

TOMAZ TADEU DA SILVA

MEDIÇÃO COM REFERÊNCIA A CRITÉRIO E  
MEDIÇÃO COM REFERÊNCIA A NORMA:  
UMA COMPARAÇÃO

Dissertação apresentada aos Cursos de Pós-Graduação  
em Educação da Universidade Federal do Rio  
Grande do Sul, para a obtenção do  
título de Mestre em Educação

PORTO ALEGRE

1977

MEDIÇÃO COM REFERÊNCIA A CRITÉRIO E  
MEDIÇÃO COM REFERÊNCIA A NORMA :  
UMA COMPARAÇÃO

160584

À memória de meu pai

À minha mãe

ORIENTADOR DA DISSERTAÇÃO :

Dr. Ray Arthur Chesterfield

Doctor of Philosophy in Education,

University of California,

Los Angeles, USA

## SUMÁRIO

ÍNDICE DE TABELAS E ANEXOS.....	xi
APRESENTAÇÃO.....	xiii
RESUMO.....	xv
ABSTRACT.....	xv
1. INTRODUÇÃO.....	1
1.1. O contexto do problema.....	3
1.2. A formulação do problema.....	4
2. REVISÃO DA LITERATURA.....	5
2.1. Medição com referência a critério e Medição com referência a norma: Características distintivas.....	7
2.2. Efeitos de procedimentos de avaliação sobre o comportamento dos estudantes.....	12
3. HIPÓTESES.....	17
4. METODOLOGIA.....	21
4.1. Definição das variáveis.....	23
4.2. Sujeitos.....	23
4.3. Instrumentos.....	24
4.4. Procedimentos.....	25
4.5. Tratamento estatístico.....	26
5. RESULTADOS E DISCUSSÃO.....	28
5.1. Hipótese 1.....	29
5.2. Hipótese 2.....	32
6. CONCLUSÕES E IMPLICAÇÕES.....	35
7. SUGESTÕES PARA OUTROS ESTUDOS.....	39
8. REFERÊNCIAS BIBLIOGRÁFICAS.....	43
9. ANEXOS.....	49

# ÍNDICE DE TABELAS E ANEXOS

## TABELAS

TABELA 1 - Distribuição de frequência dos pontos obtidos no pré-teste de rendimento pelos grupos MRC e MRN.....	29
TABELA 2 - Resultado do teste da hipótese nula sobre rendimento dos grupos MRC e MRN, antes da instrução.....	29
TABELA 3 - Distribuição de frequência dos pontos obtidos no pós-teste de rendimento pelos grupos MRC e MRN.....	30
TABELA 4 - Resultado do teste da hipótese nula 1.....	30
TABELA 5 - Distribuição de frequência dos pontos atribuídos à Escala de Atitudes, pelos grupos MRC e MRN.....	32
TABELA 6 - Resultado do teste da hipótese nula 2.....	33

## ANEXOS

ANEXO 9.1. - Escala de mensuração de atitudes para com a disciplina..	51
ANEXO 9.2. - Teste de Rendimento.....	55
ANEXO 9.3. - Plano de curso da disciplina em que se realizou o experimento.....	65
ANEXO 9.4. - Lista de objetivos de uma das 5 unidades da disciplina em que se realizou o experimento.....	71
ANEXO 9.5. - Teste de unidade de uma das 5 unidades da disciplina em que se realizou o experimento.....	75
ANEXO 9.6. - Ficha de Informação dos Resultados do grupo Medição com Referência a Critério.....	79
ANEXO 9.7. - Ficha de Informação dos Resultados do grupo Medição com Referência a Norma.....	83

## APRESENTAÇÃO

A idéia de que o ensino deve ser planejado e desenvolvido de maneira sistemática vem ganhando adeptos até mesmo em setores os mais inespérados. Cada vez mais um maior número de pessoas vai se convencendo de que a instrução pode ser atacada cientificamente. Por outro lado, o que se convencionou chamar de "tecnologia instrucional" recebe crescentes contribuições derivadas da pesquisa educacional. Tornar operacionais essas contribuições e fazê-las chegar até aqueles que, nas salas-de-aulas, solicitam-nas com interesse crescente deve ser uma das tarefas principais dos que se dedicam à investigação educacional e sua difusão.

A avaliação é um dos componentes de um sistema instrucional que se insere perfeitamente nas tendências acima. A teoria e a pesquisa educacionais vem elaborando novos enfoques e propondo procedimentos de avaliação mais adequados a concepções mais modernas de educação, enquanto a demanda por parte dos usuários dos sistemas de instrução de novos conhecimentos sobre este tema aumenta de maneira constante, como resultado dos inúmeros problemas concretos encontrados na prática diária.

Foi a partir da verificação dessa demanda que, como professor, passei a me interessar pelo tema "avaliação". Tendo a oportunidade, como aluno de um curso de Pós-Graduação em Educação, de dar uma parcela de contribuição para resolver algum aspecto do problema, era natural que escolhesse o tema para desenvolver a dissertação.

Dessa forma, a preocupação com o relacionamento entre objetivos e interpretação de resultados de avaliação, juntamente com a descoberta de que educadores americanos desenvolviam um enfoque que procurava solucionar esse problema, levaram à decisão de selecionar o tema específico "Medição com Referência a Critério". Espero que, tanto os resultados que respondem diretamente ao problema formulado, quanto os procedimentos de avaliação desenvolvidos especialmente com a finalidade de servir de suporte à sua solução, sejam de alguma utilidade aos que labutam na tarefa diária das salas-de-aulas e aos que se dedicam a pesquisas nessa área.

Difícil relacionar todos os que contribuíram, de alguma forma, para a realização deste trabalho. Algumas pessoas, entretanto, devem ser mencionadas:

PROF. DR. RAY A. CHESTERFIELD, cuja objetividade, compreensão e inteligência forneceram a dose exata de orientação: nem o "deixar-fazer", nem o autoritarismo inibidor, mas o meio-termo produtivo.

PROF. ALFREDO JOSÉ DA VEIGA NETO, Diretor da Faculdade de Ciências e Educação de Criciúma, pela compreensão, pelo apoio e sobretudo, pela amizade, proporcionadora de estimulante convívio intelectual.

PROF. RODEVAL JOSÉ ALVES, Diretor-Executivo da Fundação Educacional de Criciúma, pelo constante estímulo e boa vontade em procurar uma difícil conciliação entre serviço e tarefas de pesquisa que não prejudicasse as últimas.

PROFESSORA DALVA DE LUCCA DIAS, Chefe do Departamento de Ciências Pedagógicas da Faculdade de Ciências e Educação de Criciúma, pelo exemplo de interesse caloroso pelos problemas educacionais e pelas facilidades administrativas concedidas.

MARIA INÊS ROCHA, pelo fornecimento do componente afetivo indispensável e pelo apoio seguro nas horas mais críticas do trabalho.

Agradeço ainda : à Coordenação e Professores dos Cursos de Pós-Graduação em Educação da Universidade Federal do Rio Grande do Sul; à CAPES; à Fundação Educacional de Criciúma e à Faculdade de Ciências e Educação de Criciúma.

Tomaz Tadeu da Silva

## RESUMO

O presente estudo objetivou comparar um sistema de avaliação baseado na teoria conhecida como MEDIÇÃO COM REFERÊNCIA A CRITÉRIO (MRC) com um sistema baseado na teoria tradicional de avaliação, a MEDIÇÃO COM REFERÊNCIA A NORMA (MRN), em relação a seus efeitos sobre o comportamento dos estudantes. Mais especificamente, procurou-se comparar os dois sistemas em relação a seus efeitos sobre o rendimento e as atitudes dos estudantes para com a disciplina de estudos.

Foram formuladas as hipóteses de que o rendimento de estudantes submetidos ao sistema MRC seria maior que o rendimento de estudantes submetidos ao sistema MRN e a atitude daqueles mais favoráveis que a atitude desses. Para testá-las usou-se um experimento, em que os sujeitos, estudantes de Pedagogia, em número de 33, foram divididos aleatoriamente em dois grupos. Ambos os grupos participaram de um mesmo programa instrucional, em que diferia apenas a forma de interpretação dos resultados dos testes das unidades em que foi dividido o programa. Um dos grupos recebeu a comunicação desses resultados em termos de MRC, o outro em termos de MRN.

Não se verificou diferença significativa de rendimento entre o grupo submetido ao sistema de avaliação baseado na MRC e o grupo submetido ao sistema de avaliação baseado na MRN. Da mesma forma não foi verificada diferença significativa de atitudes para com a disciplina entre os dois grupos.

## ABSTRACT

The present study was aimed at comparing an evaluation system based on the theory known as Criterion Referenced Measurement (CRM) with a system based on the traditional evaluation theory, the Norm Referenced Measurement (NRM), concerning their effects over the students behavior. In short, it was sought to compare the two systems in relation to their effects over the achievement and students attitude toward the subject.

It was hypothesized that the achievement of students submitted to the CRM system should be greater than the achievement of students submitted to the NRM system and the attitude of those should be more favorable than the attitude of the latter. To test them it was used an experiment in which the subjects, 33 students of Pedagogy, were divided, at random, into two groups. Both took part in the same instructional program, which differed only in the interpretation of the test results of the units. One of the groups received the communication of these results in terms of CRM, the other, in terms of NRM.

With regard to the students achievement between the group submitted to the evaluation system based on CRM and the group submitted to the system based on NRM, no significant difference was revealed. It also was not indicated significant difference of attitude toward the subject, between the two groups.

# 1. INTRODUÇÃO

## 1.1. O CONTEXTO DO PROBLEMA

A formulação explícita de objetivos de ensino e o relacionamento desses com o processo de avaliação têm-se constituído em temas dominantes da literatura recente sobre modelos de instrução (Block, 1971; Mager, 1972; Popham & Baker, 1972; Nagel & Richman, 1974; Gronlund, 1975). As palavras de Gronlund (1976, p.7) constituem um exemplo da ênfase que se tem dado a esses temas: "O primeiro passo, tanto no ensino quanto na avaliação, é o de determinar os produtos de aprendizagem a serem esperados da instrução.... Que tipo de produto de aprendizagem está sendo buscado?..... Somente através da identificação de objetivos instrucionais e de sua formulação clara em termos de comportamentos específicos podemos fornecer uma direção ao processo de ensino e preparar o palco para a pronta avaliação dos produtos de aprendizagem."

Simultaneamente tem-se procurado, com frequência, relacionar o processo de avaliação em educação ao de tomada de decisão (Stufflebean, 1971; Mehrens & Lehmann, 1973), destacando-se o primeiro como encarregado de fornecer a informação necessária ao último: "Enquanto a tomada de decisões visa interferir no curso de uma ação, a avaliação é o mecanismo de apoio a essa decisão, e, enquanto tal, livre de conteúdos, objetivos e alternativas: a avaliação não é julgamento, mas é processo que permite o julgamento, a decisão. O que se oferece numa avaliação são informações úteis para a tomada de decisão." (Araújo e Oliveira, 1974, p.318).

Como decorrência dessas duas concepções têm surgido proposições de procedimentos de avaliação mais ajustados aos princípios delas derivados. Essas proposições estão apoiadas no pressuposto de que os procedimentos tradicionais têm falhado em adequar-se às necessidades colocadas pelas duas tendências antes descritas, na medida em que não fornecem o tipo de informação exigido por um modelo de instrução que baseie suas decisões no conhecimento do grau de consecução, por parte dos estudantes, de objetivos bem determinados (Glaser, 1971; Popham & Husek, 1971; Bastos & Swyter, 1974).

Alguns desses procedimentos estão reunidos no enfoque conhecido como MEDIÇÃO COM REFERÊNCIA A CRITÉRIO (MRC), em oposição ao tradicional, conhecido como MEDIÇÃO COM REFERÊNCIA A NORMA (MRN). A MRC propõe o uso de medidas de aproveitamento escolar que tomem como referência um determinado padrão de comportamentos e habilidades especificados antes da instrução, ao invés do desempenho de um grupo-norma, como é feito no tratamento tradicional. Para usar as palavras de um dos teóricos desse enfoque, Popham (1971, p.20), "a MRC é aquela que é usada para verificar o status de um indivíduo com respeito a algum critério, isto é, um padrão de comportamento. Esta medição é descrita como "com referência a critério" porque o estudante é comparado com algum critério estabelecido, antes que com outros estudantes. O significado de um escore não depende da comparação com outros. Quer-se saber o que o indivíduo pôde fazer, não como ele se situa em comparação com outros."

Tem-se feito especulações em torno das vantagens e conveniências (e por outro lado, das limitações e inconveniências) da adoção extensiva, na prática educacional, de sistemas de avaliação baseados na MRC (Popham, 1971; Ebel, 1972a), tendo-se feito propostas também sobre formas de operacionalização das teorias de MRC (Gronlund, 1973; Gorth et alii, 1975). Não se encontrou, entretanto, na literatura disponível nenhum estudo a respeito dos efeitos de um programa de ensino baseado na MRC sobre o comportamento do estudante, embora haja suficientes proposições teóricas a enfatizarem a influência de procedimentos de avaliação sobre a aprendizagem (Karmel, 1974, p.22; Tyler, 1974, p.114), entre as quais pode-se citar, como típica, a de Cook (1951, p.4): "A avaliação educacional é um fator de facilitação da aprendizagem na medida em que contribui decisivamente para estabelecer situações de aprendizagem apropriadas às necessidades, habilidades e potencialidades de cada estudante; para

diagnosticar e corrigir dificuldades específicas de aprendizagem; para a motivação e direção das experiências de aprendizagem e para o desenvolvimento e manutenção de habilidades."

Não se pode subestimar a importância de se investigar em que medida um determinado componente de um sistema de ensino influencia o comportamento dos estudantes e em que sentido, já que a finalidade da educação é produzir mudanças no repertório de comportamentos do estudante (Chadwick, 1975, p. 61). Se é verdadeira a proposição de Tyler (1974, p.14), sobre um desses componentes, que diz: "os alunos são influenciados no seu estudo pela espécie de avaliação que será feita", então parece razoável perguntar-se sobre as possibilidades oferecidas pela MRC no sentido de influenciar positivamente o comportamento dos estudantes.

É nesse contexto que se formula o problema ao qual o presente estudo buscou dar uma resposta. Para sumariar: a MRC é um novo enfoque de avaliação derivado das necessidades específicas colocadas por determinados modelos de instrução; embora se reconheça a influência de procedimentos de avaliação sobre o comportamento dos estudantes, não existem estudos conhecidos na literatura disponível a respeito dos efeitos de um programa de avaliação baseado na MRC sobre aquele comportamento.

