

Uma retrospectiva histórica sobre questões de avaliação matemática

Maria Regina Gomes da Silva*

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

RESUMO

Neste artigo, apresentamos considerações acerca da avaliação matemática, em especial, a partir do artigo de Jeremy Kilpatrick (1993) que, elaborando um referencial histórico, parte das concepções de Vives e Huarte, autores renascentistas, passando pela proposta dos “testes de inteligência” de Binet, Simon e Vaney, trajetória que, segundo Kilpatrick, é “transportada” para os testes padronizados de desempenho. Enquanto Thorndike empenha-se para definir uma padronização desses testes, Tyler procura correlacionar os testes com o currículo escolar, na medida em que formula testes práticos de mensuração associando objetivo, conteúdo e comportamento; Bloom organiza os objetivos em uma taxonomia. Apesar de criticada, a perspectiva de habilidades mentais e da organização da matemática de modo linear e hierárquico tem sido muito poderosa nas práticas de avaliação do século XX, caracterizada fortemente na prática dos testes padronizados de desempenho, amplamente adotada, por exemplo, nos Estados Unidos e transportada para o Brasil.

Unitermos: retrospectiva histórica, avaliação, testes padronizados, transferência de modelos.

INTRODUÇÃO

A avaliação matemática vem recebendo atenção desde Vives e Huarte de San Juan, autores renascentistas que revolucionaram as concepções relativas às funções cognitivas e à avaliação, como ponto de partida para a subsequente escolha profissional do aluno.

O humanista e educador Juan Luís Vives (apud Kilpatrick, 1993, p. 31-32) distinguia quatro funções cognitivas: a função de receber as ima-

*Departamento de Matemática - Faculdade de Ciências - UNESP, Av. Eng Luiz Edmundo C. Coube, s/n - CEP 17033-360 - Bauru - SP

gens sensoriais (imaginação), a função de retê-las (memória), a função de combiná-las (fantasia) e, finalmente, a função de avaliá-las (juízo). Predisposição compreendia cognição, juntamente com sensação perceptual, características físicas e natureza emocional. Vives acreditava que o ensino deveria ser adaptado às várias propensões apresentadas pelos alunos, e propôs um “juízo de agudeza de espírito” como meio para julgar e avaliar os poderes intelectuais e o caráter. Tomou a revolucionária posição de que os alunos deveriam receber acompanhamento e ensino individuais. O aluno admitido em uma escola deveria ser observado por alguns meses enquanto brincava ou conversava, e o boletim pessoal resultante deveria ser revisado a intervalos regulares por membros do corpo docente.

A proposta de Vives, do juízo da agudeza de espírito, provavelmente motivou seu compatriota o racionalista e filósofo natural Juan Huarte de San Juan (*apud* Kilpatrick, 1993, p. 31-32) a escrever o seu “Exame da agudeza de espírito para as ciências”. Neste trabalho, que tem sido denominado “um ensaio precursor do *behaviorism*”, Huarte, assim como Vives, estava preocupado em combinar as habilidades dos alunos com as disciplinas que queriam seguir. Huarte via três faculdades mentais: memória, compreensão e imaginação, sendo que: memória indicaria maiores predisposições para a matemática; compreensão, para as ciências físicas, e imaginação, para matemática avançada. Mais racionalista e filósofo natural do que humanista, via o aluno primordialmente como alguém que possui habilidades naturais que devem ser estimuladas e cultivadas e menos como alguém que deve receber ensinamento diretamente do professor. Nem Vives nem Huarte davam importância à matemática como uma matéria prática, mas ambos consideravam que a mesma exigia tanto raciocínio como memória.

