

UNIVERSIDADE FEDERAL DO CEARÁ
NICOLINO TROMPIERI FILHO

**ANÁLISE DOS RESULTADOS DA AVALIAÇÃO DO SAEB/2003,
VIA REGRESSÃO LINEAR MÚLTIPLA**

Fortaleza-Ceará
2007

UNIVERSIDADE FEDERAL DO CEARÁ
NICOLINO TROMPIERI FILHO

**ANÁLISE DOS RESULTADOS DA AVALIAÇÃO DO SAEB/2003,
VIA REGRESSÃO LINEAR MÚLTIPLA**

Tese apresentada no Curso de Doutorado em Educação Brasileira, do Programa de Pós-graduação em Educação Brasileira, da Faculdade de Educação, da Universidade Federal do Ceará-UFC, como requisito para obtenção do título de Doutor em Educação.

Orientador: Brendan Coleman McDonald - Doutor

Fortaleza-Ceará
2007

"Lecturis salutem"

**Ficha Catalográfica elaborada por
Telma Regina Abreu Camboim – Bibliotecária – CRB-3/593
tregina@ufc.br
Biblioteca de Ciências Humanas – UFC**

T766a Trompieri Filho, Nicolino.
Análise dos resultados da avaliação do SAEB/2003, via regressão linear múltipla / por Nicolino Trompieri Filho. – 2007.
--- f. : il.; 31 cm.
Cópia de computador (printout(s)).
Tese(Doutorado) – Universidade Federal do Ceará, Faculdade de Educação, Programa de Pós-Graduação em Educação Brasileira, Fortaleza(CE), 27/08/2007.
Orientação: Prof. Dr. Brendan Coleman McDonald.
Inclui bibliografia.

1-MATEMÁTICA – ESTUDO E ENSINO – CEARÁ. 2-ESTUDANTES DO ENSINO FUNDAMENTAL – AVALIAÇÃO – CEARÁ. 3-RENDIMENTO ESCOLAR – CEARÁ. 4- APRENDIZAGEM. 5-EDUCAÇÃO E ESTADO – CEARÁ. 6-ANÁLISE DE REGRESSÃO. I- McDonald, Brendan Coleman, orientador. II- Universidade Federal do Ceará. Programa de Pós-Graduação em Educação Brasileira. III- Título.

41/07

CDD(21ª ed.) 372.7044098131

**UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE EDUCAÇÃO-FACED
PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO
BRASILEIRA
DOUTORADO EM EDUCAÇÃO**

NICOLINO TROMPIERI FILHO

**ANÁLISE DOS RESULTADOS DA AVALIAÇÃO DO SAEB/2003, VIA
REGRESSÃO LINEAR MÚLTIPLA**

Tese submetida à apreciação da Comissão Examinadora como parte dos requisitos para obtenção do título de Doutor em Educação, outorgado pela Universidade Federal do Ceará-UFC. Encontra-se à disposição dos interessados na Biblioteca do Centro de Humanidades da mencionada instituição.

A citação de qualquer trecho desta tese é permitida, desde que seja feita em conformidade com as normas da ética e da atitude científica.

TESE DEFENDIDA E APROVADA EM: 27/08/2007

Banca Examinadora

Brendan Coleman McDonald

Orientador e Presidente da Banca Examinadora

Maria Isabel Filgueiras Lima Ciasca - Dr.
Membro da Banca Examinadora - UFC

Raimundo Benedito do Nascimento – Dr.
Membro da Banca Examinadora - UFC

Ana Maria Fontenele Catrib - Dra.
Membro da Banca Examinadora - UNIFOR

Maria Gilvanise de Oliveira Pontes - Dra.
Membro da Banca Examinadora - UECe.

AGRADECIMENTOS

- Ao professor Brendan Colemann McDonald,
- Aos professores Maristela Lage Alencar, Meiricele Calíope Leitinho, Rui Verlaine Oliveira Moreira, Raimundo Benedito do Nascimento pelas contribuições valiosas na realização deste trabalho.
- Aos colegas do Departamento de Fundamentos da Educação/FACED/UFC, pela amizade e bons estímulos durante a realização do trabalho final.
- Aos funcionários do Departamento de Fundamentos da Educação/FACED-UFC, pela colaboração desinteressada nos momentos solicitados.
- Ao professor José Anchieta E. Barreto, cuja amizade e ceticismo me levou a concluir este trabalho.
- À professora Maria Isabel Filgueiras Lima Ciasca, pela paciência, incentivo e apoio para a realização deste trabalho.
- Ao professor Hermínio Borges Neto, coordenador do programa de pós-graduação, pelo apoio recebido.
- Aos professores Ana Maria Fontenele Catrib, professora da UNIFOR, e Maria Gilvanise de Oliveira Pontes, professora da UECE, que gentilmente aceitaram compor a Banca Examinadora.
- Ao professor Vianney Mesquita, pela adequação estilística e apropriação gramatical deste relatório de pesquisa.
- Ao funcionário Francisco Pereira Leite, pelo apoio recebido.
- A todos os que, direta ou indiretamente, colaboraram comigo na realização do curso e do presente estudo final.
- Um agradecimento em especial ao professor Manuel Carvalho Alexandrino, que através de troca de idéias e seus conhecimentos em informática, foram fundamental na realização deste trabalho.

***“A humanidade coloca sempre a si mesma
apenas as tarefas que pode resolver.”***

Karl Marx

RESUMO

O presente estudo teve como objetivo identificar nos resultados do SAEB/2003, no Ceará, um modelo multilinear com variáveis que garantam o maior grau de explicação do rendimento em matemática nos testes aplicados nas amostras de 4ª série da escola pública; 4ª série da escola particular; 8ª série da escola pública e 8ª série da escola particular. Tomando-se, entre as variáveis do diretor, do professor, do aluno e das condições físicas da escola, as variáveis que se relacionavam significativamente com o rendimento no teste de matemática em cada uma das amostras, com os recursos do software SPSS (Statistical Package for the Social Sciences), uma regressão linear em cada amostra, com o método “stepware”. Obtiveram-se quatro modelos lineares múltiplos com variáveis que permitem um maior grau de explicação do rendimento no teste em matemática em cada uma das amostras. Concluiu-se que a análise desse tipo permite a formulação de políticas públicas capazes de superar a busca da melhoria do ensino por intervenções pontuais.

Palavras-chave: rendimento em matemática; avaliação; modelo linear múltiplo.

ABSTRACT

This study had as its objective the identification in the results of SAEB/2003, here in Ceará, of a multilinear model with variables that gave a greater degree of explanation for the results in mathematics of tests given in the randomic sample of the fourth years students in public schools, the fourth year students in private schools, the eight year students in public schools and the eight year students in private schools. The following variables were taken into account: the director, the teacher, the student, the physical state of the school as variables that had significant relevance with the results of the tests in mathematics in each one of the samples, using the resources of the software SPSS (Statistical Package for the Social Sciences), a linear regression in each sample, using the method "stepwise". Four multilinear models with variables were obtained that permitted a greater degree of explanation for the results of the mathematic tests in each sample. The study concluded affirming that an analysis of this type permits the formulation of public policies capable of overcoming the search for better forms of teaching with punctual interventions.

Key words: mathematic results; assessment; multiple linear model.

RESUMEN

Este estudio tuvo como objetivo identificar en los resultados del SAEB 2003, en el Ceará, un modelo multilíneo con variables que garantizaran el mayor grado de explicación del rendimiento en matemática en los tests aplicados en las muestras de 4º año de la escuela pública y 8º año de la escuela privada. Tomándose, entre las variables del director, del profesor, del alumno y de las condiciones físicas de la escuela, las variables que se relacionaban significativamente con el rendimiento en el test de matemática en cada una de las muestras, con los recursos del software SPSS (Statistical Package for the School Sciences), una regresión lineal en cada muestra, con el método "stepwise". Se observaron varios modelos lineales múltiples con variables que permiten un grado mayor de explicación del rendimiento en el test de matemáticas en cada una de las muestras. Conclúyese que el análisis de cada tipo permite la formulación de políticas públicas capaces de superar la búsqueda de la mejora de la enseñanza mediante intervenciones puntuales.

Palabras-chave: rendimiento en matemáticas; evaluación; modelo lineal múltiple.

LISTA DE SIGLAS

CCI: CURVA CARACTERÍSTICA DO ITEM

SPAECE: Sistema Permanente de Avaliação do estado do Ceará

FUNDEB: Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização do Magistério

INEP: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

MEC: Ministério da Educação

PCA: Parâmetros Curriculares em Ação

SAEB: Sistema Avaliação da Educação Básica

SEDUC: Secretaria da Educação Básica do estado do Ceará

TCT: Teoria Clássica dos Testes

TRI: Teoria de Resposta ao Item

VIF: valor da inflação da variância

LISTA DE QUADROS

Quadro Nº 1: Categorias científicas e principais características.....	24
Quadro Nº 2 – Classificação das Escalas de Medidas de Stevens.....	38
Quadro Nº 3 – Classificação das Escalas de Medidas de Torgerson.....	41
Quadro Nº 4 - Categorias da Análise do SAEB 2003.....	74
Quadro Nº 5 - Coeficientes de correlação significativos ($P < 0,05$ ou $0,001$) das variáveis do aluno, do professor, o diretor e da escola com a variável <i>proficiência</i> <i>em matemática</i> no arquivo de 4ª série das escolas públicas.....	81
Quadro Nº 6 - Coeficientes de correlação significativos ($P < 0,05$ ou $0,001$) das variáveis do aluno, do professor, do diretor e da escola com a variável <i>proficiência</i> <i>em matemática</i> no arquivo de 4ª série das escolas particulares.....	81
Quadro Nº 7 - Coeficientes de correlação significativos ($P < 0,05$ ou $0,001$) das variáveis do aluno, do professor, do diretor e da escola com a variável <i>proficiência</i> <i>em matemática</i> no arquivo da 8ª série das escolas públicas.....	82
Quadro Nº 8 - Coeficientes de correlação significativos ($P < 0,05$ ou $0,001$) das variáveis do aluno, do professor, do diretor e da escola com a variável <i>proficiência</i> <i>em matemática</i> no arquivo da 8ª série das escolas particulares.....	83
Quadro Nº 9 - Homogeneidade das Variâncias.....	84
Quadro Nº 10 - Verificação Posterior à Regressão de Violação dos Pressupostos para a sua Realização (Método “Stepwise”) – Resultados: 4ª série.....	87
Quadro Nº 11 - Verificação Posterior à Regressão de Violação dos Pressupostos para a sua Realização (Método “Stepwise”) - Resultados: 8ª série.....	87
Quadro Nº 12 - Regressão (Método “Stepwise”) - Resultados: 4ª série.....	87
Quadro Nº 13 - Regressão (Método “Stepwise”) - Resultados: 4ª série.....	87
Quadro Nº 14 - Regressão (Método “Stepwise”) - Resultados: 8ª série.....	89
Quadro Nº 15 - Regressão (Método “Stepwise”) – Resultados: 8ª série.....	89

LISTA DE GRÁFICOS

Gráfico 1: modelo estatístico de proposições e variáveis.....	30
Gráfico 2: Comprimento (propriedade do objeto).....	38
Gráfico 3: Exemplo de uma CCI de um item.....	53
Gráfico 4: Exemplo de localização do parâmetro "a" numa CCI.....	53
Gráfico 5: Exemplo de localização do parâmetro "b" numa CCI.....	54
Gráfico 6: Exemplo de localização do parâmetro "c" numa CCI.....	55
Gráfico 7: Exemplo de curvas de diversos itens com diferentes valores dos parâmetros.....	55
Gráfico 8: Desempenho dos alunos da 4ª série em Matemática Ensino Fundamental.....	61
Gráfico 9: Desempenho dos alunos da 8ª série em Matemática Ensino Fundamental.....	61
Gráfico 10: Desempenho dos alunos da 3ª série em Matemática Ensino Médio.....	62
Gráfico 11: Nulo.....	72

SUMÁRIO

INTRODUÇÃO.....	13
1 PESQUISA E AVALIAÇÃO QUANTITATIVAS E METODOLOGIA POSITIVISTA - ABORDAGEM FORMAL E ABORDAGEM DIALÉTICA DAS CATEGORIAS QUANTIDADE E QUALIDADE.....	17
2 TEORIA DA MENSURAÇÃO	34
2.1 DEFINIÇÃO DE TERMOS.....	34
2.2 DESENVOLVIMENTO DA TEORIA DA MENSURAÇÃO.....	35
2.3 TEORIA CLÁSSICA DOS TESTES (TCT): O MODELO LINEAR CLÁSSICO DE MENSURAÇÃO.....	43
2.4 TEORIA DA RESPOSTA AO ITEM (TRI).....	50
2.4.1 PRESSUPOSTOS DA TRI.....	51
2.4.2 CURVA CARACTERÍSTICA DO ITEM (CCI).....	52
3 O SISTEMA NACIONAL DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA-SAEB.....	60
3.1 ANÁLISE DE REGRESSÃO LINEAR MÚLTIPLA.....	65
3.1.1 ANÁLISE DE REGRESSÃO LINEAR MÚLTIPLA.....	65
3.1.2 CONDIÇÕES A SEREM CUMPRIDAS PARA A REALIZAÇÃO DA ANÁLISE DE REGRESSÃO MÚLTIPLA.....	71
3.2 METODOLOGIA.....	74
3.2.1 POPULAÇÃO.....	74
3.2.2 AMOSTRA.....	75
3.2.3 INSTRUMENTOS.....	75
3.2.4 ORGANIZAÇÃO DOS DADOS DO SAEB 2003 NO CEARÁ.....	75
3.2.5 TRATAMENTO DOS DADOS.....	76
4 APRESENTAÇÃO E ANÁLISE DOS DADOS.....	78
4.1 DESCRIÇÃO DA AMOSTRA DA QUARTA SÉRIE.....	78
4.1.1 PREPARAÇÃO DOS DADOS PARA A ANÁLISE DE REGRESSÃO.....	79
4.1.2 LINEARIDADE.....	83
4.1.4 HOMOGENEIDADE DAS VARIÂNCIAS.....	83
4.1.5 REALIZAÇÃO DAS REGRESSÕES.....	84
5 CONSIDERAÇÕES CONCLUSIVAS.....	91
BIBLIOGRAFIA.....	94
APÊNDICES.....	98

INTRODUÇÃO

O período em que prestei assessoria à SEDUC, na área de pesquisa e avaliação educacional, permitiu-me perceber que os projetos de intervenção nas políticas educacionais, aqueles próprios da secretaria, com recursos de seu orçamento, ou com recursos de instituições externas centravam-se em alguns fatores sob a expectativa de atingir o processo educacional como um todo. No entanto, os resultados têm sido pífios. Um exemplo desse contexto foi o projeto “Internet nas Escolas”, com grande publicidade na mídia e nos muros das escolas públicas, com esperança de que o problema da aprendizagem estaria resolvido. Entretanto, ao serem olhados os resultados das avaliações realizadas ao longo do tempo, a partir de 1990 pelo SAEB, bem como pelo SPAECE, o que se verifica é que em matemática e português, os resultados são muito baixos e vêm caindo a cada avaliação ou, quando muito, se mantendo no mesmo patamar.

Apesar das avaliações do SAEB e do próprio SPAECE levantar um número grande de dados referentes aos alunos avaliados, a seus professores e aos diretores de suas respectivas escolas, observa-se, nos relatórios dessas avaliações, que a análise dos resultados está concentrada no desempenho dos alunos nos testes de rendimento escolar aplicados. Por outro lado, em alguns desses relatórios, encontram-se análises relacionando, isoladamente, algumas variáveis medidas no diretor, no professor e no aluno, outras referentes aos equipamentos e estado de conservação de instalações dos prédios escolares com o desempenho dos alunos nos testes.

Esse conhecimento pontual do relacionamento de variáveis, tomadas isoladamente, com o rendimento dos alunos nos testes aplicados tem gerado políticas de intervenção que não apresentam os resultados desejados. Isso decorre de uma visão fragmentada do processo educacional, gerando o que SAVIANI (2000) denomina como “*flutuações da consciência pedagógica*”; ora privilegia-se o fator sociológico, ora o fator psicológico, ora o fator econômico, dentre outros.

A educação, como qualquer processo social, é uma totalidade inserida numa totalidade maior que é a própria sociedade sendo que os diversos fatores que a caracterizam, além de interagirem entre si, interagem com os fatores mais gerais, determinantes dos demais processos sociais.

Por outro lado, o Sistema de Avaliação da Educação Básica-SAEB disponibiliza os arquivos com os dados e sua aplicação. Existem alguns trabalhos realizados em outros estados que fazem análises conjuntas da relação das variáveis medidas pelos SAEB com o rendimento dos alunos nos testes aplicados, permitindo uma visão de totalidade do processo educacional, evitando intervenções pontuais que, historicamente, têm caracterizado a definição de políticas públicas na área de educação.

Desde a primeira avaliação do SAEB em 1990, as aferições do rendimento dos alunos têm mostrado a situação degradante da educação básica brasileira. Observa-se que é dado um valor mínimo aos resultados mostrados por esse tipo de avaliação em larga escala. Os resultados têm servido mais aos meios de comunicação para ressaltarem o fracasso do sistema educacional, principalmente no ataque à escola pública.

No entanto, apesar da riqueza de informações produzidas pelo SAEB, nota-se também uma grande rejeição por se apresentarem os dados quantitativamente. Possivelmente essa rejeição é constituída por dois fatores principais.

O primeiro situa-se na vertente do processo de negação do tecnicismo que permeou a educação durante a ditadura instalada a partir de 1964.

O segundo é resultado de um processo educacional que não ensina. Nas disciplinas básicas, matemática e língua portuguesa, as deficiências do processo de ensino e aprendizagem têm gerado pessoas que não sabem escrever com coerência, não sabem interpretar corretamente um texto, não sabem processar as mais simples operações matemáticas necessárias à vida diária.

Diante desse contexto, é negado o uso da metodologia quantitativa sob a alegação de que se trata de uma metodologia reacionária (*sic*) e de caráter

positivista. Quando é pedida uma justificativa para essa afirmação, pessoas que a fazem, não são capazes de justificá-la.

Assim, o objetivo geral do trabalho é encontrar entre o conjunto das variáveis, medidas pelo SAEB/2003, que se relacionam significativamente com rendimento dos alunos da amostra no teste de matemática aplicado, na 4^a e 8^a séries do ensino fundamental no Ceará, um subconjunto de variáveis que proporcionam um maior grau de explicação no rendimento do teste, bem como mostrar que, necessariamente, pesquisa e avaliação com metodologia quantitativa não se fundamentam no positivismo.

Em função do exposto, cuidou-se no primeiro capítulo em mostrar que a abordagem quantitativa dos fenômenos sociais não se fundamenta, necessariamente, no positivismo, no sentido de justificar a análise dos resultados do SAEB através de comparações entre as metodologias quantitativa e qualitativa, buscando evidenciar como falsa a associação entre o positivismo, pesquisa e avaliação educacional realizadas com metodologia quantitativa.

No segundo capítulo, são discutidas as duas abordagens das categorias quantidade e qualidade: a abordagem formal e a abordagem dialética, visando mostrar que a disputa entre metodologia quantitativa e metodologia qualitativa, observada nas ciências humanas, traduz-se num falso debate.

No terceiro capítulo, é tratada a teoria da mensuração, sendo descritos os dois modelos prevalentes na pesquisa e na avaliação educacional, o modelo linear clássico de mensuração e a teoria de resposta ao item, objetivando situar a análise dos dados do SAEB/2003.

A descrição do Sistema Nacional de Avaliação Educacional Básica é contemplada no quarto capítulo, enfatizando a evolução e os procedimentos usados para o tratamento dos dados. No quinto capítulo, é feita a descrição metodológica dos dados do SAEB/2003, no Estado Ceará, utilizados neste trabalho: a análise de regressão linear múltipla permitindo selecionar, no conjunto das variáveis medidas, um subconjunto de variáveis que apresentam um maior grau de explicação no rendimento em matemática nas 4^a e 8^a séries do ensino fundamental, na rede pública e privada do sistema educacional do

Estado. Tal prática de abordagem pode fornecer subsídios para a formulação de políticas educacionais capazes de melhorar significativamente o rendimento escolar.

O sexto capítulo trata da metodologia utilizada no trabalho, envolvendo a descrição da população e da amostra, além dos instrumentos utilizados para o levantamento dos dados, como os arquivos disponibilizados pelo INEP foram organizados e tratados para o estudo, bem como o tratamento dos dados.

Os capítulos sete e oito, respectivamente, apresentam e analisam os dados e fazem considerações conclusivas.

Diante dessa exposição, estabeleceu-se como objetivo geral relacionar um subconjunto de variáveis que proporcionam um maior grau de explicação no rendimento do teste de matemática aplicado na 4ª e 8ª séries do ensino fundamental, no SAEB/2003, evidenciando, também, que pesquisa e avaliação com metodologia quantitativa não se fundamentam necessariamente no positivismo. Para se atingir este fim, foram definidas duas vertentes de ordem específica:

1. mostrar que pesquisa e avaliação em educação, com metodologia quantitativa, não se relacionam necessariamente com o positivismo;
2. construir, com os dados das variáveis medidas pelo SAEB/2003, um modelo linear múltiplo com variáveis entre as que apresentaram relação significativa com rendimento em matemática em alunos de 4ª e 8ª séries do ensino fundamental do Estado do Ceará, seus professores dessa disciplina, diretores e características físicas das respectivas escolas, garantindo um maior grau de explicação desse rendimento.

1 PESQUISA E AVALIAÇÃO QUANTITATIVAS E METODOLOGIA POSITIVISTA: ABORDAGEM FORMAL E ABORDAGEM DIALÉTICA DAS CATEGORIAS QUALIDADE E QUANTIDADE

A ruptura de GALILEU (1973) com a tradição aristotélica consistiu na negação do geocentrismo e, principalmente, na afirmação de que os modelos matemáticos seriam a própria expressão da realidade. As qualidades primárias seriam todas elas quantitativas; assim, esse entendimento fixou-se com a visão de que os fenômenos poderiam ser entendidos pela análise matemática (descrição do tipo geométrico) que permitiria desvendá-los.

