

Roberto José Gonçalves autor

PLANEJAMENTO, ELABORAÇÃO E APLICAÇÃO DE
UM TESTE DE DESEMPENHO CLÍNICO, BASEADO
NO PEG (PERFORMANCE EVALUATION GUIDES),
PARA A AVALIAÇÃO QUALITATIVA DE ALUNOS DE
GRADUAÇÃO DA FOP - UNICAMP

Tese apresentada à Faculdade de
Odontologia de Piracicaba, da Uni-
versidade Estadual de Campinas,
para obtenção do título de Livre-
Docente no Departamento de Odon-
tologia Social - Área de Orientação
Profissional, Odontologia Legal e
Deontologia.

PIRACICABA — 1980

UNICAMP
BIBLIOTECA CENTRAL

"Senhor, que importa que eu esteja na cumieira da casa
ou em seus alicerces contanto que eu seja fiel, bem no
meu lugar, na tua construção".

MICHEL QUOIST , "Poemas para rezar".

A Marisa, pelo incentivo e compreensão,

a Maria Lucia, Adriana e Roberto

dedico, com amor, este trabalho.

H O M E N A G E N S

Ao Prof. Dr. Plínio Alves de Moraes, Magnífico Reitor da Universidade Estadual de Campinas.

Ao Prof. Dr. Antonio Carlos Neder, Diretor da Faculdade de Odontologia de Piracicaba, da Universidade Estadual de Campinas e, sobretudo, ex-colega de Turma e amigo.

Ao Prof. Dr. Luiz Valdrighi, Diretor Associado da Faculdade de Odontologia de Piracicaba, da Universidade Estadual de Campinas meu contemporâneo acadêmico e amigo, pelo apoio e incentivo.

Ao Prof. Dr. Andrés José Tumang, Chefe do Departamento de Odontologia Social da Faculdade de Odontologia de Piracicaba, da Universidade Estadual de Campinas pela acolhida sempre sincera e desinteressada amizade.

Ao Prof. Dr. José Merzel, Coordenador do Projeto PADES/FOP, orientador de minha Tese de Doutorado, pelo apoio na consecução de mais este trabalho.

A G R A D E C I M E N T O S

Ao Prof. Dr. Edy Franceschi Piedade, Professor Livre-Docente Responsável pela Área de Orientação Profissional do Departamento de Odontologia Social da FOP-UNICAMP, pela sugestão do tema, apoio imprescindível e amizade fraterna.

Aos colegas: Prof. Dr. Edy Franceschi Piedade, Prof. Ass. Dr. Sérgio Francisco Mazzonetto e CD Osvaldo Di Hipólito Jr., por terem participado, como examinadores, das avaliações efetuadas e, também, pela amizade, compreensão e companheirismo sempre demonstrados.

À Prof. Dra. Sonia Vieira, Titular da Área de Bioestatística do Departamento de Odontologia Social, colega inestimável, pela análise estatística dos resultados do presente trabalho e pelas valiosas sugestões apresentadas durante todo o transcorrer do mesmo.

À Prof. Ivani Aparecida Lombardo, pelas úteis sugestões no planejamento e elaboração do Instrumento de avaliação objeto desta pesquisa.

Ao Bel. Ronaldo Seichi Wada, pelo processamento dos dados obtidos neste trabalho.

À Sra. Maria Aparecida Simoni, pela cooperação constante durante a redação dessa Tese.

À Sra. Ivani Guidolim Gerolla, pela revisão das referências bibliográficas.

Ao Sr. Antonio José Leite Ferraz, pelo serviço de datilografia.

Ao Sr. Sebastião Rodrigues de Barros, pela impressão do trabalho.

A todos os colegas e funcionários das Clínicas Odontológicas e do Departamento de Odontologia Social da FOP-UNICAMP, pelo incentivo e compreensão sempre presentes.

Este trabalho contou com o auxílio financeiro
do Projeto PADES/FOP.

S U M Á R I O

	<u>Página</u>
1. INTRODUÇÃO	1
2. REVISTA DA LITERATURA	6
3. MATERIAL E MÉTODOS	21
3.1. FORMULÁRIOS "A" (CRITÉRIOS DE DESEMPENHO)	22
3.1.1. CRITÉRIOS DE DESEMPENHO	23
3.1.2. ESCORES	24
3.1.3. DEFINIÇÕES OU INDICADORES DE DESEMPENHO ..	25
3.1.4. EXCEÇÕES	27
3.1.5. RAZÕES OU FUNDAMENTOS DOS CRITÉRIOS	28
3.2. FORMULÁRIO "B" (FORMULÁRIO CLÍNICO)	28
3.2.1. COLUNA DE REGISTRO DOS ESCORES	29
3.2.2. COLUNA DE REGISTRO DAS CONSIDERAÇÕES	30
3.2.2.1. TEMPO PARA EXECUÇÃO DA FUNÇÃO CLÍNICA	30
3.2.2.2. DIFICULDADES PARA EXECUÇÃO DA FUNÇÃO CLÍNICA	31
3.2.2.3. AJUDA DO EXAMINADOR	32
3.2.2.4. ORIENTAÇÃO DO EXAMINADOR	32
3.3. FORMULÁRIO "C" (NOTAS FINAIS)	33
3.4. CONJUNTOS DE FOLHAS DE INSTRUÇÕES	35
3.5. IMPLEMENTAÇÃO DO INSTRUMENTO DE AVALIAÇÃO	35
3.6. APLICAÇÃO DO INSTRUMENTO DE AVALIAÇÃO	37
3.7. ADMINISTRAÇÃO DAS AVALIAÇÕES	39
4. RESULTADOS	41
5. DISCUSSÃO	56

6. CONCLUSÕES	71
7. REFERÊNCIAS BIBLIOGRÁFICAS	73
8. ANEXOS	76
ANEXO 1 - CRITÉRIOS DE DESEMPENHO (Formulário A ₁)	77
ANEXO 2 - CRITÉRIOS DE DESEMPENHO (Formulário A ₂)	78
ANEXO 3 - CRITÉRIOS DE DESEMPENHO (Formulário A ₃)	79
ANEXO 4 - FORMULÁRIO CLINICO (FORMULARIO B)	80
ANEXO 5 - NOTAS FINAIS (FORMULÁRIO C)	81
ANEXO 6 - INSTRUÇÕES AOS ALUNOS	82
ANEXO 7 - INSTRUÇÕES AOS EXAMINADORES	85

1. INTRODUÇÃO

O tema mais penetrante no sistema educacional brasileiro nos últimos anos tem circulado, possivelmente, em torno do desejo de se avaliar rigorosamente a qualidade de nossas atividades de ensino.

Como sói acontecer com a maioria das instituições congêneres do País, a Faculdade de Odontologia de Piracicaba da Universidade Estadual de Campinas, através de seu corpo docente e discente, também se incorpora a esse desejo.

A FOP - UNICAMP, entretanto, desde seus primórdios vem se ressentindo da falta de um eficiente sistema de avaliação, sobretudo de desempenho clínico de seus alunos, que fornecesse informações sobre a qualidade de seu processo de ensino - aprendizagem.

Essa lacuna era sentida por todas as Disciplinas, dos vários Departamentos, que, isoladamente, ministravam conhecimentos a seus alunos e proporcionavam o desenvolvimento de habilidades clínicas nos mesmos.

Em 15 de junho de 1972 criou-se a Disciplina-Programa de Clínica Integrada, hoje denominada Clínica Odontológica, com subdivisões em Clínicas Odontológicas I, II, III e IV. Ela foi criada através da Portaria 07/72 e de conformidade com o Parecer 840/70 do Conselho Federal de Educação.

Esta Disciplina tem caráter interdepartamental e nela o aluno aplica de forma integral os conhecimentos adquiridos e as habilidades desenvolvidas, prévia e isoladamente, nas diferentes áreas do curso de Odontologia.

Com o advento da Clínica Odontológica, o problema da avaliação recrudesceu. Já era difícil avaliar o ensino clínico parcelado de cada Disciplina componente do curso de Odontologia, malgrado o esforço de seus professores nesse sentido. A avaliação em Clínica Odontológica, então, tornou-se pouco objetiva; isto se explica porque nela cada aluno trata do paciente como um todo, executando nele os mais variados tipos de tratamento odontológico.

No início de 1975, a FOP-UNICAMP, por iniciativa

de seus professores, se lançou a reestruturação de seu currículo de graduação, em termos de objetivos comportamentais. O conteúdo programático desse currículo foi dividido em funções básicas, tarefas e passos de tarefas.

A Clínica Odontológica, ponto de partida dessa reforma curricular, formulou, então, seus objetivos comportamentais em consonância com os objetivos gerais da Faculdade. Dentre os objetivos formulados, a avaliação do aproveitamento dos alunos apareceu como fecho imprescindível do processo ensino-aprendizagem.

Tornou-se mais evidente a preocupação do corpo docente da FOP-UNICAMP em procurar métodos de avaliação, realmente eficientes, que fossem capazes de mensurar, objetivamente, o desempenho clínico de seus alunos.

Listou-se, então, os seguintes critérios qualitativos, segundo os quais os alunos eram e ainda são avaliados:

- a. apresentação do aluno
- b. organização do trabalho
- c. dedicação
- d. cumprimento do horário
- e. responsabilidade
- f. capacidade técnica
- g. relações humanas

Além de serem avaliados, qualitativamente, em seu desempenho clínico como um todo, através desses critérios, os alunos também são, ocasionalmente, submetidos a provas teóricas sobre a matéria ministrada pelas áreas que compõem a Clínica Odontológica.

Sentiu-se que tais critérios são pouco explícitos, além de subjetivos e, como consequência, pouco eficazes.

A falta de explicitação dos critérios se deve a ausência de definições claras e precisas do desempenho esperado do aluno; a subjetividade dos mesmos dificulta sua mensuração prática. Tais fatos tornam difícil, senão impossível, a aplicação desse sistema de avaliação, de maneira uniforme, por diferentes examinadores. Acrescente-se, como agravantes, a subjetividade, a diversificação de pontos de vista e os preconceitos dos vários professores que atuam como avaliadores. Isto difi-

culta a interpretação e aplicação avaliativa dos critérios enumerados.

Esses indicadores de desempenho e critérios subjetivos pouco compreensíveis aumentam a tensão entre os alunos ao serem avaliados; com efeito, os mesmos não sabem, precisamente, quando e porque estão cometendo erros. Ressalta à vista que este sistema de avaliação qualitativa carece de confiança plena.

Em 1977, PIEDADE ²⁵, em sua tese de Livre-Docência, enfatiza que além do componente qualitativo da avaliação, deve existir o quantitativo. Nessa tese afirma: "o processo de avaliação do aprendizado é realmente importante para melhorar o nível de instrução, e deixar assegurado um nível mínimo de treinamento, para indicar de maneira justa a competência do aluno".

Esse trabalho veio dar novas luzes ao problema da avaliação das atividades clínicas dos alunos da FOP-UNICAMP.

Esse autor utilizou dados extraídos das atividades clínicas dos alunos da Clínica Odontológica, os quais foram submetidos a análise estatística; concluiu que: "a programação das atividades clínicas dos alunos da FOP, nos anos de 1974, 1975 e 1976 não satisfaz, pois apresentou distribuição distorcida e demonstrou a necessidade de infra-estrutura administrativa para que se adotem meios mais rápidos para a correção dessas distorções". Propôs, ademais, um conjunto de indicadores que mostraram-se válidos quando aplicados na avaliação quantitativa do aluno. Propôs, finalmente, a distribuição equitativa de atividades clínicas semelhantes, para cada aluno da Clínica Odontológica.

Como consequência desse trabalho e sob a supervisão do referido autor, implantou-se, junto a Clínica Odontológica, um programa computadorizado através do qual, mensalmente, as atividades clínicas de cada aluno são quantificadas; posteriormente são distribuídas, percentualmente, no que se convencionou chamar "composição clínica" do aluno. Esta "composição clínica" expressa, em percentuais, a quantidade de cada tipo de atividade clínica que o aluno concluiu no período. Assim, os alunos, ao receberem os relatórios computadorizados, cientificam-se da quantidade e diversidade de atividades clínicas que produziu no período, bem como de sua "composição clínica". Paralelamente, um relatório semelhante, com os dados gerais da classe, é afixado, em

local bem visível, para que cada aluno compare suas atividades clínicas com a dos demais colegas. O conhecimento desses dados constituem um êmulo para que cada aluno, com a colaboração de seu professor supervisor, procure, para o próximo período, uniformizar e melhorar, quantitativamente, suas atividades clínicas.

No campo da avaliação qualitativa, diversos são os testes de desempenho clínico relatados na literatura especializada. Dentre eles, julgamos que o que mais se aplicaria às nossas condições de trabalho e a forma de ensino por objetivos comportamentais é o proposto por HUNTER ¹⁷ (PEG: Performance Evaluation Guides).

Duas razões fundamentais nos levaram a realizar o presente trabalho:

1. o atual estágio em que se encontra o processo de avaliação clínica qualitativa dos alunos de graduação da FOP-UNICAMP.

2. o reconhecimento de que o mesmo se reveste de importância capital na melhoria do sistema de ensino-aprendizagem desta casa de ensino.

O objetivo terminal do presente trabalho é elaborar um instrumento de avaliação qualitativa capaz de aferir o desempenho clínico dos alunos de graduação da FOP-UNICAMP; tal instrumento, baseado no PEG, deverá adaptar-se às atuais condições de ensino-aprendizagem dos mesmos.

Cremos, além disso, que esse trabalho admite os seguintes objetivos intermediários:

1. Aplicar esse instrumento clínico de avaliação, como amostra de trabalho, para medir o desempenho final de alunos do 7º semestre de graduação na execução de cavidades Classe I para amálgama.

2. Avaliar sua validade e consistência através da obtenção estatística de coeficientes de confiança entre pares de avaliadores, que possam expressar o grau de concordância ou calibração entre os mesmos.

3. Verificar a necessidade de treinamento dos avaliadores, para a aplicação desse instrumento de avaliação.

4. Testar a utilidade instrucional do mesmo.

5. Estabelecer, por extrapolação, a possibilidade de adaptação desse instrumento de avaliação para aferir o desempenho final das demais funções clínicas que compõem o elenco da Clínica Odontológica da FOP-UNICAMP.

2. REVISTA DA LITERATURA

Inicialmente descreveremos, resumidamente, alguns sistemas compreensíveis de avaliação clínica, existentes na literatura.

No sistema de notas FELKNER ⁹, a avaliação clínica deve ser entendida como um teste de amostra de trabalho. Ele prevê um julgamento controlado do desempenho dos examinandos sob condições que, como afirma EBEL ⁸, simulam a situação real de trabalho (nesse caso, a prática odontológica). Segundo AUSUBEL ⁴, os testes de amostra de trabalho proporcionam com maior propriedade a mensuração direta e válida da capacidade clínica dos examinandos que outros tipos de testes como, por exemplo, os de perguntas curtas ou discussão de exames clínicos.

Os problemas dos testes de amostra de trabalho, segundo FITZPATRICK & MORRISON ¹¹, são: ausência de critérios de desempenho bem definidos e sistemas de atribuição de escores e de notas deficientes. Esses dois problemas apontados tornam a avaliação pouco confiável e limitam sua validade. De acordo com MACHEN ¹⁹, confiabilidade ou coeficiente de confiança de um sistema ou instrumento de avaliação é a consistência com que ele avalia o que foi destinado a medir ou avaliar; validade de um sistema ou instrumento de avaliação é a verificação de que ele mede ou avalia aquilo para o qual foi planejado e construído.

Segundo PATRIDGE & MAST ²⁴, dois tipos de coeficientes de confiança podem ser determinados nos estudos de avaliação clínica: o coeficiente de confiança intra-avaliadores, definido como a consistência dos avaliadores consigo mesmos e o coeficiente de confiança inter-avaliadores, que é a consistência entre os avaliadores no julgamento dos mesmos critérios.

FITZPATRICK & MORRISON ¹¹, afirmam, ainda, que o treinamento dos avaliadores é considerado essencial para a avaliação de amostras de trabalho; esse treinamento pode diminuir a idiosincrasia e os preconceitos do avaliador, fazendo aumentar os coeficientes de confiança intra e inter-avaliadores.

ABOU-RASS ³ desenvolveu um sistema bem compreensível para a avaliação clínica em odontologia; define oito fatores críticos no desenvolvimento de um sistema de avaliação clínica

ca: 1) determinação da filosofia da avaliação e metas, 2) determinação de critérios e padrões de desempenho, 3) validação do conteúdo e dos padrões de desempenho, 4) desenvolvimento e planejamento de instrumentos para erros críticos e manuais de avaliação, 5) treinamento dos estudantes e examinadores, 6) validação e teste de campo dos instrumentos e manuais, 7) implementação da avaliação e 8) aplicação da avaliação.

Esse sistema foi desenvolvido e testado por BURRELL & RASMUSSEN ⁵, na avaliação pré-clínica e clínica de endodontia; não foram obtidos, entretanto, os coeficientes de confiança nessa pesquisa.

O teste de desempenho clínico (PEG) é um método, desenvolvido por HUNTER ¹⁷, destinado a avaliar, qualitativamente, atividades clínicas de alunos de graduação dos cursos de Odontologia. Ele consiste de oito componentes: 1. Objetivo da avaliação, 2. Sistema de atribuição de escores, 3. Pontos ou tarefas a avaliar, 4. Critérios de desempenho, 5. Escores, 6. Definições ou indicadores, 7. Exceções e 8. Razões ou fundamentos.

Procuraremos, agora, descrever como esses componentes foram abordados pelos diversos autores.

Quanto ao objetivo da avaliação, FELKNER ⁹, ABOU-RASS ³ e HUNTER ¹⁷ enfatizaram a importância de sua determinação antes de se planejar um método de avaliação, porquanto o objetivo afeta tanto a elaboração do sistema de registro dos escores quanto a dos critérios de desempenho. PATRIDGE & MAST ²⁴, corroborando com essas afirmativas, opinaram que objetivos de verificação, em avaliação clínica, necessitam somente de decisões que permitam validar desempenhos "satisfatórios" ou "insatisfatórios"; para objetivos de monitorar a qualidade de cuidados com o paciente, pode ser mais apropriado o sistema de atribuição de escores que identifique a ação a ser tomada, por exemplo, "aceitável" ou "necessita modificações".

Para MACKENZIE ²⁰, a avaliação com objetivos de aprendizado dos alunos deve possuir um sistema que permita diagnosticar a retro-informação oferecida, pelo mesmo, a esses estudantes.

HUNTER ¹⁷ cita, como exemplos, os seguintes objetivos da avaliação: diagnosticar os erros dos alunos, diagnosti

car os acertos dos alunos, avaliar a dificuldade do aprendizado, simular um exame e avaliar o desempenho clínico final.

Quanto aos sistemas de atribuição de escores, EBEL ⁸ sugere que são de dois tipos: global e analítico. A atribuição de escores global registra o desempenho de uma tarefa ou função clínica como um todo. Os componentes individuais, que identificam o desempenho parcelado da tarefa ou função clínica, não tem seus escores registrados pois, de preferência, é avaliado o desempenho total. Um exemplo de avaliação de desempenho global é o sistema de avaliação "olhada rápida e grau", descrito por VANEK ²⁹, o qual tem sido tradicional em educação odontológica. Ao contrário da anterior, a atribuição de escores analítica permite a avaliação, em separado, de cada uma das pequenas partes que compõem o desempenho total.

Ambos sistemas permitem uma avaliação confiável; os métodos globais de avaliação clínico-odontológica, entretanto, vem sendo gradualmente suplantados pelos analíticos. O sistema analítico fornece maior retro-informação aos estudantes e compele o avaliador a observar fatores específicos do desempenho dos mesmos.

No planejamento do sistema de atribuição de escores para uma avaliação clínica analítica há, segundo PATRIDGE & MAST ²⁴ três decisões a tomar: 1. o nível de análise, isto é, o número de dimensões a serem julgadas, 2. o número de pontos da escala a ser usada no registro de cada julgamento e 3. como combinar os escores individuais para obter um escore global ou grau.

Quanto ao número de dimensões a serem julgadas, esses autores afirmam que nada existe na literatura especializada. A esse respeito, HUNTER ¹⁷ afirma que os pontos ou tarefas a avaliar em um desempenho clínico dependem do nível de aprendizado do estudante objeto da avaliação. Assim, um estudante que se encontra no estágio inicial de seu desenvolvimento profissional pode requerer que sejam avaliados muito mais pontos de seu desempenho, do que os de um estudante mais adiantado.

Quanto ao número de pontos da escala a ser usada no registro de cada julgamento, dois estudos compararam a utilização de escalas de dois pontos com o uso de escalas de múltiplos pontos.

Assim, HOUPPT & KRESS ¹⁶ compararam o uso de uma escala de dois pontos ("correto" e "incorreto") com a utilização de uma escala de cinco pontos ("correto" e "incorreto", com três pontos indefinidos entre os mesmos) para a avaliação de preparos cavitários em oito diferentes critérios, tais como: "remoção de fissuras" e "retenção". Encontraram que os coeficientes de confiança intra e inter-avaliadores foram altos quando se usou a escala de dois pontos, para avaliar os oito critérios. Os coeficientes de confiança dos escores globais (computados pela soma dos escores dos oito critérios) não diferiram, significativamente, para as duas escalas. A acuracidade de mensuração, definida pela uniformidade entre avaliadores especializados, foi maior quando se utilizou a escala de dois pontos, em relação ao uso da escala de cinco pontos.

Já o estudo reportado por HINKELMAN & LONG ¹⁵ comparou o uso de uma escala de três pontos com a utilização de uma escala de dois pontos, na avaliação de cinco critérios do preparo cavitário para bloco metálico. Os escores, na escala de três pontos, foram definidos como: "não são necessários melhoramentos", "cl clinicamente aceitável", e "cl clinicamente inaceitável e incorrigível". Os preparos foram, independentes e previamente, avaliados com a utilização da escala de três pontos; a escala de dois pontos (sistema passável-falho), também usada para avaliar os preparos, foi obtida da escala de três pontos convertendo-se os escores "não são necessários melhoramentos" e "cl clinicamente aceitável" em "passável" e "cl clinicamente inaceitável e incorrigível" em "falho". Em duas avaliações, antes e depois do treinamento de uma semana, de 60 preparos (30 antes e 30 depois), realizadas por quatro avaliadores, os coeficientes de confiança inter-examinadores foram, respectivamente, de 8% e 10% maiores, utilizando-se a escala de dois pontos. Ressalve-se, entretanto, que ambas escalas de pontos conferiram altos coeficientes de confiança inter-avaliadores.