## 1.2. A FORMULAÇÃO DO PROBLEMA

As considerações feitas na sub-seção anterior permitem que se formule o seguinte problema: Quais os efeitos exercidos sobre o comportamento do estudante por um programa de ensino que use a MRC? Mais especificamente: Quais os efeitos de um programa de ensino que use a MRC sobre o rendimento e a atitude dos estudantes para com a disciplina de estudo, comparados com os de um programa que use a MRN?

## 2. REVISÃO DA LITERATURA

## 2.1. MEDIÇÃO COM REFERÊNCIA A CRITÉRIO E MEDIÇÃO COM REFERÊNCIA A NORMA ; CARACTERÍSTICAS DISTINTIVAS

Há uma tendência recente nos meios educacionais para relacionar o processo de avaliação em educação ao processo de tomada de decisão (Gronlund, 1976, p.4; Mehrens & Lehmann, 1973, p.3; Lindvall & Nitko, 1975, p.18). Várias são as situações educacionais que exigem tomada de decisão: "Como os alunos deveriam ser agrupados para uma aprendizagem mais efetiva? Em que medida os alunos estão atingindo os mínimos essenciais do curso? Em que ponto uma revisão seria mais benéfica?" (Gronlund, 1976, p.5); "Em que nível de um programa colocar um indivíduo? Em que medida um programa está sendo eficaz? Qual a competência de um indivíduo em relação a uma área particular?" (Gorth et alii, 1975, p.12); "Como o estudo de um determinado objetivo, tópico ou unidade, por um aluno, seria melhor conduzido? Quando um aluno dominou um objetivo instrucional específico?" (Lindvall & Nitko, 1975, p.18). Essas situações variam desde as que afetam milhares de pessoas (por exemplo, decisões a respeito de fundos federais) até as que afetam apenas uma pessoa (a decisão sobre se João reverá ou não sua lista de sinônimos) (Mehrens & Lehmann, 1973, p.3). A avaliação seria, segundo esse raciocínio, um meio de fornecer ao educador as informações úteis e necessárias para uma adequada tomada de decisão em situações que a requeressem, como as exemplificadas. Uma definição formal é fornecida por Stufflebean et alii (1971, p.40): "A avaliação educacional é o processo de delinear, obter e fornecer informação útil para o julgamento de decisões alternativas."

Uma implicação imediata da definição de avaliação dada por Stufflebean et alii é a de que a natureza da informação exigida é função direta do tipo de decisão. Por outro lado, os tipos de decisões possíveis dependem do Modelo de Instrução adotado. Um modelo que procure maximizar a aprendizagem de todos os estudantes certamente exigirá decisões de natureza bastante diversa das de um modelo que enfatize as diferenças individuais dos aprendizes. Por tanto, cada um desses modelos será compatível com um determinado enfoque de avaliação; daí a importância de se fazer uma breve análise do modelo ao qual melhor se adapte o enfoque de avaliação que é o objeto principal deste estudo.

A crescente massificação do ensino tem levado os educadores a delinear modelos de instrução baseados no pressuposto de que as diferenças de aprendizagem devem ser minimizadas e de que isto efetivamente pode acontecer desde que se conceda a cada estudante as oportunidades apropriadas a seu estilo de aprendizagem (Bloom, 1971; Glaser & Nitko, 1971; Block & Anderson, 1975). Um dos exemplares dessa corrente é a chamada "Aprendizagem para o Domínio". Segundo esse modelo a maioria dos estudantes pode dominar a maioria das tarefas de uma determinada área se lhe são concedidos tempo e oportunidades adequadas (Block, 1971).

Os procedimentos específicos de tais modelos são variáveis; entretanto, alguns elementos comuns podem ser ressaltados. Subjacente a todos está a definição de aprendizagem como um processo de transição do aprendiz de um estado inicial para um estado final arbitrariamente especificado, constituindo-se estes estados em conjuntos de comportamentos (Glaser & Nitko, 1971, p.627). A tarefa da instrução constitui-se em fornecer as condições para que essa transição ocorra. Essa concepção da aprendizagem e da instrução importa em que se destaque componentes bem definidos no processo: 1. O estado final (produtos de aprendizagem) deve ser previamente definido, em termos comportamentais; 2. É realizada uma diagnose detalhada do estado inicial do aprendiz; 3. São fornecidas alternativas educacionais adaptadas aos perfis de entrada do estudante; 4. À medida que o estudante aprende, seu desempenho é controlado e repetidamente avaliado; 5. A instrução e a aprendizagem ocorrem de forma cibernética, seguindo o desempenho e as seleções do estudante; 6. O sistema coleta informações a fim de aperfeiçoar-se (Glaser & Nitko, 1971, p.631).

São centrais nessa concepção as idéias de que devem ser tomadas decisões instrucionais em relação a cada indivíduo, antes, durante e após a instrução, de forma que seja possível controlar e corrigir o curso em cada ponto do processo e de que cada indivíduo assume uma parte das decisões, uma vez que se lhe dá oportunidade de selecionar o caminho a seguir, como se deduz do 5º componente de Glaser & Nitko (1971, p.631), já citado. Isto conduz à necessidade de que se obtenha constante informação sobre cada um dos comportamentos que fazem parte do perfil final desejável. A prática e a teoria tradicionais de avaliação educacional têm falhado em atender a esse requisito (Hambleton & Novik, 1973, p.150; Gorth, 1975, p.36), o que levou alguns educadores a proporem procedimentos de avaliação que o satisfizessem mais adequadamente.

A teoria clássica de medição educacional, referida aqui como **MEDIÇÃO COM REFERÊNCIA A NORMA (MRN)**, está baseada em dois pressupostos que não se ajustam às características dos modelos de instrução que procuram obter o máximo de competência de cada indivíduo: 1. Um bom instrumento de medição deve ser capaz de estabelecer discriminações refinadas entre os indivíduos de um determinado grupo em relação à capacidade medida: "Um dos principais objetivos da redação de itens para medições com referência a norma é torná-los tão discriminadores quanto possível. Através do controle de fatores como discriminabilidade entre respostas alternativas e vocabulário, procura-se maximizar a discriminação e portanto aumentar a variabilidade" (Payne, 1974, p.294); 2. A medição de uma determinada capacidade deve ser feita em relação a um grupo-norma: "É impossível estabelecer níveis ou tipos definidos de desempenho dos estudantes em aspectos como aproveitamento de objetivos instrucionais, capacidades mentais gerais e específicas, atitudes e interesses, sem relacionar o desempenho de cada um ao dos outros. A capacidade de soletrar de um estudante de uma dada idade cronológica ou série escolar, por exemplo, não pode ser interpretada significativamente como bem sucedida ou não através de um padrão absoluto. Podemos saber o grau em que esta capacidade é estimável somente comparando-a com a capacidade de soletrar de outros estudantes" (Remmers et alii, 1966, p.37).

Esses pressupostos são incompatíveis com os modelos de instrução descritos principalmente na medida em que aqueles são desenvolvidos com a finalidade de levar o maior número possível de estudantes à proficiência de determinados desempenhos previamente estabelecidos (Glaser, 1971a, p.6). Ora, um teste baseado na teoria clássica de medição deve obter a maior variância possível nos escores. Itens respondidos por 80% dos estudantes, por exemplo, são sistematicamente eliminados desse tipo de teste. Isso significa que mesmo itens que se refiram a conceitos ou habilidades considerados muito importantes numa determinada área serão, com o tempo, eliminados do teste (Popham, 1974, p.614).

Um enfoque alternativo de avaliação vem se desenvolvendo com a finalidade de preencher as lacunas deixadas pela teoria clássica de medição, resultantes sobretudo de sua origem histórica, ligada à Psicologia, e consequentemente às preocupações com medição das diferenças individuais, como se pode deduzir da evolução dos procedimentos de avaliação (Thorndike, 1970, p.13; Ebel, 1972b, p.11). Esse enfoque é conhecido como **MEDIÇÃO COM REFERÊNCIA A CRITÉRIO (MRC)**, em oposição ao clássico, **MEDIÇÃO COM REFERÊNCIA A NORMA (MRN)**. Uma definição formal de cada um deles faz-se necessária.

Uma MRC determina a posição de aproveitamento de um indivíduo em relação a um critério de desempenho pré-determinado, isto é, indica o grau de correspondência entre o que um indivíduo faz e o continuum de aproveitamento correspondente (Glaser, 1971, p.8; Popham & Husek, 1971, p.20; Mager, ... 1975, p.22).

Uma MRN determina a posição de aproveitamento de um indivíduo em relação ao aproveitamento de outros indivíduos, isto é, compara o desempenho do indivíduo com o do grupo (Glaser, 1971, p.8; Popham, 1971, p.20; Mager, 1975, p.21).

Fazer a correspondência entre o que o indivíduo faz e o conti.

num de aproveitamento correspondente supõe a definição precisa de objetivos . Nesse ponto é que se verifica o encontro maior entre a MRC e os modelos de instrução que enfatizam a maximização do domínio de todos os objetivos por todos os estudantes. A maioria das decisões tomadas em tal modelo estão associadas ao alcance de objetivos específicos de aprendizagem : quais objetivos são desejáveis, quais dos objetivos desejáveis fazem parte do repertório do estudante antes de sua entrada no sistema, em que ponto da seqüência de aprendizagem relacionado com um determinado conjunto de objetivos deve o estudante ser colocado, em relação a quais objetivos o estudante precisa de instrução adicional remediadora? Essas decisões básicas relacionadas com objetivos já eram ressaltadas na década de 40 por Tyler em obra clássica. Tyler (1974, p.1) destacava especificamente quatro tipos principais de decisões envolvidas no planejamento e desenvolvimento de um programa de instrução : 1. Que objetivos educacionais deve-se procurar atingir? 2. Que experiências educacionais podem ser oferecidas que tenham probabilidade de alcançar esses objetivos? 3. Como organizar eficazmente essas experiências educacionais? 4. Como se pode ter certeza de que esses objetivos estão sendo alcançados?

Por seu turno, a MRN não descarta a definição de objetivos como uma das fases do planejamento de instrumentos de medição, como se pode observar em qualquer texto clássico (Ebel, 1972b, por exemplo). Entretanto, como a interpretação dos resultados não é feita em relação aos objetivos, mas ao desempenho do grupo, e como o critério final da seleção de itens é seu poder de discriminação (Popham, 1974, p.614), essa fase tende a ser relegada a segundo plano pela MRN.

A formulação de objetivos, portanto, é uma fase essencial aos procedimentos de MRC. É para essa e outras fases de delineamentos concretos que se dirige agora a análise.

Por se constituir em uma tendência relativamente recente, raros são os relatos sobre delineamentos práticos de MRC. Já se pode notar, entretanto, alguns esforços de educadores em levar à prática os princípios da MRC. Entre esses pode-se destacar as tentativas de Gronlund e de Gorth et alii.

Gronlund (1973, p.3) lista uma série de tarefas que devem ser levadas a termo pelo elaborador de instrumentos de MRC : 1. definir e delimitar claramente um domínio de tarefas de aprendizagem; 2. definir os objetivos de instrução em termos comportamentais; 3. definir claramente os padrões de desempenho; 4. amostrar adequadamente o desempenho do estudante dentro de cada área; 5. selecionar os itens de teste na base de quão bem reflitam os comportamentos especificados nos objetivos instrucionais; 6. adotar um sistema de correção e comunicação que descreva adequadamente o desempenho do estudante em tarefas claramente definidas.

Todas as fases de Gronlund acima, exceto as 5 e 6, podem ser seguidas pela MRN sem que ela perca seu caráter, desde que o resultado obtido pelo estudante seja interpretado exclusivamente em termos do desempenho do grupo. O que distingue os dois enfoques, porém, é que as fases de 1 a 4 não precisam ser seguidas pela MRN, enquanto que são condições necessárias à elaboração de bons instrumentos de MRC.

Gronlund detalha cada uma das fases, indicando uma série de sugestões práticas para a elaboração de testes de MRC. Entretanto, apesar da ênfase no estabelecimento de objetivos e na interpretação dos resultados em relação a eles, pode-se dizer que suas indicações são bastante conservadoras, no sentido de que a maioria delas consta de manuais tradicionais de MRN.

Mais elaborados são os procedimentos descritos por Gorth et alii (1975). Esses autores sugerem uma estratégia sofisticadíssima de MRC, incluindo o uso de computadores. A partir da divisão da estrutura instrucional em unidades administrativas (módulos, por exemplo) e da formulação de objetivos e sua distribuição pelas unidades é proposto um delineamento de avalia -

ção que reflete as necessidades do tomador de decisão por tipos e quantidades específicos de informação a certos intervalos de tempo. Os autores apresentam quatro desses delineamentos, os quais refletem os diferentes tipos de decisões a serem tomadas: Delineamento de Unidade, Delineamento de Unidade Móvel, Delineamento Padrão e Delineamento Padrão/Unidade. Para uma possível comparação do procedimento adotado no presente estudo, descreve-se com algum detalhe os delineamentos de Gorth et alii.

O Programa Instrucional hipotético desses autores, para fins de descrição dos delineamentos, está dividido em 8 unidades de duas semanas de duração cada uma. Cada unidade tem 6 objetivos, estando disponível um conjunto de 6 itens de teste para cada objetivo.

O Delineamento de Unidade consiste em um pré-teste geral do Curso, um pós-teste de domínio de cada unidade e um pós-teste geral de Curso. Cada forma do pré-teste e do pós-teste gerais contém um item para cada objetivo do curso, sendo o pré-teste administrado à metade dos estudantes e o pós-teste à outra metade, obtendo-se desta forma uma estimativa imperfeita do rendimento individual e uma indicação mais fidedigna do rendimento do grupo no início e no final do curso. Cada pós-teste de unidade contém 4 itens para cada um dos 6 objetivos, gerando assim informação sobre cada estudante, com os dados reunidos fornecendo informação sobre o rendimento do grupo na unidade.

O Delineamento de Unidade Móvel consiste em um pré-teste e um pós-teste gerais de curso e em duas formas de teste, aplicadas ao final de cada unidade. O pré e o pós-teste gerais têm conteúdo idêntico aos do delineamento anterior. Os testes de unidade, entretanto, cobrem objetivos de três unidades: da recém completada, da unidade anterior e da seguinte. Portanto, tem-se assim, dados sobre retenção e dados pré e pós-instrucionais de cada unidade. O teste administrado após cada unidade, em duas formas, contém seis itens relacionados à unidade anterior, 36 itens relacionados à unidade recém terminada e 6 itens relacionados à unidade seguinte.

O Delineamento Padrão consiste em 10 formas de teste com 24 itens cada. Cada uma delas cobre os objetivos de todo o curso, contendo 3 itens de cada uma das oito unidades, com cada item relacionado a um objetivo diferente. No começo do curso, cada grupo, correspondendo a um décimo dos estudantes, responde a uma forma diferente do teste. Ao final da segunda semana o processo é repetido, cada grupo respondendo a uma forma diferente da que respondeu anteriormente e assim por diante, dez vezes durante o curso. Desta maneira cada estudante, durante o curso, responde uma e somente uma vez a cada forma. À medida que o curso avança obtém-se menos dados pré-instrucionais e mais dados sobre retenção, enquanto a quantidade de dados pós-instrucionais é constante.