Esta corrente filia-se à tradição empirista de um conhecimento baseado na observação sensorial da realidade exterior à idéia. Ao mesmo tempo, desenvolveu-se outra corrente (racionalista), baseada na observação do homem e de sua própria mente. Tal introspecção pretendia também obedecer às leis matemáticas, não no sentido quantitativo, mas em relação à definição clara e distinta dos conceitos e ao relacionamento a se estabelecer entre eles (DESCARTES, 1999).

Constituíam-se como método a descrição algébrica ou geométrica que essa corrente cartesiana propunha para o novo referencial. A matemática aparece, desde então, como uma linguagem do entendimento comum a ambas as correntes, explicando no realismo as operações da matéria e, no racionalismo, as das idéias. Mas tanto numa tendência como na outra, a origem das idéias se situa na realidade exterior à mente. Uma caracteriza-se como realismo objetivo e a outra como realismo subjetivo.

Essa condição, característica das duas correntes, trata formalmente as categorias do entendimento e, conseqüentemente, vêem-nas separadas, o que vai dar origem ao que está no fundo da discussão que ocorre nas ciências humanas entre metodologia quantitativa e metodologia qualitativa. KANT (1985), quando apresenta as categorias *a priori* do entendimento como totalidades, o faz de forma isolada. Em sua tábua constam quatro categorias que KANT (1985) enumera

“1 - da quantidade (unidade, pluralidade e totalidade); 2 - da qualidade (realidade, negação e limitação) 3 - da relação (inerência e subsistência [substantia et accidens], causalidade e dependência [causa e efeito], comunidade [ação recíproca entre o agente e paciente]), 4 – da modalidade (possibilidade-impossibilidade e existência-não existência). Esta é pois a lista de todos os conhecimentos, originariamente puros, da síntese que o entendimento ‘a priori’ contém em si, e apenas graças aos quais é um entendimento puro; só mediante eles pode compreender algo no diverso da intuição, isto é, pode pensar num objecto. (Op Cit, p.110-111).

HEGEL (1995), com sua dialética idealista supera a separação formal entre as categorias. No caso das categorias qualidade e quantidade, trata-as em relacionamento antitético, com a tríade qualidade-quantidade-medida. A oposição quantidade-qualidade é superada na medida.

Os lados da relação são ainda quantos [quanta] imediatos, e as determinações qualitativas e quantitativas são ainda exteriores uma a outra. Mas, segundo sua verdade – de que o quantitativo mesmo é relação para consigo em sua exterioridade, ou seja, que o ser-para-si e a indiferença da indeterminidade - , um quantitativo é a medida – adendo: (...) Primeiro tínhamos, como conceito de quantidade, a qualidade supra-sumida, isto é, a determinidade apenas exterior, não idêntica ao ser, mas indiferente ao seu respeito. (...) É esse conceito o que está também na base da definição comum de grandeza na matemática: “aquilo que pode ser aumentado (Op cit, p.212).

Portanto, na visão idealista de HEGEL (1995), a primeira determinação do espírito sobre o ser-em-si, é qualitativa, que leva a uma determinação quantitativa. Por exemplo, quando o espírito se volta para um objeto, ele põe uma categoria de cor (qualidade), levando isso a uma distinção do objeto em relação aos demais que o cercam. Tal fato conduz a uma percepção do objeto como um volume (quantidade), tornando possível determinar metricamente o seu volume (medida).

Tal perspectiva deu condições a MARX (2003), de operar dialeticamente em uma base realista sem negar a subjetividade, pois

[...] não se julga o que um indivíduo é pelo que ele imagina de si próprio, tão-pouco se pode julgar uma tal época de revolucionamento a partir da sua consciência, mas se tem, isso sim, de explicar esta consciência a partir das contradições da vida material, do conflito existente entre forças produtivas e relações de produção sociais. Uma

formação social nunca decai antes de estarem desenvolvidas todas as forças produtivas para as quais é suficientemente ampla, e nunca surgem relações de produção novas e superiores antes de as condições materiais de existência das mesmas terem sido chocadas no seio da própria sociedade velha. Por isso a humanidade coloca sempre a si mesma apenas as tarefas que pode resolver, pois que, a uma consideração mais rigorosa, se achará sempre que a própria tarefa só aparece onde já existem, ou pelo menos estão no processo de se formar, as condições materiais da sua resolução.

O entendimento das condições materiais do conhecimento humano permeia o próprio caminho da construção do conhecimento, sua existência demanda da produção das próprias condições materiais de vida. Por outro lado, um indivíduo ao se imaginar o que é, pela sua própria condição de existir, deve imaginar de si próprio pelo que produz, pois tão-pouco pode julgar um revolucionamento a partir da sua consciência, sem ter a própria consciência de si. Entretanto, se tem a explicação dessa consciência, a partir das contradições da vida material, do conflito existente entre forças produtivas e relações de produção sociais, o levam a um estágio superior dentro da produção histórica do conhecimento, como condição da existência.

Diz ainda MARX,

[...] Como, em geral, em toda a ciência histórica, social, ao observar o desenvolvimento das categorias econômicas há que ter sempre presente que o sujeito - neste caso a sociedade burguesa moderna - é algo dado tanto na realidade como na mente; e que, por conseguinte, essas categorias exprimem formas e modos de existência, amiudadamente simples aspectos desta sociedade, deste sujeito; e que, portanto, mesmo do ponto de vista científico, esta sociedade não começa a existir de maneira nenhuma apenas a partir do momento em que se começa a falar dela como tal. Uma regra a fixar, pois dá-nos elementos decisivos para o [nosso] plano [de estudo].

Como se percebe, a sociedade burguesa, ao apropriar-se da produção material do homem, que se desenvolve historicamente, transcende apropriando-se da produção do conhecimento. Se para HEGEL, o espírito se volta para um objeto, a “coisa em si”, lançando-se sobre ele e mudando seu caráter qualitativo para “coisa para si”, para MARX o processo é inverso, pois o caráter qualitativo permanece na “coisa para si”, mas como idéia de reflexo do objeto sobre o espírito. Assim, MARX situa sua teoria do conhecimento no

realismo objetivo. Por exemplo, em MARX a matéria existe na natureza como matéria bruta. O homem ao aplicar trabalho na matéria bruta, retirando-a para determinado uso, esta matéria retirada muda de qualidade, transformando-se em matéria prima.

[...] O concreto é concreto porque é a síntese de múltiplas determinações e, por isso, é a unidade do diverso. Aparece no pensamento como processo de síntese, como resultado, e não como ponto de partida, embora seja o verdadeiro ponto de partida, e, portanto, também, o ponto de partida da intuição e da representação. No primeiro caso, a representação plena é volatilizada numa determinação abstrata; no segundo caso, as determinações abstratas conduzem à reprodução do concreto pela via do pensamento. Eis por que Hegel caiu na ilusão de conceber o real como resultado do pensamento que, partindo de si mesmo se concentra em si mesmo, se aprofunda em si mesmo e se movimenta por si mesmo; ao passo que o método que consiste em elevar-se do abstrato ao concreto é, para o pensamento, apenas a maneira de se apropriar do concreto, de o reproduzir na forma de concreto pensado; porém, não é este de modo nenhum o processo de gênese do concreto em si. Com efeito, a mais simples categoria (...) supõe uma população, população essa que produz em condições determinadas; supõe ainda um certo tipo de família, ou de comunidade, ou de Estado, etc. Tal valor não pode existir nunca senão sob a forma de relação unilateral e abstrata, no seio de um todo concreto e vivo já dado.

Entender a coisa posta no mundo como concreto, inclusive a tudo aquilo que vem da interação do sujeito com o mundo concreto, para MARX é perceber e se apropriar do processo da gênese do concreto em si, pois a mais simples categoria derivada dessa relação mundo-sujeito histórico, supõe uma forma de relação unilateral e abstrata que define o concreto vivo e já vivido, principalmente quando se refere à sobrevivência no mundo material. Na verdade, a síntese de múltiplas determinações aparece no pensamento como processo de síntese e como resultado, e não como ponto de partida, muito embora, o verdadeiro ponto de partida, é o ponto de partida da intuição e da representação. Em qualquer caso, a representação plena é dissolvida numa determinação abstrata, que conduz à reprodução do concreto pela via do pensamento.

... Assim, para a consistência filosófica - que considera que o pensamento que concebe é o homem real, e que, portanto, o mundo só é real quando concebido - para esta consciência, é o movimento das categorias que lhe aparece com um verdadeiro ato de produção (o qual recebe do exterior um pequeno impulso, coisa que esta consciência só muito a contragosto admite <que produz o mundo. Isto é exato (embora aqui nos vamos encontrar com uma nova tautologia>, na medida em que a totalidade concreta, enquanto totalidade do pensamento, enquanto concreto do pensamento é in fact um produto do pensamento, do ato de conceber; não é de modo nenhum, porém, produto do conceito que pensa e se gera a si próprio e que atua fora e acima da intuição e da representação; pelo contrário, é um produto do trabalho de elaboração, que transforma a intuição e a representação em conceitos. O todo, tal como aparece na mente como um todo pensamento, é produto da mente que pensa e se apropria do mundo do único modo que lhe é possível; modo que difere completamente da apropriação desse mundo na arte, na religião, no espírito prático. O sujeito real conserva a sua autonomia fora da mente, antes e depois, pelo menos durante o tempo em que o cérebro se comporte de maneira puramente especulativa, teórica. Por consequência, também no método teórico é necessário que o sujeito - a sociedade - esteja constantemente presente na representação como ponto de partida. (Marx; 2003; p.3-5).

Nessa perspectiva, a consistência do entendimento que considera que o pensamento que concebe a produção histórica do conhecimento é a própria consciência do homem real que, no mundo real, concebido aquilo que são as condições materiais de sua própria existência. Para MARX, esta consciência é o movimento das categorias que lhe aparece com um verdadeiro ato de produção que produz o mundo material e objetivo. Isto é exato na medida em que a totalidade concreta, enquanto totalidade do pensamento, enquanto concreto do pensamento, no ato de conceber e no ato de produzir o conhecimento não é, por nenhuma hipótese, produto do conceito ou da idéia de que o homem que pensa e se gera a si próprio e que atua fora e acima da intuição e da representação. De outra forma, tudo que é um produto do trabalho de elaboração, que transforma a intuição e a representação em conceitos, tem que ser considerado como categoria de produção, dentro da apropriação desse mundo, seja na arte, na religião, no espírito prático etc.

A metodologia quantitativa tanto é usada na pesquisa, quanto na avaliação educacional. Apesar da rejeição ao neopositivismo (positivismo lógico, empirismo lógico) a metodologia quantitativa é erroneamente

identificada como baseada em pressupostos do positivismo. Esta falsa concepção permeia a disputa entre os praticantes da metodologia qualitativa e aqueles que usam a metodologia quantitativa.

Na literatura pertinente, verifica-se, com grande incidência, referir-se à metodologia qualitativa como sendo pós-positivista.

Visando contribuir para o esclarecimento de algumas concepções falsas, discutir-se-ão, a seguir, os tópicos:

- a) comparações entre as metodologias quantitativas e as qualitativas;**
- b) a falsa associação entre o positivismo e pesquisa e avaliação educacional utilizando metodologia quantitativa;**
- c) a superação dialética da dicotomia quantidade-qualidade.**

É comum se observar, presentes nas justificativas sobre a utilização de uma metodologia qualitativa, textos (livros e artigos em publicações científicas) com comparações entre os dois tipos de metodologia. Por exemplo, como atribuição à metodologia quantitativa de fundamento no positivismo, pode-se citar como exemplo BERGER e LUCKMANN (1996) que identificam explicitamente pesquisa quantitativa como sendo baseada no positivismo

“Os positivistas utilizam metodologias empíricas tomadas das ciências naturais para investigar os fenômenos. Estratégias quantitativas servem ao ideal da ciência positiva por proverem a agregados de dados o rigor, a precisão e a verificabilidade, através da testagem estatística de hipótese empírica” (Op. Cit., p. 10).

MERRIAM (1998) atribui características positivistas aos métodos quantitativos:

“Na forma positivista da pesquisa (...) o conhecimento obtido através da pesquisa experimental é objetivo e quantificável (...) a partir de uma perspectiva positivista você deve começar hipotetizando... Você pode então delinear um tipo de intervenção [na realidade]. Você configura o experimento controlando tantas variáveis quanto possível e então mede os resultados.” (Op. Cit., p. 4).

GUBA (1990) defende o fim da disputa entre qualidade e quantidade envolvendo pesquisadores educacionais através do diálogo entre as duas tendências. Apesar disso, ele identifica a metodologia quantitativa como positivista.

No desdobramento deste posicionamento, LAUDAN (1977) considera que o debate envolvendo a oposição entre os dois tipos de metodologia está situado tendo por base a noção de paradigma introduzida por KUHN (1975), que levou à polarização de diferenças metodológicas, à possibilidade de transição entre paradigmas rivais minimizada pelo fato que os defensores de paradigmas diferentes situam-se também em padrões metodológicos diferentes e não têm valores cognitivos idênticos ao paradigma a que se opõem.

Assim, teorias rivais baseadas em paradigmas diferentes são radicalmente incomensuráveis. A baixa determinação de um conjunto de dados significa que as regras ou críticas de avaliação da ciência não escolhem uma única teoria para a exclusão de todos os seus oponentes.

FEYRABEND (1977) afirma que muitos cientistas bem sucedidos violam normas e cânones chamados de científicos. Acredita que o consenso não é desejável para o progresso da ciência; a ciência progride através de um questionamento sem fim de sua fundamentação, tendo essa arguição assentada na filosofia natural pré-socrática.

Nas comparações entre características da pesquisa quantitativa com a pesquisa qualitativa as categorias científicas que mais se apresentam na literatura são:

- a) objetiva (quantitativa) X subjetiva (qualitativa);
- b) particularista (quantitativa) X holística (qualitativa);
- c) orientada para o resultado (quantitativa) X orientada pelo processo (qualitativa);
- d) categorias fixas (quantitativa) X categorias emergentes (qualitativa);
- e) realidade estática (quantitativa) X realidade fluída (qualitativa);
- f) mecânica (quantitativa) X criativa (qualitativa);
- g) formulística (quantitativa) X interpretativa (qualitativa);

- h) reducionista (quantitativa) X expansionista (qualitativa);
- i) fundamentada em teoria (quantitativa) X não fundamentada em teoria (qualitativa);

Quadro Nº 1: Categorias científicas e principais características das metodologias quantitativa e qualitativa

METODOLOGIA	
QUANTITATIVA	QUALITATIVA
OBJETIVA	SUBJETIVA
PARTICULARISTA	HOLÍSTICA
ORIENTADA PARA O RESULTADO	ORIENTADA PELO PROCESSO
CATEGORIAS FIXAS	CATEGORIAS EMERGENTES
REALIDADE ESTÁTICA	REALIDADE FLUÍDA
MECÂNICA	CRIATIVA
FORMULÍSTICA	INTERPRETATIVA
REDUCIONISTA	EXPANSIONISTA
FUNDAMENTADA EM TEORIA	NÃO FUNDAMENTADA EM TEORIA

Essas comparações baseiam-se na concepção incorreta que associam a metodologia quantitativa ao paradigma positivista/neopositivista bem como se limitam à percepção formal, tratando isoladamente duas categorias antitéticas.

A fim de se examinar a relação entre positivismo e metodologia quantitativa torna-se necessário caracterizar o positivismo.

KOLAKOWSKI (1972), identifica como positivismo

“(...) uma atitude normativa, regulando como nós devemos usar termos como ‘conhecimento’, ‘ciência’, ‘cognição’ e ‘informação’ (...) as regras positivistas distinguem entre as disputas filosóficas e científicas aquelas que podem ser perseguidas com proveito e aquelas que não têm nenhuma chance de serem levadas a cabo e consequentemente não merecem ser tomadas em consideração” (Op. Cit., p. 11)

Segundo esse autor, são quatro as regras mais importantes do positivismo que devem ser observadas pela ciência no sentido de descartar

questões que são falsamente formuladas ou envolvem conceitos ilegítimos, são: a *Regra do fenomenismo*, a *Regra do nominalismo*, a *Regra que recusa julgamento de valor e afirmações normativas* e a *Regra da crença na unidade da ciência através do método*.

Na primeira, a *regra do fenomenismo*, a distinção entre essência e fenômeno deve ser eliminada da ciência. Entende-se por essência a natureza íntima das coisas, aquilo que faz com que uma coisa seja o que é e, por fenômeno, aquilo que pode ser percebido pelos sentidos. Para o positivismo somente tem sentido para a ciência aquilo que se manifesta através da experiência, “*opiniões relativas a entidades ocultas das quais coisas experimentadas são supostamente a manifestação, são sem sentido*” KOLAKOWSKI (1972, p 11).

Segundo a *Regra do nominalismo*, qualquer “*insight*” formulado em termos gerais não pode se referir senão a objetos individuais concretos. O mundo que nós conhecemos é uma coleção de fatos individuais observados. O nominalismo não considera a existência real de gênero e espécie,

“[...] *um sistema ordenando nossas experiências deve ser tal que ele não introduz na experiência mais entidades além daquelas contidas na experiência (...) essas abstrações são mais ou menos que meios em ações humanas que servem para organizar experiências (...) elas não têm nenhuma experiência separada*” (KOLAKOWSKI, 1972; p. 15).

Pela terceira, a *Regra que recusa julgamento de valor e afirmações normativas*: a experiência não produz qualidade como bom, mau, belo ou outras. Essas qualidades provêm de julgamentos de valor. Qualquer experiência não nos pode obrigar aceitar afirmações contendo comandos ou proibições baseadas em valores que nos digam para fazer algo ou para não o fazer.

Por último, a *regra da crença na unidade da ciência através do método*, segundo a qual os métodos para aquisição do conhecimento válido, em todas as espécies da experiência são os mesmos. As diferenças qualitativas entre ciências particulares são reflexo de um estágio particular do desenvolvimento histórico dessas ciências. A evolução reduzirá todos os domínios do conhecimento a uma única ciência. Desde COMTE (1852) supõe-se que essa

unidade em uma única ciência dar-se-á pelo método tendo como modelo a Física, uma vez que entre todas as disciplinas experimentais a Física desenvolveu os métodos mais exatos de descrição.

Segundo HACKING (1983) e Yu (2002), os temas principais do positivismo são: *Ênfase na verificação*, *Pró-observação*, *Negação da relação causa-efeito*, *Explicação reducionista*, *Negação de entidades teóricas*, *Negação da metafísica*.

De acordo com a *Ênfase na verificação*, uma afirmação é sem sentido se não é possível verificá-la ou se o critério de verificação não é claro. Esta noção aplicada de modo radical faz com que as afirmações da moral, da ética, da estética, etc. sejam consideradas não verificáveis e, desta forma, sem sentido. O princípio da verificação torna-se nada mais que uma definição de sentido e verificabilidade que são intercambiáveis. Os métodos estatísticos não provêm à verificação um sentido lógico do positivismo. A lógica da testagem estatística de uma hipótese não é verificar se ela é verdadeira ou falsa. Sua lógica é encontrar a possibilidade de se estender uma relação observada através de uma amostra à população representada pela amostra. A lógica da testagem de hipótese baseia-se em aceitar ou deixar de aceitar a hipótese nula (que nega a existência na população da relação observada na amostra), dada certa probabilidade.

Na *Pró-observação*, enquanto o verificacionismo estabelece os sentidos do conhecimento, a observação empírica é uma metodologia específica para verificação. Para os positivistas, a realidade refere-se à experiência. A realidade é reflexo na consciência do mundo sentido, assim o positivismo não implica numa posição de realismo. No entanto, a visão da pesquisa quantitativa como tendo por base ontológica a existência de uma realidade objetiva não conduz necessariamente à aceitação do positivismo como uma forma de realismo. O positivismo lógico é visto, segundo LAUDAN (1977 e 1996) como um tipo de convencionalismo, relativismo e subjetivismo.

Para o positivismo lógico, mesmo se exista uma realidade última, objetiva, não se tem contato direto com essa realidade, pois a única coisa que interessa é com o que se está em contato através da experiência.

A metodologia quantitativa não é necessariamente objetiva. A abordagem Bayseana opera inicialmente com probabilidade subjetiva que envolve uma crença de chance assentada em algum conhecimento prévio. Essa probabilidade subjetiva será mais tarde corrigida por uma probabilidade posterior baseada em dados observados. Essa noção parece apoiar a asserção de que a pesquisa quantitativa existe numa realidade objetiva. Entretanto, de acordo com a teoria bayseana da probabilidade somente duas probabilidades são verdadeiras, zero – quando o evento esperado não ocorre, e, um – quando o evento esperado ocorre. Antes da ocorrência do evento a probabilidade é um grau de crença, logo é subjetiva. Embora processos de interação possam conduzir a probabilidades subjetivas e a algum grau de convergência, a utilização de procedimentos bayseanos não afirma a existência de razões para substanciar uma verdade como conclusão; em vez disso, se afirma que há razões para modificar uma crença à luz de novas evidências.

Com relação à *Negação da relação causa-efeito*, atribuem-se incorretamente inferências causais ao positivismo lógico. No entanto NATION (1997) e GLINER e MORGAN (2000) afirmam a aceitação da relação causa-efeito como própria do positivismo lógico.

Segundo HUME (1980), uma das fontes do positivismo, a noção de causa nada mais é que a percepção de uma associação regular entre dois fenômenos, nada garante que, quando se observa a ocorrência de um fenômeno **A** seguido pela ocorrência de um fenômeno **B**, implique em que **A** cause **B**. A relação de causa-efeito se escrita matematicamente por uma função, do 1º grau dá-nos:

$$y = ax + b,$$

com a permutação das duas variáveis, tem-se:

$$x = \frac{y - b}{a}$$

Dessa forma **x** não pode ser visto como causa de **y**, porque as posições de **x** e **y** podem ser trocadas na equação. Assim, a relação de causa-efeito como

uma propriedade ontológica não é aceita pelos positivistas. Mesmo quando positivistas operam com essa relação isso nada mais é do que um mero fio condutor para a produção do conhecimento, como postula ZETTERBERG (1970).