Cabe salientar que na China, já em 2200 antes de Cristo, estabeleceu-se a prática de estar oficiais de três em três anos. O candidato ao serviço civil, em 1151 antes de Cristo, tinha de se submeter a exames em vários campos, inclusive em matemática. Ao redor de 1370 depois de Cristo, exames competitivos abertos estavam bem estabelecidos na China, e candidatos que eram bem sucedidos em seus exames escritos, no nível do distrito, continuavam a submeter-se a outros testes na capital da província e então, se bem sucedidos, em Pequim. O sistema só foi abolido em 1905, mas, antes, já havia sido elogiado e copiado na Europa e nos Estados Unidos, onde exames competitivos foram adotados como meio de seleção, não somente para o serviço civil, mas também para o serviço militar e para a universidade (Kilpatrick, 1993, p. 36).

Na Idade Média, as “maratonas públicas”, assumem papel importante no surgimento do sistema de provas, quando os alunos deveriam demonstrar domínio da matéria, ao utilizar textos padrões e fundamentar seus argumentos (contra ou a favor) com pensamentos de autoridades. Vives, que havia estudado em Paris de 1509 a 1512, acreditava que as maratonas promoviam a exibição pessoal em detrimento da “busca cole-

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

tiva da verdade”. Apesar disso, nos séculos XVII e XVIII, na Inglaterra e na Irlanda, os exames continuavam a manifestar-se como uma mistura de maratona e “catequismo” (sistema de perguntas e respostas). O sistema de provas escritas, nas universidades européias, parece ter se desenvolvido em princípios do século XVIII (Kilpatrick, 1993, p. 36-37).

A IDÉIA DAS HABILIDADES MENTAIS

Ainda segundo Kilpatrick (1993, p. 33-36), também os “testes mentais” deram atenção especial à avaliação, quando Francis Galton, desde 1869, em sua obra *Hereditary Genius*, tenta aplicar as teorias de Darwin à psicologia. Na virada do século, Alfred Binet, juntamente com Theodore Simon, dedicaram-se a avaliar tarefas complexas que se aproximavam mais da vida diária.

O teste de inteligência Binet-Simon, que viria influenciar a construção de numerosos testes, organizava a habilidade mental em uma escala indo do menor para o maior, com base no ponto a partir do qual o desempenho positivo começava a ser ultrapassado pelo desempenho negativo. Tarefas eram organizadas em crescente ordem de dificuldade, e o examinado recebia o resultado. Subseqüentes baterias de testes passaram a embasar o modo de contagem para produzir resultados baseados no número de respostas corretas, mas a noção de escala permaneceu.

Em 1905, no mesmo ano em que o teste Binet-Simon foi publicado, Binet abriu a primeira escola laboratorial francesa para pedagogia experimental em Paris. Estimulou o diretor da escola, seu amigo e colaborador Vaney, a produzir testes de desempenho que poderiam servir de “barômetros” do ensino.

O primeiro teste de Vaney avaliava o desempenho em aritmética, numa escala de notas de 1 a 7, e continha itens que alunos “médios” deveriam ser capazes de realizar com êxito. Binet e Vaney utilizavam os resultados dos testes dentro de um processo de identificação de crianças com deficiência mental. Nos anos que se seguiram, surgiu um debate quanto à natureza da inteligência e sobre o que estava sendo medido por testes de inteligência, sendo que a idéia de medir tarefas segundo o grau de dificuldade como um indicador de perícia mental permaneceu relativamente incontestável.

O número crescente de estudantes estimulou a prática dos exames orais que, além de, com freqüência, se caracterizarem por certa superficialidade, exigiam longos períodos de espera dos alunos. De 1770 a 1840, percebe-se, com a crescente importância da solução de problemas matemáticos, a predominância gradual das provas escritas com conseqüente detrimento das orais. Oxford e outras universidades seguiram o exemplo de Cambridge neste sentido, mas as provas orais não perderam sua primazia em muitas universidades, permanecendo, de uma ou outra forma, até hoje.

O *boom* dos exames orais, na Inglaterra, teve lugar na segunda metade do século XIX. O governo inglês, em 1862, começou um esquema “prêmios segundo os resultados”, isto é, as escolas recebiam dinheiro conforme o desempenho de seus alunos em leitura, redação e aritmética. A crescente dificuldade de pagar o enorme número de examinadores externos contribuiu para voltarem as provas escritas, louvadas não só por sua eficiência, como também porque pareciam proporcionar, adicionalmente, tratamento igualitário, livre da parcialidade e favoritismo dos examinadores.