Para HUNTER ¹⁷, a escolha do sistema de atribuição de escores a ser utilizado em avaliação clínica depende do grau de harmonia (uniformidade ou calibração) que se espera entre os avaliadores e do número e tipo de informações que se deseja coletar. Assim, se a expectativa da avaliação é de grande uniformidade entre os avaliadores, deve-se usar escalas de poucos pontos como, por exemplo, a de dois pontos (passável-falho, aceitável-

vel-não aceitável, satisfatório-insatisfatório, etc). Já, se deseja-se grande quantidade de informações deve-se usar escalas de múltiplos pontos como, por exemplo, o sistema numeral (0 - 5, 0 - 10 , etc), embora saiba-se que a uniformidade inter-avaliadores será prejudicada.

Esse autor afirma, também, que a utilização do sistema de dois escores pode provocar a perda de informações importantes. Cita como exemplos: se o polimento realizado, pelo aluno, em uma restauração for ótimo e não apenas aceitável? Da mesma forma, se uma restauração precisa ser refeita e é julgada apenas como inaceitável, não se distinguirá de uma outra que, em bora inaceitável, não precisa ser reposta.

A validação empírica de um sistema com quatro escores, utilizado na avaliação clínica de restaurações, que demonstrou altos coeficientes de confiança foi reportada por RYGE & SNYDER ²⁶.

Embora nesse estudo não se tenha variado o número de pontos da escala utilizada os quatro escores usados foram, operacionalmente, bem definidos. Após seleção e treinamento dos avaliadores, esses autores testaram o sistema na avaliação de 991 restaurações dentais; obtiveram 92% de concordância inter-avaliadores e 89% de concordância intra-avaliadores (baseada no reexame de uma amostra de 10% das restaurações realizadas).

Outro aspecto a ser focalizado, considerando-se o número de pontos das escalas de escores, é a utilidade instrucional dessas escalas.

HOUPT & KRESS ¹⁶ e HINKELMAN & LONG ¹⁵ sugerem que embora a escala de dois pontos proporcione altos coeficientes de confiança sua utilização tem desvantagens instrucionais pois não permite a retro-informação ao aluno, no sentido de incrementar melhoramentos em seu desempenho.

Outra decisão a ser tomada no planejamento de um sistema de marcação de escores é combinar os escores individuais para a obtenção de um escore global e, se necessário, dar pesos diferentes aos escores individuais.

Entretanto, a atribuição de pesos aos escores tem sido utilizada, principalmente, em avaliações de habilidades cognitivas, segundo KILLIP & LEWIS Jr. ¹⁸, obtendo-se pouca impor-

tância no aumento do coeficiente de confiança ou validade das mesmas.

Já a atribuição de pesos aos critérios de desempenho, em avaliação clínico-odontológica, tem recebido pouca atenção dos pesquisadores.

Dois estudos foram efetivados, atribuindo-se pesos diferentes aos critérios de desempenho.

GAINES e colab.¹⁴ atribuíram pesos aos critérios, ao avaliarem o desempenho de estudantes durante a raspagem dental em manequim. O critério "apoio no arco correto", por exemplo, tinha um peso de, somente 0,5 (assinalado, arbitrariamente, pelos professores), enquanto que o critério "utilização adequada da face cortante" tinha um peso 4. Os critérios tiveram, então, seus escores assinalados através de uma escala de três pontos de finidos como: 3 (mais do que 90% do tempo de desempenho), 2 (mais do que 75% do tempo) e 1 (menos do que 75% do tempo). A avaliação sistemática, incluindo-se a atribuição de pesos, proporcionou uma melhoria significativa do sistema "olhada rápida e grau" utilizado, já que os coeficientes de confiança inter-avaliadores aumentaram de 3% para 65%. Entretanto, os autores não puderam determinar se esse aumento deveu-se ao uso de critérios de desempenho, ao treinamento, ao uso da escala de três pontos, à atribuição de pesos aos escores ou a uma interação dessas variáveis.

No método documentado por BURRELL & RASMUSSEN⁵, utilizou-se um sistema de atribuição de escores, com pesos, na avaliação de procedimentos endodônticos. Assim, foram atribuídos pontos dobrados àqueles passos que envolviam medida de comprimento, instrumentação e condensação; foram atribuídos pontos extras ao valor dos passos de selamento coronário e preenchimento do canal, multiplicando-o pelo fator 0,5.

Os autores concluíram que a atribuição de pesos a certos escores foi necessária, já que alguns passos do procedimento são mais importantes que outros.

Como exemplos de outros sistemas de atribuição de pesos, HOUPPT & KRESS¹⁶ os atribuíram, igualmente, a todas as dimensões, no cômputo do escore global; já RYGE & SNYDER²⁶ utilizavam o escore mais baixo, dos atribuídos a cada um dos critérios individuais, para determinar o escore global.

Quanto aos critérios de desempenho, PATRIDGE & MAST ²⁴ definiram-nos como sendo descrições do comportamento esperado com as quais o desempenho de cada estudante pode ser comparado.

Embora não trabalhando na área odontológica, ABELLES ¹ identificou quatro meios para o desenvolvimento dos critérios de desempenho: 1. examinando-se experiências anteriores na avaliação do comportamento, 2. analisando-se os livros de texto usados no ensino do comportamento, 3. fundamentando-se os itens de avaliação numa estrutura teórica pré-existente na área a ser avaliada e 4. obtendo-se uma amostra da análise do desempenho, feita por especialistas.

Poucos trabalhos têm sido reportados sobre a confirmação estatística dos critérios de desempenho.

SALVENDY e colab. ²⁷, utilizando dois examinadores, avaliaram Classe I para amálgama usando uma escala de três pontos ("inaceitável", "aceitável" e "excelente") para detectar cinco aspectos: qualidade geral, deficiências marginais, excessos marginais, forma anatômica e textura da superfície. A análise regressiva indicou que um dos critérios, forma anatômica, era responsável por 56% de variância na avaliação de qualidade geral.

Outro método geralmente aceito para a identificação de critérios de desempenho é a análise de tarefas. A análise de tarefas é o fracionamento de uma tarefa em suas partes, ou passos, descrevendo-se as relações entre essas partes.

O método de análise de tarefas descrito por MAST & SPEDDING ²¹, resultou na identificação de três tipos de critérios de desempenho: mensurável, sensorial e processual. ABOU-RASS ² também descreve o desenvolvimento de critérios de desempenho a partir da análise de tarefas de procedimentos endodônticos.

MACKENZIE ²⁶ listou seis linhas mestras para a identificação de critérios de desempenho: 1. validade, 2. interpretação consistente, 3. cobertura (integralidade), 4. precisão, 5. utilidade e 6. utilidade instrucional.

HUNTER ¹⁷ afirma que se os critérios de desempenho forem superiores a sete, torna-se difícil utilizá-los em ava

liação clínica; sugere a eliminação de critérios que: 1. o exame clínico não confirma sua importância, 2. não se relacionam claramente com o bem estar do paciente, 3. chocam-se com outros critérios (nesse caso eliminá-lo ou consolidá-lo), 4. se aplicam raramente e, conseqüentemente, não são importantes, 5. em alguns casos não se aplicam e 6. são difíceis de detectar.

Quanto ao teor dos critérios de desempenho, há dois estudos que comparam os efeitos do conteúdo subjetivo e objetivo dos mesmos sobre os coeficientes de confiança inter e intra-avaliadores.

Assim, HOUPPT & KRESS ¹⁶ compararam a acuracidade e os coeficientes de confiança intra e inter-examinadores da avaliação do desempenho clínico de restaurações realizadas por grupos que se utilizaram de uma escala de 5 pontos, para o julgamento de cada uma das oito dimensões; essas avaliações foram feitas sob duas condições: sem critérios de desempenho para cada um dos cinco pontos, por exemplo, "proteção pulpar" e com definição dos critérios de desempenho para cada um dos cinco pontos, por exemplo, "proteção pulpar - toda dentina exposta cobertta". A presença dos critérios de desempenho não aumentaram os coeficientes de confiança intra-avaliadores, antes ou depois do treinamento, tanto nos escores individuais quanto nos totais. Entretanto, os coeficientes de confiança inter-avaliadores, medidos nos critérios individuais, aumentaram significativamente com a utilização de critérios de desempenho. Houve somente um moderado ganho nos coeficientes de confiança obtidos dos escores totais.

GAINES e colab. ¹³ desenvolveram duas escalas para a avaliação de escultura em cera, cada uma com oito sub-escalas, por exemplo, "forma externa vestibular". A primeira escala tinha três critérios com teor subjetivo ("fora dos padrões", "médio" e "pobre") para cada sub-escala. A segunda escala utilizou três frases com conteúdo objetivo como, por exemplo, "completamente lisa", "algumas rugosidades" e "maiores entalhes e ranhuras". Sete avaliadores examinaram, independentemente, as esculturas realizadas por oito estudantes, escolhidos ao acaso, utilizando-se da escala subjetiva. Três meses mais tarde, os mesmos avaliadores examinaram diferentes esculturas efetuadas por oito estudantes, também escolhidos ao acaso, utilizando-se da escala objetiva.

Análises de variância, realizadas em separado, para as duas escalas demonstraram diferenças significantes entre os avaliadores quando se utilizou a escala subjetiva; o uso da escala objetiva, entretanto, não acarretou diferenças significantes entre os examinadores. Assim, os coeficientes de confiança inter-avaliadores aumentaram de 26%, com o uso dos critérios subjetivos, para 56% quando se utilizaram os critérios objetivos.

Vários outros estudos experimentais, como os de GAINES e colab.¹⁴, FERNANDEZ¹⁰ e NATKIN & GUILD²³, manipularam os critérios de desempenho enquanto, simultaneamente, introduziram outras variáveis impedindo, deste modo, conclusões acerca dos efeitos desses critérios.

Uma outra variável no teor dos critérios de desempenho é a utilização dos mesmos com conteúdo negativo. Uma das maneiras de se empregar essa variável é avaliar o desempenho dos alunos através de seus erros críticos.

A técnica do erro crítico consiste, segundo ABOU-RASS⁷, na identificação de possíveis erros no desempenho de uma habilidade, os quais podem, crítica ou significantemente, afetar o resultado desse desempenho.

NATKIN & GUILD²³ reportaram um sistema de atribuição de escores utilizando-se de erros críticos para avaliar desempenhos em endodontia.

Quatro tipos de erros foram descritos: leves, moderados, sérios e críticos. O registro de um erro, de uma dessas categorias, foi realizado com base no efeito que este erro poderia ter no prognóstico do caso clínico. Dois aspectos do sistema podem ser notados. Primeiramente, esse sistema de atribuição de escores limita a possibilidade dos avaliadores na variação do registro dos mesmos. Com efeito, o sistema especifica a possível categoria de erros para um dado critério. Assim, por exemplo, o avaliador não pode classificar a projeção do cone de prata por mais do que 2mm no interior da câmara pulpar como um erro moderado, sério ou crítico; este é um erro que somente pode ser classificado como leve. Em segundo lugar, os tipos de erros que um aluno pode cometer durante o desempenho de uma habilidade foram enumerados para beneficiar os avaliadores.

Esses escores (o número e tipo de erros cometidos) foram convertidos em letras que expressavam o grau ou nota. Assim, por exemplo, um "C" significava um erro sério e um ou mais erros moderados ou leves. Encontrou-se que, embora a variação média dos graus assinalados para os estudantes fosse pequena, este melhoramento não foi estatisticamente significativo. Houve, entretanto, um aumento estatisticamente significativo no percentual de vezes nos quais todos os graus assinalados estavam dentro de dois pontos, indicando alto coeficiente de confiança inter-avaliadores.

Um instrumento de avaliação reportado por ABOU-RASS ² tinha critérios redigidos negativamente. Sua razão para o uso da forma negativa foi que os critérios positivos, embora melhores, são difíceis de construir e freqüentemente são muito longos.

RYGE & SNYDER ²⁶ usaram critérios com redação positiva e negativa.

Em seu teste de desempenho clínico (PEG), HUNTER ¹⁷ acrescenta definições ou indicadores aos critérios de desempenho. Justifica tal atitude afirmando que boas definições dos critérios melhoram a uniformidade, o desempenho e o moral dos alunos; também fornecem retro-informações, aos mesmos, no sentido de incrementar progressos em seus desempenhos, além de ajudar os professores a melhorar a qualidade do ensino.

Continua, afirmando que as definições dos critérios de desempenho devem descrever um procedimento mensurável; essas definições operacionalizadas relatam desempenhos que permitem julgar quando as definições estão sendo bem cumpridas. Nessas definições operacionalizadas deve-se descrever a utilização, de preferência, de instrumentos e/ou materiais odontológicos comuns como meios de mensuração. Pode-se, também, utilizar diagramas, fotos ou modelos; a descrição de características visuais ou táteis pode ser usada quando nenhum meio instrumental é apropriado.

Quanto ao uso de definições positivas ou negativas, esse autor afirma que depende de alguns fatores; para alunos que estão iniciando o treinamento clínico, as definições positivas são melhores porque os mesmos mais erram do que acertam; nesse caso os acertos, que ocorrem menos, são mais facilmente avaliados. Para alunos mais avançados, entretanto, as defini-

ções negativas ou de erros são mais eficientes porque sucede o inverso; acrescenta, ainda, que as definições de erros são melhor trabalhadas.

Prossegue o autor, afirmando que a definição dos critérios de desempenho influe decisivamente no número de escores empregados em um sistema de avaliação. Assim, o número de escores a serem utilizados depende do número de definições que se enquadram em cada critério de desempenho; vale dizer que para cada definição necessitamos de um escore. Pelo mesmo raciocínio chega-se à conclusão que cada critério de desempenho admite tantas definições quantos sejam o número de escores utilizados.

Se utilizarmos, entretanto, um sistema de atribuição de escores de 2 pontos, por exemplo, "satisfatório" - "insatisfatório", poderemos descrever apenas uma definição por critério de desempenho; ou redigiremos uma definição para desempenho satisfatório ou escreveremos uma definição para desempenho insatisfatório. Se o desempenho do estudante não atende a definição proposta não registramos o respectivo escore.

Esse mesmo autor acrescenta, ainda, exceções e razões aos critérios de desempenho.

As exceções são descritas e utilizadas quando as definições dos critérios de desempenho não se aplicam a determinados casos clínicos incomuns. Assim, por exemplo, as cavidades preparadas podem ser mais profundas do que os parâmetros exigidos pela definição, se houver extremo desgaste oclusal do dente. Do mesmo modo, a cavidade pode não ter a forma exigida pela definição do desempenho, quando o ataque da cárie não o permitir.

As razões ou fundamentos de um critério de desempenho justificam a sua inclusão numa avaliação clínica. Elas podem enfatizar os princípios básicos que delimitam o critério.

Sem esses fundamentos básicos os alunos passam a ser meros artesãos. Assim, por exemplo, a "fratura do dente" (razão ou fundamento) pode ser causada pelo erro "esmalte sem suporte" (critério).

O mesmo HUNTER ¹⁷ sugere para a atribuição do escore global ou grau do aluno, além do cômputo dos escores dos

vários critérios individuais, a inclusão de considerações; assim grau = escores + considerações. As considerações são fatores anormais que afetam o desempenho do aluno como, por exemplo, ajuda docente, tempo excessivo para completar a tarefa, dificuldades locais que dificultam a execução da técnica, etc.

O autor justifica essa inclusão argumentando que se dois alunos executam procedimentos clínicos semelhantes e recebem idênticos escores, nos vários critérios de desempenho, nem sempre devem receber o mesmo grau; pode ocorrer que um deles tenha incorrido em uma ou mais considerações, isto é: demorou excessivamente para completar a tarefa e/ou solicitou ajuda docente para completá-la; nesse caso, esse aluno deve receber um grau menor. Do mesmo modo, o aluno que encontrou dificuldades locais, ao longo de seu desempenho, deve receber um grau maior. Exemplos dessas dificuldades locais seriam: excesso de saliva, dificuldade de visão do campo operatório, etc.

Finalmente, esse autor fornece um exemplo hipotético de graus = escores + considerações:

Superior (grau): todos os critérios com escore aceitável (escores) no primeiro atendimento (consideração). Louvor (grau): nenhum critério com escore insatisfatório (escores); desempenho completado dentro de 1 ou 2 horas se o caso for extremamente difícil (considerações). Mais que aceitável (grau): nenhum critério com escore incorrigível ou falho (escores) e nenhuma assistência ou orientação docentes (considerações). Aceitável (grau): nenhum critério com escore insatisfatório (escores) em nenhum dos três primeiros pacientes atendidos (considerações).

Uma vez estudados os sistemas de atribuição de escores e os critérios de desempenho, as próximas variáveis a serem consideradas são: a extensão e o tipo de treinamento recebidos pelos avaliadores na utilização dos sistemas de avaliação.

Muitos trabalhos que incluem o treinamento dos avaliadores, comparam esse treinamento com a falta do mesmo.

Assim, FULLER ¹² relatou o exame mais completo do efeito do treinamento sobre os coeficientes de confiança. Ele examinou quatro pares de avaliadores e suas consistências na avaliação de preparos de Classe II para amálgama e/ou preparos de

Classe III para outro. As três condições experimentais foram: 1. o tradicional sistema "olhada rápida e grau", 2. um modelo utilizando-se "checklist" e 3. um modelo utilizando-se "checklist" e treinamento. O programa de treinamento durou duas horas e consistia de discussões dos conceitos e técnicas ministrados aos estudantes, dos fatores humanos que afetam a avaliação do trabalho dos estudantes, além de explicações de técnicas de avaliação e prática em avaliação.

O modelo utilizando-se "ckecklist" e treinamento não provocou aumento significativo dos coeficientes de confiança inter-avaliadores, comparativamente ao modelo utilizando-se somente "ckecklist".

ABOU-RASS ² reportou o refinamento de uma escala de dois pontos de um instrumento de avaliação endodôntica, baseada em modelos previamente desenvolvidos. Antes do treinamento a uniformidade (calibração) entre quatro avaliadores, em 20 casos avaliados, variou de 32% (para condensação lateral) a 75% (para instrumentação). A uniformidade global (em todos os critérios) foi de 54%.

O treinamento foi descrito como seis seminários, realizados durante o período de oito semanas, nos quais se enfatizaram a definição, significado, medição e avaliação de cada passo do desempenho.

Os mesmos examinadores avaliaram 30 casos após esse treinamento. A uniformidade global aumentou para 77% e as individuais variaram de 65% (para condensação lateral) a 79% (para instrumentação). Durante um curso subsequente, no qual quatro examinadores avaliaram 378 casos, encontrou-se uma uniformidade global média de 79,83%. Tanto a uniformidade global quanto a uniformidade nos critérios individuais aumentaram como consequência do treinamento.

Do estudo de HOUPPT & KRESS ¹⁶ o treinamento influiu muito pouco. Esse treinamento foi descrito como a informação aos examinadores do modo correto de avaliar, imediatamente após haverem executado as avaliações. O treinamento ocasionou um leve aumento nos coeficientes de confiança intra-avaliadores, nas avaliações realizadas pelas assistentes odontológicas; entretanto, não aumentou esses coeficientes de estudantes e professores que atuaram, também, como avaliadores. Do mesmo modo,

o treinamento não teve efeito significativo sobre os coeficientes de confiança inter-avaliadores.

NATKIN & GUILD ²³ reportaram um aumento estatisticamente significativo dos coeficientes de confiança (medidos através de todos os desvios da média), após a realização de oito sessões de treinamento. Cada sessão de treinamento consistiu do exame de 10 projetos de estudantes, escolhidos ao acaso, discutindo-se os critérios estabelecidos para esses projetos.

HINKELMAN & LONG ¹⁵ constataram que o treinamento não proporcionou aumento significativo na porcentagem de concordâncias inter-examinadores, tanto com a utilização de escalas de avaliação de 2 pontos, quanto com o uso de escalas de 3 pontos. O tipo e extensão do treinamento não foram definidos.

HUNTER ¹⁷ não se refere, especificamente, a treinamento para a utilização de seu sistema de avaliação; fala, entretanto, em sessões de calibração na clínica.

Nessas sessões, os estudantes não são, propriamente, avaliados; avalia-se os examinadores e o sistema de avaliação. Assim, cada examinador acompanha o desempenho clínico, nos mesmos seis pacientes, dos mesmos seis alunos; registram os escores de todos os critérios individuais, bem como as considerações que cercam os desempenhos. Posteriormente, um escore global ou grau é atribuído, pelos diferentes examinadores, a cada desempenho.

Confrontam-se, então, os diversos graus atribuídos e verifica-se a porcentagem de calibração entre os examinadores.

O autor afirma que, tipicamente, ocorre o seguinte: 1. alguns examinadores são mais flexíveis que outros, ao avaliarem alguns critérios individuais e 2. certos critérios de desempenho proporcionam maior calibração que outros. 80% de calibração entre os examinadores é considerada ótima; se houver muita divergência entre eles pode-se modificar o teste, reduzindo o número de escores e categorias de graus empregados, treinar os examinadores e/ou revisar as definições de desempenho.

PATRIDGE & MAST ²⁴ oferecem as seguintes recomendações para melhorar a metodologia de documentação dos estudos de avaliação clínica:

1. Identificar o objetivo da avaliação.
2. Descrever o sistema de atribuição de escores utilizado.
3. Descrever e dar exemplos dos critérios de desempenho usados e como esses critérios se originaram.
4. Examinar e reportar os coeficientes de confiança intra e inter-avaliadores.
5. Descrever o treinamento dos avaliadores e reportar os coeficientes de confiança intra e inter-avaliadores, após o treinamento.
6. Identificar o número e tipo de avaliadores.
7. Coletar e reportar evidências de validade para a avaliação.
8. Coletar e reportar evidências da praticabilidade do sistema de avaliação, isto é: tempo requerido para o desenvolvimento dos critérios e para cada avaliação.