HOYLE (1995) apresenta três critérios a serem entendidos para a validação de inferências causais: direcionamento, isolamento e associação. O direcionamento refere-se à influência de n variáveis independentes ($n = 1; 2; 3; \dots$) sobre a variável tomada como dependente. A isolamento refere-se a que interferências estranhas e erros de mensuração devem ser isolados do estudo de modo que a relação observada possa ser explicada pelo que é proposto teoricamente e, finalmente, na associação, as variáveis independentes e a variável dependente são correlacionadas matematicamente.

Para estabelecer a direção das variáveis, aplica-se a lógica, (por exemplo, a altura das pessoas não se relaciona com seu desempenho acadêmico), teoria (por exemplo, a colaboração entre membros de um grupo na realização de uma tarefa) e, também, o delineamento da pesquisa (por exemplo, outras explicações possíveis são mantidas fora do experimento).

Para se garantir o critério de isolamento, o processo de mensuração deve ser levado a cabo no sentido de garantir a validade, a precisão e a redução dos erros de mensuração. Por outro lado, a interferência de variáveis estranhas ao estudo deve ser controlada. Por fim, a associação é obtida através de técnicas estatísticas para se calcular a associação entre variáveis. Mesmo quando se observa uma forte associação matemática, a inferência causal pode não fazer sentido se os critérios de direcionalidade e isolamento não são estabelecidos previamente.

Dessa forma, a negação da relação de causa-efeito como propriedade ontológica é incompatível com muitas áreas do método quantitativo, tais como os experimentos aleatórios, os quase-experimentos, a teoria do constructo latente, a modelagem de equações estruturais e a meta-análise dentre outros, a explicação da mesma forma que ocorre com a causalidade é falsamente identificada com positivismo lógico. LANGENBACH, VAUGHN, & AAGAARD

(1994), afirmam que a pesquisa quantitativa é baseada na filosofia positivista e busca explicar a causa das mudanças em fatos sociais.

Isso vai de encontro com a posição positivista que afirma que a pesquisa permite descrever o que acontece, mas não explica ou prescreve. No entanto, o objetivo maior da metodologia quantitativa é explicar, mostrando relações complexas entre variáveis ou variáveis de constructo.

Na *Explicação reducionista*, para o positivismo, o conhecimento resume-se à descrição da realidade. Segundo HEMPEL (1965), para se explicar um evento **S**, aplica-se uma lei **L**; e um fato particular **F** é observado para ligar **S** e **L**. Para o autor, este raciocínio dedutivo é descritivo por natureza, por exemplo: **L** → o homem age de forma racional.

F → entre várias opções, uma delas é mais eficiente para se atingir o objetivo desejado.

S → o homem, agindo racionalmente, escolhe a opção que lhe permite atingir o objetivo desejado.

A abordagem descritiva do positivismo é mais compatível com a pesquisa qualitativa de orientação fenomenológica que a pesquisa quantitativa. Na pesquisa quantitativa, a explicação depende de uma formulação teórica sobre a natureza da relação entre as variáveis. A natureza dessa relação pode ou não ser causal. A partir do referencial teórico pode ser sem sentido calcular correlações parciais, enquanto que em outras situações tais estatísticas têm sentido. Assim, a explicação da pesquisa quantitativa não é inteiramente descritiva. De acordo com KELLEY (1998), freqüentemente, há uma lacuna entre o “*explanandum*” e a explicação (hipótese). Uma boa explicação é capaz de preencher a lacuna, por exemplo:

X → alunos do ensino fundamental provenientes de famílias com renda familiar alta apresentam melhor rendimento escolar que alunos oriundos daquelas com renda familiar baixa.

Para validar essa explicação, outras exposições são necessárias para preencher lacunas nela existentes. Deve-se mostrar que:

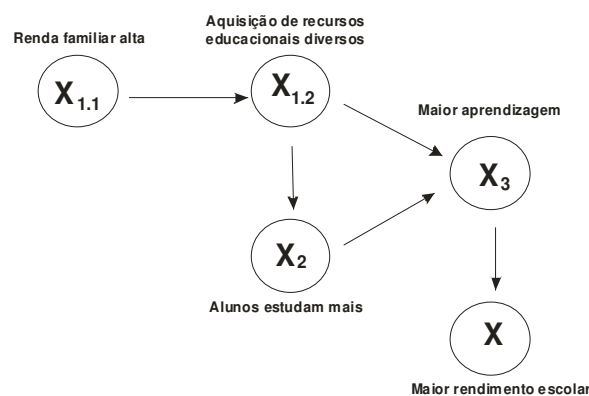
X₁ → famílias que tenham renda familiar alta adquirem mais recursos para a educação de seus filhos, tais como materiais de leitura, computador etc.

X_2 → alunos que provém de famílias com alta renda estudam mais.

X_3 → alunos que estudam mais e são expostos a recursos educacionais diversificados aprendem melhor.

Expressando as proposições acima sob a forma de variáveis mensuráveis, pode-se delinear um modelo estatístico como o representado no diagrama a seguir:

Diagrama 1: Modelo estatístico de proposições e variáveis



A modelagem estatística permite preencher as lacunas entre “*explanandum*” e explicação. A pesquisa quantitativa também busca a explicação, não se reduzindo à simples descrição da realidade, logo, não se enquadra na noção positivista de reducionismo na explicação.

Na *Negação de entidades teóricas*, o positivismo é cético em relação a entidades teóricas não observáveis, como variável latente, fator etc. Pesquisadores que trabalham com metodologia qualitativa fazem uma associação com metodologia quantitativa e o paradigma positivista. No entanto, a pesquisa quantitativa nas ciências humanas opera com conceitos teóricos abstratos. A análise fatorial e o escalonamento multidimensional podem servir como base para a elaboração teórica. Em psicometria e em avaliação educacional, constructos como auto-imagem, motivação intrínseca, são sempre hipotetizados. A validade de constructo é vista como um produto da ligação entre constructos teóricos e dados observáveis. Se a negação de entidades teóricas fosse própria da metodologia quantitativa, pesquisas operando em psicologia e em educação deveriam ser eliminadas por falta de validade, uma

vez que constructos não são definidos operacionalmente, mas funcionam como conceitos abertos [SALVUCCI, WALTER, COLEY, FINK e SABA (1997)].

Com respeito à *Negação da metafísica*, o positivismo nega a metafísica e a existência de uma realidade transcendental. O mundo da ciência é um mundo da vida diária, onde idéias, plantas, animais, estrelas etc., existem. Em filosofia, a metafísica refere-se a um outro mundo. Quando o positivismo refere-se ao mundo como constituído pelo que é acessível aos sentidos (fenômeno), isso está em contradição com sua negação da metafísica, uma vez que a sua visão de mundo supõe uma visão ontológica. A metodologia quantitativa não rejeita necessariamente a existência metafísica. Embora o método estatístico seja considerado empírico, a testagem estatística é uma comparação entre estatísticas observadas e distribuições teóricas. Distribuições, amostragens teóricas não são reproduzidas empiricamente. As distribuições teóricas e grande parte da matemática estão na esfera da metafísica.

Além desses temas próprios tanto ao positivismo clássico como ao positivismo lógico, pode-se fazer uma distinção entre as duas correntes, uma vez que o positivismo lógico acrescenta uma ênfase na análise lógica da linguagem e, em relação aos métodos quantitativos, adota a teoria frequentista de probabilidade de VON MISES (1928 e 1957) *apud* TORANZOS, 1962).

O neopositivismo agrega a esses temas do positivismo clássico a Ênfase na lógica, análise lógica da linguagem e reducionismo.

O positivismo lógico enfatiza exageradamente a análise lógica da linguagem seguindo o atomismo lógico de RUSSELL (1938): os fenômenos complexos podem ser expressos em termos da linguagem matemática e a matemática pode ser reduzida à lógica. Os positivistas lógicos desenvolveram métodos de análise lingüística baseados no reducionismo analítico, nos quais proposições descrevendo afinidades entre fenômenos observados são partidas em seus componentes o necessário e o suficiente para que a relação ocorra. Embora já se tenha mencionado que a metodologia opera também com constructos teóricos, ela não é de nenhum modo uma redução com uma única direção de eventos para dados números, com finalidade de construir modelos matemáticos. Em vez disso, eventos, dados e teoria constituem uma forma

auxiliar de retroalimentação positiva. Quando CRONBACH e MERHL (1955) propuseram o conceito de validade de constructo, eles partiram do pressuposto de que os constructos, hipotéticos, iluminam a natureza dos dados a serem coletados. Por sua vez, os dados obtidos com a aplicação do instrumento de medida são usados para revisar a própria teoria que se viu como ponto de partida.

Segundo MORENTE (1967) desde a Idade Média, nenhuma corrente enfatizou tanto a lógica quanto o positivismo lógico.

Desse modo, opera-se, mesmo formalmente, em um sistema aberto. Por sua vez o atomismo lógico de Russel produz um modelo matemático derivado de uma análise formal de um sistema fechado. Ele não está preocupado com o que a realidade é ou se os objetos geométricos existem, mas está interessado em descobrir as relações lógicas entre objetos. Alguns estatísticos aceitam a afirmação de que todos os modelos são falsos, apoiando-se no fato de que os dados empíricos não se ajustam perfeitamente ao modelo matemático. Assim sendo, todo modelo tem algum grau de falsidade.

A *Probabilidade frequentista* é tratada no capítulo II, sobre a teoria da mensuração.

Aristóteles descreve dez categorias, entre elas as de quantidade e a de qualidade. As categorias aristotélicas são vistas formalmente, isto é, mesmo quando duas se negam, como no caso quantidade e qualidade, essa negação é uma negação formal, exterior. A negação formal de uma em função da outra, não traz qualquer implicação para existência de qualquer uma das duas.

Já HEGEL (1995), em sua dialética, trata as categorias, entre elas quantidade e qualidade, como uma tríade na qual a terceira categoria é uma síntese das outras duas que se opõem, negando-se e afirmando-se. Para este filósofo idealista, a primeira determinação do espírito é uma determinação qualitativa que leva a uma determinação quantitativa; quando o espírito se volta para o objeto, a primeira determinação que ele lança sobre o objeto é qualitativa, por exemplo, uma *cor*. Isto leva à imposição quantitativa, uma vez que, através da cor, pode-se distinguir o objeto dos demais que o cercam, o

que, por sua vez leva a uma determinação qualitativa, conduzindo o sujeito a uma síntese, a *medida*.

Na lógica dialética, não tem sentido tratar isoladamente as categorias, entre elas quantidade e qualidade. Se, a uma pessoa informa-se sobre um sujeito o numeral 80, para ela isto não terá sentido; porém se se disser que esse sujeito tem 80 Kg de massa, ela saberá que esta informação se refere ao peso do seu corpo. Logo, quando se tem quantidade e qualidade unidas, isso passa a ter sentido objetivo. Assim, a medida se constitui numa qualidade quantificada (Hegel, 1995).

Como, tanto a metodologia quantitativa quanto a metodologia qualitativa, são realistas, a primeira sob a forma de realismo objetivo e a outra sob a forma de realismo subjetivo, nas duas estão implícitas quantidade e qualidade. A avaliação que é uma forma de pesquisa que vai além da descrição dos fenômenos, se diferencia da pesquisa como está implícito no termo que a define. Esse “*vai além*” se completa por um julgamento de valor.

MARX (2003), ao colocar a dialética hegeliana “de cabeça para baixo”, cria as condições, na pesquisa e na avaliação, por estas terem base empírica, para se operar dialeticamente como teoria do conhecimento, como metodologia e como lógica. Cabe ainda ressaltar, que a lógica dialética não nega a lógica formal, porém a supera, vai além.

2 TEORIA DA MENSURAÇÃO

2.1 DEFINIÇÃO DE TERMOS

A teoria da mensuração opera com conceitos que, na linguagem comum, dão margem a diferentes interpretações. Daí a necessidade de defini-los. Entende-se por:

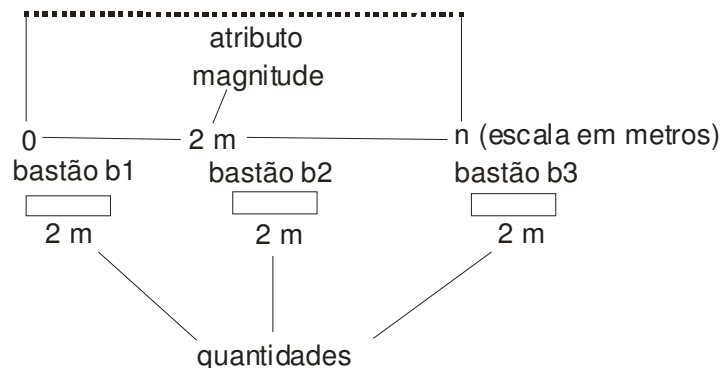
- a) *objeto* ou *sistema* – coisa que possui uma ou mais propriedades;
- b) *atributo* – qualquer propriedade mensurável do objeto;
- c) *propriedade mensurável* – propriedade do objeto capaz de apresentar graduações;
- d) *magnitude* – particularização de um atributo;
- e) *quantidade* – uma instância particular de uma magnitude singular.

Exemplificando, apoiando-se em TORGERSON (1958, p.26), o conceito bastão de madeira tem por extensão uma classe de objetos que têm comprimento como uma de suas propriedades, o atributo refere-se ao contínuo de comprimento que os objetos dessa classe podem apresentar, por exemplo, na escala métrica com valores no intervalo $[0; n]$. A magnitude é o ponto do contínuo que corresponde a um dado comprimento de bastão de madeira. Se, na classe constituída por bastões de madeira, existem bastões aos quais se pode associar uma mesma magnitude, neles, o comprimento se apresenta com quantidades iguais.

Quando uma magnitude pode ser particularizada por uma posição temporal, espacial ou espacial-temporal ou, quando sendo uma relação, pode ser particularizada em consideração a um par de termos entre os quais se situa, então, a magnitude, assim particularizada, é chamada uma quantidade. (RUSSELL, 1938; p.167).

Graficamente, ter-se-ia:

Gráfico 2: Comprimento (propriedade do objeto)



A distinção entre os conceitos de magnitude e de quantidade é fundamental. A construção das escalas de mensuração se sustenta em postulados referentes a relações entre quantidades ou entre magnitudes. As regras de correspondência devem ser apropriadas para cada caso. No primeiro, a ênfase situa-se no sistema empírico, enquanto que, no segundo caso, no sistema formal.

2.2 DESENVOLVIMENTO DA TEORIA DA MENSURAÇÃO

A teoria da mensuração desenvolveu-se fundamentada gnosiologicamente no idealismo platônico. Platão admite três classes de número: os *números ideais*, são os números em si, pertencem ao mundo das idéias e dos quais procedem os demais números; os *números matemáticos*, são os conceitos, desprovidos de matéria e dotados de quantidade e que se encontram na mente e os *números sensíveis*, que se encontram unidos aos objetos do mundo material. Assim, um atributo do objeto (propriedade mensurável) tem existência *a priori* na mente como magnitude que pode ter sua quantidade determinada. O objeto da mensuração é fazer corresponder à propriedade, tal como se apresenta no objeto, uma magnitude que identifique com precisão a quantidade da propriedade no objeto.

HÖLDER em 1901 (*apud* FERNÁNDEZ, 1992) formulou condições para se levar a mensuração a cabo através dos axiomas de quantidade, que

especificam as condições que as quantidades deveriam atender para serem usadas em equações matemáticas. Esses axiomas foram expressos em termos de quantidades reais de elementos específicos. São eles:

- a) axiomas referentes à identificação de ordem ou de igualdade:
 - a.1) $a > b$, ou $a < b$, ou $a = b$;
 - a.2) se $a > b$ e $b > c$, então $a > c$;
 - a.3) para cada elemento a existe um elemento a' , distinto de a , porém, com a mesma quantidade tal que $a = a'$; identidade.
 - a.4) Se $a > b$ e $b > b'$, então $a > b'$; transitividade.
 - a.5) Se $a = b$, então $b = a$; simetria.
 - a.6) Para cada a existe um b tal que $a > b$.
- b) Axiomas referentes à condição para que se possam somar quantidade para se obter uma nova quantidade.
 - b.1) para cada a e b existe um c tal que $c = a + b$;
 - b.2) $a + b > a'$;
 - b.3) $a + b = a' + b'$;
 - b.4) $a + b = b + a$;
 - b.5) $(a + b) + c = a + (b + c)$
 - b.6) se $a < b$, existe um número n tal que $na > b$.

CAMPBELL, em uma série de trabalhos publicados no período 1920-1938 (FERNÁNDEZ; 1992, p.233), desenvolveu a primeira exposição dos fundamentos lógicos da mensuração. Campbell define mensuração como *a atribuição de numerais para representar propriedades de sistemas materiais, que são números, em virtude de leis que governam essas propriedades* (TORGERSON; 1958, p.13). considera a mensuração como o isomorfismo entre a idéia de magnitude e a quantidade da propriedade a ser medida; demonstrando-se o isomorfismo quando se prova que as quantidades atendem aos axiomas de HÖLDER. Mesmo na mensuração de propriedades físicas dos objetos, isso se torna problemático se, por um lado, os axiomas de HÖLDER podem ser comprovados empiricamente para propriedades como massa, comprimento ou outros, podendo-se ordenar e somar quantidades, por outro lado, muitas outras propriedades físicas não atendem aos axiomas que se

referem à soma de quantidades para se obter uma nova quantidade, como é o caso da densidade e da temperatura. Segundo FERNÁNDEZ (1992; p.234),

*[...] Esse tratamento da teoria da mensuração assume um ponto de vista muito restrito daquilo que se pode considerar mensuração. Havia somente um tipo de escala tolerável; a **quantidade** tinha um conjunto de propriedades e estas eram estritamente não negociáveis. As magnitudes teriam que as cumprir ou não teriam que ser medidas... O sistema, claro está, não oferecia a possibilidade de incluir atributos psicológicos.*

Em 1940, um comitê designado pela Associação Britânica para o Progresso da Ciência concluiu que a mensuração em educação e em psicologia não era da mesma espécie da mensuração em física.

*STEVENS, referindo-se às declarações contraditórias no relatório final do referido comitê, conclui que talvez a concordância possa melhor ser alcançada se reconhecermos que a mensuração existe sob uma variedade de formas e que as escalas de numeração situam-se em classes definidas. Essas classes são determinadas tanto pelas operações empíricas envolvidas no processo de **medir**, quanto pelas propriedades formais (matemáticas) das escalas. (KEEVES; 1997, p.713).*

Eliminando a restrição de que as medidas teriam que obedecer necessariamente aos axiomas de quantidade, STEVENS, em 1946, traz uma inovação fundamental à teoria da mensuração, fundamentado na possibilidade de se medirem propriedades cuja mensuração não seria possível na visão de CAMPBELL.

STEVENS (1967), definindo mensuração como *a atribuição de numerais a objetos ou eventos de acordo com leis*, e tomando os postulados de ordem de HUNTGTON¹ como capazes de garantir o isomorfismo entre idéias de magnitude e a de quantidade, distingue os diferentes tipos de atribuição (tipos de escala), classificando os níveis de medida de acordo com a regra de atribuição de numerais que torna invariante a forma da escala.

¹ Esses postulados são: sejam as magnitudes **a**, **b** e **c**.

a) Se **a** \neq **b**, então, ou **a** > **b** ou **a** < **b**.

b) Se **a** < **b**, então **a** \neq **b**.

c) Se **a** < **b** e **b** < **c**, então **a** < **c**.

O problema então, é tornar explícitas: a) as várias regras para a atribuição de numerais; b) as propriedades matemáticas (ou estrutura do grupo) das escalas resultantes, e c) as operações estatísticas aplicáveis às mensurações feitas com cada tipo de escala... As escalas são possíveis em primeiro lugar, porque há um certo isomorfismo entre o que nós podemos fazer com os aspectos dos objetos e as propriedades das séries de numerais. (STEVENS, Op. Cit., p.4-5).

A eliminação da restrição de CAMPBELL permitiu garantir a existência de quatro classes possíveis de atribuição de numerais, isto é, quatro tipos de escala: nominal, ordinal, intervalar e de razão.

Segundo STEVENS (1967), a escala de medida determina que tipo de técnica estatística é aplicável às medidas que a mesma produz.

O quadro a seguir resume a classificação de STEVENS (1967):

Quadro Nº 2 – Classificação das Escalas de Medidas de Stevens

TIPO DE ESCALA	OPERAÇÕES EMPÍRICAS BÁSICAS (*)	ESTRUTURA MATEMÁTICA DO GRUPO (**)	ESTATÍSTICAS PERMISSÍVEIS
1 – Nominal	Determinação de igualdade	Grupo permutativo $x' = f(x)$ significa qualquer substituição uma a uma	Número de casos Moda-Coeficiente de contingência
2 – Ordinal	Determinação de maior ou menor	Grupo isotônico $x' = f(x)$, onde $f(x)$ é qualquer função monótona crescente	Mediana Percentis Correlação ordinal
3 – Intervalar	Determinação de igualdade de intervalo	Grupo linear $X' = ax + b, a > 0$	Média aritmética Desvio padrão Correlação produto-momento
4 – Proporcional (ou de Razão)	Determinação de igualdade de proporções	Grupo de similaridade $x' = cx, c > 0$	Coeficiente de variação

(*) Operações básicas para a criação de uma escala.

(**) Transformações matemáticas que deixam invariante a forma da escala.

As restrições ao emprego de técnicas estatísticas conforme o tipo de escala levantou uma polêmica entre especialistas em psicometria. Argüia-se se era possível, com dados obtidos com escalas ordinais, o uso de técnicas estatísticas que requeriam dados obtidos com escalas intervalares ou com escalas proporcionais. Em oposição à visão de STEVENS, uma variedade de psicometristas

[...] por exemplo, BURKE e ANDERSON, entre outros, argüiram que MCNEMAR e HAYS – **apud** KEEVES (1997) mostrou que as estatísticas se aplicam a números e não a coisas e que as propriedades

formais das escalas de medidas, como tais, não teriam influência na escolha das estatísticas (BAKER, HARDYCK e PETRINOVICH; 1967, p.17).

