise absolutamente individualizada de cada item, representante de cada unidade de script. A interação entre sistemas, ou seja, entre diferentes estímulos que compõem o processo de tomada de decisão, pode ser comprovado pela presença, em nossa análise, da interação item-item.

O método que utilizamos, ou seja, o cálculo e a análise do escore por situações, é apenas uma forma indireta de constatar que há essa interação. Alguns estudos G mais elaborados, que considerassem mais um componente de variância ou faceta para a interação item-item, poderiam quantificar esse componente de erro no escore observado.

Portanto, o alfa por itens subestima as fontes de erro no escore do teste. Porém, a análise isolada de cada item permite diferentes interpretações sobre os resultados obtidos, propondo correções no instrumento para novas aplicações subseqüentes, por meio de alterações apenas nos itens, e não nas situações como um todo.

É interessante observar que, se considerássemos apenas o escore por situações, seria a retirada de uma situação inteira (situação onze) que aumentaria a consistência interna dos resultados. A qualidade isolada dos itens deixaria de ser considerada para a tomada desta decisão.

Ainda assim, o alfa de 0.80 torna a reprodutibilidade do escore por situações inferior à do escore por itens. Isso se deve principalmente ao número de escores por situações (doze escores com adequada consistência interna), que é bastante inferior ao número de escores por itens (sessenta e um itens com adequada consistência interna). O número de questões de um teste é um dos principais fatores relacionados a mudanças no coeficiente alfa Cronbach (CRONBACH e SHAVELSON, 2004).

Entretanto, conforme introduzido anteriormente, o cálculo da consistência interna, a partir da teoria clássica de mensuração, é insuficiente para análise mais detalhada da precisão de um teste que pretende de alguma maneira predizer o desempenho profissional dos examinandos. A teoria moderna, por meio da aplicação de estudos G, desenvolvida por Cronbach e sua equipe, a partir de 1972, permite uma melhor compreensão quanto a formatos de testes que pretendem ir além de uma simples medida cognitiva ou de conhecimento absoluto.

A evidência de incerteza ligada ao escore de uma escola ou de um indivíduo é necessária para se evitar o excesso de interpretação dos dados. A análise G permite inferir o grau de dúvida ligado ao escore. Determinar um escore absoluto pode penalizar o examinando ou a escola. O grau de incerteza tolerável em uma avaliação dependerá das recompensas e penalidades associadas com a aprovação ou reprovação do desempenho aferido (CRONBACH et al., 1997).

Cronbach ampliou o conceito de precisão de uma medida educacional por meio da introdução da teoria G. A partir de então, não foi mais preciso decompor a variação de um escore observado em somente dois componentes, como era feito a partir da teoria clássica (SHAVELSON e WEBB, 1991). A variação de um escore não seria mais atribuída somente às diferenças verdadeiras entre os examinandos (escore verdadeiro) e à variação devido a múltiplas fontes de erro, sistemáticas ou ao acaso, que incluíam as variáveis omitidas, a interação entre os examinandos e os componentes de avaliação, além de outras contribuições de medidas individuais que não seriam de interesse da proposta de avaliação do teste.

Esse conjunto de medidas obtidas por um escore, que está além dos interesses de medida do instrumento, pode ser considerado como fontes de erro, que podem ser dissecadas e quantificadas pelo método de análise de variâncias (ANOVA). Com isso, as estimativas provenientes da teoria G podem prescrever novas decisões para próximas construções ou modificações do teste, controlando as variações das fontes de erro, o que foi chamado de teoria D.

A teoria G (“Generalizability theory”) é um método estatístico que envolve o conceito de dependência (“dependability”) de uma mensuração comportamental. Esse conceito se refere à precisão com que o escore observado de um examinando pode ser generalizável para uma média de escore que este mesmo examinando obteria em todas as condições, questões ou itens que o teste seria capaz de formular.

As condições para que essa hipótese seja válida seriam que os conhecimentos, as habilidades e as competências envolvidas para a resolução do teste fossem relativamente estáticas (“steady state”) e que qualquer diferença de escore obtido por um determinado examinando em diferentes ocasiões fosse resultado de fontes de erro, e não de variações individuais de desempenho. Não haveria, assim, o efeito de memória, de esquecimento ou de aprendizagem sobre os resultados de uma nova aplicação do teste.

Como uma medida real em um teste educacional jamais preencherá todas essas condições para ser completamente generalizável, os escores observados sempre apresentam fontes de erro. A precisão absoluta é sempre buscada, mas nunca alcançada. É preciso se conhecer esse universo de componentes de variância do escore (facetas), que podem estar prejudicando a precisão da medida.

Quanto mais complexa uma análise, maior o número de facetas que podem ser consideradas em um estudo G. Quanto mais ampla, também, a tentativa de generalização dos resultados obtidos, maior o número de fontes de erro que devem ser considerados. Se quisermos generalizar somente para os itens que podem ser desenvolvidos para este tipo de teste, podemos considerar somente a faceta “itens” como uma fonte de erro. No caso de considerarmos a generalização para outros grupos de examinandos, precisamos considerar a faceta grupos como mais uma fonte de erro para o escore observado.

Nesta aplicação da teoria G, o objetivo foi considerar o componente itens como uma única fonte de erro na medida do instrumento. Por isso, o desenho da análise pode ser denominado de “um universo com uma faceta”. Com isso, poderíamos analisar as qualidades e as limitações do instrumento, quanto às características psicométricas de seus itens ou situações.

Ainda assim, um desenho de estudo com uma única faceta tem quatro fontes de variabilidade. Uma dessas fontes seriam as diferenças sistemáticas nos desempenhos obtidos pelos examinandos (componente *p* da variância ou escore universal), que representa o objetivo de nossa avaliação e refletiria a medida do raciocínio clínico em contextos de incerteza.

A segunda fonte de variabilidade surge a partir das diferenças nos níveis de dificuldade entre os itens ou situações, dependendo do tipo de escore considerado. Quanto maior a variação no grau de dificuldade entre os itens, mais difícil é a generalização dos resultados da aplicação do teste, ou seja, menos próximo da média dos escores do universo se torna o escore observado.

A terceira fonte de variação do escore aparece a partir de experiências de vida ou vivências educacionais que os examinandos poderiam trazer para a resolução do teste. Se um dos participantes da pesquisa já experimentou alguma situação semelhante àquela descrita no teste, a resolução dos itens desta situação se torna mais fácil e mais significativa para este participante.

Esta interação entre os itens ou as situações e o examinando pode trazer mais variações de desempenho individual ao longo de todo o teste, ou seja, o desempenho em um item ou em uma situação anterior não será preditivo de desempenhos seguintes ao longo do teste. Esta variabilidade também dificulta a generalização por itens ou por situações.

A última fonte de variabilidade em um desenho com uma única faceta seria formada por um conjunto de fenômenos ou componentes, sistemáticos ou ao acaso, capazes de produzir imprecisão na medida, como lapsos de memória ou desatenção, diferentes examinandos resolvendo o teste em locais e em dias diferentes, um toque de telefone celular ou uma conversa durante a resolução, entre tantos outros. Entretanto, o cálculo do terceiro e do quarto componentes não podem ser estatisticamente desemaranhados. Consequentemente, o cálculo do componente "*p* x *i*" considera esses dois componentes em conjunto.

Inicialmente, na aplicação do estudo G, a faceta grupo foi considerada uma faceta fixa, ou seja, não foi uma fonte de erro para o escore. Com isso, buscamos generalizar os resultados especificamente para cada um dos grupos, isoladamente, e não para todos os examinandos, como se fossem um conjunto único.

Essa estratégia também permite comparações sobre determinados comportamentos de resposta que poderiam ser diferentes entre os grupos. Observar essas diferenças e discuti-las levaria a conclusões a respeito de outros indícios de validade ou de acurácia dos resultados, bem como poderia inferir algumas conclusões sobre as teorias que fundamentam o instrumento. Segundo Shavelson e Webb (1991), não há sentido em nivelar as condições de uma faceta fixa se as conclusões sobre esse nivelamento são de pouco interesse para cada condição desta faceta e para os objetivos da pesquisa.

Desse modo, o estudo G com desenho P/I para os escore por itens, considerando o 61 itens com adequada consistência interna, demonstrou, de uma maneira geral, uma predominância da participação do componente de resíduo ("*p* x *i*") como fonte de variabilidade do escore (98% para os especialistas e 89,3% para os estudantes).

Conforme discussão anterior, esse componente de variância é constituído por uma série de fontes de erro. Porém, provavelmente, um dos mais preponderantes é a interação entre os examinandos e os itens, que passa a ganhar importância para a interpretação do escore neste TCS.

Essa interação se reflete em algumas interpretações sobre os resultados obtidos. Primeiramente, ela fornece indícios de que o desempenho individual de cada examinando ao longo do teste foi bastante variável ou instável, ou seja, imprevisível. Parece que a pontuação em um determinado item não foi preditiva do desempenho em itens subseqüentes. Caso o raciocínio clínico em situações de incerteza seja, de fato, caso-específico, como se afirmou em outras pesquisas sobre esse tema, esse comportamento ao longo do teste pode ser uma fonte de validade para a hipótese de medida desta competência pelo teste por nós desenvolvido.

Além disso, um grande componente de interação entre pessoas e itens indica que os examinandos utilizam-se freqüentemente de experiências educacionais e profissionais anteriores para a tomada de decisões durante a resolução do teste. Essa conclusão fundamenta a hipótese de que o instrumento é capaz de medir os scripts que são formados a partir de vivências individuais. Porém, esse resultado limita o uso do TCS como instrumento que pode ser aplicado isoladamente para a avaliação dos resultados de uma intervenção ou de um programa educacional, por exemplo. O processo de ensino-aprendizagem do raciocínio clínico em contextos de incerteza parece seguir pressupostos de complexidade, ou seja, é errático, não-linear, interativo, de difíceis previsões ou planejamentos.

Entretanto, deve se considerar também que o componente *p* x *i* também inclui diferenças de escore devido a outras fontes de erro, como a interação item-item, que foi mais bem observada após a análise do escore por situações.

Outro elemento desse componente poderia ser a mudança freqüente de ocasiões para a aplicação do teste, já que cada especialista, por exemplo, resolveu o teste em um dia e em um local diferente. Os estudantes participaram da pesquisa em vários grupos menores, o que diminuiu a variabilidade de ocasiões e ainda fez com que conseguíssemos aplicar o instrumento sempre em ocasiões iguais ou semelhantes, já que todos estudavam na mesma universidade.

Assim, o componente de variância "*p* x *i*" foi a principal fonte de erro no escore obtido por essa pesquisa. Isso não significa que ele seja o mais importante. Não se pode considerar também que ele deve ser eliminado. O nível de importância de um componente de variância do escore depende dos objetivos de avaliação. Contudo, esse fato pode trazer limitações ao instrumento, pois pode dificultar a interpretação e os objetivos de avaliação do desempenho dos examinandos. Todavia, não cabe a essa pesquisa responder definitivamente a esta questão.

Como foi colocado anteriormente, as interpretações provenientes do componente *"p* x *i"* de variabilidade do escore estão de acordo com algumas hipóteses sobre o raciocínio clínico do especialista e do raciocínio clínico em contextos de incerteza. A interação entre o profissional e a situação que ele vivencia e o significado dessa interação passam a ser essenciais no processo de ensino-aprendizagem desta competência. Parece que esse raciocínio é baseado em experiências práticas anteriores, difíceis de serem planejadas em propostas curriculares.

Cabe detalhar ainda que o componente de resíduo para os especialistas (98%) apresentou participação maior como fonte de variabilidade do escore neste grupo do que no grupo de estudantes (89,3%). Essa diferença parece confirmar a hipótese de que os especialistas utilizariam com maior freqüência e importância o reconhecimento de padrões para a tomada de decisões, com base em suas vivências anteriores, do que os estudantes. O processo de ensino-aprendizagem dessa competência, entre os especialistas,

parece ser ainda mais errático, imprevisível e não-linear. Baseia-se, portanto em experiências, circunstâncias ou acontecimentos invariavelmente ao acaso e incontroláveis. Nesse ponto, pode-se supor que o planejamento de propostas de ensino deste tipo de raciocínio é de difícil previsão quanto a seus resultados.

Portanto, se é uma característica inerente ao teste a participação da interação *"p* x *i"* na formação do escore, como assumir o TCS como uma medida válida e precisa? Não se pode esquecer que fontes de erro no escore reduzem os valores do coeficiente de alfa de Cronbach e do coeficiente G. Com isso, torna-se difícil crer que o TCS possa ser utilizado como medida única ou isolada em exames que exijam pontuações mais rigorosas ou precisas, com capacidade de aprovação ou reprovação de estudantes ou instituições.

Porém, reconhece-se que, se um instrumento é somente uma parte de um processo de avaliação mais amplo e abrangente, ele pode apresentar coeficientes de alfa e coeficientes G menores, sem prejudicar os resultados dos examinandos. Nesse caso, a precisão de um teste como um todo é maior que a precisão de suas partes.

Essa constatação vai de encontro com propostas mais recentes e cada vez mais freqüentes de avaliação em finais de cursos, títulos de especialidade, entre outros, que exigem maiores níveis de precisão e validade, mas que se utilizam de recursos ou formatos de teste nem sempre tão confiáveis quanto o desejado. Nesse sentido, o TCS se insere adequadamente como mais uma opção de formato de teste educacional na área de saúde, que pode fazer parte de um exame mais amplo, como fragmento de um processo que avalie múltiplas competências.

E quanto ao significado do componente *i* de variabilidade para o escore por itens? Qual a sua importância? De acordo com a teoria G, a diferença no grau de dificuldade entre os itens não provoca erro no escore, se este for considerado de forma relativa, ou seja, para classificar os examinandos conforme suas posições em relação aos outros do mesmo grupo.

Mas, e se for preciso saber o quanto o desempenho de um indivíduo, de um grupo ou de uma escola é bom ou ruim? Como saber o escore absoluto de cada examinando e medir sua distância em relação aos outros? É possível se

definir notas de corte? Pode-se dizer quem foi aprovado ou quem foi reprovado pelo teste? A resposta a essas perguntas não pode ser esclarecida pela teoria clássica de medidas de um teste.

Todavia, por meio de considerações sobre componente *i* da variância e do coeficiente G absoluto, é possível ter indícios sobre a capacidade do escore em responder a essas questões. Mais ainda, pelos cálculos do estudo G, pode-se definir o erro padrão e o intervalo de confiança do escore, que são medidas de interpretação para a definição de escores absolutos e da distância entre os examinandos. O erro-padrão de um escore segue sempre uma distribuição de normalidade e pode definir a probabilidade de uma nota estar errada e o quanto ela pode estar errada. Ou seja, é possível saber o quanto o escore absoluto está preciso e qual a chance de ele estar injusto.

De qualquer forma, é ainda um desafio para as pesquisas sobre o TCS definir escores absolutos para a interpretação dos resultados. Por enquanto, o teste foi desenvolvido para a análise de grupos, de forma relativa, considerando-se o desempenho de uns em relação aos outros. Em nossa pesquisa, se pudéssemos considerar os escores absolutos para os examinandos, poderíamos concluir que a participação do componente *i* como fonte de erro do escore foi muito maior para os estudantes (10,7%) do que para os especialistas (1,7%).

Essa diferença demonstra que a variabilidade no grau de dificuldade dos itens é diferente entre os grupos. Ainda mais, essa variabilidade no grau de dificuldade dos itens influencia mais o escore dos estudantes do que o dos especialistas. Sendo assim, para se aumentar a precisão do escore dos estudantes, deveríamos homogeneizar o grau de dificuldade dos itens para esse grupo, priorizando itens de moderada dificuldade.

A partir dos cálculos das variâncias para as três fontes envolvidas (*p*, *i*, "*p* x *i*"), obteve-se a porcentagem de variância de 4,4% para o componente *p* entre os especialistas e 6,1% entre os estudantes. Porcentagens relativamente baixas em relação aos outros componentes para o componente *p* indicam que o escore individual por examinando variou pouco entre os examinandos de um mesmo grupo. Sendo maior esta variabilidade entre os estudantes, pode-se inferir que as respostas deste grupo podem estar mais dispersas ao longo da escala

Likert, embora possa haver outros motivos para que isso aconteça. Parece que o desempenho no teste é menos homogêneo entre os novatos. Os cálculos de frequências de opções de resposta, que fora discutido anteriormente, também já tinham comprovado esta hipótese.

Isso se torna mais evidente quando comparamos as análises obtidas após a formação dos escores por situações. Nesse estudo, o componente *p* para os especialistas foi de 11,3% para o escore por situações. Para os estudantes, esse componente foi de 12,5%. Embora tenham se mantido as posições relativas de diferença entre esses valores, observam-se que eles são superiores às porcentagens do componente *p* para os itens.

Com essa análise, podemos inferir que o escore observado por situações, apresenta maior variabilidade do que o escore por itens, para cada examinando, dentro dos grupos. Nesse sentido, pode ser que o escore por situações diminua o efeito de tendência central das respostas na escala Likert, que é uma das principais fontes de erro em testes que usam essa escala.

Além disso, os dados com o escore por situações confirmam ainda mais que o raciocínio clínico é predominantemente contexto-específico. O componente *p* x *s* para o escore por situações permanece alto (91,6% para os especialistas e 82,2% para os estudantes), embora relativamente inferior às porcentagens deste componente no escore por itens.

Nessa comparação, há ainda alguns detalhes evidentes. Parece claro que, observando todas as porcentagens dos componentes de variabilidade, pode se inferir que o componente "*p* x *s*" foi menor no escore por situações do que o componente "*p* x *i*" para o escore por itens. Assim, espera-se que os outros componentes de variabilidades tenham aumentado. Com relação ao componente *p*, algumas interpretações já foram discutidas anteriormente, demonstrando que o escore universal por situações (componente p) é proporcionalmente maior do que o escore universal por itens.

Está nítido, portanto, que estes componentes estão inter-relacionados, não só em números, mas também em significados e interpretações. No escore por situações, conseguimos reduzir a principal fonte de erro, tanto relativo quanto absoluto, para a medida do teste, que é o resíduo do escore ("*p* x *i* "). Se esse

resíduo é menor, pode-se dizer que a interação entre situações e pessoas é menor do que a interação entre itens e pessoas? É difícil afirmar isso, comparando-se números absolutos de medidas relativas. A participação relativa do resíduo sobre o escore foi menor. Em números absolutos, não se pode ter certeza se isso reduziu a interação entre os itens e as pessoas.

Ainda nessa análise, o componente "*p* x *s*" para os especialistas continua sendo mais alto do que para os estudantes no escore por situações, o que demonstra que esse grupo interage mais com os casos ou situações do que os estudantes, talvez por familiaridade com os contextos envolvidos, conforme comprovações anteriores sobre as teorias de raciocínio do expert.

De uma maneira geral, um resíduo menor poderia aumentar a precisão do instrumento, o que se refletiria em coeficientes G mais elevados. Conforme será discutido adiante, isso não foi constatado, provavelmente porque o componente *s* foi mais elevado do que o *i* para ambos os grupos, o que prejudicou a elevação dos coeficientes G absolutos. O número baixo de situações comparado ao número de itens também prejudicou a precisão do escore por situações e reduziu ainda o coeficiente G relativo, que é matematicamente e significativamente semelhante ao coeficiente de alfa de Cronbach. Resultados semelhantes já haviam sido observados nos cálculos de consistência interna.

Todavia, outra interpretação pode ser considerada na redução do resíduo encontrada no cálculo do escore por situações. Com essa estratégia, buscamos reduzir a interação item-item. Se houve redução no componente equivalente ao "*p* x *i*", ou seja, o componente "*p* x *s*" com esse novo método, um dos motivos pode ser a redução da influência dessa interação sobre o escore observado. A confirmação dessa hipótese pode confirmar estatisticamente a presença de interações entre os itens. Mais estudos são necessários para a comprovação do significado destes achados.

Uma conclusão como essa permitiria fundamentar que realmente essa interação item-item é importante na formação do escore e pode gerar erros em sua medida, já que as instruções para a resolução do teste determinam que os itens sejam independentes. Porém teorias advindas da psicologia da educação reconhecem que essa independência de fato jamais existiria. De qualquer

maneira, a influência desse fenômeno sobre uma medida mais precisa de um instrumento poderia ser amenizada. A redução do número de itens por situação clínica pode reduzir essa interação.

Portanto, a mudança do escore por itens para o escore por situações parece reduzir o resíduo no escore. Entre outras fontes de erro que podem ter prejudicado a formação do escore por itens, a interação item-item parece ser relevante. Esse dado pode contribuir para novas propostas de pesquisa sobre o TCS, buscando melhores qualidades psicométricas à formação de seu escore.

Quanto ao componente *s* de variabilidade para o desenho P/S do escore observado por situações, esse se mostrou relativamente maior do que o componente *i* do escore por itens. Mantendo-se a porcentagem maior desse componente no grupo de estudantes (17,8%) em comparação com o grupo de especialistas (8,4%), o aumento dessa porcentagem em ambos os grupos em relação ao componente *i* indica que o instrumento com escore por situações parece ter questões mais difíceis do que o instrumento com escore por itens. Parece que é mais difícil acertar a avaliação por situações nos dois níveis de experiência. Nesse sentido, lidar com as situações como um todo parece ser um desafio maior do que fragmentá-las em itens.

Muitas discussões podem ainda ser levantadas sobre o real significado desta última conclusão. As teorias sobre o raciocínio humano confirmam a necessidade cognitiva de se fragmentar o problema para a tomada de decisões. Considerando essa hipótese em nosso instrumento, esta análise da comparação do componentes i e s de variabilidade do escore confirmam a validade cognitiva do TCS.

Além disso, esse dado indica que provavelmente o grau de dificuldade das situações influencia mais o escore obtido do que o grau de dificuldade dos itens. Além disso, caso buscássemos formar escores absolutos para as situações, teríamos mais dificuldade utilizando o escore por situações, já que o componente *s*, equivalente ao componente *i* ou faceta instrumental, foi particularmente influente na formação desse escore. Esse resultado reflete-se em quedas no coeficiente G absoluto para o escore por situações, como será discutido adiante.

E, por final, quanto aos coeficientes G obtidos, já se disse que os coeficientes G relativos são semelhantes em cálculo e significância aos coeficientes de alfa de Cronbach. Porém, de uma maneira geral, seus valores tendem a ser inferiores aos do coeficiente alfa, pois levam em consideração várias fontes de erro ao mesmo tempo, refletindo em maior precisão e rigor em sua medida.

Considerando-se o coeficiente G para o escore por itens, seus valores foram maiores para o grupo de estudantes (0,816 e 0,798) em relação ao grupo de especialistas (0,741 e 0,738). Cabe lembrar que esse cálculo considerou os 61 itens com melhor consistência interna e que a pequena diferença entre os coeficientes relativo e absoluto, principalmente para o grupo de especialistas reflete a relativa pequena participação do grau de dificuldade dos itens na formação dos escores. O nível de resíduo também afetou esse resultado.

Entretanto, a interpretação dos coeficientes G não é necessariamente a mesma do coeficiente de alfa (CRONBACH et al, 1997 e 2004). Segundo Shavelson e Webb (1991), a teoria clássica, por meio da interpretação do coeficiente de alfa, considera apenas medidas paralelas, assumindo que a média dos escores ao longo dos itens são iguais, assim como suas variâncias. Nessas circunstâncias, a variância devido ao item somente pode ser igual à zero. Não se leva em conta, com esse método de análise, a variabilidade de dificuldade entre os itens.

Assim, o coeficiente alfa reflete a capacidade de replicação dos resultados. O coeficiente G vai além, ao considerar outras fontes de erro, como a variabilidade dos itens, tornando capaz a generalização dos resultados para todos os itens possíveis. Por isso, a teoria clássica é uma teoria sobre diferenças individuais, ou seja, que se preocupa com as posições relativas dos examinandos. Decisões absolutas quanto ao escore somente podem ser tomadas com base na teoria moderna.

Os resultados em exames como o de vestibular, por exemplo, podem ser analisados pelo coeficiente de alfa. Nessas situações, o que importa são as posições relativas entre os vestibulandos. Porém, se é preciso decidir sobre o desempenho em determinado processo de ensino-aprendizagem, ou aprovar

determinado estudante, curso ou instituição, ou se busca a real medida de competência, são as posições absolutas que devem ser consideradas.

Portanto, a teoria G distingue entre interpretações absolutas e relativas de uma medida comportamental. Ao assumir dois valores de coeficientes, é capaz de colocar invariavelmente essa questão como um fato que afeta os resultados dos indivíduos e das instituições em avaliações educacionais. Considerando-se o escore por situações, os valores dos coeficientes G foram inferiores tanto para os especialistas como para os estudantes, em relação aos valores obtidos com o escore por itens. Esse resultado poderia ser mais bem apurado se considerarmos todos os examinandos conjuntamente, com o objetivo de compará-lo com o coeficiente de alfa de Cronbach para o escore por itens para todos os examinandos. Esses valores foram inferiores aos resultados obtidos para o alfa, confirmando a necessidade de se aumentar o número de itens para que se possa generalizar melhor os resultados.

Conclui-se que desenvolver um escore absoluto é ainda um desafio para o TCS. Por enquanto, ele permanece como um teste experimental, utilizado em pesquisas que buscam compreender o processo de aquisição de competência ao longo da formação continuada de médicos. Pode diferenciar grupos, ou até indivíduos, mas somente com relação as suas posições relativas

Entretanto, o nível de performance absoluto dos examinandos pode ser significativo e importante em muitas circunstâncias. Isto tem particular significado em muitos testes educacionais que pretendem certificar que as pessoas estão aptas para o exercício profissional. Por isso, novos meios para se aumentar a quantidade e a diversidade de situações e outros estudos buscando escolher somente itens ou situações com níveis semelhantes de dificuldade precisam ser elaborados. O escore por situações pode ser um novo caminho possível e viável para a investigação dessas fontes de evidência.

6.4 Os resultados em relação às pesquisas anteriores sobre o TCS

Este estudo foi baseado na metodologia descrita por Charlin et al. desde 1998, sobre o desenvolvimento de um teste de concordância de scripts. Particularmente, procurou-se um desenho de pesquisa que considerasse dois momentos diametralmente diferentes quanto à formação profissional médica, os estudantes de graduação e os especialistas, para a comparação quanto ao desempenho na resolução do teste.

Os dados encontrados, com isso, assemelham-se aos resultados de duas pesquisas anteriores sobre o TCS. Seus respectivos testes, um deles na área de ginecologia-obstetrícia (CHARLIN et al, 1998a) e o outro na área de radiologia (CHARLIN et al, 1998b), já haviam observado que a média dos escores obtidos pelos especialistas é superior à média dos escores obtidos pelos estudantes. Assim como nosso estudo, também essas duas pesquisas anteriores observaram que a variação na pontuação dos examinados tem relação inversa com o nível de experiência dos mesmos.

Entretanto, nossa pesquisa apresenta limitações quanto às suas conclusões em relação a estas duas primeiras publicações. Ambas formaram três grupos para a comparação: estudantes, residentes e especialistas. Com isso, conseguiram analisar aspectos sobre a eliminação do efeito intermediário de uma avaliação sobre os resultados do TCS. Nossos resultados, por envolverem somente dois grupos, em nada podem contribuir para a confirmação dessa hipótese.

