

# DESENVOLVIMENTO DE PROJETOS DE AVALIAÇÃO DO SISTEMA EDUCACIONAL NO ESTADO DE SÃO PAULO

BERNARDETE GATTI\*

Fiz um levantamento dos estudos avaliativos de desempenho escolar em redes de ensino desenvolvidos especificamente no Estado de São Paulo e verifiquei que, durante muitos anos, vários estudos foram realizados, mas de pequeno porte e praticamente sem impacto nenhum. Foi a partir de 1988, por iniciativa do Ministério da Educação, que São Paulo iniciou um processo de avaliação um pouco mais amplo, com amostras de alunos mais representativas da sua própria rede – embora se tratasse ainda de estudos preliminares que envolveram um número não muito grande de crianças do sistema.

Portanto, falarei mais do que aconteceu de 1988 para cá, quando avaliações começaram a ser realizadas num nível um pouco mais amplo. Considero que houve quatro estudos nessa fase que vou chamar de "estudos preliminares" que tiveram alguma importância para criar em São Paulo certa competência em relação ao que se pode chamar de "avaliação de rendimento escolar nos sistemas de ensino."

Foram estudos que permitiram análises bastante apuradas, não só quanto ao problema curricular envolvido, como também quanto ao tipo de prova e ao tipo de criança que existe nas redes e como estas lidam com esse tipo de prova, além de examinar alguns problemas de política e outros

---

\*Coordenadora do Depto de Pesquisas Educacionais-FCC. e Profª da Pontifícia Universidade Católica de São Paulo (PUC)

relativos, por exemplo, a características de diretores ou professores das escolas. Esses estudos preliminares foram, como já disse, realizados por proposta do Ministério da Educação dentro de um projeto mais amplo, em que se envolveram praticamente todos os Estados da Federação durante quatro anos.

Em 1987, o MEC convocou para uma reunião alguns pesquisadores para discutir a viabilidade de iniciar projetos de pesquisa em avaliação de desempenho escolar e para discutir a possibilidade de montagem de um Sistema Nacional de Avaliação. A importância desse momento foi que a iniciativa do MEC envolveu uma discussão não só com os técnicos, mas com todos os Secretários de Educação da época, que foram convidados a ir a Brasília participar de um Seminário onde essas questões foram discutidas. Vale dizer que o Secretário de Educação de São Paulo – na verdade, os Secretários, já que São Paulo teve no período vários Secretários e, por outro lado, foram várias reuniões – nunca compareceram. Sempre foi enviado algum técnico substituindo algum Secretário.

De qualquer maneira, o Ministério conseguiu realizar convênios com todos os Estados para que se pudesse ter avaliações com base em amostras de cada Estado. Decidiu-se avaliar naquele momento as disciplinas de Português com Redação, Matemática e Ciências, nas crianças de 1<sup>as.</sup>, 3<sup>as.</sup>, 5<sup>as.</sup> e 7<sup>as.</sup> séries do primeiro grau.

Esses estudos, como disse, eram exploratórios e foram feitos inicialmente em 10 capitais, depois em 20 cidades representativas desses Estados participantes e, em outro momento, em 39 grandes cidades, incluindo São Paulo só nesta terceira etapa.

A perspectiva do Ministério com essas avaliações, segundo os documentos publicados, foi a de, além de obter informações gerais sobre desempenho escolar, subsidiar as Secretarias de Educação com informações relativas ao desenvolvimento dos programas de ensino, fornecendo elementos para reflexão sobre problemas de currículo ou questões de formação e treinamento de professores.

O Ministério acreditava, uma vez firmado o convênio com os Secretários e tendo estes participado de várias reuniões, que os Estados ao receberem os resultados, promoveriam uma seqüência de discussões sobre esses dados com os diversos segmentos educacionais envolvidos. No entanto, esses dados parecem ter caído num vazio, apesar do aparente comprometimento das Secretarias de Educação.

O MEC, também nesse período, não desenvolveu nenhum trabalho mais intenso de reflexão sobre os resultados com as Secretarias. A imprensa chegou a comentar em vários momentos esses dados e o impacto ficou por conta, sobretudo, dos maus resultados divulgados. No entanto, creio que o impacto desses estudos foi muito importante, sob o ponto de vista de se criar nas instâncias administrativas, quer no Ministério, quer em alguns

Estados, alguma competência em questões de avaliação em equipes que estavam envolvidas com esses estudos. Também começaram a subsidiar a compreensão dos problemas ligados à repetência escolar e à evasão.