O Delineamento Padrão/Unidade consiste em 6 testes gerais e 4 conjuntos (de 2 formas cada) de testes de unidade. Cada forma do teste geral inclui 3 itens para cada unidade do curso e é respondida por grupos constituídos de 1/6 dos estudantes em 6 ocasiões, durante o curso, de forma similar à do delineamento anterior. Cada conjunto de teste de unidade cobre 2 unidades consecutivas, com cada forma contendo 2 itens por objetivo. A metade dos estudantes responde a uma das formas ao final de cada duas unidades e a outra metade à outra forma.

Embora a descrição dos delineamentos de Gorth et alii tenha sido demasiado sucinta é possível perceber claramente o aproveitamento das potencialidades da amostragem de objetivos, de itens, de estudantes e de ocasiões para gerar dados que se pretende adequados ao tipo de decisão que se necessita. Esse aproveitamento é crucial em MRC, principalmente por causa da enorme massa de dados que é gerada em relação a cada objetivo. "Poder-se-ia imaginar", diz Millman (1974, p.32), "o fornecimento de todos os itens em um pequeno domínio a todos no grupo; mas mesmo que isso fosse possível, raramente valeria a pena administrar todo item a toda pessoa em todas as ocasiões. Os

custos em tempo e esforço para examinadores e examinandos seriam extraordinariamente proibitivos."

Tendo-se definido a MRC, exposto o raciocínio que a justifica e descrito procedimentos específicos, é conveniente agora apresentar também algumas considerações contrárias à sua adoção. A exposição desses argumentos facilitará o confronto com o raciocínio favorável já apresentado.

Um dos mais notáveis críticos é Ebel (1972a, p.147), para quem a exigência da definição clara de objetivos específicos e de critérios de desempenho constitui-se em uma séria limitação da MRC. Segundo este autor a MRC pode ser praticável naquelas áreas que enfatizam o cultivo de um grau alto de habilidade no exercício de um número limitado de competências. Em áreas em que a ênfase é sobre o conhecimento e a compreensão o uso efetivo da MRC parece muito menos provável. Pois o conhecimento e a compreensão, argumenta ele, consistem em uma estrutura complexa que deve sua potência e beleza a uma infinidade de delicadas fibras de relacionamento. O conhecimento não é constituído de pedaços discretos que possam ser definidos e identificados separadamente. O problema está, segundo o mesmo autor, em que as habilidades básicas constituem somente uma pequena fração do que as escolas ensinam e do que as várias pessoas estão interessadas em aprender e somente essas admitem o estabelecimento de um critério de domínio. Habilidades mais complexas, compreensões e apreciações não são, na experiência de quase todas as pessoas, meramente uma questão de gradação. Nenhuma, a não ser as mais simples delas, pode jamais ser completamente dominada por alguém. Portanto, conclui Ebel, qualquer critério de domínio será provavelmente um tanto imperfeito e arbitrário. Logicamente, olhando de um outro ponto de vista, esses argumentos demonstram um posicionamento desse autor a favor da MRN.

Gronlund (1973, p.16) parece concordar com Ebel. Esse autor distingue dois tipos de objetivos: objetivos de domínio e objetivos de desenvolvimento. Os primeiros são objetivos referentes ao mínimo essencial que todo estudante deve dominar em determinada área de aprendizagem e os últimos são aqueles que ultrapassam a complexidade de um nível mínimo de proficiência. Para Gronlund a MRC é bastante adequada para objetivos de domínio pelas seguintes razões: 1. os produtos de aprendizagem são relativamente simples (conhecimento e habilidades básicas); 2. o domínio de comportamento é bastante limitado (o mínimo essencial); 3. a aprendizagem é frequentemente sequencial por natureza; 4. o escore na forma de porcentagem correta fornece um relato significativo (indica o grau de progresso em relação ao domínio completo). Estas condições tornam possível definir mais claramente o domínio de comportamento a ser testado, a amostrar mais adequadamente as tarefas de aprendizagem a serem incluídas no teste e a estabelecer mais facilmente padrões de critério para julgar e relatar o desempenho do estudante.

Por outro lado, na opinião de Gronlund, a dificuldade em usar os testes com referência a critério no nível de desenvolvimento (isto é, aprendizagem além do mínimo essencial) resulta em grande extensão da inexistência das condições citadas, neste nível. Ao invés disso, os produtos de aprendizagem são complexos (e.g., compreensão, habilidades de pensamento), o domínio de tarefas de aprendizagem é virtualmente ilimitado e a aprendizagem raramente é feita através de uma seqüência clara de estágios. Em adição, os objetivos instrucionais representam metas em direção às quais trabalha-se antes que metas a serem integralmente alcançadas, pois aqui a ênfase está no desenvolvimento contínuo da compreensão e das habilidades. Cada estudante é encorajado a empenhar-se para obter o máximo nível de aproveitamento e excelência de que ele é capaz, antes que o domínio de algum conjunto pré-determinado de mínimos essenciais.

A arbitrariedade da seleção de um critério de domínio é também a objeção feita por Mueller (1973, p.6) à MRC. Selecionar tal critério, para ele, já constitui um problema em objetivos relacionados a habilidades sim-

ples, tornando-se uma dificuldade praticamente intransponível em objetivos mais complexos.

Essas colocações conduzem a questões de filosofia da educação e da estrutura do conhecimento, sujeitas a discussões que escapam ao foco deste estudo. Importa aqui indagar-se das conseqüências da adoção na prática educacional do dia-a-dia de procedimentos de avaliação baseados na MRC.

## 2.2. EFEITOS DE PROCEDIMENTOS DE AVALIAÇÃO SOBRE O COMPORTAMENTO DOS ESTUDANTES

Uma dessas conseqüências diz respeito ao comportamento dos estudantes. Pois, se a MRC fornece informações úteis para a tomada de decisão em situações relacionadas à aprendizagem do estudante, parece útil perguntar - se qual a sua influência direta sobre o rendimento e as atitudes do mesmo, já que alguns teóricos, como se verá a seguir, apontam a avaliação como responsável por uma parte da variância do rendimento e das atitudes em relação ao estudo.

Essa é a opinião de Tyler (1974, p.14), para quem a avaliação exerce uma poderosa influência sobre a aprendizagem. "A Enquete dos Regentes da Universidade do Estado de Nova Iorque", diz este autor, "mostrou que as "Regent's Examinations", que são os instrumentos de avaliação do Estado, têm mais efeito sobre o que se ensina naquela unidade da União do que os planos de curso de estudo como tais. Os alunos são influenciados no seu estudo pela espécie de avaliação que será feita, e mesmo os professores são influenciados na ênfase que dão a determinados aspectos do ensino pela espécie de avaliação que esperam seja feita."

Observações semelhantes são feitas por Kuethe (1974, p.145), embora refiram-se de maneira específica a notas. Para este autor "as notas possuem alguns dos atributos dos bons instrumentos de motivação... As notas podem motivar os alunos a fazerem novos esforços, proporcionando-lhes um conhecimento objetivo dos resultados." É fácil ver que estas proposições podem ser estendidas à avaliação em geral.

Opiniões similares de Karmel (1974, p. 22) sobre testes também podem ser estendidas à avaliação, de modo geral. Segundo este autor, "os testes bem construídos podem melhorar a aprendizagem contribuindo para que o estudante forme hábitos de estudo e dirija sua energia intelectual para os objetivos desejados. Os resultados de teste podem revelar as áreas em que cada estudante obtém melhor ou pior desempenho e também atuar como motivadores em estudos futuros."

O aspecto motivador da avaliação é ressaltado também por Cook (1951, p.4), que lista entre as funções da avaliação a de facilitação da aprendizagem. Para ele a avaliação é importante fonte de motivação e fator de direção da aprendizagem.

No mesmo trabalho (p.38) esse autor diz que a condição motivante dos testes apresenta-se através de três funções: 1. a função energizante, para aumentar o nível geral de atividade e esforço; 2. a função diretiva, para guiar a atividade variável e persistente do organismo para canais desejáveis e 3. a função seletiva, para determinar as respostas que serão fixadas e as que serão eliminadas.

É interessante examinar com algum detalhe essas funções componentes da condição motivante dos testes. Este exame poderá auxiliar a compreender

der e a desvendar os mecanismos do relacionamento entre avaliação e comportamento do estudante.

Quanto à função energizante, Cook (1951, p.38) observa que a extensão com que os exames aumentam o nível geral de atividade de aprendizagem e o esforço é atestado pelas sessões de estudo intensivo que precedem os exames. Os exames, segundo ele, determinam em grande parte quando os alunos estudam, o que estudam e como estudam.

Em relação à função diretiva, diz o mesmo autor, que é difícil enfatizar demasiado a importância dos procedimentos de exame na determinação do que os professores ensinam e como ensinam, do que os alunos aprendem e como aprendem. Por exemplo, ao se prepararem para os exames tipo ensaio, os estudantes tendem a esboçar e a organizar o material sistematicamente em grandes unidades enfatizando relacionamentos, tendências e reações pessoais. Ao se prepararem para o tipo de amostragem ampla dos exames objetivos fatuais, os estudantes enfatizam os detalhes sobre fatos, nomes, datas e os resultados de experimentos específicos.

Quanto à função seletiva dos exames é ainda o mesmo autor (p. 41) quem assinala que a extensão com que os testes auxiliam a fixar os comportamentos corretos e desejáveis e a eliminar os erros dependem não somente da natureza do instrumento mas também de como ele é corrigido e da ênfase posta nos erros individuais e no trabalho de recuperação feito no processo de acompanhamento. A evidência lógica e experimental, continua Cook, indica que na situação de exame quanto mais imediato e direto é o conhecimento do estudante de onde e porque ele está correto e porque ele está errado, maior a tendência a fixar as respostas corretas.

Dentro da mesma linha de pensamento Mehrens & Lehmann (1973, p.11) destacam que a avaliação auxilia o estudante a : 1. compreender a comunicação dos objetivos do professor; 2. aumentar a motivação; 3. encorajar bons hábitos de estudo e 4. providenciar retro-alimentação que identifique seus pontos fortes e seus pontos fracos. Para esses autores, um aspecto importante da aprendizagem é que o estudante deve estar consciente de seus pontos fracos e seus pontos fortes. A avaliação e a retro-alimentação subsequente podem exercer um importante papel em guiar os esforços futuros do estudante.

Dessas proposições pode-se deduzir que a interpretação explícita dos resultados de testes em termos de comportamento, fornecida pela MRC, teria maiores probabilidades de influir positivamente sobre a motivação do estudante por causa do sentimento de auto-controle proporcionado. Em consequência essa condição motivante talvez influenciasse diretamente seu rendimento, suas atitudes e estratégias de direção de sua própria aprendizagem. Por outro lado, pode-se especular sobre a hipótese de que a natureza da interpretação dos resultados proporcionada pela MRC forneça uma retro-alimentação mais adequada ao estudante sobre seu desempenho, auxiliando-o, portanto, a direcionar melhor seus esforços de aprendizagem. Se é verdade, segundo Thorndike & Hagen (1970, p.17) que "quanto mais adequadas e exatas sejam as informações que, acerca de si mesmo, podemos oferecer ao indivíduo, tanto mais facilmente tomará estas decisões que o ajudem a tratar seus próprios problemas", então parece lógico especular sobre as possibilidades oferecidas nesse sentido pela MRC, tanto em relação ao rendimento cognitivo quanto a atitudes.

A área de atitudes exige uma análise especial, principalmente porque parece ser um domínio a que se tem dado pouca atenção na prática educacional. A importância de se incluir o desenvolvimento de atitudes positivas no rol de objetivos a serem atingidos por um programa instrucional é ressaltada por Mager (1971, p.25) na seguinte cadeia de argumentos : "1. o aprender para o futuro; isto é, o propósito da instrução é facilitar certo tipo de conduta em um ponto posterior ao término da instrução; 2. a probabilidade de que o aluno ponha em prática seus conhecimentos é afetada por sua atitude a favor ou contra a disciplina de estudos; o que o desgosta tende a esquecer; 3. umas

pessoas influem em outras. Os professores e os demais exercem influência sobre a atitude para com a disciplina e sobre a própria aprendizagem; 4. um objetivo pelo qual há de esforçar-se, portanto, é que o estudante deixe a influência do professor com uma atitude tão favorável quanto possível para com a sua disciplina. Assim ele auxiliará o estudante a aumentar ao máximo a possibilidade de que ele recorde o que lhe foi ensinado e de que esteja disposto a aprender mais sobre a mesma disciplina." Para este mesmo autor (p.52) há três fatores principais de influência sobre as atitudes para com uma disciplina: as condições que a rodeiam; as consequências de pôr-se em contato com ela e a forma em que os demais reagem a respeito da mesma.

Existe uma certa lógica em pensar-se nas possibilidades oferecidas por um sistema de avaliação baseado na MRC, no sentido de cercar a disciplina de condições mais atrativas para o estudante. A MRC oferece ao estudante um conhecimento mais completo e preciso de seu aproveitamento, permitindo com isso que ele tome uma série de decisões instrucionais. É razoável, portanto, supor-se que essa oportunidade de participação instrucional aumente a probabilidade de que o estudante desenvolva uma atitude mais favorável para com a disciplina de estudo.

Sumariando o que foi revisto até aqui sobre o relacionamento entre avaliação e comportamento do estudante, pode-se dizer que, apesar das restrições colocadas por seus opositores, pode-se esperar que algumas das funções da avaliação educacional sejam melhor cumpridas pela MRC. Por interpretar os resultados em termos de desempenho explícito e portanto fornecer ao estudante um meio útil para seu auto-controle e auto-correção pode-se prever que a MRC exercerá efeitos positivos sobre a conduta do estudante, seu rendimento e suas atitudes. Uma análise lógica da MRC demonstra que ela possui as propriedades colocadas por Dressel (apud Payne, 1974, p.9): "a avaliação é mais suscetível de conduzir à aprendizagem quando providencia e encoraja a auto-avaliação e é mais significativa para a aprendizagem quando permite e encoraja o exercício da iniciativa individual."

Parece haver pouca correspondência entre as colocações teóricas sobre o relacionamento entre avaliação e comportamento do estudante e estudos experimentais na mesma direção. Entre esses raros estudos podem ser citados dois.

