

A Avaliação de Sistemas Educacionais: a Experiência Inglesa(1)

Caroline Gipps(2)

Introdução

O Reino Unido é um país pequeno, principalmente se comparado com o Brasil. Tem quatro áreas distintas, mas interligadas: a Inglaterra, a Escócia, o País de Gales e a Irlanda do Norte; uma população de 58 milhões e 400 mil habitantes; 4.455 escolas secundárias (que atendem a alunos de 11 a 15/16 anos) e 23.408 escolas primárias (que se destinam a crianças de 5 a 11 anos). O ensino obrigatório começa aos 5 anos de idade e termina aos 16, embora a maioria dos jovens continue sua educação ou seu treinamento técnico depois. A Escócia e a Irlanda do Norte possuem sistemas educacionais bem diferentes dos da Inglaterra e do País de Gales; estes últimos têm o mesmo currículo nacional e a mesma estrutura de avaliação, mas, no País de Gales, o galês é incluído como língua principal ou como a segunda língua. Neste artigo, concentro-me na experiência e nos progressos ocorridos na Inglaterra.

1 Título original em inglês: *Assessment in the Education System: the English Experience*. Tradução de Dean de Abreu Azevedo; revisão técnica da tradução de Maria José do Amaral Perneira.

2 Professora e pesquisadora do Instituto de Educação da Universidade de Londres, onde ocupa o cargo de Dean of Research. Sua formação inicial é na área de Psicologia, e trabalhou como professora primada antes de ingressar na carreira universitária. É presidenta da British Educational Research Association (Associação Britânica de Pesquisa Educacional) e, entre as inúmeras pesquisas que publicou, encontram-se várias na área de Avaliação. A *fair test? Arseument, Achievement and Equo*, escrito juntamente com Patrícia Murphy, ganhou o prêmio SCSE de melhor livro sobre Educação publicado em 1994.

O Currículo Nacional

A Lei da Reforma Educacional de 1988 introduziu, pela primeira vez na história recente, um currículo nacional para o ensino obrigatório de 5 a 16 anos, junto com um programa nacional de avaliação para alunos com idades de 7, 11, 14 e 16 anos.

O currículo nacional foi planejado de forma que todos os alunos em idade de escolarização obrigatória façam o mesmo curso, no qual Inglês, Matemática e Ciências constituem o núcleo, e História, Geografia, Tecnologia, uma língua estrangeira moderna, Arte, Música e Educação Física - consideradas também como disciplinas básicas - formam um núcleo ampliado. Para todas essas disciplinas o currículo está definido em lei: os regulamentos descrevem como "programas de estudo" os assuntos, as habilidades e os processos a serem ensinados, e como "objetivos de rendimento escolar" os conhecimentos, as habilidades e a compreensão que se espera dos alunos quando atingem certos estágios da escolarização. Os estágios são definidos como Estágios-Chave 1 (de 5 a 7 anos), 2 (de 7 a 11 anos), 3 (de 11 a 14 anos) e 4 (de 14 a 16 anos).

O National Curriculum 5-16: A Consultation Document (publicado pelo Department of Education - DES', em 1987) já antecipava a Lei da Reforma Educacional. Naquele documento, foi feita a seguinte declaração sobre avaliação e padrões de melhoria:

"Os alunos devem ter as mesmas oportunidades onde quer que curse a escola, e os padrões de aproveitamento escolar devem ser elevados em toda a Inglaterra e em todo o País de Gales.

(-) Um currículo nacional baseado em procedimentos claros de avaliação ajudará a melhorar os padrões de aproveitamento:

- (i) garantindo que todos os alunos estudem um leque amplo e equilibrado de assuntos (-)*
- (ii) estabelecendo objetivos claros para aquilo que as crianças devem saber em todas as disciplinas estudadas (-)*
- (iii) avaliando o progresso em direção àqueles objetivos e o desempenho alcançado nos vários estágios, de modo que os alunos possam progredir ainda mais quando estiverem indo bem e recebam ajuda quando não estiverem (-)*

(...) Os objetivos, em termos de rendimento escolar, serão estabelecidos para as três disciplinas nucleares, ou seja, para Matemática, Inglês e Ciências. Essa é uma forma essencial e comprovada de elevar os padrões de aproveitamento. Os objetivos devem ser suficientemente específicos para que alunos, Professores, pais e outros tenham uma idéia clara do que se espera, e para que constituam uma boa base de avaliação. "
(DES, 1987, p. 3-4, 9,10).