Conjuntamente, estas duas pesquisas, além da nossa, foram as únicas a utilizar a teoria G para a análise de seus resultados. Uma pesquisa mais recente, realizada por Gagnon et al. (2008), ainda está em fase de publicação exclusivamente on-line. Porém, os resultados dessas duas análises mais anteriores parecem ser diferentes dos nossos. Ambas as pesquisas realizaram estudos G com universo com uma única faceta, no formato I/P, com o objetivo único de observar a interação entre os examinandos e os itens. Sendo assim, não

havia, nestas pesquisas, o objetivo de generalizar qualquer uma das características do instrumento.

Contrariamente ao que foi encontrado por nós, estas análises demonstraram uma relação inversa entre o resíduo e o nível de experiência dos grupos. Com isso, também o componente relacionado à dificuldade dos itens também apresentou comportamento diferente do nosso, apresentando uma relação direta com o nível de experiência dos grupos.

A explicação para tal diferença nas interpretações ainda não parece totalmente clara. É preciso se reconhecer que a teoria G é altamente complexa e há pouca experiência com ela entre os estatísticos (CROSSLEY et al., 2002). As múltiplas opções de desenhos de análise com essa teoria e as mudanças conceituais nos tipos de facetas consideradas podem justificar as diferenças encontradas entre o nosso estudo e os anteriormente publicados.

O estudo com questões de radiologia apresentava 49 questões e o de ginecologia obstetrícia, 50 questões. Estes são números inferiores ao número de itens de nosso instrumento, porém são próximos do número de 61 itens que foram selecionados por meio da análise de consistência interna empregada por nós. Nesse sentido, também, nossa análise psicométrica foi mais rigorosa em relação a estas pesquisas anteriores, pois considerou apenas itens com correlação parcial-total acima de +0,19. Os estudos anteriores consideraram qualquer item que apresentasse correlação que não fosse negativa.

Ao se comparar o coeficiente de alfa de Cronbach obtido na análise dos escores destas pesquisas anteriores, encontramos algumas semelhanças notáveis. O alfa de Cronbach para o teste inteiro de radiologia foi de 0,83 e no teste inteiro de ginecologia-obstetrícia, de 0,882. Esses valores foram próximos aos obtidos nessa pesquisa, que foi de 0,842. Porém, quando retirados os itens com baixa correlação parcial-total, nosso teste se aproximou em número de itens destes outros dois testes, mas com valor de alfa concordante com apenas um deles, de 0,882.

Estas comparações confirmam a importância da realização de testes piloto com um grande número de itens para a formação de um banco de itens com boas qualidades psicométricas para aplicações subseqüentes. Novos

desenhos de pesquisa testando diretamente esta hipótese poderiam confirmar a validade desta conclusão.

Com relação ao painel de referência, muitas questões podem ser levantadas. O número de integrantes deste painel é diferente nos diversos estudos publicados sobre o TCS. O estudo específico que buscou responder a esta questão, publicado por Gagnon el al., em 2005, concluiu que qualquer número acima de dez integrantes seria suficiente para garantir uma consistência interna adequada e uma boa correlação entre as amostras de painéis de referência simuladas pelo estudo e o painel de referência total, que, neste caso, foi formado por 38 especialistas. As análises provenientes desta pesquisa também demonstraram que existiam ganhos apenas marginais de consistência interna por meio do aumento do painel hipotético, com números acima de 20 integrantes.

Com uma atenção maior sobre os dados desta pesquisa, é preciso observar, entretanto, que o coeficiente de alfa de Cronbach entre os especialistas foi de, no máximo, 0,76, para o total de 38 integrantes. Como se disse anteriormente, o valor de alfa pode não ser o mais adequado para avaliar a consistência interna de um instrumento de avaliação de competências. Além disso, este valor de alfa, obtido para o grupo inteiro de especialistas, pode ser adequado para alguns objetivos de avaliação, mas não para outros.

Pode-se levantar a hipótese, também, de que o número de especialistas do painel de referência deva ser diferente, de acordo com os objetivos de avaliação e com grupos a serem avaliados. Outros estudos com desenhos comparativos de diferentes propostas em número e qualidade sobre esse painel irão discutir mais sobre essa questão.

Considerando-se o estudo G realizado com o escore por itens, nosso coeficiente G relativo para o painel de referência foi de 0,74, o mesmo valor do coeficiente alfa do painel testado com 20 integrantes, por Gagnon et al., em 2005. Este autor considerou esse número de vinte especialistas como possivelmente o mais adequado em sua pesquisa, já que o aumento desse número não alterou significativamente a consistência interna de seus resultados.

Portanto, nossos dados com vinte e um especialistas replicam aqueles obtidos para o painel de referência nesse estudo anterior. Entretanto, o número de itens considerados para o valor de alfa naquele estudo foi 73, superior ao nosso, que foi de 61. Este fato adiciona valor ao nosso instrumento em comparação àquele construído por esse autor, pois precisamos de menos itens para obter a mesma capacidade de replicação em nossos resultados.

Quanto ao nível de experiência exigido para a composição do painel de referência, já fora comentado que este dependeria dos objetivos da avaliação. Para a comparação com um grupo de estudantes de graduação, simplesmente ser especialista na área de avaliação seria suficiente para a formação do escore agregado.

O estudo com itens sobre ginecologia-obstetrícia considerou os especialistas da área em atividade, sem especificações. O estudo sobre radiologia apresentou como critério de inclusão o mínimo de três anos de experiência na área. Ambos convidaram especialistas de apenas um centro médico. Charlin et al, em 2002, formaram um grupo de especialistas de uma única faculdade, sem outros critérios adicionais. Gagnon et al., em 2005, formaram um painel de referência com médicos de família franceses de um único centro universitário, sem especificar outras características deste grupo. Gagnon et al., em 2006, convidaram trinta geriatras com dedicação exclusiva e atuação prática nessa área, de dois centros universitários, para compor o painel.

Portanto, nosso estudo traz novidades sobre a composição do painel de referência. Seu caráter multi-cêntrico, com geriatras de oito instituições, busca uma diversidade de formações profissionais e opiniões que não havia sido procurada até então. Não há, por hora, nenhuma padronização dos critérios de inclusão para a formação do painel de referência. Como discutido anteriormente, eles irão depender dos objetivos da pesquisa e dos grupos a serem comparados.

Este estudo também utilizou o critério de no mínimo cinco anos de experiência na área de geriatria para a formação desse grupo. Provavelmente esse tempo não é suficiente para a formação de expertise nessa área (CROSKERRY, 2005). Os guias de recomendação sobre o TCS (FOURNIER et al., 2008), sugerem que os TCS com objetivos gerais ou com painéis que possam

ser utilizados em diversas comparações constituam grupos de especialistas com profissionais que sem dúvida representem o corpo de jurados de uma avaliação nesta área de atuação.

Nesse sentido, nosso painel de referência foi formado por profissionais com idades e tempos de formação diferentes. Porém essa característica não invalida a representatividade deste painel diante dos objetivos propostos. Todavia, painéis mais restritos, com critérios de inclusão mais seletivos, buscando especificamente os formadores de opinião dentro da área de atuação que se quer avaliar, seriam necessários para exames mais rigorosos. O risco desta rigorosidade seria um excesso de seletividade a ponto de comprometer a busca por um número adequado de integrantes no painel, o que poderia levar perdas de outras qualidades no instrumento.

Esta pesquisa foi inédita em seu desenho ao utilizar-se de um teste piloto com especialistas sobre o nível de concordância de respostas entre os mesmos. Com isso, não esperávamos reduzir todos os problemas que poderiam surgir na primeira aplicação de um teste construído experimentalmente. Buscávamos, apenas, eliminar ou corrigir os itens que, teoricamente, poderiam gerar pouca incerteza sobre o raciocínio clínico dos examinandos.

A idéia de se analisar a concordância entre as respostas dos especialistas surgiu de discussões teóricas sobre o raciocínio clínico em contextos de incerteza e sobre o teste de concordância scripts que indicavam que as respostas muito discordantes ou muito concordantes entre si poderiam indicar os itens com pouca capacidade para gerar incertezas. Mantendo-se esses itens no teste, provavelmente não se conseguiria diferenciar os grupos conforme o nível de experiência.

Entretanto, essa conclusão não foi encontrada por Charlin et al., em 2006, quando se analisou o efeito da variabilidade de respostas dentro do painel de referência sobre a capacidade do teste em diferenciar os grupos de acordo com seus níveis de experiência. Embora se tenha comprovado que a variabilidade de repostas deste painel foi um componente-chave para o poder de discriminação deste TCS, foram os itens com maior variabilidade que apresentaram essa maior capacidade discriminante.

Esta descoberta é contrária, portanto, à hipótese que fora levantada anteriormente, ou seja, de que uma grande variabilidade de repostas poderia significar um ruído nas mesmas e não contribuiria para a avaliação proposta. Entretanto, conforme discutido nesta publicação anterior, provavelmente o teste elaborado por Charlin et al. naquela ocasião não continha itens com grande dispersão de respostas ao longo da escala Likert, o que foi comprovado também por uma análise posterior de seus achados. Ou seja, o teste já não continha, desde seu início, itens com grande variabilidade de respostas entre os especialistas, sendo diferente daquele que foi desenvolvido inicialmente por nós.

É preciso reconhecer que o grupo de pesquisadores liderados por Charlin já tem experiência de pelo menos 10 anos com o TCS. Esse teste era inédito para nós.

Estabelecendo um paralelo com a metodologia aqui aplicada, o conceito de variabilidade nas respostas entre os especialistas se aproxima daquilo que denominamos de concordância entre as respostas dos especialistas, a partir de uma metodologia de análise elaborada por Cicchetti et al., em 1997. Sendo assim, uma pequena variabilidade seria compatível com uma alta concordância em nossa análise do teste piloto, bem como uma grande variabilidade poderia ser equivalente a uma concordância baixa em nosso estudo.

Porém, conforme apontado anteriormente, os itens com concordância de respostas muito baixa em nosso estudo apresentavam uma dispersão de repostas ao longo da escala Likert muito alta e, conforme confirmado por Charlin et al, em 2006, esse fato representa erros de construção no teste, o que provocaria problemas de interpretação pelos especialistas durante a resolução. Sendo assim, as respostas com grande variabilidade neste estudo publicado anteriormente, de fato, poderiam representar as respostas com concordância justa ou boa em nossa metodologia. Essa pesquisa anterior, portanto, confirma que são esses itens, com respostas com concordância justa ou boa entre os especialistas, que possuem o maior poder de discriminação, de acordo com o nível de experiência profissional.

Esta coincidência no critério para a inclusão de itens na aplicação final do teste reforça nossa estratégia de retirar ou reformular os itens com concordâncias de respostas baixas ou muito altas. Entretanto, os métodos estatísticos para a análise de variabilidade entre as respostas não foram os mesmos empregados para a análise de concordância que fizemos. Nesse sentido, a semelhança dos dois métodos é, antes de tudo, apenas conceitual.

Nosso estudo também é restrito por ter utilizado apenas um pré-teste com cinco geriatras com poucos anos de experiência. O estudo anterior havia elaborado um painel de referência com trinta especialistas em atividade para essa intenção semelhante, embora não se tenha descrito seus níveis de experiência.

Outra análise feita por Charlin et al., 2006, nesta pesquisa demonstrou que os itens com baixa variabilidade de respostas entre os especialistas têm respostas mais freqüentes nos extremos da escala Likert (equivalente às respostas 1 e 5 de nossa pesquisa). Esta descoberta vai de encontro à hipótese de que os extremos da escala Likert são respondidos com menor freqüência pelos especialistas, que tendem a escolher mais as opções de respostas centrais da escala.

Com isso, confirmou-se ainda mais que os itens com respostas com menor variabilidade são semelhantes a testes de múltipla escolha tradicionais, pois geram pouca incerteza, mas que, por também apresentarem como resultado as respostas mais freqüentes nos extremos da escala, podem ser úteis, pelo menos para equilibrar o número de respostas ao longo das opções da escala Likert no painel de referência. Esta hipótese já havia sido levantada por nós no tópico anterior de discussão.

Por final, é preciso discorrer sobre a pesquisa publicada por Bland et al., em 2005. Trata-se da primeira pesquisa publicada sobre o TCS por um grupo não supervisionado ou influenciado diretamente pelos pesquisadores que iniciaram os estudos sobre este instrumento de avaliação na Universidade de

Montreal, no Canadá. Esse autor coloca a hipótese de que as avaliações médicas precisam se basear em uma única resposta certa. Sua pesquisa, portanto, busca invalidar as características únicas do método de escore agregado proposto para o TCS.

Para isso, Bland et al. testaram a aplicação de TCS na área de nefrologia em 85 examinandos. Além do tradicional escore agregado a partir das respostas de 8 especialistas, o autor analisou outros 4 escores alternativos. Um método eliminou a escala Likert com 5 opções de resposta e considerou apenas três destas opções. Outros três métodos eliminaram a escala Likert e empregaram um escore com uma única resposta certa.

Entre os resultados encontrados, duas entre as quatro alternativas para o escore tradicional demonstraram performances semelhantes entre os examinandos àquelas obtidas no método de escore agregado para uma escala Likert com 5 pontos. Nenhum dos métodos apresentou o efeito intermediário nas respostas. Sendo assim, o autor concluiu que a opção de escore com uma única resposta certa para uma escala com três opções de resposta pode ser válida para o TCS.

Conforme nossos achados e as conclusões baseadas em publicações anteriores sobre o método de escore agregado e o sobre o TCS, não estamos certos quanto a uma opção de escore com uma única resposta certa para a avaliação do raciocínio clínico em situações de incerteza. Em primeiro lugar, é preciso se reconhecer que uma avaliação e seu método de pontuação e ranqueamento pode ter diversos objetivos explícitos e implícitos.

Nenhuma avaliação, sob o ponto de vista educacional, é exclusivamente classificatória ou somativa. Sempre há a possibilidade do papel formativo da aplicação de um exame. O programa de avaliações de uma instituição guia substancialmente a aprendizagem dos estudantes. Com o TCS não há razão para julgar que esse sentido seja diferente. Sendo assim, seu método de escore implicitamente pode trazer diferentes significados aos examinandos.

Portanto, as características psicométricas não são os únicos objetivos de um bom procedimento de escore, diferentemente do que foi defendido por Bland. O método de escore agregado não só avalia, mas também ensina. Ao demonstrar que cada resposta tem um valor diferente e que cada profissional desenvolve um raciocínio único, o TCS com o método de escore agregado representa um avanço nos métodos objetivos e numéricos de avaliação educacional, discordando da visão ainda predominante em medicina de que o processo de tomada de decisões é simplificadamente dividido em opções certas ou erradas.

6.5 Considerações sobre a importância de uma avaliação do raciocínio clinico em contextos de incerteza na formação médica atual

O teste de concordância de script pode ser mais uma opção para a avaliação da formação médica atual. Segundo as Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina (MINISTÉRIO DA EDUCAÇÃO, 2001), uma das principais competências para o exercício da medicina é a capacidade para a tomada de decisões. Este elemento pode ser detectado e diferenciou os examinandos por meio do instrumento que foi elaborado.

Reconhece-se também que o volume do conhecimento médico, exames diagnósticos e opções terapêuticas esta cada vez maior (BORNSTEIN e EMIER, 2001). Assim, alguns desafios educacionais estão lançados. Como integrar as evidências científicas com as decisões práticas? Como ensinar o processo de raciocínio clínico em contextos de incerteza em uma era em que se acredita que a dúvida pode ser sanada simplesmente com o advento da medicina baseada em evidências?

É preciso lembrar que a maioria dos erros médicos se deve a equívocos de raciocínio. Não são, portanto, conseqüência da falta de conhecimento propriamente dito (GANDHI et al, 2006). Ainda, a freqüência de erros diagnósticos parece não ter se modificado ao longo das últimas décadas. Um estudo demonstrou que o número de erros diagnósticos não se modificou entre 1960 e 1980 em um hospital universitário norte-americano, apesar da introdução de novas tecnologias como a tomografia computadorizada e outros exames subsidiários (GOLDMAN et al. 1983). De fato, a freqüência de diagnósticos equivocados pode subir pelo excesso de confiança em novos procedimentos.

As pesquisas sobre o uso de computadores para a melhora dos diagnósticos também têm demonstrado pequenos benefícios, mais importantes em estudantes do que em residentes e médicos mais experientes. Em algumas circunstâncias, a consulta ao computador foi prejudicial e causou falha diagnóstica (FRIEDMAN et al. 1999).

O raciocínio a partir do teorema de Bayes, no qual se baseiam freqüentemente os programas de informática, pode ser limitado pela falta de confiabilidade dos dados para caracterizar o paciente como único. Mesmo quando os dados estão disponíveis, o complexo processo de interpretar súbitas interações no processo de saúde-doença de um paciente pode ser cercado de dúvidas. Os testes mais exuberantes e as combinações exaustivas de dados da literatura poderão gerar uma abundância de relatos falso-positivos e conclusões conflitantes, que podem aumentar, ao invés de diminuir, a carga cognitiva sobre os médicos (REDELMEIER, 2005).

Algumas das possibilidades para se aprimorar o processo de raciocínio clínico poderiam ser o auto-conhecimento das fontes de equívocos desse raciocínio ou os treinamentos práticos (BORNSTEIN e EMIER, 2006). A estratégia conhecida como metacognição enfatiza a necessidade de se refletir e discutir continuamente sobre as decisões tomadas na prática diária. Mais ainda, coloca que é preciso persistência para se corrigir as possíveis falhas sistemáticas do raciocínio.

Porém, alguns teóricos do processo de tomada de decisão em medicina têm se fixado somente em teorias normativas, quase mecanicistas, sobre o raciocínio. Esta ênfase tem pouca aplicação prática e em cenários reais (CROSKERRY, 2003).

Aprender a partir dos erros é difícil para o médico, mas necessário. Os educadores precisam reconhecer o impacto que os erros de raciocínio tem no processo de ensino-aprendizagem da medicina. Poderiam ainda refutar o caráter inevitável destes erros cognitivos e desmistificar o pessimismo que cerca as propostas que podem amenizá-los

A prática médica não é fácil por sua inerente e inevitável incerteza. Uma observação recorrente é que as pessoas cometem erros quando encontram problemas complexos. Porém, os erros não acontecem ao acaso. Por meio da aprendizagem, a habilidade diagnóstica de um astuto clínico pode ainda ser muito melhor do que o desempenho de um programa de computador com um conhecimento enciclopédico em medicina (BERNER et al, 1994).

Ambos, médico e paciente devem ser parceiros no processo de tomada de decisões diagnósticas e terapêuticas. Porém, o longo e árduo processo de treinamento médico é socialmente construído para desenvolver sobre o médico um forte senso de confiança em face da incerteza. Os estudantes de medicina precocemente aprendem que é desvantajoso duvidar demais. Pior poderia ser a demonstração dessas dúvidas aos pacientes (WEED, 1999).

Um estudo sobre o uso de recursos na área de saúde demonstrou que cada desvio-padrão no aumento da ansiedade gerada pela incerteza por parte dos médicos correspondia a um aumento significativo de 17% na média de custos em saúde (ALLISON et al., 1998). A solução para a questão das incertezas envolvidas em decisões sobre a saúde e a vida dos pacientes não envolve somente conhecimentos e cognição. As emoções e os afetos parecem exercer uma influência significativa nesta questão (CROSKERRY, 2005).

É preciso se reconhecer em educação médica que a maioria dos profissionais não utiliza guias de recomendação, conhecidos como “guidelines”, em sua prática diária (CABANA et al., 1999). A grande parte dos ensaios clínicos em que se baseiam esses documentos que buscam nortear a tomada de decisões exclui os pacientes muito idosos ou com comorbidades, enquanto muitos pacientes na prática clínica apresentam outras características que podem determinar as decisões diagnósticas e terapêuticas. Esses “guidelines” podem ser aplicados a pacientes similares àqueles estudados em grandes pesquisas clínicas. Porém, os sub-grupos ou outros pacientes que se desviam das características estudadas, por se aproximarem da complexidade envolvida em fenômenos ou doenças distantes destes protocolos, podem não se beneficiar de organogramas previamente determinados (GARFIELD e GARFIELD, 2000).

Portanto, o julgamento clínico, por meio de processos ainda a serem mais compreendidos, como o raciocínio clínico em contextos de incerteza, pode estar cercado pela ciência, mas também é envolvido pela criatividade e pela arte (FELSTEIN, 1994). Por isso, parece improvável tornar completamente explícitos todos os aspectos da competência profissional para esse raciocínio (TONELLI, 1998). O desenvolvimento do expertise em medicina está em parte associado com uma crescente habilidade em se encontrar novas soluções diante de ciladas do passado. Mas é o futuro que determinará a validade destes recursos.

Para não se cometer os mesmos equívocos do passado, é preciso se avaliar as atitudes do presente. O tempo e a capacidade cognitiva impõem pressões e limites à precisão de nossas decisões. Parece ser impossível que um teste escrito alcance toda a complexidade dessa realidade. Nesse sentido, o TCS apenas se aproxima desses fenômenos. O raciocínio clínico mais adequado seria aquele capaz de reduzir ao mínimo a probabilidade de erros médicos. Importante seria também se esse raciocínio reduzisse os custos de um sistema de saúde cada vez mais caro.

Assim os sistemas de avaliação precisam usufruir de diversas opções de testes baseados em fundamentos cognitivos e educacionais bem elaborados, na busca pela compreensão e pelo aprimoramento deste complexo elemento, ou seja, a competência médica. Avaliar o raciocínio clínico em contextos de incerteza pode ainda contribuir para o direcionamento de estratégias para o processo de ensino-aprendizagem de tal competência. Novos estudos em avaliação educacional, buscando delimitar o impacto dessas novas metodologias de avaliação em longo prazo, portanto, serão bem vindos.

6.6 Perspectivas para novos desenhos de pesquisa sobre o TCS

Muito a respeito do TCS deve ainda ser pesquisado na busca de elementos que reconheçam ainda mais a sua validade, a confiança em seus resultados, a viabilidade de sua aplicação para diversos objetivos de avaliação e a possibilidade de acompanhamento progressivo da formação continuada da competência para o raciocínio clínico.

Uma maior diversidade de estudos sobre o painel de referência parece ainda ser necessária, não só a respeito do número de seus integrantes, como também das características profissionais e demográficas dos mesmos. A formação de um banco de itens para aplicações subseqüentes que garantam melhores qualidades psicométricas ao teste também precisa ser mais analisada.

Outras pesquisas importantes sobre esta metodologia poderiam comparar o desempenho neste teste em relação a outras diversas metodologias de avaliação, buscando mais dados a respeito da validade de constructo do TCS. Comparações de desempenho de estudantes em diferentes TCS, em diferentes áreas de atuação, poderia esclarecer se a competência para o raciocínio clínico se

mantém de alguma maneira apesar de se reconhecer que em grande parte este raciocínio é caso-específico. Os dados de pesquisas como essas poderiam confirmar algumas hipóteses de validade do instrumento.

As evidências quanto ao valor preditivo do TCS precisam ainda ser estudadas com mais cautela. Comparações de desempenho para o raciocínio clínico por meio dos resultados em avaliações, ou pelo reconhecimento por colegas de trabalho ou por pacientes, ou até mesmo por meios de análises de precisão quanto à tomada de decisões diagnósticas ou terapêuticas, poderiam sugerir conclusões a respeito da validade preditiva do resultado em TCS sobre o desempenho em atividades da prática profissional diária.

Por fim, alguns estudos longitudinais sobre a evolução das características do raciocínio clinico ao longo da vida profissional poderiam contribuir com conclusões sobre o que se espera do desempenho de estudantes e profissionais em determinados momentos da formação médica. Os conceitos sobre essa expectativa trariam consigo o desafio de se determinar escores absolutos para o TCS, necessários para diversos objetivos de avaliação educacional, sem os quais pouco se pode concluir sobre a real diferença de desempenho entre os examinandos ou entre os diferentes momentos da formação profissional.

7. CONCLUSÕES

O processo de elaboração, aplicação e análise desta proposta de avaliação do raciocínio clínico em contextos de incerteza permite uma série de conclusões.

A construção deste instrumento pareceu relativamente simples, quando comparada com a de outros testes escritos tradicionais. Porém, reconhece-se que pessoas com experiências prévias com essa construção podem conseguir a elaboração de situações e itens mais adequados.

A aplicação desta proposta em dois momentos diferentes da formação profissional foi capaz de originar resultados, análises e discussões importantes para o aprofundamento da compreensão sobre o TCS.

Assim, a proposta de análise a partir de um escore formado pela somatória dos escores por itens de cada situação, chamado de escore por situações, é inédita em pesquisas sobre o TCS. Nosso estudo foi o primeiro a levantar essa questão e a tentar medir sua influência sobre os resultados obtidos. Novos estudos direcionados para essa hipótese são necessários para se dimensionar o verdadeiro valor deste questionamento.

Inicialmente, apresentamos à educação médica brasileira uma metodologia de avaliação já em parte validada e pesquisada em outros países. Esperamos ter ido além, contribuindo com o complexo processo de validação deste instrumento, construindo-o, aplicando-o e analisando-o, a partir dos princípios da teoria de scripts.

Portanto, aprofundamos a discussão sobre o potencial e as limitações do TCS. Com isso, também aprendemos muito sobre as dificuldades e o rigor que devem guiar a elaboração de uma proposta de avaliação, bem como a análise de seus resultados.

As propostas de avaliação que buscam captar os sentidos da educação e da prática médica atual precisam se reconhecer diante dos processos de ensino-aprendizagem voltados para o ensino de competências. Encontramos este desafio diante de nós. Esperamos tê-lo superado de alguma maneira.

8. REFERÊNCIAS

ALLISON. et al. The Association of physician attitudes about uncertainty and risk taking with resource use in a medicare HMO. **Med. Decis. Making.** 1998; 18: p. 320-329.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association, 1999.

BARROWS HS; BENNETT K. The diagnostic skill of the neurologist. **Arch. of Neurol.** 1972, 27: p. 273-277.

BERNARDO & SMITH. (1993). **Bayesian theory,** John Wiley & Sons.

BERNER ES. et al. Performance of four computer-based diagnostic systems. **N. Engl. J. Med.** 1994; 330: p. 1792-1796.

BLAND AC; KREITER CD; GORDON JA. The Psychometric Properties of Five Scoring Methods Applied to the Script Concordance Test. **Acad. Med.** 2005, 80 (4): p. 395-399.

BORNSTEIN BH; EMLER C. Rationality in medical decision making: a review of the literarure on doctors´ decision-making biases. **J. Eval. Clin. Pract.** 2001; 7 (2): p. 97-107.

BRAILOVSKY C. et al. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. **Med. Educ.** 2001; 35: p. 430-436.

BISQUERRA R; SARRIERA JC; MARTINEZ F (orgs). **Introdução à Estatística: enfoque informático com o pacote estatístico SPSS.** Porto Alegre: Artmed, 2004.

CABANA MD. et al. Why don´t physicians follow clinical practice guidelines? A framework for improvement. **JAMA.** 1999; 282: p. 1458-1465.

CHARLIN B. et al. Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. **Teach Learn Med.** 2002, 14 (3): p. 150-156.

CHARLIN B. et al. Script questionnaires: their use for assessment of diagnostic knowledge in radiology. **Med. Teach.** 1998b; 20 (6): p. 567-571.