Os estudos, realizados em 87, 88 e 89, aconteceram sobretudo em escolas públicas, embora o projeto pretendesse envolver escolas particulares. Só em 1991 se conseguiu realizar os estudos das escolas particulares, usando os mesmos instrumentos e princípios que foram utilizados nas outras avaliações.

O grande problema nesse momento, que creio foi importante resolver, foi a questão da escolha dos tópicos a avaliar em cada disciplina e série. No Brasil, não temos um currículo mínimo obrigatório e as variações regionais e locais são muito grandes. A saída para essa questão, parece-me, foi muito importante, pois envolveu a discussão, não com especialistas ou com pessoal que faz guias curriculares, mas sim com professores em exercício nas redes em diferentes Estados do país; envolveu a análise dos livros didáticos mais adotados pelos professores, para ver o que realmente se dava em sala de aula.

Acho que essa opção foi interessante. A própria aplicação das provas mostrou o quanto esses conteúdos estavam realmente mais próximos daquilo que o professor fazia na sala de aula do que os guias curriculares existentes, na medida em que se recebiam avaliações sucessivas dos professores sobre as provas, e que essas provas depois foram discutidas inclusive com algumas associações de professores em alguns Estados.

Dentro desses estudos preliminares, no Estado de São Paulo, foram avaliadas 1.198 crianças das diferentes séries. Depois, quando se realizou aqui no Estado e em outros Estados o estudo com as escolas particulares (1991), a amostra dessas escolas sofreu graves problemas. Primeiro, porque as escolas particulares não aceitavam a avaliação, e o MEC teve que fazer uma série de negociações para conseguir que elas fossem avaliadas. Segundo, porque no momento da aplicação das provas algumas escolas colocavam vários impecilhos para sua aplicação. Apesar do compromisso assinado com o Ministério, na hora da aplicação das provas escolas particulares dificultavam a entrada do pesquisador dizendo que tinham decidido que não iriam mais participar da avaliação ou inventando algum tipo de desculpa. Então a amostra ficou diminuída no que se referiu às escolas particulares. Mesmo assim se conseguiu avaliar um número significativo de alunos nessas escolas, no âmbito destes primeiros estudos. Os dados obtidos estão publicados e várias análises foram possíveis de serem desenvolvidas.

Outro programa que teve nesse momento impacto no desenvolvimento de uma certa competência nessa área, sendo também desdobramento do programa do Ministério da Educação, foi um projeto de avaliação de desempenho de alunos de 3ª série do 2º grau. Observem que até aqui não houve qualquer iniciativa do Estado de São Paulo, apenas do Ministério da

Educação. Esse projeto, embora dentro da proposta do Ministério, teve estímulo do Banco Mundial. Foi um estudo amostral feito em 1990, visando a levantar informações sobre uma amostra de alunos da rede pública e privada a nível de 2º grau, para uma melhor compreensão dos problemas que afetavam esse nível de escolaridade.