Um deles, de autoria de Almeida (1976), comparou o resultado de aprendizagem em três áreas de estudo de dois grupos de estudantes: um deles submetido a procedimentos de avaliação convencional, o outro a procedimentos de avaliação formativa. Constatou o autor que os resultados de aprendizagem dos alunos sob o modelo de avaliação formativa foram superiores nas três disciplinas aos do grupo sob processo de avaliação convencional, embora do ponto de vista estatístico, só se tenha verificado diferença significativa em uma delas. Observou ainda o mesmo estudo a evidência de que o modelo de avaliação formativa incrementou a maior participação dos alunos nas situações de aprendizagem em classe, e os levou a estudarem mais que o habitual e permitiu o avanço mais rápido do programa.

O outro, de autoria de Chassot (1976) comparou o resultado de aprendizagem e o grau de satisfação de dois grupos de estudantes: um submetido a testes objetivos, o outro a questões de resposta livre. Os resultados mostraram que os escores obtidos por um mesmo aluno, com um ou outro tipo de instrumento, estavam altamente correlacionados e que não foi verificada diferença significativa entre os escores finais dos dois grupos de alunos. Constatou ainda o estudo deste autor que o grau de satisfação dos alunos foi maior quando forma usadas provas objetivas.

Não se pode discutir a consistência de resultados de apenas dois estudos. Pode-se, entretanto, notar que existe algum indício de que os procedimentos de avaliação são responsáveis por uma parte da variância de alguns aspectos do comportamento do estudante. Certamente serão necessárias ain-

da muitas pesquisas, para que, dos resultados acumulados possa se extrair alguma generalização válida.

Se já são raros os estudos relacionando avaliação e comportamento do estudante, os que relacionam MRC e conduta estudantil inexistem, ao menos na bibliografia disponível. É possível que isto se deva, principalmente, ao fato de ser esta uma idéia recente. O presente artigo contribui para uma tentativa de, a partir das proposições teóricas expostas, preencher uma parte dessa lacuna.

### 3. HIPÓTESES

Da fundamentação exposta foram deduzidas as seguintes hipóteses, que direcionaram o presente estudo :

HIPÓTESE 1 : Alunos submetidos a um programa de ensino que use a MRC alcançam rendimento significativamente maior que alunos submetidos a um programa que use a MRN.

HIPÓTESE 2 : Alunos submetidos a um programa de ensino que use a MRC têm atitudes significativamente mais favoráveis para com a disciplina de estudo que alunos submetidos a um programa que use a MRN.

## 4. METODOLOGIA

Para responder às perguntas colocadas pelo problema e testar as hipóteses resultantes fez-se uso de um estudo tipo "experimental", optando-se para a Hipótese 1 por um delineamento que Campbell & Stanley (1973, p. 32) chamam de "delineamento de grupo de controle pré-teste-pós-teste", e para a hipótese 2 por um delineamento que os mesmos autores (p.55) chamam de "delineamento de grupo de controle com pós-teste unicamente".

Em ambos os delineamentos os sujeitos são distribuídos aleatoriamente por dois grupos. No primeiro é feita uma medição da variável dependente antes e depois do tratamento, enquanto que no segundo tipo de delineamento a variável independente é medida somente após o tratamento. Os detalhes da forma em que este delineamento foi aplicado ao presente estudo são descritos na subseção 4.4.

Se - R 01 X 02  
Se - R 03 04

Se - R X 01  
Se - R 02

↓  
comparação  
na

## 4.1. DEFINIÇÃO DAS VARIÁVEIS

### 4.1.1. VARIÁVEL INDEPENDENTE : TIPO DE AVALIAÇÃO

4.1.1.1. Medição com referência a critério - procedimento de avaliação em que a interpretação dos resultados de aproveitamento do estudante é feita em relação a um critério de desempenho pré-determinado.

4.1.1.2. Medição com referência a norma - procedimento de avaliação em que a interpretação dos resultados de aproveitamento do estudante é feita em relação ao grupo, considerado como norma.

### 4.1.2. VARIÁVEIS DEPENDENTES

4.1.2.1. Rendimento - grau em que os objetivos de todo o curso são atingidos pelo estudante, conforme medido por um teste de 50 itens objetivos administrado ao fim do programa instrucional ( Anexo 9.2 ).

4.1.2.2. Atitude para com a disciplina - conjunto dos sentimentos dos estudantes em relação a disciplina em questão, conforme medidos por uma Escala de Atitudes tipo Likert ( Anexo 9.1 ).

## 4.2. SUJEITOS

Foram sujeitos do experimento 36 alunos da segunda série do curso de Pedagogia de uma Faculdade de Educação do interior catarinense. A seleção de alunos do 3º grau deveu-se principalmente à possibilidade de uma maior compreensão dos mecanismos introduzidos pelos procedimentos por parte de alunos desse nível. Uma lista de objetivos, por exemplo, pouco significado teria para um aluno do 2º grau e menos ainda para um aluno do 1º grau. A escolha do curso de Pedagogia também está relacionada a este critério, já que os sujeitos haviam estudado previamente o tema "Objetivos Educacionais". Pensou-se ainda na oportunidade de difusão da inovação e de uma maior receptividade por parte de futuros especialistas em Educação.

Estava-se consciente de que os sujeitos não constituíam uma amostra representativa por causa das peculiaridades apontadas. Entretanto, como se explicou, a existência dessas peculiaridades era uma imposição necessária, ao menos no atual estágio de desenvolvimento da MRC (difusão quase nula). É importante ter este fato em mente para a análise dos dados.

Os sujeitos eram todos do sexo feminino, com idade média de 23 anos, provenientes de cursos de formação de magistério para o 1º grau, pertencentes a classes sócio-econômicas de renda média em torno de três salários mínimos e com instrução paterna média ao nível de 1º grau (informações fornecidas pela Secretaria da Faculdade).

### 4.3. INSTRUMENTOS

Usou-se dois tipos de instrumentos: um teste de rendimento, para medição da variável dependente RENDIMENTO e uma Escala de Atitudes tipo Likert, para medição da variável dependente ATITUDE PARA COM A DISCIPLINA. Ambos os instrumentos foram pré-testados com 23 sujeitos de características similares às dos sujeitos deste estudo, já que pertenciam a uma outra turma do mesmo curso, com apenas um semestre de adiantamento.

#### 4.3.1. Teste de rendimento

O Teste de Rendimento (Anexo 9.2) é formado por 50 itens de escolha múltipla, referentes aos objetivos e conteúdos trabalhados no Programa Instrucional (Anexo 9.3) a que foram submetidos os sujeitos. O teste foi elaborado pelo autor especialmente para uso neste estudo.

Procurou-se maximizar a validade de conteúdo do teste através da elaboração de uma tabela de especificações. Distribuindo-se os itens de acordo com a importância relativa dos conteúdos e objetivos e do tempo instrucional a eles dedicado, e de se esperar tenha-se garantido uma amostragem adequada.

A fidedignidade foi estimada na fase de pré-testagem, usando-se o Método das Metades. A partir de uma única aplicação do teste, calculou-se o coeficiente de correlação de Spearman entre os escores dos itens pares e os itens ímpares, estimando-se o coeficiente de fidedignidade do teste inteiro através da Fórmula de Spearman-Brown, obtendo-se um coeficiente de fidedignidade de 0,80. Como observa Vianna (1973, p.166), um coeficiente de fidedignidade deve ser julgado em função da importância da decisão a ser tomada, sendo aceitável, entretanto, para a maioria das decisões, um coeficiente mínimo de 0,70. Por esse critério o coeficiente obtido é bastante aceitável.

#### 4.3.2. Escala de atitudes para com a disciplina

A Escala de atitudes para com a disciplina (Anexo 9.1) é uma escala do tipo Likert, com 20 itens. Esses itens foram elaborados a partir de uma escala semelhante construída por Koff (1975) para medir atitudes de alunos do 2º grau para com a disciplina de Biologia e de uma escala para medir atitudes para com qualquer disciplina, devida a Remmers (apud Payne, 1974, p.173).

Era importante para o estudo que os itens da escala fossem constituídos de sentenças que representassem de fato atitudes dos estudantes para com a disciplina. Devido à inexistência de um critério externo contra o qual pudesse ser verificada esta suposição (validação de critério) e à impossibilidade de realizar uma laboriosa validação de construto, concentrou-se a

preocupação numa validação de conteúdo. Supõe-se tê-la alcançado, em certa medida, por um lado, pela maneira com que foi elaborada a escala de Koff : a alunos de terceira do 2º grau ( portanto, de características educacionais próximas aos dos sujeitos deste estudo) pediu-se que escrevessem 3 ou 4 sentenças, expressando seus sentimentos em relação à Biologia; a partir destas sentenças foi construída a escala final. Por outro lado, supõe-se estar reforçando a validade assim obtida ao introduzir alguns itens de escala elaborada por especialista.

A fidedignidade foi estimada na fase de pré-testagem, usando-se o método do teste-reteste. A escala foi aplicada aos 23 sujeitos que se submeteram à pré-testagem, em duas ocasiões, num intervalo de 20 dias. A partir dos dois conjuntos de escores, calculou-se o coeficiente de Spearman, obtendo-se um coeficiente de fidedignidade de 0,79. Pelo critério de Vianna, descrito na subseção anterior, pode-se considerá-lo aceitável.

#### 4.4. PROCEDIMENTOS

Para testar as hipóteses formuladas elaborou-se um programa instrucional da disciplina Medidas Educacionais, constante de 5 unidades instrucionais, com duração de um semestre letivo de 15 semanas ( Anexo 9.3). Devem ser ressaltados aqui a importância e a dificuldade da formulação de delineamentos de MRC para a prática educacional. A operacionalização dos conceitos teóricos apresenta uma série de problemas que, embora à margem do objetivo central do estudo, o autor tentou resolver. Por isto, o programa instrucional resultante, principalmente no que diz respeito ao componente "avaliação", representa um produto tão importante para o autor quanto os resultados que dão resposta direta ao problema colocado pelo estudo.

Todos os sujeitos participaram, simultaneamente e no mesmo local, de todas as atividades de aprendizagem coletivas. As aulas foram conduzidas por um único professor, o autor.

Ao final de cada uma das cinco unidades foi aplicado um teste. Os itens dos testes de unidade foram elaborados em estrita correspondência com os objetivos da unidade, cuja lista mimeografada era entregue a cada aluno antes de seu início. Cuidou-se para que os itens constituíssem uma amostra representativa dos objetivos. No processo de fazer corresponder o item usou-se o método descrito por Mager (1975). Por este método gerador de itens, procura-se fazer com que o item expresse tão fielmente quanto possível o comportamento descrito no objetivo. O Anexo 9.4 mostra uma lista de objetivos de uma das unidades e o anexo 9.5 o teste correspondente.

Para a comunicação dos resultados dos testes de unidade dividiu-se a turma, aleatoriamente, em dois grupos. Ambos recebiam os resultados através de uma Ficha de Informação dos Resultados (FIR). Um dos grupos, constituído de 16 estudantes, recebia a FIR em termos dos objetivos fixados para a unidade ( Anexo 9.6), isto é, eram submetidos a uma Medição com Referência a Critério. Em relação a cada objetivo o aluno era informado se o havia atingido ou não. Este grupo foi chamado de Grupo MRC.

O outro grupo, com 17 estudantes, recebia a FIR em termos do desempenho do próprio grupo, isto é, eram submetidos a uma Medição com Referência a Norma. Para isto usou-se uma norma de porcentagem ( Anexo 7.7). Este grupo foi chamado de Grupo MRN. Além disto, os dois grupos eram igualmente informados dos resultados em termos do sistema regulamentar da Instituição ( uma nota, variando de 0 a 10).

1 Sem  
devidado

A variável dependente RENDIMENTO foi medida imediatamente antes e imediatamente após o programa instrucional, através de instrumento já descrito. A variável dependente ATITUDE PARA COM A DISCIPLINA foi medida imediatamente após o programa instrucional, através de instrumento já descrito.

#### 4.5. TRATAMENTO ESTATÍSTICO

Cada sujeito recebeu um escore no teste de rendimento, resultante da atribuição de 1 ponto para itens corretamente respondidos e 0 ponto para itens incorretamente respondidos. Não se adotou nenhum procedimento de correção para adivinhação.

O escore para a escala de atitudes foi obtido da maneira usual. Os itens favoráveis recebiam os seguintes pontos, conforme a opção selecionada: Concordo plenamente - 5 pontos; Concordo - 4; Indeciso - 3; Discordo - 2; Discordo plenamente - 1. Os itens desfavoráveis eram assim contados: Concordo plenamente - 1 ponto; Concordo - 2; Indeciso - 3; Discordo - 4; Discordo plenamente - 5. Os itens neutros não contribuíam em nada para o escore final.

Para fins de teste estatístico foram formuladas as seguintes hipóteses nulas, em correspondência com as hipóteses anteriormente formuladas:

HIPÓTESE NULA 1: Não existe diferença significativa de rendimento entre alunos submetidos a um programa instrucional que use a MRC e alunos submetidos a um programa instrucional que use a MRN.

HIPÓTESE NULA 2: Não existe diferença significativa de atitude para com a disciplina de estudo entre alunos submetidos a um programa instrucional que use a MRC e alunos submetidos a um programa instrucional que use a MRN.

Considerando-se a natureza ordinal das escalas usadas para medir as variáveis dependentes e a independência das amostras que constituem os grupos MRC e MRN, usou-se, para testar as hipóteses nulas acima, o teste U de Mann-Whitney (Conover, 1971, p.224). Adotou-se, para efeitos de decisão, sobre a rejeição ou não das hipóteses nulas o nível de significância de 0,05.

## 5.1. HIPÓTESE 1

Embora a atribuição de sujeitos aos dois grupos tenha sido realizada de forma aleatória, o teste de rendimento ( Anexo 9.2 ) foi aplicado também antes do programa instrucional. A Tabela 1 mostra a distribuição de frequência dos pontos obtidos pelos dois grupos nesse pré-teste. Como se observa, as distribuições são semelhantes, ambas concentradas no extremo inferior, embora a média do grupo MRC seja um pouco maior.

TABELA 1

Distribuição de frequência dos pontos obtidos no pré-teste de rendimento pelos grupos MRC e MRN

Nº de pontos	Frequência	
	MRC	MRN
9-12	1	2
13-16	3	9
17-20	7	4
21-24	4	2
25-28	1	-
N	16	17
Média	17,9	15,7
Desvio-padrão	4,1	3,3

Aplicado o teste U de Mann-Whitney à essa distribuição, não se verificou diferença significativa, conforme mostra a Tabela 2. Deste modo, pode-se dizer que os dois grupos, MRC e MRN, eram equivalentes, ao início da instrução, quanto à variável "rendimento".