3 órgão cujas funções se assemelham a algumas do nosso Ministério da Educação.

O programa nacional de avaliação vem sendo, pois, considerado crucial para o currículo nacional, pois é através dele que tem sido vista a possibilidade de uma melhoria nos padrões atingidos pelos alunos.

O primeiro passo para desenvolver o currículo nacional e o programa de avaliação foi a instituição do Grupo de Trabalho em Avaliação e Testes - TGAT(4). O relatório desse grupo de trabalho (DES, 1988) apresentava um esboço de estrutura do currículo que deveria orientar a proposta curricular de todas as disciplinas. As disciplinas foram divididas em vários módulos, em termos de objetivos de rendimento escolar a serem atingidos, articulados em uma série de níveis progressivos. A série de níveis foi formulada de modo a possibilitar progressão: a maioria dos alunos com 7 anos ou mais estaria no nível dois do sistema, enquanto a maioria dos alunos de 11 anos ou mais estaria no nível quatro, e assim por diante. Em cada um dos níveis, os objetivos de rendimento foram descritos através de uma série de critérios de atingimento do rendimento esperado, de forma a constituir a estrutura básica de um sistema de avaliação referenciado a critério.

O modelo subjacente ao currículo e à avaliação nacionais é, portanto, um modelo de currículo baseado em objetivos (concretizado em objetivos de rendimento, descritos com a especificação de critérios de atingimento), que incentiva os professores a se concentrar nas habilidades e nos processos (estabelecidos nos programas de estudo) ligados a um sistema de avaliação referenciado a critério.

Dois métodos principais são usados nesse sistema de avaliação: o primeiro parte de testes e tarefas externos, ou seja, de **testes e tarefas padronizados de avaliação(5)**, e o segundo é a própria avaliação informal que os professores fazem do rendimento do aluno, ou seja, a **avaliação do professor(6)**. Para realizá-la, os professores fazem uma avaliação do nível de rendimento de cada aluno, tendo em vista a escala de níveis referentes aos objetivos de rendimento previstos para as disciplinas nucleares. Os professores podem fazer essas avaliações da forma que desejarem, mas a observação, a avaliação informal regular e a guarda de exemplares de trabalhos são todas ações incentivadas. Quanto aos **testes e tarefas padronizados de avaliação**, foram planejados originalmente como avaliações complexas do desempenho, nas quais cada criança estaria envolvida em atividades relacionadas ao currículo.

Em função da confiança depositada na avaliação do professor, o relatório-TGAT sugeriu que se realizasse também um processo complexo de discussões em grupo, através do qual as avaliações dos professores pudessem alinhar-se em torno de um padrão comum. A combinação dos resultados desses dois tipos de avaliação - a feita pelo professor e os testes e tarefas padronizados - tem sido uma área controversa; no início, a regra era que, quando a consecução de um objetivo fosse avaliada tanto pelo professor quanto por testes e tarefas padronizados e os resultados diferissem, o resultado desses últimos devia ser "preferido".

(4) Task Group on Assessment and Testing-TASK.

(5) Standard Tests and Standard Assessment Tasks-SATs.

(6) Teacher Assessment - ta.

Atualmente, os resultados da avaliação feita pelo professor e dos testes e tarefas padronizados têm de ser apresentados em separado.

A primeira avaliação para crianças de 7 anos em Inglês, Matemática e Ciências aconteceu em 1991, e a segunda, em 1992. Na época em que deveria ser feita a avaliação de 1993, o alcance do programa de avaliação e o tempo que demandava foram dramaticamente reduzidos; além disso, em muitas escolas, em consequência de um boicote nacional organizado pelos três maiores sindicatos de professores, ou os testes não foram aplicados, ou os resultados não foram encaminhados às autoridades educacionais locais⁽⁷⁾. O boicote continuou em 1994, mas, em 1995, todas as escolas fizeram a avaliação nacional no fim dos Estágios-Chave 1, 2 e 3.

Boicote e revisão

Em 1992 e no começo de 1993, houve muitos debates no âmbito dos sindicatos dos professores e das associações profissionais para definir se os professores deviam ou não boicotar a avaliação. Em março de 1993, um dos sindicatos de professores propôs boicotar toda e qualquer avaliação do currículo nacional por causa da grande sobrecarga de trabalho (aproximadamente 116 horas) gerada pela aplicação e pela correção dos testes dos alunos de 14 anos. Essa decisão foi declarada legal pelos tribunais e outros sindicatos aderiram ao boicote. Assim, embora a avaliação do currículo nacional estivesse instituída em grande parte das escolas primárias, muitos professores de crianças de 7 anos não aplicaram os testes em 1993. Em consequência do boicote, o governo criou, sob a direção de Sir Ron Dearing, um comitê para rever todo o currículo nacional e o programa de avaliação, com o objetivo expresso de encontrar formas de simplificar este último.

