

# **OS PROFESSORES PODEM USAR A AVALIAÇÃO PARA MELHORAR O ENSINO?\***

## **CAN TEACHERS USE ASSESSMENT TO IMPROVE LEARNING?**

Paul Black\*\*

### **Resumo**

Este artigo discute dois diferentes tipos de propósitos da avaliação: a avaliação formativa, que serve para apoiar a aprendizagem dos alunos, e a avaliação sumativa, que é usada para saber o que os alunos aprenderam e para conceder uma nota em certificados ou diplomas. A avaliação formativa preocupa-se tanto com as interações entre professores e alunos, como com outros aspectos essenciais no processo avaliativo, dentre os quais o a articulação do planejamento dos professores com as necessidades de aprendizagem dos alunos. Os professores que estão acostumados apenas a falar para os alunos em vez de envolvê-los em um diálogo podem achar difícil mudar. Os alunos também precisam mudar de uma atitude passiva para um envolvimento ativo na aprendizagem. A avaliação formativa pode desgastar-se se professores e alunos estiverem muito preocupados com testes. Tal preocupação pode levá-los a focalizar o preparo para os testes em vez de criar bons hábitos de aprendizagem, que, de fato, poderiam ser uma melhor preparação para trabalhar melhor com os próprios testes.

**Palavras-chave:** Avaliação formativa. Avaliação sumativa. Testes.

### **Abstract**

This article discusses two different purposes of assessment: formative assessment is designed to support pupils' learning, whilst summative assessment is designed to review what has been learnt, perhaps to record it in certificates or diplomas. Formative assessment is concerned with the frequent interactions between teacher and pupils which are essential if the teacher's plans can be matched to the learning needs of the pupils. Teachers who are accustomed to simply telling pupils, rather than engaging them in dialogue, find it hard to change. Pupils also have to change from passive reception to active engagement in the learning. Formative work can be undermined if pupils or teachers are worry too much about summative tests; such worry can lead them to focus entirely on practising for the tests and not on the good habits of learning which would in fact be the best preparation for doing well in them.

**Keywords:** Formative assessment. Summative assessment. Tests.

### **A importância da avaliação formativa**

Promover o aprendizado das crianças é o principal objetivo das escolas. A avaliação está no coração desse processo. A avaliação pode garantir um quadro no qual os objetivos educacionais podem ser estabelecidos e o progresso dos alunos mapeado e expressado. Pode também promover uma base para planejar os próximos passos educacionais em resposta às necessidades das crianças. Quando estimulamos o diálogo entre os professores, a avaliação pode melhorar as habilidades profissionais e ajudar a escola como um todo a fortalecer o aprendizado dos conteúdos curriculares e, por meio disso, a duração do seu alcance. (DES, 1987).

A mensagem geral desse artigo é a de que a boa avaliação formativa pode ser uma ferramenta poderosa para aumentar os padrões do aprendizado, mas, em geral, isso tem sido pouco explorado nas escolas. Aí está, portanto, uma boa oportunidade para a melhoria. Entretanto, essa oportunidade

não pode ser conquistada apenas por modestas mudanças no trabalho existente.

A característica distintiva da avaliação formativa é a de que as informações obtidas por meio da avaliação são usadas, tanto pelos professores quanto pelos alunos, para modificar o seu trabalho visando a torná-lo mais eficaz. Existe pouco sentido em coletar informações a não ser para utilizá-las na prática e, uma vez que as informações certamente revelam a heterogeneidade das necessidades de aprendizado de uma turma, a ação necessária deve incluir alguma forma de ensino diferenciado. Esse ponto foi colocado muito claramente por Perrenoud (1991). A sua afirmação "as pessoas não desejam saber coisas sobre as quais não podem interferir" implica que a consciência da própria inabilidade para agir de acordo com evidências pode ser uma razão pelas quais tais evidências (sobre a heterogeneidade das necessidades de aprendizagem) não são coletadas.

\* Tradução do artigo "Can teachers use assessment to improve learning?", publicado no British Journal of Curriculum & Assessment, v. 5, n. 2, p. 7-11, 1995. Artigo traduzido e publicado com a autorização do autor e da Editora Hodder and Stoughton. Tradução de Fernando Zan Vieira. Revisão técnica de Jefferson Mainardes e Ana Cláudia Gomes.

\*\* Professor emérito do King's College (Londres). E-mail: [paul.black@kcl.ac.uk](mailto:paul.black@kcl.ac.uk)

Se a educação deve explorar ao máximo os benefícios potenciais das avaliações dos professores, então um trabalho de desenvolvimento extensivo com os mesmos é necessário para implementar a base de uma avaliação formativa efetiva nos programas de ensino. Tal trabalho deve envolver programas de longa duração para modificar o entendimento e os papéis dos professores e alunos, com uma ênfase específica no desenvolvimento da compreensão e da responsabilidade dos alunos para a sua autoavaliação.