Em 1889, William Torrey Harris levantou a questão de que os exames deveriam ser realizados pelo corpo docente, e não por avaliadores externos. Hendell Harris, um professor de Haverford, acreditava, ao contrário, que uma banca de professores externos (não preocupados com “resultados democráticos”, como os professores dos alunos) deveria ser responsabilizada pelas provas, “juntando somente os bons peixes em seus recipientes adequados, jogando os ruins fora”. Harris enfatizava que a primeira coisa a ser mudada nos exames deveria ser o examinador, defendendo os exames como um meio de revelar diferenças individuais (Kilpatrick, 1993, p. 34).

TESTES DE DESEMPENHO

Testes de desempenho nos Estados Unidos começaram em 1845, quando “O Comitê da Escola de Boston” (orgulhoso do sucesso alcançado por suas escolas e sob pressão de Horace Mann, Secretário da Comissão de Educação de Massachussetts), realizou uma pesquisa escolar geral, que deveria indicar que as escolas estavam executando a tarefa que o estado esperava delas, em vista das generosas contribuições financeiras que haviam recebido.

Os examinadores publicaram uma tabela classificando as escolas de redação e de gramática de Boston segundo a média de desempenho de uma amostra de seus alunos em uma série de testes escritos. Embora tenham se apressado em destacar que as classificações só ofereciam medidas aproximadas do desempenho intelectual dos alunos e eram medidas inadequadas dos méritos da escola, isso não impediu que os administradores utilizassem testes de desempenho na comparação de escolas e professores. Por volta de 1870, testes escritos de desempenho eram amplamente usados em cidades grandes e certos estados, sendo que os resultados eram ocasionalmente publicados em jornais. Na época em que os Estados Unidos entraram na Primeira Guerra Mundial, mais de 200 testes de desempenho encontravam-se disponíveis para uso nas escolas primárias e secundárias, com onze disponíveis só em aritmética (Resnick, 1982 *apud* Kilpatrick, 1993, p. 34).

Em 1888, quatrocentos eminentes homens e mulheres, dirigindo-se às escolas, protestaram contra o “sacrifício da educação pelas provas”,

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

mas as universidades também eram criticadas por educarem com “um corpo sem alma”. Os argumentos de Vives contra maratonas, conseqüentemente contra o domínio do “competir” em detrimento do compreender, reviveram. A tensão causada pelas provas era angustiante para algumas pessoas que consideravam falha grave, dentro do sistema, a não promoção da aprendizagem permanente, aspecto que a educação deve valorizar. Nessa mesma época, nos Estados Unidos, os superintendentes escolares, professores e presidentes das universidades começaram a pressionar, exigindo reformas no sistema de avaliação, já ultrapassado.

Educadores começaram a preocupar-se quanto à forma como estes testes estavam sendo usados e começaram a requerer padronização. Se os testes deveriam ser utilizados para fins de comparação, algum tipo de padronização fazia-se imprescindível. Uma primeira tentativa para representar escalas da contagem dos testes de desempenho e fornecer normas para cada classificação de nota - seguindo o caminho de Binet e Vaney - foi efetuada por Edward Thorndike e seus alunos, que, de 1908 a 1916, desenvolveram testes de desempenho em aritmética e outras disciplinas.

Independentemente das questões de construção e padronização de testes, havia também a questão de como os testes se relacionavam com o currículo escolar. Neste campo projetou-se Ralph Tyler que, em 1931, relatou seus esforços dentro de vários cursos universitários. Seu método foi de formular os objetivos do curso e definir cada objetivo em termos do comportamento do aluno. Coletava situações nas quais os alunos revelariam a presença ou ausência do objetivo. Apresentaria, então, uma situação ao estudante, avaliando suas reações à luz do objetivo proposto. Procurou determinar a objetividade e confiabilidade da avaliação, esforçando-se para melhorá-la. Para tanto, ao mesmo tempo, sempre que necessário, desenvolvia outros métodos praticáveis de mensuração. Adicionalmente, revelou que cada objetivo de curso tinha dois aspectos, de conteúdo e de comportamento.