CHARLIN B; TARDIF J; BOSHUIZEN HPA. Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. **Acad. Med.** 2000; 75: p. 182-190.

CHARLIN B. et al.. The diagnosis Script questionnaire: a new tool to assess a specific dimension of clinical competence. **Adv. Healt. Sci. Educ.** 1998a; 3: p. 51-58.

CHARLIN B; VLEUTEN CVD. Standardized Assessment of Reasoning in Contexts of Uncertainty: The Script Concordance Approach. **Eval. Heal Prof.** 2004; 27 (3): p. 304-319.

CHARLIN B. et al. The Script Concordance Test: A Toll to Assess the Reflective Clinician. **Teach Learn Med.** 2000; 12 (4): p. 189-195.

CHOON-HUAT KOH G; KHOO HE; WONG ML; KOH D. The effects of problem-based learning during medical school on physician competency: a systematic rewiew. **CMAJ**. 2008; 178 (1): p. 34-41.

CICCHETTI DV; SHOWALTER D; ROSENHECK R. A new method for assessing interexaminer agreement when multiple ratings are made on a single subject: applications to the assessment of neuropsychiatric symptomatology. **Psych. Res.**1997; 72: p. 51-63.

CLAUSER BE. Recurrent issues and recent advances in scoring performance assessments. **Appl. Psych. Measurement.** 2000, 24 (4): p. 310-324.

CROCKER LM; ALGINA J. **Introduction to Classical and Modern Test Theory.** New York: Holt, Rinehart and Winston, 1986.

CRONBACH; LJ. Coefficient alpha and the internal structure of tests. **Psychometrika** 1951; 16: p. 297-335.

CRONBACH LJ; MEEHL PE, Construct Validity in Psychological **Tests. Psychological Bulletin.** 1955, 52: p. 281-302.

CRONBACH LJ. et al. Generalizability analysis for performance assessment of student achievement or school effectiveness. **Educational and Psychological Measurement**. 1997, 57 (3): p. 373-399.

CRONBACH LJ; SHAVELSON RJ. My current thoughts on coefficient alpha and successor procedures. **Educational and Psychological Measurement**. 2004, 64 (3): 391-418.

CRONBACH LJ. et al. **The dependability of behavioral measurements: Theory of generalizability of scores and profiles**. New York: John Wiley. 1972.

CROSKERRY P. **Diagnostic Failure: A cognitive and affective approach. In Advances in patient safety: from research to implementation.** Vol. 2, Concepts and Methodology. AHRQ Publications. Rockville MD: Agency for Helthcare Research and Quality, 2005.

CROSKERRY P. The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize them. **Acad. Med.** 2003; 78: p. 775-780.

CROSSLEY J; DAVIES H; HUMPHRIS G; JOLLY B. Generalizability: a key to unlock professional assessment. **Med. Educ.** 2002; 36: p. 972-978.

DANS PE. Looking for answers in all the wrong places. **Ann. Int. Med.** 1993, 119 (8): p. 855-857.

DOWNING SM. Reliability: on the reproducibility of assessment data. **Med. Educ.** 2004; 38: 1006-1012.

DOWNING SM. Validity: on the meaningful interpretation of assessment data. **Med. Educ.** 2003; 37: p. 830-837.

ELSTEIN AS. et al. Comparison of physicians decisions regarding estrogen replacement therapy for menopausal women and decisions derived from a decision analytic model. **JAMA.** 1986; 80: p. 246-258.

ELSTEIN AS; SHULMAN LS; SPRAFKA SA. **Medical Problem solving: An analysis of clinical reasoning.** Cambridge, MA: Harvard University Press, 1978.

EPSTEIN RM; HUNDERT EM. Defining and Assessing Professional Competence. **JAMA.** 2002; 287 (2): p. 226-235.

ERICSSON KA. An expert-perfomance perspective of research on medical expertise: the study of clinical performance. **Med. Educ.** 2007, 41: p. 1124-1130.

FELSTEIN AR. “Clinical Judgement” revisited: the distraction of quantitative models. **Ann. Intern. Med.** 1994; 120: p. 799-805.

FOUNIER JP; DEMEESTER A; CHARLIN B. Script Concordance Test: Guidelines for Construction. **BMC Medical Informatics and Decision Making.** 2008, 8: p. 18

FREUDNHEIM E. **Chronic Care in America: A 21st Century Challenge.** Princenton. New Jersey: Rober + Wood Foundation, 1996.

FRIEDMAN CP. Enhancement of clinicians diagnostic reasoning by computer-based consultation: A multiple study of 2 systems. **JAMA.** 1999; 282: p. 1851-1856.

GAGNON R. et al. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? **Med Educ.** 2005; 39: p. 284-291.

GAGNON R. et al. Script Concordance Testing: more cases or more questions? **Adv. Health Sci Educ Theory Pract.** 2008 May 15 (online publication). No prelo.

GAGNON R. et al. The Cognitive Validity of the Script Concordance Test: A Processing Time Study. **Teach Learn Med.** 2006; 18 (1): p. 22-27.

GANDHI TK. et al. Missed and delayed diagnosis in the ambulatory setting: a study of closed malpractice claims. **Ann. Int. Med.** 2006; 145: p.488-496.

GARFIELD FB; GARFIELD JM. Clinical Judgment and Clinical Practice Guidelines. **Int. J. Tech. Assess. H. Care.** 2000; 16 (4): p. 1050-1060.

GILL TM, Geriatric medicine: it's more than caring for old people. **The American Journal of Medicine.** 2002; 113 (1): p. 85-90.

GILLIES J, SHEIHAN M. Practical reasoning and decision making – Hippocrates´ problem. **The British Journal of General Practice.** 2002; June: p. 518-519.

GOLDMAN L. et al. The value of the autopsy in three medical eras. **NEJM.** 1983; 308: p.1000-1005.

GRANT J, MARSDEN P. The structure of memorized knowledge in students and clinicians: an explanation for diagnostic expertise. **Med. Educ. 1987**; 21 (2): p. 92-98.

GRANT J, MARSDEN P. Primary knowledge, medical education and consultant expertise. **Med. Educ.** 1988; 22: p. 173-179.

HASSEBROCK F. et al. When less is more – representation and selective memory in expert problem solving. **Am J Psych.** 1993; 106: p. 155-189.

HIPÓCRATES. **Aforismos, Antologia.** São Paulo: Martin Claret, 2007.

HOBUS PP. et al. Contextual factorsin the activation of first diagnostic hypotheses: expert-novices differences. **Med. Edu.** 1987; 21: p. 471-476.

HODGES . at al. OSCE Checklists do not capture increasing levels of expertise. **Acad Med.** 1999; 74: p. 1129-1134.

IBGE. **Pesquisa Nacional por Amostra de Domicílios (PNDA).** Rio de Janeiro; 1999.

JAYAWICKRAMARAJAH PT. Problems for problem based learning - a comparative study of documents. **Med. Edu.** 1996; 30: p. 272 – 282.

KANE MT; CROOKS TJ; COHEN AS. Validating measures of performance. **Educ Measure Issues Prac.** 1999; 18: p. 5-17.

LAMPERT JB. **Na transição paradigmática da educação médica: o que o paradigma da integralidade atende que o paradigma flexneriano deixou de lado.** Boletim da ABEM. 2003; 31: p. 4-5.

LIKERT, R. A Technique for the Measurement of Attitudes. **Arch. Psy.** 1932, Vol. 140, junho.

MCGUIRE C. Medical Problem Solving: a critique of the literature. **J. Med. Edu.** 1985; 60: p. 587-585.

METERRISIAN SH. A novel method of assessing clinical reasoning in surgical residents. **Surg. Innov.** 2006; 13 (2): p. 115-119.

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.** Brasília, DF; 2001.
Disponível em: <http://portal.mec.gov.br/cne/arquivos/pdf/CES04.pdf.>. Acesso em 25 fev. 2006.

MORIN E; ALMEIDA MC; CARVALHO EA (orgs). **Educação e Complexidade: os sete saberes e outros ensaios.** 3ª ed. São Paulo: Cortez, 2005.

NEAME . et al. Problem solving in undergraduate medical students. **Med. Decis. Mak.** 1985; 5 (3): p. 311-325.

NEUFELD VR. Et al. Clinical problem solving by medical students: a cross sectional and longitudinal analysis. **Med. Educ.** 1081; 15: p. 315-322.

NORMAN GR. Objective measurement of clinical performance. **Med. Educ.** 1985; 19: p. 43-47.

NORMAN GR. Et al. Measuring physician's performance by standardized patients. **J. Med. Edu.** 1985; 60: p. 925-934.

NORMAN G; SWASON DB; CASE SM. Conceptual and methodological issues in studies comparing assessment formats. **Teach. Learn. Med.** 1996; 8: p. 208-216.

REDELMEIER DA. The cognitive psychology of missed diagnoses. **Ann. Inter. Med.** 2005; 142: p.115-120.

RICHARDSON RJ. **Métodos quantitativos e qualitativos.** São Paulo: Atlas, 1989.

RIDDERIKHOFF J. Medical problem solving: an exploration of strategies. **Med. Edu.** 1993; 25: p. 196-207.

ROUND A. Introduction to clinical reasoning. **J. Eval. Clin. Pract.** 2001; 7 (2): p. 109-117.

SCHMIDT HG; NORMAN GR; BOSHUIZEN HP. A cognitive perspective on medical expertise: theory an implication. **Acad. Med.** 1990; 65 (10): p. 611-621.

SHAVELSON RJ; WEBB NM. **Generalizabillity theory: a primer.** Thousand Oaks, California, USA: SAGE Publications Ltd, 1991.

STEWART AL. et al. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. **JAMA.** 1989; 262: p. 907-913.

TONELLI MR. The philosophical limits of evidence-based medicine. **Acad. Med.** 1998; 73: p. 1234-1240.

VERNON DT; BLAKE RL. Does problem based learning work? A meta-analysis of evaluative research. **Acad. Med.** 1993; 68: p. 550-563.

VON NEUMAN J; MORGENSTERN O. **Theory of Games and Economic Behavior.** Princeton: Princenton University Press, 1947.

WEED LL. Clinical Judgment Revisited. **Meth. Inform. Med.** 1999; 38: p. 279-286.

WOLF FM. et al. Differential diagnosis and the competing hypothesis heuristic: a practical approach to judgement under uncertainty and Bayesian probability. **JAMA.** 1985; 253: p. 1858-1862.

9. APÊNDICES

Apêndice 1 – Estrutura de item para avaliação de raciocínio para diagnóstico

Se você está pensando em	E você encontra	A hipótese se torna
(uma hipótese diagnóstica)	*(um novo dado clínico, estudo de imagem ou resultado laboratorial)*	1 2 3 4 5

Legenda

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

Apêndice 2 – Estrutura de item para avaliação do raciocínio para investigação

Se você está considerando solicitar	E você encontra	A investigação se torna
(um teste diagnóstico)	*(um novo dado clínico, estudo de imagem ou resultado laboratorial)*	1 2 3 4 5

Legenda

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

Apêndice 3 – Estrutura de item para avaliação para o raciocínio para o tratamento

Se você está considerando prescrever	E você encontra	A prescrição se torna
(uma opção terapêutica)	*(um novo dado clínico, estudo de imagem ou resultado laboratorial)*	1 2 3 4 5

Legenda

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

Apêndice 4 - Exemplo de caso clínico seguido de um grupo de itens sobre diagnóstico em geriatria (extraído de Gagnon et al, 2006, p. 24, com tradução livre)

Leia cuidadosamente o caso clínico e a escala seguinte:

Um idoso de 82 anos de idade está em avaliação na sala de emergência há 18 horas. Ele apresenta febre. Ele estava apático até há 2 horas atrás. Desde então, ele se apresenta agitado, gritando que assaltantes invadiram a sala de emergência. Ele retirou seu acesso venoso. Sua prescrição indica que ele está recebendo antibióticos endovenosos para pneumonia e medicamentos para insuficiência cardíaca congestiva. Seu médico ambulatorial refere que sua pontuação no último mini-exame do estado mental realizado há 3 meses atrás foi de 23 pontos em 30. Ele sofre de prostatismo e não urina há 12 horas. Ele sempre foi constipado.

Se você está pensando em	E o paciente ou acompanhante relatam, ou você encontra no exame clínico	A hipótese se torna
Delirium	***Alucinações visuais***	**1 2 3 4 5**

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

Apêndice 5 – Questionário sobre as características dos especialistas e suas opiniões sobre o teste

Numeração na Pesquisa: _____

Nome: _____

idade: _____ Sexo: fem ( )1 masc ( ) 2

Ano de formatura: _____ Ano de término da especialização e/ou obtenção do título de especialista em geriatria: _____

Realiza atividades de ensino nessa área: sim ( ) 1 não ( ) 2

Se sim, qual:

_____

_____

Onde exerce atividade clínica atualmente? (pode-se assinalar mais que uma)

Consultório/ambulatório ( )1 Hospital ( ) 2 Assistência Domiciliar ( ) 3

Instituição de Longa Permanência ( ) 4

Outra – Qual? _____

Quais foram as dificuldades encontradas na resolução do teste?

_____

_____

_____

_____

_____

Ele trouxe situações duvidosas na prática médica?

( ) sim ( ) não

Ele representou um desafio para você? Por quê?

_____

_____

_____

_____

_____

Apêndice 6 – Termos de Consentimento Livre e Esclarecido

Termo de Consentimento – FORMATO PARA O GRUPO DE CONSTRUÇÃO DO TESTE

Título da Pesquisa

A AVALIAÇAO DO RACIOCÍNIO CLÍNICO EM CONTEXTOS DE INCERTEZA: O DESENVOLVIMENTO DE UM TESTE A PARTIR DE SITUAÇÕES EM GERIATRIA

Responsável: Ronaldo Delmonte Piovezan

Você está sendo convidado a participar como voluntário em uma pesquisa. O objetivo principal da pesquisa é **desenvolver, aplicar e analisar um instrumento de avaliação para o raciocínio clínico em situações de incerteza, com enfoque em tópicos relacionados à geriatria, em língua portuguesa, baseado no teste de concordância de script, em um contexto brasileiro de aprendizado.**

Por meio da análise do teste que será elaborado por um grupo do qual você fará parte, junto com as respostas dos outros participantes da pesquisa, poderemos conhecer melhor como se processa e se desenvolve o raciocínio clínico em situações práticas ao longo da formação profissional do médico.

As descobertas e discussões a partir deste instrumento poderão aprofundar o conhecimento a respeito do ensino voltado para habilidades e competências na educação médica.

A – Para a elaboração do teste, em uma entrevista, serão expostos a você os princípios e a estrutura do teste de concordância de script. Em seguida, será feita uma seqüência de encontros, quando você deverá trazer, individualmente, a elaboração de uma proposta por escrito que inclua tópicos em geriatria que sejam relevantes para a construção de situações clínicas problemáticas, mal definidas ou duvidosas.

Para cada uma dessas situações, deverão ser especificados os seguintes tópicos: hipóteses diagnósticas consideradas, estratégias de investigação e opções de tratamento. Além disso, cada um dos integrantes do grupo deverá especificar as perguntas que eles fariam para solucionar cada problema, o exame físico que realizariam e o que esperariam para cada hipótese, além dos exames que solicitariam. Por final, deverão também, especificar as informações clínicas, positivas ou negativas, que eles buscariam em cada caso.

A partir do material recolhido serão construídos, em encontros com os outros integrantes dessa fase de elaboração, os itens do teste.

B – Em qualquer etapa do estudo, você terá acesso aos profissionais responsáveis para o esclarecimento de eventuais dúvidas.
O pesquisador principal é o médico Ronaldo Delmonte Piovezan, que pode ser encontrado no endereço Rua Francisco de Castro, 105 – Vila Clementino – São Paulo – SP – Telefone 55754848, ou pelo e-mail: rdpiovezan@gmail.com.
Se você tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP) – Rua Botucatu, 572 – 1º andar – conjunto 14, telefone: 55711062 - FAX: 55397162 – E-mail: cepunifesp@epm.br;

C – É garantida a liberdade de retirada de consentimento a qualquer momento e de deixar de participar do estudo;

D – O pesquisador Ronaldo Delmonte Piovezan se compromete a não identificar o portador das resoluções do teste que será utilizado como instrumento para a coleta de dados, garantindo assim a confidencialidade;

E – Não há despesas pessoais para o participante em qualquer fase da pesquisa. Também não existirá compensação financeira relacionada à sua participação;

F – O pesquisador se compromete a utilizar os dados coletados somente para esta pesquisa;

G – Acredito ter sido suficientemente esclarecido a respeito das informações que li sobre a pesquisa a ser realizada. Eu discuti com o pesquisador Ronaldo Delmonte

Piovezan sobre a minha decisão em participar deste estudo. Ficaram claros para mim quais são os propósitos do estudo e os procedimentos para a coleta de dados, bem como a garantia de confidencialidade, Ficou também claro que minha participação é isenta de despesas. Concordo voluntariamente em participar desta pesquisa e poderei retirar meu consentimento a qualquer momento, antes ou durante o mesmo, sem penalidades ou qualquer prejuízo.

O(a) geriatra ---------- declara que dá plena autorização a Ronaldo Delmonte Piovezan para proceder às investigações sobre **o instrumento de avaliação para o raciocínio clínico em situações de incerteza, com enfoque em tópicos relacionados à geriatria.**

Assinatura do(a) geriatra: ----------
Cidade: ---------- Estado: ---------- Data:--- de ---- de-----

Termo de Consentimento – FORMATO PARA O GRUPO DE ESPECIALISTAS PARA A FORMAÇÃO DO PAINEL DE REFERÊNCIA

Título da Pesquisa

A AVALIAÇAO DO RACIOCÍNIO CLÍNICO EM CONTEXTOS DE INCERTEZA: O DESENVOLVIMENTO DE UM TESTE A PARTIR DE SITUAÇÕES EM GERIATRIA

Responsável: Ronaldo Delmonte Piovezan

Você está sendo convidado a participar como voluntário em uma pesquisa. O objetivo principal da pesquisa é **desenvolver, aplicar e analisar um instrumento de avaliação para o raciocínio clínico em situações de incerteza, com enfoque em tópicos relacionados à geriatria, em língua portuguesa, baseado no teste de concordância de script, em um contexto brasileiro de aprendizado.**

Por meio da análise de suas respostas ao teste, junto com as respostas dos outros participantes da pesquisa, poderemos conhecer melhor como se processa e se desenvolve o raciocínio clínico em situações práticas ao longo da formação profissional do médico.

As descobertas e discussões a partir deste instrumento poderão aprofundar o conhecimento a respeito do ensino voltado para habilidades e competências na educação médica.

A – Para a coleta de dados será aplicado um teste a partir de situações clínicas em geriatria, sob forma de casos clínicos por escrito. Seguindo cada caso, há itens que relacionam uma hipótese diagnóstica, uma estratégia de investigação ou uma decisão terapêutica a uma nova informação sobre o caso. Por último há uma escala do tipo Likert para cada item, que deverá ser respondida de acordo com ou grau de concordância ou discordância que você acredite haver entre a nova informação e a hipótese, investigação ou terapêutica consideradas.

Após a resolução do teste, o pesquisador principal fará a você perguntas de um questionário sobre características demográficas, formação educacional e práticas docentes e assistenciais. As questões irão englobar informações sobre idade, sexo, ano de graduação, função docente, ano de obtenção de título de especialista e locais de prática clínica atual.
Em seguida este questionário solicitará a sua indicação de itens do teste que considerou confusos ou inadequados, que então poderão ser descartados ou reformulados. Também serão feitas perguntas sobre as dificuldades na resolução do teste e se este colocou situações duvidosas ou desafiadoras para você.

B – Em qualquer etapa do estudo, você terá acesso aos profissionais responsáveis para o esclarecimento de eventuais dúvidas.
O pesquisador principal é o médico Ronaldo Delmonte Piovezan, que pode ser encontrado no endereço Rua Francisco de Castro, 105 – Vila Clementino – São Paulo – SP – Telefone 55754848, ou pelo e-mail: rdpiovezan@gmail.com.
Se você tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP) – Rua Botucatu, 572 – 1º andar – conjunto 14, telefone: 55711062 - FAX: 55397162 – E-mail: cepunifesp@epm.br;

C – É garantida a liberdade de retirada de consentimento a qualquer momento e de deixar de participar do estudo;

D – O pesquisador Ronaldo Delmonte Piovezan se compromete a não identificar o portador das resoluções do teste que será utilizado como instrumento para a coleta de dados, garantindo assim a confidencialidade;

E – Não há despesas pessoais para o participante em qualquer fase da pesquisa. Também não existirá compensação financeira relacionada à sua participação;

F – O pesquisador se compromete a utilizar os dados coletados somente para esta pesquisa;

G – Acredito ter sido suficientemente esclarecido a respeito das informações que li sobre a pesquisa a ser realizada. Eu discuti com o pesquisador Ronaldo Delmonte Piovezan sobre a minha decisão em participar deste estudo. Ficaram claros para mim quais são os propósitos do estudo e os procedimentos para a coleta de dados, bem como a garantia de confidencialidade, Ficou também claro que minha participação é isenta de despesas. Concordo voluntariamente em participar desta pesquisa e poderei retirar meu consentimento a qualquer momento, antes ou durante o mesmo, sem penalidades ou qualquer prejuízo.

O (a) geriatra ---------- declara que dá plena autorização a Ronaldo Delmonte Piovezan para proceder às investigações sobre **o instrumento de avaliação para o raciocínio clínico em situações de incerteza, com enfoque em tópicos relacionados à geriatria.**

Assinatura do (a) geriatra: ----------
Cidade: ---------- Estado: ---------- Data:--- de ---- de-----

Termo de Consentimento – FORMATO PARA O GRUPO DE ESTUDANTES

Título da Pesquisa

A AVALIAÇAO DO RACIOCÍNIO CLÍNICO EM CONTEXTOS DE INCERTEZA: O DESENVOLVIMENTO DE UM TESTE A PARTIR DE SITUAÇÕES EM GERIATRIA

Responsável: Ronaldo Delmonte Piovezan

Você está sendo convidado a participar como voluntário em uma pesquisa. O objetivo principal da pesquisa é **desenvolver, aplicar e analisar um instrumento de avaliação para o raciocínio clínico em situações de incerteza, com enfoque em tópicos relacionados à geriatria, em língua portuguesa, baseado no teste de concordância de script, em um contexto brasileiro de aprendizado.**

Por meio da análise de suas respostas ao teste, junto com as respostas dos outros participantes da pesquisa, poderemos conhecer melhor como se processa e se desenvolve o raciocínio clínico em situações práticas ao longo da formação profissional do médico.

As descobertas e discussões a partir deste instrumento poderão aprofundar o conhecimento a respeito do ensino voltado para habilidades e competências na educação médica.

A – Para a coleta de dados será aplicado um teste a partir de situações clínicas em geriatria, sob forma de casos clínicos por escrito. Seguindo cada caso, há itens que relacionam uma hipótese diagnóstica, uma estratégia de investigação ou uma decisão terapêutica a uma nova informação sobre o caso. Por último há uma escala do tipo Likert para cada item, que deverá ser respondida de acordo com ou grau de concordância ou discordância que você acredite haver entre a nova informação e a hipótese, investigação ou terapêutica consideradas;

B – Em qualquer etapa do estudo, você terá acesso aos profissionais responsáveis para o esclarecimento de eventuais dúvidas.
O pesquisador principal é o médico Ronaldo Delmonte Piovezan, que pode ser encontrado no endereço Rua Francisco de Castro, 105 – Vila Clementino – São Paulo – SP – Telefone 55754848, ou pelo e-mail: rdpiovezan@gmail.com.
Se você tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP) – Rua Botucatu, 572 – 1º andar – conjunto 14, telefone: 55711062 - FAX: 55397162 – E-mail: cepunifesp@epm.br;

C – É garantida a liberdade de retirada de consentimento a qualquer momento e de deixar de participar do estudo;

D – O pesquisador Ronaldo Delmonte Piovezan se compromete a não identificar o portador das resoluções do teste que será utilizado como instrumento para a coleta de dados, garantindo assim a confidencialidade;

E – Não há despesas pessoais para o participante em qualquer fase da pesquisa. Também não existirá compensação financeira relacionada à sua participação;

F – O pesquisador se compromete a utilizar os dados coletados somente para esta pesquisa;

G – Acredito ter sido suficientemente esclarecido a respeito das informações que li sobre a pesquisa a ser realizada. Eu discuti com o pesquisador Ronaldo Delmonte Piovezan sobre a minha decisão em participar deste estudo. Ficaram claros para mim quais são os propósitos do estudo e os procedimentos para a coleta de dados, bem como a garantia de confidencialidade, Ficou também claro que minha participação é isenta de despesas. Concordo voluntariamente em participar desta pesquisa e poderei retirar meu consentimento a qualquer momento, antes ou durante o mesmo, sem penalidades ou qualquer prejuízo.

O (a) estudante ---------- declara que dá plena autorização a Ronaldo Delmonte Piovezan para proceder às investigações sobre **o instrumento de avaliação para o raciocínio clínico em situações de incerteza, com enfoque em tópicos relacionados à geriatria.**

Assinatura do (a) estudante: ----------

Cidade: ---------- Estado: ---------- Data:--- de ---- de-----

Apêndice 7 – Parecer do Comitê de Ética Institucional

UNIFESP
PAULISTA
ESCOLA MEDICINA

Universidade Federal de São Paulo
Escola Paulista de Medicina

Comitê de Ética em Pesquisa
Hospital São Paulo

São Paulo, 27 de abril de 2007.

CEP 0406/07

IImo(a). Sr(a).
Pesquisador(a) RONALDO DEMONTE PIOVEZAN
Co-Investigadores: Nildo Alves Baptista (Orientador)
Disciplina/Departamento: CEDESS da Universidade Federal de São Paulo/Hospital São Paulo
Patrocinador: Recursos Próprios.