3. MATERIAL E MÉTODOS

O planejamento, elaboração e aplicação de um instrumento de avaliação qualitativa, com o objetivo de se avaliar o desempenho clínico dos alunos de graduação da FOP-UNICAMP, basearam-se na metodologia proposta por HUNTER ¹⁷ (PEG: Performance Evaluation Guides). Tal instrumento deveria se adaptar às atuais condições de ensino-aprendizagem desta Faculdade; com esse propósito introduziram-se no PEG uma série de modificações, que atendessem a essas condições.

Como decisão inicial, optou-se por sua aplicação na avaliação do desempenho clínico final dos alunos de graduação; essa escolha se norteou na necessidade premente de se consolidar um sistema de avaliação destinado a aferir o grau de treinamento clínico dos alunos da última série do curso de graduação para, posteriormente, adaptá-lo às séries precedentes. Essa decisão foi coerente com a estratégia empregada na reforma curricular da FOP-UNICAMP, que se processou, também, no sentido inverso, isto é, das últimas às primeiras séries do curso de graduação.

Escolheu-se, então, como população alvo, alunos do 7º semestre de 1980. Era necessário eleger, como amostra de trabalho, para ser avaliada, uma função clínica básica que, a esse nível de aprendizagem, fosse de execução consciente por todos os alunos da Clínica Odontológica III (DC-701). Optou-se, por conseguinte, pela avaliação do desempenho clínico de cavidades Classe I para amálgama. Dentre essas cavidades, entretanto, selecionaram-se, para a avaliação, apenas as que decorrem de cáries de sulcos e fissuras nas superfícies oclusais de pré-molares. Esse tipo de atividade clínica, além de ser de ocorrência muito freqüente nas Clínicas Odontológicas, facilita a uniformidade de aplicação e o planejamento e construção do instrumento de avaliação.

Como se decidiu avaliar o desempenho final do aluno, dessa função clínica básica (Cavidade Classe I para amálgama) escolheram-se, apenas, as tarefas consideradas críticas. Essas tarefas críticas são as que influem decisivamente no resultado final do desempenho clínico, bem como se relacionam com a integridade dental e o bem estar do paciente.

Foram três as tarefas críticas selecionadas para

avaliação:

- A. Preparo de cavidades Classe I para amálgama.
- B. Proteção pulpar (forramento).
- C. Restauração do dente - Classe I com amálgama.

O objetivo da avaliação foi o de diagnosticar os erros dos alunos. A decisão de se avaliar os erros críticos, configurados pelos critérios de desempenho, foi tomada em virtude de se inferir que, a esse nível de aprendizagem (7º semestre), os alunos acertam mais do que erram, no desempenho clínico dessas tarefas.

A partir das premissas expostas planejou-se e construiu-se o instrumento de avaliação que se compõe de cinco formulários e dois conjuntos de folhas de instruções.

Os três primeiros formulários se destinam aos critérios de desempenho; o quarto é o formulário clínico e o quinto o formulário de notas finais. Os dois conjuntos de folhas de instruções se destinam: um aos alunos e outro aos examinadores.

3.1. FORMULÁRIOS "A" (CRITÉRIOS DE DESEMPENHO)

Construíram-se três formulários "A" que visam diagnosticar os erros críticos mais comuns, dos alunos, na execução clínica das três tarefas selecionadas.

Assim, o Formulário A₁ (Anexo 1) se destina a avaliar os erros cometidos durante o preparo de cavidades Classe I para amálgama; o Formulário A₂ (Anexo 2), aos erros cometidos durante a proteção pulpar (forramento) e, finalmente, o Formulário A₃ (Anexo 3) aos erros críticos detectáveis ao longo da restauração do dente - Classe I com amálgama.

Para o julgamento dos diversos critérios de desempenho, constantes dos três formulários "A", utilizou-se o mesmo sistema de atribuição de escores.

O sistema analítico empregado utilizou-se de uma escala de quatro escores no registro do julgamento dos diversos

critérios de desempenho.

Os quatro escores usados na avaliação dos critérios de desempenho e suas respectivas descrições são os seguintes:

C = CORRIGÍVEL : Identifica um erro crítico que permite, sempre, que o aluno o corrija.

Ic = INACEITÁVEL-CORRIGÍVEL : Configura um erro crítico que, embora inaceitável, ainda permite correções por parte do aluno.

If = INACEITÁVEL-FALHO : O erro é considerado inaceitável embora não comprometa, seriamente, a estrutura dental do paciente; entretanto não permite correções por parte do aluno.

F = FALHA CLÍNICA : Representa um erro clínico grave e incorrigível, que compromete seriamente a estrutura dental do paciente.

Descreveremos, agora, os demais componentes utilizados nos Formulários "A".

3.1.1. CRITÉRIOS DE DESEMPENHO

No presente trabalho, utilizou-se, para a identificação dos critérios de desempenho, a análise de tarefas. Essa análise se baseou na "Análise de funções e tarefas clínicas" elaborada pelo corpo docente da FOP-UNICAMP²⁸ e empregada no atual processo de ensino-aprendizagem desta casa de ensino.

Assim, fracionaram-se as três tarefas críticas, a serem avaliadas, em suas partes ou passos. Cada passo de tarefa, por sua vez, foi subdividido em um ou mais critérios de desempenho configurados por determinados erros críticos, passíveis de serem cometidos pelos alunos avaliados.

Deste modo, por exemplo, a tarefa 0072 - Preparo de cavidades Classe I para amálgama (vide ANEXO 1) foi fracionada em três passos de tarefa: a) delimitação de contorno cavitário, b) preparo de formas de resistência e retenção e c) remoção de tecido cariado. Destacando-se, apenas para exemplificar, o passo de tarefa (a) delimitação de contorno cavitário, verifica-se que o mesmo foi subdividido em quatro critérios de desempenho re

presentados pelos seguintes erros críticos: abertura excessiva, abertura insuficiente, crista marginal enfraquecida e cúspides enfraquecidas.

Os demais erros críticos ou critérios de desempenho, que compõem essa tarefa, foram obtidos através dos mesmos procedimentos descritos.

Da mesma forma foram identificados e listados os erros críticos das tarefas 0074 - Proteção pulpar (forramento) e 0075 - Restauração do dente - Classe I com amálgama (vide ANEXOS 2 e 3).

A seleção e listagem dos critérios de desempenho constantes dos Formulários "A" nortearam-se nos princípios gerais dos preparos cavitários e restaurações para amálgama descritos por MONDELLI e colab.²² e na experiência clínica do próprio autor do presente trabalho.

Após a seleção e listagem, os critérios de desempenho foram analisados e revistos por dois especialistas da área de Dentística Operatória da FOP-UNICAMP.

3.1.2. ESCORES

Já vimos que o sistema de atribuição de escores empregado utilizou-se de uma escala de quatro escores no registro do julgamento dos diversos critérios de desempenho que compõem os Formulários "A".

Para o registro do julgamento de cada critério, entretanto, utilizaram-se, sempre, um ou dois desses escores. Assim, a cada critério de desempenho atribuiu-se, apenas, um escore para cada definição do mesmo. Deste modo, à maioria dos critérios atribuiu-se somente um escore, porquanto aos mesmos se enquadrava apenas uma definição de desempenho. A alguns critérios, que admitiam duas definições, atribuíram-se dois escores, isto é, um para cada definição.

Os escores atribuídos foram: C = Corrigível, Ic = Inaceitável-Corrigível, If = Inaceitável-falho e F = Falha Clínica.

Quando ocorrer o registro do escore "C", o avalia

dor deve dar orientação ao aluno no sentido de que corrija o erro praticado. O aluno, por sua vez, deve, sempre, tentar tais correções; se acerta, não é novamente penalizado; se comete novos erros ou repete o mesmo erro, é penalizado com o registro dos escores representativos dos erros praticados.

O mesmo procedimento deve ser seguido, pelo avaliador, quando do registro do escore "Ic".

Quando o aluno não comete erro, o procedimento é considerado aceitável e não há registro de escores. O escore global ou grau, que significa o julgamento da função clínica básica como um todo, será atribuído através da utilização do Formulário "B" (Formulário clínico).

3.1.3. DEFINIÇÕES OU INDICADORES DE DESEMPENHO

As definições ou indicadores de desempenho, constantes dos Formulários "A", são descrições avaliativas, objetivas e mensuráveis, que permitem, ao avaliador, observar a ocorrência de erros críticos no desempenho do aluno. Esses indicadores de desempenho, portanto, fornecem parâmetros precisos, aos examinadores, de quando detectar os erros praticados e, aos alunos, de quando e porque estão cometendo erros.

Procurou-se, sempre que possível, elaborar indicadores de desempenho com descrições avaliativas que se utilizavam, como meios de mensuração, de instrumentos e/ou materiais odontológicos comuns. A descrição avaliativa de características visuais foi utilizada quando nenhum meio instrumental e/ou material eram apropriados.

Assim, por exemplo, na elaboração do indicador do critério "abertura excessiva", do Formulário "A₁" (vide ANEXO 1), utilizou-se, como meio de mensuração, a broca para alta rotação 57 (cilíndrica lisa). A escolha dessa broca se deveu ao fato de que sua ponta ativa mede 0,8mm de diâmetro; essa medida foi coerente para aferir o tamanho da abertura cavitária realizada pelo aluno. Com efeito, os parâmetros usados no indicador do critério "abertura excessiva" mostraram, na prática, boas relações de proporcionalidade com os diâmetros vestibulo-linguais de

prémolares (8,5mm em média) e molares (11mm em média), inferidos das pesquisas reportadas por DELLA SERRA & FERREIRA ⁷.

Na elaboração de nove dos indicadores de desempenho descritos nos Formulários "A" (vide ANEXOS 1, 2 e 3), utilizaram-se, como meios de mensuração, instrumentos e material odontológicos comuns (broca 57, explorador, calcador e carbono para articulação); na construção dos doze indicadores restantes, empregou-se a descrição avaliativa de características visuais. Pode-se observar, entretanto, que em alguns indicadores utilizaram-se, além de meios de mensuração instrumentos ou materiais, descrições avaliativas de características visuais.

Os indicadores elaborados nessa pesquisa descrevem, sempre, características de desempenhos insatisfatórios que fornecem padrões para o julgamento dos vários erros passíveis de serem cometidos pelos alunos.

Assim, na avaliação dos critérios, quando os indicadores se aplicam aos desempenhos dos alunos os erros críticos estão configurados e seus respectivos escores são registrados.

Ao contrário, se os desempenhos dos estudantes não se enquadram aos indicadores dos critérios avaliados, os mesmos não cometem os erros críticos em julgamento. Nesse caso seus procedimentos são considerados satisfatórios e não haverá registro de escores.

Por vezes, os indicadores elaborados nesse trabalho descrevem mais de uma característica de desempenho insatisfatório. Assim, por exemplo, o indicador "hemorragia; polpa visível" (vide ANEXO 1), foi utilizado para o julgamento do erro crítico ou critério de desempenho "exposição mecânica". Nesse indicador são descritas duas características de desempenhos insatisfatórios: "hemorragia" e "polpa visível". Pode acontecer, entretanto, "exposição mecânica" apenas com "polpa visível", sem "hemorragia".

Entretanto, para efeito de avaliação do critério de desempenho "exposição mecânica", a ocorrência de apenas uma ou de ambas características descritas pelo indicador, configuram o erro crítico e o respectivo escore é registrado.

O mesmo procedimento foi adotado na avaliação dos

demaís critérios que se utilizam de indicadores com mais de uma característica de desempenho insatisfatório.

Na elaboração dos indicadores de desempenho constantes dos Formulários "A", a exemplo do que ocorreu com a seleção e listagem dos critérios de desempenho, empregaram-se a experiência clínica do próprio autor e os princípios gerais dos preparos cavitários e restaurações para amálgama descritos por MONDELLI e colab.²². Aproveitou-se, também, como referência, os exemplos de indicadores de desempenho elaborados por CHARBENEAU⁶.

3.1.4. EXCEÇÕES

As exceções descritas nos Formulários "A" (vide ANEXOS 1, 2 e 3) foram empregadas quando os indicadores dos critérios de desempenho não se aplicavam a determinados casos clínicos incomuns.

Assim, por exemplo, quando a cárie é muito profunda ou há extremo desgaste oclusal (exceções), no dente a ser preparado, os parâmetros exigidos pelo indicador de desempenho do critério cavidade profunda não servem para configurar o respectivo erro crítico. Isto significa que, para esses casos incomuns (exceções), mais que 1/2 da ponta ativa da broca 57 podia penetrar na cavidade, em relação ao bordo cavo-superficial, junto as cristas marginais M e D, sem que se configurasse o erro crítico cavidade profunda. Ocorre que a cavidade ou já era profunda por sua própria característica inicial (cárie profunda) ou tornou-se profunda, após o preparo, devido ao extremo desgaste oclusal do dente objeto da avaliação.

Para cada indicador de desempenho descreveram-se uma ou mais exceções. Determinados indicadores, entretanto, não admitiam nenhuma exceção.

Para efeito de avaliação dos diversos critérios, a ocorrência de uma ou mais exceções em determinados indicadores de desempenho, eximiam os alunos dos respectivos erros críticos. Nesses casos os escores, que configuravam esses erros críticos, não eram registrados.

3.1.5. RAZÕES OU FUNDAMENTOS DOS CRITÉRIOS

As razões ou fundamentos dos critérios de desempenho ou erros críticos descritos nos Formulários "A" (vide ANEXOS 1, 2 e 3) justificaram a inclusão desses critérios no instrumento de avaliação. Com efeito, essas razões ou fundamentos enfatizavam os princípios básicos que delimitavam os critérios. Desse modo, os erros críticos listados ou não atendiam aos fundamentos básicos do preparo, forramento e restauração dos dentes ou poderiam provocar falhas indesejáveis na preservação e utilização dos mesmos após a restauração.

Assim, por exemplo, o erro crítico cavidade profunda não atendia à razão ou fundamento básico proteção pulpar; do mesmo modo, a falta de material, na condensação do amálgama, poderia provocar infiltração marginal e/ou recidiva de cárie após a restauração do dente. Já o erro crítico contacto oclusal alto, ao teste de oclusão, além de não atender ao fundamento função mastigatória, certamente provocará trauma oclusal quando da utilização do dente restaurado.

Para cada erro crítico, ou critério de desempenho, utilizado correspondiam sempre, uma ou mais razões ou fundamentos. A presença de apenas uma razão ou fundamento, entretanto, já justificava a inclusão do respectivo erro crítico no instrumento de avaliação.

Ao contrário, um erro crítico que não correspondesse a determinada razão ou fundamento não deveria ser, como não o foi, incluído entre os critérios de desempenho dos Formulários "A", pois seria considerado dispensável.

3.2. FORMULÁRIO "B" (FORMULÁRIO CLÍNICO)

O formulário clínico (Formulário "B" - ANEXO 4) é o utilizado, na Clínica, para o registro de escores e das considerações atribuídos, pelos examinadores, ao desempenho do aluno na função básica avaliada. Ele possui, em sua parte superior, espaços para os dados de identificação do aluno, da Clínica a que pertence e do dente objeto da avaliação; possui, também, lugar reservado aos nomes do examinador e do paciente no qual se

executará a função clínica a avaliar; tem, finalmente, um local para o número do prontuário geral do aludido paciente.

O corpo central do formulário clínico se compõe: de uma coluna, à esquerda, para o registro dos escores e seu respectivo total; de outra coluna, à direita, para o registro das considerações e seu correspondente total.

3.2.1. COLUNA DE REGISTRO DOS ESCORES

Nessa coluna foram listados todos os erros críticos ou critérios de desempenho utilizados nos Formulários "A" com seus escores correspondentes. Esses erros e seus respectivos escores foram enumerados em blocos independentes, segundo a tarefa a que pertenciam, com o objetivo de facilitar tanto o registro desses escores, durante o processo de avaliação, quanto a sua totalização.

Decidiu-se, para facilitar a obtenção do escore global ou grau, atribuir pesos negativos diferentes aos quatro escores representativos dos diversos erros críticos listados. Esses pesos negativos, expressos em números, visavam penalizar, hierarquicamente, os erros considerados mais graves.

Assim, foram atribuídos os seguintes pesos negativos aos escores utilizados no presente instrumento de avaliação:

C = Corrigível = - 1

Ic = Inaceitável-corrigível = - 2

If = Inaceitável-falho = - 3

F = Falha clínica = - 5

A diferença maior do valor negativo entre os escores If e F (-2) foi atribuída em virtude de se considerar que F representava um erro crítico de grande gravidade, que comprometia a estrutura dental do paciente.

Os pesos negativos numéricos atribuídos foram listados, junto a seus respectivos escores, na coluna de registro dos mesmos. O total dos escores deverá ser registrado, em espaço próprio, ao final dessa coluna; deverá ser obtido pela soma

dos vários pesos negativos correspondentes aos escores assinalados durante o processo de avaliação do desempenho dos alunos.

3.2.2. COLUNA DE REGISTRO DAS CONSIDERAÇÕES

Nessa coluna foram listadas, também em blocos independentes, as considerações que podem afetar o desempenho dos alunos durante o processo avaliativo. Essas considerações também influem na atribuição do escore global ou grau dos alunos; são fatores anormais que podem cercar o desempenho da função clínica básica avaliada.

A exemplo dos escores, os componentes de algumas considerações também receberam pesos negativos diferentes; esses pesos negativos, expressos em números, foram atribuídos aos componentes dessas considerações procurando-se penalizá-los, hierarquicamente, segundo o grau de suas influências na avaliação global do desempenho do aluno.

Assim, as seguintes considerações foram relacionadas e utilizadas no processo avaliativo:

3.2.2.1. TEMPO PARA EXECUÇÃO DA FUNÇÃO CLÍNICA

Dados obtidos, sem comprovação estatística, junto à Clínica Odontológica da FOP-UNICAMP, pelo próprio autor e vários professores da referida Clínica, demonstraram que o tempo máximo utilizado, em média, pelos alunos, na execução de uma restauração Classe I com amálgama foi de 30 minutos. Baseado nesses dados, estipulou-se um tempo máximo de 30 minutos para a consecução da função clínica avaliada, sem penalização. A cada 5 minutos, além do tempo máximo, foram atribuídos pesos negativos, em progressão aritmética dois a dois, até -14, que representava um tempo de 61 minutos ou mais utilizado para a execução da função clínica avaliada.

Esses pesos negativos, com seus respectivos intervalos de tempo, constituíram uma escala de tempo para execução da função clínica, incluída no primeiro bloco da coluna de registro das considerações. Nesse mesmo bloco reservaram-se locais

para o registro do horário inicial e final de execução da função clínica avaliada, bem como do tempo real utilizado, em minutos, para a realização da mesma. Ao final desse bloco, destinou-se um lugar para a anotação do tempo utilizado, pelos examinadores, para a execução de cada avaliação.

3.2.2.2. DIFICULDADES PARA EXECUÇÃO DA FUNÇÃO CLÍNICA

Certos fatores negativos inerentes ao paciente ou ao instrumental e material utilizados, podem dificultar a execução, pelo aluno, da função clínica objeto da avaliação.

Essas dificuldades, para serem contornadas, exigem mais tempo para o desempenho da aludida função pelo aluno. Uma das maneiras de se evitar essas dificuldades seria a seleção prévia dos pacientes nos quais seriam executadas as funções clínicas a avaliar. Como, todavia, queremos utilizar um instrumento para avaliar qualquer caso clínico, devemos levar em conta as dificuldades mencionadas, que podem ser assim enumeradas:

- A) Dentes com desvios acentuados de seu longo eixo.
- B) Dificuldades de visão direta ou indireta do dente a ser preparado.
- C) Paciente que não colabora.
- D) Língua excessivamente grande (principalmente em preparo de dentes inferiores).
- E) Excesso de salivagem (não controlável por medicamentos).
- F) Deficiências no material e/ou instrumental utilizados.
- G) Outras dificuldades (notadas pelos examinadores).

As dificuldades constatadas, pelos avaliadores, são assinaladas, pelos mesmos, nos quadrados correspondentes inseridos no segundo bloco da coluna de registro das considerações.

A cada dificuldade anotada são dados, ao aluno, mais 5 minutos para a execução da função clínica; o acréscimo total de tempo para o desempenho da função clínica é registrado no espaço próprio do formulário; esse acréscimo é deduzido, pos

teriormente, do tempo utilizado pelo aluno. Essa dedução, finalmente, é realizada antes da marcação dos pontos negativos que correspondem ao tempo total empregado pelo estudante.

3.2.2.3. AJUDA DO EXAMINADOR

Um dos examinadores executa parte da função clínica avaliada. Essa ajuda influi, negativamente, na atribuição do escore global ou grau do aluno; a ajuda é dada por solicitação expressa pelo aluno ou para preservar a integridade dental do paciente.

A cada ajuda dos examinadores será atribuído um peso de -5, até o máximo de três ajudas que corresponderá ao peso de -15. Acima desse limite não serão oferecidas mais ajudas ao aluno e um dos examinadores concluirá a função clínica; nesse caso ou se atribui grau 0 (zero) ao aluno ou se cancela a referida avaliação.

A aparente severidade na atribuição de pesos negativos às ajudas, prestadas ao aluno, se deve ao fato de ser quase inadmissível, a esse nível de aprendizado clínico, a necessidade das mesmas, mormente para preservar a integridade dental do paciente.

As ajudas oferecidas pelos examinadores são registradas, por todos os avaliadores, nos locais apropriados do terceiro bloco da coluna de registro das considerações.

3.2.2.4. ORIENTAÇÃO DO EXAMINADOR

Um dos examinadores orienta (ajuda verbal) parte da função avaliada. Essa orientação também pesa na atribuição do grau do aluno; da mesma forma que a consideração anterior, a orientação do examinador é prestada por solicitação expressa pelo aluno ou para preservar a integridade dental do paciente.