TABELA 2

Resultado do teste da hipótese nula sobre rendimento dos grupos MRC e MRN, antes da instrução

$U_{\text{observado}}$	$U_{\text{crítico}}$	Hipótese nula
90	89	Não rejeitada

Nível de significância : 0,05

A distribuição de frequência dos pontos obtidos pelos sujeitos dos grupos MRC e MRN é apresentada na Tabela 3. Essas distribuições diferem em alguns aspectos. A distribuição dos pontos do grupo MRN apresenta uma maior variabilidade, como se pode concluir comparando suas amplitudes totais (MRC=5,2; MRN=7,8). Além disso a distribuição do grupo MRN tende a se concentrar um pouco mais na sua parte superior, como demonstra uma comparação de suas medianas (a do grupo MRC está no intervalo 25-28 e a do grupo MRN, no intervalo 29-32).

**TABELA 3**

Distribuição de frequência dos pontos obtidos no pós-teste de rendimento, pelos grupos MRC e MRN

Nº de pontos	Frequência	
	MRC	MRN
9-12	-	1
13-16	-	1
17-20	-	1
21-24	5	2
25-28	6	3
29-32	1	5
33-36	3	2
37-40	1	2
N	16	17
Média	27,9	26,9
Desvio-padrão	5,2	7,8

Essas diferenças, entretanto, são pequenas, como o é a diferença entre as médias (MRC = 27,9; MRN = 26,9). O teste U, de Mann-Whitney, conforme mostra a Tabela 4, confirma essa observação: a diferença entre as distribuições de frequência dos grupos MRC e MRN não é significativa. A hipótese nula 1, testada unidirecionalmente ao nível de significância de 0,05, não pôde ser rejeitada.

**TABELA 4**

Resultado do teste da hipótese nula 2

U observado	U crítico	Hipótese nula
133	89	Não rejeitada

Nível de significância : 0,05

## 5. RESULTADOS E DISCUSSÃO

## 5.1. HIPÓTESE 1

Embora a atribuição de sujeitos aos dois grupos tenha sido realizada de forma aleatória, o teste de rendimento ( Anexo 9.2 ) foi aplicado também antes do programa instrucional. A Tabela 1 mostra a distribuição de frequência dos pontos obtidos pelos dois grupos nesse pré-teste. Como se observa, as distribuições são semelhantes, ambas concentradas no extremo inferior, embora a média do grupo MRC seja um pouco maior.

TABELA 1

Distribuição de frequência dos pontos obtidos no pré-teste de rendimento pelos grupos MRC e MRN

Nº de pontos	Frequência	
	MRC	MRN
9-12	1	2
13-16	3	9
17-20	7	4
21-24	4	2
25-28	1	-
N	16	17
Média	17,9	15,7
Desvio-padrão	4,1	3,3

Aplicado o teste U de Mann-Whitney à essa distribuição, não se verificou diferença significativa, conforme mostra a Tabela 2. Deste modo, pode-se dizer que os dois grupos, MRC e MRN, eram equivalentes, ao início da instrução, quanto à variável "rendimento".

TABELA 2

Resultado do teste da hipótese nula sobre rendimento dos grupos MRC e MRN, antes da instrução

U observado	U crítico	Hipótese nula
90	89	Não rejeitada

Nível de significância : 0,05

A distribuição de frequência dos pontos obtidos pelos sujeitos dos grupos MRC e MRN é apresentada na Tabela 3. Essas distribuições diferem em alguns aspectos. A distribuição dos pontos do grupo MRN apresenta uma maior variabilidade, como se pode concluir comparando suas amplitudes totais (MRC=5,2; MRN=7,8). Além disso a distribuição do grupo MRN tende a se concentrar um pouco mais na sua parte superior, como demonstra uma comparação de suas medianas ( a do grupo MRC está no intervalo 25-28 e a do grupo MRN, no intervalo 29-32).

**TABELA 3**

Distribuição de frequência dos pontos obtidos no pós-teste de rendimento, pelos grupos MRC e MRN

Nº de pontos	Frequência	
	MRC	MRN
9-12	-	1
13-16	-	1
17-20	-	1
21-24	5	2
25-28	6	3
29-32	1	5
33-36	3	2
37-40	1	2
N	16	17
Média	27,9	26,9
Desvio-padrão	5,2	7,8

Essas diferenças, entretanto, são pequenas, como o é a diferença entre as médias ( MRC = 27,9; MRN = 26,9 ). O teste U, de Mann-Whitney, conforme mostra a Tabela 4, confirma essa observação : a diferença entre as distribuições de frequência dos grupos MRC e MRN não é significativa. A hipótese nula 1, testada unidirecionalmente ao nível de significância de 0,05, não pode ser rejeitada.

**TABELA 4**

Resultado do teste da hipótese nula 2

$U_{\text{observado}}$	$U_{\text{crítico}}$	Hipótese nula
133	89	Não rejeitada

Nível de significância : 0,05

Esse resultado parece levar à conclusão de que os procedimentos de avaliação não são responsáveis pela variância de rendimento do estudante, apesar de seu papel central em um sistema de ensino e do raciocínio desenvolvido na Seção 2, apoiado em especulações teóricas. Algumas interpretações, entretanto, podem ser desenvolvidas com o fim de introduzir matizes nessa conclusão.

Uma delas baseia-se numa análise dos dois procedimentos comparados, tais como foram usados no experimento. Recorde-se que os dois procedimentos, pela definição fornecida e pela sua operacionalização, diferiam apenas na fase "interpretação dos resultados". Nas demais fases da avaliação os dois grupos recebiam o mesmo tratamento. Ou seja, ambos recebiam antes de cada unidade a lista de objetivos e respondiam aos mesmos itens de teste, que aliás, eram elaborados segundo normas da MRC. Pode-se, então, perguntar se não se teria observado uma diferença significativa de rendimento entre os dois grupos se as diferenças entre os dois procedimentos, MRC e MRN, fossem maximizadas. Isto é, se apenas o grupo MRC tivesse recebido a lista de objetivos (afinal, esse é um aspecto enfatizado pela MRC) e se os testes de unidade tivessem sido elaborados segundo os princípios próprios de cada um dos sistemas.

A carência de pesquisas relacionando avaliação e comportamento do estudante, e mais particularmente, os procedimentos de MRC e comportamento do estudante, tornam difícil a interpretação dos resultados ora obtidos. O raciocínio lógico desenvolvido na Seção 2, entretanto, pode servir de pano de fundo para algumas interpretações.

Segundo aquele raciocínio a MRC teria maiores probabilidades de fornecer ao estudante um meio de tomar suas próprias decisões instrucionais e, portanto, controlar sua própria aprendizagem. As especulações teóricas de Tyler (1974), Karmel (1974), Cook (1951), Mehrens & Lehmann (1973), que podem ser sintetizadas na afirmação de Dressel (apud Payne, 1974, p.9), "a avaliação é mais suscetível de conduzir à aprendizagem quando providencia e encoraja a auto-avaliação e é mais significativa para a aprendizagem quando permite e encoraja o exercício da iniciativa individual", pareciam tornar razoável a previsão de que o auto-controle proporcionado ao estudante pela MRC teria efeitos favoráveis sobre seu rendimento e atitudes.

Pode-se, entretanto, indagar se o grupo MRC usou efetivamente o conhecimento dos resultados em termos de objetivos para melhorar seu desempenho. Os estudantes do grupo MRC, mesmo que tivessem analisado seu desempenho em cada unidade, através da Ficha de Informação de Resultados, teriam tido a oportunidade de usar essa análise para melhorá-lo? Analisando-se os procedimentos empregados, pode-se observar que havia apenas um teste de unidade, não tendo os estudantes oportunidade de executarem atividades de recuperação e de se apresentarem para novo teste, no qual o conhecimento dos resultados pudesse ser imediatamente usado. Esse conhecimento, fornecido após o término de cada unidade, teve oportunidade de ser usado apenas no teste final, após o término das 5 unidades do programa instrucional. Nesse ponto reside talvez uma das principais falhas do estudo.

Mesmo que se tivesse oferecido oportunidade aos estudantes para usarem o conhecimento dos resultados, permaneceria a pergunta: Eles usaram efetivamente o conhecimento dos resultados fornecido? Em que medida cada estudante do grupo MRC deteve-se em uma análise da Ficha de Informação dos Resultados? Um controle dessa variável teria sido útil para responder a essas questões.

Uma variável sobre a qual não se pode deixar de especular é o tempo, considerada, aliás, como uma variável chave nos estudos de Likert (.... 1975, p,100) sobre sistemas administrativos. Uma inovação leva algum tempo até

que seja internalizada pelos seus usuários. A novidade que os procedimentos de MRC representou para os sujeitos deste experimento deve ter levado certo tempo para ser compreendida, inicialmente, e um outro período até que o seu uso pudesse ser integrado ao seu comportamento de rotina. Afinal, as normas tradicionais, que deram direção a toda a vida escolar do estudante, não poderiam ser deixadas de lado de um mês para o outro, por um simples ato de vontade. É lícito, portanto, perguntar-se se o tempo de quatro meses, que foi o tempo de duração do experimento, não teria sido insuficiente para que se observasse os efeitos que, segundo o raciocínio em que estavam apoiadas as hipóteses, estava se prevendo.

Um aspecto a ser observado é o de que o fato de não se ter verificado diferença de rendimento entre os dois grupos pode deslocar a análise para outros pontos de comparação entre os procedimentos de MRC e MRN. Se eles são equivalente quanto a essa variável, a decisão de usar ou não a MRC na prática educacional, dependerá, então, de considerações em torno de outras variáveis. Uma dessas comparações poderia ser entre as possibilidades de um e outro procedimento no sentido de fornecer um meio mais eficiente de melhorar o ensino e a aprendizagem posteriores, verificando qual deles proporciona um conhecimento mais preciso e adequado da situação do estudante.

## 5.2. HIPÓTESE 2

As distribuições de frequência dos pontos atribuídos pelos sujeitos dos grupos MRC e MRN à Escala de Atitudes é apresentada na Tabela 5. Uma comparação entre essas distribuições revela que elas são bastante semelhantes. As médias, 77,4 do grupo MRC e 79,7 do grupo MRN, estão muito próximas e da mesma forma os desvios-padrão, de 10,7 para o grupo MRC e de 9,3 para o grupo MRN, com os pontos desse último apresentando uma variabilidade levemente maior. Deve-se notar que ambas as distribuições apresentam a amplitude total de 40.

TABELA 5

Distribuição de frequência dos pontos atribuídos à Escala de Atitude pelos grupos MRC e MRN

Nº de pontos	Frequência	
	MRC	MRN
51-55	1	1
56-60	1	-
61-65	-	-
66-70	1	1
71-75	2	2
76-80	5	4
81-85	2	4
86-90	2	4
91-95	2	1
N	16	17
Média	77,4	79,7
Desvio-padrão	10,7	9,3

O teste U, de Mann-Whitney, aplicado às distribuições, conforme mostra a Tabela 6, confirma as observações feitas. A hipótese nula 2, testada unidirecionalmente ao nível de significância de 0,05, não pôde ser rejeitada.

TABELA 6

Resultado do teste da hipótese nula 2

$U_{\text{observado}}$	$U_{\text{crítico}}$	Hipótese nula
126	89	Não rejeitada

Nível de significância : 0,05

Desse modo, não pôde ser aceita a hipótese alternativa. Isto é, não se verificou diferença significativa de atitudes para com a disciplina, entre os grupos MRC e MRN, devendo-se notar, entretanto, que ambos os grupos apresentaram um grau alto de atitudes favoráveis, já que o número máximo de pontos era 95 e as médias aproximam-se de 80.

A hipótese de que um sistema de avaliação baseado na MRC desenvolveria no estudante uma atitude mais favorável que um sistema de avaliação baseado na MRN apoiava-se em um raciocínio dedutivo. A MRC fornece um conhecimento detalhado dos resultados, não vincula a interpretação do desempenho do estudante ao de outros e propicia ao estudante um poderoso meio de autocontrole e auto-direção da aprendizagem. Esses fatores, previa-se, constituiriam condições que poderiam permitir uma aproximação favorável à disciplina de estudo, para usar a nomenclatura de Mager (1971, p.52).

Pode-se tentar o fornecimento de explicações alternativas para o resultado de não-diferença encontrado. Uma delas consiste em se especular sobre os efeitos do sistema de notas da Instituição, que não foi abandonado ao longo do experimento. Esse sistema adota uma escala de notas que variam de 0 a 10, atribuídas em função do rendimento absoluto do estudante. Como tal não pode ser caracterizado como um procedimento de MRN, mas tampouco como de MRC. Os sujeitos estavam habituados (dois anos, pelo menos) com esse sistema. Portanto, até o momento do experimento ele dirigia sua aprendizagem por ele. Como, durante o experimento, esse componente continuou presente, tanto na disciplina em que o experimento foi realizado, como nas outras que se desenvolviam paralelamente, pode-se pensar na hipótese de o estudante não ter levado em consideração o procedimento introduzido pelo experimento, inclusive por que ele não tinha maiores consequências sobre sua vida escolar, ao contrário do procedimento institucional.

Uma explicação similar a uma já tentada em relação à Hipótese 1 pode ser desenvolvida também aqui. Se uma atitude mais favorável do grupo MRC fosse função, entre outras variáveis, de uma maior participação em decisões sobre sua própria aprendizagem, então também seria lógico supor-se que a intensidade dessa atitude deveria ser proporcional ao grau de participação. Entretanto, como já se comentou, os estudantes do grupo MRC receberam um conhecimento mais detalhado dos resultados, mas tiveram pouca oportunidade de usá-lo. Essa pouca oportunidade refere-se ao teste final de rendimento, que abrangia as 5 unidades do programa, já que não havia chance de recuperação nas unidades. Esse teste final, porém, foi administrado em uma ocasião distante das ocasiões em que os estudantes tinham obtido o conhecimento dos resultados das unidades, o que pode ter anulado possíveis efeitos benéficos.

Um outro ponto a ser considerado é o da diferença efetiva entre os procedimentos comparados. Por imposição da metodologia adotada os dois procedimentos diferiam apenas no aspecto "interpretação dos resultados". As outras fases eram coincidentes e ambos os procedimentos obedeceram a um planejamento e elaboração cuidadosos. Isto significa inclusive que os procedimentos de MRN estariam altamente contaminados de preceitos de MRC. É possível, portanto, que a direção das atitudes e sua intensidade estivessem mais influenciadas pelo cuidado e atenção igualmente dedicados a ambos os procedimentos pelo experimentador que pela sua natureza diferente.