PARECER DO COMITÊ DE ÉTICA INSTITUCIONAL

Ref: Projeto de pesquisa intitulado: **"A avaliação do raciocínio clínico em contextos de incerteza: o desenvolvimento de um teste a partir de situação em geriatria".**

CARACTERÍSTICA PRINCIPAL DO ESTUDO: ESTUDO DE APLICAÇÃO E ANÁLISE DE INSTRUMENTO EDUCACIONAL - TRANSVERSAL E QUANTITATIVO.
RISCOS ADICIONAIS PARA O PACIENTE: SEM RISCO.
OBJETIVOS: Desenvolver, aplicar e analisar um instrumento de avaliação para o raciocínio clínico em situações de incerteza, com enfoque em tópicos relacionados à geriatria, em lingua portuguesa, baseado no teste de concordância de script, em um ambiente de aprendizado brasileiro.
RESUMO: O projeto envolverá o desenvolvimento, a aplicação e a análise de um teste escrito que tem como objetivo avaliar o raciocínio clínico em contextos de incerteza envolvendo situações da prática geriátrica. Para isto, 3 geriatras da UNIFESP participarão do projeto como colaboradores para a construção do teste. Outros cinco geriatras da UNIVERSIDADE formarão o grupo para o teste piloto. Mais dezoito geriatras ligados à centros de ensino em geriatria no estado de São Paulo constituirão o painel de especialistas, que serão referências para a comparação do desenvolvimento do raciocínio clínico com os estudantes. Por último, setenta estudantes do internato da UNIFESP serão convidados a resolver o teste. A partir dos resultados colhidos, será feita uma análise estatística descritiva dos grupos e comparativa entre os mesmos, buscando subsídios de validade do instrumento.
FUNDAMENTOS E RACIONAL: Estudo bem fundamentado buscando aprofundar o conhecimento a respeito do ensino voltado para habilidades e competências na educação médica.
MATERIAL E MÉTODO: Estão descritos todos os instrumentos que serão aplicados no desenvolvimento do projeto.
TCLE: TCLE adequado, de acordo com as normas estabelecidas pela Res.196/96.
DETALHAMENTO FINANCEIRO: PROJETO SEM FINANCIAMENTO ESPECÍFICO.
CRONOGRAMA: 15 MESES.
OBJETIVO ACADÊMICO: MESTRADO.
ENTREGA DE RELATÓRIOS PARCIAIS AO CEP PREVISTOS PARA: **26/4/2008** e **26/4/2009**.
O Comitê de Ética em Pesquisa da Universidade Federal de São Paulo/Hospital São Paulo **ANALISOU** e **APROVOU** o projeto de pesquisa referenciado.
1. Comunicar toda e qualquer alteração do projeto e termo de consentimento livre e esclarecido. Nestas circunstâncias a inclusão de pacientes deve ser temporariamente interrompida até a resposta do Comitê, após análise das mudanças propostas.
2. Comunicar imediatamente ao Comitê qualquer evento adverso ocorrido durante o desenvolvimento do estudo.
3. Os dados individuais de todas as etapas da pesquisa devem ser mantidos em local seguro por 5 anos para possível auditoria dos órgãos competentes.
Atenciosamente,

Prof. Dr. José Osmar Medina Pestana
Coordenador do Comitê de Ética em Pesquisa da
Universidade Federal de São Paulo/ Hospital São Paulo

Apêndice 8 - Instruções para a Resolução

(referente às quatro páginas a seguir)

1. O teste é baseado em casos clínicos seguidos de proposições. Cada proposição é considerada um item do teste e não uma alternativa a ser escolhida;

2. Em cada item há um painel de respostas que vai de 1 até 5. Cada um desses números indica um significado de resposta possível, conforme a legenda que se segue abaixo de cada grupo de itens;

3. Após a descrição do caso, há três possíveis agrupamentos de itens. Cada grupo tem no máximo 5 itens. O primeiro grupo é sobre hipóteses diagnósticas. O segundo, sobre ações investigativas. E o terceiro, sobre opções terapêuticas. Nem todos os casos possuem estes três grupos de questões;

4. Cada item ou questão é independente do anterior ou do posterior. Ou seja, a informação contida em um item é uma suposição e não uma informação do caso. Portanto ela não influencia a resposta aos outros itens;

5. As respostas devem ser feitas preferencialmente com um círculo ao redor do número escolhido. Em caso de mudança da resposta, pode-se riscar a escolhida anteriormente e circular uma nova escolha;

6. Não há resposta certa para cada item. Os especialistas que responderem o teste estarão contribuindo para a formação de um painel de referência, que servirá de comparação com outros grupos estudados. Isto significa que estaremos avaliando o raciocínio (e não o conhecimento dos mesmos). A resposta de cada especialista tem um valor único. O valor final de cada uma das opções de resposta depende do número de especialistas que responderem a aquela opção. Esse valor servirá de comparação com os estudantes que serão testados;

7. Na página a seguir, segue-se um modelo de cada tipo de item. Depois, há um caso com itens explicativos para serem resolvidos como treino.

Item para avaliação do raciocínio para diagnóstico

Se você está pensando em	**E você encontra**	**A hipótese se torna**
(uma hipótese diagnóstica)	*(um novo dado clínico, estudo de imagem ou resultado laboratorial)*	1 2 3 4 5

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

Item para avaliação do raciocínio para investigação

Se você está considerando solicitar	**E você encontra**	**A investigação se torna**
(um teste diagnóstico)	*(um novo dado clínico, estudo de imagem ou resultado laboratorial)*	1 2 3 4 5

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

Item para avaliação do raciocínio para o tratamento

Se você está considerando prescrever	**E você encontra**	**A prescrição se torna**
(uma opção terapêutica)	*(um novo dado clínico, estudo de imagem ou resultado laboratorial)*	1 2 3 4 5

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

Exemplo de caso clínico seguido de um grupo de itens sobre diagnóstico em geriatria

Leia cuidadosamente o caso clínico e a escala seguinte:

Um idoso de 82 anos de idade está em avaliação na sala de emergência há 18 horas. Ele apresenta febre. Ele estava apático até há 2 horas atrás. Desde então, ele se apresenta agitado, gritando que assaltantes invadiram a sala de emergência. Ele retirou seu acesso venoso. Sua prescrição indica que ele está recebendo antibióticos endovenosos para pneumonia e medicamentos para insuficiência cardíaca congestiva. Seu médico ambulatorial refere que sua pontuação no último mini-exame do estado mental realizado há 3 meses atrás foi de 23 pontos em 30. Ele sofre de prostatismo e não urina há 12 horas. Ele sempre foi constipado.

Se você está pensando em	**E o paciente ou acompanhante relatam, ou você encontra no exame clínico**	**A hipótese se torna**
Delirium	Alucinações visuais	1 2 3 4 5

1 : praticamente descartada
2 : menos provável
3 : nem mais, nem menos provável
4 : mais provável
5 : praticamente certa

Legenda que orienta as respostas

Escala Likert para ser respondida

Se você está pensando em solicitar	**E você encontra no exame clínico**	**A investigação se torna**
Ecocardiograma Transtorácico	Edema e dor à palpação de panturrilha direita	1 2 3 4 5

1 : contra-indicada totalmente
2 : possivelmente prejudicial
3 : nem mais, nem menos útil
4 : útil
5 : absolutamente necessário

Observe que esta legenda é diferente da anterior

Se você está pensando em prescrever	**E você encontra nos exames colhidos**	**A prescrição se torna**
Furosemida por via endovenosa	Creatinina de 1,5 mg/dl, Uréia de 80 mg/dl e potássio de 4,2 mEq/l	1 2 3 4 5

1 : contra-indicada totalmente
2 : possivelmente prejudicial
3 : nem mais, nem menos útil
4 : útil
5 : absolutamente necessário

Apêndice 9 - Tabelas referentes aos dados e aos escores obtidos

Tabela 9. Respostas obtidas a partir da aplicação do teste no grupo piloto

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1P	4	3	4	3	4	2	4	4	4	3
2P	4	3	3	3	4	4	3	3	5	3
3P	4	4	3	4	4	3	5	5	3	3
4P	4	3	4	3	5	4	4	4	4	3
5P	3	4	4	2	5	2	5	4	5	2

	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
1P	5	4	5	4	3	4	3	3	3	5
2P	5	5	4	3	3	4	3	5	4	5
3P	1	5	5	5	3	4	3	2	2	4
4P	5	5	3	2	2	3	3	3	3	5
5P	5	2	2	3	3	5	2	2	4	5

	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30
1P	4	3	3	3	4	4	4	5	4	3
2P	4	3	3	4	5	4	3	4	4	4
3P	4	2	1	2	5	2	2	5	3	1
4P	4	2	2	3	4	2	4	4	4	4
5P	4	5	4	4	5	2	4	5	5	3

	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40
1P	4	2	4	3	4	5	5	4	4	2
2P	4	3	2	2	5	3	3	3	4	3
3P	5	3	4	2	3	3	5	3	3	3
4P	4	3	5	3	3	3	4	3	3	1
5P	4	2	1	4	3	5	5	4	4	1

	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50
1P	3	5	2	5	3	3	4	4	4	2
2P	3	3	3	3	3	3	4	3	4	4
3P	2	5	3	5	3	2	4	4	3	2
4P	1	5	2	2	3	3	4	4	4	4
5P	1	5	1	3	3	2	4	4	4	4

	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60
1P	3	3	3	4	3	3	3	3	3	4
2P	4	3	4	4	3	4	3	4	3	4
3P	4	4	4	3	3	4	4	3	2	3
4P	3	3	3	3	3	5	3	3	3	3
5P	4	3	4	4	2	5	5	2	3	4

Tabela 9. Respostas obtidas a partir da aplicação do teste no grupo piloto

	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70
1P	3	4	2	2	3	2	3	2	3	4
2P	3	4	3	2	3	3	4	4	3	4
3P	3	4	3	2	3	2	4	5	5	4
4P	3	4	4	3	3	3	4	4	3	3
5P	2	4	4	2	2	2	4	4	4	4

	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80
1P	2	4	4	3	1	3	3	2	1	3
2P	2	4	4	3	2	3	3	2	2	4
3P	2	4	4	4	2	4	3	4	2	3
4P	3	3	2	4	2	3	3	3	4	3
5P	2	5	4	5	4	4	2	4	4	4

	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90
1P	4	1	3	3	3	3	2	2	4	4
2P	4	1	3	4	3	3	3	2	4	3
3P	4	3	2	3	3	2	2	3	4	4
4P	3	2	4	3	3	3	2	2	4	3
5P	3	2	4	3	3	2	1	2	4	4

	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100
1P	3	4	3	2	3	3	3	4	5	5
2P	3	4	4	4	3	3	5	4	5	4
3P	3	4	4	2	4	4	3	4	3	4
4P	3	4	4	3	3	4	3	3	4	4
5P	3	4	4	2	3	4	3	5	5	4

	Q101	Q102	Q103	Q104	Q105	Q106	Q107	Q108	Q109	Q110
1P	3	2	3	4	3	5	4	5	3	2
2P	4	2	4	4	3	2	3	3	3	2
3P	3	2	2	3	2	4	2	2	4	1
4P	5	4	4	4	4	3	3	3	4	1
5P	4	2	2	2	3	2	3	2	4	2

	Q111	Q112	Q113	Q114	Q115
1P	2	3	3	4	3
2P	2	5	3	3	3
3P	3	3	2	5	1
4P	4	3	3	5	3
5P	2	4	2	5	4

Tabela 10. Respostas obtidas a partir da aplicação do teste no grupo de especialistas

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
E1	4	4	4	3	4	3	4	3	2	2
E2	3	3	4	4	4		4	3	4	2
E3	3	3	5	3	5	4	5	4	3	3
E4	4	3	4	2	3	4	5	5	3	5
E5	4	3	2	4	4	4	3	4	4	4
E6	3	4	3	2	4	3	5	3	3	3
E7	4	3	3	3	3	3	5	3	5	2
E8	3	3	4	3	2	3	5	4	3	3
E9	4	3	4	4	4	4	4	4	5	5
E10	3	3	4	2	4	3	4	4	4	3
E11	4	4	4	4	4	4	5	4	5	4
E12	4	4	4	2	4	5	5	5	4	4
E13	4	3	2	3	4	4	4	4	4	4
E14	3	3	4	3	4	4	5	3	5	5
E15	4	4	4	3	4	4	5	3	4	3
E16	4	3	4	3	4	4	4	4	3	5
E17	3	3	3	3	3	4	4	3	4	4
E18	3	3	4	3	4	3	4	3	3	3
E19	4	4	4	3	4	4	4	4	3	4
E20	4	3	4	4	5	3	5	4	4	2
E21	3	4	4	4	4	3	4	3	3	3

	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
E1	2	4	2	3	2	3	3	4	4	2
E2	4	2	4	2	4	4	3	4	3	2
E3	2	5	1	2	4	3	2	4	3	2
E4	4	5	2	2	3	3	2	4	2	4
E5	2	4	4	3	3	2	3	4	3	1
E6	4	3	4	2	3	2	3	4	2	1
E7	4	4	2	2	3	3	3	3	3	3
E8	5	5	1	2	2	3	2	3	2	2
E9	4	4	3	3	3	4	2	4	3	2
E10	3	4	4	2	3	3	3	2	3	2
E11	5	5	2	4	3	4	2	4	4	2
E12	5	4	2	2	2	3	3	4	4	2
E13	4	4	2	2	4	2	2	3	3	2
E14	4	4	1	1	1	3	2	4	4	1
E15	2	4	3	2	2	4	3	4	4	1
E16	5	4	3	2	2	3	2	4	3	1
E17	2	4	4	3	3	3	2	4	4	1
E18	4	4	3	2	2	3	3	4	3	4
E19	1	2	1	2	2	3	2	4		1
E20	5	4	2	2	2	3	3	4	3	2
E21	4	4	3	3	3	3	2	4	3	1

Tabela 10. Respostas obtidas a partir da aplicação do teste no grupo de especialistas

	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30
E1	4	3	4	3	4	4	3	4	5	2
E2	4	3	4	3	3	4	4	3	4	2
E3	2	2	4	3	3	4	4	4	5	1
E4	4	3	4	3	5	5	4	4	3	3
E5	4	2	4	3	3	4	4	4	4	2
E6	3	3	5	3	3	5	4	3	4	2
E7	4	3	3	3	4	4	3	3	4	3
E8	4	3	3	4	5	5	5	4	4	2
E9	4	3	3	3	3	4	4	4	4	3
E10	4	2	4	3	3	4	4	3	4	2
E11	4	3	4	2	5	3	5	5	5	2
E12	5	4	4	3	3	4	3	4	5	2
E13	4	3	4	3	3	4	3	3	5	2
E14	3	2	4	3	5	5	5	5	4	2
E15	5	3	3	4	4	4	3	4	4	3
E16	4	2	4	4	3	5	4	3	4	2
E17	4	2	2	3	4	4	3	4	4	2
E18	4	3	3	3	3	4	4	4	4	2
E19	3	2	4	3	4	5	4	3	4	3
E20	4	2	3	3	3	4	3	3	4	2
E21	3	3	4	3	3	3	3	3	4	2

	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40
E1	4	4	3	2	4	3	4	3	3	2
E2	5	3	4	2	2	4	3	4	4	2
E3	3	5	3	2	3	3	4	3	4	1
E4	3	2	1	2	4	4	4	4	3	4
E5	4	2	1	1	2	4	3	2	3	3
E6	3	3	2	1	4	4	3	4	4	3
E7	3	2	3	2	4	3	3	2	3	2
E8	3	4	2	1	4	3	4	3	4	3
E9	4	2	2	2	2	4	3	4	4	3
E10	2	2	3	2	3	3	3	3	3	3
E11	4	3	3	1	3	4	4	2	1	1
E12	3	4	2	2	4	4	5	4	5	3
E13	3	2	2	2	3	3	2	3	3	3
E14	4	1	1	2	4	4	3	3	3	1
E15	4	1	2	1	4	4	4	3	3	3
E16	5	2	2	1	3	2	3	3	5	3
E17		2	2	2	4	2	3	3	4	4
E18	3	2	3	2	3	3	3	3	4	3
E19	3	2	2	1	3	4	2	2	1	3
E20	4	5	4	2	4	4	4	3	4	3
E21	4	2	2	2	3	4	4	3	4	3

Tabela 10. Respostas obtidas a partir da aplicação do teste no grupo de especialistas

	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50
E1	2	4	4	3	5	4	4	3	4	3
E2	2	4	4	3	4	4	4	4	3	3
E3	4	4	4	3	4	4	4	3	3	3
E4	3	4	4	4	4	4	4	3	4	3
E5	3	4	2	3	4	4	2	4	4	2
E6	1	4	4	3	4	4	2	2	3	4
E7	3	4	4	3	4	5	2	2	3	3
E8	4	4	5	4	5	4	2	2	4	2
E9	3	4	4	4	3	4	4	4	4	3
E10	2	2	4	2	3	3	3	3	3	2
E11	4	2	1	4	5	3	2	3	2	2
E12	2	4	5	4	5	4	4	2	4	2
E13	1	3	4	3	4	3	2	3	3	2
E14	4	5	4	4	4	4	2	2	4	3
E15	1	4	4	4	4	4	2	4	4	4
E16	3	4	4	3	4	4	2	2	2	2
E17	1	4	4	3	3	4	2	4	4	3
E18	2	4	4	3	4	4	2	3	3	3
E19	3	3	4	3	4	4	2	2	3	3
E20	1	3	4	3	4	3	2	2	3	2
E21	1	4	4	3	4	4	4	4	3	3

	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60
E1	4	3	4	3	4	3	4	3	3	4
E2	2	4	3	4	4	4	4	4	4	3
E3	2	4	3	3	4	3	4	3	3	4
E4	2	2	3	4	4	4	3	4	3	3
E5	2	3	3	4	3	4	3	4	3	4
E6	3	4	2	2	4	3	3	2	4	3
E7	2	3	3	4	4	3	3	4	3	2
E8	1	3	3	5	5	3	4	1	4	5
E9	2	3	3	4	4	3	4	2	4	2
E10	2	2	3	4	3	3	3	3	3	4
E11	3	4	2	4	2	3	5	4	4	3
E12	2	4	3	5	4	4	5	2	4	5
E13	2	3	3	3	3	4	3	3	4	4
E14	2	3	3	4	3	4	4	1	4	2
E15	2	2	3	4	4	3	4	4	3	2
E16	2	4	3	2	3	4	4	2	3	3
E17	2	4	4	4	4	3	5	5	3	3
E18	2	4	3	3	4	3	4	4	4	3
E19	1	2	3	3	4	3	4	1	3	4
E20	3	3	3	3	5	4	5	5	3	3
E21	4	2	3	4	3	3	4	3	4	3

Tabela 10. Respostas obtidas a partir da aplicação do teste no grupo de especialistas

	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70
E1	3	3	3	5	2	4	3	3	2	3
E2	3	2	3	4	2	4	2	3	3	4
E3	3	3	3	5	3	3	3	3	3	3
E4	4	2	3	4	3	3	3	4	1	3
E5	4	2	3	4	3	3	2	2	1	4
E6	4	2	3	1	2	4	2	3	1	3
E7	3	3	3	3	3	3	3	3	3	3
E8	4	2	3	4	1	4	2	3	5	3
E9	4	2	4	4	3	4	2	4	4	3
E10	3	2	2	4	3	3	3	3	2	3
E11	4	3	4	4	2	2	3	3	4	4
E12	5	2	3		5	5	1	5	1	4
E13	4	2	3	3	3	3	3	4	2	3
E14	4	3	2	4	3	4	2	3	4	2
E15	4	4	2	4	4	3	2	3	4	4
E16	4	2	3	4	2	3	2	4	4	3
E17	4	1	3	2	2	4	2	3	1	3
E18	4	2	3	3	3	3	3	4	4	3
E19	4	2	3	4	2	4	2	2	3	3
E20	4	5	4	4	2	3	3	4	4	4
E21	4	4	3	4	3	3	3	3	2	4

	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80
E1	4	3	4	4	5	3	3	3	4	3
E2	4	3	4	4	4	2	4	3	4	4
E3	4	2	3	3	5	3	5	1	5	2
E4	3	3	3	3	5	3	2	3	4	4
E5	3	2	3	4	4	2	3	2	4	4
E6	3	2	3	3	5	2	1	2	4	2
E7	3	3	4	3	1	3	3	3	4	3
E8	3	1	3	5	5	2	1	3	3	4
E9	4	3	4	3	4	3	2	3	4	3
E10	3	3	2	3	4	3	2	2	3	2
E11	4	3	4	5	5	3	1	1	4	4
E12	4	1	4	5	5	2	2	3	5	4
E13	4	2	4	3	4	3	1	2	4	3
E14	4	2	3	3	5	2	2	3	4	4
E15	4	2	4	5	5	3	3	2	4	4
E16	4	2	4	5	5	4	1	2	4	4
E17	3	1	4	4	5	2	1	2	4	3
E18	3	2	4	3	5	3	2	3	4	4
E19	2	2	4	4	5	3	2	2	5	2
E20	4	1	4	5	5	2	1	2	4	4
E21	4	4	4	4	4	3	2	2	4	4

Tabela 10. Respostas obtidas a partir da aplicação do teste no grupo de especialistas

	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92
E1	4	4	4	3	3	4	4	2	5	4	4	2
E2	3	3	4	4	4	3	4	3	4	3	3	4
E3	3	3	4	3	3	3	4	2	3	3	3	4
E4	3	3	4	3	4	3	3	3	5	5	4	3
E5	3	3	4	2	4	4	3	3	4	3	3	4
E6	3	3	4	3	4	4	3	2	4	5	3	1
E7	3	3	3	4	3	4	5	2	4	4	3	3
E8	3	4	4	1	3	5	4	1	5	3	3	1
E9	4	3	4	3	4	4	3	3	4	3	3	2
E10	3	3	3	3	4	2	4	3	4	3	3	2
E11	3	3	4	2	4	5	4	4	5	4	3	2
E12	2	3	5	2	4	4	4	1	5	4	4	4
E13	3	3	4	2	3	3	4	1	3	3	2	2
E14	3	4	4	3	4	4	5	4	5	3	4	2
E15	3	3	4	4	4	4	4	2	4	3	3	2
E16	3	4	4	2	4	4	3	3	5	3	3	4
E17	3	3	4	3	3	3	3	3	5	4	3	4
E18	3	4	4	3	3	3	3	3	4	3	4	4
E19	3	3	4	3	3	3	3		3	4	3	3
E20	3	3	4	2	3	3	3	2	5	4	1	1
E21	3	3	4	3	4	4	4	2	5	4	4	4

	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100	Q101	Q102	Q103	Q104
E1	3	3	3	2	4	3	2	2	4	2	4	2
E2	4	4	4	2	3	4	2	4	5	4	4	2
E3	3	3	4	4	3	5	2	2	4	3	4	3
E4	4	3	3	3	3	2	4	4	3	3	5	3
E5	4	4	4	3	4	2	4	2	1	3	4	3
E6	2	2	4	1	2	2	3	2	3	3	4	3
E7	4	3	4	3	4	4	4	2	2	4	4	2
E8	3	4	2	1	4	2	5	4	2	2	4	1
E9	3	4	4	3	3	3	5	3	4	2	4	3
E10	3	2	3	2	3	2	4	4	2	3	4	3
E11	4	3	4	1	3	2	4	1	2	4	5	1
E12	3	3	2	1	3	3	1	1	2	4	5	4
E13	3	2	4	2	3	3	3	4	5	3	~~5~~	3
E14	4	2	2	5	4	2	2	2	2	4	4	2
E15	4	4	4	2	4	2	2	1	2	4	5	4
E16	4	3	4	5	3	2	3	3	4	3	4	2
E17	4	4	2	5	1	5	3	4	4	1	5	5
E18	3	3	3	3	3	2	4	3	3	2	4	2
E19	4	2	4	3	3	2	4	4	4	2	4	2
E20	4	1	4	3	4	2	2	4	5	1	5	2
E21	4	3	4	3	4	3	3	4	2	4	4	3

Tabela 11. Respostas obtidas a partir da aplicação do teste no grupo de estudantes

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
A1	3	4	3	2	4	3	4	5	4	4	2	5	4	1	2
A2	3	3	3	2	4	4	4	5	4	4	2	5	2	2	2
A3	3	4	3	2	4	5	4	4	4	3	3	4	4	2	2
A4	3	5	3	4	4	4	3	4	5	4	2	3	3	4	3
A5	3	4	3	4	4	3	4	4	4	5	2	4	3	1	2
A6	2	5	1	4	2	3	2	4	5	1	2	2	2	1	3
A7	1	5	2	4	4	3	3	4	3	5	4	1	2	3	2
A8	3	4	4	2	4	3	5	4	4	4	4	4	5	3	2
A9	3	4	3	4	4	4	4	4	5	4	1	4	3	1	1
A10	3	4	4	4	4	3	5	3	5	3	5	5	4	2	3
A11	3	4	4	4	3	3	4	3	5	3	1	4	3	3	3
A12	2	5	4	4	4	2	2	3	4	4	4	2	4	1	3
A13	3	3	3	5	5	4	5	5	4	5	5	2	2	2	3
A14	3	4	3	4	4	4	5	4	5	4	2	4	4	1	2
A15	3	3	2	4	1	4	5	3	5	3	3	1	1	2	3
A16	1	4	3	4	4	4	4	3	5	2	4	1	3	1	4
A17	2	4	4	3	4	4	5	5	5	5	2	1	4	3	2
A18	2	3	3	4	2	4	4	4	4	4	2	4	4	2	3
A19	3	5	2	4	3	4	3	4	4	3	2	1	4	1	2
A20	3	4	4	2	3	3	5	4	5	4	1	3	5	2	2
A21	2	4	4	4	3	4	5	5	4	4	2	4	3	2	2
A22	3	3	3	4	3	3	3	3	4	4	3	2	3	2	2
A23	3	4	3	4	3	3	1	3	4	4	1	3	2	2	3
A24	2	5	3	4	4	4	4	4	3	5	2	2	5	2	2
A25	4	4	1	3	4	4	4	5	5	4	2	4	3	3	1
A26	4	4	3	4	4	3	5	4	5	5	4	2	4	2	3
A27	2	5	4	2	2	3	4	5	5	4	2	2	2	1	3
A28	3	4	3	3	5	4	5	4	5	4	1	4	3	2	2
A29	2	5	3	4	4	4	5	5	5	4	2	4	4	2	2
A30	4	4	3	4	2	4	5	5	4	1	1	4	3	4	2
A31	3	4	2	4	5	3	4	3	3	3	2	3	4	2	2
A32	3	4	4	3	4	3	4	4	5	4	3	4	4	4	2
A33	4	5	3	4	4	4	4	5	5	5	1	5	1	2	2
A34	3	4	2	4	2	4	4	4	3	4	3	4	4	2	1
A35	4	4	2	2	4	4	4	3	4	4	3	5	4	4	3
A36	4	3	2	5	4	4	4	4	2	3	1	2	1	4	4
A37	3	4	2	4	4	4	4	5	5	3	3	2	4	4	2
A38	2	4	4	4	4	3	4	4	5	4	2	2	4	3	1
A39	3	4	3	3	4	4	4	5	5	4	3	2	2	4	2
A40	3	4	3	4	4	4	3	3	4	4	1	2	2	3	1
A41	3	4	4	3	3	4	5	5	4	4	3	2	2	3	2

Tabela 11. Respostas obtidas a partir da aplicação do teste no grupo de estudantes