A cada orientação dos examinadores será atribuído um peso de -2 até o máximo de três, que corresponderá ao peso de -6. Acima desse limite não serão oferecidas mais orientações ao aluno. Um dos examinadores concluirá a função clínica se houver riscos para a integridade dental do paciente; nesse caso se

procederá de maneira semelhante à prevista para a atribuição do grau quando houver mais de três ajudas dos examinadores.

Não se incluem nessas considerações as orientações oferecidas aos alunos, no sentido de que corrijam os erros praticados, quando da atribuição dos escores "C" e/ou "Ic".

As orientações prestadas pelos examinadores serão assinaladas, por todos os avaliadores, em locais próprios do quarto bloco da coluna de registro das considerações.

O total das considerações deverá ser anotado no último bloco dessa coluna e será obtido pela soma dos pesos negativos assinalados nos blocos precedentes.

O grau ou escore global da função clínica avaliada é obtido pela soma dos totais dos escores e considerações. É expresso em pontos negativos e significa a somatória de todos os pesos atribuídos aos escores e às considerações.

A nota final do aluno será atribuída consultando-se o Formulário "C", como veremos.

O registro do grau e da nota final do aluno é feita na porção inferior do Formulário "B", onde também são registradas as assinaturas do examinador e do aluno.

3.3. FORMULÁRIO "C" (NOTAS FINAIS)

O grau, atribuído ao aluno, é o julgamento global do desempenho do mesmo no Teste como um todo.

O registro de um escore avalia o desempenho, do aluno, em um único critério. Já a atribuição do grau julga o desempenho, do estudante, em toda a função clínica avaliada, incluindo as considerações que a envolveram. Assim, por exemplo, dois alunos executaram procedimentos clínicos semelhantes e receberam idênticos escores, nos vários critérios de desempenho. Eles devem receber o mesmo grau? Provavelmente não. Porque? Devido as considerações que acompanharam o julgamento da função clínica, ou sejam, fatores anormais como: tempo excessivo para completar a função, ajuda dos examinadores, etc.

Assim, para tornar mais objetivo e abrangente o grau atribuído ao aluno, utilizou-se a equação: grau = total dos escores + total das considerações. Esse grau é expresso em pontos negativos e significa a somatória de todos os pesos negativos atribuídos aos escores e às considerações.

Baseado nessas assertivas, construiu-se o Formulário "C" - Notas Finais (ANEXO 5) através do qual os graus atribuídos, aos alunos, serão transformados em notas finais.

Nesse Formulário a coluna da esquerda contém as notas finais, escalonadas de 0,0 a 10,0, que correspondem às faixas de pontos negativos que representam a somatória dos totais dos pesos atribuídos aos escores e às considerações (graus), registrados na coluna da direita.

Partiu-se da nota 10,0 a qual corresponde o grau 0 (zero). Isto significa que o aluno não cometeu nenhum erro crítico e nem incorreu em nenhuma consideração.

Desta nota máxima partiu-se para notas cada vez menores que correspondem a faixas de pontos negativos, gradativamente, crescentes. Assim, chegou-se a nota 0,0 (zero) a qual correspondem pesos negativos iguais ou superiores a -41.

Pela análise da coluna de registro dos escores (Formulário "B" - ANEXO 4), pode-se verificar que se forem atribuídos todos os pesos negativos, correspondentes à assinalação de todos os escores listados, o seu total será de -41; isto significará que o aluno cometeu todos os erros críticos possíveis sem, todavia, repetí-los. Sua nota final será 0,0 (zero), porquanto seu grau atingiu o valor negativo máximo, independentemente do acréscimo dos pesos negativos das considerações. Se, entretanto, além de cometer todos os erros possíveis o aluno repetir alguns deles, esse total dos escores atingirá, cumulativamente, valores superiores a -41; contudo, sua nota final continuará sendo 0,0 (zero).

O aluno pode cometer novos erros ou repetir os mesmos erros na tentativa de corrigir falhas ocorridas quando da atribuição de escores "C" e/ou "Ic"; nesses casos, como já vimos, é penalizado com o registro de novos escores representativos dos erros praticados. Tais fatos significam que a prática

de erros quando da tentativa de correções faz com que o escore total do aluno aumente, progressivamente, podendo atingir valores superiores a -41; conseqüentemente, sua nota final diminui, paulatinamente, podendo atingir 0,0 (zero), já que os pontos negativos de seu grau aumentam, gradativamente, mesmo sem o acréscimo dos pesos negativos das considerações.

Do exposto, pode se depreender que tanto a prática de erros quando da tentativa de correções quanto o recebimento de pesos negativos das considerações aumentam, progressivamente, o valor negativo do grau do aluno e, conseqüentemente, diminuem sua nota final.

3.4. CONJUNTOS DE FOLHAS DE INSTRUÇÕES

Elaboraram-se dois conjuntos de folhas de instruções destinados a implementar a aplicação do Instrumento de avaliação proposto por esta pesquisa.

O primeiro contém folhas de instruções dirigidas aos alunos (ANEXO 6), com informações detalhadas de como proceder antes, durante e depois do processo avaliativo.

O segundo, de interesse dos examinadores (ANEXO 7), compõem-se de folhas com instruções minuciosas de como os avaliadores devem utilizar os componentes do aludido instrumento na avaliação do desempenho clínico dos alunos da DC - 701.

3.5. IMPLEMENTAÇÃO DO INSTRUMENTO DE AVALIAÇÃO

O instrumento de avaliação construído compõe-se, como vimos, de cinco formulários e dois conjuntos de folhas de instruções. Entretanto, para sua implementação e aplicação, esse instrumento, após impressão em "off-set", foi dividido em dois blocos destinados: um aos alunos e o outro aos examinadores.

O bloco reservado aos alunos era constituído de cópias dos Formulários "A" (Critérios de desempenho - ANEXOS 1, 2 e 3) e do conjunto de folhas de instruções dirigidas aos mesmos (ANEXO 6). Já o bloco destinado aos examinadores compunha-

se de cópias dos Formulários "A", do Formulário "B" (Formulário Clínico - ANEXO 4) e do conjunto de folhas de instruções de interesse dos mesmos (ANEXO 7).

As cópias do Formulário "C" (Notas Finais - ANEXO 5) foram mantidas com o Administrador do Teste; entretanto, eram fornecidas, mediante solicitação, aos alunos e examinadores interessados.

Convocaram-se todos os alunos da Clínica Odontológica III (DC - 701) para uma reunião em que se explanaram, aos mesmos, as diretrizes gerais tanto do instrumento de avaliação construído, quanto da sua aplicação. Na oportunidade, solicitou-se a colaboração voluntária de todos esses alunos para que se submetessem à aplicação do referido instrumento.

Paralelamente, convidaram-se três professores, que compõem o corpo clínico das Clínicas Odontológicas da FOP-UNICAMP, para, juntamente com o autor, atuarem como examinadores; o autor do presente trabalho, também professor integrante do aludido corpo clínico, além da função de examinador, desempenhou a tarefa de Administrador do Teste.

Os quatro examinadores, utilizados na avaliação do desempenho clínico dos alunos da DC - 701, possuem, no mínimo, 10 (dez) anos de experiência clínica e docente. Nenhum deles, entretanto, tinha grande experiência em avaliação clínica.

Para a implementação do instrumento de avaliação, os quatro examinadores se reuniram sob a coordenação do Administrador do Teste. Distribuiu-se, na oportunidade, a cada um dos examinadores um bloco contendo cópias da parte do Instrumento reservada aos mesmos.

Nessa reunião discutiu-se, exaustivamente, a respeito de todos esses componentes do instrumento de avaliação; enfatizou-se, sobretudo, a interpretação avaliativa das características dos indicadores e dos critérios de desempenho; planejaram-se, finalmente, detalhes do processo de aplicação do instrumento de avaliação.

Dentre os alunos da DC - 701 escolheram-se, ao acaso, para a avaliação os que possuíam pacientes com cáries de sulcos e fissuras (cavidade Classe I) na superfície oclusal de molares e/ou pré-molares. Esses alunos foram escolhidos, paula

tinamente, segundo o número de vagas em cada sessão de avaliação. Aproximadamente uma semana antes de cada sessão de avaliação, era distribuído, a cada aluno selecionado para a respectiva sessão, um bloco contendo cópias da parte do Instrumento destinada ao mesmo; na mesma ocasião eram-lhe fornecidas películas radiográficas para que radiografassem o dente a ser preparado durante o processo avaliativo; as radiografias obtidas tinham o objetivo de verificar se o processo de cárie, do dente objeto da avaliação, localizava-se apenas em sua superfície oclusal.

Durante a semana que antecedia a sua sessão de avaliação, os alunos selecionados mantinham, se julgassem necessário, reuniões com o Administrador do Teste para esclarecerem possíveis dúvidas quanto ao conteúdo, implementação e aplicação do Instrumento.

3.6. APLICAÇÃO DO INSTRUMENTO DE AVALIAÇÃO

O Instrumento, elaborado no presente trabalho, foi aplicado em várias sessões de avaliação realizadas, predominantemente, em dependências da Clínica Odontológica da FOP - UNICAMP, nos próprios boxes dos alunos; a eleição deste local foi feita pelos próprios alunos que preferiram ser avaliados em seus costumeios ambientes de trabalho.

Por vezes, entretanto, algumas sessões de realizaram na Unidade Clínica Piloto - sala 3-47 (conjunto de equipamentos odontológicos simplificados); a escolha deste local, com a anuência dos alunos interessados, foi feita porque o mesmo oferecia maior tranquilidade e espaço de trabalho aos examinadores.

Assim, foram examinados 21 alunos na Clínica Odontológica e 9 na Unidade Clínica Piloto, perfazendo um total de 30 alunos avaliados. Esses alunos avaliados receberam, à guisa de identificação, números de 1 a 30.

Na avaliação desses 30 alunos da DC - 701 utilizouse, sempre, 3 ou 4 examinadores. Esses 4 examinadores foram identificados, no presente trabalho, pelas letras A, B, C e D.

Nem sempre foram usados os 4 (quatro) examinadores.

res, porquanto um deles tinha horário funcional diferente dos demais. Esse examinador participou, apenas, da avaliação dos 9 alunos que trabalharam na Unidade Clínica Piloto. Assim, os quatro examinadores e os respectivos números de avaliações em que participaram constam do quadro abaixo.

EXAMINADOR	Nº DE AVALIAÇÕES
A	29*
B	30
C	30
D	9

* O examinador A deixou de participar de uma avaliação realizada na Unidade Clínica Piloto.

Cada avaliação realizada desenvolveu-se dentro dos parâmetros propostos pelas instruções aos alunos (ANEXO 6) e aos examinadores (ANEXO 7) e pelas informações contidas nos Formulários "A" (ANEXOS 1, 2 e 3) e "B" (ANEXO 4).

O tempo utilizado para a execução da função clínica em cada avaliação foi obtido através do uso de um cronômetro. Assim, no momento em que os examinadores registravam o horário inicial de execução da aludida função clínica, um cronômetro era acionado pelo Administrador do Teste. Esse cronômetro era paralizado a cada interrupção do trabalho do aluno, destinada à avaliação de cada tarefa clínica. Ao término da avaliação de cada tarefa, o cronômetro era novamente acionado. Procedeu-se dessa maneira até o registro, pelos examinadores, do horário final de execução da função clínica, quando o cronômetro foi paralizado, pela última vez, na respectiva avaliação.

Obteve-se, assim, o tempo real utilizado para a execução da função clínica em cada avaliação. Esse tempo foi registrado, pelo Administrador do Teste, no local apropriado do Formulário "B", após a conclusão de cada avaliação.

Para a aplicação do instrumento de avaliação não se programaram sessões específicas de treinamento dos examinadores. Contudo, na implementação do referido instrumento, reali-

zou-se uma reunião, entre os examinadores, na qual se enfatizou a interpretação avaliativa das características dos indicadores e dos critérios de desempenho, bem como plenajaram-se detalhes do processo de avaliação. Essas variáveis também foram discutidas, pelos examinadores, após algumas sessões de avaliação.

3.7. ADMINISTRAÇÃO DAS AVALIAÇÕES

Após cada sessão de avaliação, o Administrador do Teste procedeu ao preenchimento final dos vários Formulários Clínicos (Formulários "B") resultantes das avaliações realizadas na respectiva sessão. Inicialmente, os Formulários de cada avaliação (3 ou 4, conforme o número de examinadores que atuaram na respectiva avaliação) foram grampeados em um único conjunto. Procedeu-se, então, ao preenchimento final de cada Formulário "B" integrante de cada conjunto.

Em primeiro lugar efetuou-se a soma dos sub-totais 1, 2 e 3, bem como do total dos escores, registrando-os nos locais adequados.

Verificou-se, a seguir, se haviam sido registradas, nos quadrados correspondentes, dificuldades encontradas para a execução da função clínica avaliada; anotou-se, então, o acréscimo de tempo para a execução da função clínica (5 (cinco) minutos para cada dificuldade assinalada). Esse acréscimo foi deduzido do tempo real utilizado para a execução da função clínica avaliada. De posse desse resultado, anotou-se, na escala de tempo para execução da função clínica, o respectivo peso negativo correspondente.

Quando não haviam dificuldades assinaladas, registrava-se o peso negativo correspondente ao tempo real utilizado para a execução da função clínica avaliada.

O tempo utilizado, pelos examinadores, para a execução de cada avaliação foi obtido da seguinte forma: subtraía-se o tempo inicial do tempo final de execução da função clínica; obtinha-se, assim, o tempo total utilizado na execução da função clínica, inclusive o tempo destinado à avaliação; desse tempo total subtraía-se o tempo real utilizado na execução da função clínica avaliada, restando o tempo empregado, pelos exami

nadores, para a consecução de cada avaliação. Esse tempo foi registrado, pelo Administrador do Teste, em local apropriado do Formulário "B".

O passo seguinte foi somar os pesos negativos assinalados nos blocos das considerações, obtendo-se o total das considerações; esse total foi anotado no último bloco da coluna de registro das considerações.

A seguir, atribuiu-se, em pontos negativos, o grau ou escore global da função clínica avaliada; esse grau foi alcançado pela soma dos totais dos escores e das considerações.

Finalmente, registrou-se a nota final do aluno consultando-se, previamente, o Formulário "C".

Dessa maneira, executou-se o preenchimento final de todos os conjuntos de Formulários Clínicos (Formulários "B"), correspondentes às 30 avaliações realizadas.

De posse desses conjuntos de Formulários Clínicos, devidamente preenchidos, foi feito o processamento dos dados obtidos para determinar:

1. O coeficiente de confiança inter-avaliadores, por aluno avaliado.

2. O efeito do treinamento sobre os coeficientes de confiança inter-avaliadores.

3. O grau de uniformidade ou calibração entre os examinadores no registro dos escores.

4. O grau de uniformidade ou calibração, entre os examinadores, no registro dos escores segundo o critério de desempenho avaliado.

5. A adequação das definições (indicadores de desempenho) dos critérios de desempenho avaliados.

6. Quais os erros críticos mais freqüentes no desempenho da função clínica avaliada.

7. A validade da atribuição de pesos negativos diferentes aos escores e a algumas considerações na obtenção da nota final do aluno e grau de calibração, entre os examinadores, na concessão dessas notas finais.

4. RESULTADOS

Processados os dados, obtiveram-se, inicialmente, os coeficientes de confiança entre os pares de examinadores, por aluno avaliado.

Neste trabalho, os coeficientes de confiança inter-avaliadores são definidos pela razão, expressa em porcentagem, entre o número de vezes em que houve concordância, entre os pares de examinadores, no registro dos escores e o total dos escores possíveis de serem assinalados.

Os coeficientes de confiança obtidos são apresentados na Tabela 1.

Observando-se essa tabela, verifica-se que os coeficientes de confiança inter-avaliadores, por aluno avaliado, são altos. A variação desses coeficientes está entre o valor mínimo de 76,19% e o valor máximo de 100%.

Também foram obtidas as médias dos coeficientes entre os pares de examinadores, registradas na parte inferior da Tabela 1. Essas médias foram calculadas sobre os coeficientes de confiança alcançados, pelos pares de examinadores, em todas as avaliações em que participaram. As médias mais altas foram obtidas pelos pares de examinadores A - C, A - B e B - C (95,08%, 94,25% e 91,74%, respectivamente). As médias mais baixas foram alcançadas pelos pares de examinadores C - D, B - D e A - D (87,30%, 88,89% e 89,88%, respectivamente). Entretanto, pode-se verificar que todas as médias dos coeficientes de confiança inter-avaliadores podem ser consideradas altas, isto é, acima de 80%.

A ordem crescente dos números que identificam os alunos avaliados, também apresentada na Tabela 1, corresponde à ordem em que foram feitas as avaliações.

Pode-se admitir que a medida em que se processaram as avaliações, houve um treinamento gradual e natural dos examinadores; esse treinamento progressivo e natural deve ter sido aperfeiçoado pelas discussões efetuadas, como vimos, entre os examinadores, após algumas sessões de avaliação.

Da mesma Tabela 1 pode-se inferir, entretanto, que o treinamento dos examinadores, ao longo das avaliações, não in-

TABELA 1 - NÚMERO DE CONCORDÂNCIAS ENTRE OS EXAMINADORES NO REGISTRO DOS ESCORES E COEFICIENTE DE CONFIANÇA, EM PORCENTAGEM, SEGUNDO O ALUNO AVALIADO E O PAR DE EXAMINADORES.

ALUNO AVALIA DO	P A R D E E X A M I N A D O R E S											
	A - C		A - B		B - C		A - D		B - D		C - D	
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
01	--	--	--	--	18	85,71	--	--	21	100,00	18	85,71
02	19	90,48	21	100,00	19	90,48	--	--	--		--	
03	20	95,24	18	85,71	19	90,48	--	--	--		--	
04	19	90,48	19	90,48	17	80,95	20	95,24	19	90,48	18	85,71
05	20	95,24	20	95,24	19	90,48	19	90,48	18	85,71	20	95,24
06	19	90,48	20	95,24	18	85,71	18	85,71	17	80,95	18	85,71
07	19	90,48	20	95,24	18	85,71	18	85,71	19	90,48	20	95,24
08	19	90,48	18	85,71	20	95,24	--		--		--	
09	20	95,24	19	90,48	20	95,24	--		--		--	
10	20	95,24	20	95,24	21	100,00	--		--		--	
11	20	95,24	16	76,19	17	80,95	--		--		--	
12	21	100,00	21	100,00	21	100,00	--		--		--	
13	20	95,24	19	90,48	18	85,71	--		--		--	
14	20	95,24	21	100,00	20	95,24	--		--		--	
15	21	100,00	21	100,00	21	100,00	18	85,71	18	85,71	18	85,71
16	18	85,71	20	95,24	19	90,48	19	90,48	18	85,71	18	85,71
17	21	100,00	21	100,00	21	100,00	--		--		--	
18	21	100,00	21	100,00	21	100,00	--		--		--	
19	21	100,00	21	100,00	21	100,00	--		--		--	
20	20	95,24	20	95,24	19	90,48	20	95,24	19	90,48	19	90,48
21	18	85,71	21	100,00	18	85,71	19	90,48	19	90,48	16	76,19
22	19	90,48	20	95,24	18	85,71	--		--		--	
23	21	100,00	17	80,95	17	80,95	--		--		--	
24	21	100,00	21	100,00	21	100,00	--		--		--	
25	20	95,24	21	100,00	20	95,24	--		--		--	
26	21	100,00	21	100,00	21	100,00	--		--		--	
27	21	100,00	18	85,71	18	85,71	--		--		--	
28	21	100,00	21	100,00	21	100,00	--		--		--	
29	20	95,24	18	85,71	19	90,48	--		--		--	
30	19	90,48	20	95,24	18	85,71	--		--		--	
MÉDIA	95,08		94,25		91,74		89,88		88,89		87,30	

terferiu nos coeficientes de confiança obtidos. Com efeito, pode-se observar que os coeficientes de confiança inter-avaliadores não aumentaram em função da ordem de avaliação. Esse dado pode ser deduzido porquanto esses coeficientes apresentaram uma tendência de ocorrência irregular, mais facilmente visualizada nos gráficos apresentados nas Figuras 1 e 2. Esses curvogramas apresentam, graficamente, o número de concordâncias no registro dos escores, de cada par de examinadores, em função da ordem em que foram feitas as avaliações. Essas concordâncias representam, em última análise, os coeficientes de confiança inter-avaliadores.

Para determinar o grau de uniformidade ou calibração entre os examinadores no registro dos escores, foi organizada a Tabela 2.

Nessa tabela distribuíram-se as avaliações e seus respectivos percentuais, segundo o número de concordâncias entre os examinadores no registro dos escores, por par de examinadores que realizaram as respectivas avaliações.

Observando-se a Tabela 2, verifica-se que a distribuição das avaliações, segundo o número de concordâncias por par de examinadores, se situa entre 16 e 21 para um total de 21 concordâncias possíveis. Esta concentração da distribuição das avaliações em seus valores mais altos sugere alta uniformidade ou calibração, entre todos os examinadores, no registro dos escores.

Observa-se, também, que para os pares de examinadores A - C, A - B e B - C (75% dos avaliadores) os percentuais mais altos, do maior número de avaliações efetuadas por esses pares, correspondem ao número máximo de concordâncias possíveis (21). Isto revela altíssima calibração entre esses examinadores.

Já para os pares A - D, B - D e C - D, dos quais fazia parte o examinador D, os percentuais mais altos correspondem, respectivamente, a 18 e 20, 18 e 19 e 18 concordâncias. Embora com números de concordâncias um pouco menores, esta constatação também indica alto grau de calibração entre esses examinadores.

Para avaliar o grau de uniformidade ou calibração, entre os examinadores, no registro dos escores segundo o critério de desempenho avaliado, elaborou-se a Tabela 3. Nela apresentam-se o número de discordâncias entre os examinadores no re-

FIGURA 1 - CURVOGRAMAS DO NÚMERO DE CONCORDÂNCIAS, NO REGISTRO DOS ESCORES, DE CADA PAR DE EXAMINADORES, EM FUNÇÃO DA ORDEM DE AVALIAÇÃO.