Os defensores da posição de STEVENS ficaram conhecidos como teóricos da escala de mensuração fraca (weak measurement), em torno da questão, o debate estendeu-se por toda a década de 50 e ocorre até hoje. A posição de BURKE (In: MUÑIZ, 1991) é reveladora do lado dos teóricos da estatística robusta (strong measurement). Demonstrando que as propriedades de uma escala de mensuração envolvem correspondências entre conjuntos de axiomas sobre objetos e números, com relações e operações apropriadas e que os métodos estatísticos começam e terminam com números, deduz, dessas duas proposições, que as propriedades do conjunto de números de uma escala de medida não têm nenhum efeito sobre a escolha das técnicas estatísticas para representar e interpretar as medidas fornecidas pela escala.

BAKER, HARDYCK e PETRINOVICH (1967, p.18-19), considerando que, para STEVENS e seguidores, testes paramétricos de significância tais como os testes **t de Student** e **F de Snedecor**, somente podem ser usados como medidas obtidas com escalas nas quais transformações lineares são admissíveis (escalas intervalares e de razão) vão buscar empiricamente resposta à questão se é possível *tornar decisões corretas sobre a natureza da realidade e negligenciar da escala de medida quando se aplicam testes estatísticos*. O debate é deslocado para o campo empírico quando, até então, desenvolvia-se somente no campo teórico.

Visando responder empiricamente à questão que, na verdade, refere-se à possibilidade ou não do uso de testes estatísticos paramétricos, com dados obtidos com escalas ordinais quando, na visão de Stevens, a utilização desses testes supõe, pelo menos, o nível intervalar de mensuração, esses autores, a partir de uma distribuição de escores intervalares construíram, através de transformações não lineares, uma série de distribuições com dados ordinais e geraram, em computador, amostras aleatórias com esses dados transformados. Observaram que o uso de medidas ordinais, não afeta seriamente os valores observados de **t** para os níveis de significância de **5%** e **1%** quando comparados com valores de **t** referentes às medidas intervalares

correspondentes e que as restrições de STEVENS (1967) somente são válidas para as medidas da estatística descritiva.

Esta visão pode ser correta se consideram determinações específicas únicas de uma estatística no sentido descritivo (...) mas, é incorreta quando aplicada ao problema da inferência estatística (...) Os dados presentes indicam que estatísticas robustas como o teste t são mais que adequados para serem usados com medidas fracas e, com algumas reservas menores, as possibilidades estimadas a partir da distribuição t são pouca afetadas pelo tipo de escala de medida usada. (Op. Cit., p.23)

Os autores sugerem que, no caso de distribuições assimétricas e amostras com tamanhos diferentes, deve-se usar somente teste bicaudal.

TORGERSON (1958) refuta a definição de mensuração de STEVENS (atribuição de numerais a objetos ou eventos) afirmando que são as propriedades dos objetos que são passíveis de serem medidas:

... a mensuração pertence a propriedades dos objetos, e não aos objetos. Então um bastão de madeira não é mensurável em seu uso do termo, embora seu comprimento, peso, diâmetro e dureza possam sê-lo. (TORGERSON; 1958, p.14).

Para TORGERSON (*op cit*), o problema da mensuração refere-se ao processo e à racionalidade envolvidas na construção de uma escala e às propriedades que lhe podem ser atribuídas. Considerando o tipo de representação numérica da escala, esse autor classifica as escalas de medida de acordo com as características dos números, capazes de garantir o isomorfismo entre as magnitudes e as quantidades.

Essas características são:

- a) de **ordem** – se $a \# b$, então $a > b$ ou $a < b$;
- b) de **distância** – se $a \# b \# c \# d$, então $(a - b) > (c - d)$ ou $(a - b) < (c - d)$;
- c) de **origem** – a escala possui ou não uma origem única indicada por **zero**.

Qualquer escala apresenta a característica de ordem e pode, também, apresentar, ou uma das duas outras, ou ambas. Assim, a classificação de Torgerson é feita em função das características de distância e de origem, em uma escala.

- 1) se $a \neq b \neq c$, pode-se ou não determinar se $(b - a) > (c - b)$ ou $(b - a) = (c - b)$, ou $(b - a) < (c - b)$;
- 2) ou o **zero** indica ausência do atributo (a escala possui origem natural), ou o **zero** é relativo, não indicando ausência do atributo (a escala possui origem não natural), ou não apresenta **zero**.

O quadro a seguir resume a classificação de TORGERSON (1958):

Quadro Nº 3 – Classificação das Escalas de Medidas de Torgerson

CARACTERÍSTICAS DE DISTÂNCIA	CARACTERÍSTICAS DE ORIGEM	
	NÃO NATURAL	NATURAL
NÃO DETERMINÁVEL	ESCALA ORDINAL	ESCALA ORDINAL COM ORIGEM NATURAL
DETERMINÁVEL	ESCALA INTERVALAR	ESCALA DE RAZÃO

TORGERSON (1958) não considera as escalas do tipo nominal da classificação de Stevens como medida, pois, de acordo com sua definição de mensuração, os valores fornecidos por uma escala nominal não são números que traduzem quantidades, os numerais utilizados são simplesmente símbolos indicativos de categorias e permitem somente a classificação de objetos em função de propriedades não mensuráveis. As escalas dos tipos ordinal, intervalar e de razão são as mesmas da classificação de STEVENS. Quanto às escalas ordinais com origem natural, elas constituem um tipo escala ordinal com um zero absoluto, indicando ausência no objeto da propriedade sob mensuração.

Na mensuração de atributos como atitudes, estética, preferência e valores, a origem natural ocorre dentro das séries e pode ser descrita com um ponto neutro onde todos os estímulos ou indivíduos em uma direção são favoráveis, agradáveis, apreciados ou desejados, conforme o caso, enquanto que todos aqueles no outro lado são desfavoráveis, desagradáveis, menosprezíveis ou indesejáveis (...) para esses atributos (...) os números atribuídos aos elementos em um lado da origem natural possuem sinal negativo e aqueles no outro lado têm sinal positivo (Op cit p.30).

Tanto na classificação de TORGERSON (1958) quanto na de STEVENS (1967), os quatro tipos de escala restringem, em grupos diferentes, a atribuição de números. A restrição à atribuição de números implica na quantidade de informação contida na medida obtida por um tipo de escala. Assim, a medida de uma propriedade obtida com uma escala ordinal possui menos informação que a medida da mesma propriedade obtida por uma escala intervalar. Esta, por sua vez, contém menos informação que a medida fornecida por uma escala de razão. As escalas de razão, por atenderem a todos os axiomas de HÖLDER, permitem o nível mais elevado de mensuração e, conseqüentemente, maior sobre a propriedade medida no objeto. TORGERSON (1958) exemplifica com a mensuração do raio e da área de conjunto de círculos, visando expressar matematicamente a relação entre essas duas propriedades de círculo. Com as medidas dos raios e áreas tomadas com uma escala ordinal, a área é uma função monótona do raio. *Isto é, se a área ou raio ou ambos são medidas com uma escala ordinal, nós podemos determinar somente que a relação entre os dois é monótona* (Op. Cit. 1958, p.21). Com uma escala ordinal com origem natural, a área é uma função monótona do raio a partir da origem; se o raio é igual a **zero** a área é igual a **zero**. Com uma escala de razão, a área é proporcional ao quadrado do raio. Neste caso, é possível verificar a existência de uma constante de proporcionalidade, em todos os círculos, a razão entre a área e o quadrado do raio é igual ao número π .

Enquanto CAMPBELL (1930), [apud MUÑIZ 1992] considerava as relações empíricas como depositárias de toda informação relevante, (as quantidades deveriam atender aos axiomas de HÖLDER), TORGERSON (1958) e STEVENS (1967) deslocam a ênfase para as propriedades da escala: as magnitudes devem atender aos axiomas de ordem de HUNTINGTON, no caso de STEVENS (1967) ou às características da representação numérica capazes de garantir o isomorfismo entre as magnitudes e quantidades em TORGERSON (1958).

Independente de a ênfase estar no sistema empírico ou no formal, esses autores argumentam como condição necessária à mensuração, a relação

axiomática entre os dois sistemas. Ocorre aí uma inversão de prioridade, pois, a identificação da natureza das relações entre magnitudes da escala e as quantidades (entidades empíricas) não é uma consequência da escolha de um tipo de escala para se realizar a mensuração de uma propriedade e sim um pré-requisito.

LORD e NOVICH (1968, p.17), ao desenvolverem a generalização do modelo clássico de mensuração rompem com a concepção idealista platônica *um procedimento para a atribuição de números (scores, medidas) a propriedades específicas de unidades experimentais de modo a caracterizar e preservar relações do domínio empírico*, fundamentam-se numa visão realista.

Os autores acrescentam à classificação de Stevens em quinto nível de mensuração, constituído pela escala absoluta que apresenta as características das escalas de razão e, adicionalmente, a unidade de mensuração é fixa. Qualquer operação de contagem produz medida absoluta. Ao atributo sob mensuração são associados números de modo que todas as propriedades dos números refletem propriedades do atributo. Nesse caso, nenhuma transformação é permissível, qualquer transformação destrói alguma propriedade da escala.

Em relação às escalas nominais, é feita uma restrição; somente as escalas dicotômicas produzem medida. Escalas nominais com três ou mais categorias são simplesmente escalas classificatórias, podendo, no entanto, ser dicotomizadas, tornando-se uma das categorias como referência.

2.3 TEORIA CLÁSSICA DOS TESTES (TCT): O MODELO LINEAR CLÁSSICO DE MENSURAÇÃO

A utilização de testes na mensuração em Psicologia e em Educação impôs a necessidade de um modelo que permitisse estimar, através de técnicas estatísticas, o erro da medida, uma vez que este não é observado diretamente.

O modelo linear clássico, desenvolvido por Spearman no início do século XX, tem por objetivo central a estimação do erro cometido quando se utiliza um

teste para medir variáveis psicológicas, no entanto, o modelo pode ser estendido a situações de mensuração com os mais variados tipos de instrumento de medida.

O modelo linear clássico estabelece que a medida (**X**), atribuída por meio de instrumento de medida a um sujeito em uma dada propriedade, é igual à soma de dois componentes: a medida verdadeira (**X_v**) do sujeito nessa propriedade (**X_v** é desconhecida) e o erro da medida (**e**) cometido no processo de mensuração², isto é, **X = X_v + e**.

Na descrição do modelo acima, o termo **atribuída** é um termo chave para se compreender a fundamentação platônica do conceito de medida. O conceito de medida verdadeira era definido sintaticamente (isto é no mundo real) como a medida obtida através de um teste com um infinito de itens. Nesse teste ideal, os itens abarcariam, em seu conjunto, todos os estímulos possíveis de gerar todas as possíveis medidas. A propriedade por ele medida poderia se apresentar, através da medida, em qualquer ponto de uma escala contínua.

O fato da medida verdadeira de uma propriedade não poder ser observada diretamente cria um problema de ordem epistemológica gerando os três pontos de vista sobre o conceito de medida verdadeira encontrados na literatura sobre a importância prática e teórica desse conceito. Segundo LOEVINGER (*apud* LORD e NOVICH; 1968, p.27-28), a medida verdadeira, não sendo diretamente observável, o conceito é desprovido de importância prática. No segundo ponto de vista, para THORNDICK (*apud* LORD e NOVICH; 1969 p.27), o fato do conceito não ser diretamente observável, reveste-o de um caráter místico e, portanto, não apresenta importância teórica. O terceiro ponto de vista é que o conceito é útil teoricamente e permite a obtenção de resultados práticos.

Esta não é uma posição metafísica, nós não patrocinamos uma teoria da mensuração que postule inúmeras proposições que são incapazes de verificação prática. Nós usamos a noção de escore verdadeiro porque achamos que ela é útil teoricamente e produz implicações que podem ser verificadas na prática efetiva. (Lord e Novick; 1969 p.27-28).

² São fontes do erro: o instrumento, o observador e o objeto sob mensuração.

Em relação ao conceito de medida verdadeira, desenvolveram-se três concepções. A primeira, com fundamentação platônica, define a medida verdadeira a partir de uma situação específica de mensuração (**m**), na qual um sujeito (**s**) apresenta, na propriedade (**p**) sendo medida, uma única quantidade verdadeira (X_v) – a medida verdadeira (**s**) em (**p**). Qualquer mensuração particular de (**p**) em (**s**) na mensuração (**m**) conduz a um valor (**X**) – a medida observada – que, em geral, difere da quantidade X_v devido ao erro de medida. O valor **X** é observável, mas o valor X_v não o é. Na física clássica, essa concepção parece ser razoável. Se a situação de mensuração é especificada com precisão, há especificação dos procedimentos de mensuração e da escala de medida utilizada, é correto falar na velocidade verdadeira da luz, no peso verdadeiro de um objeto, entre outros. Por outro lado, nas ciências humanas, cujas teorias são elaboradas com base em construtos, a concepção platônica de medida verdadeira não é capaz de sustentar uma fundamentação axiomática satisfatória para essa teoria.

A segunda concepção a se desenvolver tem fundamentação empirista, com base na teoria de probabilidades de Von Mises. Ainda é a mais freqüente na literatura sobre mensuração em Psicologia e em Educação. Von Mises, em reação ao subjetivismo e apriorismo da teoria clássica de probabilidades, elaborou uma estruturação da teoria análoga àquela das ciências naturais, dando ao conceito de probabilidade uma fundamentação empírica através do conceito de *coletivo*.

Coletivo constitui uma sucessão infinita de observações que realizam suas condições:

[...] as freqüências relativas dos atributos particulares de elementos singulares do coletivo tendem para um limite definido; e (...) este limite não é afetado por seleções localizadas (princípio do azar). O valor limite da freqüência relativa de um atributo dado, admitindo que é independente de qualquer seleção localizada, será chamada 'a probabilidade' deste atributo; dentro do coletivo dado, só é possível falar de probabilidade com respeito a um coletivo definido com propriedade. (TORANZOS; 1969, p.22).

TORANZOS (1962) apresenta as três objeções mais freqüentes na literatura à teoria de VON MISES: 1) a propriedade como valor limite para o

número de provas tendendo para o infinito não fundamenta empiricamente o conceito, pois, um experimento com as n provas tendendo para o infinito não é realizável, uma vez que não é possível prolongar idealmente o experimento porque os processos regidos pelo acaso, se prolongados racionalmente, perdem o caráter aleatório, que é sua essência; 2) a segunda condição do conceito de coletivo postula a aleatoriedade, introduzindo um elemento subjetivo que rompe, dessa forma, com o fundamento empirista e 3) a supressão dos conceitos de probabilidade de um caso isolado e de probabilidade **a priori** restringe o cálculo de probabilidade; esses conceitos têm se revelado muito fecundo em aplicações.

A concepção empirista de medida verdadeira entende-a como valor limite se, em uma dada situação (**m**) de mensuração, é feita, no sujeito (**s**), uma série sucessiva de mensurações da propriedade (**p**), a média das medidas X_i ($i = 1, 2, 3, n; n \rightarrow \infty$) de (**p**) em (**s**), com probabilidade **1**, converge para um valor constante; esse valor constante é a medida verdadeira de (**s**) em (**p**).

A terceira posição, desenvolvida por LORD e NOVICH (1968) como base para a generalização da teoria clássica, fundamenta-se na teoria axiomática intuitiva de probabilidades, desenvolvida por Frechet, Cramer e outros (TORANZOS, 1969).

A teoria axiomática intuitiva de probabilidades apresenta uma fundamentação axiomática própria da Matemática, contendo ao lado dos axiomas formais, outros de caráter existencial. Segundo TORANZOS (1969), a adoção de suposições que, apesar de não se harmonizarem com a perfeição e a rigidez axiomática, como, por exemplo, o conceito de aleatoriedade como idéia primitiva, permite fundamentar uma teoria capaz de aplicações à realidade. A teoria também parte, como a de Von Mises, da noção de freqüência, diferindo, no entanto, por evitar a passagem ao limite com o número de mensurações tendendo ao infinito na situação de mensuração.

TORANZOS (1969) reproduz a definição de probabilidade de FRECHET e verifica-se que, numa determinada categoria de provas, quando se calculam as freqüências de acontecimento fortuito em diferentes grupos de provas, essas freqüências diferem pouco, praticamente, quando cada um dos grupos

se constitui de numerosas provas. As freqüências de um acontecimento fortuito **E** nos grupos, cada um composto de numerosas provas, todas pertencentes a uma mesma categoria **C**, são valores experimentais de uma constante física, determinada pela natureza do acontecimento **E**, e pela categoria **C**. É uma constante física que chamamos: a probabilidade do acontecimento **E** na categoria de provas **C**.

Em outras palavras, tomando-se uma seqüência constituída por **K** experimentos aleatórios idênticos, cada um deles com um número grande de provas, observa-se que as freqüências relativas de um evento, com ocorrência possível em cada prova, distribuem-se em torno de um valor constante; a esse valor denomina-se probabilidade do evento **E**.

Segundo LORD e NOVICH (1968), os modelos sustentados por processos de mensuração que descrevem o comportamento de variáveis psicológicas e, por extensão educacionais, são modelos probabilísticos devido ao problema do erro da medida. Assim, esses modelos podem ser representados sistematicamente por:

$$x = \varphi(\theta_i) + e$$

Onde **x** é uma variável dependente observável, θ_i é um vetor de variáveis independentes, φ é uma função conhecida (não necessariamente linear) que relaciona **x**, θ e **e**, o erro, um conjunto de efeitos não associados a θ .

Tomando-se o modelo linear

$$X = X_V + e$$

—
a medida verdadeira de X_V é definida como a esperança matemática de uma série de valores observados de uma variável. A série pode ser constituída por medidas da variável tomadas em mensurações repetidas, em uma única vez, em cada sujeito de uma amostra.

O modelo generalizado tem por base três postulados:

- 1) a medida verdadeira (X_V) é a esperança matemática de uma série de medidas empíricas:

$$X_V = E(x)$$

- 2) Não existe correlação entre as medidas verdadeiras e seus respectivos erros de medida: $\rho(X_V; e) = 0$.

Não há razão plausível para se pensar que o tamanho do erro de medida esteja associado sistematicamente ao tamanho da medida verdadeira. O erro se distribui aleatoriamente.

- 3) Definindo-se instrumentos paralelos como dois ou mais instrumentos que apresentam a mesma variância dos erros de medidas e as mesmas medidas verdadeiras (pode-se dizer, em linguagem comum, que dois ou mais instrumentos são paralelos se medem a mesma propriedade, apesar de variarem na forma), postula-se que os erros de medida em um instrumento (**J**) não se correlacionam com seus erros de medida em um instrumento paralelo (**K**), $\rho(e_j; e_k) = 0$.

Se a mensuração é feita corretamente com os instrumentos paralelos, não existe razão para que as medidas covariem sistematicamente umas com as outras.

O modelo permite deduções imediatas de importância teórica e prática. Essas deduções estão formuladas para os valores paramétricos da população.

- 1) O erro de medida se define como a diferença entre a medida empírica e a medida verdadeira:

$$E = X - X_V$$

- 2) A esperança matemática dos erros de medida é **zero**:

$$E(x) = 0$$

- 3) A média das medidas empíricas (μ_x) é igual à média das medidas verdadeiras:

$$\mu_x = \mu_{x_v}$$

- 4) As medidas verdadeiras não covariam com os erros de medidas:

$$Cov(X_V; e) = 0$$

- 5) A covariância entre as medidas verdadeiras e as medidas empíricas é igual à variância das medidas verdadeiras:

$$\mathbf{Cov(X; X_v) = Var(X_v)}$$

- 6) A covariância entre as medidas empíricas de dois testes é igual à covariância entre as medidas verdadeiras:

$$\mathbf{Cov(X_1; X_2) = Cov(X_{v1}; X_{v2})}$$

- 7) A variância das medidas empíricas é igual à variância das medidas verdadeiras mais a variância do erro de medida:

$$\mathbf{Var(X) = Var(X_v) + Var(e)}$$

- 8) A correlação entre as medidas empíricas e os erros de medida é igual ao coeficiente entre o desvio padrão dos erros e o desvio padrão das medidas empíricas:

$$\mathbf{\rho(X;e) = \sigma_e / \sigma_x}$$

para **K** testes paralelos.

- 9) As médias das medidas empíricas são iguais:

$$\mathbf{\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k}$$

- 10) As variâncias das medidas empíricas são iguais:

$$\mathbf{\sigma^2(X_1) = \sigma^2(X_2) = \dots = \sigma^2(X_k)}$$

- 11) As inter-correlações entre as medidas empíricas são iguais:

$$\mathbf{\rho(X1; X2) = \rho(X1; X3) = \dots = \rho(X1; Xk)}$$

SANTISTEBAN (1990, p.26) acentua a importância atual do modelo linear. Diz ela:

A introdução deste modelo permitiu que se desenvolvesse uma das teorias mais profícuas no campo da Psicometria.(...) Ainda que posteriormente tenham sido desenvolvidas novas teorias, a teoria

baseada no modelo de SPEARMAN continua sendo influente em nosso tempo ...

2.4 TEORIA DA RESPOSTA AO ITEM (TRI)

As origens da TRI vêm dos trabalhos pioneiros de RICHARDSON (1936), LAWLEY (1944), TUCKER (1946), LORD (1952 e 1953) e BIRNBAUM (1957 e 1968), ampliando-se sobretudo com as contribuições de BIRNBAUM, LORD e NOVICK (1968). A elaboração de programas de computadores para análises com a TRI incrementou ainda mais a utilização desta teoria, principalmente por facilitar os cálculos que seriam muito complexos se fossem realizados manualmente. Conquistou inúmeros adeptos nas áreas de psicologia e educação devido ao fato de *“oferecer recursos mais flexíveis e eficazes na confecção, análise e apresentação dos resultados de prova que quaisquer outros recursos equivalentes, derivados a partir da teoria clássica”* (FLETCHER, 1994; p.21).