O passo seguinte foi dado por Benjamin Bloom (1956 *apud* Kilpatrick, 1993, p. 35) e seus colegas, que organizaram os objetivos em uma taxonomia (dedicada a Tyler) que tentava refletir a preocupação dos professores de serem lógicos e internamente consistentes, além de refletirem a psicologia atual, e serem, ao mesmo tempo, neutros e compreensivos. A taxonomia, relacionada a “comportamentos educacionais”, e organizada em ordem ascendente do simples para o complexo, aplicar-se-ia a todas as matérias escolares, enquanto se dividiam os objetivos segundo três “domínios”: cognitivo, afetivo e psicomotor, sendo que o cognitivo recebia a maior atenção. Evidências de pesquisas deveriam apoiar as declarações sobre o terreno cognitivo no sentido de que “à medida que os comportamentos tornam-se mais complexos, o indivíduo torna-se mais consciente de sua existência”, assim como as aptidões e habilidades que se encontram nos níveis mais altos da taxonomia, são aprendidos mais eficientemente do que o conhecimento nos níveis mais baixos (Kilpatrick, 1993, p. 42).

A taxonomia de Bloom tem sido frequentemente vista como servindo à matemática de uma forma deficiente. Ormell (1974 *apud* Kilpatrick, 1993, p. 36), por exemplo, apontou em uma forte crítica à taxonomia: “as categorias de Bloom são extremamente amorfas em relação à matemática. Eliminar o ponto central da disciplina, e tentar implementá-las, pelo menos no nível das classes mais adiantadas, é um exercício permanente em escolhas arbitrárias”. Entretanto, desde sua publicação, variantes da taxonomia de Bloom para o terreno cognitivo têm ajudado a fornecer esquemas para a construção e análise de muitos testes de desempenho matemático. Estudos como o “Estudo Nacional Longitudinal de Aptidões Matemáticas” e a “Avaliação Nacional de Progresso Educacional” nos Estados Unidos, assim como o “Primeiro e Segundo Estudos Matemáticos Internacionais”, têm desenvolvido itens e escalas dos testes em matrizes de “Conteúdo pelo processo” nas quais a dimensão do conteúdo abarca uma classificação de tópicos matemáticos e a dimensão do processo é uma versão da taxonomia de Bloom (Kilpatrick, 1993, p. 35).

Com o florescimento de provas dentro e fora das escolas e universidades, educadores começaram a preocupar-se com a confiabilidade e objetividade das avaliações. Surgiu a ciência da mensuração educacional em busca de um “novo tipo” mais objetivo, de avaliação, diferente dos ensaios da época.

TESTES OBJETIVOS PADRONIZADOS

Nos Estados Unidos, sob a liderança de Joseph Rice e Thorndike, respostas curtas, falso / verdadeiro, e testes de múltipla escolha, começaram a ser utilizados. Nesse ínterim, na Europa, particularmente na Inglaterra, os educadores se perguntavam o que é que os exames estavam avaliando, e, se as notas dadas eram justas. Em 1910, Philip Hartog propôs em um discurso à “Royal Academy of Art” que fosse realizada uma pesquisa sobre exames, o que, por falta de fundos, não se concretizou (Kilpatrick, 1993, p. 41).

Em 1925, uma comissão conjunta da “Association of College Presidents” da Pensilvânia e do “State Department of Public Instruction” do mesmo estado, começou a considerar a relação entre a escola secundária e currículos universitários. Queriam verificar como alunos poderiam ser melhor preparados para a faculdade, e como as faculdades poderiam planejar o ensino com base no que os estudantes haviam aprendido anteriormente. Convidaram a “Carnegie Foundation for the Advancement of Teaching” a conduzir a pesquisa, que ficou sendo conhecida como o “Pennsylvania Study”. O estudo partiu da premissa que, enquanto na Europa se esperava que o crescimento intelectual dos alunos fosse sistemático e crescente ao longo de linhas claramente determinadas, nos Estados Unidos o programa dos alunos caracterizava-se como uma “colcha de re-

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

talhos” de cursos não relacionados, em que os objetivos eram medidos por créditos acumulados, mais do que através de êxitos examinados. O “Pennsylvania Study” necessitou mais de dez anos para sua realização e tornou-se um ponto de referência no desenvolvimento de testes objetivos padronizados para uso nas avaliações escolares e universitárias.