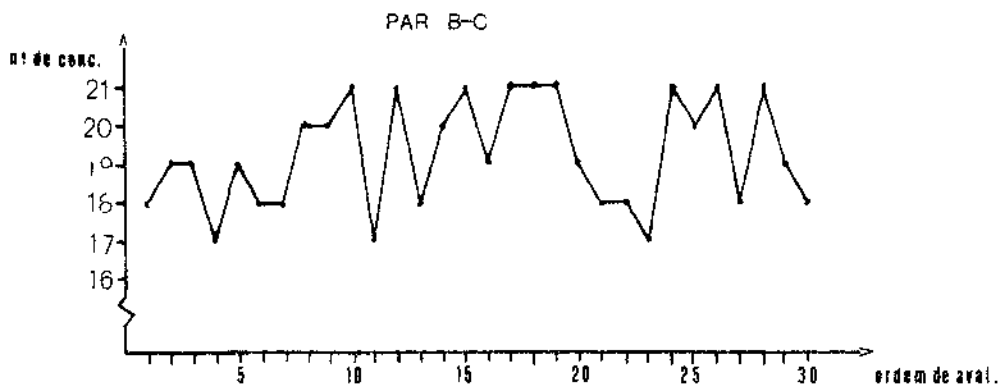
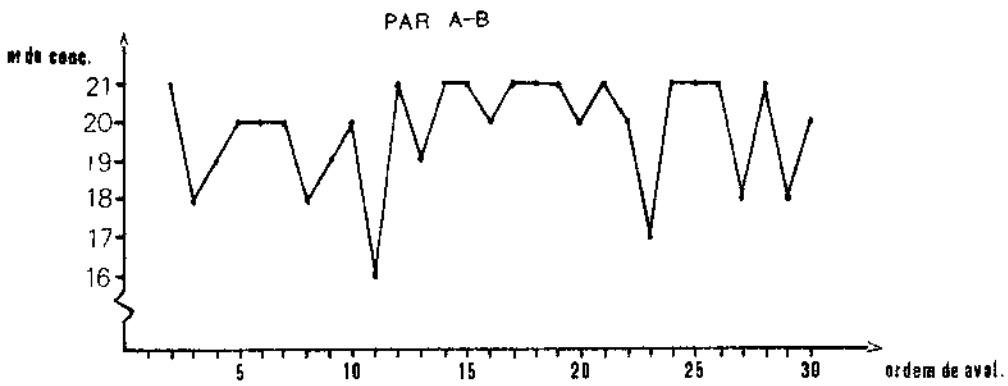
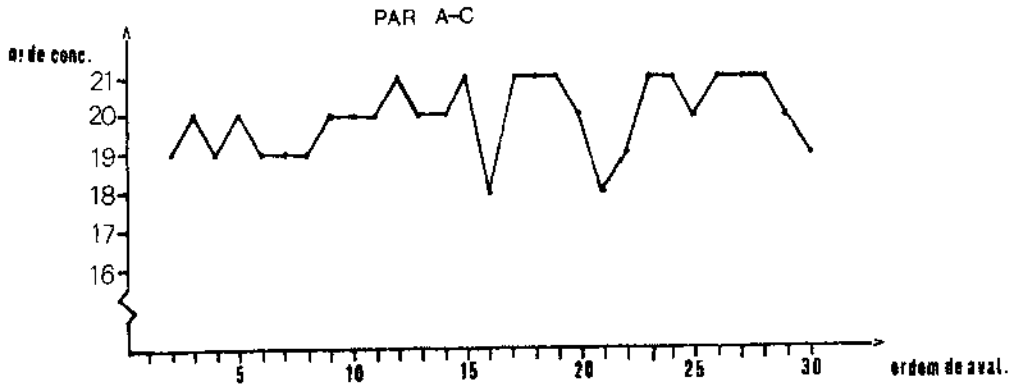


FIGURA 2 - CURVOGRAMAS DO NÚMERO DE CONCORDÂNCIAS, NO REGISTRO DOS ESCORES, DE CADA PAR DE EXAMINADORES, EM FUNÇÃO DA ORDEM DE AVALIAÇÃO.

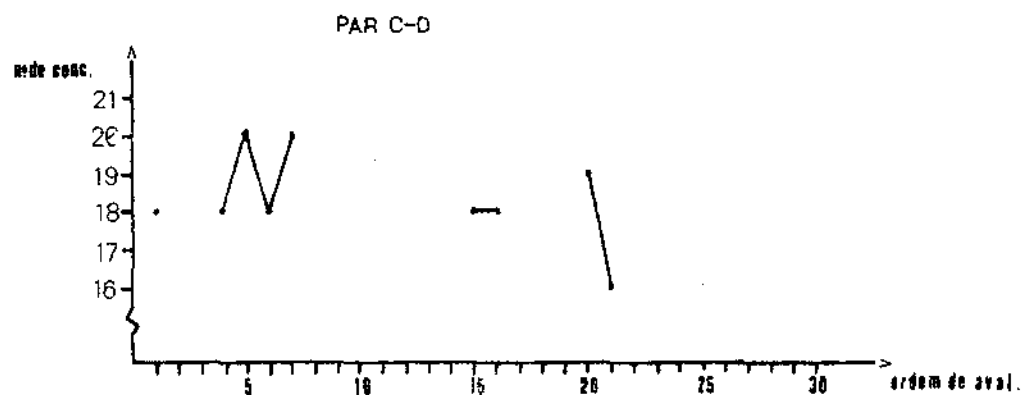
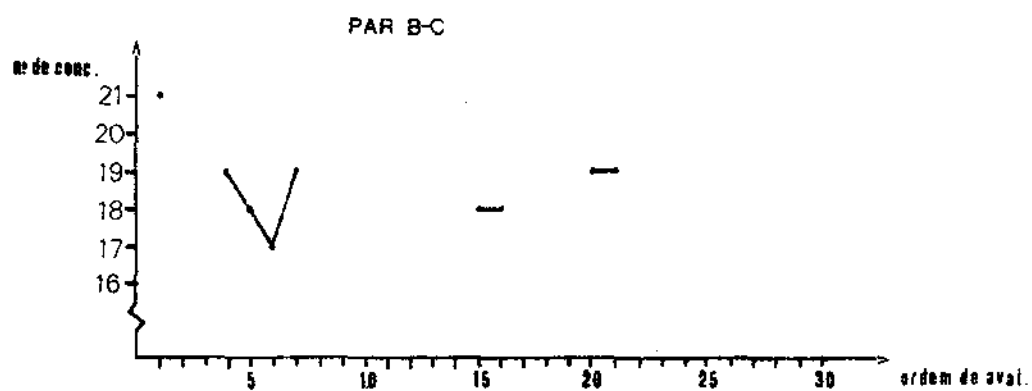
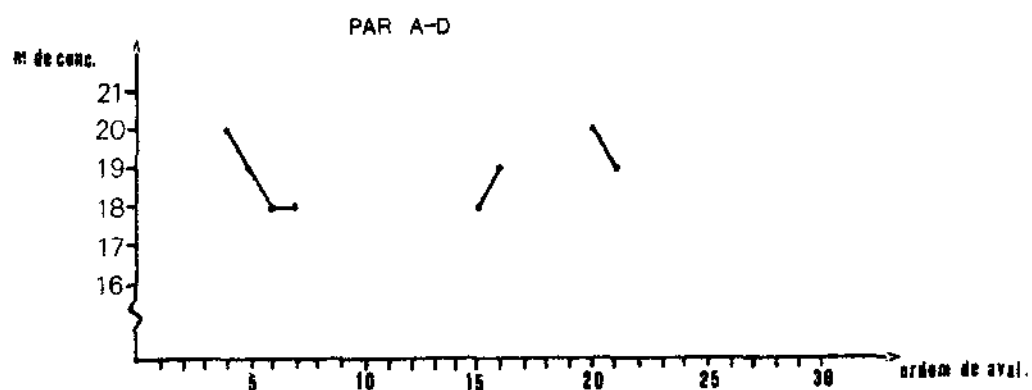


TABELA 2 - DISTRIBUIÇÃO DAS AVALIAÇÕES, EM PORCENTAGEM, SEGUNDO O NÚMERO DE CONCORDÂNCIAS ENTRE OS EXAMINADORES NO REGISTRO DOS ESCORES, POR PAR DE EXAMINADORES.

P A R D E E X A M I N A D O R E S								
A - C			A - B			B - C		
Nº DE CONC.	Nº DE AVAL.	PORCEN TUAL	Nº DE CONC.	Nº DE AVAL.	PORCEN TUAL	Nº DE CONC.	Nº DE AVAL.	PORCEN TUAL
16	0	0,00	16	1	3,45	16	0	0,00
17	0	0,00	17	1	3,45	17	3	10,00
18	2	6,90	18	4	13,79	18	8	26,67
19	8	27,59	19	2	6,90	19	6	20,00
20	9	31,03	20	9	31,03	20	4	13,33
21	10	34,48	21	12	41,38	21	9	30,00
TOTAL	29	100,00	29	100,00		30	100,00	

P A R D E E X A M I N A D O R E S								
A - D			B - D			C - D		
Nº DE CONC.	Nº DE AVAL.	PORCEN TUAL	Nº DE CONC.	Nº DE AVAL.	PORCEN TUAL	Nº DE CONC.	Nº DE AVAL.	PORCEN TUAL
16	0	0,00	16	0	0,00	16	1	11,11
17	0	0,00	17	1	11,11	17	0	0,00
18	3	37,50	18	3	33,33	18	5	55,55
19	2	25,00	19	3	33,33	19	1	11,11
20	3	37,50	20	1	11,11	20	2	22,22
21	0	0,00	21	1	11,11	21	0	0,00
TOTAL	8	100,00	9	99,99		9	99,99	

gistro dos escores das avaliações em que participaram e frequência relativa, em porcentagem, segundo o critério de desempenho avaliado e o par de examinadores. É importante ressaltar que o número de discordâncias, entre pares de examinadores, a exemplo do número de concordâncias, também é um indicador do grau de calibração entre esses avaliadores; entretanto, esse indicador é mais refinado que o de número de concordâncias. Assim, a Tabela 3 foi construída com a finalidade precípua de verificar o grau de calibração inter-examinadores no registro dos escores de cada critério de desempenho avaliado.

Para isso, os números de discordâncias entre os pares de examinadores, nos vários critérios de desempenho avaliados, foram transformados em percentuais considerando-se o número de avaliações em que esses pares participaram.

Pela análise da tabela 3, pode-se observar que são poucos os percentuais acima de 20%, para os diversos critérios de desempenho de todos os pares de examinadores.

Isto sugere que, na avaliação dos vários critérios ou erros críticos, houve baixo grau de discordâncias inter-examinadores no registro dos respectivos escores. Pelo raciocínio inverso, deduz-se que houve alta calibração, inter-examinadores, no registro dos escores da maioria dos critérios de desempenho avaliados.

Ocorre, entretanto, que os poucos percentuais acima de 20% concentram-se em determinados critérios de desempenho avaliados. Um exemplo marcante disso é o erro crítico "excesso de material forrador"; nesse critério ocorreram percentuais de discordâncias, acima de 20%, entre quase todos os pares de examinadores (com exceção do par C - D). O mesmo ocorre, em escala decrescente, com os critérios de desempenho: "crista marginal enfraquecida" (entre os pares A - D, B - D e C - D), "cavidade pouco profunda" (entre os pares A - B, B - D e C - D), "superfície oclusal abaixada" (entre os pares A - D, B - D e C - D), "abertura excessiva" (entre os pares B - C e C - D), "má orientação das paredes cavitárias" (entre o par C - D), "cavidade profunda" (entre o par B - D) e, finalmente, "excesso de mercúrio" (entre o par C - D).

Observa-se, também, que, com exceção do erro crítico "excesso de material forrador", nos demais erros críticos,

TABELA 3 - NUMERO DE DISCORDÂNCIAS ENTRE OS EXAMINADORES NO REGISTRO
DOS ESCORES DAS AVALIAÇÕES EM QUE PARTICIPARAM E FREQUÊN-
CIA RELATIVA, EM PORCENTAGEM, SEGUNDO O CRITÉRIO DE
DESEMPENHO AVALIADO E O PAR DE EXAMINADORES.

CRITÉRIO DE DESEMPENHO	PAR DE EXAMINADORES					
	A - C	A - B	B - C	A - D	B - D	C - D
	Nº %	Nº %	Nº %	Nº %	Nº %	Nº %
ABERTURA EXCESSIVA	4 13,79	3 10,34	8 26,67	0 0,00	1 11,11	3 33,33
ABERTURA INSUFICIENTE	1 3,45	0 0,00	1 3,33	0 0,00	0 0,00	0 0,00
CRISTA MARG. ENFRAQUECIDA	3 10,34	1 3,45	4 13,33	3 37,50	3 33,33	3 33,33
CÚSPIDES ENFRAQUECIDAS	1 3,45	0 0,00	1 3,33	0 0,00	0 0,00	1 11,11
MÁ ORIENTAÇÃO DAS	1 3,45	1 3,45	2 6,66	1 12,50	1 11,11	2 22,22
PAREDES CAVITÁRIAS	1 3,45	0 0,00	1 3,33	0 0,00	0 0,00	0 0,00
CAVIDADE PROFUNDA	2 6,90	4 13,79	6 20,00	1 12,50	2 22,22	1 11,11
CAVIDADE POUCO PROFUNDA	5 17,24	6 20,69	5 16,67	1 12,50	3 33,33	3 33,33
CÁRIE REMANESCENTE	0 0,00	1 3,45	1 3,33	0 0,00	0 0,00	0 0,00
EXPOSIÇÃO MECÂNICA	0 0,00	0 0,00	0 0,00	0 0,00	0 0,00	0 0,00
ELEIÇÃO INADEQUADA	0 0,00	0 0,00	0 0,00	0 0,00	0 0,00	0 0,00
COB.INSUF.PAREDE PULPAR	1 3,45	1 3,45	0 0,00	1 12,50	1 11,11	1 11,11
EXCESSO MAT. FORRADOR	7 24,14	8 27,59	9 30,00	3 37,50	4 44,44	1 11,11
EXCESSO DE MERCÚRIO	1 3,45	3 10,34	4 13,33	2 25,00	1 11,11	2 22,22
FALTA DE MATERIAL	1 3,45	1 3,45	2 6,66	0 0,00	0 0,00	0 0,00
SUPERFÍCIE OCLUSAL	0 0,00	0 0,00	1 3,33	3 37,50	3 33,33	4 44,44
ABAIXADA	0 0,00	0 0,00	0 0,00	1 12,50	1 11,11	1 11,11
SUP. OCLUSAL ALTA	2 6,90	2 6,90	3 10,00	0 0,00	0 0,00	1 11,11
ANATOMIA OCLUSAL IN CORR.	0 0,00	2 6,90	2 6,66	0 0,00	0 0,00	0 0,00
CONTACTO OCLUSAL ALTO	0 0,00	1 3,45	1 3,33	0 0,00	0 0,00	0 0,00
REST. FRATURADA, PERDIDA	0 0,00	1 3,45	1 3,33	0 0,00	0 0,00	0 0,00

que tiveram três ou dois percentuais de discordância acima de 20%, houve a participação do avaliador D entre os pares discordantes.

As seis colunas verticais da Tabela 3 identificam as discordâncias, no registro dos escores, em todos os critérios de desempenho, e seus respectivos percentuais, por pares de examinadores que participaram das avaliações.

Pela análise dos percentuais, acima de 20%, dessas colunas pode-se verificar que o par mais discordante, nos vários erros críticos avaliados, foi o C - D (discordâncias em 6 critérios). Segue-se, em ordem decrescente do número dessas porcentagens de discordância, os pares B - D (em 5 critérios), A - D (em 4 critérios), A - B e B - C (em 2 critérios) e A - C (em 1 critério).

Reforça-se, novamente, a afirmativa de que a maioria dos critérios de desempenho avaliados proporcionaram alto grau de concordância, entre pares de examinadores, no registro dos respectivos escores. Com efeito, o maior número de porcentagens de discordâncias, acima de 20%, em 21 possíveis, foi em 6 critérios (28,57%) e ocorreu, apenas, com um par de examinadores (C - D).

Percebe-se, facilmente, a participação do avaliador D entre os pares discordantes nos maiores números de erros críticos avaliados (6, 5 e 4).

Para verificar se as definições ou indicadores de desempenho relativos aos critérios avaliados foram adequadamente descritos e utilizados, elaborou-se a Tabela 4. Nessa tabela distribuíram-se os alunos avaliados segundo a concordância unânime dos examinadores no registro dos escores ou a existência de pelo menos um examinador discordante, segundo o critério de desempenho avaliado. Dessa forma obtiveram-se, na primeira coluna numérica, os números de alunos avaliados em que houve concordância unânime dos examinadores no registro dos escores; na última coluna apresentam-se os respectivos percentuais de alunos em que ocorreram essas concordâncias unânimes.

A concordância unânime, dos examinadores, no registro dos escores dos erros críticos em julgamento é um bom indicador de que as definições dos respectivos critérios de desempenho são adequadas.

TABELA 4 - DISTRIBUIÇÃO DOS ALUNOS AVALIADOS SEGUNDO A CONCORDÂNCIA UNÂNIME DOS EXAMINADORES NO REGISTRO DOS ESCORES OU A EXISTÊNCIA DE PELO MENOS UM EXAMINADOR DISCORDANTE, SEGUNDO O CRITÉRIO DE DESEMPENHO AVALIADO.

CRITÉRIO DE DESEMPENHO	Nº DE ALU- NOS EM QUE HOVE CON- CORDÂNCIA UNÂNIME.	PELO MENOS UM EXAMINA- DOR DISCOR- DANTE.	PORCENTUAL DE ALUNOS EM QUE HOU- VE CONCOR- DÂNCIA UNÂ- NIME.
ABERTURA EXCESSIVA	22	8	73,33
ABERTURA INSUFICIENTE	28	2	93,33
CRISTA MARGINAL ENFRAQUECIDA	24	6	80,00
CÚSPIDES ENFRAQUECIDAS	29	1	96,67
MÁ ORIENTAÇÃO DAS	28	2	93,33
PAREDES CAVITÁRIAS	29	1	96,67
CAVIDADE PROFUNDA	24	6	80,00
CAVIDADE POUCO PROFUNDA	21	9	70,00
CÁRIE REMANESCENTE	29	1	96,67
EXPOSIÇÃO MECÂNICA	30	0	100,00
ELEIÇÃO INADEQUADA	30	0	100,00
COBERTURA INSUF. PAREDE PULPAR	28	2	93,33
EXCESSO DE MATERIAL FORRADOR	18	12	60,00
EXCESSO DE MERCÚRIO	25	5	83,33
FALTA DE MATERIAL	28	2	93,33
SUPERFÍCIE OCLUSAL	26	4	86,67
ABAIXADA	29	1	96,67
SUPERFÍCIE OCLUSAL ALTA	26	4	86,67
ANATOMIA OCLUSAL INCORRETA	28	2	93,33
CONTACTO OCLUSAL ALTO	29	1	96,67
RESTAURAÇÃO FRATURADA, PERDIDA	29	1	96,67

A observação dos percentuais da última coluna da Tabela 4, permite deduzir que houve concordância unânime, dos examinadores, no julgamento da grande maioria dos critérios de desempenho avaliados; com efeito, os percentuais de alunos em que houve concordância unânime, dos examinadores, no registro dos escores foram, em sua quase totalidade, iguais ou superiores a 80%, para os vários critérios de desempenho. Constituíram-se exceções, apenas, os critérios: "abertura excessiva" (73,33%), "cavidade pouco profunda" (70,00%) e "excesso de material forrador" (60,00%).

Pode-se, então, inferir que a grande maioria dos critérios avaliados tiveram indicadores de desempenho bem definidos e perfeitamente mensuráveis.

Para se verificar quais os erros críticos mais freqüentes, bem como quais os critérios de desempenho que devem merecer maior atenção, tanto dos examinadores como dos alunos, durante o processo avaliativo da função clínica em análise, elaborou-se a Tabela 5.

Nessa tabela fez-se a distribuição dos alunos, para os quais houve concordância unânime dos examinadores, segundo o registro ou não dos escores e o critério de desempenho avaliado.

Assim, na primeira coluna numérica vê-se o número de alunos avaliados para os quais houve concordância unânime, dos examinadores, no registro dos escores nos critérios de desempenho correspondentes. Verifica-se que para o erro crítico "cavidade pouco profunda" houve o registro unânime dos escores para o maior número de alunos dentre os 30 avaliados (12); sua respectiva porcentagem, registrada na terceira coluna numérica, foi de 40,00%. Isto indica, sem dúvida, que esse erro crítico foi o de ocorrência mais freqüente dentre os detectados durante o desempenho da função clínica avaliada.

Do mesmo modo, em ordem decrescente de freqüência de ocorrência, os erros "excesso de material forrador" e "cobertura insuficiente da parede pulpar" tiveram, respectivamente, registro unânime dos escores para 7 e 6 dentre todos os alunos avaliados (23,33% e 20,00%). As porcentagens, iguais ou superiores a 20%, indicam, também, a ocorrência bastante freqüente dos erros críticos mencionados.

TABELA 5 - DISTRIBUIÇÃO DOS ALUNOS AVALIADOS, PARA OS QUAIS HOUE CONCORDÂNCIA UNÂNIME DOS EXAMINADORES, SEGUNDO O REGISTRO OU NÃO DOS ESCORES E O CRITÉRIO DE DESEMPENHO AVALIADO.