### Avaliação formativa e sumativa

Um dos principais obstáculos ao desenvolvimento da avaliação formativa é que ela é apenas uma das três funções da avaliação, que são: a) a assistência direta ao ensino, b) a certificação individual dos estudantes, e c) a responsabilização pública das instituições e dos professores (*accountability*).

A segunda e a terceira função influenciam a primeira. A função de responsabilização pode afetar o *status* e o emprego dos professores, enquanto que a função da certificação influencia as “oportunidades de vida” dos seus alunos. Para atender o propósito da responsabilização não há necessidade de se testar cada estudante individualmente. Apesar disso, na prática, muitos gestores do sistema educacional fazem uso dos mesmos testes para dar um resultado para cada aluno individualmente, bem como para obter as médias do desempenho das classes ou das escolas. Muito do investimento em avaliações e testes, seja em operações práticas ou em pesquisa e desenvolvimento, tem sido empregado para as funções de certificação e responsabilização, negligenciando a função formativa.

Os testes externos, aplicados pelo governo ou outras agências, podem orientar o trabalho da sala de aula e, assim, distorcer o ensino no sentido de que as condições para uma adequada avaliação formativa podem deixar de existir. Os professores acreditam que preparar os alunos para um teste específico, ao focalizar a concentração nos procedimentos e exigências da tarefa, levaria à obtenção de escores mais altos. Porém, os efeitos negativos dos testes externos sobre o ensino são bem conhecidos: a aprendizagem é distorcida, pois os professores concentram a atenção sobre aqueles tópicos do conteúdo fáceis de serem testados; portanto, o nível cognitivo da sala de aula é rebaixado; os alunos têm de trabalhar muito rapidamente e superficialmente para uma aprendizagem efetiva; os métodos criativos e inovadores, bem como os conteúdos, são reduzidos. Este ensino mais superficial e limitado desmotiva muitos alunos e, por outro, lado, recompensa os que trabalham mais focalizados naquilo que o sucesso dos testes exige.

Dentro do contexto até agora indicado, não é surpreendente que muitos professores vejam a avaliação com aversão e suspeita. Exemplos do uso da avaliação formativa como um poderoso componente em um programa no qual um conjunto de planos e práticas de ensino são definidos para ajudar os alunos a aprenderem e tornarem-se melhores aprendizes no futuro são difíceis de serem encontrados. A imagem geral que emerge da observação da prática é o negligenciamento da avaliação. (ver BLACK, 1993a).

As evidências sobre a situação atual das práticas avaliativas dos professores mostram que há necessidade de se estabelecer um grande investimento na formação continuada de docentes para que se possa estabelecer a avaliação formativa. As mudanças não são feitas facilmente. Tentativas de fortalecer a avaliação no currículo nacional do Reino Unido tiveram sucesso limitado, em parte porque os professores interpretaram o seu papel avaliativo apenas em termos de avaliação sumativa,<sup>1</sup> ou seja, as novas exigências para as avaliações dos professores foram entendidas como uma demanda para realizar seus próprios testes sumativos (HARLEN; QUALTER, 1991; SCOTT, 1991). Butler e Beasley (1987) relataram que em Queensland, em 1972, a mudança da certificação externa para aquela baseada na escola teve pouco efeito na prática de sala de aula. Foi apenas após uma mudança liderada pelo Estado em 1985 do enfoque sumativo baseado em norma para o enfoque formativo em critério de referência, que as práticas lentamente mudaram. Uma característica do trabalho desses autores foi que os professores consideraram difícil efetivar o *feedback*<sup>2</sup> a partir dos resultados da avaliação. Torrance (1986) apresenta uma conclusão similar sobre os professores na Grã-Bretanha.

<sup>1</sup> Avaliação sumativa é o teste formal realizado com o objetivo de saber o que foi efetivamente aprendido pelos alunos, bem como para se obter uma nota. A avaliação sumativa sintetiza o desenvolvimento dos alunos em um tempo específico, por exemplo, ao final de uma unidade de trabalho, ao final de um bimestre ou trimestre, etc. Os resultados dessa avaliação formal podem ser usados como diagnóstico para identificar pontos fracos e dificuldades dos alunos para, a partir disso, desenvolver a avaliação formativa. Já a avaliação chamada formativa enfatiza o acompanhamento do processo de aprendizagem dos alunos com o objetivo de planejar como auxiliá-los. Alguns estudiosos de avaliação consideram que a distinção entre avaliação sumativa e formativa é falsa, uma vez que as informações obtidas por meio da avaliação sumativa podem subsidiar a avaliação formativa e todas as formas de avaliação possuem algum elemento formativo. (Nota dos revisores)

