

AVALIAÇÃO E O AVALIADOR EDUCACIONAL: DEPOIMENTO

HERALDO MARELIM VIANNA¹

Uma experiência só faz sua irrupção quando está sendo dita. E se não for dita é, por assim dizer, não-existente.

Hannah Arendt²

O presente trabalho visa a definir nossa posição em relação a alguns problemas de avaliação educacional. Parece-nos impossível uma discussão da totalidade das questões ligadas à avaliação educacional; desse modo, a argumentação será feita a partir de nossa vivência nessa área, especialmente no período que vai de 1962, quando, mais ou menos formalmente, iniciamos atividades educacionais diretamente ligadas à avaliação, ao ano de 1994, em que começamos a consignar nossas reflexões sobre o assunto, com vistas a registrar a compreensão que possuímos dos diversos temas enfocados.

Ao término de nossa formação acadêmica para o magistério (1952), depois de estudos sobre áreas nem sempre relacionadas com a realidade do sistema educacional brasileiro, incluindo, também, matérias pedagógicas, tínhamos uma noção bastante restrita do campo da avaliação, limitada, especificamente, ao

¹ Pesquisador-Senior da Fundação Carlos Chagas, São Paulo – SP.

² Carta de Hannah Arendt a Mary McCarthy, em 31/05/71. In: **Entre Amigas — A correspondência de Hannah Arendt e Mary McCarthy (1949-1975)**. Rio de Janeiro: Delume Dumará, 1995.

rendimento escolar. O assunto, apresentado como um simples tópico da didática geral, era transmitido no curto espaço de uma aula de cinquenta minutos, ao ser analisada a questão da verificação da aprendizagem. Uma discussão bastante esquematizada da avaliação formal e informal era oferecida sem uma análise aprofundada dos seus fundamentos e de suas implicações, limitando-se, assim, a noções sumárias e gerais sobre como proceder em sala de aula, usando reproduções simplificadas de experiências supostamente consagradas pela tradição pedagógica. Após dez anos de atividades de magistério, em diferentes níveis de ensino, vimo-nos diante de um curso formal de avaliação educacional (1962), ao qual chegamos com algumas mudanças em nossos posicionamentos em decorrência do ensaio de Ebel³ (1951) e, especialmente, dos vários trabalhos da obra coletiva organizada e editada por Lindquist⁴ (1951), que nos deram uma visão mais consistente da avaliação, segundo uma perspectiva quantitativa. Essa abordagem se consolidou com um curso em 1962, ministrado a partir do manual de avaliação da Força Aérea Norte-Americana (USAF), que nos colocou em contacto com o pensamento psicométrico, que, no nosso caso, se desenvolveria, inicialmente, com base no pensamento de Guilford⁵ (1946) e Flanagan⁶ (1951), autores de repercussão em diferentes setores da sociedade educacional norte-americana. O curso por nós freqüentado não foi, efetivamente, de avaliação educacional, em seu sentido mais amplo, concentrou-se, apenas, em testes e medidas, especialmente na construção de instrumentos, na discussão de tipos de validade e, sobretudo, na questão da fidedignidade, com seus diferentes métodos de cálculo. Fomos, assim, introduzidos nos fundamentos estatísticos das medidas, o que nos levou a procurar novos conhecimentos

³ EBEL, R.L. Writing the test item. In: LINDQUIST, E. F. (Ed.) **Educational Measurement**. Washington, D. C. American Council on Education. 1951.

⁴ LINDQUIST, E. F. (Ed.) **Educational Measurement**. Washington, D.C. American Council on Education. 1951

⁵ GUILFORD J. P. New standards for test evaluation. **Educational and Psychological Measurement**, 6, 4, 1946, p. 432 e segs.

⁶ FLANAGAN, J. C. The use of comprehensive rationales in test development. **Educational and Psychological Measurement**, 11, 1951, p. 151 e segs.

nessa área por intermédio de Garrett⁷ (1962) e, posteriormente, utilizando a obra de Guilford⁸ (1965), entre outras, num esforço de autodidatismo, como ocorre com bastante freqüência em nosso contexto educacional, particularmente em assuntos relacionados com a avaliação. Observamos que, decorridos mais de 30 anos da realização desse curso, sempre que se fala em curso/seminário/treinamento sobre avaliação o entendimento é que o mesmo incidirá, obrigatoriamente, em tecnologia da construção de questões objetivas e organização de testes, o que reflete uma falsa concepção do que seja avaliação.

A partir desse curso, assumimos a responsabilidade de orientar um programa de avaliação em uma academia militar que ministrava cursos em nível de 2º grau. Alguma coisa aconteceu: as avaliações, abrangendo todas as áreas do Ensino Médio, passaram a ser mensais e não mais bimestrais, procuramos capacitar professores na construção de itens, introduzimos, ainda que de forma precária, por falta de **hard/software**, um sistema de análise estatística das questões, segundo a teoria clássica, e os resultados das provas (testes objetivos) passaram a ser apresentados sob a forma de escores padronizados em função do desempenho do grupo, por intermédio de uma escala de **estatinos**, que foi criada por Flanagan durante a Segunda Guerra Mundial (1939-1945), a partir da curva normal — modelo esse que, hoje reconhecemos, causou enormes malefícios às ciências humanas, inclusive à educação. A metodologia do seu cálculo foi a apresentada por Durost e Prescott (1966)⁹, a qual, por sua simplicidade, permitia que alunos e professores situassem os desempenhos de cada sujeito em relação ao grupo total. Isso representava um avanço, ainda que seja discutível o modelo seguido, baseado em uma curva de probabilidade para fenômenos aleatórios e relativos a um grande número de indivíduos.

⁷ GARRETT, H. E. *A Estatística na Psicologia e na Educação*. Trad. Eva Nick. Rio de Janeiro. Fundo de Cultura (2 vols.). 1946 .

⁸ GUILFORD., J. P. *Fundamental Statistics in Psychology and Education*. 4th Edition. New York. McGraw-Hill Book Co., 1965.

⁹ DUROST, W. N. e PRESCOTT, G. A. *Essentials of measurement for teachers*. New York. Harcourt, Brace, and World. 1962.