## 6. CONCLUSÕES E IMPLICAÇÕES

O problema colocado pelo presente estudo perguntava qual o efeito sobre o rendimento e atitudes do estudante de um programa de ensino baseado nos princípios da Medição com Referência a Critério. À falta de estudos disponíveis que relacionassem avaliação e comportamento do estudante, partiu-se de considerações teóricas para formular as hipóteses de que procedimentos de avaliação baseados na MRC teriam o efeito de causar um maior rendimento e atitudes mais favoráveis do estudante em relação à disciplina que os procedimentos tradicionais de MRN. Dos resultados e discussão podem ser extraídas as seguintes conclusões:

1. Não se pôde observar diferença significativa de rendimento entre estudantes submetidos aos dois tipos de tratamentos : MRC e MRN.

2. Não se observou também diferença de atitudes para com a disciplina entre os estudantes do grupo que recebeu o tratamento de avaliação baseado na MRC e os estudantes do grupo que recebeu o tratamento de avaliação baseado na MRN.

3. Os resultados obtidos devem ser interpretados à luz de duas considerações : a. ambos os procedimentos foram cuidadosamente planejados e elaborados, o que pode ter contribuído para que os dois grupos apresentassem resultados igualmente altos nas duas variáveis dependentes; b. talvez não se tenha aproveitado com propriedade as possibilidades de retro-alimentação oferecidas pela MRC.

Aceitando-se o resultado obtido, apesar das limitações expostas, algumas implicações poderiam ser levantadas :

1. A teoria da MRC não seria afetada por tal resultado, primeiro, porque não houve diferença a favor do outro procedimento, e segundo porque os teóricos dessa corrente partem do princípio de que ela é a única adequada aos modelos instrucionais da chamada "tecnologia educacional"(Glaser,1971).

2. A seleção do procedimento a adotar na prática diária deve basear-se em outros critérios que não os que constituíram as variáveis dependentes deste estudo. Um dos critérios principais, alternativos, poderia ser o da possibilidade de melhorar o ensino, pela realimentação concedida ao professor ou ao elaborador de programas.

3. Uma implicação não diretamente relacionada ao resultado, mas não menos importante, diz respeito à viabilidade de uso de um sistema de MRC na prática. A enorme massa de dados torna o sistema praticamente impossível de ser manipulado sem o auxílio do computador.

## 7. SUGESTÕES PARA OUTROS ESTUDOS

A análise das limitações deste estudo e a experiência obtida apontam as seguintes sugestões para pesquisas futuras :

1. Dever-se-ia considerar a possibilidade de usar amostras de tamanho maior que a que se usou. Naturalmente, nesse caso, os dados gerados somente serão manipuláveis por meio de computador. Ainda quanto à amostra, dever-se-ia tentar superar a limitação da falta de representatividade, usando estudantes de outros níveis e/ou cursos.
2. Outra variável que deverá merecer consideração de estudos futuros é o tempo. Os efeitos de um sistema não usual poderão levar mais tempo que o usado no experimento para se manifestar.
3. Futuros estudos deveriam procurar uma operacionalização da MRC em que o estudante tivesse uma oportunidade real de usar os resultados obtidos e comunicados em termos de objetivos para corrigir possíveis pontos fracos.
4. Estudos de desenvolvimento poderiam pesquisar esquemas de amostragem de itens, estudantes e ocasiões na linha de Gorth et alii (1975) e de Hively(1974), verificando sua viabilidade.
5. Um tipo de estudo, de formulação e desenvolvimento, que não pode ser relegado a segundo plano, é o da busca de formas de comunicação de resultados aos alunos e de sistemas de sumariar dados obtidos da MRC que sejam úteis ao professor e ao elaborador de programas instrucionais. Somente dessa forma a avaliação assumirá a sua função principal de melhorar o ensino e a aprendizagem.
5. Problemas sérios a serem superados por pesquisas metodológicas referem-se à análise de itens, validade e fidedignidade de testes de MRC, já que os métodos tradicionais de encarar esses aspectos estão baseados na variabilidade dos escores.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

1. ALMEIDA, A.B. *Efeitos de um modelo de avaliação formativa no processo de aprendizagem*. Porto Alegre, Dissertação de Mestrado apresentada aos Cursos de Pós-Graduação em Educação da Universidade Federal do Rio Grande do Sul, 1976.
2. ARAÚJO E OLIVEIRA, J.B. Avaliação de sistemas e tomada de decisões. *Revista Brasileira de Estudos Pedagógicos*, 60(135): 318-29, jul/set 1974.
3. BASTOS, L. DA R. & SWYTER, L. Avaliação com referência a norma e a critério. *Revista Brasileira de Estudos Pedagógicos*, 60(135): 305-9, jul/set 1974.
4. BLOCK, J.H. Ed. *Mastery learning. Theory and practice*. New York, Holt, Rinehart and Winston, 1971.
5. BLOCK, J.H. & ANDERSON, L.W. *Mastery learning in classroom instruction*. New York, Macmillan, 1975.
6. BLOOM, B.S. et alii. *Handbook on formative and summative evaluation of student learning*. New York, McGraw-Hill, 1971.
7. CAMPBELL, D.T. & STANLEY, J.C. *Diseños experimentales e cuasiexperimentales en la investigación social*. Buenos Aires, Amorrortu, 1973.
8. CHADWICK, C. *Tecnologia educacional para el docente*. Buenos Aires, Paidós, 1975.
9. CHASSOT, A.I. *Comparação de dois instrumentos de avaliação: questões de respostas livres vs. questões objetivas*. Porto Alegre, Dissertação de Mestrado apresentada aos Cursos de Pós-Graduação em Educação da Universidade Federal do Rio Grande do Sul, 1976.
10. CONOVER, W.J. *Practical nonparametric statistics*. New York, John Wiley, ... 1971.
11. COOK, W.W. The functions of measurement in the facilitation of learning. In: LINDQUIST, E.F. Ed. *Educational measurement*. Washington, D.C., ACE, .... 1951, p.3-45.
12. EBEL, R. L. Some limitations of criterion-referenced measurement. In : BRACHT, G.H. et alii. Eds. *Perspectives in educational and psychological measurement*. Englewood Cliffs, Prentice-Hall, 1972a, p.144-9.
13. \_\_\_\_\_. *Essentials of educational measurement*. Englewood Cliffs, Prentice-Hall, 1972b.
14. GLASER, R. Instructional technology and the measurement of learning outcomes: some questions. In: POPHAM, W.J. Ed. *Criterion-referenced measurement. An introduction*. Englewood Cliffs, Educational Technology Publications, 1971, p.5-14.
15. GLASER, R. & NITKO, A.J. Measurement in learning and instruction. In: THORNDIKE, R.L. Ed. *Educational measurement*. Washington, D.C., ACE, 1971.
16. GORTH, W.P. et alii. *Comprehensive achievement monitoring. A criterion-referenced evaluation system*. Englewood Cliffs, Educational Technology Publications, 1975.
17. GRONLUND, N.E. *Preparing criterion-referenced tests for classroom instruction*. New York, Macmillan, 1973.
18. \_\_\_\_\_. *A formulação de objetivos comportamentais para as aulas*. Rio, Rio, 1975.

19. GRONLUND, N.E. *Measurement and evaluation in teaching*. New York, Macmillan, 1976.
20. HAMBLETON, R.K. & NOVIK, M.R. Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 10 (3): 159-70, Fall 1973.
21. HIVELEY, W. Introduction to domain-referenced testing. In: HIVELEY, W. Ed. *Domain-referenced testing*. New Jersey, Educational Technology Publications, 1974, p.5-15.
22. KARMELO, L.J. *Medición y evaluación escolar*. Mexico, Trillas, 1974.
23. KOFF, E.D. Mensuração de atitudes para com a disciplina de Biologia. Trabalho apresentado na disciplina "Testes e Medidas" dos Cursos de Pós-Graduação da UFRGS, Porto Alegre, 1975. (Inédito).
24. KUETHE, J.L. *O processo ensino-aprendizagem*. Porto Alegre, Globo, 1974.
25. LIKERT, R. *A organização humana*. São Paulo, Atlas, 1975.
26. LINDVALL, C.M. & NITKO, A.J. *Measuring pupil achievement and aptitude*. New York, Harcourt Brace Jovanovich, 1975.
27. MAGER, R. *Actitudes positivas en la enseñanza*. Mexico, Pax-Mexico, 1971.
28. \_\_\_\_\_. *Objetivos para o ensino efetivo*. Rio, SENAI, 1972.
29. \_\_\_\_\_. *Medición del intento educativo*. Buenos Aires, Guadalupe, 1975.
30. MEHRENS, W.A. & LEHMANN, I.J. *Measurement and evaluation in education and psychology*. New York, Holt, Rinehart and Winston, 1973.
31. MILLMAN, J. Sampling plans for domain-referenced tests. In: HIVELEY, W. Ed. *Domain-referenced testing*. Englewood Cliffs, Educational Technology Publications, 1974, p.31-42.
32. MUELLER, D.J. The mastery model and some alternative models of classroom instruction and evaluation: an analysis. *Educational Technology*, 13(5): 5-10, May 1973.
33. NAGEL, T.S. & RICHMAN, P.T. *Ensino para competência*. Porto Alegre, Globo, 1974.
34. PAYNE, D.A. *The assessment of learning. Cognitive and affective*. Lexington, D.C. Heath, 1974.
35. POPHAM, W.J. Ed. *Criterion-referenced measurement. An introduction*. Englewood Cliffs, Educational Technology Publications, 1971.
36. \_\_\_\_\_. An approaching peril: cloud-referenced tests. *Phi Delta Kappan*, 55(9): 614-5, May 1974.
37. POPHAM, W.J. & BAKER, E.L. *Los objetivos de la enseñanza*. Buenos Aires, Paidós, 1972.
38. POPHAM, W.J. & HUSEK, T.R. Implications of criterion referenced measurement. In: POPHAM, W.J. Ed. *Criterion-referenced measurement. An introduction*. Englewood Cliffs, Educational Technology Publications, 1971, p.17-37.
39. REMMERS, H.H. et alii. *A practical introduction to measurement and evaluation*. New York, Harper and Row, 1966.

40. STUFFLEBEAN, D.L. et alii. *Educational evaluation and decision making*. Itasca, F.E. Peacock, 1971.
41. TYLER, R.W. *Princípios básicos de currículo e ensino*. Porto Alegre, Globo, 1974.
42. THORNDIKE, R.L. & HAGEN, E. *Tests y técnicas de medición en psicología y educación*. Mexico, Trillas, 1970.
43. VIANNA, H.M. *Testes em educação*. São Paulo, Ibrasa, 1973.

## 9. ANEXOS

**9.1. ESCALA DE MENSURAÇÃO DE ATITUDES PARA COM A DISCIPLINA**

PSEUDÔNIMO : \_\_\_\_\_ GRUPO : \_\_\_\_\_

ESCALA DE MENSURAÇÃO DE ATITUDES PARA COM A DISCIPLINA

INSTRUÇÕES : As sentenças abaixo referem-se à disciplina "Medidas Educacionais I". Decida qual o seu grau de concordância ou discordância com cada uma delas, assinalando com um X no espaço correspondente.

1. Esta disciplina é a mais indesejável do semestre.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
2. Não tenho nenhum gosto ou desgosto definido por esta disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
3. Eu adoro estudar esta disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
4. Esta disciplina é de grande valor.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
5. Eu detesto esta disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
6. Todo estudante beneficia-se em cursar esta disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
7. Esta disciplina é muito prática.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
8. É um castigo para qualquer um cursar esta disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
9. Eu não aconselharia ninguém a cursar esta disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
10. Esta disciplina é uma perda de tempo.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
11. Vale a pena gastar o meu tempo estudando esta disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
12. Esta disciplina é uma boa disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
13. Esta disciplina é proveitosa para todos os que a cursam.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
14. Eu não estou interessado nesta disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )
15. Não vejo razão para interessar-me por esta disciplina.  
Concordo plenamente( )Concordo( )Indeciso( )Discordo( )Discordo plenamente( )

16. Eu realmente gosto desta disciplina.

Concordo plenamente ( ) Concordo ( ) Indeciso ( ) Discordo ( ) Discordo plenamente ( )

17. Eu passaria muito bem sem esta disciplina.

Concordo plenamente ( ) Concordo ( ) Indeciso ( ) Discordo ( ) Discordo plenamente ( )

18. Não vejo nenhum valor nesta disciplina.

Concordo plenamente ( ) Concordo ( ) Indeciso ( ) Discordo ( ) Discordo plenamente ( )

19. Não tenho nenhuma vontade de estudar esta disciplina.

Concordo plenamente ( ) Concordo ( ) Indeciso ( ) Discordo ( ) Discordo plenamente ( )

20. Eu odeio esta disciplina.

Concordo plenamente ( ) Concordo ( ) Indeciso ( ) Discordo ( ) Discordo plenamente ( )

## 9.2. TESTE DE RENDIMENTO

TESTE DE RENDIMENTO DE "MEDIDAS EDUCACIONAIS I"

INSTRUÇÕES : Todos os itens deste teste são do tipo "escolha múltipla". Para cada um deles você deve assinalar na grade de respostas a alternativa que você selecionar como correta, cobrindo o círculo correspondente. Não serão computados os itens rasurados.

1. O critério principal para a seleção de técnicas de avaliação de resultados educacionais é
  - A. a precisão da medição dos comportamentos exibidos pelos estudantes.
  - B. a objetividade da medição dos comportamentos exibidos pelos estudantes.
  - C. sua adequação ao tipo de produto de aprendizagem medido.
  - D. sua adequação ao tipo de técnicas e métodos de ensino usados.
2. Um dos princípios gerais de avaliação diz que "a avaliação requer uma ampla gama de técnicas de avaliação" principalmente porque
  - A. cada procedimento de ensino requer uma técnica de avaliação adequada.
  - B. cada transformação de comportamento desejada exige uma técnica de avaliação adequada.
  - C. o uso de uma única técnica conduz a hábitos viciados e indesejáveis de estudo.
  - D. o uso de uma única técnica não atende ao princípio consagrado das diferenças individuais.
3. O que deve ser feito em primeiro lugar no processo de avaliação?
  - A. A seleção do tipo de item apropriado.
  - B. A formulação e seleção de objetivos.
  - C. A determinação da validade do instrumento.
  - D. A determinação da fidedignidade do instrumento.

Os itens de 4 a 6 envolvem exemplos de avaliação formativa, somativa e de ambas. Selecione, para cada um dos itens, a alternativa correta, dentre as seguintes :

- A. Avaliação formativa.
  - B. Avaliação somativa.
  - C. Avaliação formativa e somativa.
4. Após a primeira parte da terceira unidade os alunos foram submetidos a um teste abrangendo apenas a matéria estudada naquela parte. Além de identificar dificuldades específicas que merecerão atendimento especial o professor obteve uma nota que juntamente com outras contribuirá para o grau final.
  5. Ao final do curso os alunos responderam a um teste de 50 itens que decidiu sua aprovação ou não.
  6. Ao final da aula o prof. Carlos aplicou um breve teste para verificar em que medida os alunos estavam dominando os conceitos trabalhados naquela aula.