Do intercâmbio entre Frederick Keppel (da “Carnegie Corporation”) e Paul Monroe (do “International Institute of Teachers College”, da Universidade de Colúmbia) e da idéia de fazer algo pela educação na Europa, algo que afetasse vários países, surgiu a “International Examinations Inquiry”, conduzida de 1931 a 1938 (Kandel, 1936, Monroe, 1931, 1936, 1939 *apud* Kilpatrick, 1993, p. 40). Na primeira “Conference on Examinations” (Eastbourne, Inglaterra, maio de 1931), os delegados da França e da Alemanha manifestaram seu ceticismo quanto à habilidade dos testes objetivos avaliarem a *culture générale* ou *allgemeine Bildung*. Os delegados da Inglaterra e dos Estados Unidos, por sua parte, mais ansiosos em experimentar testes objetivos e novos esquemas de atribuição de notas, expressavam sua insatisfação com as práticas de avaliação vigentes. Outras conferências foram realizadas em junho de 1935 e setembro de 1938. A Segunda Guerra Mundial pôs fim à “International Examinations Inquiry” e aos planos do final do encontro no sentido de estabelecer uma organização permanente para estudar sistemas de provas em todo o mundo. Segundo estimativa de Kilpatrick (1993), é difícil julgar como vários países modificaram seus sistemas de provas em consequência da “Inquiry”, mas as metas de tornarem as avaliações mais confiáveis e objetivas, evitando divergências nas notas, foram reconhecidas como desejáveis.

OS TESTES, DIMINUINDO A IMPORTÂNCIA DA APRENDIZAGEM E DISTORCENDO O ENSINO

A visão hierárquica das habilidades e, conseqüentemente, de comportamento e de objetivos, parece estar provocando práticas educacionais infelizes, como, por exemplo, a separação em uma fase de transmissão de conhecimento e outra fase, não conectada a esta, quando, então, entram a aplicação ao raciocínio e a solução de problemas. “Hierarquizações” de objetivos educacionais, embora destaquem habilidades mais elevadas de organização, paradoxalmente, sugerem que a fase de aquisição do conhecimento é a primeira dentro de uma seqüência.

Desde que é relativamente mais fácil avaliar conhecimentos que processos de pensamento, mudanças dentro do sistema educacional dificilmente ocorrem. O êxito na solução de um problema de matemática, por exemplo, pressupõe não só o uso de habilidades cognitivas, como também uma leitura social e cultural correta da situação. “Nossas condições sociais compreendem uma quantidade enorme de componentes que interagem, de forma que nunca são inteiramente especificadas, pré-determi-

nadas, antecipadas ou permitidas” (PopKewitz, 1988 *apud* Kilpatrick, 1993, p. 43).

Na Conferência final sobre exames, Desclos colocou uma questão fundamental: “Não existe o perigo de que, se nos comprometemos ao ideal de primeiro estabelecer mensurações e, então, conformar nossos ensinamentos a tais medidas, estarmos baixando o nível do que estamos ensinando, ou estarmos nos limitando em nosso esforço de ensinar naqueles níveis mais elevados, que tanto ansiamos por conseguir?” Os defensores dos testes objetivos nas Conferências argumentavam dizendo que os mesmos, longe de baixarem o nível do currículo ou distorcerem o ensino, capacitariam professores a diagnosticar dificuldades dos alunos, individualizando a aprendizagem. Provas, utilizando ensaios, observações clínicas etc., ofereceriam informações adicionais, a serem registradas no boletim do aluno (Monroe, 1939 *apud* Kilpatrick, 1993, p. 43).