Em 1994 houve um segundo boicote, levado a cabo por apenas um sindicato de professores; mas muitos professores de escolas primárias não aplicaram os testes-uma versão modificada para as crianças de 7 anos e testes-piloto para os alunos de 11 anos - ou os aplicaram, mas não apresentaram os resultados.

Os principais resultados da Revisão Dearing são:

- uma simplificação do currículo;
- a suspensão da divulgação da classificação das escolas no que se refere aos alunos de 7 e 14 anos (embora o governo tenha mantido a divulgação da classificação das escolas primárias usando os resultados dos alunos de 11 anos);
- a apresentação dos resultados da avaliação feita pelos professores junto com os resultados de testes (ao invés de os primeiros serem omitidos, como ocorria antes);
- a substituição dos relatórios focados no rendimento por descrições mais abrangentes.

⁷ Local Education Authorities-LEAs.

Descrições de objetivos por níveis/padrões

Para simplificar a base referenciada a critério, quase 1.000 descrições de objetivos de rendimento foram reduzidas a 200 descrições de objetivos por níveis. Um exemplo dessas descrições é apresentado a seguir:

Objetivo 2. • Aritmética e Álgebra Nível 2 Os aluno contam séries de objetos corretamente e falem mentalmente contas de adição e subtração de fatores até 10. Começaram a entender o valor da localização de cada dígito num número e usam esse conhecimento para ordenar números até 100. Escolhem a operação adequada ao resolver problemas de adição e subtração. Identificam e usam metades e quartos, como a metade de um retângulo e um quarto de oito objetos. Reconhecem seqüências de números, inclusive números pares e ímpares.

Essas descrições de objetivos são semelhantes às usadas em algumas partes da Austrália e àquelas que estão sendo desenvolvidas nos Estados Unidos.

A passagem para as descrições de objetivos por níveis foi feita em razão da sobrecarga inegável constituída pelo número colossal de descrições de objetivos de rendimento das disciplinas nucleares do currículo nacional. No entanto, uma objeção às descrições por níveis é que elas são globais demais para serem usadas como critério de avaliação: se os professores tiverem de utilizá-las com propósitos de avaliação que não se resumam a um procedimento muito abrangente e intuitivo, talvez tenham de decompô-las; haverá também necessidade de exemplos, de forma a ajudar os professores em classe a fazer avaliações, e não apenas descrições. Discussões com os professores também serão muito importantes, principalmente nos dois primeiros anos, enquanto o novo sistema se estabelece. Cabe lembrar, enfim, que quando descrições desse tipo, ou ligeiramente mais elaboradas, são usadas em outros países, ou em outros programas de avaliação, o fornecimento de exemplos é automático.

Apresentação de resultados e classificações

Os resultados da avaliação do currículo nacional têm de ser apresentados para as idades de 7, 11 e 14 anos; o desempenho dos alunos de 16 anos é avaliado através do exame público que dá o Certificado Geral de Educação Secundária - GCSE(8). Os resultados que cada aluno obtém são confidenciais para ele, para seus pais e para seus professores; os resultados em nível de escola, por sua vez, são apresentados publicamente para as idades de 11 e 16 anos. A publicação dos resultados deve fazer parte de um relatório mais amplo feito pela escola a respeito de seu trabalho como um todo; neste sentido, o relatório TGAT sugeriu que esses informes incluíssem um relatório geral por área, "para indicar a natureza das influências sócio-econômicas e outras que, sabidamente, afetam as escolas". Apesar dessas recomendações, entretanto, ultimamente listagens relativas ao desempenho, ou classificações, têm sido publicadas em nível local e nacional.