A Fundação Getúlio Vargas, no Rio de Janeiro, possuía um centro de estudos de avaliação, sob a supervisão de Ruth Schaeffer e a orientação técnica de Nícia Maria Bessa, e nessa época (1965) promoveu a vinda de especialistas de fama mundial para que transmitissem suas experiências em avaliação. Um deles, Frederick B. Davis¹⁰, ministrou curso sobre medidas, do qual participamos, que seguiu orientação tipicamente norte-americana: — testes padronizados, seu uso e interpretação dos resultados. Desta vez, entretanto, foi dado destaque a fórmulas punitivas para a tentativa de acerto casual, muitas vezes apresentadas por intermédio de sofisticação matemática, mas que em nada contribuem para a melhoria do processo de avaliação, como hoje em dia é reconhecido. Esse assunto, que parece preocupar a tantas pessoas não inteiramente identificadas com a mensuração educacional, constou de um amplo ensaio escrito pelo próprio Davis para o **Educational Measurement** na edição de 1951, organizada por Lindquist. Esse tema — punição para a tentativa de acerto casual — foi suprimido da edição seguinte, sob responsabilidade de Thorndike (1971)¹¹, por ser inteiramente irrelevante, na nossa opinião. A presença nessa mesma época, na Fundação Getúlio Vargas, de Anne Anastasi, cujo livro básico (1968)¹² teria importância em nossa formação teórica, no final de 1969, e o desenvolvimento por Nícia M. Bessa, em meados da década de 60, de um teste inspirado no **Iowa Basic Skills**, para ser utilizado no então Estado da Guanabara, foram fundamentais para a nossa compreensão do processo de medida, especialmente no que diz respeito à complexidade das chamadas habilidades básicas de crianças da Escola Fundamental.

Ao mesmo tempo que adquirimos **expertise** nessa área, no ano de 1965 procuramos socializar esses conhecimentos através de cursos de curta duração ministrados na Secretaria de Estado da Educação do Estado de São Paulo e em diversas instituições do

¹⁰ DAVIS, F. B. **Educational measurement and their interpretation**. Belmont, Calif. Wadsworth Publ. Co. Inc. 1964.

¹¹ THORNDIKE, R. L. (Ed.) **Educational Measurement** Washington, DC. American Council on Education, 1971.

¹² ANASTASI, A. **Psychological Testing**. Third Edition. New York. The MacMillan Co. 1968

Estado de Minas Gerais, orientando professores na elaboração de instrumentos — objetivos/não-objetivos — para uma avaliação do rendimento escolar que considerasse inclusive suas implicações sociais — repetência/evasão —, evitando, assim, a realização de um trabalho quase sempre precário e muitas vezes destituído de fundamentação teórica. As intensas atividades nessa área, em diferentes regiões, no período de 1965-67, evidenciaram que a situação não diferia da que conhecêramos há 15 anos, ao terminarmos o curso de bacharelado exigido para a prática do magistério.

O período seguinte, abrangendo os anos de 1967 a 1969, foi marcante em virtude das experiências vivenciadas no exterior, especialmente em Universidades norte-americanas e, em menor grau, em instituições francesas. Se já conhecíamos alguns trabalhos fundamentais da bibliografia norte-americana, passamos a observar mais detalhadamente a prática da avaliação em centros universitários. A **Universidade de Michigan** (Ann Arbor, Mich.), por intermédio do **English Language Institute** (ELI), em 1967, e sob a orientação de John Upshur, proporcionaria um trabalho de especialização em medidas do domínio do inglês como segunda língua para estrangeiros. A partir do trabalho de Thorndike e Hagen (1961)¹³, iniciamos estudos para uma maior fundamentação estatística dos instrumentos de medida, graças à obra de Thorndike (1949)¹⁴, em que aborda grande variedade de problemas psicométricos ligados à teoria clássica das medidas, especialmente à questão da validade preditiva, assunto raramente considerado em nosso contexto, inclusive no processo de seleção de recursos humanos para a Universidade. A experiência de Ann Arbor, junto ao ELI, serviu para mostrar que é possível dominar conceitos estatísticos básicos sem um envolvimento mais aprofundado da análise matemática; contudo, para a compreensão da moderna teoria dos testes e de sua fundamentação estatística, uma formação quantitativa é realmente

¹³ THORNDIKE, R. L. e HAGEN, E. **Measurement and Evaluation in Psychology and Education**. New York. John Wiley and Sons. 1961.

¹⁴ THORNDIKE, R. L. **Personnel Selection-Test and Measurement Techniques**. New York. John Wiley and Sons, Inc. 1949.

indispensável, a fim de compreender os diferentes modelos matemáticos utilizados no estudo das características humanas.

O ano de 1968 foi rico de novas experiências, que decorreram de estágio no **Centre International d'Études Pédagogiques** (Sèvres, Paris) e em um liceu-piloto na cidade de Toulouse, em situações bastante diferenciadas das que foram experimentadas no contexto norte-americano. Ambas as experiências foram, entretanto, igualmente válidas para a nossa formação na área da avaliação, ao longo de um processo que ainda se desenvolve a cada experiência vivenciada, geradora de conhecimentos que sempre se expandem e renovam. Assim, o avaliador, no dia-a-dia das suas atividades profissionais, vive contínua construção do conhecimento. A experiência de Sèvres foi válida porque possibilitou acesso a outros centros educacionais e permitiu entrar em contato com especialistas ligados à análise quantitativa de características humanas. Sèvres deu-nos a oportunidade de conhecer Miallaret e Pham (1967)¹⁵ e Barbut (1967)¹⁶, que nos levaram a uma reflexão sobre o alto nível do preparo exigido na formação dos professores. Ambos os livros fogem à orientação norte-americana, pois demandam conhecimentos de matemática avançada, como álgebra de matrizes, fazendo, assim, com que essas obras, destinadas a educadores, tenham, no nosso contexto educacional, uma audiência bastante restrita, sobretudo no caso da obra de Barbut, de grande importância para a compreensão do fundamento matemático de certos modelos usados em educação. Os livros desses autores serviram para uma melhor compreensão do significado da linguagem matemática na análise de problemas das ciências humanas e anteciparam, por outro lado, situações difíceis com que nos depararíamos, ao longo dos anos, à medida que novas perspectivas de conhecimento técnico surgiam em nossa vida profissional. A experiência de Toulouse, ao estabelecermos contatos com a cultura pedagógica dos liceus franceses, na parte

¹⁵ MIALLARET, G. e PHAM, D. *Statistique a l'Usage des Éducateurs*. Paris. Presses Universitaires de France. 1967.