	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30
A1	3	1	4	4	1	4	3	4	3	5	5	4	3	4	2
A2	3	1	4	1	1	4	4	4	4	5	3	3	3	5	2
A3	3	2	3	4	2	5	3	3	3	5	5	5	3	5	2
A4	3	2	3	5	4	5	3	2	3	3	4	4	3	5	2
A5	3	1	4	4	5	4	4	4	3	5	4	3	3	4	2
A6	3	1	3	4	5	1	4	5	2	5	4	5	4	5	3
A7	3	1	4	3	5	5	3	3	4	4	5	4	1	3	3
A8	4	2	3	4	2	4	3	4	2	4	4	3	3	5	3
A9	3	4	4	4	5	2	3	4	3	3	5	5	4	5	3
A10	3	2	4	4	2	3	3	4	3	3	5	5	4	4	2
A11	4	2	3	4	1	3	3	3	2	3	4	3	3	4	2
A12	2	2	4	4	1	5	3	2	3	4	5	4	4	4	3
A13	3	5	4	5	1	5	3	4	4	4	5	4	3	4	3
A14	3	2	4	4	2	4	3	4	4	5	4	3	4	4	3
A15	3	1	4	4	1	5	2	4	2	3	4	3	5	4	2
A16	2	2	4	3	2	4	4	2	3	3	5	5	4	4	3
A17	3	1	4	5	1	5	3	4	4	3	4	5	4	5	3
A18	4	2	3	5	2	4	3	4	3	5	5	4	3	5	2
A19	3	3	3	4	1	2	4	2	2	4	5	3	4	5	2
A20	3	2	4	4	4	4	3	2	4	3	4	4	3	5	2
A21	4	2	4	4	5	4	3	4	3	5	5	4	4	5	3
A22	3	2	4	4	2	4	3	4	3	3	4	3	3	4	3
A23	3	2	3	3	1	4	3	4	3	4	4	3	3	5	3
A24	3	2	4	4	5	4	3	4	4	3	5	5	3	4	2
A25	1	2	4	4	4	3	3	4	4	3	3	3	3	5	2
A26	2	2	4	4	4	5	3	4	3	4	5	5	5	5	2
A27	3	2	4	4	1	4	2	1	3	3	5	3	4	5	2
A28	3	2	3	4	1	4	2	4	3	3	5	4	4	4	2
A29	2	3	4	2	1	4	3	4	4	3	5	4	4	5	3
A30	3	3	4	4	5	3	4	4	4	3	4	5	5	5	2
A31	3	2	4	4	1	5	3	4	2	4	5	5	3	5	2
A32	4	2	3	4	2	4	3	5	4	4	4	4	4	5	2
A33	3	3	4	4	1	3	4	4	4	5	5	2	3	4	2
A34	3	2	4	4	2	3	3	4	3	3	4	4	4	5	4
A35	3	3	3	3	4	3	3	3	4	4	4	3	4	5	3
A36	4	2	1	3	4	3	3	4	4	4	3	4	5	2	5
A37	3	2	4	4	2	2	3	4	4	4	4	4	3	5	1
A38	3	2	4	4	4	2	4	4	2	4	5	3	3	5	3
A39	3	3	4	4	2	4	3	5	3	4	5	4	3	5	3
A40	3	3	4	5	2	4	4	4	3	5	5	4	3	4	3
A41	3	2	4	4	2	2	3	4	3	3	3	4	3	4	3

Tabela 11. Respostas obtidas a partir da aplicação do teste no grupo de estudantes

	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45
A1	3	2	2	2	2	3	3	2	4	3	2	4	4	4	4
A2	4	1	2	1	2	2	5	2	3	2	2	5	4	5	2
A3	4	3	2	2	3	4	3	2	3	5	1	3	2	3	4
A4	3	4	2	2	2	4	4	3	3	2	3	4	2	3	4
A5	3	1	2	1	3	4	4	4	4	3	4	5	4	4	2
A6	3	1	2	1	2	4	1	1	3	4	3	5	3	2	4
A7	2	1	5	2	2	2	3	2	3	1	2	5	4	1	2
A8	2	4	5	1	2	2	2	3	4	2	4	5	3	3	4
A9	3	4	4	2	2	4	4	3	1	2	2	4	4	5	2
A10	3	1	1	3	3	4	3	4	5	3	3	5	4	3	2
A11	3	3	2	2	3	4	3	3	4	2	3	4	4	4	4
A12	3	1	2	1	3	3	3	2	3	4	2	5	4	3	2
A13	3	2	4	1	4	5	3	1	4	3	4	4	4	3	1
A14	3	2	2	2	3	4	3	3	3	3	4	4	3	3	4
A15	3	5	2	2	2	2	4	1	5	3	4	4	4	3	4
A16	1	4	2	2	4	4	4	2	4	2	2	4	4	3	4
A17	4	1	2	2	4	4	3	2	3	3	3	5	4	2	1
A18	3	4	2	2	4	4	3	2	4	3	2	4	4	2	4
A19	3	3	2	2	3	4	2	1	4	3	2	5	4	2	1
A20	5	1	2	2	2	4	2	3	5	1	1	4	4	2	2
A21	4	3	2	3	4	3	4	1	2	2	3	4	3	2	4
A22	3	3	2	1	2	4	3	1	3	3	3	4	4	3	3
A23	3	1	2	2	3	4	3	3	3	1	1	4	4	3	4
A24	3	1	2	1	3	3	4	2	3	3	3	3	4	3	4
A25	3	4	1	1	1	3	4	4	3	4	3	4	3	3	4
A26	4	5	2	4	2	4	4	2	5	3	1	5	4	3	1
A27	3	4	2	2	2	4	2	1	2	2	2	1	5	2	5
A28	3	5	2	2	3	4	4	2	4	3	2	5	4	4	5
A29	3	1	2	2	3	4	4	3	3	1	1	4	4	3	5
A30	2	1	2	2	3	2	3	2	3	2	3	4	3	3	4
A31	3	5	1	2	3	4	3	4	4	3	1	4	4	3	2
A32	3	4	2	2	3	2	4	3	3	3	2	4	3	5	4
A33	3	4	2	1	3	3	4	1	1	3	1	5	3	2	5
A34	3	4	2	2	3	4	4	3	4	3	1	4	4	4	5
A35	5	2	2	2	2	4	4	3	3	4	3	4	4	3	4
A36	3	5	4	1	3	4	5	3	1	3	4	4	4	3	5
A37	3	4	1	1	3	2	3	3	3	3	3	4	2	4	4
A38	3	1	2	1	3	3	3	4	4	3	1	4	4	3	4
A39	3	5	3	1	4	3	2	1	3	3	2	5	4	3	3
A40	3	5	4	1	2	2	3	4	3	3	3	5	4	3	4
A41	4	4	2	1	3	2	3	4	3	2	2	4	4		2

Tabela 11. Respostas obtidas a partir da aplicação do teste no grupo de estudantes

	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60
A1	4	2	2	2	3	4	3	3	4	4	4	4	5	4	2
A2	5	4	4	2	1	4	3	3	3	4	5	4	2	4	2
A3	4	4	4	4	2	1	3	3	3	3	3	5	2	4	2
A4	5	3	4	4	3	2	3	3	4	4	3	3	5	3	2
A5	4	2	4	2	3	2	4	4	3	4	3	4	1	4	3
A6	4	5	4	4	1	1	1	3	3	5	4	5	4	4	5
A7	4	1	4	4	3	3	3	3	3	3	5	3	4	5	3
A8	5	1	4	3	3	3	2	4	3	4	3	4	2	4	3
A9	3	4	3	4	3	4	4	1	3	4	4	5	3	3	4
A10	4	2	4	4	3	1	2	3	3	4	3	5	4	4	3
A11	4	2	1	4	3	1	2	3	4	4	3	5	2	4	3
A12	4	4	1	3	3	1	4	3	3	3	3	4	3	4	4
A13	4	2	4	4	3	1	3	4	3	3	4	3	4	4	4
A14	4	2	2	4	3	2	3	3	4	4	3	4	4	4	2
A15	4	3	1	4	2	3	4	5	3	3	3	4	2	4	4
A16	3	4	4	4	3	2	4	3	4	4	3	5	3	4	4
A17	3	2	2	4	3	2	4	3	4	3	4	4	4	5	3
A18	3	4	2	4	2	2	3	3	4	4	3	4	2	4	4
A19	4	4	5	1	1	2	2	3	4	4	3	4	2	5	1
A20	4	2	3	4	2	2	4	2	3	4	4	5	1	4	2
A21	4	2	4	2	3	1	3	1	4	3	3	4	4	4	3
A22	3	1	2	2	4	1	3	3	3	3	3	3	3	4	4
A23	4	2	4	4	3	2	3	3	4	4	3	3	3	3	3
A24	3	2	2	4	2	2	4	3	3	4	3	4	4	4	3
A25	4	3	4	4	3	2	3	4	4	4	3	4	3	5	3
A26	4	2	5	4	2	1	5	3	4	5	3	5	3	4	5
A27	5	2	1	4	3	2	5	3	4	2	3	4	2	4	4
A28	3	2	4	4	2	2	4	2	4	5	3	4	4	3	1
A29	5	4	4	4	2	4	5	4	4	4	3	5	2	5	2
A30	4	4	3	4	3	3	4	3	4	3	3	4	3	5	2
A31	4	1	4	4	3	2	2	3	3	3	3	4	4	4	2
A32	4	1	4	3	3	2	2	3	4	4	3	4	4	4	3
A33	3	1	3	4	3	2	4	4	4	2	5	4	5	5	3
A34	4	4	4	3	3	2	2	4	3	4	4	4	4	4	2
A35	4	4	2	1	3	3	4	3	3	3	4	3	3	4	2
A36	5	4	4	1	3	4	4	4	3	4	4	4	4	4	2
A37	4	2	2	2	3	1	2	3	3	4	3	4	4	4	1
A38	4	4	5	3	3	2	3	4	3	4	3	5	5	5	1
A39	4	5	4	3	4	1	1	3	4	4	4	4	5	4	3
A40	4	4	2	4	3	2	3	3	3	3	3	5	4	4	2
A41	3	2	3	3	2	3	4	4	4	3	3	4	3	4	2

Tabela 11. Respostas obtidas a partir da aplicação do teste no grupo de estudantes

	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75
A1	4	2	4	3	3	3	3	2	4	3	4	2	4	3	4
A2	4	1	4	2	3	5	1	2	3	3	4	1	4	3	5
A3	4	2	4	3	3	2	2	4	3	2	4	3	5	3	4
A4	5	3	3	2	3	3	3	3	1	4	4	2	4	2	1
A5	4	2	4	4	3	3	2	3	4	3	4	2	4	3	4
A6	5	5	1	3	1	4	2	2	1	2	4	1	1	3	5
A7	4	1	3	2	3	4	1	2	2	2	4	3	3	3	5
A8	4	1	3	3	3	3	3	3	4	4	5	2	4	3	3
A9	4	2	3	2	2	3	2	2	2	4	4	4	5	3	4
A10	4	1	4	4	3	3	2	3	1	4	3	2	4	3	5
A11	4	2	4	3	3	2	1	3	4	4	4	2	4	3	5
A12	4	2	2	2	3	4	2	2	2	2	4	2	3	4	5
A13	4	2	3	3	4	4	2	3	3	4	4	4	4	4	5
A14	4	2	3	4	4	4	2	3	2	4	3	3	4	4	4
A15	4	1	4	4	4	3	2	3	4	3	4	2	4	3	5
A16	4	1	4	3	3	3	4	3	4	4	4	2	4	3	4
A17	5	3	1	4	3	3	2	3	4	4	5	2	5	5	5
A18	5	2	2	4	4	3	3	3	4	4	4	2	4	4	5
A19	4	1	2	3	4	2	1	3	1	2	3	1	3	4	5
A20	4	1	1	4	3	2	3	3	4	4	4	3	4	4	5
A21	3	3	2	4	3	3	3	3	4	2	4	2	3	3	4
A22	4	1	2	3	4	3	2	3	5	4	4	2	3	4	5
A23	4	2	2	4	4	3	2	2	2	4	4	3	2	4	4
A24	4	2	4	4	3	4	2	1	4	3	4	2	3	4	5
A25	3	4	4	4	2	2	1	3	4	4	4	1	4	3	4
A26	4	1	4	5	1	5	2	3	5	4	4	2	3	3	5
A27	4	1	4	3	3	2	2	3	1	2	1	2	2	4	5
A28	2	1	5	4	2	3	1	2	4	3	4	2	4	3	5
A29	4	1	4	5	2	4	3	3	2	4	4	2	5	3	2
A30	4	2	4	3	2	3	2	3	4	4	4	2	4	4	5
A31	2	2	3	4	3	3	2	3	4	3	5	1	4	4	5
A32	4	2	4	4	3	3	3	3	4	4	4	3	4	3	4
A33	4	4	1	4	1	5	2	3	5	4	4	2	3	3	3
A34	3	2	2	4	3	3	2	3	5	4	4	2	4	3	3
A35	4	1	4	3	3	3	4	4	4	4	5	2	4	3	4
A36	4	1	4	4	3	2	1	3	4	4	4	2	4	3	2
A37	4	2	4	4	2	2	2	2	4	2	4	2	4	3	4
A38	4	4	4	4	3	4	4	1	4	4	2	1	4	3	5
A39	4	2	2	4	3	4	1	1	4	3	4	2	4	3	3
A40	5	1	4	4	3	3	3	3	5	4	5	3	4	3	5
A41	4	2	4	4	2	3	3	3	4	4	4	3	2	4	4

Tabela 11. Respostas obtidas a partir da aplicação do teste no grupo de estudantes

	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90
A1	4	2	3	4	4	3	3	4	3	4	4	4	4	3	4
A2	2	1	3	5	3	3	2	4	1	4	3	3	2	5	5
A3	2	2	3	4	4	3	3	4	3	4	4	4	2	5	4
A4	4	4	2	4	4	3	3	4	2	4	2	4	3	4	3
A5	3	2	2	4	4	3	3	4	4	4	3	5	2	5	4
A6	2	4	3	4	3	2	3	4	4	4	3	3	2	5	3
A7	2	1	2	5	2	3	3	4	3	4	3	3	3	5	3
A8	4	2	2	4	4	4	3	3	2	2	3	3	2	4	3
A9	2	2	3	4	3	2	3	4	3	4	3	1	3	4	4
A10	2	1	2	4	5	4	2	4	3	4	4	3	3	5	5
A11	2	2	2	4	4	4	3	5	4	4	3	3	2	4	4
A12	2	4	3	5	3	4	3	4	3	4	3	3	3	5	4
A13	3	2	2	4	4	3	4	4	3	4	4	4	3	3	4
A14	4	2	3	4	4	3	3	4	2	3	4	3	2	4	4
A15	2	4	2	4	3	2	4	4	2	4	3	3	1	5	5
A16	2	2	3	4	4	3	4	4	4	4	3	3	3	4	4
A17	2	1	2	5	4	2	3	4	4	5	5	4	2	5	3
A18	3	4	2	4	3	4	3	4	2	4	3	4	3	5	4
A19	2	4	3	5	5	3	4	2	4	4	4	5	5	5	4
A20	2	2	2	4	3	3	3	4	3	4	3	3	2	4	3
A21	1	2	2	4	1	3	4	4	3	4	3	4	3	4	3
A22	2	2	2	4	3	3	3	3	4	4	3	3	2	3	3
A23	2	2	2	4	2	3	3	4	3	3	3	3	3	2	4
A24	4	1	2	5	3	3	3	3	4	5	3	3	4	5	4
A25	4	4	2	4	3	1	1	3	1	4	4	3	3	3	3
A26	3	2	5	4	2	3	3	4	2	5	4	4	3	5	3
A27	2	1	2	4	4	3	1	4	5	5	5	2	4	5	5
A28	2	1	2	4	4	4	3	4	3	4	5	4	3	5	4
A29	4	4	2	5	4	4	4	3	3	3	4	3	3	5	3
A30	4	2	2	4	4	4	3	4	4	4	5	5	2	5	5
A31	4	4	2	5	3	3	4	5	3	4	5	4	3	5	3
A32	4	4	2	4	2	2	3	4	4	4	3	4	2	4	3
A33	1	2	2	4	4	2	2	4	4	4	3	1	1	4	4
A34	3	4	2	4	4	4	3	3	3	4	4	3	3	4	4
A35	4	4	3	3	3	3	2	4	2	3	3	3	3	2	3
A36	4	3	1	4	4	4	4	3	2	4	4	1	3	4	3
A37	2	2	3	4	3	4	3	4	3	4	3	4	3	4	4
A38	1	1	2	4	4	2	3	3	3	4	3	3	1	5	5
A39	4	3	2	5	4	4	5	4	2	4	3	4	2	4	4
A40	4	4	2	4	4	4	3	4	3	4	3	3	3	4	4
A41	3	4	2	4	3	3	3	4	4	3	4	3	1	4	3

Tabela 11. Respostas obtidas a partir da aplicação do teste no grupo de estudantes

	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100	Q101	Q102	Q103	Q104
A1	3	2	4	4	2	3	3	2	4	2	2	4	4	3
A2	3	4	4	4	2	3	2	2	5	1	1	2	2	2
A3	3	2	4	2	2	3	3	4	4	4	2	5	1	3
A4	3	4	2	4	5	2	3	3	4	3	4	3	3	4
A5	3	2	4	4	4	4	3	2	2	4	4	4	2	4
A6	3	4	1	4	4	4	3	2	2	2	1	5	2	2
A7	3	2	4	2	4	3	3	2	5	1	4	3	5	5
A8	2	3	2	4	4	3	4	4	5	3	3	3	2	3
A9	3	4	4	3	2	4	3	2	5	5	5	5	5	5
A10	3	2	4	3	3	3	3	2	5	5	1	4	2	1
A11	4	2	3	2	4	2	2	2	5	2	1	3	4	2
A12	4	1	4	3	4	3	4	4	4	2	1	4	3	4
A13	3	4	4	3	4	4	4	2	4	3	2	5	4	4
A14	3	2	4	2	3	2	4	2	4	4	1	4	4	3
A15	3	2	4	2	4	3	3	2	4	1	1	5	5	4
A16	3	4	3	4	4	3	3	4	4	4	4	4	4	2
A17	2	1	4	3	4	3	2	3	5	1	2	3	2	2
A18	3	2	4	2	4	3	3	4	4	4	2	4	3	3
A19	2	4	4	4	4	3	3	2	5	1	4	4	5	2
A20	3	2	4	2	2	1	2	4	4	2	2	4	4	2
A21	3	2	3	3	3	4	2	2	4	2	3	3	5	4
A22	3	4	2	4	4	3	3	2	2	3	1	3	5	2
A23	4	2	3	3	4	3	2	2	3	2	2	4	3	2
A24	3	1	4	3	4	3	3	2	5	1	2	2	2	1
A25	3	4	3	3	4	3	3	1	4	3	3	4	4	3
A26	3	4	4	4	4	4	3	2	4	4	2	5	1	5
A27	4	2	3	2	3	3	2	2	5	4	2	4	3	2
A28	4	2	3	4	4	3	3	2	5	1	3	3	1	2
A29	3	4	3	4	4	2	4	2	5	2	5	5	2	2
A30	3	3	4	3	4	3	4	2	2	2	2	4	1	2
A31	3	2	4	4	4	2	3	2	5	2	4	3	4	2
A32	4	2	4	4	4	3	3	2	4	2	2	4	4	2
A33	3	4	3	3	4	3	2	2	3	2	4	3	3	5
A34	3	2	4	3	4	3	3	2	4	4	4	3	2	4
A35	3	4	4	4	3	3	2	4	5	2	4	3	2	3
A36	3	4	3	3	4	4	3	2	4	2	5	2	4	5
A37	3	4	4	3	4	3	3	2	4	1	4	2	4	2
A38	3	1	4	3	4	3	4	2	4	2	1	5	2	3
A39	3	2	4	4	2	3	3	1	5	1	1	2	1	4
A40	3	4	3	3	4	4	3	4	5	2	1	4	4	4
A41		2		2	4	2	3	2	4	2	1	2	4	2

Tabela 12. Freqüência das opções de resposta para cada questão, entre os especialistas, para a formação do escore

Resposta	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	2	4	1	0	0	0	1	4
3	9	14	3	11	3	8	1	9	8	7
4	12	7	15	6	15	11	10	10	8	6
5	0	0	1	0	2	1	10	2	4	4

Resposta	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
1	1	0	4	1	1	0	0	0	0	8
2	5	2	7	14	8	3	11	1	3	10
3	1	1	5	5	9	14	10	3	11	1
4	9	14	5	1	3	4	0	17	6	2
5	5	4	0	0	0	0	0	0	0	0

Resposta	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30
1	0	0	0	0	0	0	0	0	0	1
2	1	8	1	1	0	0	0	0	0	15
3	4	12	6	17	12	2	8	9	1	5
4	14	1	13	3	5	13	10	10	15	0
5	2	0	1	0	4	6	3	2	5	0

Resposta	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40
1	0	2	3	7	0	0	0	0	2	3
2	2	11	10	14	3	2	2	4	0	3
3	9	3	6	0	8	7	10	12	8	13
4	8	3	2	0	10	12	8	5	9	2
5	2	2	0	0	0	0	1	0	2	0

Resposta	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50
1	6	0	1	0	0	0	0	0	0	0
2	5	2	1	1	0	0	13	8	2	8
3	6	3	0	13	3	4	1	7	10	11
4	4	15	17	7	14	16	7	6	9	2
5	0	1	2	0	4	1	0	0	0	0

Resposta	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60
1	2	0	0	0	0	0	0	3	0	0
2	14	5	2	2	1	0	0	4	0	4
3	3	8	17	6	6	13	6	5	11	9
4	2	8	2	11	12	8	11	7	10	6
5	0	0	0	2	2	0	4	3	0	2

Tabela 12. Freqüência das opções de resposta para cada questão, entre os especialistas, para a formação do escore

Resposta	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70
1	0	1	0	1	1	0	1	0	5	0
2	0	12	3	1	8	1	10	2	4	1
3	5	5	15	3	10	11	10	12	4	13
4	15	2	3	13	1	8	0	6	7	7
5	1	1	0	2	1	1	0	1	1	0

Resposta	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80
1	0	4	0	0	1	0	7	2	0	0
2	1	9	1	0	0	8	8	10	0	4
3	8	7	6	9	0	12	4	9	2	5
4	12	1	14	6	6	1	1	0	16	12
5	0	0	0	6	14	0	1	0	3	0

Resposta	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90
1	0	0	0	1	0	0	0	3	0	0
2	1	0	0	6	0	1	0	7	0	0
3	18	16	2	11	9	8	9	8	3	11
4	2	5	18	3	12	10	10	2	8	8
5	0	0	1	0	0	2	2	0	10	2

Resposta	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100
1	1	3	0	1	0	4	1	0	1	3
2	1	7	1	5	4	5	1	12	6	6
3	13	3	8	9	4	8	11	5	5	3
4	6	8	12	6	13	1	8	2	7	9
5	0	0	0	0	0	3	0	2	2	0

Resposta	Q101	Q102	Q103	Q104
1	1	2	0	2
2	8	5	0	8
3	3	7	0	8
4	6	7	14	2
5	3	0	7	1

Tabela 13. Escores calculados para cada opção de resposta, em cada questão

Resposta	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0,13	0,36	0,07	0	0	0	0,09	0,57
3	0,75	1	0,2	1	0,2	0,73	0,1	0,9	1	1
4	1	0,5	1	0,55	1	1	1	1	1	0,86
5	0	0	0,07	0	0,13	0,09	1	0,2	0,5	0,57

Resposta	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
1	0,2	0	0,57	0,07	0,11	0	0	0	0	0,8
2	0,56	0,14	1	1	0,89	0,21	1	0,06	0,27	1
3	0,2	0,07	0,71	0,35	1	1	0,91	0,18	1	0,1
4	1	1	0,71	0,07	0,33	0,28	0	1	0,54	0,2
5	0,56	0,28	0	0	0	0	0	0	0	0

Resposta	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30
1	0	0	0	0	0	0	0	0	0	0,07
2	0,07	0,67	0,08	0,06	0	0	0	0	0	1
3	0,28	1	0,46	1	1	0,15	0,8	0,9	0,07	0,33
4	1	0,08	1	0,18	0,42	1	1	1	1	0
5	0,14	0	0,08	0	0,33	0,46	0,3	0,2	0,33	0

Resposta	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40
1	0	0,18	0,3	0,5	0	0	0	0	0,22	0,23
2	0,22	1	1	1	0,3	0,17	0,2	0,33	0	0,23
3	1	0,27	0,6	0	0,8	0,58	1	1	0,89	1
4	0,89	0,27	0,2	0	1	1	0,8	0,42	1	0,15
5	0,22	0,18	0	0	0	0	0,1	0	0,22	0

Resposta	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50
1	1	0	0,06	0	0	0	0	0	0	0
2	0,83	0,13	0,06	0,08	0	0	1	1	0,2	0,72
3	1	0,2	0	1	0,21	0,25	0,08	0,88	1	1
4	0,67	1	1	0,54	1	1	0,54	0,75	0,9	0,18
5	0	0,07	0,12	0	0,29	0,06	0	0	0	0

Resposta	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60
1	0,14	0	0	0	0	0	0	0,43	0	0
2	1	0,63	0,12	0,18	0,08	0	0	0,57	0	0,44
3	0,21	1	1	0,55	0,5	1	0,55	0,71	1	1
4	0,14	1	0,12	1	1	0,62	1	1	0,91	0,67
5	0	0	0	0,18	0,17	0	0,36	0,43	0	0,22

Resposta	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70
1	0	0,08	0	0,08	0,1	0	0,1	0	0,71	0
2	0	1	0,2	0,08	0,8	0,09	1	0,17	0,57	0,08
3	0,33	0,42	1	0,23	1	1	1	1	0,57	1
4	1	0,17	0,2	1	0,1	0,72	0	0,5	1	0,54
5	0,07	0,08	0	0,15	0,1	0,09	0	0,08	0,14	0

Tabela 13. Escores calculados para cada opção de resposta, em cada questão

Resposta	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80
1	0	0,57	0	0	0,07	0	0,88	0,2	0	0
2	0,08	1	0,07	0	0	0,67	1	1	0	0,33
3	0,67	0,78	0,43	1	0	1	0,5	0,9	0,13	0,42
4	1	0,11	1	0,67	0,43	0,08	0,13	0	1	1
5	0	0	0	0,67	1	0	0,13	0	0,19	0

Resposta	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90
1	0	0	0	0,09	0	0	0	0,38	0	0
2	0,06	0	0	0,54	0	0,1	0	0,88	0	0
3	1	1	0,11	1	0,75	0,8	0,9	1	0,3	1
4	0,11	0,31	1	0,27	1	1	1	0,25	0,8	0,73
5	0	0	0,06	0	0	0,2	0,2	0	1	0,18

Resposta	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100
1	0,08	0,38	0	0,11	0	0,5	0,09	0	0,14	0,33
2	0,08	0,88	0,08	0,56	0,31	0,63	0,09	1	0,86	0,67
3	1	0,38	0,67	1	0,31	1	1	0,42	0,71	0,33
4	0,46	1	1	0,67	1	0,13	0,73	0,17	1	1
5	0	0	0	0	0	0,38	0	0,17	0,29	0

Resposta	Q101	Q102	Q103	Q104
1	0,13	0,29	0	0,25
2	1	0,86	0	1
3	0,38	1	0	1
4	0,75	1	1	0,25
5	0,38	0	0,5	0,13

Tabela 14. Escores individuais e totais para os especialistas

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
E1	1	0,5	1	1	1	0,73	1	0,9	0,09	0,57
E2	0,75	1	1	0,55	1	0	1	0,9	1	0,57
E3	0,75	1	0,07	1	0,13	1	1	1	1	1
E4	1	1	1	0,36	0,2	1	1	0,2	1	0,57
E5	1	1	0,13	0,55	1	1	0,1	1	1	0,86
E6	0,75	0,5	0,2	0,36	1	0,73	1	0,9	1	1
E7	1	1	0,2	1	0,2	0,73	1	0,9	0,5	0,57
E8	0,75	1	1	1	0,07	0,73	1	1	1	1
E9	1	1	1	0,55	1	1	1	1	0,5	0,57
E10	0,75	1	1	0,36	1	0,73	1	1	1	1
E11	1	0,5	1	0,55	1	1	1	1	0,5	0,86
E12	1	0,5	1	0,36	1	0,09	1	0,2	1	0,86
E13	1	1	0,13	1	1	1	1	1	1	0,86
E14	0,75	1	1	1	1	1	1	0,9	0,5	0,57
E15	1	0,5	1	1	1	1	1	0,9	1	1
E16	1	1	1	1	1	1	1	1	1	0,57
E17	0,75	1	0,2	1	0,2	1	1	0,9	1	0,86
E18	0,75	1	1	1	1	0,73	1	0,9	1	1
E19	1	0,5	1	1	1	1	1	1	1	0,86
E20	1	1	1	0,55	0,13	0,73	1	1	1	0,57
E21	0,75	0,5	1	0,55	1	0,73	1	0,9	1	1