8 General Certificate of Secondary Education - ME.

Nível primário

• *Estágio-Chave 1 (avaliação de crianças de 7 anos)*

A avaliação do Estágio-Chave 1 começou em 1991. Níveis crescentes de discussão e cooperação entre os colegas foram observados nas escolas que pesquisamos⁽⁹⁾ nos primeiros estágios de sua implementação: por parte dos professores de alunos de outras faixas etárias, ao ajudar os professores do Ano 2 - isto é, os professores de crianças de 7 anos - que sentiram estar diante de uma tarefa hercúlea; e entre os professores do Ano 2, durante todo o estudo, ao negociar significados para as avaliações do Estágio 1 - SOAS - e padronizar critérios. O envolvimento dos diretores no apoio à sua equipe de professores significou que havia um sentimento de "estar no trabalho juntos", o que também ajudou a desenvolver formas cooperativas de trabalho. A ajuda mútua por ocasião da aplicação dos testes e tarefas padronizados e da comparação entre os resultados destes e aqueles obtidos na avaliação realizada por eles próprios tirou os professores de suas salas de aula e levou-os a trabalhar em contato uns com os outros (GIPPS et al., 1992, 1995).

Somente no segundo ano da avaliação nacional é que o governo conseguiu mostrar que os padrões tinham se elevado: John Patten, ministro da Educação na época, fez uma declaração à imprensa no final de 1992 anunciando a elevação dos padrões, uma vez que a porcentagem da população com 7 anos de idade atingindo níveis mais altos aumentara em leitura, ortografia e Matemática ("Os resultados dos alunos de 7 anos mostram melhoria nos padrões", nota à imprensa do DFEE em 21 de dezembro de 1992). Patten tomou esse fato como prova de que o currículo nacional estava "dando certo". É claro que essa elevação dos níveis de desempenho deveu-se ao fato de os professores ensinarem mais intensamente o que era exigido pelos testes e tarefas padronizados: pontuação, ortografia, caligrafia e aritmética mental. Deram mais atenção ao "básico" em 1992 e 1993, e isso apareceu nos níveis mais elevados de desempenho das crianças. A experiência de aplicar os testes e tarefas padronizados também ajudou os professores a entender os critérios de avaliação e as próprias avaliações do Estágio 1, de modo que pudessem ensinar de uma forma mais dirigida à avaliação dos anos seguintes. É claro que, dadas as mudanças do currículo e das tarefas de avaliação havidas em cada um dos três anos, além da falta de confiabilidade dos resultados (**SHORROCKS** et al., 1992), as mudanças na qualidade do desempenho devem ser tratadas com prudência.

Os professores de nossa pesquisa não ensinavam diretamente para provas específicas (embora as crianças estivessem andando "rápido" na aritmética mental para conseguir fazer as somas da prova em cinco segundos); estavam ensinando áreas do currículo, de modo que as crianças pudessem fazer avaliações a respeito desses tópicos. A diferença entre a situação que observamos e a da maioria dos outros países; onde ensinar para a avaliação é prática comum, é que, na Inglaterra e no País de Gales, temos um currículo imposto, assim como uma avaliação imposta. Há, portanto, além dos itens testados, algo mais que os professores devem ensinar.

9 national Assessment in Primary Schools: an Evaluation: E.S.R.C.1990.1996.
(10) Stage one Assessments - SoAs.

O que nosso estudo mostrou claramente foi uma mudança na prática da avaliação que os professores fazem das crianças (apesar da variação individual), que passou de uma abordagem intuitiva para outra baseada em evidências e trabalhos escritos. Além disso, é claro que muitos professores ficaram mais capacitados em termos de avaliação, não se limitando a operar técnicas de um sistema imposto.

• **Estágio-Chave 2 (avaliação de crianças de 11 anos)**

A avaliação do Estágio-Chave 2 começou realmente em 1995 e, em sua maior parte, tem sido feita a partir de testes padronizados, e não de tarefas de avaliação. Os resultados passarão a ser divulgados em listas de classificação a partir de 1997; e já é possível prever o efeito bem rotineiro do ensinar-para-o-teste.

Esse efeito foi observado diretamente nas seguintes formas: em 1995, as escolas-piloto envolvidas desde 1993 na implantação dos testes do Estágio-Chave 2 informaram que, em consequência desse trabalho, mudaram suas formas de organização: ao invés de privilegiarem o ensino conjunto com vistas a um determinado objetivo, partindo do trabalho com tópicos que abrangessem todo o currículo, como faziam antes, passaram para o ensino focado mais em cada disciplina; além disso, **todos** os professores de nossa pesquisa fizeram algo para preparar seus alunos para os testes como, por exemplo, revisão de Ciências, provas de Matemática e testes cronometrados de Inglês.