¹⁶ BARBUT, M. *Mathématiques des Sciences Humaines*. 2 vols. Paris. Presses Universitaires de France. 1967.

relativa à avaliação, mostrou a independência acadêmica dos professores, a sua capacidade de decidir sobre o futuro dos seus alunos, a severidade — muitas vezes exagerada — dos seus exames, que podem gerar frustrações, como as decorrentes do célebre “**baccalauréat**”, que foram parcialmente responsáveis pelos acontecimentos de maio de 1968, cujo significado não foi percebido por De Gaulle, então presidente da França, que considerou tudo simplesmente um *chienlit*, mas cujas raízes penetravam na área educacional e em um sistema de avaliação que precisava ser renovado, como deve ocorrer em toda e qualquer estrutura educacional, segundo a nossa percepção.

Ainda no final de 1968, tivemos uma seqüência de novas experiências que serviram para consolidar antigas vivências e proporcionar novas visões no campo da avaliação educacional, a partir da realização do Mestrado na **Michigan State University** (East Lansing, Mich.), sob a orientação de Robert L. Ebel. As disciplinas do *major* — **Educational Research Methods, Testing and Grading, Problems of Measurement, Quantitative Methods, Standardized Tests, Advanced Quantitative Methods, Principles of Measurement e Psychological Testing** — foram a base para o domínio da parte substantiva da avaliação educacional, segundo uma perspectiva eminentemente quantitativa, e que foi complementada por um *minor* que fundamentou essa prática, por intermédio de outros cursos, e lhe deu significado: **School Learning, Growth and Behavior, Philosophy of Education e Problems of Higher Education**. A experiência de uma universidade norte-americana é marcante e, no nosso caso pessoal, sentimos que, apesar de altamente competitiva e **estressante**, exerceu um papel formativo e consolidou antigos conhecimentos. Foi importante, neste novo contexto, o papel de Robert L. Ebel, que nos mostrou que avaliação não é apenas análise estatística, a partir de instrumentos construídos segundo os princípios da tecnologia, que oferecem resultados fidedignos, mas uma atividade que envolve seres humanos e pode ter influência sobre seus destinos, no plano da realização pessoal e profissional. Foi a partir desse momento que começamos a nos preocupar mais seriamente com certos conceitos carregados de abstração, mas

que constituem o cerne de toda a avaliação: valores, critérios, objetivos, normas, significância prática, entre outros. A própria avaliação pareceu-nos um conceito abstrato, como reconhecem Madaus, Scriven e Stufflebean (1993)¹⁷.

A universidade norte-americana é **book-oriented**, exigindo um grande volume de leituras em tempo relativamente reduzido. Em alguns casos, a leitura de dois livros semanais era uma situação comum; por outro lado, um número igualmente grande de artigos era incluído entre as **assigned readings**, aspecto que, no nosso caso, merece destaque, por explicar grande parte das influências recebidas durante nossa formação profissional. Em um dos cursos de Robert L. Ebel — **Problems of Measurement** — tivemos oportunidade de ler alguns autores clássicos na área de medidas, como S.S. Stevens, B. O. Baker, R. M. W. Travers, e F. M. Lord (**Teoria das medidas**); E. F. Gardner, R. L. Ebel, F. M. Lord, W. H. Angoff, E. F. Lindquist, R. T. Lenon (**Normas**); G. F. Kuder e M. W. Richardson, P. J. Rulon, C. J. Hoyt, R. L. Ebel, L. J. Cronbach, E. D. Cureton, P. Horst, F. M. Lord (**Fidedignidade**); C. I. Mosier, R. L. Ebel, H. O. Gulliksen, L. J. Cronbach e P. E. Meehl, E.D. Cureton, D. T. Campbell e D. W. Fiske, L. Sechrist, P. E. Meehl e A. Rosen, R. B. Cattell (**Validade**); J. C. Flanagan, A. P. Johnson, W. G. Findley, M. D. Engelhart, M. W. Richardson, O. K. Buros e L. J. Cronbach (**Análise de Itens**), cujos artigos foram publicados na antologia organizada por Mehrens e Ebel¹⁸. A excelência deste material consolidou conceitos e deu-nos um embasamento teórico, além, naturalmente, do domínio de um instrumental necessário ao trabalho que desenvolveríamos após nosso regresso ao Brasil, o que realmente ocorreu, especialmente no período de 1970-86.

Ao discutirmos aspectos de nossa formação acadêmica, especialmente em relação a influências teóricas, não poderíamos omitir o curso de W. A. Mehrens, que adotou como texto básico

¹⁷ MADAUS, G. F.; SCRIVEN, M. S. e STUFFLEBEAM, D. F. **Evaluation Models – Viewpoint on Educational and Human Services Evaluation**. Boston, Kluwer-Nijhoff Publishing. 1993.

¹⁸ MEHRENS, W. A. e EBEL, R. L. **Principles of Educational and Psychological Measurement, a book of selected readings**. Rand McNally and Co. Chicago, Ill. 1967.

o livro de D. Magnusson, professor na Universidade de Estocolmo¹⁹. A obra oferece uma visão bastante lúcida da teoria clássica dos testes, além de ser uma excelente revisão da estatística dos testes, em um nível de matematização suportável. Outras leituras foram igualmente exigidas, conforme a tradição universitária norte-americana, obrigando-nos a uma imersão nas obras de J. P. Guilford, G. A. Ferguson, H. Gulliksen, E. E. Ghiselli, R. L. Thorndike, L. J. Cronbach, F. M. Lord, A. Anastasi e Q. N. Nemar. Houve, portanto, um grande entrosamento entre os cursos ministrados por Ebel e Mehrens, contribuindo, assim, tendo em vista a sua natureza quantitativa, para que nos aprofundássemos na psicometria associada à teoria clássica dos instrumentos de medida, segundo nosso planejamento inicial e a expectativa da instituição brasileira a que nos ligáramos (**Fundação Carlos Chagas**), quando regressássemos do exterior.

Ainda em relação à nossa vivência em uma instituição norte-americana, com auxílio da Fundação Ford, gostaríamos de acentuar um aspecto importante ligado ao trabalho prático, sempre exigido, aos exercícios constantes, muitas vezes diários, e cobrados para correção e comentários, e as exposições orais, seguidas de debates, com avaliações, inclusive a controversa **peer evaluation**, em alguns casos. A realização de trabalhos, apresentados aos professores em **blue books** (cadernos padronizados para a realização de exercícios, comuns nas escolas americanas, inclusive na **high school** e nos **colleges**), cria, naturalmente, um ambiente de pressão e competitividade, sobretudo considerando que muitas avaliações tinham os resultados expressos **on the curve**, ou seja, em função do desempenho do grupo. Tudo isso concorre para alimentar o caldo de cultura de um dos elementos mais caros à sociedade americana, a valorização do desempenho e o constante provar que, mesmo não sendo o primeiro, a pessoa se situa no **top**, entre os melhores, os mais bem dotados e sucedidos, revelando, assim, um outro aspecto da vida social e intelectual da própria universidade: – o culto da meritocracia. A experiência, entretanto,

¹⁹ MAGNUSSON, D. **Test Theory**. Addison-Wesley Publishing Co. Reading, Mass. 1967.

foi extremamente válida e a esse ambiente voltamos em diferentes momentos, que, admitimos, foram sempre enriquecedores, quantitativa e qualitativamente, em nossas atividades profissionais.