O mesmo autor afirma que:

[...] na relação que estabelece entre as características dos itens e as características operacionais da prova composta desses itens. Em sua essência, esta relação é invariável e permanente, não dependendo do número de itens da prova, do subconjunto de itens selecionados ou das habilidades das pessoas que participam da prova. Quaisquer outros resultados comparáveis advindos da teoria clássica servem apenas para testes hipotéticos de uma infinidade de itens aplicados a uma única população padrão (Op Cit p.23).

A medida feita através da TRI parte da suposição de que existe no sujeito um traço (uma característica individual que determina sua forma de responder ao teste) que possui uma relação probabilística com cada um dos itens utilizados. Os parâmetros de cada item não dependem, em absoluto, dos outros itens do teste. Ao contrário, a pontuação do teste se faz em função das respostas do sujeito a cada item e dos seus respectivos parâmetros. Colocam-se, assim, itens e pessoas na mesma escala de desempenho. Com isso, pode-se afirmar se as pessoas são mais ou menos hábeis e se os itens são mais ou menos difíceis na mesma escala de uma característica latente subjacente. O

modelo de resposta que fundamenta a teoria da resposta ao item permite ao analista prever a probabilidade de acerto de uma pessoa com determinada habilidade representada pelo resultado da prova (FLETCHER, 1994; p.23)

A relação entre os valores da variável que é medida pelo item e a probabilidade de acertá-lo é descrita por uma função matemática denominada Curva Característica do Item (CCI), que relaciona a probabilidade de êxito em um item com a capacidade medida pelo conjunto de itens que o teste contém (Gaviria, 2001: p.7). São as propriedades desta função que fazem a diferença entre a TRI e TCT.

2.4.1 PRESSUPOSTOS DA TRI

1º Unidimensionalidade

Ao elaborar um teste, é definido, previamente, o traço ou traços que se deseja avaliar e se supõe que estes possuem as dimensões necessárias para descrever a característica estudada no indivíduo. Segundo SANTISTEBAN (1990), cada indivíduo examinado será representado por um vetor no qual cada um dos componentes indica o nível do indivíduo nesse componente ou dimensão. Espera-se que os sujeitos que têm o mesmo nível no traço tenham as mesmas pontuações no teste.

Geralmente determina-se que um só traço θ é necessário para explicar ou dar conta da atuação do indivíduo no teste. Assim, geram-se os modelos unidimensionais. Para comprovação da **unidimensionalidade**, o método mais utilizado é a Análise Fatorial.

2º Independência local dos itens

Se for comprovada a **unidimensionalidade** do item, por conseqüência, haverá independência local entre as respostas dadas pelos sujeitos e entre os itens. De acordo com SANTISTEBAN (1990), a probabilidade de que um examinando responda corretamente a um item do teste não vem afetada pelas

respostas dadas a outros itens, sempre e quando se considerem estas pontuações condicionadas a um certo dado.

A independência local entre itens pode ser constatada através da determinação de que a probabilidade de um sujeito acertar n itens é igual ao produto das probabilidades de acertar cada um deles (Fernandez, 1990).

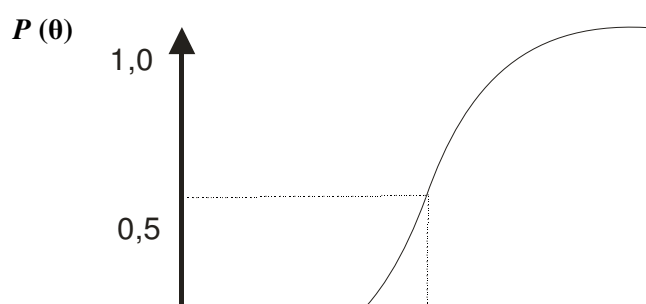
A independência local entre sujeitos é constatada comprovando-se que o rendimento de um sujeito que responde a um teste é independente do rendimento dos outros. Segundo Gaviria (1995), a pontuação de cada sujeito é caracterizada por seu nível no traço latente. Assim, a probabilidade de uma resposta correta de um examinando não dependerá de quantos outros examinandos estejam colocados no mesmo nível do traço ou em níveis diferentes.

2.4.2. CURVA CARACTERÍSTICA DO ITEM (CCI)

A CCI caracteriza o item, pois representa seus parâmetros típicos. Apresenta-se em três modelos diferentes, sendo utilizados de acordo com os dados disponíveis. Nos três modelos de CCI podem ser observados parâmetros que caracterizam as qualidades técnicas dos itens, independentes da população investigada. Esses parâmetros representam a dificuldade, o poder discriminante e a probabilidade de acerto casual ao item.

A CCI utilizada na TRI é do tipo "S", que representa o modelo idealizado. No eixo das abscissas está indicado o nível do sujeito na variável observada (traço latente) designada pela letra grega (θ), variando de $-\infty$ a $+\infty$. A probabilidade de responder corretamente o item, $P(\theta)$, está indicada no eixo das ordenadas e varia de 0 a 1. A curva desenvolvida na função demonstra a probabilidade de acerto do item para distintos valores de (θ).

Gráfico 3: Exemplo de uma CCI de um item



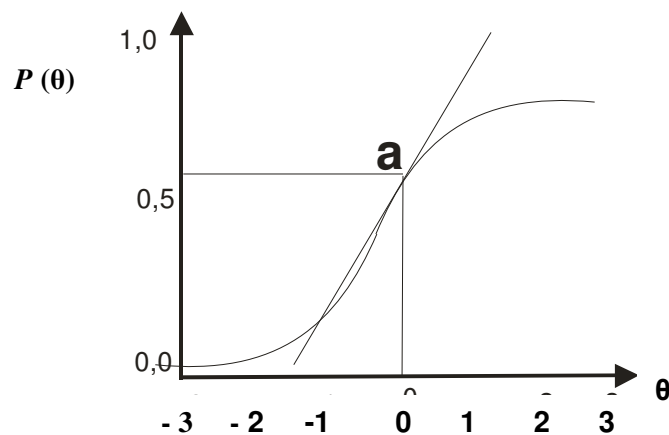
- Parâmetros apresentados na CCI

A CCI representa graficamente os três parâmetros idealizados pela TRI para um item ideal: a discriminação, a dificuldade e a probabilidade de acerto ao acaso.

1. Parâmetro "a" (discriminação): Segundo Fernandez (1990, p.21) este parâmetro é determinado pelo ponto máximo da inclinação da reta; seu valor é proporcional a esta pendência e, quanto maior a inclinação, maior será o índice de discriminação.

O gráfico 4 mostra a localização do parâmetro "a" numa CCI.

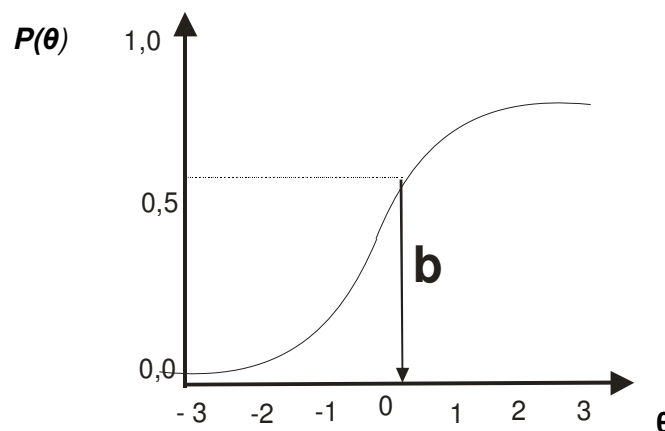
Gráfico 4: Exemplo de localização do parâmetro "a" numa CCI



2. Parâmetro "b" (dificuldade): é um parâmetro de posição do item na escala de (θ), no eixo das abscissas. Informa em que parte da escala de (θ) se encontra o ponto de inclinação máxima da curva. A probabilidade de responder corretamente ao item varia de 0 a 1. Se o valor for alto, dizemos que o item é difícil e se é baixo, então, o item é fácil.

O parâmetro de dificuldade do item é, portanto, o valor que informa em qual parte da escala de (θ) se dá a transição desde uma maior probabilidade de responder incorretamente ao item a uma maior probabilidade de responder corretamente (Gaviria, 1995; p.8-9). O gráfico 5 é um exemplo de onde o parâmetro "b" pode se localizar numa CCI.

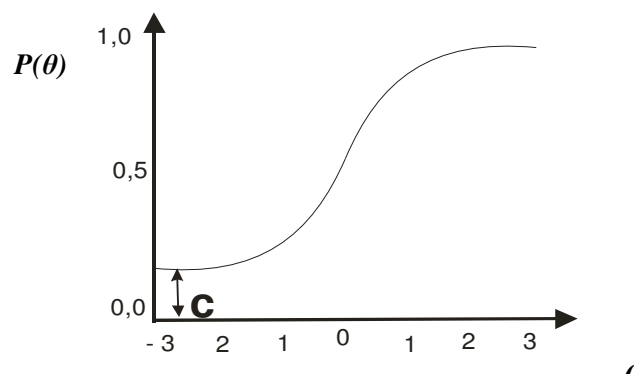
Gráfico 5: Exemplo de localização do parâmetro "b" numa CCI



3. Parâmetro "c" (acerto ao acaso): representa a probabilidade de acertar o item ao acaso, "quando não se tem certeza da resposta certa" Gaviria (1995, p.16) citando Lord diz que os examinandos, na incerteza, buscam indícios indiretos que poderiam orientá-los na localização da opção correta. O parâmetro "c" é o valor da $P(\theta)$, quando $(\theta) = -\infty$, ou ainda, é o valor assintótico da CCI quando θ tende a $-\infty$.

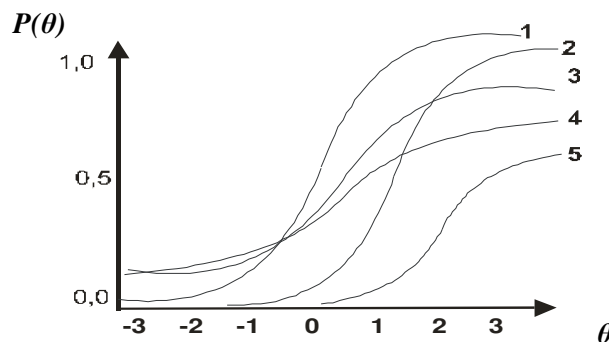
No gráfico 6 é demonstrado um exemplo de localização do parâmetro "c" numa CCI.

Gráfico 6: Exemplo de localização do parâmetro "c" numa CCI.



Para ilustrar o exposto são a seguir representados graficamente (gráfico 7) curvas de cinco itens diferentes para que sejam observadas as diversas localizações dos parâmetros descritos.

Gráfico 7: Exemplo de curvas de diversos itens com diferentes valores dos parâmetros.



Analisando os Itens representados na CCI acima, observa-se que:

- o item 1 é o mais fácil, visto que foi respondido corretamente por Indivíduos que estão a menos de 1 desvio da média dos respondentes, possui também alto poder discriminativo para os valores de θ a partir de -1 e o parâmetro de acerto ao acaso é zero (0);
- o item 2 é de média dificuldade, possui alto poder discriminativo (observar a inclinação da curva) para os valores de 0 em torno de 0,50 e os parâmetros de acerto ao acaso é zero (0);
- os itens 3 e 4 são difíceis, poucos discriminativos e com valores do parâmetro de acerto ao acaso em torno de 0,10; merecendo uma análise mais detalhada;
- o item 5 é um item difícil, com baixo poder discriminativo e o valor do parâmetro de acerto ao acaso é zero (0).

2.2. Modelos da curva característica dos itens Os modelos da CCI foram produzidos a partir dos três parâmetros dos itens (dificuldade, discriminação e acerto ao acaso). sendo descritos através de funções logísticas

a) Modelo logístico de um parâmetro ou modelo de Rasch: proposto originalmente pelo matemático GEORGE RASCH. Segundo FERNANDEZ (1990) é o modelo mais utilizado em virtude da sua simplicidade matemática. Pressupõe que a resposta a um item só depende da competência do sujeito e da dificuldade do item. O único parâmetro do item a ser observado neste modelo é **b**, a dificuldade, sendo expressa matematicamente da seguinte forma:

$$P_i(\theta) = \frac{e^{D(\theta + b_i)}}{1 + e^{D(\theta + b_i)}}$$

$P_i(\theta)$ = probabilidade de acertar o item num determinado nível θ .

θ = Valores da variável medida.

b_i = índice de dificuldade do item i .

e = Base dos logaritmos neperianos (2,72)

D = Constante³

b) Modelo logístico de dois parâmetros: BIRNBAUM (1957 e 1968) foi o primeiro a propô-lo, sendo anterior ao de RASCH. Neste modelo. trabalha-se com dois parâmetros do item o **índice de dificuldade “b”** e o **índice de discriminação “a”**. Sua fórmula é dada por:

$$P_i(\Theta) = \frac{e^{Da_i(\Theta + b_i)}}{1 + e^{Da_i(\Theta + b_i)}}$$

³ A constante **D** faz com que a curva logística coincida com a curva normal, onde **a = 1**, com uma margem de erro de **0,01**. Logo, este modelo supõe que todos os itens têm a mesma pendente, portanto, o mesmo **índice de discriminação**, sendo adotado o valor **1** (GAVIRIA, 1995; p.13).

$P_i(\theta)$ = Probabilidade de acertar o item i num determinado nível θ .

θ = Valores da variável medida.

b_i = Índice de dificuldade do item i .

a = Índice de discriminação do item i .

e = Base dos logaritmos neperianos (2,72).

Da_i = Constante de valor de 1,7; para aproximação da normal acumulada (FERNANDEZ, 1990; p.32).

c) **Modelo logístico de três parâmetros:** originado também dos estudos de BIRNBAUM. Nesse modelo, estão incluídos os parâmetros a , b e c . O parâmetro c , relativo aos acertos ao acaso é, nesse modelo, descrito tecnicamente como o valor de $P(\theta)$, para um valor de $\theta = -\infty$. A equação referente ao modelo é:

$$P_i(\theta) \cdot c_i + (1 - c) = \frac{e^{Da_i(\theta + b_i)}}{1 + e^{Da_i(\theta + b_i)}}$$

$P_i(\theta)$ = Probabilidade de acertar o item i num determinado nível θ .

θ = Valor da variável medida.

b_i = Índice de dificuldade do Item i .

a_i = Índice de discriminação do item i .

c_i = índice de acertos ao acaso do item i .

e = Base dos logaritmos neperianos (2,72).

D = Constante (1,7).

Para sintetizar as descrições e informações até aqui apresentadas pode-se afirmar que a TRI pretende:

- resolver o problema da dependência da medida em relação ao instrumento utilizado e à variância das propriedades deste em função dos objetivos medidos;
- obter medições que sejam invariáveis quanto aos testes;
- dotar os instrumentos de parâmetros invariáveis quanto aos sujeitos.

Segundo Lord (1980), a teoria da resposta ao item não contradiz a teoria clássica dos testes uma vez que ela acrescenta pressupostos adicionais que permitem resolver situações que a teoria clássica não soluciona. A TRI constitui-se num marco importante na mensuração em psicologia e educação e apesar do seu carácter complementar à TCT, sua formulação pouco utiliza da formulação dessa teoria.

Para ANDRADE, TAVARES & VALLE (2000, p.35) com referência ao processo de equalização da TRI, o caminho lógico é tornar comparável, o que no caso da TRI, significa colocar parâmetros de itens vindo de provas distintas ou habilidades de respondentes de diferentes núcleos, na mesma meta, isto é, numa escala comum, tornando os e/ou as habilidades comparáveis.

Além de permitir trabalhar os resultados de um único grupo, fazendo um único teste, a TRI garante, através do processo de equalização, traduzir em uma única variável nas situações a seguir:

- um único grupo, dividido em dois ou mais subgrupos, com cada subgrupo fazendo um teste sem qualquer item comuns aos testes dos outros subgrupos.
- um único grupo, dividido em dois ou mais subgrupos, com cada subgrupo fazendo dois ou mais testes com alguns itens comum aos testes dos demais subgrupos.
- Dois ou mais grupos, fazendo um mesmo teste.
- Dois ou mais grupos, cada um fazendo um teste, sem nenhum item comum entre eles.
- Dois ou mais grupos, fazendo cada um teste com alguns itens comuns com os demais grupos.

O objetivo central da TRI, como se percebe, é obter medidas que sejam invariantes em relação aos instrumentos empregados e permitir a construção

de instrumentos de medida cujas propriedades não dependam dos objetos medidos e sejam invariantes com respeito aos sujeitos avaliados, sendo que sua grande contribuição à teoria da mensuração se localiza na possibilidade de obter medidas invariantes em relação aos instrumentos utilizados e aos sujeitos envolvidos.

3 O Sistema Nacional de Avaliação da Educação Básica-SAEB

O Sistema de Avaliação da Educação Básica-SAEB foi criado em 1988. Trata-se de um sistema de avaliação em larga escala visando coletar dados sobre alunos, professores e diretores de escolas públicas e privadas do Brasil, para oferecerem subsídios à formulação e monitoramento de políticas públicas na educação. É uma ação do governo federal sob a responsabilidade do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira–Inep, com a participação das Secretarias Estaduais e Municipais de Educação.

Já foram realizados oito levantamentos (1990, 1993, 1995, 1997, 1999, 2001, 2003 e 2005). Os resultados preliminares do SAEB foram divulgados em fevereiro de 2007, estando ainda, para este ano, prevista a nona avaliação.

A partir da Portaria nº 931/MEC, de 21 de março de 2005, o SAEB passou a ser constituído por dois processos, a Avaliação Nacional da Educação Básica-ANEB e a Avaliação Nacional do Rendimento Escolar-ANRESC. A avaliação da ANEB é divulgada como o nome de SAEB, mantendo as mesmas características das avaliações anteriormente realizadas pelo Inep. Já a ANRESC é mais extensa e detalhada, sendo realizada de forma censitária em cada unidade escolar do País, e é divulgada sob o nome de PROVA BRASIL. Foi realizada pela primeira vez em 2005.

Em 1995, foram introduzidas inovações metodológicas: a Teoria da Resposta ao Item-TRI, para a análise dos resultados dos testes aplicados pelo SAEB, o modelo de amostragem matricial de itens e os questionários contextuais. Estas inovações permitiram tornar comparáveis os resultados a partir daquele ano.

Em 1997, foram incorporadas as matrizes de referência para a construção dos itens dos testes, tendo como ponto de partida os conteúdos comuns às propostas curriculares dos estados e da consulta a professores das capitais estaduais de Língua Portuguesa e Matemática, atuando nas redes de ensino estadual e privada nas 4ª e 8ª séries do ensino fundamental e na 3ª série do ensino médio. Foram examinados também os livros didáticos mais utilizados nessas séries. Os gráficos a seguir mostram o desempenho dos alunos avaliados desde a edição de 1995 do SAEB (ano da introdução das modificações metodológicas que permitiram comparações entre as avaliações

a cada ano de realização), nas 4ª e 8ª séries do ensino fundamental e na 3ª série do ensino médio, em Matemática no Estado do Ceará e no Nordeste.

Gráfico 8: Desempenho dos alunos da 4ª série em Matemática Ensino Fundamental

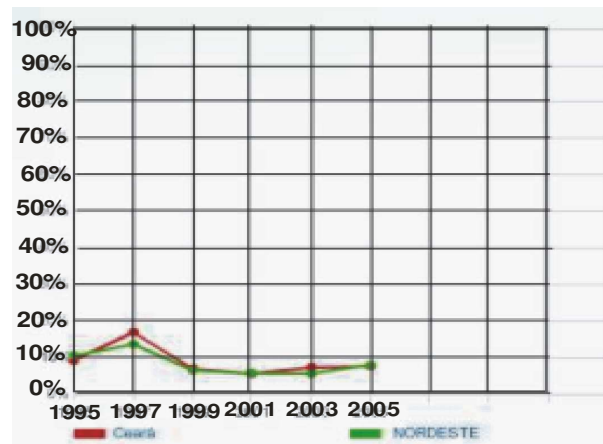


Gráfico 9: Desempenho dos alunos da 8ª série em Matemática Ensino Fundamental

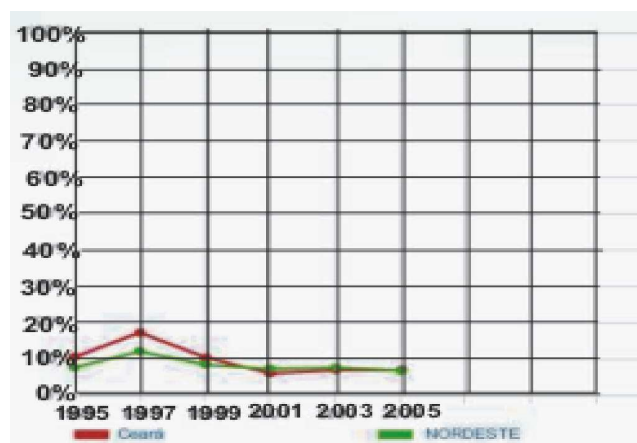
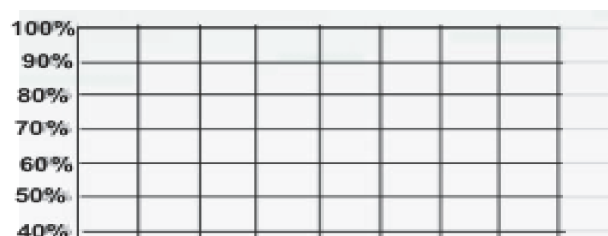


Gráfico 10: Desempenho dos alunos da 3ª série em Matemática Ensino Médio



Para a avaliação em Matemática, as amostras, em 2003, foram constituídas no Ceará por 1375 alunos, 92 diretores e 92 escolas de 4ª série do ensino fundamental; 1455 alunos, 81 diretores e 81 escolas de 8ª série do ensino fundamental; 980 alunos, 42 diretores e 42 escolas de 3ª série do ensino médio.