Descobrimos também, no entanto, que os professores, em muitos casos, mudaram sua **forma de ensinar**. Fizemos entrevistas em profundidade com 29 professores do Estágio-Chave 2 em 1995. Dos 29 entrevistados, 15 disseram adotar diferentes formas de ensinar para circunstâncias diferentes. Descreveram os momentos em que se tornavam basicamente "transmissores" de conhecimento, os momentos em que deveriam encorajar as crianças a descobrir sozinhas e os momentos em que sentiam estar "construindo o aprendizado conjuntamente". Nenhum dos professores endossou a instrução direta ou a abordagem de "transmissão" para ensinar seus alunos, mas poucos a repudiaram abertamente; limitam essa abordagem a certos objetivos. Que a metade de nossa amostra tenha defendido explicitamente a abordagem mista do ensino é interessante, à luz dos comentários mais gerais de nossos professores sobre o efeito do currículo nacional sobre sua prática de ensino. Embora nenhum professor acredite que as crianças aprendam exclusivamente com a transmissão de fatos, um quarto deles sentiu-se impelido a enfatizar mais isso, por causa da quantidade de trabalho demandada pelo currículo nacional ou, em alguns casos, pelos testes.

Os efeitos nas habilidades de avaliação dos professores do Estágio-Chave 2 foram menos marcados, até agora, do que nos professores do Estágio-Chave 1. Isso se deve, a nosso ver, às menores exigências quanto à avaliação feita pelo próprio professor e à natureza mais padronizada dos testes. Mas, devido à grande repercussão gerada pela publicação dos resultados, houve um efeito mais evidente em sua prática organizacional e educacional, a fim de que atingissem classificações *melhores*.

Em ambos os Estágios-Chave da escola primária, podemos discernir muito claramente um ensino voltado para a avaliação; contudo, como o currículo é amplo e se requer que as habilidades do "processo" sejam verificadas por meio da avaliação feita pelo próprio professor, esse efeito não foi até agora particularmente negativo ou redutor.

• *Estágio-Chave 3 (avaliação de alunos com 14 anos)(11).*

O currículo do Estágio-Chave 3 foi introduzido em 1989, em primeiro lugar nas disciplinas de Ciências e Matemática. O currículo de Inglês entrou em vigor no ano seguinte, em 1990. Esse Estágio-Chave abrange três anos (7^{os.}, 8^{os.} e 9^{os.} anos) e, por isso, a primeira avaliação de fim de estágio referente a ele deveria ser realizada em 1992. Durante esse período, entretanto, houve considerável controvérsia sobre a forma da avaliação de fim de estágio, controvérsia esta que refletia grande parte do debate que houvera no Estágio-Chave 1 sobre a extensão, a complexidade e o papel da avaliação realizada pelo próprio professor. Os órgãos responsáveis, de acordo com o relatório TGAT, produziram extensas tarefas padronizadas de avaliação a serem realizadas em sala de aula, sendo que uma delas, para a disciplina Inglês, exigia cerca de seis semanas para que os alunos a terminassem (ver GIPPS, 1992, para detalhes sobre essas abordagens). Em 1991, o ministro da Educação, Kenneth Clarke, arruinou o trabalho-piloto ao descrevê-lo como "um nonsense elaborado" e ao dizer:

'Estou persuadido de que o processo será mais fácil e de que os resultados inspirarão mais confiança se a avaliação for feita principalmente sob a forma de testes escritos curtos'
(DAUGHERTT, 1995, p. 53).

Em 1992, assim, houve a aplicação-piloto de testes escritos curtos de Inglês, Matemática e Ciências. Em 1993, os testes, a essa altura elaborados em grande parte por juntas examinadoras, caíram por terra, em função dos boicotes levados a cabo pelos professores. Embora a sobrecarga de trabalho constituísse a base legal do boicote, sua verdadeira causa foi a série de objeções que os professores faziam aos testes de Inglês, tanto por seu conteúdo quanto pela ameaça de publicação dos resultados nacionais, apesar de os itens utilizados na avaliação não terem sido adequadamente pré-testados. Em consequência da Revisão Dearing de 1995, os testes foram corrigidos externamente e depois devolvidos às escolas.