Após a experiência universitária norte-americana, e sem uma fase de transição adaptativa à nossa cultura, iniciamos atividades na Fundação Carlos Chagas no final de 1969, centradas na seleção para a Universidade e na de recursos humanos qualificados, especialmente para agências governamentais. Enfrentamos uma situação inédita, reflexo da massificação do ensino e o conseqüente afluxo de grande número de estudantes às portas da universidade em busca não apenas do saber, mas sobretudo de uma qualificação profissional. Ao mesmo tempo que procurávamos socializar conhecimentos e a **expertise** desenvolvidos nos Estados Unidos, tínhamos que defender posições em uma controvérsia inteiramente sem sentido: a falsa dicotomia prova discursiva **versus** prova objetiva.

Procuramos desenvolver competência, inclusive entre professores universitários, na construção de itens e no planejamento e montagem de provas objetivas, que eram o instrumento adequado para enfrentar uma situação de exame de massa. Ainda que ataques a esse procedimento fossem originários de segmentos conservadores da comunidade acadêmica, havia repercussões na sociedade, que acreditava na argumentação nem sempre consistente das "autoridades". Isso levou-nos a escrever artigos para divulgação na mídia²⁰, tentando equacionar o problema diante do quadro revelado pela alta relação candidato/vaga no acesso ao ensino de 3º Grau, e a redigir trabalhos de caráter técnico para as instituições que realizavam seus exames por intermédio da Fundação Carlos Chagas e para a própria comunidade acadêmica²¹.

²⁰ Acerto casual em provas objetivas. **O GLOBO**, 04/09/73. A Seleção de Candidatos através de Provas Objetivas. **Folha de São Paulo**, 25/12/73. O que a prova de redação realmente mede? **Folha de São Paulo**, 19/10/75.

²¹ Emprego e características de provas objetivas. **Ciência e Cultura**, vol. 22, nº 3, 1970. Os vestibulares refletem toda a problemática da educação. **Mundo Econômico**, vol. IV, nº 5, 1971.

As atividades na área psicométrica intensificaram-se a partir do final de 71, com o aumento do número de instituições que passaram a integrar a Fundação Carlos Chagas, exigindo o deslocamento para diversos pontos do território nacional, especialmente no Nordeste e no extremo Sul, a fim de participar de treinamentos e realizar cursos regulares, em nível de graduação (Faculdade de Filosofia, Ciências e Letras de Araraquara, atualmente integrando a Universidade Estadual de São Paulo) e em nível de pós-graduação, na Universidade Federal do Rio de Janeiro, na área de medidas educacionais, o que resultou na publicação de um livro sobre o assunto²², posteriormente editado, também, em espanhol.

A preocupação maior em **Testes em Educação** foi com aspectos ligados ao planejamento dos instrumentos de medida, à redação de objetivos instrucionais operacionais, à tecnologia da construção de questões objetivas e discursivas, ao problema da validade (curricular e preditiva), às questões ligadas à fidedignidade dos resultados e à análise de itens. Tudo isso em linguagem acessível a leitores sem maiores experiências com a estatística dos testes. Aos poucos, o problema da medida da capacidade de expressão escrita começou a adquirir uma dimensão maior, inclusive com o apoio do Ministério da Educação, que, a partir de 1975, oficializou a sua utilização nos exames de acesso à Universidade, com vistas a “resguardar” a língua nacional, problema analisado por alguns setores à luz da segurança nacional. Não nos esqueçamos de que, à época, vivíamos em pleno regime militar (1964-85).

²² **Testes em Educação**. Editora IBRASA. São Paulo. 1973. **Los Testes en la Educación**. Ediciones Universidad de Navarra SA. EUNSA. Pamplona, Espanha.1983.

Ainda que convencidos da eficiência das provas objetivas na seleção de grande número de estudantes, passamos a nos preocupar com o problema da medida da expressão escrita²³, analisando experiências realizadas no exterior, especialmente no **Educational Testing Service - ETS**, com particular destaque para o trabalho de Godshalk, Swineford e Coffman²⁴, realizando estudos empíricos sobre a fidedignidade dos corretores e dos resultados, inclusive usando vários critérios, determinando a validade preditiva e concorrente de provas objetivas na medida da expressão escrita, e verificando, entre outros aspectos, as diversas contribuições da pesquisa educacional para a compreensão dessa medida em situação de exame de massa, como é o rito de passagem do "vestibular". Todo o material escrito (artigos e pesquisas) foi posteriormente publicado em forma de livro²⁵.

Simultaneamente, outras atividades foram igualmente realizadas, como a análise do possível impacto dos testes sobre o sistema educacional brasileiro²⁶, tendo em vista a afirmação, aliás não confirmada empiricamente, de que o exame vestibular estaria

²³ Medida da Expressão Escrita. **Didata**, nº 4, 1976. Redação e medida da expressão escrita: algumas contribuições da pesquisa educacional. **Cadernos de Pesquisa**. São Paulo, Fundação Carlos Chagas, nº 16. Flutuação de julgamentos em provas de redação. **Cadernos de Pesquisa**. São Paulo, Fundação Carlos Chagas, nº 19, 1976. Aplicação de critérios de correção em provas de redação. **Cadernos de Pesquisa**. São Paulo, Fundação Carlos Chagas, nº 26, 1978. Medida da Expressão Escrita e Prova Objetiva: um estudo preliminar de validade. **Cadernos de Pesquisa**. São Paulo, Fundação Carlos Chagas, nº 38, 1981. Comunicação e Expressão no acesso à Universidade: uma experiência diversificada. **Educação e Seleção**. São Paulo, Fundação Carlos Chagas, nº 4, 1981. Validade de conteúdo de uma prova de Comunicação e Expressão: análise de alguns problemas. **Educação e Seleção**. São Paulo, Fundação Carlos Chagas, nº 4, 1981. Redação e Medida da Expressão Escrita: algumas contribuições da pesquisa educacional. **Educação e Seleção**. São Paulo, Fundação Carlos Chagas, nº 6, 1982. Dupla correção em provas de redação. In: **Comunicação e Expressão**. IBRASA. São Paulo, 1983. Provas e Testes no Concurso Vestibular. **Educação e Seleção**. São Paulo, Fundação Carlos Chagas, nº 12, 1985.