Nas amostras para a avaliação do desempenho dos alunos em Matemática foram aplicados, em cada série envolvida:

a) nos alunos:

a.1 → testes de Matemática;

a.2 → questionários para coletar informações sobre o contexto social, econômico, cultural dos alunos e a trajetória de sua escolarização.

b) nos professores: um questionário

c) nos diretores: um questionário

Os questionários do professor e diretor têm por objetivo levantar a formação profissional, práticas pedagógicas, nível sócio-econômico e cultural, estilos de liderança, climas acadêmico e disciplinar da escola, recursos pedagógicos disponíveis, infra-estrutura, recursos humanos e forma de gestão.

d) Questionário sobre as condições de infra-estrutura da escola, preenchidos pelo aplicador do teste.

Na construção da matriz de referência, envolveram-se também análises de professores especialistas nas áreas do conhecimento a serem avaliadas. Isso pressupõe a existência de competências cognitivas e habilidades a serem

desenvolvidas pelo aluno no processo de ensino-aprendizagem. Os conteúdos associados às competências e habilidades desejáveis para cada série, em cada disciplina, foram subdivididos em partes menores, cada uma especificando o que os itens das provas do SAEB devem medir. Estas unidades são denominadas “descritores”. Cada descritor dá origem a diferentes itens e, a partir das respostas dadas a eles, verifica-se o que os alunos sabem e conseguem fazer com os conhecimentos adquiridos.

Cada teste envolveu 169 itens de múltipla escolha, sendo aplicados cobrindo toda a Matriz de Referência. Esse conjunto de itens é dividido em 13 blocos compostos por 13 itens. Os blocos são agrupados, então, de três em três, formando 26 cadernos diferentes de prova. Apesar de se estar avaliando uma ampla gama de conteúdos, somente 39 questões são respondidas por cada aluno.

Os instrumentos são aplicados na amostra em cada estado pelas Secretarias Estaduais de Educação, utilizando-se dados cadastrais atualizados. Com relação às Secretarias Municipais de Educação, estas verificam se os dados das escolas selecionadas conferem com as informações do Censo Escolar, além de comunicar às escolas a sua participação no Saeb. Servem também de elo entre as Secretarias Estaduais de Educação e as escolas da rede municipal.

A amostra é selecionada aleatoriamente, tomando-se como base o cadastro no censo educacional do MEC e de acordo com o plano amostral do Saeb. A seleção está dividida em duas partes: na primeira, as escolas são identificadas em função dos extratos de interesse (municipal, estadual ou particular; interior ou capital; rural ou urbana); e em segunda, dentro dos extratos, as escolas são selecionadas aleatoriamente e de forma independente, dentro desses extratos.

A partir da avaliação realizada em 1995, tornou-se possível comparar resultados entre estados, entre os anos em que a avaliação foi feita e entre séries. Isto se tornou possível com o tratamento dos resultados obtidos utilizando-se a Teoria da Resposta ao Item (TRI) que permite estimar as habilidades dos alunos independentemente do conjunto específico de itens

respondido e a inserção de itens comuns aos testes aplicados em cada período. Da mesma forma, apesar de cada aluno responder somente a um bloco com 39 itens do total de 169 aplicados, através do processo de equalização, próprio da TRI, pode-se avaliar o desempenho do conjunto de sujeitos testados em função das habilidades e competências medidas através dos 169 itens.

Os resultados dos testes também são analisados com base na Teoria Clássica dos Testes (TCT). Nesse sentido, cabe ressaltar que o SAEB não tem como objetivo avaliar escolas, mas o sistema educacional como um todo. Os resultados obtidos são encaminhados para os gestores do sistema (Ministério da Educação, Secretarias Estaduais e Municipais de Educação), para que as informações possam ser utilizadas no processo de planejamento e de elaboração de políticas públicas na área educacional. Ao MEC, cabe apontar problemas e sugerir linhas de ação às Secretarias Estaduais e Municipais de Educação. É válido lembrar que alguns dos programas específicos criados pelo MEC para atacar os problemas da educação, e que vêm produzindo efeitos significativos sobre a gestão das escolas e sobre o aprendizado dos alunos, foram produtos da avaliação do SAEB.

Por outro lado, os resultados da avaliação são apresentados em uma escala de desempenho capaz de descrever, em cada nível, competências e habilidades que os alunos são capazes de demonstrar. A escala é única para cada disciplina e permite apresentar, em uma mesma métrica, resultados de desempenhos dos estudantes de todas as séries (4a e 8a séries do Ensino Fundamental e 3a série do Ensino Médio) e anos de aplicação dos testes (1995, 1997, 1999 e 2001). Pela escala, podem-se verificar quantos alunos já construíram as competências e habilidades desejáveis para cada uma das séries avaliadas, quantos ainda estão em processo de construção, quantos estão abaixo do nível que seria desejável para a série e quantos estão acima do nível que seria esperado.

Os alunos e escolas não são identificados na divulgação dos resultados, embora a coleta de informações e dados sejam individualizados, seus resultados são estendidos para os diversos grupos da população investigada.

3.1 ANÁLISE DE REGRESSÃO LINEAR MÚLTIPLA

A análise de regressão múltipla é um dos métodos de análise multivariada que permite, através de uma amostra \mathbf{n} , tomada em uma população \mathbf{P} , estimar a relação entre um conjunto de \mathbf{k} variáveis \mathbf{X}_i ($i = 2, 3, \dots, k$), com medida em \mathbf{P} , denominadas variáveis independentes ou preditoras, e um único critério de medida em \mathbf{P} , medido em escala métrica, denominado variável dependente (\mathbf{Y}) ou, simplesmente, critério. É um método que permite estimar valores assumidos por \mathbf{Y} , partindo-se do conhecimento dos valores assumidos pelas variáveis \mathbf{X}_i .

3.1.1 ANÁLISE DE REGRESSÃO LINEAR MÚLTIPLA

A análise de regressão linear múltipla é um caso da análise de regressão múltipla no qual se assume a ligação linear entre as variáveis independentes e a variável dependente em um plano \mathbf{n} -dimensional (hiperplano), segundo uma função linear geral.

$$\mathbf{Y} = \beta_0 + \beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2 + \dots + \beta_k\mathbf{X}_k + \mathbf{e}$$

Na expressão, observa-se que \mathbf{Y} é formada pela adição de duas componentes:

- a) uma componente funcional $\mathbf{Y} = \boldsymbol{\varphi}(\mathbf{X}_i)$ ou regressão de \mathbf{Y} sobre \mathbf{X}_i , representando a influência das variáveis independentes \mathbf{X}_i sobre a variável dependente \mathbf{Y}_i , definindo uma reta no hiperplano, representativa da lei estatística;
- b) uma componente aleatória \mathbf{e} , denominada erro ou resíduo, que representa a influência de outras variáveis que não as variáveis \mathbf{X}_i e denominadas variáveis latentes em \mathbf{Y} , incluindo-se em \mathbf{e} , também os erros de medida. Está implícito (Teoria Clássica da Mensuração) que os valores \mathbf{e} se anulam na média: $\mathbf{E}(\mathbf{e}) = 0$. Pelo teorema do limite central, quaisquer que sejam os tipos de distribuição de um grande número de variáveis, a sua soma tem distribuição normal, logo a

componente aleatória \mathbf{e} tem distribuição normal com média $\mathbf{0}$ e variância σ^2 , assim, $\mathbf{e} \Rightarrow N(\mathbf{0}; \sigma^2)$, logo, a probabilidade de se obter um valor de \mathbf{Y} , dada a medida para cada X_i , $\mathbf{P}(\mathbf{Y}|\mathbf{X}_i)$ segue uma distribuição normal, sendo mais verossímeis os valores que estão mais próximos da média. Ainda, segundo a Teoria Clássica, o erro \mathbf{e} independe de seus próprios valores em outros j instantes da mensuração: a não auto-correlação dos resíduos

$\mathbf{Cov}(\mathbf{e}_i; \mathbf{e}_j) = \mathbf{E}(\mathbf{e}_i; \mathbf{e}_j) = 0 \quad \forall, i \neq j \in i = 1, 2, \dots, n$. Essa idéia, teoricamente justificável, garante a característica de eficiência para os estimadores \mathbf{b}_i dos coeficientes β_i da função linear múltipla geral. Subentende-se que o modelo sendo bem especificado, o resíduo \mathbf{e} não deverá apresentar nenhuma componente sistemática.

Como todo modelo teórico, no caso da análise de regressão linear múltipla, supõe-se o cumprimento de condições básicas:

- 1) especificação formal do modelo, ligando a variável dependente \mathbf{Y} às k variáveis independentes \mathbf{X}_i . As funções mais comuns em estudos, envolvendo a análise de regressão múltipla, além da função linear múltipla, têm sido as funções hiperbolóide, exponencial n -dimensional e polinomial que, através de transformação, são conduzidas à forma linear múltipla como se segue:

- a) função hiperbolóide

$$\mathbf{Y} = \beta_0 \mathbf{X}_1^{\mathbf{B}_1} \mathbf{X}_2^{\mathbf{B}_2} \dots \mathbf{X}_k^{\mathbf{B}_k}$$

é transformada fazendo-se

$$\mathbf{LY} = \mathbf{Z}$$

$$\mathbf{L} \beta_0 = \mathbf{A}$$

$$\mathbf{LX}_1 = \mathbf{W}_1$$

$$\mathbf{LX}_2 = \mathbf{W}_2$$

...

...

...

$$\mathbf{LX}_k = \mathbf{W}_k$$

L indica logaritmo natural que, aplicado à expressão da função, resulta:

$$LY = L\beta_0 + \beta_1 LX_1 + \beta_2 LX_2 + \dots + \beta_k LX_k$$

então,

$$Z = A + B_1 W_1 + b_2 W_2 + \dots b_k W_k$$

b) função exponencial n-dimensional

$$Y = \beta_0 \cdot e^{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}$$

Fazendo-se

$$LY = Z$$

$$L\beta_0 = A$$

Como,

$$LY = L\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Então,

$$Z = A + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

c) Função Polinomial

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2^2 + \dots + \beta_k X_k^k$$

Fazendo-se

$$X_1 = Z_1; X_2^2 = Z_2; \dots; X_k^k = Z_k$$

tem-se

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2^2 + \dots + \beta_k Z_k^k$$

2) Y é uma variável dependente e aleatória formada pela soma de uma função linear com k variáveis

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

B_k (k= 0, 1, 2, ..., n) são os parâmetros da função linear múltipla

3) Os valores médios de Y , condicionados aos valores das variáveis independentes, constituem uma função hiperplana das variáveis

$$E(Y | X_1, X_2, \dots, X_k) = \varphi(X_1, X_2, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

4) A variável aleatória e tem distribuição normal com média zero e variância

$$\sigma^2 \Rightarrow N(0; \sigma^2)$$

5) Por via de consequência, pela 2ª condição

$$Y \Rightarrow N[(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k), \sigma^2]$$

6) A variável aleatória e é não auto-correlacionada e de variância constante (homocedasticidade)

$$\text{COV}(e_i; e_j) = E(e_i; e_j) = \sigma^2 \text{ constante, para } i = j \text{ e } 0 \text{ para } i \neq j$$

7) As variáveis independentes X_1, X_2, \dots, X_k não se correlacionam entre si quando tomadas duas a duas

$$|P_{X_i, X_j}| < 1, \forall j \neq i$$

A estimativa dos parâmetros β da reta de regressão consiste em, dada a função populacional

$$Y = \beta_0 + \sum_{i=1}^k \beta_i X_i$$

determinar, com uma amostra de tamanho n , os estimadores b_0 e b_j de modo que eles possam representar adequadamente os parâmetros populacionais β_0 e β_j .

No caso do modelo linear geral, utiliza-se o método dos mínimos quadrados que gera estimadores eficientes.

Definido o modelo na população de tamanho **N**, têm-se os valores de **Y** para os N elementos da população dados por

$$Y_i = \beta_0 + \sum_{j=1}^k \beta_j X_{i;j}$$

(i = 1, 2, ... k; j = 1, 2, ... N)

Definido o modelo na população de tamanho **N**, têm-se os valores de **Y** para os N elementos da população dados pela expressão acima:

Com os valores de **Y** e **X_i** em uma amostra de tamanho **n** tem-se

$$Y_1 = \beta_0 + \beta_1 X_{1;1} + \beta_2 X_{2;1} + \dots + \beta_k X_{k;1}$$

$$Y_2 = \beta_0 + \beta_1 X_{1;2} + \beta_2 X_{2;2} + \dots + \beta_k X_{k;2}$$

... ..

A função a ser ajustada pelo método dos mínimos quadrados é dada por

$$\sum_{i=1}^k [y_i - y_{i(est)}]^2 = [y_i - b_0 - b_1 X_{1;j} - b_2 X_{2;j} - \dots - b_k X_{k;j}] = \text{mínimo}$$

onde

y_i - valores de **y** na amostra

y_{i (est)} - valores de **y** estimados através do modelo

Para o termo constante **b₀**, tendo como condição necessária

$$\frac{\delta \sum_{i=1}^n [y_i - y_{i(est)}]^2}{\delta b_0} = 0 \rightarrow \forall i \in I = \{1, 2, 3, \dots, k\}$$

E como condição de suficiência

$$\frac{\delta^2 \sum_{i=1}^n [y_i - y_{i(est)}]^2}{\delta b_0^2} = 0$$

Os estimadores dos k parâmetros restantes (b_1, b_2, \dots, b_n) são obtidos, resolvendo-se k equações com k incógnitas, obtidas pela condição necessária, atendida a condição de suficiência.

A significância de cada estimador obtido é verificada com o teste "t" de Student, fazendo-se o contraste da hipótese nula $H_0: \beta = 0$ com hipótese alternativa $H_1: \beta \neq 0$. Aceitando-se H_1 , conclui-se que a variável, cujo parâmetro foi testado, é uma das variáveis explicativas do modelo. A análise de variância permite verificar a adequação do ajustamento ao modelo (bondade do ajustamento) com a hipótese nula $H_0: \beta_i = 0$, contra a hipótese alternativa H_1 : pelo menos um dos parâmetros β_i é diferente de zero.

Outro resultado importante obtido com a análise de variância é o poder explicativo da variação das variáveis X_i sobre a variação de Y , o coeficiente de determinação, indicado por R^2 .

O coeficiente de determinação vale o quadrado do coeficiente de correlação múltipla entre y, x_1, x_2, \dots, x_k

$$R^2 = \frac{\text{soma dos quadrados devidos à regressão}}{\text{soma dos quadrados dos resíduos}}$$

Como R^2 tende a ser uma estimativa otimista da qualidade do modelo ajustado à população, a estatística $R^2(\text{ajustado})$ busca corrigir o R^2 amostral para refletir com melhor exatidão o ajustamento do modelo à população. O $R^2(\text{ajustado})$ é dado por:

$$R^2(\text{ajustado}) = \frac{R^2 - k(1 - R^2)}{N - k - 1}$$

Onde

R^2 - coeficiente de determinação

k - número de variáveis independentes no modelo

n - tamanho da amostra

3.1.2 Condições a serem cumpridas para a realização da análise de regressão múltipla

3.1.2.1 INCORPORAÇÃO DE VARIÁVEIS NÃO MÉTRICAS

As variáveis a serem incluídas na análise de regressão são variáveis métricas. A inclusão de variáveis não métricas é feita através de transformações das categorias de medida (níveis) dessas variáveis em **k** variáveis dicotômicas, atribuindo-se à categoria o valor **1** se esta se apresenta como medida da variável não métrica e, em caso contrário, atribuindo-se o valor **0**. Na transformação da variável não métrica, incluem-se somente **(k – 1)** categorias uma vez que a categoria não incluída é perfeitamente determinada pelas demais (uma singularidade).

3.1.2.2 TAMANHO DE AMOSTRA ADEQUADO

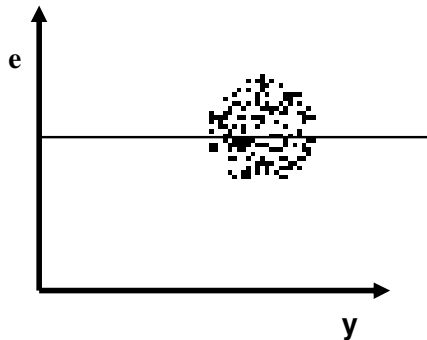
O tamanho da amostra afeta a generalização dos resultados em função da razão entre o número de observações e o número de variáveis independentes envolvidas. A razão não deve ficar abaixo de **5** para **1**, ou seja, para cada variável independente o número de observações deve ser igual ou maior que **5**. Consequentemente, o tamanho da amostra deve ser igual ou maior que 5 vezes o número de variáveis independentes. Se o tamanho da amostra for menor que essa relação, corre-se o risco de “superajustar” o modelo com a amostra, fazendo com que os resultados sejam específicos à amostra, minimizando a capacidade de generalização para a população.

3.1.2.3 NORMALIDADE DA VARIÁVEL DEPENDENTE

A variável dependente tem distribuição normal.

3.1.2.4 HOMOGENEIDADE DAS VARIÂNCIAS (HOMOCEDESTICIDADE)

A homocedasticidade refere-se à constância dos resíduos associados aos valores das variáveis independentes. Observa-se, previamente, a existência de homocedasticidade com o **Teste de Levene** que mede a igualdade de variâncias para um par de variáveis, ou posteriormente, através de um gráfico, representando a relação entre os resíduos e os valores da variável dependente estimados pelo modelo obtido. Não ocorre violação dessa condição quando se obtém um gráfico nulo do tipo Gráfico Nulo



Onde

e – eixo dos resíduos

y – eixo dos valores da variável dependente, estimados pelo modelo

3.1.2.5 LINEARIDADE

Cada variável independente relaciona-se linearmente com a variável dependente. Verifica-se uma relação linear através da análise de variância.

3.1.2.6 INDEPENDÊNCIA DOS TERMOS DE ERRO

O valor previsto pela regressão não está relacionado com qualquer outra previsão. Se os resíduos forem independentes o padrão será aleatório. Verifica-se essa condição através de uma representação gráfica do valor estimado da variável independente do modelo obtido e os resíduos. Essa condição é atendida quando se obtém um gráfico nulo.

3.1.2.7 NORMALIDADE DOS TERMOS DE ERRO

O resíduo tem distribuição normal com média $\mu = 0$ e variância σ^2 . Observa-se o atendimento da condição com uso do gráfico de probabilidade normal, onde os resíduos padronizados são comparados com a distribuição normal.

3.1.2.8 COLINEARIDADE E MÚLTICOLINEARIDADE

As variáveis independentes não se correlacionam entre si. A ocorrência de colinearidade ou de multicolinearidade é avaliada através de:

- a) identificação dos índices de condição com valor maior que **30**.

Para todos os índices de condição maiores que **30** identificar as variáveis com proporção da variância maior que **0,90**.

A existência de colinearidade ou de multicolinearidade quando, para um índice de condição ≥ 30 , existem duas ou mais variáveis com proporção da variância $> 0,90$; ou

- b) valor de tolerância e o fator de inflação da variância dizem o grau em que cada variável independente é afetada pelas demais variáveis independentes. Um valor de referência comum é uma tolerância de **0,10** correspondendo a $V_{if} > 10$.

$$V_{if} = \frac{1}{\text{Tolerância}}$$

Tolerância – quantia de uma variável independente não explicada pelas outras variáveis independentes.

3.2 METODOLOGIA

O trabalho é de natureza quantitativa, utilizando os dados obtidos pela avaliação do SAEB/2003, cujos dados levantados referem-se às características do professor, do diretor, da escola e dos alunos de 4ª e 8ª séries do ensino fundamental, cursando à época a disciplina matemática.

3.2.1 População

A população objeto da avaliação pelo SAEB, em nível nacional, no ano de 2003, foi definida como constituída

[...] “por todos os alunos matriculados em 2003 nas escolas urbanas constantes do Censo Escolar de 2002 em uma das três séries de interesse, (4ª e 8ª séries do Ensino Fundamental e 3ª série do Ensino Médio), exclusive os alunos das turmas multisseriadas e de aceleração. Foram incluídos também no universo da 4ª série do ensino fundamental do ensino rural com 10 ou mais alunos na série.” (SAEB-2003).

Os extratos de interesse para o SAEB 2003 foram definidos a partir dos critérios a seguir:

Quadro 4 - CATEGORIAS DA ANÁLISE DO SAEB 2003

Critério	Número	Descrição
Série	3	4ª EF, 8ª EF e 3ª EM
Estado	27	26 estados + Distrito Federal
Rede	4	Federal, estadual, municipal e particular
Localização	4	Região metropolitana das capitais ou capital, cidades grandes não metropolitanas, outras cidades, rural (apenas para a 4ª série do ensino fundamental)

Fonte: INEP/SAEB – 2003.

3.2.2 Amostra

A amostra utilizada para o estudo foi constituída pela porção da amostra nacional do SAEB, selecionada aleatoriamente, referente ao Estado do Ceará, com os alunos da 4ª e 8ª séries do ensino fundamental que responderam ao teste de matemática, seus respectivos professores, seus diretores e escolas.

3.2.3 Instrumentos

Foram aplicados pelo SAEB, para cada série objeto da avaliação, os seguintes instrumentos:

- a) nos alunos:
 - a.1 → testes de Matemática;

a.2 → questionários para coletar informações sobre o contexto social, econômico, cultural dos alunos e a trajetória de sua escolarização.

b) Nos professores: um questionário;

c) nos diretores: um questionário;

Os questionários do professor e diretor, tendo por objetivo levantar a formação profissional, práticas pedagógicas, nível sócio-econômico e cultural, estilos de liderança, climas acadêmico e disciplinar da escola, recursos pedagógicos disponíveis, infra-estrutura, recursos humanos e forma de gestão.

d) Questionário sobre as condições de infra-estrutura da escola, preenchidos pelo aplicador do teste.

3.2.4 Organização dos dados do SAEB/2003 no Estado do Ceará

Os arquivos da avaliação do SAEB/2003 foram cedidos pelo INEP. Em relação à avaliação em matemática, totalizavam quatro arquivos:

- um arquivo com dados dos alunos;
- um arquivo com dados do professor;
- um arquivo com dados dos diretores;
- um arquivo com dados das escolas;

Construíram-se dois arquivos (um para 4ª série e outro para a 8ª série). Nos dois arquivos, tomando-se como referência o arquivo do aluno, os dados dos outros três arquivos do SAEB foram pareados para cada aluno com os dados do seu professor de matemática, do diretor de sua escola e os dados da sua respectiva escola. Esses arquivos foram construídos, usando-se o “software” **SPSS** (Statistical Packwage for the Social Sciencies) para “Windows”, versão **13,0**.