	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
E1	0,56	1	1	0,35	0,89	1	0,91	1	0,54	1
E2	1	0,14	0,71	1	0,33	0,28	0,91	1	1	1
E3	0,56	0,28	0,57	1	0,33	1	1	1	1	1
E4	1	0,28	1	1	1	1	1	1	0,27	0,2
E5	0,56	1	0,71	0,35	1	0,21	0,91	1	1	0,8
E6	1	0,07	0,71	1	1	0,21	0,91	1	0,27	0,8
E7	1	1	1	1	1	1	0,91	0,18	1	0,1
E8	0,56	0,28	0,57	1	0,89	1	1	0,18	0,27	1
E9	1	1	0,71	0,35	1	0,28	1	1	1	1
E10	0,2	1	0,71	1	1	1	0,91	0,06	1	1
E11	0,56	0,28	1	0,07	1	0,28	1	1	0,54	1
E12	0,56	1	1	1	0,89	1	0,91	1	0,54	1
E13	1	1	1	1	0,33	0,21	1	0,18	1	1
E14	1	1	0,57	0,07	0,11	1	1	1	0,54	0,8
E15	0,56	1	0,71	1	0,89	0,28	0,91	1	0,54	0,8
E16	0,56	1	0,71	1	0,89	1	1	1	1	0,8
E17	0,56	1	0,71	0,35	1	1	1	1	0,54	0,8
E18	1	1	0,71	1	0,89	1	0,91	1	1	0,2
E19	0,2	0,14	0,57	1	0,89	1	1	1	0	0,8
E20	0,56	1	1	1	0,89	1	0,91	1	1	1
E21	1	1	0,71	0,35	1	1	1	1	1	0,8

Tabela 14. Escores individuais e totais para os especialistas

	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30
E1	1	1	1	1	0,42	1	0,8	1	0,33	1
E2	1	1	1	1	1	1	1	0,9	1	1
E3	0,07	0,67	1	1	1	1	1	1	0,33	0,07
E4	1	1	1	1	0,33	0,46	1	1	0,07	0,33
E5	1	0,67	1	1	1	1	1	1	1	1
E6	0,28	1	0,08	1	1	0,46	1	0,9	1	1
E7	1	1	0,46	1	0,42	1	0,8	0,9	1	0,33
E8	1	1	0,46	0,18	0,33	0,46	0,3	1	1	1
E9	1	1	0,46	1	1	1	1	1	1	0,33
E10	1	0,67	1	1	1	1	1	0,9	1	1
E11	1	1	1	0,06	0,33	0,15	0,3	0,2	0,33	1
E12	0	0,08	1	1	1	1	0,8	1	0,33	1
E13	1	1	1	1	1	1	0,8	0,9	0,33	1
E14	0,28	0,67	1	1	0,33	0,46	0,3	0,2	1	1
E15	0	1	0,46	0,18	0,42	1	0,8	1	1	0,33
E16	1	0,67	1	0,18	1	0,46	1	0,9	1	1
E17	1	0,67	0,08	1	0,42	1	0,8	1	1	1
E18	1	1	0,46	1	1	1	1	1	1	1
E19	0,28	0,67	1	1	0,42	0,46	1	0,9	1	0,33
E20	1	0,67	0,46	1	1	1	0,8	0,9	1	1
E21	0,28	1	1	1	1	0,15	0,8	0,9	1	1

	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40
E1	0,89	0,27	0,62	1	1	0,58	0,8	1	0,89	0,23
E2	0,22	0,27	0,2	1	0,3	1	1	0,42	1	0,23
E3	1	0,18	0,62	1	0,8	0,58	0,8	1	1	0
E4	1	1	0,3	1	1	1	0,8	0,42	0,89	0,15
E5	0,89	1	0,3	0,5	0,3	1	1	0,33	0,89	1
E6	1	0,27	1	0,5	1	1	1	0,42	1	1
E7	1	1	0,62	1	1	0,58	1	0,33	0,89	0,23
E8	1	0,27	1	0,5	1	0,58	0,8	1	1	1
E9	0,89	1	1	1	0,3	1	1	0,42	1	1
E10	0,22	1	0,62	1	0,8	0,58	1	1	0,89	1
E11	0,89	0,27	0,62	0,5	0,8	1	0,8	0,33	0,22	0
E12	1	0,27	1	1	1	1	0,1	0,42	0,22	1
E13	1	1	1	1	0,8	0,58	0,2	1	0,89	1
E14	0,89	0,18	0,3	1	1	1	1	1	0,89	0
E15	0,89	0,18	1	0,5	1	1	0,8	1	0,89	1
E16	0,22	1	1	0,5	0,8	0,17	1	1	0,22	1
E17	0	1	1	1	1	0,17	1	1	1	0,15
E18	1	1	0,62	1	0,8	0,58	1	1	1	1
E19	1	1	1	0,5	0,8	1	0,2	0,33	0,22	1
E20	0,89	0,18	0,2	1	1	1	0,8	1	1	1
E21	0,89	1	1	1	0,8	1	0,8	1	1	1

Tabela 14. Escores individuais e totais para os especialistas

	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50
E1	0,83	1	1	1	0,29	1	0,54	0,88	0,9	1
E2	0,83	1	1	1	1	1	0,54	0,75	1	1
E3	0,67	1	1	1	1	1	0,54	0,88	1	1
E4	1	1	1	0,54	1	1	0,54	0,88	0,9	1
E5	1	1	0,06	1	1	1	1	0,75	0,9	0,72
E6	1	1	1	1	1	1	1	1	1	0,18
E7	1	1	1	1	1	0,06	1	1	1	1
E8	0,67	1	0,12	0,54	0,29	1	1	1	0,9	0,72
E9	1	1	1	0,54	0,21	1	0,54	0,75	0,9	1
E10	0,83	0,13	1	0,08	0,21	0,25	0,08	0,88	1	0,72
E11	0,67	0,13	0,06	0,54	0,29	0,25	1	0,88	0,2	0,72
E12	0,83	1	0,12	0,54	0,29	1	0,54	1	0,9	0,72
E13	1	0,2	1	1	1	0,25	1	0,88	1	0,72
E14	0,67	0,07	1	0,54	1	1	1	1	0,9	1
E15	1	1	1	0,54	1	1	1	0,75	0,9	0,18
E16	1	1	1	1	1	1	1	1	0,2	0,72
E17	1	1	1	1	0,21	1	1	0,75	0,9	1
E18	0,83	1	1	1	1	1	1	0,88	1	1
E19	1	0,2	1	1	1	1	1	1	1	1
E20	1	0,2	1	1	1	0,25	1	1	1	0,72
E21	1	1	1	1	1	1	0,54	0,75	1	1

	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60
E1	0,14	1	0,12	0,55	1	1	1	0,71	1	0,67
E2	1	1	1	1	1	0,62	1	1	0,91	1
E3	1	1	1	0,55	1	1	1	0,71	1	0,67
E4	1	0,63	1	1	1	0,62	0,55	1	1	1
E5	1	1	1	1	0,5	0,62	0,55	1	1	0,67
E6	0,21	1	0,12	0,18	1	1	0,55	0,57	0,91	1
E7	1	1	1	1	1	1	0,55	1	1	0,44
E8	0,14	1	1	0,18	0,17	1	1	0,43	0,91	0,22
E9	1	1	1	1	1	1	1	0,57	0,91	0,44
E10	1	0,63	1	1	0,5	1	0,55	0,71	1	0,67
E11	0,21	1	0,12	1	0,08	1	0,36	1	0,91	1
E12	1	1	1	0,18	1	0,62	0,36	0,57	0,91	0,22
E13	1	1	1	0,55	0,5	0,62	0,55	0,71	0,91	0,67
E14	1	1	1	1	0,5	0,62	1	0,43	0,91	0,44
E15	1	0,63	1	1	1	1	1	1	1	0,44
E16	1	1	1	0,18	0,5	0,62	1	0,57	1	1
E17	1	1	0,12	1	1	1	0,36	0,43	1	1
E18	1	1	1	0,55	1	1	1	1	0,91	1
E19	0,14	0,63	1	0,55	1	1	1	0,43	1	0,67
E20	0,21	1	1	0,55	0,17	0,62	0,36	0,43	1	1
E21	0,14	0,63	1	1	0,5	1	1	0,71	0,91	1

Tabela 14. Escores individuais e totais para os especialistas

	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70
E1	0,33	0,42	1	0,15	0,8	0,72	1	1	0,507	1
E2	0,33	1	1	1	0,8	0,72	1	1	0,57	0,54
E3	0,33	0,42	1	0,15	1	1	1	1	0,57	1
E4	1	1	1	1	1	1	1	0,5	0,71	1
E5	1	1	1	1	1	1	1	0,17	0,71	0,54
E6	1	1	1	0,08	0,8	0,72	1	1	0,71	1
E7	0,33	0,42	1	0,23	1	1	1	1	0,57	1
E8	1	1	1	1	0,1	0,72	1	1	0,14	1
E9	1	1	0,2	1	1	0,72	1	0,5	1	1
E10	0,33	1	0,2	1	1	1	1	1	0,507	1
E11	1	0,42	0,2	1	0,8	0,09	1	1	1	0,54
E12	0,07	1	1	0	0,1	0,09	0,1	0,08	0,71	0,54
E13	1	1	1	0,23	1	1	1	0,5	0,57	1
E14	1	0,42	0,2	1	1	0,72	1	1	1	0,08
E15	1	0,17	0,2	1	0,1	1	1	1	1	0,54
E16	1	1	1	1	0,8	1	1	0,5	1	1
E17	1	0,08	1	0,08	0,8	0,72	1	1	0,71	1
E18	1	1	1	0,23	1	1	1	0,5	1	1
E19	1	1	1	1	0,8	0,72	1	0,17	0,57	1
E20	1	0,08	0,2	1	0,8	1	1	0,5	1	0,54
E21	1	0,17	1	1	1	1	1	1	0,507	0,54

	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80
E1	1	0,78	1	0,67	1	1	0,5	0,2	1	0,42
E2	1	0,78	1	0,67	0,43	0,67	0,13	0,2	1	1
E3	1	1	0,43	1	1	1	0,13	0	0,19	0,33
E4	0,67	0,78	0,43	1	1	1	1	0,2	1	1
E5	0,67	1	0,43	0,67	0,43	0,67	0,5	0,07	1	1
E6	0,67	1	0,43	1	1	0,67	0,88	0,07	1	0,33
E7	0,67	0,78	1	1	0,07	1	0,5	0,2	1	0,42
E8	0,67	0,57	0,43	0,67	1	0,67	0,88	0,2	0,13	1
E9	1	0,78	1	1	0,43	1	1	0,2	1	0,42
E10	0,67	0,78	0,07	1	0,43	1	1	0,07	0,13	0,33
E11	1	0,78	1	0,67	1	1	0,88	0	1	1
E12	1	0,57	1	0,67	1	0,67	1	0,2	0,19	1
E13	1	1	1	1	0,43	1	0,88	0,07	1	0,42
E14	1	1	0,43	1	1	0,67	1	0,2	1	1
E15	1	1	1	0,67	1	1	0,5	0,07	1	1
E16	1	1	1	0,67	1	0,08	0,88	0,07	1	1
E17	0,67	0,57	1	0,67	1	0,67	0,88	0,07	1	0,42
E18	0,67	1	1	1	1	1	1	0,2	1	1
E19	0,08	1	1	0,67	1	1	1	0,07	0,19	0,33
E20	1	0,57	1	0,67	1	0,67	0,88	0,07	1	1
E21	1	0,11	1	0,67	0,43	1	1	0,07	1	1

Tabela 14. Escores individuais e totais para os especialistas

	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92
E1	1	0,31	1	1	0,75	1	1	0,88	1	0,73	0,46	0,88
E2	0,2	1	1	0,27	1	0,8	1	1	0,8	1	1	1
E3	0,2	1	1	1	0,75	0,8	1	0,88	0,3	1	1	1
E4	0,2	1	1	1	1	0,8	0,9	1	1	0,18	0,46	0,38
E5	0,2	1	1	0,54	1	1	0,9	1	0,8	1	1	1
E6	0,2	1	1	1	1	1	0,9	0,88	0,8	0,18	1	0,38
E7	0,2	1	0,11	0,27	0,75	1	0,2	0,88	0,8	0,73	1	0,38
E8	0,2	0,31	1	0,09	0,75	0,2	1	0,38	1	1	1	0,38
E9	1	1	1	1	1	1	0,9	1	0,8	1	1	0,88
E10	0,2	1	0,11	1	1	0,1	1	1	0,8	1	1	0,88
E11	0,2	1	1	0,54	1	0,2	1	0,25	1	0,73	1	0,88
E12	0,06	1	0,06	0,54	1	1	1	0,38	1	0,73	0,46	1
E13	0,2	1	1	0,54	0,75	0,8	1	0,38	0,3	1	0,08	0,88
E14	0,2	0,31	1	1	1	1	0,2	0,25	1	1	0,46	0,88
E15	0,2	1	1	0,27	1	1	1	0,88	0,8	1	1	0,88
E16	0,2	0,31	1	0,54	1	1	0,9	1	1	1	1	1
E17	0,2	1	1	1	0,75	0,8	0,9	1	1	0,73	1	1
E18	0,2	0,31	1	1	0,75	0,8	0,9	1	0,8	1	0,46	1
E19	0,2	1	1	1	0,75	0,8	0,9	0	0,3	0,73	1	0,38
E20	0,2	1	1	0,54	0,75	0,8	0,9	0,88	1	0,73	0,08	0,38
E21	0,2	1	1	1	1	1	1	0,88	1	0,73	0,46	1

	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100	Q101	Q102	Q103	Q104
E1	0,67	1	0,31	0,63	0,73	0,42	0,86	0,67	0,75	0,86	1	1
E2	1	0,67	1	0,63	1	0,17	0,86	1	0,38	1	1	1
E3	0,67	1	1	0,13	1	0,17	0,86	0,67	0,75	1	1	1
E4	1	1	0,31	1	1	1	1	1	0,38	1	0,5	1
E5	1	0,67	1	1	0,73	1	1	0,67	0,13	1	1	1
E6	0,08	0,56	1	0,5	0,09	1	0,71	0,67	0,38	1	1	1
E7	1	1	1	1	0,73	0,17	1	0,67	1	1	1	1
E8	0,67	0,67	0,31	0,5	0,73	1	0,29	1	1	0,86	1	0,25
E9	0,67	0,67	1	1	1	0,42	0,29	0,33	0,75	0,86	1	1
E10	0,67	0,56	0,31	0,63	1	1	1	1	1	1	1	1
E11	1	1	1	0,5	1	1	1	0,33	1	1	0,5	0,25
E12	0,67	1	0,31	0,5	1	0,42	0,14	0,33	1	1	0,5	0,25
E13	0,67	0,56	1	0,63	1	0,42	0,71	1	0,38	1	0,5	1
E14	1	0,56	0,31	0,38	0,73	1	0,86	0,67	1	1	1	1
E15	1	0,67	1	0,63	0,73	1	0,86	0,33	1	1	0,5	0,25
E16	1	1	1	0,38	1	1	0,71	0,33	0,75	1	1	1
E17	1	0,67	0,31	0,38	0,09	0,17	0,71	1	0,75	0,29	0,5	0,13
E18	0,67	1	0,31	1	1	1	1	0,33	0,38	0,86	1	1
E19	1	0,56	1	1	1	1	1	1	0,75	0,86	1	1
E20	1	0,11	1	1	0,73	1	0,86	1	0,38	0,29	0,5	1
E21	1	1	1	1	0,73	0,42	0,71	1	1	1	1	1

Tabela 14. Escores individuais e totais para os especialistas

	Escore Total	Aproveitamento (%)
E1	80,91	77,8
E2	83,95	80,72
E3	80,5	77,5
E4	83,89	80,7
E5	84,33	81,08
E6	78,73	75,7
E7	80,98	77,87
E8	73,24	70,42
E9	88,04	84,65
E10	80,42	77,33
E11	71,12	68,38
E12	70,27	67,57
E13	82,61	79,43
E14	78,92	75,88
E15	83,23	80,02
E16	87,56	84,19
E17	78,65	75,63
E18	92,09	88,55
E19	80,52	77,42
E20	80,86	77,55
E21	88,72	85,31

Tabela 15. Escores individuais e totais para os estudantes

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
A1	0,75	0,5	0,2	0,36	1	0,73	1	0,2	1	0,86	0,56	0,28
A2	0,75	1	0,2	0,36	1	1	1	0,2	1	0,86	0,56	0,28
A3	0,75	0,5	0,2	0,36	1	0,09	1	1	1	1	0,2	1
A4	0,75	0	0,2	0,55	1	1	0,1	1	0,5	0,86	0,56	0,07
A5	0,75	0,5	0,2	0,55	1	0,73	1	1	1	0,57	0,56	1
A6	0	0	0	0,55	0,07	0,73	0	1	0,5	0	0,56	0,14
A7	0	0	0,13	0,55	1	0,73	0,1	1	1	0,57	1	0
A8	0,75	0,5	1	0,36	1	0,73	1	1	1	0,86	1	1
A9	0,75	0,5	0,2	0,55	1	1	1	1	0,5	0,86	0,2	1
A10	0,75	0,5	1	0,55	1	0,73	1	0,9	0,5	1	0,56	0,28
A11	0,75	0,5	1	0,55	0,2	0,73	1	0,9	0,5	1	0,2	1
A12	0	0	1	0,55	1	0	0	0,9	1	0,86	1	0,14
A13	0,75	1	0,2	0	0,13	1	1	0,2	1	0,57	0,56	0,14
A14	0,75	0,5	0,2	0,55	1	1	1	1	0,5	0,86	0,56	1
A15	0,75	1	0,13	0,55	0	1	1	0,9	0,5	1	0,2	0
A16	0	0,5	0,2	0,55	1	1	1	0,9	0,5	0,57	1	0
A17	0	0,5	1	1	1	1	1	0,2	0,5	0,57	0,56	0
A18	0	1	0,2	0,55	0,07	1	1	1	1	0,86	0,56	1
A19	0,75	0	0,13	0,55	0,2	1	0,1	1	1	1	0,56	0
A20	0,75	0,5	1	0,36	0,2	0,73	1	1	0,5	0,86	0,2	0,07
A21	0	0,5	1	0,55	0,2	1	1	0,2	1	0,86	0,56	1
A22	0,75	1	0,2	0,55	0,2	0,73	0,1	0,9	1	0,86	0,2	0,14
A23	0,75	0,5	0,2	0,55	0,2	0,73	0	0,9	1	0,86	0,2	0,07
A24	0	0	0,2	0,55	1	1	1	1	1	0,57	0,56	0,14
A25	1	0,5	0	1	1	1	1	0,2	0,5	0,86	0,56	1
A26	1	0,5	0,2	0,55	1	0,73	1	1	0,5	0,57	1	0,14
A27	0	0	1	0,36	0,07	0,73	1	0,2	0,5	0,86	0,56	0,14
A28	0,75	0,5	0,2	1	0,13	1	1	1	0,5	0,86	0,2	1
A29	0	0	0,2	0,55	1	1	1	0,2	0,5	0,86	0,56	1
A30	1	0,5	0,2	0,55	0,07	1	1	0,2	1	0	0,2	1
A31	0,75	0,5	0,13	0,55	0,13	0,73	1	0,9	1	1	0,56	0,07
A32	0,75	0,5	1	1	1	0,73	1	1	0,5	0,86	0,2	1
A33	1	0	0,2	0,55	1	1	1	0,2	0,5	0,57	0,2	0,28
A34	0,75	0,5	0,13	0,55	0,07	1	1	1	1	0,86	0,2	1
A35	1	0,5	0,13	0,36	1	1	1	0,9	1	0,86	0,2	0,28
A36	1	1	0,13	0	1	1	1	1	0,09	1	0,2	0,14
A37	0,75	0,5	0,13	0,55	1	1	1	0,2	0,5	1	0,2	0,14
A38	0	0,5	1	0,55	1	0,73	1	1	0,5	0,86	0,56	0,14
A39	0,75	0,5	0,2	1	1	1	1	0,2	0,5	0,86	0,2	0,14
A40	0,75	0,5	0,2	0,55	1	1	0,1	0,9	1	0,86	0,2	0,14
A41	0,75	0,5	1	1	0,2	1	1	0,2	1	0,86	0,2	0,14

Tabela 15. Escores individuais e totais para os estudantes

	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24
A1	0,71	0,07	0,89	1	0	1	0,54	0,8	1	1	1	1
A2	1	1	0,89	1	0	1	0	0,8	1	0,08	1	0,18
A3	0,71	1	0,89	1	1	0,18	0,54	1	0	1	0,46	1
A4	0,71	0,07	1	1	1	0,18	0	0,2	0	1	0,08	1
A5	0,71	0,07	0,89	1	0	1	0,54	0	1	0,08	1	1
A6	1	0,07	1	1	0	0,18	0,54	0	0	0,08	0,08	0,06
A7	1	0,35	0,89	1	0	1	1	0	0	1	0,46	0,18
A8	0	0,35	0,89	0,28	1	0,18	0,54	1	1	1	1	0,06
A9	0,71	0,07	0,11	1	0	1	0,54	0	0,07	1	1	1
A10	0,71	1	1	1	1	1	0,54	1	0,28	1	1	1
A11	0,71	0,35	1	0,28	1	0,18	0,54	0,8	0,28	1	0,46	0,06
A12	0,71	0,07	1	0,21	1	1	0,54	0,8	0	1	0,08	1
A13	1	1	1	1	0	1	0	0,8	0	1	1	0,18
A14	0,71	0,07	0,89	1	1	1	0,54	1	1	1	1	0,18
A15	0,57	1	1	1	0	1	0,54	0,8	0	0,67	1	0,06
A16	0,71	0,07	0,33	0,21	1	1	1	1	1	0,08	0,08	1
A17	0,71	0,35	0,89	1	0	1	0	0,8	0	1	1	0,18
A18	0,71	1	1	0,28	1	0,18	0	1	1	1	1	1
A19	0,71	0,07	0,89	1	0,91	0,18	0,54	0,8	0,07	0,08	0,08	0,06
A20	0	1	0,89	1	1	1	0,54	0,2	1	1	0,08	0,18
A21	0,71	1	0,89	0,28	1	1	0,54	0	1	1	1	1
A22	0,71	1	0,89	1	1	1	0,54	1	1	1	1	1
A23	1	1	1	1	1	0,18	1	0,8	1	1	1	1
A24	0	1	0,89	1	1	1	0,54	0	1	1	1	0,18
A25	0,71	0,35	0,11	0	1	1	0,54	0,2	0,28	1	1	0,18
A26	0,71	1	1	0,21	1	1	0,54	0,2	0	1	1	1
A27	1	0,07	1	1	1	1	0,54	0,8	1	0,67	0	1
A28	0,71	1	0,89	1	1	0,18	0,54	0,8	1	0,67	1	1
A29	0,71	1	0,89	0,21	0,91	1	0,27	0,8	1	1	1	0,18
A30	0,71	0,07	0,89	1	0,91	1	0,54	0	0,28	0,08	1	0,18
A31	0,71	1	0,89	1	1	1	0,54	0,8	0	1	1	0,06
A32	0,71	0,07	0,89	0,28	1	0,18	0,54	1	1	1	0,08	0,18
A33	0,57	1	0,89	1	0,91	1	0,54	0,8	0,28	0,08	1	0,18
A34	0,71	1	0,11	1	1	1	0,54	1	0,28	1	1	1
A35	0,71	0,07	1	1	0,91	0,18	1	0,2	0,28	1	0,46	0,18
A36	0,57	0,07	0,33	0,28	1	0	1	0,2	0,28	1	1	0,18
A37	0,71	0,07	0,89	1	1	1	0,54	1	0,07	1	1	0,18
A38	0,71	0,35	0,11	1	1	1	0,54	0,2	0,07	0,08	1	0,06
A39	1	0,07	0,89	1	0,91	1	0,54	1	1	1	0,08	1
A40	1	0,35	0,11	1	0,91	1	0	1	1	0,08	1	1
A41	1	0,35	0,89	1	1	1	0,54	1	0,07	1	1	1

Tabela 15. Escores individuais e totais para os estudantes

	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36
A1	0,33	0,46	1	0,9	1	1	1	1	1	1	0,3	0,58
A2	0,33	0,15	0,8	0,9	0,33	1	0,89	0,18	1	0,5	0,3	0,17
A3	0,33	0,46	0,3	0,9	0,33	1	0,89	0,27	1	1	0,8	1
A4	1	1	1	0,9	0,33	1	1	0,27	1	1	0,3	1
A5	0,33	1	0,8	0,9	1	1	1	0,18	1	0,5	0,8	1
A6	0,33	1	0,3	1	0,33	0,33	1	0,18	1	0,5	0,3	1
A7	0,42	0,46	1	0	0,07	0,33	0,22	0,18	0	1	0,3	0,17
A8	0,42	1	0,8	0,9	0,33	0,33	0,22	0,27	0	0,5	0,3	0,17
A9	1	0,46	0,3	1	0,33	0,33	1	0,27	0,2	1	0,3	1
A10	1	0,46	0,3	1	1	1	1	0,18	0,3	0	0,8	1
A11	1	1	0,8	0,9	1	1	1	0,27	1	1	0,8	1
A12	0,42	0,46	1	1	1	0,33	1	0,18	1	0,5	0,8	0,58
A13	0,42	0,46	1	0,9	1	0,33	1	1	0,2	0,5	1	0
A14	0,33	1	0,8	1	1	0,33	1	1	1	1	0,8	1
A15	1	1	0,8	0,2	1	1	1	0,18	1	1	0,3	0,17
A16	1	0,46	0,3	1	1	0,33	0	0,27	1	1	1	1
A17	1	1	0,3	1	0,33	0,33	0,89	0,18	1	1	1	1
A18	0,33	0,46	1	0,9	0,33	1	1	0,27	1	1	1	1
A19	0,42	0,46	0,8	1	0,33	1	1	0,27	1	1	0,8	1
A20	1	1	1	0,9	0,33	1	0,22	0,18	1	1	0,3	1
A21	0,33	0,46	1	1	0,33	0,33	0,89	0,27	1	0	1	0,58
A22	1	1	0,8	0,9	1	0,33	1	0,27	1	0,5	0,3	1
A23	0,42	1	0,8	0,9	0,33	0,33	1	0,18	1	1	0,8	1
A24	1	0,46	0,3	0,9	1	1	1	0,18	1	0,5	0,8	0,58
A25	1	0,15	0,8	0,9	0,33	1	1	0,27	0,3	0,5	0	0,58
A26	0,42	0,46	0,3	0,2	0,33	1	0,89	0,18	1	0	0,3	1
A27	1	0,46	0,8	1	0,33	1	1	0,27	1	1	0,3	1
A28	1	0,46	1	1	1	1	1	0,18	1	1	0,8	1
A29	1	0,46	1	1	0,33	0,33	1	0,18	1	1	0,8	1
A30	1	1	0,3	0,2	0,33	1	0,22	0,18	1	1	0,8	0,17
A31	0,42	0,46	0,3	0,9	0,33	1	1	0,18	0,3	1	0,8	1
A32	0,42	1	1	1	0,33	1	1	0,27	1	1	0,8	0,17
A33	0,33	0,46	0	0,9	1	1	1	0,27	1	0,5	0,8	0,58
A34	1	1	1	1	0,33	0	1	0,27	1	1	0,8	1
A35	0,42	1	0,8	1	0,33	0,33	0,22	1	1	1	0,3	1
A36	0,42	0,15	1	0,2	0	0	1	0,18	0,2	0,5	0,8	1
A37	0,42	1	1	0,9	0,33	0,07	1	0,27	0,3	0,5	0,8	0,17
A38	0,42	0,46	0,8	0,9	0,33	0,33	1	0,18	1	0,5	0,8	0,58
A39	0,42	0,46	1	0,9	0,33	0,33	1	0,18	0,62	0,5	1	0,58
A40	0,33	0,46	1	0,9	1	0,33	1	0,18	0,2	0,5	0,3	0,17
A41	1	0,15	1	0,9	1	0,33	0,89	0,27	1	0,5	0,8	0,17