Os testes de 1996 para o Estágio-Chave 3 foram, portanto, as primeiras provas formais" aplicadas a toda a população de 14 anos: nesse estágio, foram utilizados essencialmente exames escritos convencionais. Alunos que se esperava estarem nos níveis 3-7 fizeram duas provas, o que requereu de 2 a 3 horas no total. A correção foi feita externamente às escolas, e os limites de nível foram estabelecidos de acordo com as notas finais, depois de examinado o desempenho dos candidatos. Os estudantes do nível 8, agora no topo da escala, devem fazer um prova extra que requer de uma a uma hora e meia, e que implica a elaboração de uma redação. Os alunos dos níveis 1-2 são avaliados por meio de tarefas padronizadas, ao invés de o serem por meio de testes. Há uma avaliação feita pelo professor para todas as disciplinas do currículo e para todos os alunos, e esta é apresentada em separado, junto com os resultados dos testes. Atualmente, as escolas têm de apresentar o boletim das crianças a seus pais, junto com as notas da escola e as médias nacionais

O Certificado Geral de Educação Secundária- GCSE (avaliação para alunos de 16 anos)

O exame nacional para os estudantes de 16 anos na Inglaterra, no País de Gales e na Irlanda do Norte (a Escócia tem seu próprio Diploma Padronizado), o GCSE, foi introduzido em 1988 por um governo conservador. Substituiu o exame para a obtenção do Certificado Geral de Educação de Nível O(12) destinado aos 30% mais bem classificados nas escolas, e o exame feito pelos que não conseguiam atingir o nível O para obter o Certificado de Educação Secundária(13), por um único exame que passou a incorporar características de ambos (ver KINGDON e STOBART, 1988). O GCSE teve um sucesso impressionante. No cerne deste sucesso estava a introdução de programas inteligentes de cursos e técnicas flexíveis de avaliação; igualmente importante foi a introdução de trabalhos a serem completados no decorrer do curso, longos ensaios, projetos e investigações práticas propostos por juntas examinadoras, corrigidos pelos professores e discutidos pelas próprias juntas.

Como se esperava que o GCSE fosse feito virtualmente por todos os alunos, fizeram-se necessárias técnicas de avaliação que pudessem lidar com um amplo leque de aproveitamento escolar. A abordagem adotada envolveu diferenciação. Um exame diferenciado é aquele onde diferentes módulos são deliberadamente colocados em diferentes níveis de dificuldade para vir ao encontro das necessidades de candidatos com diferentes níveis de capacidade (KINGDON e STOBART, 1988).

O GCSE satisfaz muitos dos requisitos de um bom exame por ser uma mistura de trabalhos desenvolvidos ao longo do curso e exame (determinada pela adequação ao seu objetivo); por estimular os professores a fazer inovações e a respeitar juízos profissionais; por motivar os alunos com trabalhos feitos ao longo do curso e com programas mais inteligentes que os dos exames que ele substituiu. Entretanto, os limites recentes - politicamente impostos aos trabalhos desenvolvidos ao longo do curso secundário enfraqueceram o modelo.

Há quem veja como um anacronismo a idéia de que os principais exames públicos sejam feitos aos 16 anos. Como uma das poucas nações industrializadas ainda existentes nas quais o ensino obrigatório termina aos 16, e não aos 18 anos de idade (embora, agora, a maioria continue com alguma forma de educação ou treinamento depois dos 16 anos), o GCSE é visto como um obstáculo para se chegar a uma política coerente de educação e treinamento da população de 14 a 19 anos. Os proponentes da mudança defendem uma estrutura modular de créditos que propiciaria aos estudantes fazer um misto de curso acadêmico e profissionalizante, que lhes permitiria chegar sem quebra de continuidade ao mercado de trabalho ou à educação superior.

Publicação de resultados e classificações

Conforme explicado anteriormente, a publicação dos resultados, em nível tanto local quanto nacional, foi um elemento-chave da Lei da Reforma da Educação.