A voz mais confiante e mais otimista ouvida nas Conferências da “International Examination Inquiry” foi a de Thorndike, o ardente positivista e marcante psicólogo. Hoje, ele quase parece justificado em sua fé, de que tudo o que existe pode ser medido de alguma forma. O século XX tem produzido uma prática de avaliação na educação que é dominada no mundo inteiro pela psicomетria, a mensuração da *psique*. O desafio para o século XXI, no que diz respeito aos professores de matemática, é criar uma prática de avaliação que faça mais do que medir a mente de uma pessoa e, então, prescrever um tratamento àquela mente. Necessitamos compreender como as pessoas (não à parte, mas sim integradas às suas culturas) utilizam a Matemática em vários meios sociais. Além disso, é imprescindível criar um ensino matemático capaz de proporcionar um melhor uso, maiores compensações e mais responsabilidade. Para chegar a tal ponto, será necessário transcender às horríveis visões da mente como uma hierarquia, da escola como uma máquina e da avaliação como engenharia.

Cinqüenta anos depois da Conferência, levanta-se a acusação, nos Estados Unidos, de que os testes objetivos provocaram a eliminação de outras formas de avaliação nas escolas, diminuindo a importância da aprendizagem e distorcendo o ensino. Em outros países, como na França, por exemplo, surgem perguntas quanto aos efeitos secundários indesejáveis e, algumas vezes, até perigosos, nos métodos correntes de avaliação” (Commission Internationale de l’Enseignement Mathematique, ICMI-1990 *apud* Kilpatrick, 1993, p. 44). Segundo ICMI-1990, a visão dos sistemas de exame como um meio eficiente e justo de diagnosticar habilidades e encaminhar alunos a papéis sociais adequados tem sido fortemente ofuscada.

A própria história dos Estados Unidos - um país que dá uma importância extraordinária à hierarquização das habilidades, ao espírito de competitividade, ao desempenho positivo, ao sucesso - tem demonstrado que os testes, não raras vezes, subestimam a habilidade dos indivíduos, além de serem previsores ineficientes, como no caso dos 2,3 milhões de egres-

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

dos da Segunda Guerra Mundial que, admitidos irrestritamente às universidades americanas (sem se submeterem a qualquer tipo de teste), alcançaram resultados acima da média, confirmando que “maturidade e motivação aparentemente eram mais importantes do que níveis de habilidade ou mesmo preparação acadêmica” (Gifford & O’Connor, 1992, p. 4).

Críticas adicionais têm questionado a validade da distinção entre objetivos cognitivos e afetivos, a independência de conteúdo do processo, e o significado dos objetivos isolados de qualquer contexto. Apesar de tudo, com profundas raízes em nossa história e em nossa psique, a perspectiva das habilidades mentais e, conseqüentemente, do pensamento e logros matemáticos organizados de maneira linear, hierárquica, tem sido muito poderosa na prática de avaliação do século XX.

O USO ATUAL DOS TESTES PADRONIZADOS

Nos Estados Unidos, país de modelos de alto teor globalizante, estima-se, segundo Dossey & Swafford (1993, p. 49), que o número de testes padronizados, amplamente aplicados nas escolas primária e secundária, atinge atualmente os 200 milhões por ano. Trata-se de testes “comercialmente preparados”, com itens para obter amostras do comportamento do aluno sob procedimentos uniformes, em forma de múltipla escolha.

Os objetivos dos testes padronizados de desempenho nos Estados Unidos são, supostamente, “ajudar diretores, professores e coordenadores de currículo a melhorarem o currículo e o ensino” (p. 47). Os testes padronizados, originariamente meios auxiliares do ensino, “são agora utilizados para avaliar o progresso acadêmico: para determinar níveis de desenvolvimento dos alunos; diagnosticar pontos fortes e fracos, comparar procedimentos instrucionais alternativos e servir como uma variável dependente na pesquisa educacional” (p. 50). Recorre-se aos mesmos testes para fins de prestação de contas para avaliar melhorias nas escolas, qualidade escolar global ou em nível de distrito, competência de administradores e professores. Os resultados ocupam, quase sempre, as manchetes das primeiras páginas dos jornais; afetam preços imobiliários; e, em alguns estados, podem até influir no melhor pagamento ou no reconhecimento ao mérito de certos professores.