Tabela 15. Escores individuais e totais para os estudantes

	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48
A1	1	0,33	1	1	0,83	1	1	0,54	1	1	1	1
A2	0,1	0,33	0,89	0,23	0,83	0,07	1	0	0	0,06	0,54	0,75
A3	1	0,33	0,89	0	1	0,2	0,06	1	1	1	0,54	0,75
A4	0,8	1	0,89	0,23	1	1	0,06	1	1	0,06	0,08	0,75
A5	0,8	0,42	1	1	0,67	0,07	1	0,54	0	1	1	0,75
A6	0	0	0,89	0,15	1	0,07	0	0,08	1	1	0	0,75
A7	1	0,33	0,89	0	0,83	0,07	1	0	0	1	0	0,75
A8	0,2	1	1	0,23	0,67	0,07	0	1	1	0,06	0	0,75
A9	0,8	1	0,22	0,23	0,83	1	1	0	0	0,25	0,54	0,88
A10	1	0,42	0,22	1	1	0,07	1	1	0	1	1	0,75
A11	1	1	1	0,23	1	1	1	0,54	1	1	1	0
A12	1	0,33	0,89	0,15	0,83	0,07	1	1	0	1	0,54	0
A13	1	0	1	1	0,67	1	1	1	0	1	1	0,75
A14	1	1	0,89	1	0,67	1	0	1	1	1	1	1
A15	0,8	0	0,22	1	0,67	1	1	1	1	1	0,08	0
A16	0,8	0,33	1	0,23	0,83	1	1	1	1	0,25	0,54	0,75
A17	1	0,33	0,89	1	1	0,07	1	0,08	0	0,25	1	1
A18	1	0,33	1	1	0,83	1	1	0,08	1	0,25	0,54	1
A19	0,2	0	1	1	0,83	0,07	1	0,08	0	1	0,54	0
A20	0,2	1	0,22	0	1	1	1	0,08	0	1	1	0,88
A21	0,8	0	0	0,23	1	1	0	0,08	1	1	1	0,75
A22	1	0	0,89	1	1	1	1	1	0,21	0,25	0	1
A23	1	1	0,89	0	1	1	1	1	1	1	1	0,75
A24	0,8	0,33	0,89	1	1	0,2	1	1	1	0,25	1	1
A25	0,8	0,42	0,89	0,15	1	1	0	1	1	1	0,08	0,75
A26	0,8	0,33	0,22	1	1	0,07	1	1	0	1	1	0
A27	0,2	0	0	0,23	0,83	0	0,12	0,08	0,29	0,06	1	0
A28	0,8	0,33	1	1	0,83	0,07	1	0,54	0,29	0,25	1	0,75
A29	0,8	1	0,89	0	1	1	1	1	0,29	0,06	0,54	0,75
A30	1	0,33	0,89	0,23	1	1	0	1	1	1	0,54	0,88
A31	1	0,42	1	1	1	1	1	1	0	1	0	0,75
A32	0,8	1	0,89	1	0,83	1	0	0	1	1	0	0,75
A33	0,8	0	0,22	1	1	0,07	0	0,08	0,29	0,25	0	0,88
A34	0,8	1	1	1	1	1	1	0,54	0,29	1	0,54	0,75
A35	0,8	1	0,89	0,15	1	1	1	1	1	1	0,54	1
A36	0,1	1	0,22	1	0,67	1	1	1	0,29	0,06	0,54	0,75
A37	1	1	0,89	1	1	1	0,06	0,54	1	1	1	1
A38	1	0,42	1	1	1	1	1	1	1	1	0,54	0
A39	0,2	0	0,89	1	0,83	0,07	1	1	0,21	1	0	0,75
A40	1	0,42	0,89	1	1	0,07	1	1	1	1	0,54	1
A41	1	0,42	0,89	0,23	0,83	1	1	0	0	0,25	1	0,88

Tabela 15. Escores individuais e totais para os estudantes

	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60
A1	0,2	1	0,14	1	1	1	1	0,62	1	0,43	0,91	0,44
A2	0,2	0	0,14	1	1	0,55	1	0	1	0,57	0,91	0,44
A3	0,9	0,72	0,14	1	1	0,55	0,5	1	0,36	0,57	0,91	0,44
A4	0,9	1	1	1	1	1	1	1	0,55	0,43	1	0,44
A5	0,2	1	1	1	0,12	0,55	1	1	1	0,43	0,91	1
A6	0,9	0	0,14	0	1	0,55	0,17	0,62	0,36	1	0,91	0,22
A7	0,9	1	0,21	1	1	0,55	0,5	0	0,55	1	0	1
A8	1	1	0,21	0,63	0,12	0,55	1	1	1	0,57	0,91	1
A9	0,9	1	0,14	1	0	0,55	1	0,62	0,36	0,71	1	0,67
A10	0,9	1	0,14	0,63	1	0,55	1	1	0,36	1	0,91	1
A11	0,9	1	0,14	0,63	1	1	1	1	0,36	0,57	0,91	1
A12	1	1	0,14	1	1	0,55	0,5	1	1	0,71	0,91	0,67
A13	0,9	1	0,14	1	0,12	0,55	0,5	0,62	0,55	1	0,91	0,67
A14	0,9	1	1	1	1	1	1	1	1	1	0,91	0,44
A15	0,9	0,72	0,21	1	0	0,55	0,5	1	1	0,57	0,91	0,67
A16	0,9	1	1	1	1	1	1	1	0,36	0,71	0,91	0,67
A17	0,9	1	1	1	1	1	0,5	0,62	1	1	0	1
A18	0,9	0,72	1	1	1	1	1	1	1	0,57	0,91	0,67
A19	0	0	1	0,63	1	1	1	1	1	0,57	0	0
A20	0,9	0,72	1	1	0,12	0,55	1	0,62	0,36	0,43	0,91	0,44
A21	0,2	1	0,14	1	0	1	0,5	1	1	1	0,91	1
A22	0,2	0,18	0,14	1	1	0,55	0,5	1	0,55	0,71	0,91	0,67
A23	0,9	1	1	1	1	1	1	1	0,55	0,71	1	1
A24	0,9	0,72	1	1	1	0,55	1	1	1	1	0,91	1
A25	0,9	1	1	1	0,12	1	1	1	1	0,71	0	1
A26	0,9	0,72	0,14	0	1	1	0,17	1	0,36	0,71	0,91	0,22
A27	0,9	1	1	0	1	1	0,08	1	1	0,57	0,91	0,67
A28	0,9	0,72	1	1	0,12	1	0,17	1	1	1	1	0
A29	0,9	0,72	0,14	0	0,12	1	1	1	0,36	0,57	0	0,44
A30	0,9	1	0,21	1	1	1	0,5	1	1	0,71	0	0,44
A31	0,9	1	1	0,63	1	0,55	0,5	1	1	1	0,91	0,44
A32	1	1	1	0,63	1	1	1	1	1	1	0,91	1
A33	0,9	1	1	1	0,12	1	0,08	0	1	0,43	0	1
A34	1	1	1	0,63	0,12	0,55	1	0,62	1	1	0,91	0,44
A35	0	1	0,21	1	1	0,55	0,5	0,62	0,55	0,71	0,91	0,44
A36	0	1	0,14	1	0,12	0,55	1	0,62	1	1	0,91	0,44
A37	0,2	1	0,14	0,63	1	0,55	1	1	1	1	0,91	0
A38	1	1	1	1	0,12	0,55	1	1	0,36	0,43	0	0
A39	1	0,18	0,14	0	1	1	1	0,62	1	0,43	0,91	1
A40	0,9	1	1	1	1	0,55	0,5	1	0,36	1	0,91	0,44
A41	1	0,72	0,21	1	0,12	1	0,5	1	1	0,71	0,91	0,44

Tabela 15. Escores individuais e totais para os estudantes

	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72
A1	1	1	0,2	0,23	1	1	1	0,17	1	1	1	1
A2	1	0,08	0,2	0,08	1	0,09	0,1	0,17	0,57	1	1	0,57
A3	1	1	0,2	0,23	1	0,09	1	0,5	0,57	0,08	1	0,78
A4	0,07	0,42	1	0,08	1	1	1	1	0,71	0,54	1	1
A5	1	1	0,2	1	1	1	1	1	1	1	1	1
A6	0,07	0,08	0	0,23	0,1	0,72	1	0,17	0,71	0,08	1	0,57
A7	1	0,08	1	0,08	1	0,72	0,1	0,17	0,507	0,08	1	0,78
A8	1	0,08	1	0,23	1	1	1	1	1	0,54	0	1
A9	1	1	1	0,08	0,8	1	1	0,17	0,507	0,54	1	0,11
A10	1	0,08	0,2	1	1	1	1	1	0,71	0,54	0,67	1
A11	1	1	0,2	0,23	1	0,09	0,1	1	1	0,54	1	1
A12	1	1	0,2	0,08	1	0,72	1	0,17	0,507	0,08	1	1
A13	1	1	1	0,23	0,1	0,72	1	1	0,57	0,54	1	0,11
A14	1	1	1	1	0,1	0,72	1	1	0,507	0,54	0,67	0,78
A15	1	0,08	0,2	1	0,1	1	1	1	1	1	1	1
A16	1	0,08	0,2	0,23	1	1	0	1	1	0,54	1	1
A17	0,07	0,42	0	1	1	1	1	1	1	0,54	0	1
A18	0,07	1	0,2	1	0,1	1	1	1	1	0,54	1	1
A19	1	0,08	0,2	0,23	0,1	0,09	0,1	1	0,71	0,08	0,67	0,57
A20	1	0,08	0	1	1	0,09	1	1	1	0,54	1	0,78
A21	0,33	0,42	0,2	1	1	1	1	1	1	0,08	1	1
A22	1	0,08	0,2	0,23	0,1	1	1	1	0,14	0,54	1	1
A23	1	1	0,2	1	0,1	1	1	0,17	0,507	0,54	1	0,78
A24	1	1	0,2	1	1	0,72	1	0	1	1	1	1
A25	0,33	0,17	0,2	1	0,8	0,09	0,1	1	1	0,54	1	0,57
A26	1	0,08	0,2	0,15	0,1	0,09	1	1	0,14	0,54	1	1
A27	1	0,08	0,2	0,23	1	0,09	1	1	0,71	0,08	0	1
A28	0	0,08	0	1	0,8	1	0,1	0,17	1	1	1	1
A29	1	0,08	0,2	0,15	0,8	0,72	1	1	0,507	0,54	1	1
A30	1	1	0,2	0,23	0,8	1	1	1	1	0,54	1	1
A31	0	1	1	1	1	1	1	1	1	1	0	0,57
A32	1	1	0,2	1	1	1	1	1	1	0,54	1	0,78
A33	1	0,17	0	1	0,1	0,09	1	1	0,14	0,54	1	1
A34	0,33	1	0,2	1	1	1	1	1	0,14	0,54	1	1
A35	1	0,08	0,2	0,23	1	1	0	0,5	1	0,54	0	1
A36	1	0,08	0,2	1	1	0,09	0,1	1	1	0,54	1	1
A37	1	1	0,2	1	0,8	0,09	1	0,17	1	0,08	1	1
A38	1	0,17	0,2	1	1	0,72	0	0	1	0,54	0,08	0,57
A39	1	1	0,2	1	1	0,72	0,1	0	1	1	1	1
A40	0,07	0,08	0,2	1	1	1	1	1	0,14	0,54	0	0,78
A41	1	1	0,2	1	0,8	1	1	1	1	0,54	1	0,78

Tabela 15. Escores individuais e totais para os estudantes

	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84
A1	1	1	0,43	0,08	1	0,2	1	1	0,2	1	1	1
A2	1	1	1	0,67	0,88	0,2	0,19	0,42	0,2	0	1	0,09
A3	0	1	0,43	0,67	1	0,2	1	1	0,2	1	1	1
A4	1	0	0,07	0,08	0,13	0,07	1	1	0,2	1	1	0,54
A5	1	1	0,43	1	1	0,07	1	1	0,2	1	1	0,27
A6	0	1	1	0,67	0,13	0,2	1	0,42	0,06	1	1	0,27
A7	0,43	1	1	0,67	0,88	0,07	0,19	0,33	0,2	1	1	1
A8	1	1	0	0,08	1	0,07	1	1	1	1	0,11	0,54
A9	0	1	0,43	0,67	1	0,2	1	0,42	0,06	1	1	1
A10	1	1	1	0,67	0,88	0,07	1	0	1	0	1	1
A11	1	1	1	0,67	1	0,07	1	1	1	1	0,06	0,27
A12	0,43	0,67	1	0,67	0,13	0,2	0,19	0,42	1	1	1	1
A13	1	0,67	1	1	1	0,07	1	1	0,2	0,31	1	1
A14	1	0,67	0,43	0,08	1	0,2	1	1	0,2	1	1	0,54
A15	1	1	1	0,67	0,13	0,07	1	0,42	0,06	0,31	1	0,54
A16	1	1	0,43	0,67	1	0,2	1	1	0,2	0,31	1	0,27
A17	0	0,67	1	0,67	0,88	0,07	0,19	1	0,06	1	1	0,27
A18	1	0,67	1	1	0,13	0,07	1	0,42	1	1	1	0,54
A19	0,43	0,67	1	0,67	0,13	0,2	0,19	0	0,2	0,31	0	0,27
A20	1	0,67	1	0,67	1	0,07	1	0,42	0,2	1	1	1
A21	0,43	1	0,43	0	1	0,07	1	0	0,2	0,31	1	1
A22	0,43	0,67	1	0,67	1	0,07	1	0,42	0,2	1	0,11	0,27
A23	0,07	0,67	0,43	0,67	1	0,07	1	0,33	0,2	1	1	1
A24	0,43	0,67	1	0,08	0,88	0,07	0,19	0,42	0,2	1	0,11	0,27
A25	1	1	0,43	0,08	0,13	0,07	1	0,42	0	0	0,11	0,09
A26	0,43	1	1	1	1	0,13	1	0,33	0,2	1	1	0,54
A27	0,07	0,67	1	0,67	0,88	0,07	1	1	0,2	0	1	0
A28	1	1	1	0,67	0,88	0,07	1	1	1	1	1	1
A29	0	1	0	0,08	0,13	0,07	0,19	1	1	0,31	0,11	1
A30	1	0,67	1	0,08	1	0,07	1	1	1	1	1	0,27
A31	1	0,67	1	0,08	0,13	0,07	0,19	0,42	0,2	0,31	0,06	1
A32	1	1	0,43	0,08	0,13	0,07	1	0,33	0,06	1	1	0,27
A33	0,43	1	0	0	1	0,07	1	1	0,06	0	1	0,27
A34	1	1	0	1	0,13	0,07	1	1	1	1	0,11	1
A35	1	1	0,43	0,08	0,13	0,2	0,13	0,42	0,2	0	1	0,54
A36	1	1	0	0,08	0,5	0	1	1	1	0,31	0,11	0,54
A37	1	1	0,43	0,67	1	0,2	1	0,42	1	1	1	1
A38	1	1	1	0	0,88	0,07	1	1	0,06	1	0,11	1
A39	1	1	0	0,08	0,5	0,07	0,19	1	1	0	1	0,54
A40	1	1	1	0,08	0,13	0,07	1	1	1	1	1	1
A41	0,07	0,67	0,43	1	0,13	0,07	1	0,42	0,2	1	1	0,27

Tabela 15. Escores individuais e totais para os estudantes

	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96
A1	1	1	1	0,25	0,3	0,73	1	0,88	1	0,67	0,31	1
A2	1	0,8	0,9	0,88	1	0,18	1	1	1	0,67	0,31	1
A3	1	1	1	0,88	1	0,73	1	0,88	1	0,56	0,31	1
A4	1	0,1	1	1	0,8	1	1	1	0,08	0,67	0	0,63
A5	1	0,8	0,2	0,88	1	0,73	1	0,88	1	0,67	1	0,13
A6	1	0,8	0,9	0,88	1	1	1	1	0	0,67	1	0,13
A7	1	0,8	0,9	1	1	1	1	0,88	1	0,56	1	1
A8	0	0,8	0,9	0,88	0,8	1	0,08	0,38	0,08	0,67	1	1
A9	1	0,8	0	1	0,8	0,73	1	1	1	1	0,31	0,13
A10	1	1	0,9	1	1	0,18	1	0,88	1	1	0,31	1
A11	1	0,8	0,9	0,88	0,8	0,73	0,46	0,88	0,67	0,56	1	0,63
A12	1	0,8	0,9	1	1	0,73	0,46	0,38	1	1	1	1
A13	1	1	1	1	0,3	0,73	1	1	1	1	1	0,13
A14	0,75	1	0,9	0,88	0,8	0,73	1	0,88	1	0,56	0,31	0,63
A15	1	0,8	0,9	0,38	1	0,18	1	0,88	1	0,56	1	1
A16	1	0,8	0,9	1	0,8	0,73	1	1	0,67	0,67	1	1
A17	0	0,2	1	0,88	1	1	0,08	0,38	1	1	1	1
A18	1	0,8	1	1	1	0,73	1	0,88	1	0,56	1	1
A19	1	1	0,2	0	1	0,73	0,08	1	1	0,67	1	1
A20	1	0,8	0,9	0,88	0,8	1	1	0,88	1	0,56	0,31	0,5
A21	1	0,8	1	1	0,8	1	1	0,88	0,67	1	0,31	0,13
A22	1	0,8	0,9	0,88	0,3	1	1	1	0,08	0,67	1	1
A23	0,75	0,8	0,9	1	0	0,73	0,46	0,88	0,67	1	1	1
A24	0	0,8	0,9	0,25	1	0,73	1	0,38	1	1	1	1
A25	1	1	0,9	1	0,3	1	1	1	0,67	1	1	1
A26	0	1	1	1	1	1	1	1	1	0,67	1	0,13
A27	0	0,2	0	0,25	1	0,18	0,46	0,88	0,67	0,56	0,31	1
A28	1	0,2	1	1	1	0,73	0,46	0,88	0,67	0,67	1	1
A29	0,75	1	0,9	1	1	1	1	1	0,67	0,67	1	0,63
A30	1	0,2	0,2	0,88	1	0,18	1	0,38	1	1	1	1
A31	1	0,2	1	1	1	1	1	0,88	1	0,67	1	0,63
A32	1	0,8	1	0,88	0,8	1	0,46	0,88	1	0,67	1	1
A33	1	0,8	0	0,38	0,8	0,73	1	1	0,67	1	1	1
A34	1	1	0,9	1	0,8	0,73	1	0,88	1	1	1	1
A35	0,75	0,8	0,9	1	0	1	1	1	1	0,67	0,31	1
A36	1	1	0	1	0,8	1	1	1	0,67	1	1	0,13
A37	1	0,8	1	1	0,8	0,73	1	1	1	1	1	1
A38	1	0,8	0,9	0,38	1	0,18	1	0,38	1	1	1	1
A39	1	0,8	1	0,88	0,8	0,73	1	0,88	1	0,67	0,31	1
A40	1	0,8	0,9	1	0,8	0,73	1	1	0,67	1	1	0,13
A41	0,75	1	0,9	0,38	0,8	1	0	0,88	0	0,56	1	0,63

Tabela 15. Escores individuais e totais para os estudantes

	Q97	Q98	Q99	Q100	Q101	Q102	Q103	Q104	Total	Aproveitamento (%)
A1	1	1	1	0,67	1	1	1	1	80,78	77,67
A2	0,09	1	0,29	0,33	0,13	0,86	0	1	60,27	57,95
A3	1	0,17	1	1	1	0	0	1	72,5	69,71
A4	1	0,42	1	0,33	0,75	1	0	0,25	68,76	66,12
A5	1	1	0,86	1	0,75	1	0	0,25	78,44	75,42
A6	1	1	0,86	0,67	0,13	0	0	1	51,16	49,19
A7	1	1	0,29	0,33	0,75	1	0,5	0,13	61,117	58,77
A8	0,73	0,17	0,29	0,33	0,38	1	0	1	65,85	63,32
A9	1	1	0,29	0	0,38	0	0,5	0,13	65,007	62,51
A10	1	1	0,29	0	0,13	1	0	0,25	77,05	74,09
A11	0,09	1	0,29	0,67	0,13	1	1	1	76,76	73,81
A12	0,73	0,17	1	0,67	0,13	1	0	0,25	68,107	65,49
A13	0,73	1	1	0,33	1	0	1	0,25	72,71	69,91
A14	0,73	1	1	1	0,13	1	1	1	84,117	80,88
A15	1	1	1	0,33	0,13	0	0,5	0,25	69,61	66,93
A16	1	0,17	1	1	0,75	1	1	1	75,96	73,04
A17	0,09	0,42	0,29	0,33	1	1	0	1	68,44	65,81
A18	1	0,17	1	1	1	1	0	1	81,38	78,25
A19	1	1	0,29	0,33	0,75	1	0,5	1	57,13	54,93
A20	0,09	0,17	1	0,67	1	1	1	1	72,5	69,71
A21	0,09	1	1	0,67	0,38	1	0,5	0,25	70,04	67,35
A22	1	1	0,86	0,33	0,13	1	0,5	1	72,21	69,43
A23	0,09	1	0,71	0,67	1	1	0	1	77,897	74,90
A24	1	1	0,29	0,33	1	0,86	0	0,25	73,93	71,09
A25	1	0	1	0,33	0,38	1	1	1	67,35	64,76
A26	1	1	1	1	1	0	0	0,13	67,34	64,75
A27	0,09	1	0,29	1	1	1	0	1	59,44	57,15
A28	1	1	0,29	0,33	0,38	1	0	1	77,52	74,54
A29	0,73	1	0,29	0,67	0,38	0	0	1	67,097	64,52
A30	0,73	1	0,86	0,67	1	1	0	1	72,17	69,39
A31	1	1	0,29	0,67	0,75	1	1	1	74,9	72,02
A32	1	1	1	0,67	1	1	1	1	80,32	77,23
A33	0,09	1	0,71	0,67	0,75	1	0	0,13	61,31	58,95
A34	1	1	1	1	0,75	1	0	0,25	81,12	78,00
A35	0,09	0,17	0,29	0,67	0,75	1	0	1	66,57	64,00
A36	1	1	1	0,67	0,38	0,86	1	0,13	64,52	62,04
A37	1	1	1	0,33	0,75	0,86	1	1	77,44	74,46
A38	0,73	1	1	0,67	0,13	0	0	1	67,25	64,66
A39	1	0	0,29	0,33	0,13	0,86	0	0,25	67,72	65,12
A40	1	0,17	0,29	0,67	0,13	1	1	0,25	72,23	69,45
A41	1	1	1	0,67	0,13	0,86	1	1	72,96	70,15

Tabela 16. Escore total dos especialistas após seleção de itens por consistência interna

	Escore após seleção de 61 itens	Aproveitamento (%)
E1	48,67	79,79
E2	53,25	87,30
E3	49,927	81,85
E4	50,37	82,57
E5	50,06	82,07
E6	48,15	78,93
E7	51,47	84,38
E8	44,13	72,34
E9	52,38	85,87
E10	49,7	81,48
E11	36,43	59,72
E12	40,17	65,85
E13	52,45	85,98
E14	44,41	72,80
E15	47,54	77,93
E16	52	85,25
E17	50,08	82,10
E18	56,65	92,87
E19	48,73	79,89
E20	50,12	82,16
E21	52,47	86,02

Tabela 17. Escore total dos estudantes após seleção de itens por consistência interna

	Escore após seleção de 61 itens	Aproveitamento (%)
A1	45,61	74,77
A2	27,44	44,98
A3	39,59	64,90
A4	38,33	62,84
A5	43,23	70,87
A6	22,82	37,41
A7	28,89	47,36
A8	40,21	65,92
A9	35,5	58,20
A10	41,57	68,15
A11	44,18	72,43
A12	37,75	61,89
A13	41,98	68,82
A14	50,7	83,11
A15	38,37	62,90
A16	43,08	70,62
A17	36,25	59,43
A18	48,28	79,15
A19	27,19	44,57
A20	42,93	70,38
A21	38,88	63,74
A22	40,19	65,89
A23	47,37	77,66
A24	40,83	66,93
A25	37,42	61,34
A26	33,46	54,85
A27	28,92	47,41
A28	43,56	71,41
A29	34,27	56,18
A30	37,49	61,46
A31	43,34	71,05
A32	47,3	77,54
A33	28,23	46,28
A34	48,76	79,93
A35	37,58	61,61
A36	32,88	53,90
A37	41,65	68,28
A38	38,1	62,46
A39	35,33	57,92
A40	41,7	68,36
A41	42,33	69,39