Os itens 7 e 8 envolvem exemplos de avaliação, medição e descrição qualitativa de resultados educacionais. Selecione, para cada um dos itens, a alternativa correta, dentre as seguintes :

- A. Somente medição.
- B. Avaliação.
- C. Somente descrição qualitativa.

7. Dos 60 alunos da classe, 20% receberam notas acima de 8 ( numa escala de zero a dez), 30% obtiveram notas entre 6 e 8, 30% entre 4 e 6 e 20% abaixo de 4.
8. Mais de metade dos alunos errou mais de 40% dos itens do teste. Estes alunos não podem passar à unidade seguinte, devendo-se programar um curso de recuperação especial.

Os itens de 9 a 11 referem-se a perguntas relacionadas a decisões que devem ser tomadas no processo educacional. Responda em qual das fases essenciais deste processo elas devem ser consideradas, conforme a chave seguinte:

- A. Identificação e formulação de objetivos.
  - B. Avaliação prévia.
  - C. Ensino (propriamente dito).
  - D. Avaliação.
9. Que situações que facilitam a aquisição de modificações desejáveis de comportamento devem ser selecionadas e como organizá-las?
10. Como saber se modificações desejáveis de comportamento foram alcançadas?
11. Quais modificações de comportamento são desejáveis?
12. Qual dos elementos seguintes é component e indispensável da avaliação educacional?
- A. Medição.
  - B. Descrição qualitativa dos alunos.
  - C. Juízo de valor.
  - D. Descrição quantitativa dos alunos.
13. A avaliação educacional pode ser definida como um processo sistemático para determinar
- A. quais mudanças de comportamento são desejáveis e devem ser atingidas.
  - B. em que medida os objetivos fixados devem ser atingidos.
  - C. em que medida os objetivos fixados podem ser atingidos.
  - D. em que medida as mudanças de comportamento desejadas foram atingidas.
14. O processo de avaliação está intimamente relacionado ao processo ensino-aprendizagem porque
- A. os procedimentos de ensino-aprendizagem visam produzir modificações desejáveis de comportamento no estudante e as técnicas de avaliação determinar o que deve ser modificado.
  - B. tanto o processo ensino-aprendizagem quanto o processo de avaliação devem estar baseados nos antecedentes educacionais do aluno.
  - C. tanto o processo ensino-aprendizagem quanto o processo de avaliação contribuem para a formulação de objetivos educacionais mais valiosos.
  - D. os procedimentos de ensino-aprendizagem visam produzir modificações desejáveis de comportamento no estudante e as técnicas de avaliação determinar o grau dessas modificações.

Os itens 15 e 16 envolvem exemplos retirados do contexto educacional. Selecione, para cada um dos itens, a alternativa correta, conforme a seguinte chave:

- A. Produto.
- B. Processo.
- C. Processo e produto.
- D. Nem processo, nem produto.

15. Na metade do curso 90% dos alunos estavam aptos a escreverem um ensaio sobre José de Alencar, considerando o professor satisfeito os propósitos da quela parte.
16. Os alunos receberam instruções para lerem dois capítulos do livro "Medidas Educacionais", de A. Pereira.
17. A categoria COMPREENSÃO do Domínio Cognitivo da Taxonomia de Bloom envolve
- A. Criatividade.
  - B. Memorização.
  - C. Julgamento.
  - D. Interpretação.
18. As categorias da Taxonomia dos Objetivos Educacionais, de Bloom, estão organizadas segundo
- A. a complexidade dos objetivos, dos simples para os complexos, as classes de ordem superior independentes das de ordem inferior.
  - B. o nível de generalidade dos objetivos, dos objetivos gerais para os objetivos específicos.
  - C. o nível de generalidade dos objetivos, dos objetivos específicos para os objetivos gerais.
  - D. complexidade dos objetivos, dos simples para os complexos, as classes de ordem superior incluindo as de ordem inferior.
19. O método de formulação de objetivos preconizado por Gronlund compreende as duas primeiras etapas seguintes : Formular objetivos
- A. gerais e deles derivar os itens de teste.
  - B. gerais e deles derivar objetivos específicos.
  - C. específicos e após os gerais que os englobam.
  - D. específicos e a partir deles os itens de teste.
20. Gronlund justifica seu método de formulação de objetivos em duas etapas básicas através da seguinte proposição:
- A. Os objetivos gerais permitem a elaboração de itens mais diretamente relacionados ao processo ensino-aprendizagem, constituindo os específicos apenas uma amostra através da qual o estudante pode guiar seus esforços.
  - B. Os objetivos específicos permitem ao professor dirigir a instrução diretamente à sua consecução, constituindo os objetivos gerais apenas uma forma de formular metas educacionais mais valiosas.
  - C. Os objetivos gerais definem um universo de comportamentos, do qual os objetivos específicos constituem apenas uma amostra possível, devendo a instrução estar dirigida para a consecução dos objetivos gerais.
  - D. Os objetivos específicos põem melhor em evidência a relação entre os processos de ensino e o de avaliação, constituindo os objetivos gerais apenas um guia para a elaboração de materiais de instrução e de instrumentos de avaliação.

Os itens de 21 a 23 envolvem exemplos de objetivos educacionais. Decida a qual dos domínios seguintes eles pertencem:

- A. Cognitivo.
- B. Afetivo.
- C. Psicomotor.

21. Julgar a consistência dos argumentos de um texto científico.
22. Interpretar gráficos e tabelas estatísticas.
23. Demonstrar interesse científico, através de leituras e execução de experimentos.

Os itens de 24 a 26 envolvem formulações de objetivos educacionais gerais. Para cada um deles indique o tipo de erro cometido, segundo a seguinte chave:

- A. Formulação em termos do comportamento do professor.
- B. Formulação em termos de tópicos.
- C. Formulação em termos do processo de aprendizagem.
- D. O objetivo está bem formulado, não há erro.

24. Adquirir noções básicas de eletricidade.

25. Aplicar princípios e generalizações na solução de problemas.

26. Demonstrar aos alunos o funcionamento de alavancas simples.

Os itens de 27 a 29 envolvem exemplos de objetivos educacionais do domínio cognitivo da Taxonomia de Bloom. Decida a qual das seguintes categorias eles pertencem:

- A. Conhecimento.
- B. Compreensão.
- C. Aplicação.
- D. Análise ou Síntese ou Avaliação.

27. Formular uma teoria original de aprendizagem aplicável aos procedimentos de ensino em sala de aula.

28. Preparar representações gráficas sobre fenômenos físicos a partir de dados observados ou relatados.

29. Evocar e reconhecer fatos biológicos de importância para um entendimento sistemático dos processos biológicos.

Os itens de 30 a 31 envolvem exemplos de questões de teste. Decida a qual das categorias seguintes da Taxonomia de Bloom elas pertencem:

- A. Conhecimento.
- B. Compreensão.
- C. Aplicação.
- D. Análise ou Síntese ou Avaliação.

30. Qual das seguintes frases sobre movimento ondulatório define período?

1. a distância máxima em que uma partícula pode se deslocar de seu ponto de repouso.
2. o tempo necessário para que uma partícula efetue uma oscilação completa.
3. o número de oscilações por segundo.
4. a variação no tempo da mudança de posição, numa direção determinada.

31. Quando uma pessoa pobre justifica sua situação, reclamando que o mundo está em dívida para com ela, está

1. comportando-se como uma pessoa psicótica.
2. apresentando um sintoma paranóico.
3. sofrendo uma alucinação.
4. desenvolvendo uma reação tipicamente infantil.
5. racionalizando.

Os itens de 32 a 34 referem-se a perguntas relacionadas com as definições de validade de testes. Selecione, para cada uma delas, a resposta correta, dentre as seguintes:

- A. Validade ( de modo genérico).
- B. Validade de construto.
- C. Validade de conteúdo.
- D. Validade relacionada a critério.

32. Em que grau os resultados do teste relacionam-se com alguma outra medida de desempenho?
33. Em que grau o teste mede uma amostra representativa dos vários tópicos e comportamentos?
34. Em que grau o teste atinge os propósitos para os quais foi elaborado?
35. Qual dos seguintes enunciados melhor expressa a relação entre validade e fidedignidade?
- A. Um teste não pode ser fidedigno sem ser válido.  
 B. Um teste pode ser válido sem ser fidedigno.  
 C. Um teste tanto pode ser válido sem ser fidedigno quanto ser fidedigno sem ser válido.  
 D. Um teste não pode ser válido sem ser fidedigno.
36. Dada a seguinte tabela de contingência, por quartis, que resume os resultados de dois testes, respondidos por 32 alunos, um dos testes considerado como critério e o outro como preditor, pode-se dizer que

		CRITÉRIO			
		1	2	3	4
P R E D I C T O	1	2	2	2	2
	2	2	2	2	2
	3	2	2	2	2
	4	2	2	2	2

- A. o coeficiente de validade preditiva está mais próximo de +1 que de zero.  
 B. o coeficiente de validade preditiva está mais próximo de -1 que de zero.  
 C. o coeficiente de validade preditiva está mais próximo de zero que de +1 e -1.  
 D. nada pode se afirmar quanto ao coeficiente de validade preditiva baseado apenas nesta tabela.
37. Uma suposição implícita no conceito de validade de conteúdo é a existência de
- A. um universo de alunos com possibilidade de responder ao teste, constituindo o grupo particular apenas uma das amostras possíveis.  
 B. itens fáceis e itens difíceis, devendo-se selecionar para o teste aqueles itens que possuem dificuldade média.  
 C. itens cujas respostas estão menos sujeitas a flutuações de memória e de fatores emocionais e do ambiente.  
 D. um universo de itens possíveis de serem apresentados como estímulos, constituindo o teste uma parte representativa.
38. Os fatores devidos às condições ambientais, como luz, calor, ventilação, são considerados como variância do erro na obtenção da fidedignidade pelo método
- A. da prova repetida.  
 B. das formas equivalentes.  
 C. das metades.  
 D. de Kuder-Richardson.
39. A fidedignidade de um teste pode ser definida como a consistência
- A. dos escores atribuídos por diferentes avaliadores, de competência equivalente, às respostas aos itens do teste fornecidas por um mesmo indivíduo.  
 B. dos resultados obtidos pelo mesmo indivíduo, quando testado novamente com o mesmo instrumento ou equivalente.

- C. dos resultados obtidos pelo mesmo indivíduo, nas partes objetiva e de dissertação do teste, cuja equivalência procurou-se obter na fase de elaboração.
- D. dos escores atribuídos em ocasiões diferentes, por um único avaliador, sob condições equivalentes de correção, às respostas aos itens do teste de um mesmo indivíduo.
40. A tabela abaixo apresenta os resultados obtidos por 4 grupos de sujeitos, num mesmo teste, qualificado numa escala de 0 a 100. Qual desses conjuntos de resultados fornecerá um maior coeficiente de fidedignidade?
- |    |    |    |     |    |    |    |    |    |    |    |
|----|----|----|-----|----|----|----|----|----|----|----|
| A. | 99 | 94 | 100 | 95 | 96 | 95 | 98 | 99 | 96 | 95 |
| B. | 6  | 12 | 2   | 3  | 6  | 5  | 4  | 9  | 8  | 5  |
| C. | 7  | 99 | 42  | 87 | 66 | 53 | 33 | 22 | 11 | 75 |
| D. | 50 | 48 | 42  | 47 | 41 | 46 | 40 | 43 | 47 | 49 |
41. Qual dos seguintes testes exige um coeficiente de fidedignidade maior?
- A. De diagnóstico.
- B. Formativo.
- C. Somativo.
- D. De pré-requisitos.
42. O coeficiente de correlação entre os resultados das metades par e ímpar de um teste aplicado a um determinado grupo é igual a 0,25. O coeficiente de fidedignidade do teste inteiro é igual a
- A. 0,25.
- B. 0,30.
- C. 0,40.
- D. 0,50.
43. Qual dos seguintes métodos para obtenção do coeficiente de fidedignidade considera maior número de fatores como variância do erro?
- A. Da prova repetida com formas equivalentes.
- B. Da prova repetida.
- C. Das formas equivalentes.
- D. Da medida de consistência interna.
44. Qual dos seguintes tipos de item tem maior probabilidade de proporcionar um teste com maior validade de conteúdo, quando usado com exclusividade?
- A. Resposta curta.
- B. Dissertação.
- C. Escolha múltipla.
- D. Resposta alternativa.
45. Uma vantagem dos itens objetivos sobre as questões de dissertação é que eles
- A. fornecem a mensuração de resultados mais complexos de aprendizagem.
- B. colocam maior ênfase na evocação de informação de fatos.
- C. exigem menos tempo para preparação e correção do teste.
- D. fornecem uma amostragem mais extensiva do conteúdo e comportamentos.
46. Qual é a vantagem principal de se usar uma tabela de especificações quando se prepara um teste de aproveitamento escolar?
- A. Melhor caracterização do universo de itens.
- B. Possibilidade maior de selecionar uma amostra de cada tipo de item.
- C. Aumento da objetividade.
- D. Redução do tempo de elaboração do teste.

Os itens de 47 a 50 referem-se a defeitos de itens de escolha múltipla. Decida qual dos seguintes princípios de elaboração de itens cada um destes defeitos está violando.

- A. Os itens de teste usados devem ser determinados mediante os produtos de aprendizagem formulados antes do curso.
- B. Os itens de teste devem estar baseados numa amostra representativa do conteúdo do curso e dos produtos específicos de aprendizagem que se pretende medir.
- C. Os itens de teste devem ser construídos de tal maneira que os fatores extrínsecos não impeçam a resposta do aluno.
- D. Os itens de teste devem ser elaborados de tal maneira que o aluno obtenha a resposta correta unicamente quando haja alcançado o produto de aprendizagem que se está pretendendo medir.

- 47. Dificuldade de leitura acima do nível de capacidade de leitura do grupo.
- 48. Enfeites no suporte que não contribuem para a formulação do problema.
- 49. Coerência gramatical entre o suporte e a alternativa correta, unicamente.
- 50. Palavras comuns no suporte do item e na alternativa correta, unicamente.

9.3. PLANO DE CURSO DA DISCIPLINA  
EM QUE SE REALIZOU O EXPERI-  
MENTO

PLANO DE CURSO

DISCIPLINA : MEDIDAS EDUCACIONAIS I  
TURMA : 4º SEMESTRE DE PEDAGOGIA

SEMESTRE : 2º de 1976  
CRÉDITOS : 4

1. OBJETIVOS GERAIS

Conhecer, compreender e aplicar os termos, conceitos, princípios e procedimentos básicos de medição educacional e de elaboração de testes de aproveitamento escolar.