²⁴ Godshalk, F. L.; Swineford, F. e Coffman, W. E. **The measurement of writing ability**. College Entrance Examination Board. New York. 1966.

²⁵ **Comunicação e Expressão — problemas teóricos e práticos de avaliação**. IBRASA. São Paulo. 1983.

²⁶ Impacto dos testes sobre os sistemas e objetivos educacionais: a experiência brasileira. **Cadernos de Pesquisa**. São Paulo, Fundação Carlos Chagas, nº 27, 1978, mais tarde editado por Dockrell, W.B. In: **The Impact of Tests in Education**. IAEA. Princeton, New Jersey, 1980.

moldando, negativamente, o sistema de ensino no Brasil. Além disso, preocupava-nos, sobretudo, a falta de formação técnica dos professores em medidas educacionais, motivando um **paper** que foi apresentado em seminário internacional no Rio de Janeiro, em 1978²⁷.

O problema do acesso ao ensino superior é recorrente na educação brasileira; desse modo, no final da década de 70, voltou a ser discutido e processos alternativos foram propostos. Velhas questões — provas objetivas versus provas dissertativas — continuaram a ser discutidas **ad nauseam**²⁸ sem nenhuma comprovação, mas apenas com base em opiniões pessoais, idiosincrasias ou vieses político-acadêmicos. O mesmo assunto voltaria a ser discutido em 1986 e, novamente, tornaria a ser discutido em 1995, sem maiores conseqüências práticas. Ainda em 1986, analisamos o problema²⁹ e voltamos ao assunto em um longo ensaio sobre as origens do vestibular (1911), sua história, sua legislação, suas inovações e retrocessos — o título do trabalho reflete a nossa posição sobre o assunto (**Acesso à Universidade — caminhos da perplexidade**), depois do que passamos a nos dedicar quase inteiramente à avaliação educacional: inicialmente, na área do rendimento escolar; mais tarde, em estudos sobre aptidões e, a seguir, na avaliação de programas e sistemas de ensino.

A intensa atividade na área de seleção não nos impediu de refletir sobre outros aspectos da avaliação, especialmente sobre a avaliação formativa, para usar a feliz expressão cunhada por M.

²⁷ Development of Technical Competence of Teachers in Educational Measurement (paper). **International Council on Education for Teaching**. Rio de Janeiro, 1978.

²⁸ Processos alternativos de seleção para ingresso no ensino superior. **Cadernos de Pesquisa**. São Paulo. Fundação Carlos Chagas, nº 34, 1980. Acesso à Universidade — reflexão sobre problemas atuais. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 1, 1980.

²⁹ Acesso à Universidade — análise de alguns modelos alternativos de seleção. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 13, 1986. Acesso à Universidade — caminhos da perplexidade. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 14, 1986. Acesso à Universidade — um estudo de validade. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 15, 1987. Acesso à Universidade — uma reflexão ao longo do tempo. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 18, 1988.

Scriven, conforme veremos na exposição sobre a lógica da avaliação. O grande drama do ensino brasileiro está na reprovação em várias séries do 1º Grau, como demonstrou Sérgio Costa Ribeiro em vários de seus trabalhos, alguns dos quais editados por nós nas revistas **Educação e Seleção** (1980-89) e **Estudos em Avaliação Educacional** (1990), da Fundação Carlos Chagas. As medidas referenciadas a critério, revigoradas a partir dos trabalhos de Glaser (1963) e por influência de Ralph W. Tyler, pareceram-nos o caminho adequado para evitar a situação constrangedora da reprovação, tendo em vista que esse tipo de medida exige uma instrução individualizada e uma avaliação formativa. Chegamos a propor o uso de um coeficiente de verificação da sensibilidade ao processo instrucional³⁰, mas a proposta não teve maior repercussão junto àqueles que poderiam implementar o projeto: — os professores, por razões óbvias, especialmente falta de capacitação na área de medidas e em avaliação.

O ano de 1980 foi rico em experiências pessoais, sobretudo pela participação em um programa internacional, coordenado por William Turnbull, ex-presidente do **Educational Testing Service – ETS**, com o envolvimento da **American University of Cairo**, **The Hong Kong Examinations Authority** e a Fundação Carlos Chagas³¹ para construir um instrumento destinado a avaliar a aptidão numérica e verbal de candidatos à pós-graduação no Brasil e no exterior. Seria um instrumento em quatro línguas: — português, inglês, árabe e chinês, em versões tecnicamente equivalentes e que se inspirariam nos modelos do SAT (**Scholastic Aptitude Test**) e do GRE (**Graduate Record Examination**), sem que fossem, entretanto, uma simples reprodução desses instrumentos de comprovada validade preditiva. O projeto, infelizmente, após a pré-testagem dos instrumentos, decorridos quase dois anos de intensos trabalhos,

³⁰ A perspectiva das medidas referenciadas a critério. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 2, 1980. Medidas referenciadas a critério: — uma introdução. In: **A construção do projeto de ensino e avaliação**. Fundação para o Desenvolvimento da Educação - FDE. São Paulo. 1990.

³¹ Seleção para programas de pós-graduação — um projeto transnacional. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas. nº 2, 1980.

entrou em colapso, como decorrência do falecimento de seu coordenador e principal elemento de ligação com as agências financiadoras.

A participação neste projeto internacional e os trabalhos na área de seleção de recursos humanos evidenciaram a falta de elementos com **expertise** suficiente para o desenvolvimento de projetos nessas áreas, levando-nos a abordar o assunto³², que é de grande complexidade, tendo em vista a inexistência de centros especializados — a avaliação educacional, infelizmente, não é área de habilitação em nossas universidades — que promovam uma formação específica para a realização de diferentes atividades que pressupõem, além da experiência docente, conhecimentos diversificados, com profundo embasamento estatístico, mesmo para a realização de estudos qualitativos, como ocorre nos grandes centros universitários, como a Universidade de Illinois (Urbana), no programa de pós-graduação dirigido por Robert Stake.