Há, porém, estudos que revelam que o sistema de testagem referenciado a normas colocam em dúvida o caráter de prestação de contas dos testes e de medida do progresso do aluno, constatando-se uma tendência na disseminação do “ensinar em função do teste” e mesmo da “cola” (Cannell, 1990 apud Dossey & Swafford, 1993, p. 51). Os professores vêem-se, portanto, quase obrigados a construir suas práticas de sala-de-aula em função dos testes.

CONSIDERAÇÕES FINAIS

Apesar do resultado de pesquisas, como as expostas acima, ouvem-se vozes de políticos, e até de educadores, clamando por um sistema de testagem no Brasil, transferido de modelos políticos e educacionais de outros países, com o suposto objetivo de melhorar a qualidade do ensino, avaliando diretamente o trabalho do professor através dos resultados dos alunos nos testes, fazendo um reducionismo na medida em que o número de pontos obtidos nos testes pelo aluno é diretamente proporcional à qualidade do trabalho do professor. Segundo esse ponto de vista, o ensino vai mal porque os professores não se esforçam efetivamente.

Nesse viés, o clamor por um sistema de testagem, já parcialmente materializado por legislação federal¹ e amplamente divulgado pela mídia, nos leva a ler em uma revista de penetração em quase todos os estados brasileiros que “No próximo ano, todos os formandos em medicina, engenharia e direito - as carreiras mais procuradas no País - devem passar por um exame nacional” (ISTOÉ, 15/11/95, p.110 -111). Acrescenta-se a isto as investidas no ensino de primeiro e segundo graus registradas em reportagem do *Jornal Nacional da Rede Globo de Televisão*, de outubro de 1995, que destacava a presença do Ministro da Educação, Paulo Renato Souza, em uma escola de Santa Catarina onde o sistema de testagem estava sendo “testado”, servindo de amostragem para futura ampliação em nível nacional.

Os responsáveis pelas diretrizes políticas, adeptos da transferência de modelos políticos e educacionais de outros países, parecem ignorar a heterogeneidade e rigidez das estruturas econômicas; a concentração do progresso técnico e da riqueza acumulada; os elevados índices de desigualdade regional e social de renda; um mercado interno relativamente limitado em face de seu porte demográfico etc. É sabido que o longo período de recessão e de instabilidade econômica trouxe, como consequência, níveis crescentes de desigualdade social e regional, fazendo do Brasil um dos países mais perversos em distribuição de renda do continente - os 10% mais ricos concentram mais da metade da renda nacional. Aliado ao longo período de recessão e instabilidade econômica que trouxe reduções nas receitas fiscais e conseqüentemente nos gastos públicos com a educação, avulta-se o não-cumprimento da aplicação dos percentuais previstos para a educação. A ausência de critérios claramente estabelecidos e de controles efetivos na alocação e distribuição de recursos “tem levado a que grande parte dos recursos destinados à educação acabem por financiar as macroestruturas, chegando poucos recursos às escolas” (Brasil, 1993, p. 26). Diante dessa falta de “comprometimento”, a educação, tomada como fonte eletiva, não assumida como questão social e nem como meio de desenvolvimento econômico, depende de um compromisso político e de uma vontade política, respaldados por “medidas fiscais adequadas”, dado que não podemos mais tolerar que de “cada

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

1 Medida Provisória nº 938 - de 16 de março de 1995: Altera dispositivos de Lei nº 4024 (Leg. Fed., 1961, pág. 979; 1962, pág. 366; 1968, pág. 1433.), de 20 de dezembro de 1961, e da Lei nº 5540, de 28 de novembro de 1968, e dá outras providências. Art. 3º O Ministério da Educação e do Desporto fará realizar exames de avaliação dos conhecimentos adquiridos pelos alunos das últimas séries dos cursos de graduação das instituições de ensino superior. Lei Nº 9131, de 24 de novembro de 1995. Altera dispositivos da Lei n. 4024, de 20 de dezembro de 1961, e dá outras providências.