CRITÉRIO DE DESEMPENHO	Nº DE ALUNOS, PARA OS QUAIS HOUE CONCORDÂNCIA UNÂNIME, SEGUNDO O REGISTRO OU NÃO DOS ESCORES.		PERCENTUAL DE ALUNOS EM QUE HOUE OU NÃO REGISTRO NO ESCORE.	
	SIM	NÃO	SIM	NÃO
ABERTURA EXCESSIVA	1	21	3,33	70,00
ABERTURA INSUFICIENTE	0	28	0,00	93,33*
CRISTA MARGINAL ENFRAQUECIDA	4	20	13,33	66,67
CÚSPIDES ENFRAQUECIDAS	0	29	0,00	96,67*
MÁ ORIENTAÇÃO DAS	0	28	0,00	93,33*
PAREDES CAVITÁRIAS	0	29	0,00	96,67*
CAVIDADE PROFUNDA	0	24	0,00	80,00*
CAVIDADE POUCO PROFUNDA	12	9	40,00	30,00
CÁRIE REMANESCENTE	1	28	3,33	93,33*
EXPOSIÇÃO MECÂNICA	1	29	3,33	96,67*
ELEIÇÃO INADEQUADA	0	30	0,00	100,00*
COBERTURA INSUF. PAREDE PULPAR	6	22	20,00	73,33
EXCESSO DE MATERIAL FORRADOR	7	11	23,33	36,67
EXCESSO DE MERCÚRIO	5	20	16,67	66,67
FALTA DE MATERIAL	0	28	0,00	93,33*
SUPERFÍCIE OCLUSAL	2	24	6,67	80,00*
ABAIXADA	0	29	0,00	96,67*
SUPERFÍCIE OCLUSAL ALTA	2	24	6,67	80,00*
ANATOMIA OCLUSAL INCORRETA	1	27	3,33	90,00*
CONTACTO OCLUSAL ALTO	0	29	0,00	96,67*
RESTAURAÇÃO FRATURADA, PERDIDA	0	29	0,00	96,67*

Os três erros críticos citados, de ocorrência mais freqüente durante o desempenho da função clínica avaliada, devem merecer, por parte de examinadores e alunos, atenção especial ao longo do processo avaliativo.

As percentagens iguais ou superiores a 80% (registradas e assinaladas com asteriscos na quarta coluna numérica da Tabela 5), de alunos em que não houve registro dos escores, revelam os acertos mais freqüentes dentre todos os critérios de desempenho avaliados. Esses acertos, ou não incidência em erros críticos, ocorreram com freqüências variáveis, segundo a grandeza dos percentuais registrados; representam, também, o consenso unânime dos examinadores quanto a sua ocorrência.

Os percentuais inferiores a 80%, não assinalados com asteriscos, mostram, em valores variáveis, as freqüências mais baixas de acertos detectados, unânimemente, pelos examinadores, nos critérios avaliados correspondentes. Esses critérios devem merecer também atenção especial, de examinadores e alunos, durante as avaliações.

Com a finalidade de se avaliar a validade da atribuição de pesos negativos diferentes aos escores e a algumas considerações, sobre as notas finais dos alunos, organizou-se a Tabela 6. Nela vê-se a nota final, de cada aluno, segundo o examinador que a atribuiu e a respectiva amplitude.

Verifica-se que as notas finais imputadas, pelos vários examinadores, ao mesmo aluno são homogêneas porquanto sua amplitude (diferença entre o maior e o menor valor) foi pequena; com efeito, os maiores valores de amplitude foram 2,5 e 2,0 (para os alunos 6 e 5) que ocorreram, cada um, apenas uma única vez. Duas notas (dos alunos 15 e 20) tiveram amplitudes de 1,5. As demais notas atribuídas tiveram amplitudes iguais ou inferiores a 1,0; sua maioria, entretanto, possuía amplitudes de 0,5 e 0,0.

Alguns resultados podem ser extraídos da homogeneidade na concessão de notas finais, ao mesmo aluno, pelos vários examinadores.

Primeiramente deve-se inferir que a atribuição de pesos negativos diferentes aos escores e a algumas considerações permitiu conferir-se notas, de zero a dez, aos alunos, con-

TABELA 6 - NOTA FINAL, DE CADA ALUNO, SEGUNDO O EXAMINADOR E A RESPECTIVA AMPLITUDE.

ALUNO AVALIADO	EXAMINADOR				AMPLITUDE
	A	B	C	D	
01	---	9,5	9,0	9,5	0,5
02	8,0	8,0	7,0	---	1,0
03	9,0	9,0	9,5	---	0,5
04	3,5	4,0	4,5	4,0	1,0
05	7,0	6,5	8,5	8,0	2,0
06	8,0	8,5	9,0	6,5	2,5
07	7,5	7,5	6,5	7,0	1,0
08	6,5	7,0	7,0	---	0,5
09	9,5	9,0	9,0	---	0,5
10	7,0	6,5	6,5	---	0,5
11	7,0	6,5	7,0	---	0,5
12	9,0	9,0	9,0	---	0,0
13	8,5	8,5	7,5	---	1,0
14	8,0	8,0	8,0	---	0,0
15	8,5	8,5	8,5	7,0	1,5
16	8,0	8,5	8,0	8,0	0,5
17	9,0	9,0	9,0	---	0,0
18	7,5	7,5	7,5	---	0,0
19	8,0	8,0	8,0	---	0,0
20	6,0	6,0	7,5	6,0	1,5
21	7,0	7,0	6,0	7,0	1,0
22	7,0	7,5	7,0	---	0,5
23	5,0	4,5	5,0	---	0,5
24	9,0	9,0	9,0	---	0,0
25	8,0	8,0	8,0	---	0,0
26	6,5	6,5	6,5	---	0,0
27	9,0	9,5	9,0	---	0,5
28	7,0	7,0	7,0	---	0,0
29	9,0	9,0	8,5	---	0,5
30	7,0	7,5	7,0	---	0,5

forme o sistema vigente no ensino brasileiro.

Deduz-se, também, que essa atribuição de pesos mencionada proporcionou alta calibração, inter-examinadores, na concessão das notas finais aos alunos avaliados.

Por último, pode-se afirmar que as diferenças de valor entre as notas finais dadas aos vários alunos avaliados mostram a variação de desempenho clínico entre os mesmos.

O tempo utilizado, pelos examinadores, na realização de cada avaliação foi registrado em local apropriado do Formulário "B" (Formulário clínico) pelo Administrador do Teste, como foi descrito anteriormente.

A análise dos dados obtidos de todos os Formulários "B", utilizados nas 30 avaliações efetuadas, mostrou que foram registrados, com maior frequência, tempos em torno de 20 minutos.

5. DISCUSSÃO

A análise dos resultados apresentados na Tabela 1 permitem, inicialmente, constatar que os coeficientes de confiança inter-examinadores, por aluno avaliado, são altos; com efeito, a variação desses coeficientes está entre o valor mínimo de 76,19% e o valor máximo de 100%. Entretanto, o valor mais freqüente, dentre todos os obtidos, foi 100% (para 21 concordâncias inter-examinadores) e ocorreu 32 vezes (entre os 6 pares de examinadores formados para fins estatísticos) em 114 vezes possíveis (somando-se os totais de todas as avaliações em que participaram os 6 pares formados).

Esses coeficientes elevados sugerem alto grau de consistência, entre os pares de examinadores, no julgamento dos mesmos critérios de desempenho de cada aluno avaliado.

As médias dos coeficientes de confiança inter-examinadores, calculadas sobre os coeficientes alcançados pelos pares de examinadores em todas as avaliações em que participaram, também foram altas, isto é, acima de 87,30%. Esses resultados permitem afirmar que houve alto grau de consistência, entre os pares de avaliadores, também no julgamento de todos os critérios avaliados.

É interessante observar que o examinador D fazia sempre parte dos pares de avaliadores que obtiveram as médias dos coeficientes de confiança mais baixas. Tal fato sugere, à primeira vista, falhas no critério avaliativo do mencionado examinador. Entretanto essa diferença, em relação aos demais pares de examinadores, talvez não se deva ao critério do avaliador D mas sim às dificuldades do processo avaliativo das avaliações em que participou. Realmente, pode-se notar, pela análise dos dados da Tabela 1, que os coeficientes de confiança mais baixos, com raras exceções, também foram obtidos pelos demais pares de examinadores, nas mesmas avaliações em que participou o citado examinador.

Da observação da Tabela 1 e dos curvogramas apresentados nas Figuras 1 e 2, pode-se depreender que o treinamento dos examinadores, ao longo das avaliações, não aumentou os coeficientes de confiança inter-avaliadores. Esse fato pode ser inferido porquanto os coeficientes obtidos apresentaram uma tendência de ocorrência irregular, mas relativamente estável, em função

da ordem em que foram feitas as avaliações.

Esses resultados discordam das afirmativas de FITZPATRICK & MORRISON ¹¹ que consideram que o treinamento é essencial para a avaliação de amostras de trabalho e aumentam os coeficientes intra e inter-examinadores.

Do mesmo modo, diferem dos resultados obtidos por ABOU-RASS ² que demonstrou que tanto a uniformidade global quanto a calibração nos critérios individuais, inter-examinadores, aumentaram como consequência do treinamento; também divergem dos dados de NATKIN & GUILD ²³ que reportaram um aumento estatisticamente significativo dos coeficientes de confiança, após a realização de oito sessões de treinamento.

É importante salientar que o tipo e extensão dos treinamentos reportados por esses autores diferem do utilizado no presente trabalho, o que não invalida, a nosso ver, as comparações discutidas.

Contrariamente, FULLER ¹², relatando o exame mais completo do efeito do treinamento sobre os coeficientes de confiança, concluiu que o modelo utilizando-se "checklist" e treinamento não provocou aumento significativo dos coeficientes inter-avaliadores, comparativamente ao modelo usando-se somente "checklist".

De maneira semelhante, HOUPPT & KRESS ¹⁶ e HINKELMAN & LONG ¹⁵ constataram que o treinamento não teve efeito significativo sobre os coeficientes de confiança inter-examinadores, embora os últimos autores citados não tenham definido o tipo e extensão do treinamento utilizado.

Nossos resultados, mesmo em condições experimentais diferentes, corroboram os achados de FULLER ¹², HOUPPT & KRESS ¹⁶ e HINKELMAN & LONG ¹⁵.

Os resultados decorrentes da análise da Tabela 2 revelaram alta calibração ou uniformidade, entre os examinadores, no registro dos escores, durante a aplicação do Instrumento de avaliação em estudo. Esses resultados foram inferidos da concentração da distribuição das avaliações, em que houve concordância entre os examinadores, em seus valores mais altos (de 18 a 21 concordâncias, em 21 possíveis).

HOUPPT & KRESS ¹⁶, comparando o uso de uma escala de dois pontos com a utilização de uma escala de cinco pontos para a avaliação de oito critérios de preparos cavitários, encontraram também altos graus de calibração intra e inter-avaliadores quando se usou a escala de dois pontos.

Do mesmo modo, RYGE & SNYDER ²⁶ utilizaram uma escala de quatro pontos para a avaliação clínica de restaurações e obtiveram 92% de concordância inter-avaliadores e 89% de concordância intra-examinadores (baseada no reexame de uma amostra de 10% das restaurações realizadas).

É importante frisar que, no presente estudo, não foi possível estabelecer-se os coeficientes de confiança intra-examinadores (consistência dos avaliadores consigo mesmos) em razão das características das avaliações, realizadas por passos de tarefas. É evidente que os alunos não poderiam repetir os mesmos passos da função clínica já concluída, para que se pudesse reavaliar a consistência, intra-examinadores, no julgamento dos critérios avaliados.

No estudo reportado por HINKELMAN & LONG ¹⁵ comparou-se o uso de uma escala de três pontos com a utilização de uma escala de dois pontos, na avaliação de cinco critérios do preparo cavitário para blocos metálicos; concluiu-se que ambas as escalas de pontos conferiram altos coeficientes de confiança inter-avaliadores, à semelhança do que ocorreu na presente pesquisa.

Na mesma Tabela 2, para os pares A - D, B - D e C - D, dos quais fazia parte o examinador D, os números de concordâncias, inter-examinadores, foram um pouco menores que os dos demais pares.

Esses resultados podem sugerir que o examinador D não estava tão calibrado quanto os demais; essa pretensa falta de calibração poderia ser atribuída ao fato de que o mesmo participou de um menor número de avaliações. Todavia, como se discutiu anteriormente, esse examinador, coincidentemente, participou somente das avaliações cujos erros críticos foram mais difíceis de se detectar; isto pode ser deduzido porquanto, nessas avaliações, os demais pares de avaliadores também obtiveram, com raras exceções, os coeficientes de confiança mais baixos.

Na Tabela 3, o baixo grau de discordâncias inter-examinadores, no registro dos escores dos vários erros críticos avaliados, mostraram que houve alta calibração entre os mesmos no julgamento da maioria dos critérios de desempenho avaliados. Saliente-se que o número de discordâncias, entre pares de examinadores, é um indicador bem refinado do grau de calibração no registro dos escores de cada critério de desempenho avaliado.

Desses resultados deve-se inferir que a maioria dos critérios de desempenho ou erros críticos foram identificados e selecionados com propriedade; pode-se deduzir, ainda, que tais critérios foram bem definidos e objetivos, por permitirem bom consenso avaliativo de todos os examinadores.

HOUPPT & KRESS ¹⁶ compararam a acuracidade e os coeficientes de confiança intra e inter-examinadores na avaliação de restaurações realizadas por grupos que se utilizaram de uma escala de cinco pontos para o julgamento de cada um dos oito critérios selecionados; essas avaliações foram feitas tanto sem definição dos critérios de desempenho, para cada um dos cinco pontos utilizados, como com critérios bem definidos. Concluíram que os coeficientes de confiança inter-avaliadores, quando medidos nos critérios individuais, aumentaram significativamente com a utilização de critérios de desempenho bem definidos.

Quanto a objetividade dos critérios de desempenho GAINES e colab. ¹³, utilizando escalas subjetiva e objetiva para a avaliação de escultura em cera, demonstraram discordâncias significativas inter-examinadores, quando utilizaram a escala subjetiva; com o uso da escala objetiva aumentaram os coeficientes de confiança inter-avaliadores em 30%.

Na identificação e seleção dos critérios de desempenho utilizados procurou-se seguir, dentro do possível, os meios identificados por ABELES ¹; esses meios empregados foram: a análise dos livros de texto usados no ensino do comportamento, fundamentação dos itens de avaliação numa estrutura teórica pré-existente e análise e revisão dos critérios por dois especialistas da área de dentística operatória. Nessa identificação dos critérios de desempenho usados seguiram-se, também, as linhas mestras propostas por MACKENZIE ²⁰.

A decisão de se avaliar os erros críticos, confi-

gurados pelos critérios de desempenho, foi tomada, no presente trabalho, em virtude de se inferir que, a esse nível de aprendizagem (7º semestre), os alunos acertam mais do que erram, no desempenho clínico da função avaliada. Essa decisão pareceu-nos acertada porquanto os coeficientes de confiança e o grau de calibração, inter-examinadores, obtidos foram altos. Pode ser que em níveis de aprendizagem mais precoces seja preferível avaliar os acertos dos alunos, no desempenho da função clínica examinada.

NATKIN & GUILD ²³ reportaram um sistema de atribuição de escores utilizando erros críticos para avaliar desempenhos em endodontia. Nesse sistema os escores (número e tipo de erros cometidos), que expressavam a variação da severidade do erro praticado, foram convertidos em letras que significavam o grau ou nota do aluno.

Concluíram que houve alto coeficiente de confiança inter-avaliadores, inferido pela ocorrência de um aumento estatisticamente significativo no percentual de vezes nas quais todos os graus registrados estavam entre duas letras.

Os mesmos autores, entretanto, afirmam que esse sistema de atribuição de escores limita a possibilidade dos avaliadores na variação do registro dos mesmos; acrescentam, ainda, que os tipos de erros que o aluno pode cometer, durante o desempenho clínico, foram listados para beneficiar a tarefa avaliativa dos examinadores.

Concordamos com a primeira afirmativa desses autores. Entretanto, no presente trabalho, parece-nos que a referência à limitação dos examinadores na variabilidade do registro dos escores, para cada erro crítico avaliado, não exerceu influência negativa na aplicação do sistema empregado. Com efeito, o sistema analítico usado compunha-se de uma escala de quatro escores (C, Ic, If e F) no registro do julgamento dos diversos erros críticos avaliados. Para o registro do julgamento de cada erro crítico, entretanto, utilizou-se, sempre, um ou dois desses escores, ou seja, um para cada definição de desempenho do mesmo.

Deste modo, a maioria dos erros críticos atribuiu-se somente um escore, porquanto aos mesmos se enquadrava apenas uma definição. A alguns critérios, que admitiam dois indicadores de desempenho, atribuíram-se dois escores, isto é, um para

cada indicador.

Quando o aluno não comete erro, o procedimento é considerado satisfatório e não há registro de escores.

O objetivo da avaliação, na presente pesquisa, foi o de diagnosticar os erros dos estudantes, como indicador do desempenho clínico final dos alunos de graduação, a esse nível de aprendizagem (7º semestre). Pode-se inferir, então, que a finalidade do Instrumento de avaliação em estudo é a detecção da ocorrência ou não de erros no desempenho clínico, do aluno, da função avaliada.

As nuances dos erros ou acertos dos alunos, durante a função clínica avaliada, carece de importância a esse nível de aprendizagem; todavia pode-se verificar que os indicadores do Instrumento utilizado fornecem descrições avaliativas de desempenhos insatisfatórios que permitem aos alunos, por dedução inversa, desempenhos clínicos de alta qualidade técnica. A níveis de aprendizagem mais precoces devem ser avaliadas as minúcias do desempenho pré-clínico ou clínico dos estudantes, através do diagnóstico dos acertos dos alunos, como já discutimos.

A segunda afirmativa de NATKIN & GUILD ²³ é bastante discutível; os tipos de erros críticos listados, no Instrumento em estudo, foram identificados e selecionados através de meios criteriosos, como já discutimos anteriormente, que proporcionaram resultados satisfatórios tanto no consenso avaliativo dos examinadores, quando no desempenho dos alunos avaliados. Isto pode ser deduzido porquanto, nas 30 avaliações realizadas, não ocorreram erros que não estivessem identificados pelas definições dos critérios de desempenho listados. Acrescente-se, ademais, que o número de erros críticos listados, em qualquer Instrumento de avaliação, não é imutável. Quando um determinado critério de desempenho ou erro crítico não ocorre, num grande número de avaliações, pode ser eliminado; do mesmo modo, quando ocorre determinado erro crítico não listado, mesmo em baixa frequência, deve ser incluído entre os critérios de desempenho a avaliar.

Dos resultados obtidos pela análise da Tabela 3 verifica-se, ainda, que os poucos percentuais de discordâncias, inter-examinadores, acima de 20% concentram-se em determinados erros críticos avaliados. Isto significa que certos critérios de

desempenho, entre os quais "excesso de material forrador" é um exemplo marcante, provocam maiores discordâncias entre pares de examinadores que os demais.

HUNTER ¹⁷, ao concluir que 80% de calibração inter-examinadores, no julgamento dos vários critérios de desempenho avaliados, é considerada ótima, também afirma que alguns examinadores são mais flexíveis que outros, ao avaliarem alguns critérios individuais e que certos critérios de desempenho proporcionam maior calibração que outros. Sugere, outrossim, quando houver grandes discordâncias inter-avaliadores, a modificação do Teste, reduzindo-se o número de escores e categorias de graus empregados, treinando-se os examinadores e/ou revisando-se os indicadores de desempenho.

O maior grau de discordâncias, inter-examinadores, no julgamento do desempenho nos critérios que obtiveram percentuais acima de 20%, pode ser atribuído a uma ou mais das seguintes variáveis:

1. Os indicadores de desempenho dos respectivos critérios não estão bem definidos e/ou são pouco objetivos, permitindo interpretações avaliativas diferentes dos examinadores.

2. Existem preconceitos dos examinadores na avaliação desses critérios; geralmente são preconceitos motivados pela experiência profissional dos avaliadores, que julgam certas minúcias das definições insignificantes do ponto de vista clínico ou fazem com que os mesmos não concordem plenamente com os parâmetros propostos pelos indicadores.

3. Falta de acuracidade de um ou de ambos avaliadores no julgamento dos critérios.

4. Excesso de rigor avaliativo de um ou de ambos os examinadores no julgamento dos critérios.

5. Certos erros críticos são mais difíceis de se detectar.

Com referência ao erro crítico "excesso de material forrador", que provocou o maior percentual de discordâncias entre os pares de examinadores, podemos afirmar, quase com certeza, que houve excesso de rigor avaliativo de alguns avaliadores, contrastando com a falta de acuracidade de outros; com efeito, na avaliação desse critério, alguns examinadores assinalavam o

respectivo escore mesmo que houvesse, apenas, um único ponto das paredes cavitárias com material forrador. Por outro lado, outros examinadores não assinalavam o escore, por julgarem a aludida ocorrência insignificante do ponto de vista clínico (preconceito).

Quanto aos demais erros críticos, que tiveram três ou dois percentuais de discordâncias acima de 20%, em cujas avaliações participou o examinador D, pode-se discutir o que se segue.

Com relação ao critério "crista marginal enfraquecida", parece-nos que houve excesso de rigor avaliativo do examinador D no julgamento do mesmo, porquanto houve discordâncias com todos os demais avaliadores.

Nos critérios "cavidade pouco profunda" e "abertura excessiva" os examinadores foram discordantes na utilização do meio de mensuração (broca 57); talvez esse meio de mensuração fosse pouco exato, permitindo mensurações discrepantes.

Já no erro crítico "superfície oclusal abaixada" ocorreu rigor avaliativo do examinador D; nesse caso, especificamente, houve falta de acuracidade dos demais avaliadores. A utilização do indicador de desempenho, no julgamento desse critério, mostrou-se pouco prática.

Quanto ao erro "excesso de mercúrio" certamente houve falta de acuracidade dos examinadores, componentes dos pares discordantes, porquanto trata-se de um erro de fácil constatação.

A ocorrência das mencionadas discrepâncias no julgamento dos critérios de desempenho citados pode ser constatada durante as reuniões, entre os examinadores, realizadas após algumas sessões de avaliação.

Dos resultados obtidos do estudo das seis colunas verticais da Tabela 3 pode-se reforçar a afirmativa de que a maioria dos critérios de desempenho avaliados proporcionaram alto grau de concordância, por par de examinadores, no registro dos respectivos escores. Isto se deduz porque o maior número de percentagens de discordâncias, acima de 20%, em 21 possíveis, foi em 6 critérios (28,57%) e ocorreu, apenas, com um par de examinadores (C - D). Esse estudo também identifica os pares de

avaliadores mais discordantes, nos vários critérios de desempenho avaliados, tendo-se verificado, sempre, a participação do avaliador D entre esses pares.