Nos arquivos do INEP as variáveis estavam codificadas, porém seus títulos eram constituídos pelas perguntas dos questionários. Essas perguntas foram transformadas na forma de nome das variáveis que elas mediam.

Tomando-se as variáveis de localização da escola (urbana ou rural, na região metropolitana ou fora da região metropolitana); dependência administrativa: pública – municipal, estadual ou federal ou particular; tamanho

do município (com menos de 200 mil habitantes, com 200 mil ou mais de 200 mil habitantes). Foi realizada a análise de variância para se verificar a existência de diferenças significativas entre essas categorias no rendimento nos testes de matemática (*proficiência em matemática*, na denominação do SAEB), de modo a não influir no resultado da regressão. Observou-se a existência de diferenças significativas com grandes distâncias entre as médias das escolas públicas e das particulares. Assim, os arquivos de 4^a e 8^a séries foram desdobrados cada um em dois arquivos, um com dados das escolas públicas e outro com dados das escolas particulares.

Aqueles conjuntos de variáveis que mediam isoladamente aspectos de uma variável foram sintetizados em uma escala de escores da variável que era medida através dessas.

3.2.5 Tratamento estatístico dos dados

Nos quatro arquivos prontos, conforme descritos anteriormente, tomou-se cada variável e verificou-se sua relação com a variável *proficiência em matemática*, calculando-se o coeficiente de correlação de Pearson. Em seguida eliminaram-se aquelas variáveis que apresentaram coeficientes de correlação não significativos, ou seja, com significância $p > 0,05$ (Apêndice II).

Como o coeficiente de correlação de Pearson é calculado supondo uma relação de linearidade, testou-se a linearidade de cada uma das variáveis não eliminadas com a variável *proficiência em matemática*, o que constitui também uma pré-condição para a realização da regressão linear (Apêndice III).

Para a verificação de linearidade e homogeneidade das variâncias na distribuição da função linear entre cada uma das variáveis com a *proficiência em matemática*, utilizou-se a estatística de Levene (Apêndice IV).

Verificados os pressupostos para a realização da análise de regressão múltipla, foram realizadas quatro regressões com os dados de cada um dos arquivos, tomando-se como variável independente a variável *proficiência em matemática* e como variáveis dependentes aquelas que se relacionaram significativamente com a variável independente.

4 APRESENTAÇÃO E ANÁLISE DOS DADOS

A seguir são apresentados os resultados da proficiência em matemática nas duas amostras, bem como o comportamento dessa variável em relação às categorias de análises do SAEB: dependência administrativa da escola, localização, tamanho do município, entre outros, justificando, através das diferenças observadas, o desdobramento da amostra nas duas séries em quatro subamostras, construídas pelos alunos da 4ª série da escola pública, outra constituída pelos alunos da 4ª série de escola particular e, na 8ª série, pelos alunos de escola pública e outra pelos alunos de escola particular. Finalmente são apresentados os resultados da análise de regressão nessas quatro subamostras.

4 1 DESCRIÇÃO DA AMOSTRA DA QUARTA SÉRIE

A amostra da 4ª série na rede pública apresentou o seguinte resultado para a *proficiência em matemática*: média = **156,1**; mediana = **153**; desvio padrão = **34,86**; coeficiente de variação = **22,3%**; coeficiente de assimetria = **0,47**; valor mínimo para a proficiência = **66,5** e o valor máximo = **302,6**. Tabelas no apêndice apresentam as estatísticas da variável proficiência do aluno no teste de matemática nessas categorias e quatro diferentes estimativas para média populacional, seguidas pela representação gráfica.

A amostra da 4ª série rede particular apresentou o seguinte resultado para a *proficiência em matemática*: média = **217**; mediana = **217**; desvio padrão = **48,09**; coeficiente de variação = **22,2%**; coeficiente de assimetria = **0,23**; valor mínimo para a proficiência = **109,9** e o valor máximo = **352,4**. Tabelas no apêndice apresentam as estatísticas da variável proficiência do aluno no teste de matemática nessas categorias e quatro diferentes estimativas para média populacional, seguidas pela representação gráfica.

A amostra da 8ª série rede pública apresentou o seguinte resultado para a *proficiência em matemática*: média = **226**; mediana = **221**; desvio padrão = **45,39**; coeficiente de variação = **20,1%**; coeficiente de assimetria = **0,87**; valor mínimo para a proficiência = **127** e valor máximo = **423**. Tabelas no apêndice apresentam as estatísticas da variável proficiência do aluno no teste de

matemática nessas categorias e quatro diferentes estimativas para média populacional, seguidas pela representação gráfica.

A amostra da 8ª série rede particular apresentou o seguinte resultado para a *proficiência em matemática*: média = **302**; mediana = **307**; desvio padrão = **57,93**; coeficiente de variação = **19,2%**; o coeficiente de assimetria = **0,21**; valor mínimo para a proficiência = **156** e o valor máximo = **421**. Tabelas no apêndice apresentam as estatísticas da variável proficiência do aluno no teste de matemática nessas categorias e quatro diferentes estimativas para média populacional, seguidas pela representação gráfica.

Tomado-se a localização da escola (em região metropolitana ou não), a dependência administrativa da escola (pública ou particular), o tamanho do município da escola (menor que 200.000 habitantes ou maior que 200.000 habitantes) como variáveis independentes e a proficiência do teste de matemática como variável dependente. Realizando-se a análise da variância (Modelo Linear Geral) observou-se, na 4ª série, diferença significativa entre as médias de proficiência da escola pública e da escola privada (**F = 86,06; p < 0,01**).

Considerando esse resultado na 4ª e 8ª séries e que a amostra das escolas particulares contém somente 8 alunos fora da região metropolitana, a análise foi realizada somente com os arquivos das escolas públicas e particulares nas duas séries.

4.1.1 PREPARAÇÃO DOS DADOS PARA A ANÁLISE DE REGRESSÃO

1. As variáveis medidas em escala nominal tiveram suas categorias de medida transformadas em variáveis “mudas” (*dumny*).
2. As medidas através de seus componentes com cada componente medido através de escala de concordância do tipo [1. *concordo*; 2. *concordo em parte*; 3. *discordo* – ou 1. *sim*; 2. *não*]. Essas escalas foram reescorizadas, atribuindo-se o escore **0** à mais baixa concordância e à concordância mais alta atribuindo-se o escore mais alto. Nos dois exemplos citados, as escalas ficaram assim [0. *discordo*; 1. *concordo em parte*; 2. *concordo* – 0. *não*; 1. *sim*].

Realizada a transformação, os escores foram totalizados, constituindo a medida da variável em uma escala intervalar.

3. Ficaram no arquivo da 4ª série, **79** variáveis do aluno, **77** do professor, **139** do diretor e **20** da escola, totalizando **315** variáveis.

Nos arquivos da 8ª série ficaram **92** variáveis do aluno, **77** do professor, **141** do diretor e **20** da escola, totalizando **330** variáveis.

Como se verificou que na amostra de 4ª série: 1- somente 32 alunos (**2,3%**) estavam matriculados em escolas da zona rural; e, 2- na 4ª série a diferença entre a média de *proficiência em matemática* de alunos de escola públicas era significativamente menor que a de alunos das escolas particulares - o arquivo de 4ª série foi desdobrado em dois arquivos: um arquivo com os dados das escolas públicas e outro com os dados das escolas particulares.

Na amostra da 8ª série, verificou-se: 1-32 alunos (**2,2%**) estavam matriculados em escolas federais; e, 2- a diferença entre a média de *proficiência em matemática* nos alunos das escolas públicas foi significativamente menor que a média dos alunos das escolas particulares; então os arquivos da 8ª série foram desdobrados em dois arquivos, um com os dados das escolas públicas e o outro das escolas particulares. Ressalte-se que os 32 alunos matriculados em escolas federais apresentaram a maior média e diferente significativamente, em relação às outras três categorias de dependência administrativa. Sua diluição no arquivo das escolas públicas não produziu modificação estatisticamente significativa na média dos alunos da escola pública, dada a pequena quantidade destes alunos em relação ao tamanho do arquivo.

Construídos os quatro arquivos, suas variáveis foram correlacionadas com a *proficiência em matemática*. Aqueles que não apresentaram o coeficiente de correlação de Pearson (**r**) significativo foram eliminados do arquivo. Em seguida testou-se a linearidade entre cada variável com a variável *proficiência em matemática* (pela análise da variância) e a homogeneidade das variâncias (estatística de Levene) analisados a seguir, com os resultados totais no Apêndice.

Quadro Nº 5 - Coeficientes de correlação significativos ($P < 0,05$ ou $0,01$) das variáveis do aluno, do professor, do diretor e da escola com a variável *proficiência em matemática* no arquivo da 4ª série das escolas públicas

ARQUIVO	INTERVALOS DO COEFICIENTE DE CORRELAÇÃO	CLASSIFICAÇÃO DA CORRELAÇÃO	Nº DE VARIÁVEIS
(variáveis do aluno)	< 0,20	muito fraca	28
	0,20 – 0,39	fraca	1
	0,40 – 0,59	média	-
	0,60 – 0,79	forte	-
	≥ 0,80	muito forte	-
Variáveis do professor	< 0,20	muito fraca	19
Variáveis do diretor	< 0,20	muito fraca	50
Variáveis da escola	< 0,20	muito fraca	5
TOTAL	-	-	103

Somente uma das variáveis da 4ª série das escolas públicas apresentou coeficiente de correlação no intervalo de **0,20** e **0,39**, sendo a correlação classificada como **fraca**. As demais variáveis com correlação significativa apresentaram coeficiente de correlação menor que **0,20**, classificando-se a correlação como **muito fraca**.

Quadro Nº 6 - Coeficientes de correlação significativos ($P < 0,05$ ou $P < 0,01$) das variáveis do aluno, do professor, do diretor e da escola com a variável *proficiência em matemática* no arquivo da 4ª série das escolas particulares

	INTERVALOS DO COEFICIENTE DE CORRELAÇÃO	CLASSIFICAÇÃO DA CORRELAÇÃO	Nº DE VARIÁVEIS
Variáveis do aluno	< 0,20	muito fraca	11
	0,21 – 0,39	fraca	6
TOTAL PARCIAL			17
Variáveis do professor	< 0,20	muito fraca	19
	0,20 – 0,39	fraca	13
TOTAL PARCIAL			32
Variáveis do diretor	< 0,20	muito fraca	21
	0,20 – 0,39	fraca	40
	0,40 – 0,59	média	5
TOTAL PARCIAL			66
Variáveis da escola	< 0,20	muito fraca	3
	0,20 – 0,39	fraca	6
	0,40 – 0,59	média	1
TOTAL PARCIAL	-	-	10
TOTAL	-	-	125

Na 4ª série nas escolas particulares, somente **6** variáveis (**4,8%**), com coeficiente de correlação significativo, apresentaram correlação **média**, com coeficiente no intervalo **0,40** a **0,69**; **65** variáveis (**52%**) com correlação **fraca**,

com coeficiente no intervalo de **0,20** a **0,39**; e, **54** variáveis (**43,2%**) com correlação ***muito fraca***, coeficiente menor que **0,20**.

Quadro Nº 7 Coeficientes de correlação significativos ($P < 0,05$ ou $0,001$) das variáveis do aluno, do professor, do diretor e da escola com a variável *proficiência em matemática* no arquivo da 8ª série das escolas públicas

ARQUIVO	INTERVALOS DO COEFICIENTE DE CORRELAÇÃO	CLASSIFICAÇÃO DA CORRELAÇÃO	Nº DE VARIÁVEIS
(variáveis do aluno)	< 0,20	muito fraca	24
	0,20 – 0,39	fraca	9
	0,40 – 0,59	média	3
Total parcial	-	-	36
Variáveis do professor	< 0,20	muito fraca	14
	0,20 – 0,39	fraca	7
	0,40 – 0,59	média	1
Total parcial	-	-	22
Variáveis do diretor	< 0,20	muito fraca	24
	0,20 – 0,39	fraca	17
	0,40 – 0,59	média	10
	0,60 – 0,79	forte	1
TOTAL PARCIAL	-	-	52
Variáveis da escola	< 0,20	muito fraca	1
	0,20 – 0,39	fraca	3
	0,40 – 0,59	média	1
Total parcial	-	-	5
TOTAL	-	-	115

No arquivo das escolas públicas da 8ª série **114** variáveis apresentaram coeficiente de correlação significativo entre elas **63** variáveis (**54,8%**) com correlação ***muito fraca***. **36** variáveis (**31,3%**) com correlação ***fraca***. **15** variáveis (**13,0%**) com correlação ***média***, e **1** variável (**0,9%**) com correlação ***forte***.

Quadro Nº 8 coeficientes de correlação significativos ($P < 0,05$ ou $0,001$) das variáveis do aluno, do professor, do diretor e da escola com a variável *proficiência em matemática* no arquivo da 8ª série das escolas particulares

	INTERVALOS DO		
--	---------------	--	--

ARQUIVO	COEFICIENTE DE CORRELAÇÃO	CLASSIFICAÇÃO DA CORRELAÇÃO	Nº DE VARIÁVEIS
(variáveis do aluno)	< 0,20	muito fraca	7
	0,20 – 0,39	fraca	14
	0,40 – 0,59	média	-
Total parcial	-	-	21
Variáveis do professor	< 0,20	muito fraca	3
	0,20 – 0,39	fraca	14
	0,40 – 0,59	média	2
Total parcial	-	-	19
Variáveis do diretor	< 0,20	muito fraca	20
	0,20 – 0,39	fraca	13
	0,40 – 0,59	média	2
TOTAL PARCIAL	-	-	35
Variáveis da escola	< 0,20	muito fraca	5
	0,20 – 0,39	fraca	9
	0,40 – 0,59	média	2
Total parcial	-	-	16
TOTAL	-	-	91

No arquivo das escolas particulares da 8ª série **91** variáveis apresentaram coeficiente de correlação significativo; entre elas **35** variáveis (**38,5%**) com correlação **muito fraca**. **50** variáveis (**54,9%**) com correlação **fraca**. **6** variáveis (**6,6%**) com correlação **média**.

4.1.2 LINEARIDADE

Todas as variáveis dos dois arquivos da 4ª série e dos dois arquivos da 8ª série apresentaram relação de linearidade com a variável *proficiência em matemática*.

4.1.3 HOMOGENEIDADE DAS VARIÂNCIAS

Os quadros a seguir contêm os resultados do teste de homogeneidade das variâncias das variáveis dos arquivos com a *proficiência em matemática*, conforme o Apêndice.

Quadro 9 - HOMOGENEIDADE DAS VARIÂNCIAS

SÉRIE	Nº de variáveis	Nº DE VARIÁVEIS SEM HOMOGENEIDADE DAS VARIÂNCIAS

4ª série – Escola pública	103	33
4ª série – Escola particular	125	45
8ª série – Escola pública	114	105
8ª série – Escola particular	91	48
TOTAL	433	231

4.1.4 Realização das regressões

Tendo-se verificado que somente a homogeneidade das variâncias não ocorria em uma parte das variáveis de cada arquivo, procedeu-se às quatro regressões e, em seguida, verificou-se graficamente, através da análise de colinearidade, do valor da inflação da variância (**VIF**) e da inspeção gráfica a ocorrência de alguma violação grave dos pressupostos que tenham interferido. Os resultados estão apresentados nos quadros relacionados nas páginas 87 a 90.

Nas regressões realizadas, observa-se que na 4ª série da escola pública, que pesam negativamente o *trabalho infantil*, tanto no lar como fora deste, a *formação do professor* (professores formados em curso superior semi-presenciais, em cursos aligeirados, têm seus alunos com menor desempenho na avaliação) *alunos que já foram reprovados*, *discriminação racial*, *início tardio da vida escolar*, *nenhuma escolaridade em quem acompanha a vida do aluno*, influem negativamente no desempenho em matemática, no caso do *início tardio da vida escolar*, *alunos que fizeram a pré-escola*, tendem a apresentar melhor rendimento no teste de matemática do que aqueles que a iniciaram na 1ª, ou na 2ª ou na 3ª séries do ensino fundamental, inclusive, pesa negativamente em ordem crescente de peso, o início na 1ª série seguido pelo na 2ª série e com maior peso negativo, aqueles que somente a iniciaram na 3ª série. Em relação às alunas apresentarem menor desempenho no teste que os alunos, devem-se a um fator cultural, é comum pais e, até mesmo professores, afirmarem que “*mulher não dá para matemática*” (sic).

O fato de variável *eletrodoméstico existente na residência* pesar positivamente, essa variável é um indicador de renda, é truísmo afirmar que o desempenho está associado à renda de sua família. Outro que chama à

atenção é o peso positivo no *relacionamento com os colegas e na formação das turmas*, tendo como critério a heterogeneidade quanto ao rendimento escolar, sem abstrair a influência de outras variáveis associadas, á atuação conjunta dessas duas variáveis, possivelmente, esteja favorecendo o processo de interação entre os alunos.

Já na 4ª série da escola particular, a regressão com as variáveis que ficaram no modelo são menor número que produzem um maior nível de explicação no rendimento escolar que aquelas no modelo da escola pública, (60% nas particulares e 39% nas públicas).

Da mesma forma que nas escolas públicas, a *reprovação*, o *trabalho infantil* e a *frequência na escola* influem negativamente no teste de matemática. Também a *sobrecarga de trabalho dos professores pesa negativamente*. A *existência de livros e não livros escolares na residência do aluno*, mesmo sendo indicador fraco de renda, mostra duas coisas: a influência da renda familiar no rendimento do aluno bem como o nível cultural da família. Ainda dentro dessa perspectiva, deve-se chamar a atenção para a *disponibilidade de computadores para apoio dos alunos em atividades didáticas nas escolas particulares* ser a variável com maior peso no modelo, tendo repercussão positiva no rendimento em matemática.

Com relação à 8ª série das escolas públicas, as regressões realizadas mostraram que, nas escolas públicas, a variável de maior peso foi a de *diretores com mestrado*, também, tendo peso positivo a de *diretores com curso de especialização* e a *graduação não em licenciatura* (outro curso). Como se vê, a formação em pós-graduação tem relativa influência na gestão da escola. Presente também, influenciando, agora, negativamente a renda familiar, o tempo na direção, a forma como assumiu a direção da escola, seja por eleição ou indicação. Neste caso, apesar da importância da democratização na vida escolar, com o diretor ter sido eleito pelo voto direto da comunidade escolar, isto, efetivamente, não contribui para a melhoria da aprendizagem, mesmo porque, em muitas escolas, observam-se os vícios que ocorrem nas eleições para os diversos cargos eletivos na vida política, tais como aliciamento, compra de votos, práticas que fogem à ética.

Nas variáveis do aluno, *a reprovação, a escolaridade somente em escola pública, o início tardio na escolaridade*, bem como, *o fato de declarar que gosta de estudar matemática*, apresentam peso positivo.

Somente uma variável do professor ficou no modelo de regressão obtido: professores que declaram que *a indisciplina dificulta o trabalho de ensinar* têm seus alunos com menos rendimento no teste que os demais. Isso significa que os alunos têm mais idade nessa série, são mais difíceis de serem controlados e falta formação aos professores para lidarem com a indisciplina, além de estratégias de motivação. Finalmente, ressalta-se que os *projetos de melhoria da escola*, escolas que participam do Projeto Parâmetro em Ação do Governo federal, têm seus alunos um melhor desempenho no teste em relação aos alunos das escolas particulares.

Na 8ª série das escolas particulares, o maior peso está na variável que se refere à *existência de equipamento e estado de condições para uso* pesando também, a *disponibilidade e a utilização de equipamento e materiais*, influenciando também, positivamente, a *existência na escola de programa de apoio e reforço da aprendizagem* e o *diretor ser licenciado em matemática*. Em relação ao aluno, influi positivamente *gostar de estudar matemática* e a *escolaridade do pai*. Negativamente, o *trabalho, a idade e fazer a lição de matemática de vez em quando*.