Procuramos divulgar, ainda em 1982, o pensamento de dois personagens fundamentais na evolução teórica e na prática da avaliação educacional: Tyler e Cronbach³³. O primeiro, na década de 40, lançou os fundamentos da avaliação educacional; Cronbach, preocupou-se com a fundamentação teórica de sua prática. Acreditamos que não teríamos chegado ao ponto em que nos encontramos se não fosse a colaboração desses dois cientistas sociais, complementada, mais tarde, pela contribuição de outros, como Scriven, Stufflebeam, Stake e Guba, nas suas obras bastante diversificadas.

Ao considerar o período de 1983-84, constatamos que nossas preocupações se diferenciaram consideravelmente, envolvendo problemas psicométricos relacionados com a validade de construto em testes educacionais, seguindo, assim, a

³² Avaliação educacional — problemas gerais e formação do avaliador. **Educação e Seleção**. São Paulo, Fundação Carlos Chagas, nº 5, 1982. Qualificação técnica e construção de instrumentos de medida educacional. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 10, 1984.

³³ Avaliação educacional — algumas idéias precursoras. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 6, 1982.

linha de Lee J. Cronbach, e com a validade de critério³⁴. Nesse ano de 1983, na Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ), realizou-se um seminário que nos permitiu um contacto maior com Robert E. Stake, que nos autorizou a tradução de dois de seus **papers**: um, sobre **estudo de caso**; outro, a respeito de **problemas epistemológicos** na pesquisa qualitativa/naturalista. Assim, convivemos com a quantitativo e o qualitativo sem maiores traumas, pois julgamos ser inteiramente falsa essa dicotomia que pretende opor uma à outra³⁵.

A partir de 1986, começamos a nos envolver mais diretamente com o problema da avaliação do rendimento escolar nas escolas de 1^a e 2^a graus, desenvolvendo projetos financiados pelo Instituto Nacional de Estudos e Pesquisas Educacionais — INEP³⁶, pelo Banco Mundial e pela Secretaria de Estado da Educação do Paraná. É evidente que uma avaliação abrangendo 69 cidades dispersas pelos vários Estados do Brasil acaba por mexer com a comunidade educacional, sobretudo tendo em vista a falsa concepção do caráter punitivo da avaliação. Algumas reações revelaram preocupação com aspectos técnicos, numa reação típica de cristão-novo que se apega à ortodoxia, antes objeto de contextualização pelos recém-conversos. Vimo-nos

³⁴ Validade de construto em testes educacionais. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 8, 1983. Validade de um critério de avaliação de uma peça de prova em Mecânica Geral. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 8, 1983. Natureza das medidas educacionais. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 9, 1984.

³⁵ **Robert E. Stake**. Estudo de caso em pesquisa e avaliação educacional. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 7, 1983. **Robert E. Stake**. Pesquisa qualitativa/naturalista: — questões epistemológicas. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 7, 1983.

³⁶ Avaliação do Rendimento de Alunos de Escola do 1^o Grau da Rede Pública: uma aplicação experimental em 10 cidades. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 17, 1988. Avaliação do Rendimento de Alunos de 2^o e 4^o séries de Escolas Oficiais do Estado do Paraná. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 18, 1988. Avaliação do Rendimento de Alunos de Escolas do 1^o Grau da Rede Pública: — um estudo em 20 cidades. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 19, 1989. Avaliação do Rendimento de Alunos de Escolas do 1^o Grau da Rede Pública: — um estudo em 39 cidades. **Educação e Seleção**. São Paulo. Fundação Carlos Chagas, nº 20, 1989.

obrigados, assim, a produzir um documento³⁷ discutindo objetivos, amostragem e tratamento estatístico dos dados, sobretudo a questão de saber quando **NÃO** usar certa tecnologia, especialmente tendo em vista o real destinatário da avaliação/pesquisa — o professor em sala de aula.

Ainda que não fosse intenção aumentar o nosso espaço de atividades na área da avaliação, envolvemo-nos, por breve momento, com a avaliação institucional — área pouco desenvolvida entre nós, apesar da importância dos trabalhos realizados pela Universidade Nacional de Brasília (Isaura Belloni) e pela Pontifícia Universidade Católica de São Paulo (Ana Maria Saul), entre outros. A discussão girou, sobretudo, a partir de um texto por nós elaborado sobre a quem caberia a responsabilidade dessa avaliação³⁸.

A participação em projetos internacionais levou-nos a considerar o problema da qualidade em educação³⁹, procurando, inclusive, uma forma operacional de medi-la, considerando o **contexto** em que o processo educacional se desenvolve, as **variáveis não diretamente ligadas à escola**, mas que afetam a educação, e a **ação da escola em termos de entrada, processo e produto**. O modelo apresentado procurou demonstrar que o desempenho escolar (**e a formação de atitudes**) é um dos produtos apenas, não se justificando, assim, a concentração de trabalhos de avaliação unicamente nesse aspecto, como está ocorrendo no Brasil.

Apesar da existência de um número grande de relatórios, há carência de informações sobre aspectos relacionados ao desempenho escolar no final do 2º Grau, ou melhor, os dados coletados nos chamados concursos vestibulares poderiam preencher essa lacuna, mas não são estudados, repousam no cemitério de dados dos arquivos institucionais, ou, o que é mais grave, no arquivo morto de órgãos oficiais. A necessidade de

³⁷ A prática da Avaliação Educacional: — algumas colocações metodológicas. **Cadernos de Pesquisa**, nº 69. São Paulo. Fundação Carlos Chagas. 1989.

³⁸ Avaliação Institucional: a Universidade (texto proposta para uma discussão). **Estudos em Avaliação Educacional**, nº 1, São Paulo. Fundação Carlos Chagas, 1990.

³⁹ Medida da Qualidade em Educação — apresentação de um modelo. **Estudos em Avaliação Educacional**, nº 2. São Paulo. Fundação Carlos Chagas. 1990.

informações urgentes, por solicitação do Ministério da Educação (MEC) — Secretaria de Ensino do 2º Grau, com apoio financeiro do Banco Mundial, fez com que nos envolvêssemos na avaliação de alunos de séries terminais⁴⁰, na rede pública e privada, em quatro grandes capitais, verificando as relações entre rendimento escolar e diferentes variáveis sócio-econômicas. A avaliação, entre outros aspectos, mostrou que, quando há recursos humanos qualificados, condições materiais, metodologia adequada, recursos didáticos e interesse, entre outros aspectos, — **accountability** —, é possível um ensino eficiente nas escolas públicas, tendo o estudante condições de realizar uma boa aprendizagem, como foi observado nas escolas técnicas. A avaliação, por outro lado, evidenciou a falácia do mito da excelência da escola privada, mostrando, ao contrário, que a sua suposta qualidade nem sempre é verdadeira, sendo superada em muitos casos pela escola pública, desde que bem orientada⁴¹.