Pode-se inferir que o examinador D, via de regra, agiu com rigor avaliativo mais aguçado que os demais avaliadores. Os demais fatores que devem ter contribuído para tais discordâncias já foram discutidos anteriormente.

Pelos resultados extraídos da Tabela 4, verifica-se que a grande maioria dos critérios avaliados tiveram indicadores de desempenho adequadamente definidos e perfeitamente mensuráveis; isto se depreende porque os percentuais de alunos em que houve concordância unânime, dos examinadores, no registro dos escores foram, em sua quase totalidade, iguais ou superiores a 80%, para os vários critérios de desempenho avaliados.

Por indicadores adequadamente definidos deve-se entender descrições avaliativas cuja clareza de redação e facilidade de interpretação permitem, aos alunos e examinadores, julgarem, por comparação, a ocorrência ou não de erros críticos no desempenho da função clínica avaliada; esses indicadores, portanto, fornecem padrões precisos de desempenhos insatisfatórios aos examinadores, orientando-os quando detectar os erros praticados, e aos alunos, informando-os quando e porque estão cometendo erros.

No estudo de avaliação clínica de RYDE & SNYDER ²⁶, já discutido, embora não se tenha variado o número de pontos da escala utilizada, os quatro escores usados também foram, operacionalmente, bem definidos.

Por indicadores perfeitamente mensuráveis deve-se entender descrições avaliativas em que se utilizaram, de preferência, instrumentos e/ou materiais odontológicos comuns como meios de mensuração; esses indicadores usaram, por vezes, descrições de características visuais, quando nenhum meio instrumental e/ou material foram apropriados.

A utilização de indicadores de desempenho mensuráveis é recomendada por HUNTER ¹⁷, ao descrever o seu teste de desempenho clínico (PEG).

Esses indicadores de desempenho definidos e mensu

ráveis, obtidos na presente pesquisa, melhoram a uniformidade de julgamento avaliativo dos examinadores e o desempenho e o moral dos alunos; também fornecem retro-informações, aos estudantes, no sentido de incrementarem progressos em seus desempenhos, além de ajudarem os professores a melhorar a qualidade do ensino. Essas afirmativas podem ser inferidas pela comparação com o trabalho de HUNTER ¹⁷.

Pode-se, então, deduzir que a aplicação do Instrumento em análise, para avaliar o desempenho clínico final de alunos do 7º semestre de graduação em restaurações Classe I com amálgama, teve grande utilidade instrucional.

Em relação aos alunos, essa utilidade instrucional foi detectada de algumas formas.

Em primeiro lugar, os padrões de desempenho insatisfatórios precisos, fornecidos pelas descrições dos indicadores, orientavam os alunos no sentido de evitarem os respectivos erros críticos na execução da função clínica avaliada; essa orientação já é um importante meio de retro-informação instrucional para o estudante.

Em segundo lugar, a orientação dada pelos examinadores durante as avaliações, sobre quais os erros praticados e de como executarem as devidas correções, forneciam aos alunos subsídios instrucionais que certamente incrementaram progressos em seus desempenhos.

Finalmente, as discussões entre examinadores e alunos das falhas clínicas cometidas durante a avaliação, imediatamente após o encerramento do Teste, constituíram-se em excelente veículo de retro-informação instrucional aos estudantes.

Dessa forma, o sistema de avaliação utilizado no presente trabalho permite diagnosticar a retro-informação oferecida, pelo mesmo, aos estudantes; portanto, se nos basearmos em MACKENZIE ²⁰, tal sistema também tem, por objetivo, a aprendizagem dos alunos. Atende, outrossim, a sugestão de EBEL ⁸, de que o sistema analítico fornece maior retro-informação aos estudantes que o global, além de compelir o avaliador a observar fatores específicos do desempenho dos mesmos.

Da mesma maneira, HOUPPT & KRESS ¹⁶ e HINKELMAN &

LONG ¹⁵ sugerem a importância da retro-informação ao aluno no sentido de incrementar melhoramentos em seu desempenho, ao afirmarem que os sistemas de avaliação que se utilizam de uma escala de dois pontos tem desvantagens instrucionais. Essas desvantagens seriam a ausência dessas retro-informações.

Em relação aos examinadores, a utilidade instrucional dos indicadores de desempenho, empregados na avaliação dos alunos, configurou-se em julgamentos mais criteriosos, o que deve determinar melhorias na qualidade do ensino. Tornou-se evidente que tanto a aplicação das descrições avaliativas dos indicadores, quanto as discussões entre os examinadores, pré e pós-avaliações, das características desses indicadores, incrementaram as referidas melhorias.

Com relação aos poucos erros críticos que se constituíram em exceções aos percentuais acima de 80% de concordância unânime dos examinadores, no registro dos respectivos escores, deve-se discutir o que se segue.

Nos critérios "abertura excessiva" e "cavidade profunda", como já vimos, o meio de mensuração utilizado (broca 57), nas definições dos mesmos, talvez fosse pouco exato, permitindo medidas discordantes; esse meio de mensuração, pouco exato, pode conferir certo grau de subjetividade ao indicador de desempenho que, por sua vez, provoca subjetivismos no julgamento avaliativo de examinadores e alunos.

Com respeito ao erro crítico "abertura excessiva" deve-se acrescentar que certamente também houve preconceito avaliativo, dos examinadores, no julgamento do referido critério. Esse preconceito traduz-se pela não aceitação plena dos parâmetros oferecidos pelo indicador de desempenho, que podem conflitar com o senso clínico dos avaliadores.

Quanto ao critério "excesso de material forrador" as razões das discordâncias, entre os examinadores, no julgamento do mesmo já foram anteriormente discutidas.

Dos resultados obtidos na análise da Tabela 5, conclui-se que os erros mais freqüentes, dentre os cometidos pelos alunos durante as avaliações, foram, em ordem decrescente de ocorrência, "cavidade pouco profunda", "excesso de material for-

rador" e "cobertura insuficiente da parede pulpar"; a esses erros críticos corresponderam, respectivamente, percentagens de 40,00%, 23,33% e 20,00% de alunos avaliados para os quais houve concordância unânime, dos examinadores, no registro dos escores.

O erro crítico "cavidade pouco profunda" foi o de ocorrência mais freqüente provavelmente pelas seguintes razões:

1. O aluno tinha receio, compreensível, de aprofundar a cavidade preparada em demasia, temendo provocar exposição pulpar do dente.
2. Os padrões oferecidos pelo indicador de desempenho, correspondente ao erro crítico avaliado, poderiam divergir dos parâmetros mentalizados pelo senso clínico de alunos e examinadores.
3. O meio de mensuração utilizado na descrição do indicador de desempenho permitia mensurações discrepantes entre os examinadores e alunos avaliados.
4. Excesso de rigor avaliativo, dos examinadores, no julgamento do critério.

A ocorrência também freqüente, embora em menor escala que o anterior, dos erros críticos "excesso de material forrador" e "cobertura insuficiente da parede pulpar" deveu-se, provavelmente, aos seguintes fatores:

1. Grau de dificuldade técnica maior para a execução clínica satisfatória, pelos alunos, desses passos da tarefa proteção pulpar (forramento).
2. Os padrões dos indicadores de desempenho, desses erros críticos, são bastante rigorosos.
3. Certas minúcias das definições desses critérios são consideradas insignificantes, do ponto de vista clínico, pelos alunos avaliados.
4. Excesso de rigor avaliativo, dos examinadores, no julgamento dos critérios.

Esses três erros críticos citados, de ocorrências mais freqüentes, devem merecer, por parte de examinadores e alunos, atenção especial ao longo do processo avaliativo de cavidades Classe I para amálgama.

Da mesma Tabela 5 pode-se extrair, também, os critérios de desempenho que tiveram, em frequências variáveis, maiores percentagens de acertos (iguais ou superiores a 80%) por parte dos alunos, através do consenso unânime dos examinadores.

O erro crítico "eleição inadequada" do material forrador não foi cometido por nenhum dos 30 alunos avaliados; talvez esse critério devesse ser eliminado da avaliação clínica dos alunos, a esse nível de aprendizado. Do mesmo modo, os erros críticos cujos percentuais foram acima de 90% poderiam, numa avaliação mais rápida, também serem eliminados.

Essas afirmativas confirmam a sugestão de HUNTER ¹⁷ para a eliminação de critérios que se aplicam raramente e, conseqüentemente, não são importantes.

Com relação aos critérios de desempenho que tiveram percentuais de acertos inferiores a 80%, poder-se-ia inferir que são erros de ocorrência também freqüente. Entretanto essa afirmativa seria temerária porquanto somente três deles ("cavidade pouco profunda", "cobertura insuficiente da parede pulpar" e "excesso de material forrador") tiveram percentagens de registros unânimes de seus escores acima de 20%; acrescente-se, ademais, que o critério "cobertura insuficiente da parede pulpar" teve percentagem de acerto relativamente alta (73,33%). O mesmo ocorreu com os critérios "crista marginal enfraquecida" e "excesso de mercúrio" que tiveram percentagens de acertos de 66,67%.

Já com o erro crítico "abertura excessiva" embora a percentagem de acertos não fosse baixa (70%), a ocorrência de erros manteve-se baixa (3,33%). Isto ocorreu em virtude das discordâncias, entre os examinadores, no registro dos escores de oito das avaliações realizadas.

Embora não se possa dizer, com exatidão, que os três últimos erros críticos mencionados foram de ocorrência freqüente, os mesmos devem merecer, também, atenção especial durante o processo avaliativo de cavidades Classe I para amálgama, por parte de alunos e examinadores.

A atribuição de pesos negativos diferentes aos escores e a algumas considerações mostrou-se válida para se conferir notas finais aos alunos avaliados, conforme se depreende da

observação da Tabela 6. Com efeito, as notas finais atribuídas pelos examinadores, ao mesmo aluno, são homogêneas porquanto suas amplitudes foram pequenas ou nulas.

Algumas afirmativas podem ser extraídas desses resultados:

1. a atribuição de pesos negativos diferentes aos escores e a algumas considerações permitiu conferir notas, de zero a dez, aos alunos, conforme o sistema vigente no ensino brasileiro.

2. essa atribuição de pesos proporcionou alta calibração, inter-examinadores, na concessão das notas finais aos alunos avaliados; ressalte-se que somente essa uniformidade permite conferir, com equidade, notas aos alunos.

3. as diferenças de valor entre as notas finais atribuídas aos vários alunos avaliados mostram as variações de desempenho clínico entre os mesmos.

GAINES e colab.¹⁴ atribuíram pesos diferentes aos critérios, ao avaliarem o desempenho de estudantes durante a raspagem dental em manequim. Entretanto esses autores não puderam determinar se o aumento dos coeficientes de confiança inter-avaliadores obtidos deveu-se ao uso de critérios de desempenho definidos, ao treinamento, ao uso da escala de três pontos, a atribuição de pesos aos escores ou a uma interação dessas variáveis.

Já BURRELL & RASMUSSEN⁵ utilizaram um sistema de atribuição de pesos aos escores, na avaliação de procedimentos endodônticos; concluíram que a atribuição de pesos diferentes a certos escores foi necessária, já que alguns passos do procedimento são mais importantes que outros.

Da mesma forma, no presente estudo, atribuíram-se pesos negativos diferentes, cada vez maiores, aos escores e a algumas considerações, de acordo com o grau de severidade das falhas cometidas pelos alunos.

HOUPT & KRESS¹⁶ atribuíram pesos, igualmente, a todas as dimensões, no cálculo do escore global; já RYGE & SNYDER²⁶ utilizavam o escore mais baixo, dos atribuídos a cada um dos critérios individuais, para determinar o escore global.

O tempo utilizado, pelos examinadores, na realização de cada avaliação, obtido pela análise dos dados extraídos de todos os Formulários "B" usados nas 30 avaliações efetuadas, esteve em torno de 20 minutos.

A discussão de todos os resultados obtidos na aplicação desse Instrumento clínico de avaliação, como amostra de trabalho, para aferir o desempenho final de alunos do 7º semestre de graduação na execução de cavidades Classe I para amálgama, permite deduzir que o mesmo mostrou-se válido e consistente.

A sua validade foi comprovada porquanto ele avaliou, com eficiência, aquilo para o qual foi planejado e construído, isto é, execução clínica de cavidades Classe I para amálgama.

A sua consistência foi revelada tanto pela obtenção de altos coeficientes de confiança inter-examinadores, por aluno avaliado, quanto pelo elevado grau de calibração, no registro dos escores, entre esses mesmos avaliadores.

Também parece-nos lícito inferir que com as devidas adaptações, esse Instrumento de avaliação poderá aferir o desempenho final das demais funções clínicas que compõem o elenco da Clínica Odontológica da FOP-UNICAMP.

Finalmente, podemos deduzir que o sistema empregado atendeu a maioria dos fatores críticos definidos por ABOU-RASS ³ para o desenvolvimento de um sistema de avaliação clínica; do mesmo modo, sua metodologia de documentação, esteve em consonância com a maior parte das recomendações oferecidas por PATRIDGE & MAST ²⁴.

6. CONCLUSÕES

Fundamentando-se na análise dos resultados obtidos da aplicação de um Teste de desempenho clínico, baseado no PEG, planejado e elaborado para a avaliação qualitativa de alunos de graduação da FOP-UNICAMP, e na sua discussão, pode-se concluir:

1. O Teste de desempenho clínico aplicado mostrou-se válido porquanto avaliou, com eficiência, o desempenho final de alunos do 7º semestre de graduação na execução, como amostra de trabalho, de cavidades Classe I para amálgama.

2. Da aplicação desse Instrumento clínico de avaliação, obtiveram-se altos coeficientes de confiança inter-examinadores; a variação desses coeficientes esteve entre o valor mínimo de 76,19% e o valor máximo de 100%; entretanto, o valor mais frequente, dentre todos os obtidos, foi 100%.

3. A obtenção desses coeficientes elevados confirmam que houve alto grau de consistência, entre os pares de examinadores, tanto no julgamento dos mesmos critérios de desempenho de cada aluno avaliado, quanto na avaliação de todos os erros críticos cometidos.

4. Esse alto grau de consistência também foi revelado tanto pelo elevado grau de concordância ou calibração quanto pelo baixo grau de discordância, inter-examinadores, no registro dos escores da maioria dos erros críticos avaliados.

5. A maioria dos critérios de desempenho ou erros críticos avaliados foram identificados e selecionados com propriedade; foram, ainda, bem definidos e objetivos, permitindo ótimo consenso avaliativo entre os examinadores.

6. Certos erros críticos avaliados provocam maiores discordâncias, entre pares de examinadores, no registro dos respectivos escores que os demais.

7. A grande maioria dos critérios avaliados tiveram indicadores de desempenho adequadamente definidos e mensuráveis.

8. Foi possível detectar os erros críticos mais frequen-

temente cometidos, pelos estudantes, durante as avaliações. Esses erros devem merecer, por parte de examinadores e alunos, atenção especial ao longo do processo avaliativo de cavidades Classe I para amálgama.

9. A atribuição de pesos negativos diferentes aos escores e a algumas considerações permitiu conferir notas aos alunos, de zero a dez, conforme o sistema vigente no ensino brasileiro.

10. Essa atribuição de pesos proporcionou alta calibração, inter-examinadores, também na concessão das notas finais aos alunos avaliados; ressalte-se que somente essa uniformidade permite conferir, com equidade, notas aos alunos.

11. As diferenças de valor entre as notas finais atribuídas aos vários alunos examinados mostram as variações de desempenho clínico entre os mesmos.

12. O tempo utilizado, pelos examinadores, na realização de cada avaliação esteve em torno de 20 minutos.

13. O treinamento dos examinadores, ao longo das avaliações, não aumentou os coeficientes de confiança entre os mesmos; esse fato comprova que esse treinamento não é essencial para a aplicação do Instrumento de avaliação utilizado.

14. A aplicação desse Instrumento de avaliação teve grande utilidade instrucional; com efeito, esse sistema permitiu diagnosticar a retro-informação oferecida, pelo mesmo, aos estudantes; do mesmo modo, em relação aos examinadores, essa aplicação proporcionou subsídio valioso a melhoria na qualidade do ensino-aprendizagem.

15. Com as devidas adaptações, esse Instrumento de avaliação poderá aferir o desempenho final das demais funções clínicas que compõem o elenco da Clínica Odontológica da FOP-UNICAMP.

7. REFERÊNCIAS BIBLIOGRÁFICAS

1. ABELES, H.F. A facet-factorial approach to the construction of rating scales to measure complex behaviors. J.Educ. Measurement, 10(2), 1973. Apud PATRIDGE, M.I. & MAST, T. A., op. cit. ref. 24.
2. ABOU-RASS, M. A clinical evaluation instrument in endodontics. J.dent.Educ., 37: 22-36, 1973.
3. _____. Clinical evaluation systems: principles, developments, validations, and administrations. In: MACKENZIE, R.S. & HARROP, T.J. An instructional information exchange for dentistry in the United States. Washington, D.C., U. S. Dept. of Health, Education and Welfare, 1975. v. 6, p. 115-49.
4. AUSUBEL, D.P. Educational psychology: a cognitive view. New York, Holt, Rinehart and Winston, 1968.
5. BURRELL, W.E. & RASMUSSEN, R.H. An evaluation system for preclinical and clinical endodontics. J.dent.Educ., 41: 256-61, 1977.
6. CHARBENEAU, G.T. Rating scales for the clinical evaluation of quality of performance in restorative dentistry. In: MACKENZIE, R.S. & HARROP, T.J. An instructional information exchange for dentistry in the United States. Washington, D.C., U.S. Dept. of Health, Education and Welfare, 1975. v. 6, p. 171-4.
7. DELLA SERRA, O. & FERREIRA, F.V. Anatomia dental. São Paulo, Artes Médicas, 1970. cap. 6 e 7, p. 101-65.
8. EBEL, R.L. Measuring educational achievement. Englewood Cliffs, New Jersey, Prentice-Hall, 1965.
9. FELKNER, L.L. Development and evaluation of standardized measure for assessing diagnostic skills. J.dent.Educ., 37: 26-33, 1973.
10. FERNANDEZ, J.J. Evaluation of student performance in dental schools. Construction and validation of a scale for the evaluation of cavity preparation and silver amalgam res-

- toration in the primary dentition. Ph.D. dissertation. Chapel Hill: University of North Carolina at Chapel Hill, 1967. Apud PATRIDGE, M.I. & MAST, T.A., op.cit. ref. 24.
11. FITZPATRICK, R. & MORRISON, E.J. Performance and product evaluation. In: THORNDIKE, R.L., ed. Educational measurement. Washington, D.C., American Council on Education, 1971.
 12. FULLER, J.L. The effects of training and criterion models on interjudge reliability. J.dent.Educ., 36:19-22, 1972.
 13. GAINES, W.G.; BRUGGERS, H.; RASMUSSEN, R.H. Reability of ratings in preclinical fixed prothodontics: effect of objective scaling. J.dent.Educ., 38: 672-5, 1974.
 14. GAINES, W.G.; RASMUSSEN, R.H.; UCHELLO, E. Increasing the objectivity of clinical grading. Dent.Hyg., 49: 227-80, 1975.
 15. HINKELMAN, K.W. & LONG, N.K. Method for decreasing subjective evaluation in preclinical restorative dentistry. J.dent.Educ., 37:13-8, 1973.
 16. HOUP, M.I. & KRESS, G. Accuracy of measurement of clinical performance in dentistry. J.dent.Educ., 37:34-36, 1973.
 17. HUNTER, H. PEG: performance evaluation guides. In: MACKENZIE, R.S. & HARROP, T.J. An instructional information exchange for dentistry in the United States. Washington, D.C., 1975. v. 6, p. 3-69.
 18. KILLIP, D.E. & LEWIS JR., A. The problem of weighting student scores. J.dent.Educ., 36: 57-8, 1972.
 19. MACHEN, J.B. Guidelines for development of tests of clinical judgment. In: MACKENZIE, R.S. An instructional information exchange for dentistry in the United States. Washington, D.C., 1975. v. 7, p. 40.
 20. MACKENZIE, R.S. Defining clinical competence in terms of quality, quantity, and need for performance criteria. J.dent.Educ., 37:37-44, 1973.

21. MAST, T.A. & SPEDDING, R.H. Task analysis of a pedodontic procedure. J.dent.Educ., 39: 725-7, 1975.
22. MONDELLI, J.; ISHIKIRIAMA, A.; GALAN JR., J.; NAVARRO, M.F.L. Dentística operatoria. 3. ed. São Paulo, Sarvier, 1977. p. 13-23, 109-11, 114.
23. NATKIN, E. & GUILD, R.E. Evaluation of preclinical laboratory performance: a systematic study. J.dent.Educ., 31: 152-61, 1967.
24. PATRIDGE, M.I. & MAST, T.A. Dental clinical evaluation: a review of the research. J.dent.Educ., 42:300-5, 1978.
25. PIEDADE, E.F. Avaliação quantitativa das atividades clínicas dos alunos de Clínica Integrada da Faculdade de Odontologia de Piracicaba da Universidade Estadual de Campinas. Piracicaba, 1977. [Tese (Livre-Docência) - F.O.P. - Unicamp.]
26. RYGE, G. & SNYDER, M. Evaluating the clinical quality of restorations. J.Am.dent.Ass., 87: 369-77, 1973.
27. SALVENDY, G.; JOOST, M.G.; CUNNINGHAM, P.R.; FERGUSON, G.W.; WILKER, R.A.; DEES, R.W. Improving evaluation of amalgam restorations. J.dent.Educ., 40: 368-9, 1976.
28. UNIVERSIDADE ESTADUAL DE CAMPINAS, Faculdade de Odontologia de Piracicaba. Análise de funções e tarefas clínicas. Piracicaba, 1976.
29. VANEK, G. Objective evaluation of dental student technique products. J.dent.Educ., 33: 140-4, 1969.