Quadro 10 - Verificação Posterior à Regressão de Violação dos Pressupostos para a sua Realização (Método “Stepwise”) – Resultados: 4ª série

AMOSTRA	AJUSTAMENTO DO MODELO		NORMALIDADE DA VARIÁVEL INDEPENDENTE	HOMOGENEIDADE DAS VARIÁNCIAS HOMOCEDASTICIDADE	LINEARIDADE	INDEPENDÊNCIA E NORMALIDADE DOS TERMOS DE ERRO	COLINEARIDADE e/ou MULTICOLINEARIDADE – Fator de Inflação da Variável (VIF)
	F	α					
Escola pública	15,49	0,01	não	Não	não	não	[1,03; 1,36]
Escolas particulares	18,02	0,01	não	não	não	não	[1,08; 2,68]

Quadro 11 - Verificação Posterior à Regressão de Violação dos Pressupostos para a sua Realização (Método “Stepwise”) – Resultados: 8ª série

AMOSTRA	AJUSTAMENTO DO MODELO		NORMALIDADE DA VARIÁVEL INDEPENDENTE	HOMOGENEIDADE DAS VARIÁNCIAS HOMOCEDASTICIDADE	LINEARIDADE	INDEPENDÊNCIA E NORMALIDADE DOS TERMOS DE ERRO	COLINEARIDADE e/ou MULTICOLINEARIDADE – Fator de Inflação da Variável (VIF)
	F	α					
Escola pública	35,65	0,01	não	não	não	não	[1,05; 2,24]
Escolas particulares	33,25	0,01	não	não	não	não	[1,09; 5,38]

Quadro 12 - REGRESSÃO (Método “Stepwise”) – RESULTADOS: 4ª série

AMOSTRA	COEFICIENTE DE CORRELAÇÃO MÚLTIPLA	ERRO PADRÃO DA ESTIMAÇÃO	TAMANHO DA AMOSTRA	Nº. de VARIÁVEIS NO MODELO FINAL
Escola pública	0,52	30,09	1073	25
Escolas particulares	0,70	35,17	302	15

Quadro 13 - REGRESSÃO (Método “Stepwise”) – RESULTADOS: 4ª série

AMOSTRA	MODELO LINEAR MÚLTIPLO	GRAU DE SIGNIFICÂNCIA DOS COEFICIENTES E DA CONSTANTE	
		"t" deStudent	α
Escola pública	$Y_1 = 140,58 + 6,63X_1 - 6,41X_2 - 6,61X_3 - 18,07X_4 - 2,99X_5 - 6,18X_6 - 8,54X_7 + 5,71X_8 - 8,73X_9 + 1,18X_{10} + 8,98X_{11} + 26,36X_{12} + 9,50X_{13} - 9,30X_{14} + 5,86X_{15} - 24,10X_{16} - 24,24X_{17} + 16,47X_{18} + 8,05X_{19} + 11,40X_{20} + 4,69X_{21} + 4,26X_{22} - 6,43X_{23} - 3,78X_{24}$	[-1,97; 16,47]	[0,01; 0,05]
Escolas particulares	$Y_4 = 176,82 + 1,03X_1 - 18,30X_2 + 14,62X_3 - 13,34X_4 - 30,34X_5 + 20,01X_6 - 6,64X_7 + 14,28X_8 + 7,38X_9$	[-3,19; 23,98]	[0,01; 0,02]

VARIÁVEIS DA ESCOLA PÚBLICA: 4ª série

- Y₁ → Proficiência do aluno em matemática
- X₁ → a042_13_14f - Relacionamento com os colegas (0,140)
- X₂ → a041_046 - Em dia de aula, quanto tempo trabalha fora de casa (-0,141)
- X₃ → p005_2 - Forma que realizou o curso superior: semi-presencial (-0,118)
- X₄ → a041_045 - Em dia de aula, quanto tempo faz trabalhos domésticos em casa (-0,105)
- X₅ → a042_002 - Reprovação (-0,110)
- X₆ → a041_001 - Sexo (-0,122)
- X₇ → p003_1 - Como se considera: Branco (-0,079)
- X₈ → a041_002_3 - Você se considera: Preto (-0,083)
- X₉ → a042_001_1 - Começou a estudar: no maternal (0,093)
- X₁₀ → a041_f1 - Variáveis do aluno - Questionário 1 - Fator1: eletrodomésticos existentes na residência (-0,071)
- X₁₁ → p018_3 - Exercício de outra atividade, além do magistério, que gera renda: não (-0,087)
- X₁₂ → d008_6 - Nível de escolaridade (até a graduação): Ensino Superior - Licenciatura em Letras (0,114)
- X₁₃ → d003_2 - Como se considera: - parto (0,135)
- X₁₄ → d070 - Dificuldades de aprendizagem do aluno: são decorrentes do nível cultural dos pais dos alunos (-0,078)
- X₁₅ → d111 - Participação em programas: Escola Ativa (0,078)
- X₁₆ → a042_001_4 - Começou a estudar: na 2ª série (-0,091)
- X₁₇ → a042_001_5 - Começou a estudar: na 3ª série (-0,092)
- X₁₈ → a041_26_6 - Escolaridade da mãe: não completou o Ensino Médio (0,073)

- X_{19} → t001_02 - Número de professores de Matemática que a turma teve: dois (0,081)
 X_{20} → d033_4 - Critério utilizado para a formação de turmas: Heterogeneidade quanto ao rendimento escolar (0,095)
 X_{21} → d024 - Carga horária de trabalho na escola (0,086)
 X_{22} → a041_002_2 - Você se considera: Pardo (0,061)
 X_{23} → a042_001_3 - Quando começou a estudar: na 1ª série (0,073)
 X_{24} → a041_33_1 - Escolaridade da pessoa que acompanha a vida escolar: nenhuma (0,053)

Escolas particulares: 4ª série

Y_4 → Proficiência do aluno em matemática

- X'_1 → e005_1_005_6f - Disponibilidade de computadores (0,276)
 X'_2 → d066 - Dificuldades de aprendizagem do aluno: relacionam-se à sobrecarga de trabalho do(as) professores(as) (-0,158)
 X'_3 → d011_3 - Especialização - mínimo de 360 horas (0,151)
 X'_4 → a042_002 - Reprovação (-0,112)

- X'_5 → a042_004 - Deixou de frequentar a escola por algum tempo (-0,151)
 X'_6 → a042_006 - Gosta de estudar Matemática (0,164)
 X'_7 → a041_045 - Em dia de aula, quanto tempo faz trabalhos domésticos em casa (-0,149)
 X'_8 → p136_2 - Escolha do livro didático: O coordenador pedagógico, orientador educacional ou o diretor depois de consultar os professores (0,119)
 X'_9 → a041_019 - Além dos livros escolares, quantos livros há a casa (0,124)

Quadro 14 - REGRESSÃO (Método “Stepwise”) – RESULTADOS: 8ª série

AMOSTRA	COEFICIENTE DE CORRELAÇÃO MÚLTIPLA	COEFICIENTE DE DETERMINAÇÃO CORRIGIDO	ERRO PADRÃO DA ESTIMAÇÃO	TAMANHO DA AMOSTRA	Nº. de VARIÁVEIS NO MODELO FINAL
Escola pública	0,63	0,39	35,47	1196	22
Escolas particulares	0,79	0,60	36,64	258	12

Quadro 15 - REGRESSÃO (Método “Stepwise”) – RESULTADOS: 8ª série

AMOSTRA	MODELO LINEAR MÚLTIPLO	GRAU DE SIGNIFICÂNCIA DOS COEFICIENTES E DA CONSTANTE	
		t ^{cr} de Student	α
Escola pública	$Y_1 = 207,48 + 48,67Z_1 + 2,39Z_2 - 7,95Z_3 + 14,26Z_4 - 8,25Z_5 - 5,92Z_6 - 17,76Z_7 + 5,84Z_8 + 16,07Z_9 + 13,56Z_{10} + 17,72Z_{11} - 49,18Z_{12} + 7,08Z_{13} - 16,08Z_{14} - 19,42X_{15} + 6,61Z_{16} - 3,12Z_{17} - 2,81Z_{18} + 10,26Z_{19}$	[-5,63; 25,10]	[0,01; 0,02]
Escolas particulares	$Y_4 = 152,22 + 2,56Z_1 - 3,41Z_2 + 36,34Z_3 + 26,81Z_4 - 9,34Z_5 - 39,00Z_6 + 28,97Z_7 - 9,92Z_8 + 32,69Z_9 + 1,52Z_{10}$	[-4,75; 9,30]	[0,01; 0,02]

VARIÁVEIS DA ESCOLA PÚBLICA: 8ª série

- Y_1 → Proficiência do aluno em matemática
 Z_1 → d011_4 – Mestrado (0,219)
 Z_2 → a081_025 - Escolaridade da mãe (0,117)
 Z_3 → a082_004 - Já foi reprovado (-0,134)
 Z_4 → d129 - a082_010 - Gosta de estudar Matemática (0,148)
 Z_5 → p039 - A disciplina de estudantes desta escola dificulta o trabalho de ensinar (-0,089)
 Z_6 → a082_003_1 - Estudou somente em escola pública (-0,061)
 Z_7 → d129 - Conhece os resultados do SAEB (-0,122)
 Z_8 → a082_007_1 - Faz lição de casa de Matemática sempre ou quase sempre (0,064)
 Z_9 → d008_8 - Nível de escolaridade (até a graduação): Ensino Superior – outros (0,153)
 Z_{10} → d110 - Parâmetros em ação (0,142)
 Z_{11} → d011_3 - Especialização - mínimo de 360 horas (0,168)
 Z_{12} → d031_3 - Projeto Pedagógico da escola: Diretor elaborou uma proposta, apresentou-a aos professores para sugestões e, depois, escreveu a versão final. (-0,117)
 Z_{13} → a082_001_1 - Quando começou a estudar: na pré escola (0,060)

- Z₁₄ → d023_3 - Forma que assumiu a direção da escola: Seleção e eleição (-0,177)
- Z₁₅ → d023_2 - Forma que assumiu a direção da escola: Só eleição (-0,093)
- Z₁₆ → d004 - Salário bruto como diretor (a) desta escola (0,222)
- Z₁₇ → d007 - Renda bruta familiar (-0,105)
- Z₁₈ → d020 - Há quantos anos você é diretor(a) da escola (-0,066)
- Z₁₉ → d031_1 - Projeto Pedagógico da escola: Adotado o modelo encaminhado pela Secretaria de Educação 0,069

Escolas particulares: 8ª série

Y₄ → Proficiência do aluno

- Z₁ → d082_106f - Existência de equipamentos, espaços e estado de condição para uso (0,570)
- Z₂ → a082_007_2 - Faz lição de casa de Matemática de vez em quando (-0,198)
- Z₃ → p007 – Participa (ou) de alguma atividade de formação continuada nos 2 últimos anos (0,277)
- Z₄ → a082_010 - Gosta de estudar Matemática (0,231)
- Z₅ → idade_al - Idade do aluno (-0,151)
- Z₆ → d038 - A escola desenvolve algum programa de apoio ou reforço de aprendizagem (0,192)
- Z₇ → d008_6 - Nível de escolaridade (até a graduação): Ensino Superior - Licenciatura em Matemática (0,160)
- Z₈ → a081_048 - Em dia de aula, quanto tempo trabalha fora de casa (-0,107)
- Z₉ → a081_028 – Escolaridade do pai (0,100)
- Z₁₀ → p65_77f - Equipamentos e materiais disponíveis e utilizados (0,112)

5 CONSIDERAÇÕES CONCLUSIVAS

O Ministério da Educação-MEC, através do INEP, vem contribuindo para o desenvolvimento de uma cultura de avaliação nas diversas esferas da estrutura educacional do país. A partir de 1988, quando o SAEB foi implantado, o sistema de avaliação mantido pelo INEP se consolidou e desenvolveu sua capacidade técnica e operacional. Hoje, além do SAEB, dispõe de uma estrutura de avaliação que cobre os dois níveis de ensino da educação nacional.

A introdução da TRI na construção e análise dos testes, a partir da avaliação do SAEB em 1995, representou um salto qualitativo, tornando comparáveis os resultados das avaliações realizadas periodicamente, seja entre elas, seja entre as diversas partes que compõem a estrutura educacional (entre estados, entre dependências administrativas das escolas entre outros).

Também ocorreu uma evolução na qualidade dos itens, com a construção e criação do banco de itens, com alguns privilegiando as operações de raciocínio necessárias para sua resolução, tarefa para a qual a maioria das escolas ainda não atentou, pois diferentemente da escola tradicional, a grande função da escola, hoje, deve ser desenvolver a capacidade do espírito de operar as informações armazenadas através dos conteúdos aprendidos nas disciplinas dos currículos. Isso é fundamental para se atingir um dos objetivos considerados entre os mais importantes na formação do ser social: a formação para o exercício pleno da cidadania.

Quando se aponta para os resultados do SAEB, é comum ouvirem-se perguntas relacionadas ao porquê somente os conteúdos de português e matemática são avaliados. Justifica-se essa abordagem uma vez que essas duas linguagens se desenvolveram historicamente como aquelas que facilitam a comunicação entre as pessoas e sua utilização nos processos operatórios com o dados de realidade.

Finalmente, utilizando-se as avaliações do SAEB na perspectiva deste trabalho, torna-se possível identificar um conjunto de fatores que incidem significativamente sobre a aprendizagem e que melhor explicam as variações

na aprendizagem observadas, além de possibilitarem a elaboração e reformulação de políticas públicas voltadas para a melhoria da qualidade da educação brasileira.

Nos últimos anos, o Brasil deu um passo importante ao (praticamente) resolver a questão do acesso à escola: 97% dos jovens de 7 a 14 anos estão matriculados. Só que esses míseros 3% que estão longe de livros e cadernos correspondem a algo em torno de um milhão e quinhentas mil pessoas (logicamente, das camadas mais pobres da população). Por outro lado, conforme se pode verificar durante as análises realizadas no conjunto e subconjuntos das correlações realizadas o investir mais recursos em escolas e professores se constitui numa reformulação pontual pois afirmar-se que falta dinheiro na Educação brasileira não traduz bem a realidade, haja vista o montante que é investido nos sistemas, inclusive a partir do que determina o FUNDEB - Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização do Magistério, não traduz a realidade.

Por outro lado, os alunos de ensino básico da rede pública custam ao governo cerca de 12% da renda per capita nacional. Se compararmos com o investimento que é feito nos Estados Unidos, esse valor salta para 25%, mais que o dobro (sem contar que a renda per capita americana é muito maior que a brasileira). Os dados se refletem na realidade que todos conhecem: salários baixos para os professores, falta de material didático e infra-estrutura precária nas escolas, entre outros, como consequência da má aplicação desses recursos.

Outra consideração conclusiva, com relação à reformulação de políticas para educação, em decorrência deste estudo, refere-se à “valorização do trabalho dos professores” em que razões levam um sujeito a escolher uma profissão crucial para o desenvolvimento do seu país, embora não entendida dessa forma pela maioria, mas ao mesmo tempo não receber um reconhecimento compatível, traduz ou sugere baixa estima no desenvolvimento do seu trabalho, refletindo no rendimento da aprendizagem de seus alunos.

Os professores precisam de qualificação, tanto na área pedagógica como nos campos específicos do conhecimento, principalmente matemática e

português. A formação inicial deve, portanto, passar por reformulações profundas, o que implica garantir ao profissional um conhecimento mínimo básico para que ela tenha bom desempenho. Para tanto, é necessário, além da questão salarial, desenvolver programas que o habilitem a pesquisar, refletir, além de unir teoria e prática de maneira satisfatória. Só assim será possível aplicar, de fato, o que prescrevem os Parâmetros Curriculares Nacionais, que foram criados com o objetivo de estabelecer um mínimo denominador comum para os conteúdos oferecidos em sala de aula.

A obrigatoriedade do curso superior, por exemplo, não é garantia de qualidade, entretanto a pós-graduação, especialização e mestrado, duas das variáveis identificadas como significativas no estudo, precisam ser incluídas entre as ações voltadas à melhoria da qualidade do ensino, visando a qualidade da aprendizagem. É verdade que um grande número de instituições - sobretudo privadas - não oferece o mínimo necessário para formar novos educadores, no entanto, é necessário, também, criar uma política de certificação nacional, que mesmo não obrigatória, ofereça algum benefício aos educadores, diferenciando o profissional que estuda e investe na qualidade de seu trabalho dos demais, como em qualquer outra profissão.

Finalmente, a se estabelecerem políticas públicas necessita-se de tempo para seu amadurecimento, não se pode esperar que elas se reflitam na melhoria do ensino-aprendizagem da noite para o dia. Na verdade, pensa-se que elas têm um prazo de maturação, surgindo seus efeitos médio e/ou longo prazo.

Os desafios da Educação demandam tempo maior para serem resolvidos do que os quatro anos de mandato de prefeitos, governadores e do presidente da República. Todos os países que conseguiram vencer os problemas de acesso às escolas e de qualidade do ensino mantiveram uma agenda de continuidade de longo prazo.

BIBLIOGRAFIA

BERGER, Peter L.; LUCKMANN, Thomas. **A construção social da realidade**. 13. ed. Petrópolis: Vozes, 1996.

BAKER, B. O.; HARDYCK, C. D. & PETRINOVICH, L.F. *Weak measurements vs. strong statistics: An empirical critique of S.S. Stevens' proscriptions on statistics*. **Educational and Psychological Measurement**, 26, 291-309 Baker, Hardyck e Petrinovich; 1966.

BIRNBAUM, A. **Efficient design and use of test of a mental ability for various decision-making problems**. (Series Report No. 58-16). Washington, DC: USAF School of Aviation Medicine, 1957.

_____. **Some latent trait models and their use in inferring and examinee's ability**. In

F.M. LORD & M.R. NOVICK, **Statistical theories of mental test scores**. Reading, MA: Addison-Wesley, ch. 17-20; 1968.

BRASIL. MEC. INEP/SAEB. **Relatório Técnico 2003**. Disponível em: <http://www.inep.gov.br>.

CRONBACH LEE, J. and MEEHL, Paul E. **Construct validity in psychological tests**. First published in **Psychological Bulletin**, 52, 281-302, 1955. disponível em <http://psychclassics.yorku.ca/Cronbach/construct.htm>.

COMTE, Auguste. **Cours de philosophie positive**. Paris: Borrani et Droz Libraires, 1852.

DESCARTE, René. **Discurso do método**. São Paulo: Martins Fontes. 1999.

GUBA, E. **The paradigm dialog**. Newbury Park, Califórnia: Sage, 1990. Hegenberg. Rio de Janeiro: Francisco Alves, 1977.

FERNÁNDEZ, José Muñiz. **Teoría de respuesta a los ítems. Un nuevo enfoque en la evolución psicológica y educativa**. Madrid: Ediciones Pirámide S.A.1990.

_____. **Teoría Clásica de los Test**. Madrid: Pirámide. 1992.

FEYERABEND D, P. **Contra o método**. Tradução de Octanny S. da Mota e Leonidas

FLETCHER, P. R. A **Teoria de resposta ao item: medidas invariantes do desempenho escolar**. Ensaio. Rio de Janeiro, v.1, n. 2, p.21-8, 1994.

GARCIA MORENTE, M. **Fundamentos da filosofia**. São Paulo: Mestre Jou, 1967.

GALILEU Galilei. *Coleção Os pensadores Vol XII. – O ensaiador*; São Paulo: Editora Abril Cultural S.A. 1973.

GAVIRIA Soto, José Luis. **Teoría de resposta ao item: breve introducción a la psicometria. Principales teorías.** Brasília: CESPE/UNBB, 2001.

GLINER, J. A. & MORGAN, G.A. **Research methods in applied settings: An integrated approach to design and analysis.** Mahwah, N.J.: Lawrence Erlbaum Associates, 2000.

HACKING, I. **Representing and intervening: introductory topics in the philosophy of natural science.** New York: Cambridge University Press. 1983.
HAYS, W.L. **Statistics for Psychologists.** New York: Holt Rinehart and Winston, 1963.

Hegel G. W. F. *Enciclopédia das ciências filosóficas. Vol I; A ciência da lógica.* São Paulo: Edições Loyola, 1995.

HEMPEL, C. G. **Aspects of Scientific Explanation and the Others Essays in the Philosophy of Science.** New York: The Free Press, 1965

Hoyle, R.H. **Structural Equation Modeling: Concepts, Issues and Applications.** Thousand Oaks: Sage, 1995.

HUME, D. **Investigação sobre o Entendimento Humano**; trad. Leonel Vallandro. - São Paulo: Abril Cultural, 1980.

KEEVES, J.P. (Ed.). **Educational research, methodology, and measurement. An international handbook.** Oxford. Pergamon press, 1997.

KELLEY, David. **The Art of Reasoning in the Books**, New York: Norton & Co. 1998.

KOLAKOWISKI, Leszek. **Positivist Philosophy – From Hume to the Vienna Circle.** 1972.

KANT, Immanuel. **Crítica da razão pura.** Lisboa: Fundação Calouste Gulbenkian. 1985.

KUHN. Thomas S. **A estrutura das revoluções científicas.** São Paulo: Perspectivas S.A. 1975.

LORD, F.M. **The relation of the reliability of multiple-choice tests to the distribution or item difficulties.** *Psychometrika*, 17, 181-194. 1952.
The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13, 517-549. 1953.

LORD, F.M. & NOVICK, M.R. **Statistical theories of mental test scores.** Reading, MA: Addison-Wesley. 1968.

LANGENBACH, M., VAUGHN, C., & AAGAARD, L. **An introduction to educational research.** Boston: Allyn and Bacon. 1994.

LAUDAN L. Progress and its problems : toward a theory of scientific growth; Berkeley, CA: University of California Press. 1977.

_____. Beyond positivism and relativism: theory, method and evidence. Bold,CO: Westview Press.

LAWLEY, D.N. **The factorial analysis of multiple item tests.** *Proceedings of the Royal Society of Edinburgh, 62-A, 74-82.* 1944.

MCNEMAR, Q. **Psychological statistics** (3rd ed.). New York: Wiley. 1962.

MERRIAM, S.B. **Qualitative research and case study: applications in education.** San Francisco, CA: Jossey-Bass, 1998.

MARX. Karl. **Contribuição à Crítica da Economia Política.** São Paulo. Martins Fontes. 2003.

_____. **Introdução à Contribuição para a Crítica da Economia Política,** Cáp III – O método da economia política. 1859.

NATION, J.R. **Research methods.** Upper Saddle River, NJ: Prentice Hall, 1997.

RUSSELL, Bertrand. **Principles of mathematics.** New York: Norton, 1938

RICHARDSON, M.W. *Notes on the rationale of item analysis.* **Psychometrika,** 1, 69-76. 1936.

----- *The relation between the difficulty and the differential validity of a test.* **Psychometrika,** 1, 33-49. 1936.

SAVIANI, D. **Educação Brasileira: Estrutura e Sistema.** 8 ed. Campinas: Autores Associados, 2000. 189 p.

SALVUCCI, S; WALTER, E.; COLEY, V.; FINK, S; e MSABA, M. **Measurement error studies at the National Center for Education Statistics.** Washington D.C. Department of Education, 1997.

SANTISTEBAN Carmen R. **Psicometria,** Madrid: Ediciones Norma, 1990.

STEVENS S.S. **On the theory of scales of Measurement,** In: W. Mehrens e Robert Ebel, R. L. e D. A. Frisbie. *Essentials of Educational Measurement.* Chicago: Rand McNally & Compang. 1967.

TORGERSON, Warren S. **Theory and methods of scaling**. New York: John Wiley & Sons, 1958.

TORANZOS, Fausto I. **Estatística**. São Paulo: Ed. Mestre Jou; 1962.

TUCKER, L.R. *Maximum validity of a test with equivalent items*. **Psychometrika**, 11, 1-13. 1946.

YU, Chong Ho. **Misconceived relationships between logical positivism and quantitative research**. Seattle: American Research Association, 2002.

ZETTERBERG, Hans. **Teoria y Verificación en Sociología**, Buenos Aires, Nueva Visión, 1970.