12 General Certificate of Education Level O-GCE O.

13 Certificate of Secondary Education -CSE.

Em 1980, a Lei da Educação exigiu que as escolas secundárias publicassem os resultados da avaliação dos alunos de 16 anos e mais; aliás, um dos elementos mais importantes do programa de governo do partido conservador é que os pais que utilizam a educação fornecida pelo Estado têm de ter mais informações e opções ao decidir quais escolas seus filhos devem freqüentar. Desse modo, a partir de 1982, a publicação anual dos resultados dos exames dos alunos de 16 e 18 anos tornou-se obrigatória. Isso levou a uma crença difundida de que as escolas devem prestar mais contas às comunidades a que servem e que a publicação dos resultados dos exames ajudaria a transformar essa proposta em realidade. No entanto, as diferentes formas escolhidas pelas escolas para apresentar seus resultados tornavam as comparações ainda muito difíceis para os pais ou para a imprensa local. Mais tarde, em 1992, exigiu-se que as escolas informassem publicamente os resultados num mesmo formato, de modo a facilitar a classificação do desempenho das escolas. Além disso, o Department for Education (DFE) divulgou classificações nacionais anuais para as escolas, com base na combinação das notas dos alunos em todas as disciplinas do GCSE e dos níveis adiantados. Os problemas gerados pelo uso dos resultados "brutos" dos exames para avaliar o desempenho escolar foram descritos em detalhe e discutidos amplamente no Reino Unido (GOLDSTEIN e CUTTANCE, 1988; STOBART e GIPPS, 1997). Apesar do debate sobre a importância de se considerar como os alunos chegam e o progresso obtido por eles (para se chegar ao "valor agregado" pela escola) como a base das classificações do desempenho escolar, em 1991 e 1992 foram publicadas as classificações dos distritos escolares com base apenas no programa de avaliação nacional dos alunos de 7 anos. Não é de surpreender que o distrito no topo da classificação se localizasse numa área suburbana socialmente privilegiada, enquanto aquele que ocupou o último lugar - ferozmente atacado pela imprensa nacional tivesse uma elevada proporção de população de minorias étnicas desprivilegiadas do centro decadente da cidade, com muitos alunos tendo o inglês como segunda língua.

Em novembro de 1992, um relatório sintético sob a forma de porcentagens de alunos que conquistaram as notas A-C e A-G do GCSE foi publicado primeiramente pelos distritos escolares e pelas escolas. Foi um trabalho de grande vulto, com custos de publicação superiores a 3 milhões de libras. No entanto, são preocupantes suas interpretações errôneas e suas injustiças. Um exemplo destas últimas é a comparação direta de resultados brutos de escolas extremamente seletivas com os de escolas não-seletivas do mesmo distrito escolar. Por exemplo: a classificação de 1996 mostra que na região londrina de Southwark há três escolas nas quais mais de 97% dos alunos conseguiram GCSEs com notas A*-C, enquanto em todas as outras 14 escolas secundárias essa porcentagem foi de menos de 50%. Não será surpresa descobrir que as três primeiras são escolas de prestígio, particulares e academicamente seletivas.

Em conseqüência dessas preocupações, foram elaboradas as abordagens de valor agregado. Nos dados de valor agregado, os resultados são apresentados de forma a considerar a contribuição da escola para o progresso do aluno. Isso em geral é feito em termos de medidas de insumo e produto; por exemplo: os alunos que entram na escola aos 11 anos com um currículo nacional de nível 3, e atingem o nível 6 aos 14 anos, podem ter-se saído melhor em termos de valor agregado do que os alunos que entraram em outra escola com o nível 5 e atingiram o nível 7 aos 14 anos. Em termos de resultados brutos, o desempenho da segunda escola seria considerado melhor -mesmo que tivesse feito menos por seus alunos.

A passagem para as análises de valor agregado, portanto, parece uma reação lógica aos limites das classificações feitas de acordo com os resultados brutos dos exames e ao uso destes últimos como base de comparação entre as escolas.

No entanto, há problemas com o uso de resultados escolares médios obtidos depois do ajuste, pois eles podem variar muito e, aparentemente, de forma aleatória, quando se fazem modificações triviais nos procedimentos de ajuste. GOLDSTEIN e WOODHOUSE (1988) chegaram a essa conclusão no caso dos distritos escolares; o mesmo efeito foi encontrado na classificação de distritos escolares (ao invés de escolas) nos Estados Unidos: a alteração ou mudança de critérios com vistas à classificação e ao ajuste dos resultados produziu classificações bem diferentes dos distritos escolares (GUSKEY e KIFER, 1989). Também acontece de a ordem da classificação mudar consideravelmente de um ano para o seguinte. Além disso, o limite da classificação enquanto processo é que, independentemente dos níveis globais de desempenho, um procedimento de classificação sempre produz um primeiro e um último lugar, com poucos indícios do que isso significa em termos de desempenho.

Está claro que atualmente há um apoio político e educacional considerável no Reino Unido para as abordagens baseadas no valor agregado. No entanto, é inevitável que haja pressão por medidas "simples" que podem bem levar a problemas semelhantes, ou novos, em relação à classificação.

Conclusões

Esbocei neste artigo os progressos consideráveis da avaliação educacional na Inglaterra nos últimos dez anos. A combinação de um novo currículo nacional com um programa nacional de avaliação (e um novo exame para os estudantes de 16 anos) e outras reformas educacionais teve um impacto tremendo sobre os professores, os alunos e as escolas.