Esse trabalho foi realizado em quatro Estados, inclusive em São Paulo, e envolveu escolas técnicas federais, escolas estaduais e particulares de formação genérica, escolas estaduais e particulares com habilitação magistério (que foram avaliadas especificamente), e, escolas de formação profissional e industrial do SENAI. Foi uma amostra complexa que estratificou todos esses elementos para poder tirar informações sobre o desempenho escolar dos jovens, relacionando o tipo de curso, a rede de ensino e o turno freqüentado. A avaliação foi feita em matérias básicas: Língua Portuguesa e Matemática. Esses dados foram cruzados com informações levantadas sobre a escola e o aluno. Questionários foram aplicados para se fazer uma análise do significado desses estudos para os alunos. Aqui em São Paulo, por exemplo, a amostra da Habilitação Específica do Magistério (HEM) teve que ser muito grande, dado o número de escolas e seu espalhamento por todo o Estado, e o grande número de alunos. Neste particular, foi possível ter uma visão bastante clara do tipo de alunado que freqüentava a HEM à época aqui no Estado. Após esses estudos, que eu consideraria precursores, houve um outro, desenvolvido especificamente pela Fundação Carlos Chagas no bojo de um projeto internacional. Foi realizado no município de São Paulo através do II Programa Internacional de Avaliação do Progresso Educacional (IAEP). Foi feita uma avaliação de alunos escolarizados, com 13 anos de idade em escolas públicas e privadas, em Matemática e Ciências, visando uma comparação internacional. Foram avaliados 1.484 alunos em Matemática e 1.469 em Ciências. Houve aplicação de questionários aos diretores das escolas e questionários sobre os alunos e suas famílias. Os questionários foram cruzados com os resultados e o estudo foi bastante divulgado, tanto nacional como internacionalmente. Além disso, seus resultados causaram muita polêmica, já que o Brasil ficou situado em penúltimo lugar dentre os países que participaram da avaliação internacional. Esse estudo ofereceu oportunidade para que o debate público sobre desempenho escolar fosse um pouco mais extenso, e que alguns envolvidos com a educação no país tomassem consciência de algumas questões e problemas.

Uma das dificuldades encontradas nesse estudo relacionou-se à amostra de crianças porque, nos países participantes, a criança de 13 anos se situa num determinado nível de escolaridade. A repetência é praticamente inexistente. A série corresponde à idade. Aqui no Brasil já é bem diferente. Dada a repetência ou início tardio da escola, as crianças com essa idade podem estar desde o 1º ano primário até a 8ª série do primeiro grau, o que

dificultou muito a definição da amostra e a validação daqueles conteúdos curriculares que aqui se iria avaliar. Acabou-se limitando a aplicação das provas às crianças de 13 anos que estavam escolarizadas num pequeno conjunto de séries (5ª a 7ª) porque a prova, na verdade, era feita num nível que se esperaria de uma criança de 13 anos frequentando a escola de maneira regular e não sendo retida sucessivamente.

Esse estudo teve uma série de impactos, políticos inclusive, por causa dessa distribuição excessiva das crianças de mesma idade em um rol de oito séries – sem contar as crianças que já estavam fora da escola nessa idade. A questão de defasagem idade/série se tornou bem visível. Então, ele propiciou, a nível de alguns grupos, de alguns técnicos, quer da Secretaria de Educação do Estado de São Paulo, quer do Ministério, quer de outros Estados, a tomada de consciência no sentido de que talvez fosse necessário olhar mais atentamente estas questões: o que significa o aprendizado das crianças nas nossas escolas públicas, privadas ou de qualquer natureza, nessas condições?

O primeiro movimento que realmente houve, para se pensar um sistema de avaliação em termos especificamente de Secretaria de Educação do Estado de São Paulo, foi feito em 1992. Pensou-se em começar um processo mais consistente, que à época se esperaria, se implantaria numa seqüência regular. Esse processo de avaliação nasceu vinculado ao Projeto das Escolas-Padrão. A Secretaria de Educação do Estado de São Paulo não pensou numa avaliação do seu sistema, mas em avaliar uma política específica que era a das Escolas-Padrão. Mas já se pensou como uma possibilidade de acompanhamento constante e continuado, dado que a idéia à época era de que as Escolas-Padrão tivessem seu modelo estendido sucessivamente para todas as escolas componentes da rede.

A avaliação então realizada abrangeu todas as 306 Escolas-Padrão implantadas e referiu-se a Língua Portuguesa, Redação, Matemática, Ciências, História e Geografia. Foram submetidos a ela todos os alunos de 8ª série. A idéia era que a escola pudesse ter dados dos seus alunos para poder subsidiar o seu planejamento e o seu trabalho pedagógico. Por parte da Administração Pública, a idéia era poder ir contando, sucessivamente, durante vários anos, com dados da progressão dessas crianças que estavam vinculadas ao Projeto da Escola-Padrão. Esperava-se evidentemente que essas crianças, não à época da avaliação, mas, sucessivamente, à medida que o projeto se implementasse permanentemente, pudessem apresentar um ganho significativo em termos de aprendizagem escolar.