8. A N E X O S

CRITÉRIOS DE DESEMPENHO (Formulário A₁)

ANEXO 1

FUNÇÃO BÁSICA: 022 - Terapêutica de cárie de sulcos e fissuras (Cav. Classe I) TAREFA: 0072 - Preparo de cavidades Classe I para amálgama

CLÍNICA: DC - 701 OBJETIVO DA AVALIAÇÃO: Diagnosticar os erros dos alunos SISTEMA DE ATRIBUIÇÃO DE ESCORES: C - 1c - If - F

CRITÉRIOS DE DESEMPENHO	ESCORES	DEFINIÇÕES OU INDICADORES	EXCEÇÕES	RAZÕES DOS CRITÉRIOS
DELIMITAÇÃO DE CONTO RNO CAVITÁRIO				
Abertura excessiva	If	Pré-molares: Com a ponta ativa da broca 57 colocada no istmo da cavidade, junto a uma das paredes axiais, resta espaço maior que o diâmetro de uma broca (0,8mm). Molares: Com a ponta ativa da broca 57 colocada nos istmos da cavidade, junto a uma das paredes axiais, resta espaço maior que o diâmetro de 1 1/2 broca (1,2mm).	Remoção de cárie; restauração antiga; fratura do dente.	Preservação do dente
Abertura insuficiente	C	O preparo não engloba todas as cicatrículas, sulcos e fissuras do dente preparado.	Pacientes (idosos, com abrasão oclusal).	Remoção de cáries; Extensão preventiva.
Crista marginal enfraquecida	F	As cristas marginais M e/ou D foram comprometidas pela abertura (resta menos de 1mm de estrutura dental).	Nenhuma	Preservação do dente.
Cúspides enfraquecidas	F	Cúspides solapadas pela broca; há esmalte e/ou dentina sem apoio;	Remoção de cárie nos pontos; restauração antiga; fratura dos dentes.	Preservação do dente.
PREPARO DE FORMAS DE RESISTÊNCIA E RETENÇÃO				
Má orientação das paredes cavitárias	C	Paredes V e/ou L e/ou M e/ou D não paralelas (excessivamente divergentes ou convergentes entre si).	Processo extenso de cárie.	Forma de resistência.
	C	Parede pulpar irregular (não plana).	Cárie profunda.	
Cavidade profunda	1c	Mais que 1/2 da ponta ativa da broca 57 penetra na cavidade, em relação ao bordo cavo-superficial, junto às cristas marginais M e D.	Cárie muito profunda. Extremo desgaste oclusal.	Proteção pulpar.
Cavidade pouco profunda	C	Profundidade menor que a abertura V-L. Menos que 1/2 da ponta ativa da broca 57 penetra na cavidade, em relação ao bordo cavo-superficial, junto às cristas marginais M e D.	Cárie muito extensa no sentido V-L; restauração antiga.	Forma de retenção.
REMOÇÃO DE TECIDO CARIADO*				
Cárie remanescente	C	Coloração típica; textura amolecida, detectável com a ponta do explorador.	Filosofia de remoção incompleta de cárie p/cvitar exposição pulpar.	Remoção de cárie.
Exposição mecânica	F	Hemorragia; polpa visível	Nenhuma	Proteção pulpar.

* Em cáries extensas e/ou profundas, a remoção do tecido cariado precede as formas de contorno, resistência e retenção.

FUNÇÃO BÁSICA : 022 - Terapêutica de cárie de sulcos e fissuras (Cav. Classe I) TAREFA : 0074 - Proteção Pulpar (Forramento)

CLÍNICA : DC - 701 OBJETIVO DA AVALIAÇÃO : Diagnosticar erros dos alunos SISTEMA DE ATRIBUIÇÃO DE ESCORES : C - Ic - If - F

CRITÉRIOS DE DESEMPENHO	ESCORES	DEFINIÇÕES OU INDICADORES	EXCEÇÕES	RAZÕES DOS CRITÉRIOS
ESCOLHA DO MATERIAL FORRADOR				
Eleição inadequada	Ic	Uso de materiais forradores <u>não adequados</u> para o caso clínico.	Nenhuma	Proteção Pulpar.
FORRAMENTO				
Cobertura insuficiente da parede pulpar	C	Existem pontos da parede pulpar não cobertos pelo material forrador.	Princípio de Gabel, de 2 ou 3 pontos do amálgama apoiados em dentina, em cavidades extensas.	Proteção Pulpar.
Excesso de material forrador	C	A quantidade excessiva de material forrador faz com que a cavidade fique pouco profunda e/ou há material forrador nas paredes cavitárias, acima do nível da parede pulpar ou no bordo cavo-superficial.	Nenhuma	Forma de Retenção. Infiltração Marginal.

CRITÉRIOS DE DESEMPENHO (Formulário A₃)

ANEXO 3

FUNÇÃO BÁSICA : 022 - Terapêutica de cárie de sulcos e fissuras (Cavidade Classe I) TAREFA: 0075 - Restauração do dente Classe I c/amálgama

CLÍNICA : DC - 701 OBJETIVO DA AVALIAÇÃO: Diagnosticar erros dos alunos SISTEMA DE ATRIBUIÇÃO DE ESCORES :

C - Ic - If - F

CRITÉRIOS DE DESEMPENHO	ESCORES	DEFINIÇÕES OU INDICADORES	EXCEÇÕES	RAZÕES DOS CRITÉRIOS
CONDENSAÇÃO DO AMÁLGAMA				
Excesso de mercúrio	C	O amálgama condensado cede à pressão moderada do calcador, marcando a superfície da restauração; superfície excessivamente brilhante.	Nenhuma	Propriedades físicas do amálgama.
Falta de material	C	Evidência de degrau no bordo cavo superficial, no qual o explorador pode se prender.	Nenhuma	Infiltração marginal; recidiva de cárie.
ESCULTURA DO AMÁLGAMA				
Superfície oclusal abaixada	Ic	Paredes com esmalte e/ou dentina expostos, detectáveis com a ponta do explorador.	Nenhuma	Fratura do dente; recidiva de cárie.
	Ic	Excessiva profundidade da anatomia oclusal; dentina ou forramento expostos.	Nenhuma	Fratura do dente; recidiva de cárie.
Superfície oclusal alta	C	Sulcos e fissuras muito rasos; amálgama saliente, além do bordo cavo-superficial, visível e/ou detectável com o explorador.	Nenhuma	Preservação da anatomia oclusal; trauma oclusal.
Anatomia oclusal incorreta	Ic	Amálgama com escultura muito deficiente ou sem escultura.	Nenhuma	Função mastigatória; trauma oclusal.
TESTE DE OCLUSÃO				
Contacto oclusal alto	C	Contactos prematuros, em oclusão cêntrica ou em balanceio, detectáveis pelo paciente ou pelo teste de carbono para articulação.	Nenhuma	Função mastigatória; Trauma oclusal.
Restauração fraturada, perda	Ic	Restauração fraturada ao teste de oclusão, necessitando ser removida e refeita.	Nenhuma	Infiltração marginal; recidiva de cárie; preservação do dente.

FORMULÁRIO CLÍNICO (FORMULÁRIO B)

ANEXO 4

ALUNO _____ Nº _____ CLÍNICA _____ DENTE _____
 EXAMINADOR _____ PACIENTE _____ PG _____

REGISTRO DOS ESCORES

Assinalar os escores correspondentes ao desempenho clínico do aluno

Tarefa: 0072-Prep.cav.Classe I p/amálgama

Abertura Excessiva	If	-3
Abertura Insuficiente	C	-1
Crista Marginal Enfraquecida	F	-5
Cúspides Enfraquecidas	F	-5
Má orientação das paredes cavitárias	C	-1
	C	-1
Cavidade Profunda	Ic	-2
Cavidade pouco profunda	C	-1
Cárie Remanescente	C	-1
Exposição Mecânica	F	-5
SUB TOTAL 1		-

Tarefa: 0074 - Prot. pulpar (Forramento)

Eleição inadequada	Ic	-2
Cobertura insuf. parede pulpar	C	-1
Excesso mat. forrador	C	-1
SUB TOTAL 2		-

Tarefa: 0075-Rest.do dente-Classe I p/amalg.

Excesso de mercúrio	C	-1
Falta de material	C	-1
Superfície oclusal abaixada	Ic	-2
	Ic	-2
Superfície oclusal alta	C	-1
Anatomia oclusal incorreta	Ic	-2
Contacto oclusal alto	C	-1
Restauração fraturada, perdida	Ic	-2
SUB TOTAL 3		-

TOTAL DOS ESCORES

SUB-TOTAIS 1 + 2 + 3 (A)

REGISTRO DAS CONSIDERAÇÕES

TEMPO P/ EXECUÇÃO DA FUNÇÃO CLÍNICA (4)

Até 30 minutos	0	Horário inicial: _____ Horário final: _____ Tempo real utilizado: _____ min Tempo utilizado p/ avaliação: _____ min
31 min - 35 min	- 2	
36 min - 40 min	- 4	
41 min - 45 min	- 6	
46 min - 50 min	- 8	
51 min - 55 min	-10	
56 min - 60 min	-12	
61 min ou mais	-14	

DIFICULDADES P/ EXECUÇÃO DA FUNÇÃO CLÍNICA

Assinalar o quadrado correspondente à dificuldade encontrada. A cada dificuldade assinalada será dado, ao aluno, mais 5' para execução da função clínica.

Dentes com desvios de seu longo eixo
Dificuldade de visão do dente preparado
Paciente não colabora
Língua excessivamente grande (infer.)
Excesso de salivagem
Deficiência material ou instrumental
Outras dificuldades

ACRÉSCIMO TEMPO P/EXEC. FUN.CLIN. ____ min

AJUDA DO EXAMINADOR (5)

Assinalar o escore correspondente à ajuda do examinador.

1 ajuda	- 5
2 ajudas	-10
3 ajudas	-15

ORIENTAÇÃO DO EXAMINADOR (6)

Assinalar o escore correspondente à orientação docente.

1 orientação	-2
2 orientações	-4
3 orientações	-6

TOTAL DAS CONSIDERAÇÕES

Esta coluna não deve ser preenchida pelo examinador.

SOMA DOS PESOS NEGATIVOS ASSINALADOS:

(4) - ____ + (5) - ____ + (6) - ____ = ____ (B)

GRAU: (A) - ____ + (B) - ____ = ____

*NOTA FINAL: _____

*Consultar o Formulário C

Assinatura do Examinador

Assinatura do Aluno

ANEXO 5NOTAS FINAIS (FORMULÁRIO C)

FUNÇÃO BÁSICA: 022 - Terapêutica de cárie de sulcos e fissuras
(Cav. Classe I)

TAREFAS: 0072 - 0074 e 0075

CLÍNICA: DC - 701 SISTEMA DE ATRIBUIÇÃO DE ESCORES
Numeral (0 a 10)

Nota Final	Total dos Escores + Total das Considerações (Graus)
10,0	0
9,5	- 1 e - 2
9,0	- 3 e - 4
8,5	- 5 e - 6
8,0	- 7 e - 8
7,5	- 9 e - 10
7,0	- 11 e - 12
6,5	- 13 e - 14
6,0	- 15 e - 16
5,5	- 17 e - 18
5,0	- 19 e - 20
4,5	- 21 e - 22
4,0	- 23 e - 24
3,5	- 25 e - 26
3,0	- 27 e - 28
2,5	- 29 e - 30
2,0	- 31 e - 32
1,5	- 33 e - 34
1,0	- 35 a - 37
0,5	- 38 a - 40
0,0	- 41 ou mais

ANEXO 6INSTRUÇÕES AOS ALUNOS

Prezado aluno:

1. Muito grato por você ter concordado em se submeter à avaliação de seu desempenho clínico em cavidades Classe I para amálgama, através desse instrumento baseado no PEG.
2. Embora essa avaliação não vá influir na sua aprovação ou não na DC - 701, será efetuada da maneira mais rigorosa possível, pois servirá de importante subsídio para futura avaliação clínica de seus colegas, de séries imediatamente anteriores a sua; por outro lado, esta avaliação será muito útil para que você próprio sinta como está seu desempenho clínico em cavidades Classe I.
3. Leia, cuidadosamente, esse conjunto de folhas de instruções e os Formulários anexos. Procure o Administrador do Teste (Dr. Roberto José Gonçalves) para esclarecer possíveis dúvidas a respeito da aplicação desse instrumento de avaliação.
4. Mantenha sempre consigo, inclusive na ocasião da avaliação, os formulários A₁, A₂ e A₃ (critérios de desempenho - ANEXOS 1, 2 e 3) que lhe darão diretrizes seguras acerca de seus possíveis erros clínicos, durante o preparo, forramento e restauração do dente objeto da avaliação, e o orientarão no sentido de evitá-los.

I) Instruções Pré-Teste

1. Selecione, dentre seus pacientes de Clínica Odontológica, 1 (um) adulto, que possua cárie de sulcos e fissuras (cavidade Classe I) na superfície oclusal de molares e/ou pré-molares. Certifique-se, inclusive com radiografias, se a cárie localiza-se, apenas, na superfície oclusal dos dentes mencionados.
2. Escolha, dentre os dentes examinados, um deles para o preparo de cavidade Classe I para amálgama a ser avaliada.
3. Você será avaliado, por 3 (três) ou 4 (quatro) examinadores,

no Laboratório de Orientação Profissional (Unidade Clínica Piloto - Sala 3-47) ou no seu próprio box da Clínica Odontológica. Por isso, marque a consulta dos pacientes selecionados, para o local eleito, em dia e horário em que os examinadores possam estar presentes (o Administrador do Teste o orientará nesse sentido).

4. No dia marcado para a avaliação, procure chegar, com seu paciente, 15 minutos antes ao local do Teste.
5. Antes de começar o trabalho, verifique e organize o instrumental e material a serem utilizados, bem como a dificuldade do procedimento e o tempo previsível que irá gastar. Deixe ao alcance os formulários A_1 , A_2 e A_3 (critérios de desempenho - ANEXOS 1, 2 e 3) para eventual consulta.
6. Posicione, corretamente, o paciente segundo o dente a ser preparado; posicione-se, também, adequadamente em seu mocho.
7. Examine o dente a ser preparado e execute a anestesia indicada; providencie e instale a broca de alta rotação que irá utilizar; teste o efeito da anestesia aplicada e chame os examinadores, informando-os que irá iniciar o preparo da cavidade.

II) Instruções do Teste

1. Prepare a cavidade Classe I, segundo a técnica; ao concluir essa tarefa, mantenha no alta rotação uma broca 57, com o auxílio da qual os examinadores procederão à avaliação objetiva do preparo cavitário. Chame os examinadores para que procedam à avaliação do preparo concluído.
2. Execute as correções que forem indicadas pelos examinadores; chame-os, novamente, quando as correções forem concluídas. Não havendo mais correções a fazer, passe para a tarefa seguinte. Proceda da mesma forma se não houver correções a fazer após a conclusão do preparo, na primeira tentativa.
3. Proceda às tarefas subseqüentes, forramento e restauração do dente, chamando os examinadores, para a respectiva avaliação, após a conclusão de cada uma; as mesmas instruções do item 2 são válidas para cada uma dessas tarefas. A avaliação da ta-

refa restauração do dente será feita em duas etapas: uma após a condensação do amálgama e outra após o término da aludida tarefa.

4. Um dos Professores Examinadores pode realizar todo ou parte do procedimento (ajuda do examinador), mas somente se você pe dir ou para proteção da integridade dental do paciente; da mesma forma, e pelos mesmos motivos, poderá oferecer orientações (orientação do examinador). Saiba, entretanto, que qual quer ajuda ou orientação recebida vai influir, negativamente, no seu grau e, conseqüentemente, na sua nota final.
5. Concluída a função clínica, dê, ao paciente, as instruções necessárias quanto à preservação da restauração e dispense-o marcando ou não nova consulta.

III) Instruções Pós-Teste

1. Os examinadores poderão discutir com você, as falhas clínicas cometidas durante a avaliação, imediatamente após o Teste e solicitar-lhe que assine o Formulário Clínico (Formulário B), se você desejar, dispense, primeiramente, o paciente e logo após poderá discutir, com os examinadores, os erros cometidos.
2. A nota final da avaliação ser-lhe-á fornecida após a tabulação dos dados do Teste, tarefa que será realizada pelo Administrador do mesmo.

Piracicaba, de de 1980.

Prof. Ass. Dr. Roberto José Gonçalves
Administrador do Teste

ANEXO 7INSTRUÇÕES AOS EXAMINADORES

Prezado colega:

1. Muito grato por você ter concordado em examinar o desempenho clínico dos alunos da DC - 701 em cavidade Classe I para amálgama, através desse instrumento baseado no PEG.
2. Saiba que essa avaliação, embora não vá influir na aprovação ou não desses alunos, será realizada da maneira mais rigorosa possível, pois servirá de importante subsídio para futura avaliação clínica dos demais alunos da FOP-UNICAMP.
3. Leia, cuidadosamente, esse conjunto de folhas de instruções e os Formulários anexos. Procure o Administrador do Teste (Roberto José Gonçalves) para esclarecer possíveis dúvidas a respeito da aplicação desse instrumento de avaliação.
4. Mantenha sempre consigo, mormente na ocasião da avaliação, os Formulários A_1 , A_2 e A_3 (Critérios de desempenho - ANEXOS 1, 2 e 3), que lhe darão diretrizes seguras acerca dos possíveis erros clínicos dos alunos avaliados, durante o preparo, forramento e restauração do dente objeto da avaliação, e o orientação no sentido da anotação dos mesmos no Formulário B (Anexo 4).
5. A intenção do presente trabalho é avaliar o maior número possível de alunos da DC - 701, durante o 1º semestre de 1980; 3 (três) ou 4 (quatro) examinadores avaliarão, concomitantemente, o desempenho clínico, em cavidade Classe I para amálgama, do mesmo aluno. Cada aluno a ser examinado, executará a função clínica proposta em 1 (um) de seus pacientes de Clínica Odontológica.

I) Instruções Pré-Teste

1. As sessões de avaliação serão realizadas, a partir de 14 de março de 1980, no Laboratório de Orientação Profissional (Sala 3-47) ou nos próprios boxes dos alunos da DC - 701.

2. Serão realizadas, por sessão, tantas avaliações quantas forem possíveis.
3. No dia marcado para a avaliação, procure chegar 15 minutos antes ao local do Teste.
4. Aprove ou não os pacientes selecionados, pelo aluno, para a avaliação. Lembre-se que somente serão avaliados cavidades Classe I, localizadas nas superfícies oclusais de molares e/ou pré-molares.
5. Essa aprovação deverá ser feita através do exame clínico minucioso do dente a ser preparado; examine, inclusive, as radiografias tomadas do mesmo, verificando se não existem cáries proximais, vestibulares ou linguais. Esse exame deverá detectar todas as características iniciais do dente a ser preparado; essas características nortearão a configuração de possíveis exceções que possam se aplicar ao desempenho do caso clínico objeto da avaliação. O exame clínico deverá, ainda, diagnosticar as possíveis dificuldades para execução da função clínica a ser avaliada.
6. Providencie, com o Administrador do Teste, uma cópia do Formulário "B" (Formulário Clínico - ANEXO 4) e mantenha-o sempre consigo, preso a uma prancheta de madeira.
7. Preencha os dados de identificação do Formulário "B", bem como assinale, nos quadrados correspondentes, as dificuldades encontradas para a execução da função clínica a ser avaliada.
8. Deixe ao alcance os Formulários A₁, A₂ e A₃ (critérios de desempenho - ANEXOS 1, 2 e 3) para eventual consulta.

II) Instruções do Teste

1. Ao ser chamado pelo aluno, que o informará que irá iniciar o preparo da cavidade, anote, no espaço indicado, o horário inicial da execução da função clínica.
2. Acompanhe, dentro do possível, a execução do preparo da cavi-

dade e, quando concluída, avalie a cavidade preparada comparando-a com os parâmetros descritos nas definições ou indicadores de desempenho do respectivo Formulário (A₁ - ANEXO 1); assinale, nas caselas de pontos negativos, os escores correspondentes aos erros clínicos praticados pelo aluno; quando os escores anotados forem C e/ou Ic, informe, ao aluno, quais os erros cometidos e dê instruções para que execute as devidas correções. Anote, se houver, os escores correspondentes a novos erros ou a repetições dos mesmos erros. Não havendo correções a serem feitas, informe ao aluno para passar à tarefa seguinte.

3. Proceda da mesma forma ao avaliar as tarefas subseqüentes, isto é, forramento e restauração do dente. Lembre-se que a avaliação da tarefa restauração do dente será realizada em duas etapas: uma após a condensação do amálgama e outra após o término da aludida tarefa.
4. Não se preocupe em preencher os sub-totais 1, 2 e 3, bem como o total dos escores; isto será feito pelo Administrador do Teste.
5. Concluída a função clínica, anote, no espaço indicado, o horário final; não se preocupe em preencher o tempo real utilizado pelo aluno, bem como em assinalar os pesos negativos correspondentes ao tempo empregado pelo mesmo; isto será feito pelo Administrador do Teste, que registrará, ainda, o tempo utilizado, pelos examinadores, para a execução de cada avaliação.
6. O acréscimo de tempo para a execução da função clínica será anotado, posteriormente, pelo Administrador do Teste.
7. Assinale os pesos negativos correspondentes à ajuda e/ou orientação dos examinadores oferecidas durante a execução da função clínica avaliada. Lembre-se que, tanto a ajuda como a orientação dos examinadores serão dadas quando solicitadas pelo aluno ou para preservar a estrutura dental do paciente.
8. Não deixe que os outros examinadores vejam os escores que você assinalou e nem confronte, com os mesmos, os resultados ob-

tidos; isto é muito importante, pois poderá influenciar no resultado da avaliação.

9. O total das considerações, o grau e a nota final serão obtidos pelo Administrador do Teste; se você desejar, poderá ter conhecimento dos mesmos, posteriormente.

III) Instruções Pós-Teste

1. Solicite, ao aluno, que assine o Formulário B; faça você o mesmo.
2. Discuta, quando solicitado pelo aluno, as possíveis falhas clínicas, por ele comitidas, durante a avaliação.
3. Entregue, ao Administrador do Teste, o Formulário B devidamente preenchido e prepare-se para nova avaliação.

Piracicaba, de de 1980.

Roberto José Gonçalves
Administrador do Teste