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

1000 crianças que ingressam na 1ª série, apenas 45 concluem o ensino fundamental em oito anos e sem repetência” (Brasil, 1993, p. 87).

Sem descartar a experiência acumulada de reformas, inovações, pesquisas e o progresso em educação registrado em muitos países, não é possível que se parta da premissa de que o que é “bom” para países estrangeiros, é bom para o Brasil, embora tenha sido “uma constante nos países da América Latina a transferência de modelos políticos, econômicos, administrativos e educacionais de outros países, na expectativa da solução ou contemporização dos problemas existentes” (Demange, 1994, p. 19). No entanto, é evidente que no jogo de poderes entre colonizadores e colonizados, nada pôde impedir “a violenta exclusão social que tem orientado as elites econômicas, políticas e culturais de vários países da América Latina” (Demange, 1994, p. 20).

Na medida em que os políticos se apropriam de certa parte do discurso das pesquisas internacionais de educação para assegurar o sucesso de suas campanhas eleitorais, no Brasil discussões vêm sendo efetuadas, sem que exista a preocupação de aplicar o dinheiro gasto nas testagens (“Provões”) nacionais e estaduais, em áreas mais necessitadas dentro do sistema educacional, por exemplo, para diminuir o ‘analfabetismo matemático’, em áreas identificadas como carentes.

SILVA, Maria R. G. da. A historical view on mathematics assessment. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.

ABSTRACT

The article presents considerations on mathematics assessment based on Jeremy Kilpatrick's article, in which he elaborates a historical referential beginning with renaissance authors Vives and Huarte, going through Binet's and Vaney's "intelligence tests", a trajectory he characterizes as "transported" to the standardized performance tests. Whereas Thorndike struggles to standardize such tests, Tyler attempts to correlate them with the school curriculum as he formulates practical assessment testes by relating objective, content and behavior; Bloom, on his turn, organizes the objectives into a taxonomy. Despite criticism, the perspective involving mental skills and a linear hierarchical organization of mathematics has been very pervasive in the 20th century's assessment practices, since it strongly characterizes the practice of standardized performance tests which are widely used in countries such as the United States (Dossey & Swafford, 1993). Even though results from studies have been revealing that the assessment system puts into question the tests' character of rendering accounts and of assessing students' progress, politicians and educator have been claiming for a Brazilian assessment system transferred from other

countries' political and educational models, with the objective to improve educational quality by directly evaluating teacher's work through students' testing results.

Key Words: historical view, assessment, standardized tests, transference of models.

REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL. Ministério da Educação e do Desporto. *Plano decenal de Educação para Todos*. Brasília: MEC, 1993.

DEMANGE, Nilson Joseph. Transferência Cultural: a educação num contexto de globalização. *Pro-Posições*, Campinas, v.5, n. 3, p.19-23, 1994.

DOSSEY, John A., SWAFFORD, Jane O. Issues in Mathematics Assessment in the United States. In: *CASES of Assessment in Mathematics Education. An ICMI Study*. Dordrecht/Boston/ London: Kluwer Academic Publishers, 1993. p. 43-57.

GIFFORD, Bernard R., O'CONNOR, Mary Catherine (ed.). *Changing Assessments. Alternative Views of Aptitude, Achievement and Instruction*. Norwell Massachusetts: Kluwer, 1992.

KILPATRICK, Jeremy. The Chain in the Arrow: From the history of Mathematics Assessment. In: *Investigations into Assessment in Mathematics Education. An ICMI Study - International Commission on Mathematical Instruction*, Dordrecht/Boston/ London: Kluwer Academic Publishers, 1993. p. 31-46.

SILVA, Maria R. G. da. Uma retrospectiva histórica sobre questões de avaliação matemática. *Mimesis*, Bauru, v. 19, n. 1, p. 101-112, 1998.