Nesse período, a Secretaria de Educação desenvolveu um esforço muito grande, redigiu vários documentos encaminhados às Delegacias de Ensino e às escolas. As escolas receberam cada uma delas não só o relatório geral do rendimento das crianças no Estado, mas um perfil dos alunos da própria escola. Foram feitas várias reuniões pelos técnicos da Secretaria

tentando discutir, quer a nível da Delegacia, quer das escolas, a melhor forma de utilização desses dados. Os resultados foram interpretados psicopedagogicamente e recomendações curriculares foram feitas e consolidadas em documentos distribuídos às escolas.

Em 1993, dando continuidade a esse programa, a Secretaria propôs desenvolver algumas análises mais aprofundadas sobre aqueles dados que poderiam então subsidiar algum tipo de ação específica, mais dirigida. O primeiro estudo realizado foi a identificação das escolas em que o desempenho escolar das crianças mostrava-se superior ao conjunto mediano, em cada disciplina e no conjunto das disciplinas. Procurou-se localizar as 10, 12, 15 escolas onde esse rendimento sobressaía, e também aquelas onde ele se apresentava mais comprometido. A idéia era associar a essas questões uma análise das condições próprias de funcionamento das escolas que estariam determinando melhor capacidade de atuação, ou então, os problemas que poderiam estar havendo para que as crianças estivessem apresentando um rendimento bastante deficitário, apesar de estarem numa Escola-Padrão. Outro estudo realizado foi sobre alguns municípios específicos. No caso, foram estudados os Municípios de Santos, São José dos Campos, Campinas e São Paulo. Em São Paulo, o estudo dividiu a análise dos resultados entre a Cidade de São Paulo propriamente dita e os Municípios que compõem a Grande São Paulo, tentando fazer uma comparação e ver o que estava acontecendo nas zonas mais centrais e mais periféricas.

Uma terceira análise aprofundou os resultados obtidos no componente curricular Língua Portuguesa, identificando-se nas provas habilidades específicas, ou seja: perguntava-se se as crianças tinham maior capacidade para responder itens de informação que dependem muito de memorização, ou itens que envolviam compreensão de conceitos ou relações, ou itens de aplicação. Aqui a discussão foi muito interessante porque, ao analisar os itens componentes da prova de Língua Portuguesa, juntamente com professores da rede e alguns especialistas, se constatou uma dificuldade muito grande, em certos casos, de dizer se um item era de compreensão ou de aplicação, ou mais de compreensão, ou mais de aplicação. Isso resultou em estudos e discussões que permitiram aperfeiçoar muito a compreensão do significado desses, por assim dizer, 'tipos de habilidades'. Nesse momento, trabalhou-se com vários conjuntos de professores que se transformaram em "juízes", para classificar o item como de "informação", de "compreensão" ou "aplicação". A partir de consensos após várias discussões, pode-se examinar como as crianças se comportavam. Se fosse verdade, como se diz, que nossas escolas só desenvolvem a memória das crianças, se o decorar é realmente o objeto de trabalho dos professores, então as crianças deveriam estar muito bem nos itens de informação que exigiam uma data, um nome de uma pessoa, e informações factuais. No entanto, não foi isso o que se encontrou. Na verdade, as crianças apresentaram, pelo menos nesse estudo,

muito melhor rendimento nos itens de "aplicação", ou seja, que envolviam um problema prático e uma certa compreensão conceitual para resolvê-los, e não nos itens de informação.

O quarto estudo desenvolvido em 1993 referiu-se a uma tentativa de verificar que tipo de padrão de avaliação os itens da prova apresentavam. Seria realmente mais uma análise dos itens que foram construídos. Para isso, aplicou-se, pela primeira vez no Brasil, a **Teoria da Resposta ao Item**. Nesse momento, a teoria não foi aplicada em toda sua possibilidade, pois não foi feita uma amostra adequada ao modelo, nem foi feita uma prova nos termos do modelo. Mas foi realizada uma avaliação desses itens quanto à sua consistência aplicando o modelo logístico-matemático da Teoria da Resposta ao Item. Isto permitiu verificar o comportamento desses itens e a potencialidade dessa metodologia, em termos de estimativa de desempenho em diferentes níveis de habilidades.