O impacto sobre a prática docente em termos de ênfase no ensino de conteúdos com probabilidade de serem avaliados chegou a ser antecipado (GIPPS et al., 1994) para o grupo de 7 anos de idade. Num pequeno número de escolas abrangidas por nosso estudo, esse efeito foi sentido até nas escolas maternas. Não que as crianças tenham sido pressionadas; houve, contudo, um realinhamento do ensino. No primeiro ano de avaliação dos alunos de 11 anos, também tivemos clara evidência de que os professores prepararam os estudantes diretamente (*The Times*, 12/4/95). O impacto mais amplo, social, nas escolas primárias, ainda não é evidente; até os resultados dos exames de alunos de 11 anos serem divulgados nas listas públicas de classificação, o efeito será pequeno. No entanto, no nível secundário, já começamos a ver um número crescente de exclusões, na relutância de algumas escolas em admitir crianças com problemas de aprendizagem ou de comportamento, e no fato de escolas com excesso de matrículas estarem entrevistando os pais antes de resolver quais alunos admitir (BALL, 1994). Essas tendências podem ser todas colocadas na conta da filosofia da escolha de mercado: o dinheiro acompanha os alunos ainda mais diretamente do que antes, aponta-se como ganho o "direito" dos pais de escolherem as escolas dos filhos ("direito" que, para muitos, não passa de uma quimera), e as classificações baseadas nos resultados dos exames são tomadas como principal indicador da "qualidade" de uma escola.

Persistem duas grandes preocupações: o impacto sobre o ensino, se a avaliação do currículo nacional continuar se tornando cada vez mais formal (com mais testes padronizados e menos avaliações de desempenho; os trabalhos escolares já saíram de cena, neste nível); e a diferenciação crescente entre as escolas, se as que tiverem excesso de pedidos de matrículas selecionarem sua clientela com base no rendimento escolar e no ambiente familiar, forçando os alunos e famílias problemáticos a entrar nas escolas "problemáticas".

Está claro, no entanto, que, a partir da avaliação nacional, os professores passaram a ter objetivos e expectativas muito mais claros em relação aos alunos, principalmente no nível primário, onde não havia programas de curso, nem exames finais. Isso provavelmente vai melhorar os níveis de desempenho nos exames do currículo nuclear. Embora muitos professores lamentem a perda de autonomia, o currículo nacional e o programa de avaliação serviram para uniformizar mais as práticas escolares, o que é desejável num sistema nacional de um país pequeno.

Bibliografia

- BALL, S. *Education reform: a critical and post-structural approach*. Buckingham: Open University Press, 1994.
- DAUGHERTY, R. *Nacional Curriculum Assessment- a review of policy 1987-1994*. Londres: Falmer Press, 1995.
- DES. *The National Curriculum 5-16: a consultation document (the TGAT Report)*. Londres: DES/WO, 1987.
- _____. *Nacional Curriculum: Task Group on Assessment and Testing - a Report*. Londres: DES/Welsh Office, 1988.
- GIPPS, C. *Developing Assessment for the National Curriculum*. ULIE/Kogan Page, Bedford Ways Series, 1992.
- GIPPS, C., BROWN, M., McCALLUM, B., McALISTER, S. *Intuição or evidence?: teachers and National Assessment of Seven Years Olds*. Buckingham: Open University Press, 1995.
- _____. *Nacional Assessment at Seven: some emerging themes*. In: GIPPS, C. (Org.). *Developing Assessment for the National Curriculum*. Bedford Way Series, ULIE/Kogan Page, 1992.
- GOLDSTEIN, H., CUTTANCE, P. National Assessment & School Comparisons. *Journal of Education Policy*, v. 3, n. 2, 1988.
- GOLDSTEIN, H., WOODHOUSE, G. National Assessment and School Comparisons. *Journal of Education Policy*, v 3, n. 2, 1988.
- GUSKEY, T, KIFER, E. *Ranking School District on the basis of Statewide Test Results: is it meaningful or misleading?* Dissertação apresentada na Conferência AERA, em San Francisco, em março de 1989.
- KINGDON, M., STOBART, G. *GCSE Examined*. Londres: Falmer Press, Lewes, 1988.
- SHORROCKS, D. et al. *ENCA 1 Project Report*. Londres: SEAC, 1992.
- STOBART, G., GIPPS, C. *Assessment: a teachers' guide to the issues*. Londres: Hodder & Stoughton, 1997.