Tabela 18. Escore por situação para os especialistas

	Situações												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
E1	11,59	4,45	10,33	11,16	7,59	3,13	3,67	6,957	3,12	8,67	3,32	3,31	18,42
E2	10,95	4,19	10,59	11,32	9,37	4,24	4,52	5,99	3	8,07	4,67	3,66	19,4
E3	10,627	5	8,94	11,39	9,43	3,13	4,15	7	1,65	7,93	4,67	2,83	17,08
E4	11,61	3,47	9,49	11,34	8,58	5	5	6,09	4,2	8,08	3,15	5	20,43
E5	11,26	3,92	11,86	11,08	8,04	4,67	5	4,62	3,24	8,44	4,67	4,4	20,75
E6	11,22	3,19	9,99	12,92	6,24	4,48	3,6	6,81	2,95	7,96	3,02	2,97	16,9
E7	12,1	3,19	10,53	11,09	9,55	3,19	4,23	6,09	3,12	5,94	4,38	3,57	17,01
E8	11,85	3,45	9	10,5	7,11	3,56	3,82	5,48	2,88	5,93	3,03	3,52	15,36
E9	12,68	4,28	11,68	11,01	9,65	3,92	3,92	6,71	3,62	9,7	4,22	3,04	20,58
E10	12,75	3,97	11,41	8,85	8,28	3,71	4,2	5,457	2,53	7,21	3,42	4,63	17,79
E11	11,32	3,82	7,15	6,59	5,57	4,33	3,09	6,99	3,88	6,92	4,88	3,83	19,51
E12	11,46	4,45	9,48	9,06	7,78	2,77	1,29	5,57	3,06	6,77	3,44	2,39	15,66
E13	13,32	3,39	12,03	10,92	7,82	4,29	4,23	6,437	3,37	6,97	3,19	3,76	17,29
E14	11,47	4,34	7,61	11,17	9,02	3,2	3,92	6,51	3,87	6,96	3,21	3,64	17,68
E15	13,56	3,53	8,26	12,73	8,46	3,61	3,3	7,21	3,57	8,15	4,55	3,55	19,82
E16	13,73	4,8	10,43	11,69	7,22	4,57	4,8	7,17	3,03	7,95	5	3,42	19,4
E17	11,53	4,34	9,97	11,53	8,13	3,51	3,6	6,62	3,04	8,38	3,98	2,35	17,75
E18	13,98	4,11	12,08	13,21	9,43	4,91	4,23	7,17	4,2	7,76	3,44	4,33	19,73
E19	12,16	3,8	10,06	10,25	8,32	4,1	4,52	5,49	2,59	6,68	3,94	5	18,21
E20	12,43	4,91	10,1	12,25	6,63	3,51	4	6,28	3,62	7,8	2,57	4,59	18,58
E21	12,49	4,8	11,02	13,14	8,02	3,79	5	5,257	4,07	8,81	4,46	3,86	21,2

Tabela 19. Escore por situação para os estudantes

	Situações												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
A1	9,11	3,34	11,69	11,58	7,96	3,78	3,43	6,6	3,28	7,48	3,86	4,67	19,29
A2	11,1	2,8	7,84	5,02	5,64	3	1,47	6,31	2,36	6,05	3,98	2,71	15,1
A3	10,7	3,72	7,94	9,82	6,92	3,92	2,52	4,36	3,87	8,81	3,75	4,17	20,6
A4	8,37	2,38	9,58	9,42	9,2	2,36	4,08	5,32	2,28	7,64	2,75	3,38	16,05
A5	10,53	2,54	10,29	9,8	7,62	4,34	4,2	7,43	4,07	7,08	4,55	3,99	19,69
A6	5,62	1,72	5,69	5,99	4,49	2,28	2,05	4,53	2,42	7,91	3,67	3,66	17,66
A7	8,32	3	4,32	6,59	6,46	3,08	2,9	4,967	2,14	8,9	4,44	3,62	19,1
A8	11,44	3	7,33	6,2	7,26	3,56	4,23	5,54	3,15	7,03	2,21	2,52	14,91
A9	9,45	2,54	7,96	8,17	6,45	4,38	3,88	3,757	3,29	7,39	4,31	2,42	17,41
A10	11,48	4,54	9,52	9,51	7,33	3,99	4,2	6,92	2,62	8,08	4,19	3,29	18,18
A11	10,39	2,8	9,77	12,57	7,03	4,48	1,62	7,54	3,74	7,44	3,57	2,68	17,43
A12	8,23	3,55	8,47	8,69	7,19	4,29	3	4,857	1,61	9,43	3,84	3,57	18,45
A13	9,55	2,8	8,49	10,17	6,13	4,58	3,05	5,89	4,07	7,54	5	3,19	19,8
A14	10,59	4,54	10,64	12,36	9,9	4,35	3,82	5,597	3,28	7,8	3,75	4,36	19,19
A15	9,6	3,34	8,91	9,24	5,88	3,23	3,3	8	2,29	6,17	4,44	4,33	17,23
A16	8,33	4,21	7,52	10,98	9,01	3,37	2,43	6,97	3,87	7,01	4,34	4,17	19,39
A17	9,28	2,8	8,21	9,62	9,02	2,49	4	5,21	2,81	6,41	3,46	2,13	14,81
A18	10,95	2,46	10,29	11,03	9,62	3,22	3,3	7,21	2,62	9,07	4,44	4,17	20,3
A19	7,96	3,43	6,57	8,52	6,63	1,65	0,72	5,13	1,19	4,71	3,75	3,62	13,27
A20	9,06	3,74	8,89	8,8	7,15	2,86	3,09	6,99	3,16	8,58	3,75	2,43	17,92
A21	10,47	2,82	9,61	7,69	6,59	3,66	4,2	5,94	2,07	8,11	3,86	2,89	16,93
A22	9,23	4,54	11,3	9,15	6,12	3,37	2,53	5,78	3,16	6,46	3,75	4,19	17,56
A23	8,96	3,98	9,96	12,69	9,2	4,71	3,3	4,167	3,07	7,38	4,01	3,47	17,93
A24	8,91	3,54	10,02	10,35	9,17	4,91	3,92	6,1	1,64	5,26	4,38	3,62	14,9
A25	9,79	2,74	8,21	8,42	8,77	2,21	2,19	6,54	1,7	5,4	4,67	3,33	15,1
A26	10,9	2,95	7,78	8,72	5,29	2,92	1,54	6,11	3,46	7,74	4,67	4,13	20
A27	7,49	4,34	9,53	5,11	6,98	3,23	2,52	4,53	3,62	2,83	2,88	3,38	12,71
A28	10,74	3,52	11,31	9,91	7,66	2,08	2,9	7,17	3,62	8,93	3,68	3,62	19,85
A29	9,47	3,19	9,48	10,38	5,99	2,09	2,87	5,047	1,47	8,07	4,34	3,32	17,2
A30	8,39	3,45	6,77	9,96	8,49	3,15	3,23	7,21	3,15	6,73	4,38	4,26	18,52
A31	9,92	4,34	6,95	11,22	8,33	3,35	5	6,24	0,89	6,77	4,55	3,59	15,8
A32	11,21	3	9,28	9,49	9,38	4,91	4,2	6,75	1,61	7,81	4,01	4,67	18,1
A33	8,96	4,25	7,5	5,59	6,98	2,6	2,19	5,11	3,07	5,04	4,67	3,47	16,25
A34	9,88	4,54	9,88	11,97	7,67	3,68	4,2	5,68	3,2	8,54	4,88	5	21,62
A35	10,01	3,29	8,02	11,68	6,43	3,14	2,43	5,47	0,96	6,19	3,98	2,22	13,35
A36	8,53	2,48	5,61	9,18	6,18	3,43	2,39	6,54	2,58	6,76	4,67	3,8	17,81
A37	8,64	4,54	7,54	10,96	7,52	3,91	3,09	5,68	3,29	9,33	5	4,33	21,95
A38	9,01	3,74	6,63	11,84	7,03	1,6	2,92	5,19	2,95	6,43	4,38	4,4	18,16
A39	9,31	4,45	8,32	8,28	6,69	4,34	3,02	6	1,84	7,75	3,86	2,62	16,07
A40	8,66	3,91	8,48	9,89	8,31	2,5	4,2	5,46	2,28	9,23	4,67	2,26	18,44
A41	10,09	4,54	9,61	8,09	7,43	4,06	4	5,49	2,62	7,3	2,44	4,3	16,66

Apêndice 10 - Teste Aplicado aos Grupos Estudados

(os itens com consistência interna menor que +0,19 estão em itálico e negrito)

1 – Um senhor de 82 anos, analfabeto, residente em instituição de longa permanência há 1 ano, vinha com dificuldade para realizar as tarefas do cotidiano. Há 4 dias, desenvolveu quadro de agitação, agressividade e recusa alimentar. Apresenta história de evacuações dificultosas. Ao exame: estado geral preservado, desidratado, afebril, com pressão arterial de 140/70 mmHg, freqüência cardíaca de 82 bpm. Ausculta cardíaca e pulmonar normais. Apresenta leve dor à palpação de hipogástrio e edema de membros inferiores de +/4+.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Delirium	Mini-exame do estado mental realizado há 1 mês igual a 11 pontos em 30 possíveis	1 2 3 4 5
Demência	Atrofia cortical difusa em tomografia de crânio	1 2 3 4 5
Delirium	Perda do aparelho auditivo há 1 semana	1 2 3 4 5
Demência	Alteração do ciclo sono-vigília	1 2 3 4 5
Delirium	*Suspensão de diazepan há 1 semana*	*1 2 3 4 5*

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

Se você está considerando solicitar	**E você encontra em sua avaliação**	**A investigação se torna**
Tomografia de crânio	*Episódios de alucinações visuais e delírios*	*1 2 3 4 5*
Exame de Urina tipo I	Bexiga palpável	1 2 3 4 5
Eletrocardiograma	Freqüência respiratória de 30 incursões por minuto	1 2 3 4 5
Radiografia de Abdome	Relato de que o paciente não evacua há 2 dias e toque retal sem fezes na ampola	1 2 3 4 5
Líquido Cefalorraquidiano	*Hemograma com leucócitos totais de 20000/mm3 e 80% de neutrófilos*	*1 2 3 4 5*

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

Se você está considerando prescrever	**E você encontra em sua avaliação**	**A prescrição se torna**
Lavagem intestinal	Fezes aquosas e com muco em pequena quantidade na fralda	1 2 3 4 5
Anti-psicótico	Flutuação do nível de consciência	1 2 3 4 5
Sondagem vesical de demora	*Diurese por sondagem de alívio com volume de 200 ml*	*1 2 3 4 5*
Anti-psicótico	Tremor de repouso e lentificação dos movimentos em membro superior esquerdo	1 2 3 4 5
Sondagem naso-enteral	*Antecedente de rinossinusite recorrente*	*1 2 3 4 5*

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

2 - Um senhor de 89 anos, de origem japonesa, vive sozinho. Conta que, no último ano, vem apresentando perda de apetite e dificuldade para a mastigação de alimentos sólidos. Refere também que não passa bem quando ingere leite ou seus derivados e que não têm o hábito de ingerir frituras. Evacua com dificuldade.

Há seis meses, começou a sentir sensação de fraqueza e indisposição. Às vezes sente dores difusas pelo corpo e, eventualmente, formigamento pelos membros. Queixa-se, ainda, de que se esquece do local onde colocou objetos em casa, quase todos os dias, nos últimos 3 meses.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Síndrome Depressiva	*Níveis séricos de testosterona abaixo da normalidade*	*1 2 3 4 5*
Deficiência de Vitamina B12	Hemograma com Volume Corpuscular Médio (VCM) normal	1 2 3 4 5
Hipotireoidismo	*Colesterol total de 380 mg/dl*	*1 2 3 4 5*
Hiperparatireoidismo Secundário	Osteopenia em densitometria óssea de coluna lombar	1 2 3 4 5
Síndrome Depressiva	Escala de Depressão Geriátrica (GDS) de 15 pontos com pontuação igual a 1 ponto	1 2 3 4 5

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

3 - Paciente de 85 anos, diabético em uso de insulina, é encontrado caído no banheiro, referindo dificuldade para se levantar. Mantém-se consciente e contactuante. Sua filha liga para o serviço de resgate, que o leva ao pronto-socorro. Ele então começa ficar sonolento. Apresenta pressão arterial de 230/110 mmHg e pulso de 100 btm, rítmico.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Acidente Vascular Cerebral	Eletrocardiograma realizado na ambulância, há 15 minutos, em ritmo de fibrilação atrial. Monitor cardíaco mostrando ritmo sinusal	1 2 3 4 5
Encefalopatia Hipertensiva	Temperatura axilar de 38,3 º C	1 2 3 4 5
Delirium	*Glicemia capilar de 400 mg/dl*	*1 2 3 4 5*
Acidente Vascular Cerebral	Diferença de pressão arterial sistólica de 10 mmHg entre os braços direito e esquerdo	1 2 3 4 5

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

Se você está considerando solicitar	**E você encontra**	**A investigação se torna**
Tomografia de crânio	Hiporreflexia pupilar em olho direito e antecedente de catarata	1 2 3 4 5
Ecocardiograma transtorácico	Sopro sistólico rude em Foco Aórtico	1 2 3 4 5
Hemograma	Hematoma em região de quadril direito	1 2 3 4 5
Urocultura	Próstata estimada em 45 gramas ao toque retal	1 2 3 4 5

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

Se você está considerando prescrever	**E você encontra**	**A prescrição se torna**
Antibiótico	Leucocitúria em exame de urina	1 2 3 4 5
Diurético	*História compatível com hiperplasia benigna prostática*	*1 2 3 4 5*
Omeprazol	*Sangue vivo ao toque retal*	*1 2 3 4 5*
Hipotensor endovenoso em infusão contínua	Antecedente de acidente vascular cerebral e exame recente demonstrando estenose de carótida esquerda de cerca de 70%	1 2 3 4 5
AAS (ácido acetil-salicílico)	*Antecedente de epigastralgia com o uso de AAS*	*1 2 3 4 5*

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

4 - Paciente de 75 anos, sexo feminino, professora aposentada, vem para avaliação em consultório com queixa de perda de memória há 2 anos. Esposo refere que a paciente esquece de tomar medicações para hipertensão e não lembra de dar recados.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Síndrome Depressiva	Paciente nega tristeza ou desânimo e mantém o interesse em suas atividades	1 2 3 4 5
Demência	Paresia de membro superior direito	1 2 3 4 5
Síndrome Depressiva	*O esposo refere que a paciente está com dificuldade para começar a dormir no horário de costume, mas que não dorme durante o dia.*	*1 2 3 4 5*
Demência	Perda de interesse para a realização de atividades da vida diária	1 2 3 4 5
Síndrome Depressiva	TSH de 12 mUI/ml e T4 livre normal	1 2 3 4 5

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

Se você está considerando solicitar	**E você encontra**	**A investigação se torna**
Avaliação neuropsicológica	Mini-exame do estado mental igual a 28 pontos em 30 possíveis	1 2 3 4 5
Coleta de Líquido Cefalorraquidiano	Aumento dos espaços liquóricos e dos ventrículos laterais em tomografia de crânio, proporcional à atrofia cortical difusa presente	1 2 3 4 5
Ressonância Magnética de Encéfalo	*Paciente com antecedente de cirurgia para fratura de quadril*	*1 2 3 4 5*
Avaliação neuropsicológica	GDS (Escala de Depressão Geriátrica) de 15 itens com 7 itens positivos	1 2 3 4 5

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

Se você está considerando prescrever	**E você encontra**	**A prescrição se torna**
Antidepressivo	Irritabilidade ao conversar com a paciente	1 2 3 4 5
Inibidor da Acetil-Colinesterase	Lesão hipodensa em região parietal esquerda em tomografia de crânio	1 2 3 4 5
AAS (ácido acetil-salicílico)	Microangiopatia periventricular em tomografia de crânio	1 2 3 4 5
Terapia Ocupacional	Melhora parcial do quadro com o uso de antidepressivo	1 2 3 4 5
Antidepressivo	História de períodos de euforia, agressividade e dificuldade para dormir	1 2 3 4 5

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

5 - Um homem de 72 anos procura atendimento médico por episódios de perda de consciência nos últimos 12 meses. Refere ter realizado tomografia de crânio, eletrocardiograma e Holter de 24 horas há 6 meses, com resultados dentro da normalidade.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Hipotensão Postural	Perda de consciência não precedida de outros sintomas	1 2 3 4 5
Estenose Aórtica	Queixa de tontura que melhora ao se deitar	1 2 3 4 5
Síncope Vasovagal	Hemoglobina de 9,5 g/dl	1 2 3 4 5
Hipotensão Postural	Episódios de perda de consciência que duram cerca de cinco minutos	1 2 3 4 5
Arritmia Cardíaca	Uso contínuo de propanolol	1 2 3 4 5

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

Se você está considerando solicitar	**E você encontra**	**A investigação se torna**
Eletroencefalograma	*Relato de que os episódios ocorrem quando está em pé*	*1 2 3 4 5*
Tilt Test	Os episódios geralmente ocorrem logo após sua caminhada matinal	1 2 3 4 5
Ecocardiograma Transtorácico	Mal estar e tontura antes dos episódios	1 2 3 4 5
MAPA (monitorização ambulatorial de pressão arterial)	Varizes calibrosas em membros inferiores	1 2 3 4 5
Novo Holter de 24 horas	*Eletrocardiograma com extrassístoles ventriculares isoladas*	*1 2 3 4 5*

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

6 - Um idoso de 72 anos é admitido na sala de emergência. Há 2 meses, vem apresentando tontura às transferências da cama. Caiu para frente há duas horas, durante a noite, no seu caminho da cama ao banheiro. Foi erguido por familiares, que notaram pequena lesão contusa frontal e escoriações em joelhos. Encontra-se orientado e os sinais vitais estão normais. Faz uso de enalapril 10mg/dia e propanolol 80mg/dia, para hipertensão arterial sistêmica; amitriptilina 25mg/dia, para dor crônica em joelho direito; e doxasozina 2 mg/dia, para noctúria.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Ataque isquêmico transitório	*Eletrocardiograma demonstrando bloqueio atrioventricular total*	*1 2 3 4 5*
Instabilidade articular	Crepitações em joelhos	1 2 3 4 5
Hipotensão ortostática	Freqüência cardíaca de 64 btm em posição supina e após alguns minutos em pé	1 2 3 4 5
Doença de Parkinson	*Diminuição do balanço dos braços ao deambular*	*1 2 3 4 5*
Interação medicamentosa	Prescrição de medicamentos inalterada há mais de 1 ano	1 2 3 4 5

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

7 - Uma idosa de 79 anos de idade está internada em uma UTI, no pós-operatório imediato de artroplastia de quadril. Ela é tabagista de 40 anos-maço e apresenta dislipidemia tratada com dieta. Está discretamente desidratada e recebe analgesia com opióides por cateter epidural. Não se queixa de dor.

Passa então a apresentar leve distensão abdominal, acompanhada de náuseas. Duas horas se passam e ela começa a se queixar de dor de forte intensidade no sítio cirúrgico. Desenvolve vômitos e mal-estar. Está agora taquipnéica e descorada.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Fecaloma	Antecedente de retossigmoidectomia há 2 anos	1 2 3 4 5
Efeito adverso de medicamento	Pupilas com discreta midríase, hipo-rreagentes	1 2 3 4 5
Abdome agudo vascular	*Gasometria venosa com pH normal e bicarbonato elevado*	*1 2 3 4 5*
Efeito adverso de medicamento	Pressão Arterial de 90/60 mmHg	1 2 3 4 5
Fecaloma	Ausência de dor à palpação de abdome	1 2 3 4 5

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

8 - Uma senhora de 76 anos queixa-se de esquecimento, que a faz necessitar de ajuda do marido para não deixar a comida queimar e para encontrar objetos que guardou. Toma condução sozinha e não se perde. Faz uso de agenda para não se esquecer dos compromissos e começou a levar uma lista de compras por escrito ao supermercado. O marido afirma que o problema começou há 6 anos, após uma queda que levou à contusão na cabeça, mas sem desmaios.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Doença de Alzheimer	*Ressonância de encéfalo normal*	*1 2 3 4 5*
Hematoma Subdural Crônico	*Tomografia de crânio com aumento de espaço liquórico em região temporal direita*	*1 2 3 4 5*
Demência Vascular	Lentificação psicomotora	1 2 3 4 5
Doença de Alzheimer	*Relato de episódios de apatia e isolamento social nos últimos meses*	*1 2 3 4 5*
Demência com corpos de Lewy	*Exame neurológico normal*	*1 2 3 4 5*

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

Se você está pensando em solicitar	**E você encontra**	**A investigação se torna**
Ressonância de encéfalo	Tomografia de crânio mostrando microangiopatia periventricular	1 2 3 4 5
Hormônios tiroideanos	*Megaloblastose no hemograma*	*1 2 3 4 5*
Dosagem sérica de vitamina B12	*Antecedente de gastrectomia parcial*	*1 2 3 4 5*

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

9 - Uma senhora de 78 anos apresenta fibrilação atrial diagnosticada há 2 anos. Está em tratamento com warfarina desde então. Procura atendimento médico por queixa de sangue na urina há 2 dias. Além disso, apresenta os antecedentes de hipertensão arterial sistêmica, dislipidemia, tabagismo e osteoporose. Conta que parou de tomar o anticoagulante no início do sangramento, pois sabia que ele poderia causar hemorragias.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Neoplasia renal	Punho percussão dolorosa em região dorsal esquerda	1 2 3 4 5
Hematúria secundária ao uso de anticoagulante	*RNI = 1,5*	*1 2 3 4 5*
Neoplasia de bexiga	*Citologia em amostra de urina negativa para células neoplásicas*	*1 2 3 4 5*
Hematúria secundária ao uso de anticoagulante	Equimoses em membros superiores	1 2 3 4 5
Neoplasia renal	*Presença de dismorfismo eritrocitário em amostra de urina*	*1 2 3 4 5*

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

10 - Um homem de 82 anos é acompanhado há 1 ano por hipertensão arterial sistêmica. Vem à consulta com queixa de perda de memória, o que não interfere em suas atividades de vida diária. Está em uso de amitriptilina, 25mg ao dia, devido à queixa de desânimo e insônia há 3 meses, após falecimento da esposa.

Desde então, reside com a filha, que o deixa sozinho em casa durante o dia para trabalhar. Apresentou perda de 8 Kg nos últimos 6 meses. Demonstra lentificação nas transferências da cadeira para a maca. Aproveita para dizer que está com tosse sem expectoração há 2 meses. É tabagista 60 anos/maço. Há 2 anos, interrompeu o hábito de etilismo de 3 doses de destilados ao dia. Realizou uma cirurgia para úlcera péptica há 25 anos.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Neoplasia de pulmão	*Estertores crepitantes em base pulmonar direita*	*1 2 3 4 5*
Erro alimentar	Quedas freqüentes quando está sozinho em casa	1 2 3 4 5
Síndrome Depressiva	*Recusa em responder às questões do mini-exame do estado mental*	*1 2 3 4 5*
Anorexia secundária a medicamentos	Presença de aftas no exame bucal	1 2 3 4 5
Neoplasia gástrica	*Queixa de empachamento pós-prandial*	*1 2 3 4 5*

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

Se você está pensando em solicitar	**E você encontra**	**A investigação se torna**
Avaliação odontológica	*Perda de paladar durante as refeições*	*1 2 3 4 5*
Avaliação nutricional	IMC = 23 Kg/m2	1 2 3 4 5
Colonoscopia	*Uso de warfarina devido a tromboses venosas de repetição*	*1 2 3 4 5*
Endoscopia Digestiva Alta	*Pesquisa de sangue oculto nas fezes positiva*	*1 2 3 4 5*
Ultrassonografia de Rins e Vias Urinárias	*Exame de urina com 18 hemácias por campo, sem dismorfismo eritrocitário*	*1 2 3 4 5*

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

11 - Uma idosa de 72 anos apresenta tontura. Conta que já apresentou labirintite no passado, mas que agora, há cerca de 2 meses, está apresentando sensações diferentes. Parece que sua cabeça está flutuando, mas, às vezes, ela também roda. Apresenta zumbido no ouvido esquerdo há 1 ano e tem dificuldade para ouvir há 2 anos. Pela sensação de desequilíbrio que apresenta, restringiu muitas de suas atividades, evitando até mesmo sair para andar no quintal de sua casa.

Como antecedentes pessoais, ela apresenta hipertensão arterial sistêmica e diabetes mellitus tratados há 12 anos.

Se você está pensando em	**E você encontra**	**A hipótese se torna**
Doença de Ménière	*Uso de furosemida, 40mg/dia*	*1 2 3 4 5*
Vertigem paroxística posicional benigna	*Crises de tontura que duram cerca de 20 minutos*	*1 2 3 4 5*
Insuficiência vértebro-basilar	Queixa de visão turva eventualmente	1 2 3 4 5
Transtorno de Ansiedade	*Crises de tontura que duram cerca de 1 minuto*	*1 2 3 4 5*
Vertigem paroxística posicional benigna	*Períodos de náuseas acompanhando os momentos de vertigem*	*1 2 3 4 5*

1: praticamente descartada

2: menos provável

3: nem mais, nem menos provável

4: mais provável

5: praticamente certa

12 - Uma senhora de 92 anos está internada há 7 dias para o tratamento de pneumonia. Ela apresenta o antecedente de dois episódios de pneumonia nos últimos 6 meses. É portadora de doença de Alzheimer há 12 anos. Há um ano, foi avaliada por uma fonoaudióloga, que concluiu que a paciente deveria usar espessante para a deglutição de líquidos. Há 6 meses, a paciente não consegue mais deambular e, há 3 meses, é dependente para todas as atividades de vida diária.

Se você está pensando em indicar:	**E você encontra:**	**A conduta se torna:**
Reavaliação fonoaudiológica	*Paciente sonolenta desde o início da internação*	*1 2 3 4 5*
Videodeglutograma	*Ausência de tosse ou engasgos durante a refeição*	*1 2 3 4 5*
Sondagem nasoenteral	*Irritabilidade à manipulação e períodos de agitação psicomotora*	*1 2 3 4 5*
Dieta pastosa assistida	Avaliação nutricional indicando desnutrição com grande perda de massa muscular	1 2 3 4 5
Gastrostomia endoscópica	Familiares responsáveis pela paciente a favor de cuidados paliativos exclusivamente.	1 2 3 4 5

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

Continuação do caso anterior - Você passa visita médica pela manhã, no oitavo dia de internação, e observa que a paciente está dispnéica, com freqüência respiratória de 32 incursões por minuto. Pelos controles de enfermagem, a freqüência respiratória começou a se elevar há cerca de 10 horas. Apresenta-se torporosa, levemente desidratada, com pressão arterial de 90/60 mmHg, freqüência cardíaca de 92 btm e temperatura axilar de 35,4 º C.

Se você está considerando	**E você encontra:**	**A conduta se torna:**
Manter cuidados paliativos exclusivos	Família preocupada com o sofrimento da paciente e a favor de qualquer procedimento terapêutico disponível	1 2 3 4 5
Encaminhar a paciente para UTI e realizar qualquer medida de terapia intensiva para manter a vida da paciente	*Há 2 dias, por alterações de secreção pulmonar na avaliação fisioterápica, houve troca de antibióticos*	*1 2 3 4 5*
Manter cuidados paliativos exclusivos	Paciente mantendo-se há 4 dias sem resposta aos estímulos verbais e com resposta motora inespecífica ao estímulo doloroso	1 2 3 4 5
Encaminhar a paciente para UTI e realizar qualquer medida de terapia intensiva para manter a vida da paciente	Freqüência respiratória vai para 22 ipm após aspiração nasotraqueal realizada pelo fisioterapeuta	1 2 3 4 5

1: contra-indicada totalmente

2: possivelmente prejudicial

3: nem mais, nem menos útil

4: útil

5: absolutamente necessária

10. BIBLIOGRAFIA CONSULTADA

ASSOCIAÇÃO BRASILEIRA DE NORMAS E TÉCNICAS. **NBR 6023. Informação e documentação: referências – elaboração.** Rio de Janeiro: ABNT, ago. 2002.

CONDURÚ, M.T.; PEREIRA J.A.R. **Elaboração de Trabalhos Acadêmicos: normas, critérios e procedimentos. 3ªed. rev. ampl.** Belém: NUMA/UFPA; 2007, 264p.

ROTHER, E.T.; BRAGA M.E.R.. **Como elaborar sua tese: estrutura e referências. 2ª ed. rev. e ampl..** São Paulo: BC Gráfica e Editora Ltda; 2005, 122p.