No decorrer de 1991, fizemos uma reanálise das avaliações desenvolvidas e esta releitura permitiu-nos uma longa reflexão sobre os fatores determinantes da reprovação e da evasão nas primeiras séries do Ensino Básico⁴². Ao ouvirmos administradores e professores, positivamos a problemática da repetência — soma de mal-entendidos que leva muitos educadores a não se aperceberem de suas calamitosas implicações, inclusive financeiras, além, naturalmente, das psicológicas e pedagógicas, e dos malefícios que determina. Vivenciamos a experiência durante anos, na década de 80, ao realizarmos estudos e avaliações em mais de 300 escolas, abrangendo dezenas de cidades do País, conforme referência anterior. Estes estudos, analisados segundo uma perspectiva temporal, podem ser considerados o início de

⁴⁰ Avaliação do Rendimento Escolar de Alunos da 3ª série do 2º Grau — subsídios para uma discussão. **Estudos em Avaliação Educacional**, nº 3. São Paulo, Fundação Carlos Chagas, 1991.

⁴¹ Ver, também, para uma discussão mais ampla desse ponto, o artigo: Avaliação do Rendimento de Alunos de Escolas de 1º Grau da Rede Privada — Pontos Críticos e Convergências. **Estudos em Avaliação Educacional**, nº 7. São Paulo, Fundação Carlos Chagas, 1993.

⁴² Evasão, repetência e rendimento escolar — a realidade do sistema educacional brasileiro. **Estudos em Avaliação Educacional**, nº 4. São Paulo, Fundação Carlos Chagas, 1991.

outros mais sistemáticos, desenvolvidos em nível estadual, a partir do início dos anos 90.

Ao mesmo tempo que realizávamos diferentes trabalhos em nossa área de concentração, preocupávamo-nos com o problema da metaavaliação, procurando fazer uma análise crítica⁴³ do que de mais representativo estava sendo feito no país, chegando a uma conclusão de certa forma pessimista, porquanto, em linhas gerais, situamos a avaliação em nosso contexto educacional na fase da pré-história, por sua preocupação com problemas nem sempre relevantes, limitada a aspectos tópicos, sem maior aprofundamento das questões que interessam, efetivamente, àqueles que militam — **os professores** —, e que deveriam ser os principais destinatários não apenas das avaliações, mas, também, das pesquisas que se realizam na área da educação, pelo menos em termos teóricos. Ambas, quase sempre, destinam-se a agências financiadoras, atendendo muitas vezes a exigências meramente burocráticas. É a avaliação pela avaliação, a pesquisa pela pesquisa, sem maiores conseqüências práticas, provocadoras de mudanças no sistema de ensino, nas práticas instrucionais, na elaboração de currículos e na orientação do processo educacional, salvo, naturalmente, as exceções habituais, como é de praxe afirmar.

Iniciamos, a partir de 1991, junto à Secretaria de Estado da Educação de Minas Gerais, um amplo programa de avaliação do sistema estadual de ensino⁴⁴, no contexto de um conjunto de atividades ligadas a um programa de qualidade do ensino, parcialmente financiado pelo Banco Mundial. Após a avaliação censitária do Ciclo Básico de Alfabetização, prosseguimos avaliando outras séries do Ensino Básico (5ª e 8ª) e, depois, a

⁴³ Avaliando a avaliação: — da prática à pesquisa. **Estudos em Avaliação Educacional**, nº 5, São Paulo, Fundação Carlos Chagas, 1992.

⁴⁴ Avaliação do Ciclo Básico de Alfabetização em Minas Gerais. **Estudos em Avaliação Educacional**, nº 5, São Paulo, Fundação Carlos Chagas, 1992. Desempenho dos Alunos do CBA em Minas Gerais: análise dos resultados e identificação de pontos críticos. **Estudos em Avaliação Educacional**, nº 6, São Paulo, Fundação Carlos Chagas, 1992. Os alunos da 8ª série do Ensino Fundamental em Minas Gerais: desempenho em redação (análise quantitativa). **Estudos em Avaliação Educacional**, nº 9, São Paulo, Fundação Carlos Chagas, 1994. Atitude em relação à Ciência. **Estudos em Avaliação Educacional**, nº 10, São Paulo, Fundação Carlos Chagas, 1994.

avaliação da 2ª série do Ensino Médio e da Habilitação Magistério (3ª e 4ª séries). A avaliação, repetimos, estava inserida num conjunto de outras medidas ligadas à autonomia administrativa, financeira e pedagógica. Uma avaliação desse tipo somente faz sentido se objetiva mexer efetivamente com o sistema, sua administração e, especialmente, com a sua pedagogia, implicando alterações curriculares, a partir da identificação de pontos críticos, além de medidas efetivas para a qualificação dos professores ligados ao ensino das primeiras séries. Apesar da expansão dos programas de avaliação em todo o sistema brasileiro de ensino, nos seus vários níveis — não será mais um modismo imposto pelas agências financiadoras? —, o problema prioritário, a nosso ver, centra-se, realmente, na qualificação de professores, especialmente para o 1º Grau; depois, então, a avaliação, em diferentes áreas visando a aspectos diversos.

A revista **Estudos em Avaliação Educacional**, em seus números 6 (1992) e 9 (1994), dá uma idéia da complexidade da avaliação de sistemas de ensino, problema que nos levou a delinear uma metodologia⁴⁵, considerando aspectos técnicos e oferecendo orientações práticas para a sua concretização, partindo do modelo inicial que desenvolvemos para usar em um programa de medida da qualidade da educação⁴⁶. A experiência desta avaliação demonstrou que o trabalho baseado em população, especialmente, e não por amostragem, somente é possível se houver a colaboração total dos professores, identificados com os objetivos do trabalho, a aceitação do corpo discente, consciente da importância do trabalho para a melhoria da sua aprendizagem, e o envolvimento efetivo dos pais em todas as fases do processo, inclusive na análise dos dados e na elaboração dos relatórios finais. O modelo proposto para Minas Gerais foi, posteriormente, com as necessárias adaptações, utilizado em São Paulo e no Paraná, entre outros Estados.

⁴⁵ Desenvolvimento de um programa de avaliação do Sistema Estadual de Ensino: o exemplo de Minas Gerais. **Estudos em Avaliação Educacional**, nº 8, São Paulo. Fundação Carlos Chagas, 1993.