Realizados esses estudos complementares, esperava-se que a Secretaria de Educação do Estado de São Paulo continuasse o desenvolvimento do seu modelo de avaliação. Em 1993 estava prevista a extensão da avaliação para além das 306 Escolas-Padrão, ou seja, para todas as outras escolas que se tornaram Escolas-Padrão durante 1993, para se ter um ano base de estudo, a fim de se continuar avaliando posteriormente o desenvolvimento dos alunos dessas escolas.

No entanto, em 1993 mudou novamente o Secretário de Educação. E o novo Secretário, cauteloso, preferiu fazer algumas análises antes de dar prosseguimento à avaliação. Então, a avaliação só se fez em 1994, agora com uma proposta mais ampla. Nesta, foram incluídas não só as Escolas-Padrão (as 306 primeiras, mais uma amostra de todas as outras Escolas-Padrão), como as Não-Padrão, pois a intenção era tentar, agora, desenvolver um sistema mais amplo de avaliação do sistema que deveria abranger todos os tipos de escola do Estado de São Paulo.

Portanto, a avaliação conduzida em 1994 foi feita com uma amostra de Escolas-Padrão e Não-Padrão, abrangendo crianças de 4ª e 8ª séries. As crianças de 4ª série foram avaliadas em Língua Portuguesa, Redação e Matemática. As de 8ª séries foram avaliadas em Língua Portuguesa, Redação, Matemática, História e Geografia, o que permitiria, mediante algumas condições, uma comparação com as crianças avaliadas dois anos atrás.

Nesse momento, final de governo, foram feitas análises do desempenho por Delegacias Regionais, que à época existiam, por Delegacias de Ensino, por turno (diurno e noturno), comparando-se sempre as crianças de Escolas-Padrão e Escolas Não-Padrão. Nessa etapa também se aplicou um questionário aos alunos, assim como no ano de 92. Esse questionário foi um aperfeiçoamento daquele já aplicado anteriormente, levantando características dos alunos como sexo, tipo de escola onde estudaram, instrução dos pais, trabalho, repetência de séries, preferência por leituras, etc... Não foi

feito nenhum cruzamento desses questionários com os dados de rendimento. Eles apenas se destinaram, nesse momento, a fornecer um perfil compreensivo das condições de vida, de trabalho, familiares e de alguns valores dos alunos.

Tanto para a etapa 92 como para a de 94, a CENP preparou documento dirigido à Rede de Ensino detalhando aspectos da avaliação procedida e estimulando as Delegacias de Ensino a levar em conta os dados da avaliação para seus planos de ação pedagógica, lembrando que um programa de avaliação do rendimento dos alunos, por si só, não resolve os problemas educacionais. Enfatizou-se que se deveria enfrentar com critério e determinação os pontos críticos revelados em relação ao processo de ensino. Em ambas as etapas as provas foram analisadas e implicações pedagógicas foram apontadas, com recomendações quanto aos pontos críticos em cada disciplina avaliada.

Neste período entre 92 e 94, outro projeto de avaliação vinha sendo desenvolvido nas escolas estaduais com caráter de uma verdadeira pesquisa avaliativa: o *Projeto de Avaliação de Impacto do Ciclo Básico e da Jornada Única na Área Metropolitana de São Paulo*. A equipe de pesquisa fez um estudo longitudinal de acompanhamento de um mesmo segmento de crianças, durante 3 anos, com o objetivo complexo de analisar mudanças de aprendizagem e de características cognitivas ao longo do tempo. Esse foi um tipo de pesquisa em avaliação nunca antes desenvolvido no país, inclusive com um estudo paralelo de caráter antropológico. Este permitiu que se tenha uma compreensão profunda do que se passava no cotidiano das escolas e no desenvolvimento cognitivo das crianças. Não vou entrar em detalhes sobre esse projeto, pois ele será relatado neste Seminário, devendo ser objeto de uma discussão aprofundada.

Essas experiências havidas no Estado de São Paulo contribuíram para a formação e o aperfeiçoamento de um conjunto de pesquisadores e técnicos em avaliação de desempenho escolar em sistemas de ensino, tanto a nível de Secretaria como fora dela. Um trabalho sócio-comunicativo deveria ser desencadeado também para dar às avaliações dessa natureza o sentido que devem ter: alimento para revisões críticas porém construtivas e base para progressão. Ou seja, é preciso mudar a representação punitiva e depreciativa da avaliação e dar-lhe novo significado.