2. CONTEÚDO

2.1. O PAPEL DA AVALIAÇÃO NO ENSINO

- 2.1.1. Decisões de instrução e avaliação
- 2.1.2. Significado de avaliação e medição
- 2.1.3. A avaliação no programa escolar
- 2.1.4. Tipos de procedimentos de avaliação
- 2.1.5. Princípios gerais de avaliação

2.2. OBJETIVOS EDUCACIONAIS E AVALIAÇÃO

- 2.2.1. Objetivos educacionais como produtos de aprendizagem
- 2.2.2. Dimensões dos objetivos educacionais
- 2.2.3. Determinação dos objetivos educacionais
- 2.2.4. Classificação dos objetivos educacionais
- 2.2.5. Plano geral de avaliação
- 2.2.6. Planejamento dos testes
- 2.2.7. Relacionamento de itens de teste com objetivos educacionais

2.3. VALIDADE DOS RESULTADOS DE TESTES

- 2.3.1. Natureza da validade
- 2.3.2. Tipos de validade
- 2.3.3. Procedimentos de obtenção da validade
- 2.3.4. Fatores que influenciam a validade

2.4. FIDEDIGNIDADE DOS RESULTADOS DE TESTES

- 2.4.1. Definição
- 2.4.2. Erros de medição : fontes
- 2.4.3. Procedimentos de estimação da fidedignidade
- 2.4.4. Fatores que influenciam a fidedignidade

2.5. PRINCÍPIOS E PROCEDIMENTOS GERAIS DE ELABORAÇÃO DE TESTES

- 2.5.1. Tabelas de especificações
- 2.5.2. Tipos de testes : objetivo (resposta curta, resposta alternativa, associação, escolha múltipla), dissertação
- 2.5.3. Vantagens e limitações dos testes objetivos e de dissertação
- 2.5.4. Princípios gerais de elaboração de testes

3. PROCEDIMENTOS

As aulas constituir-se-ão basicamente de exposições do professor, complementadas por intensas leituras, realizadas em período extra-escolar.

4. AVALIAÇÃO

Ao final de cada unidade haverá um teste. Ao final do curso será aplicado um teste de rendimento global das 5 unidades.

5. CRONOGRAMA

DATA	EVENTO
02/08	PRÉ-TESTE DESENVOLV.UNIDADE 1
05/08	DESENVOLV.UNIDADE 1
09/08	DESENVOLV.UNIDADE 1
12/08	TESTE UNIDADE 1 DESENVOLV.UNIDADE 2
16/08	DESENVOLV.UNIDADE 2
19/08	DESENVOLV.UNIDADE 2
23/08	DESENVOLV.UNIDADE 2
26/08	DESENVOLV.UNIDADE 2
06/09	TESTE UNIDADE 2 DESENVOLV.UNIDADE 3
09/09	DESENVOLV.UNIDADE 3
13/09	DESENVOLV.UNIDADE 3

DATA	EVENTO
16/09	DESENVOLV.UNIDADE 3
20/09	TESTE UNIDADE 3 DESENVOLV.UNIDADE 4
23/09	DESENVOLV.UNIDADE 4
27/09	DESENVOLV.UNIDADE 4
30/09	DESENVOLV.UNIDADE 4
04/10	DESENVOLV.UNIDADE 4
07/10	TESTE UNIDADE 4 DESENVOLV.UNIDADE 5
11/10	DESENVOLV.UNIDADE 5
14/10	TESTE UNIDADE 5 REVISÃO GERAL
18/10	TESTE FINAL

## 6. BIBLIOGRAFIA

- 6.1. BLOOM, B.S. et alii. Taxionomia dos objetivos educacionais. Domínio Cognitivo. Porto Alegre, Globo, 1972.
- 6.2. BLOOM, B.S. et alii. Evaluación del aprendizaje. Troquel, Buenos Aires, 1975. Volume 1.
- 6.3. BRADFIELD, J.M. & MOREDOCK, H.S. Medidas e testes em educação. Rio, Fundo de Cultura, 1964.
- 6.4. GRONLUND, N.E. Medición y evaluación en la enseñanza. Mexico, Pax-Mexico, 1973.
- 6.5. GRONLUND, N.E. A elaboração de testes de aproveitamento escolar. São Paulo, EPU, 1974.
- 6.6. GRONLUND, N.E. A formulação de objetivos comportamentais para as aulas. Rio, Editora Rio, 1975.
- 6.7. LINDEMAN, R.H. Medidas Educacionais. Porto Alegre, Globo, 1972.
- 6.8. MAGER, R.F. Objetivos para o ensino efetivo. Rio, SENAI, 1971.
- 6.9. NOLL, V.H. Introdução às medidas educacionais. São Paulo, Pioneira, 1965.
- 6.10. VARGAS, J.S. Formular objetivos comportamentais úteis. São Paulo, EPU, 1974.
- 6.11. VIANNA, H.M. Testes em educação. São Paulo, IBRASA, 1973.

9.4. LISTA DE OBJETIVOS DE UMA DAS  
CINCO UNIDADES DA DISCIPLI -  
NA EM QUE SE REALIZOU O EXPE-  
RIMENTO

OBJETIVOS DA UNIDADE 4 : FIDEDIGNIDADE DE RESULTADOS DE TESTES

1. CONHECER O CONCEITO DE FIDEDIGNIDADE
  - 1.1. Fornecer a definição teórica de fidedignidade
  - 1.2. Fornecer a definição operacional de fidedignidade
  - 1.3. Identificar a definição teórica de fidedignidade
  - 1.4. Identificar a definição operacional de fidedignidade
2. COMPREENDER O CONCEITO DE FIDEDIGNIDADE
  - 2.1. Fornecer exemplos de fatores de erro
  - 2.2. Explicar o conceito de fidedignidade
  - 2.3. Explicar quais fatores são considerados variância verdadeira e quais variância do erro nos vários métodos de obtenção da fidedignidade
  - 2.4. Explicar os procedimentos dos vários métodos de estimação da fidedignidade
  - 2.5. Associar a variância verdadeira e a variância do erro com fatores que contribuem para cada uma delas
3. APLICAR OS VÁRIOS MÉTODOS DE OBTENÇÃO DA FIDEDIGNIDADE MEDIANTE CORRELAÇÃO
  - 3.1. Estimar a fidedignidade pelo método da prova repetida
  - 3.2. Estimar a fidedignidade pelo método das formas equivalentes
  - 3.3. Estimar a fidedignidade pelo método das provas repetidas com formas equivalentes
  - 3.4. Estimar a fidedignidade pelo método das metades
  - 3.5. Estimar a fidedignidade pelo método de Kuder-Richardson
4. CONHECER OS FATORES QUE INFLUEM NA FIDEDIGNIDADE
  - 4.1. Reconhecer os diversos fatores que exercem influência sobre a fidedignidade
  - 4.2. Dizer quais fatores exercem influência sobre a fidedignidade
  - 4.3. Identificar a relação entre validade e fidedignidade
  - 4.4. Dizer qual a relação entre validade e fidedignidade
5. COMPREENDER OS FATORES QUE INFLUEM NA FIDEDIGNIDADE
  - 5.1. Explicar a natureza da influência dos diversos fatores que exercem influência sobre a fidedignidade
  - 5.2. Explicar de que maneira os diversos fatores influem sobre a fidedignidade
  - 5.3. Dadas várias distribuições de notas, identificar a que conduz à maior fidedignidade
6. CONHECER OS CRITÉRIOS DE JULGAMENTO DE UM COEFICIENTE DE FIDEDIGNIDADE
  - 6.1. Dizer de que elementos depende o julgamento de um coeficiente de fidedignidade
  - 6.2. Identificar os elementos de que depende o julgamento de um coeficiente de fidedignidade
7. COMPREENDER OS CRITÉRIOS DE JULGAMENTO DE UM COEFICIENTE DE FIDEDIGNIDADE
  - 7.1. Explicar como decidir se um coeficiente de fidedignidade é aceitável ou não
  - 7.2. Dado um coeficiente de fidedignidade de um teste selecionar a finalidade para a qual ele poderia ser aceitável
  - 7.3. Associar exemplos de testes elaborados para determinadas finalidades com aceitabilidade; e coeficientes mais baixos ou mais altos

9.5. TESTE DE UNIDADE DE UMA DAS  
CINCO UNIDADES DA DISCIPLI-  
NA EM QUE SE REALIZOU O EX-  
PERIMENTO

TESTE DA UNIDADE 4

Os itens de 1 a 4 envolvem exemplos de fatores que podem ser considerados como fontes de variação em resultados de testes. Selecione a alternativa correta dentre as seguintes, conforme o fator possa ser considerado como variância.

- A. do erro no método do teste repetido e não no método das formas equivalentes.
- B. do erro no método das formas equivalentes e não no método do teste repetido.
- C. do erro nos métodos do teste repetido e das formas equivalentes.
- D. verdadeira nos métodos do teste repetido e das formas equivalentes.

1. Habilidade de compreensão verbal geral do indivíduo.
2. Estado de saúde do indivíduo.
3. Flutuações de memória do indivíduo.
4. O elemento chance determinando a inclusão no teste de um item determinado e não de outro.
5. Qual das descrições seguintes pode melhor ser associada ao conceito de fidedignidade?
  - A. Consistência dos escores dados por vários avaliadores diferentes ao mesmo teste de um indivíduo.
  - B. Consistência dos escores obtidos por um mesmo grupo de indivíduos em vários testes diferentes de uma mesma área.
  - C. Extensão em que as posições de um grupo de indivíduos permanecem estáveis em mais de uma aplicação do mesmo teste.
  - D. Extensão em que as posições de um mesmo grupo de indivíduos permanecem estáveis ao longo de uma série de testes diferentes.

Os itens de 6 a 8 referem-se aos diversos aspectos que são colocados em evidência pelos vários métodos de estimação da fidedignidade. Decida qual dos aspectos descritos nestes itens é colocado em evidência por cada um dos métodos:

- A. Teste repetido.
- B. Formas equivalentes.
- C. Metades.

6. Informa em que extensão os resultados de duas amostras de tarefas correspondem a um mesmo universo.
7. Informa em que extensão as posições dos indivíduos permanecem as mesmas em aplicações do teste feitas em ocasiões diferentes e separadas por um tempo apreciável.
8. Informa em que extensão partes do teste medem uma mesma habilidade.
9. Os escores obtidos por um grupo de 11 alunos em 4 testes diferentes estão descritos abaixo. Qual dos testes apresentará maior coeficiente de fidedignidade?
  - A. 21 30 29 22 25 24 28 26 20 23 27
  - B. 28 26 20 8 30 16 14 6 12 24 4
  - C. 20 12 12 19 18 11 14 13 17 15 16

D. 2 4 10 9 8 1 3 5 6 7 11

10. Qual a relação existente entre validade e fidedignidade?

- A. Um teste pode ser válido sem ser fidedigno.
- B. Um teste pode ser fidedigno sem ser válido.
- C. Não pode existir validade sem fidedignidade, nem fidedignidade sem validade.

Para os itens 11 e 12 use o seguinte código:

- A. A asserção e a razão são verdadeiras e a razão é uma justificativa válida para a asserção.
- B. A asserção e a razão são verdadeiras mas a razão não é uma justificativa válida para a asserção.
- C. A asserção é verdadeira e a razão é falsa.
- D. A asserção é falsa e a razão é verdadeira.

11. A pouca variabilidade dos escores contribue para diminuir a fidedignidade porque o teste que conduz à maior variabilidade é o de dificuldade média.

12. Manter padrões uniformes (objetivos) de correção em testes contribue para aumentar a fidedignidade porque escores baseados em critérios uniformes (objetivos) de correção refletem com mais precisão as diferenças de capacidades dos alunos e minimizam as diferenças de julgamento dos avaliadores.

13. São fornecidas abaixo as colocações de 10 indivíduos em duas aplicações do mesmo teste, feitas em ocasiões distanciadas pelo período de 15 dias. O coeficiente de fidedignidade do teste é igual a \_\_\_\_\_ (duas casas decimais depois da vírgula).

ALUNO	A	B	C	D	E	F	G	H	I	J
Ocasião 1	20	40	10	100	80	60	50	90	30	70
Ocasião 2	30	50	20	90	70	80	40	100	10	60

14. Suponha que a tabela do item anterior represente a colocação de um grupo de 10 alunos nas metades par e ímpar de um mesmo teste de 20 itens (ocasião 1 = metade ímpar; ocasião 2 = metade par). Então o coeficiente de fidedignidade do teste inteiro é igual a \_\_\_\_\_ (com duas casas decimais depois da vírgula).

15. Diga com suas próprias palavras quais elementos devem ser levados em conta ao se julgar a aceitabilidade do tamanho do coeficiente de fidedignidade de um determinado teste.

9.6. FICHA DE INFORMAÇÃO DOS RE -  
SULTADOS DO GRUPO, MEDIÇÃO COM  
REFERÊNCIA A CRITÉRIO

FUNDAÇÃO EDUCACIONAL DE CRICIÚMA  
 FACULDADE DE CIÊNCIAS E EDUCAÇÃO  
 DEPARTAMENTO DE CIÊNCIAS PEDAGÓGICAS  
 MEDIDAS EDUCACIONAIS I

FICHA DE INFORMAÇÃO DOS RESULTADOS

NOME : \_\_\_\_\_

UNIDADE 4

1. Nº DE ITENS CORRETOS : \_\_\_\_\_

2. NOTA : \_\_\_\_\_

3. CONSECUÇÃO DOS OBJETIVOS :

OBJ. GERAL	OBJ. ESPECIFICO	ITEM(NS)	ALCANÇADO	NÃO ALC.
1	1.3	5		
2	2.5	1,2,3,4		
	2.6	6,7,8		
3	3.1	13		
	3.2	14		
4	4.3	10		
5	5.3	9		
	5.4	11,12		
7	7.1	15		

9.7. FICHA DE INFORMAÇÃO DOS RE -  
SULTADOS DO GRUPO MEDIÇÃO COM  
REFERÊNCIA A NORMA

FUNDAÇÃO EDUCACIONAL DE CRICIÚMA  
FACULDADE DE CIÊNCIAS E EDUCAÇÃO  
DEPARTAMENTO DE CIÊNCIAS PEDAGÓGICAS  
MEDIDAS EDUCACIONAIS I

FICHA DE INFORMAÇÃO DOS RESULTADOS

NOME : \_\_\_\_\_

UNIDADE 4

1. Nº DE ITENS CORRETOS : \_\_\_\_\_

2. NOTA : \_\_\_\_\_

3. PERCENTIL :

∧	
∨	

∧ : PORCENTAGEM DE ALUNOS COM DESEMPENHO INFERIOR AO SEU  
∨ : PORCENTAGEM DE ALUNOS COM DESEMPENHO SUPERIOR AO SEU