⁴⁶ Ver **Estudos em Avaliação Educacional**, nº 2, 1990

Ao fazermos a revisão dos trabalhos publicados pela revista **Cadernos de Pesquisa**, ao longo de um período de 20 anos (1972-92)⁴⁷, verificamos que há um interesse geral sobre o assunto, especialmente em relação à avaliação do rendimento escolar, e grande preocupação metodológica na abordagem dos diferentes problemas, no entanto, observamos, também, que somente a partir de 1973 as questões ligadas à avaliação passaram a merecer um maior enfoque teórico e um interesse maior por metodologias qualitativas, sem, entretanto, um maior conhecimento dos fundamentos dessas mesmas metodologias, **grosso modo**. Tudo indica que já possuímos um material substancial a respeito da educação no Brasil e que é chegado o momento de uma ação direta, com vistas à alteração do presente panorama, que se revela bastante caótico.

Ao longo do tempo, procuramos seguir uma linha de coerência em relação ao que pensávamos e fazíamos, sem, entretanto, nos apegarmos a uma rígida ortodoxia, que, geralmente, conduz a caminhos pouco férteis. A presente revisão do nosso pensamento e do trabalho concretizado após 1962, levou-nos a consultar antigas anotações e livros de autores que de uma forma ou de outra contribuíram para a nossa formação. Encontramos na contracapa de um dos livros de Guilford⁴⁸, autor que fortemente nos influenciou, uma citação por nós manuscrita em 1967, extraída de obra que infelizmente não anotamos, e que reflete o pensamento positivista de William Thomson, Lord Kelvin (1824-1907), o grande físico inglês: ***"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be"*** (1883). A transcrição mostra o seu posicionamento

⁴⁷ Avaliação educacional nos *Cadernos de Pesquisa*. *Cadernos de Pesquisa*, nº 89. São Paulo. Fundação Carlos Chagas, fevereiro de 1992.

⁴⁸ Guilford, J. P. *Fundamental Statistics in Psychology and Education*. McGraw-Hill Book Co. Fourth Edition. New York. 1965.

claramente positivista e, admitimos, orientou, em grande parte, o nosso proceder, ainda que nunca tenhamos confundido as áreas bastante próximas da medida e da avaliação, podendo ser aquela — a medida — o início desta última, mas não necessariamente, porquanto outras abordagens são igualmente possíveis para a análise dos problemas da educação.

O certo é que procuramos acompanhar a evolução do pensamento docimológico e reconhecemos, com Mislevy (1993)⁴⁹, que ao longo dos anos algumas obras tiveram grande influência: **An Introduction to the Theory of Mental and Social Measurements**, de E. L. Thorndike (1919); **Interpretation of Educational Measurements**, de T. L. Kelley (1927); **Psychometric Methods**, de J. P. Guilford (1936); **Probabilistic Models for Some Intelligence and Attainment Tests**, de G. Rasch (1960/80) e, particularmente, a obra capital e verdadeiramente enciclopédica que é a **Statistical Theories of Mental Test Scores**, de F. M. Lord e M. R. Novick (1968), trabalho em que colaborou Allan Birnbaum — **Some latent trait models and their use in inferring an examinee's ability** — Parte 5, caps. 17 e segs. —, apresentando a fundamentação da Teoria da Resposta ao Item (**Item Response Theory**). Esses últimos teóricos contribuíram para dar uma nova dimensão à teoria das medidas educacionais (e psicológicas), ajudaram a compreender a relação que existe entre o nível de habilidade dos sujeitos e o escore obtido em um teste, sobretudo a partir de Rasch e outros mais que aprofundaram essa nova visão, geralmente chamada de teoria moderna dos testes, que complementa, mas não invalida, a teoria clássica, na linha de H. Gulliksen e outros.

A teoria das medidas, ao desenvolver-se, passou a exigir certo nível de sofisticação, que dificulta a compreensão daqueles que não têm, por exemplo, conforme dizem Lord e Novick (1968)⁵⁰, **“competência moderada em cálculo diferencial e integral, além de familiaridade com a linguagem e a mecânica**

⁴⁹ FREDERICKSEN, N.; MISLEVY, R. J. e BEJAR, I. I. **Test Theory for a New Generation of Tests**. Laurence Erlbaum Associates. Publishers; Hillsdale, New Jersey, 1993

⁵⁰ LORD, F. M. e NOVICK, M. R. **Statistical theories of mental test scores**. Reading, M. A. Addison-Wesley Publishing Co. 1968.

básica da estatística matemática". Isso tem contribuído para dificultar o trabalho de muitos, entre os quais nos incluímos, e afastado outros que se julgam incapazes de penetrar em um campo reservado a alguns poucos eleitos. A sofisticação de certos teóricos, sobretudo considerando o emprego cada vez mais freqüente da estatística bayesiana em pesquisas educacionais e psicológicas, demonstra que a área docimológica está a exigir novos talentos com capacitações diferenciadas para acompanhar o desenvolvimento da Psicometria, o que temos procurado fazer, apesar das nossas limitações pessoais decorrentes de uma formação predominantemente humanista.

Acreditamos que um avaliador se constrói ao longo de sua experiência profissional, que, no nosso caso particular, contribuiu para que não tivéssemos uma visão maniqueísta de aceitação irrestrita ao positivismo e rejeição incondicional às alternativas que o condenam: — pós-positivismo, teoria crítica e construtivismo⁵¹, que conflitam entre si, ainda que se unam na sua oposição ao empirismo. A partir das nossas vivências pessoais, quantitativo e qualitativo, objetivismo e subjetivismo, deixaram de ser pólos opostos, irreconciliáveis, que não podem coexistir, mas posicionamentos que se completam no exercício da avaliação educacional, onde não devem existir Ormuz e Arimã, mas um pensamento suficientemente flexível para adequar as várias metodologias à diversidade das situações. Por isso, sentimo-nos gratos a autores como Stake, Guba e Lincoln, Stufflebeam, e até mesmo aos radicais Parlett e Hamilton, entre outros, que nos ajudaram a compreender o problema da avaliação educacional e contribuíram, juntamente com Tyler, Scriven e, particularmente, Cronbach, para a formação do nosso pensamento, preocupado com o mundo das realidades, sem cultivar mitos que deformam a visão do avaliador, acentuando mais ainda os seus vieses, que sempre existem.

⁵¹ GUBA, E. G. *The Paradigm Dialog*. Newbury Park. Sage Publications, California, 1990.