<sup>2</sup> No contexto das discussões sobre avaliação formativa, *feedback* (realimentação) é o apoio e orientação que os professores podem oferecer aos alunos com o objetivo de superar lacunas e dificuldades de aprendizagem. Parte-se da ideia de que as informações obtidas por meio da avaliação devem ser consideradas como um ponto de partida para o planejamento das situações de ensino que, supostamente, ajudariam os alunos a superarem as eventuais dificuldades. Ver, a respeito, o artigo “Para uma teoria da avaliação formativa”, de Domingos Fernandes (Revista Portuguesa de Educação, v. 19, n.2, 2006). (Nota dos revisores)

Crooks (1988), em um levantamento da literatura internacional sobre o impacto das práticas avaliativas nos estudantes, contrastou os muitos efeitos positivos da boa avaliação com evidências de que a avaliação na sala de aula muitas vezes encoraja o aprendizado repetitivo e superficial, contexto no qual o papel das notas é fortemente enfatizado.

Tentativas de fortalecer a avaliação que os professores realizam podem cair na controvérsia da relação entre as funções formativa e sumativa. Embora a avaliação formativa deva ser perseguida por seu propósito principal de *feedback* do processo de aprendizagem, ela também pode produzir informações úteis para satisfazer propósitos sumativos. Isso pode ajudar a proteger a avaliação formativa das pesadas influências de testes externos de “alto risco”, aplicados pelo governo ou outras agências. Entretanto, pode haver tensão entre o papel de aconselhamento e de julgamento de um professor, e alguns estudiosos da avaliação argumentam que os mesmos instrumentos e procedimentos avaliativos não podem servir para duas funções diferentes e que a avaliação pode ser fortalecida pela separação destes papéis. (HARLEN *et al.*, 1992). Contudo, a necessidade de deixar claro que os alunos obtêm bons resultados em testes sumativos frequentemente significa que os professores não sentem que eles podem usar a avaliação formativa em um sentido positivo, e esses autores não propõem formas para resolver esse problema. (RESNICK; RESNICK, 1992; BLACK, 1993b). Embora a prática sumativa seja dominante, ela pode esconder porque os testes externos são um modelo ineficiente para a avaliação formativa, pelos seguintes motivos:

- em testes sumativos, a necessidade de um resultado único total significa que dados bastante díspares (por exemplo, para o prático e para o teórico) tem de ser adicionados de forma arbitrária: a avaliação formativa não pode fazer isso – ela se preocupa com os padrões de necessidade de aprendizagem para alunos individualmente;
- o trabalho sumativo tem de insistir em padrões de uniformidade e confiabilidade não requeridos no trabalho formativo e que inibem a liberdade e a atenção para necessidades individuais de que o trabalho formativo necessita; em particular, a prática formativa não tem que ser “justa” - sua prioridade é a identificação das necessidades de cada aluno, o que envolve tratar os alunos de formas diferenciadas;
- os propósitos sumativos podem demandar a coleta de evidências documentais por resultados e, assim, aumentar a carga de trabalho, enquanto o trabalho formativo exige ação sobre os dados, mais do que a armazenagem destes.

## Avaliação e aprendizado

O princípio chave enfatizado aqui é o de que a avaliação formativa deve estar intimamente conectada com os processos de ensino e aprendizagem. A questão foi claramente destacada por Linn (1989):

A concepção de testes úteis para decisões instrutivas feitas na sala de aula requer uma integração entre o teste e a instrução. Isso também requer uma concepção clara do currículo, dos objetivos e do processo de instrução. E requer uma teoria de instrução e de aprendizado e uma compreensão muito melhor do processo cognitivo dos aprendizes. (p. 5)

Essa necessidade de avaliação como suporte para o ensino diretamente dirigido para a melhoria do pensar tem sido explorada por Resnick e Resnick (1992). Ao se testar com questões curtas, “atomizadas” (fragmentadas), reforçam-se as pressuposições sobre o aprendizado que a psicologia atual abandonou. Um exemplo é a falsa aceção de que um assunto complexo pode ser ensinado de forma fragmentada e as partes testadas separadamente; outro é a falsa pressuposição de que uma ideia que é comum à ação em muitos contextos pode ser ensinada de forma mais econômica ao ser apresentada em um isolamento abstrato, de forma a ser empregada em muitas situações.

Se o *feedback* dado pelas frequentes avaliações deve melhorar o ensino, então deve refletir objetivos importantes. E se as informações obtidas por meio da avaliação devem ser significativas em relação ao aprendizado dos objetivos do currículo, então esta deve ser referenciada por critérios. O desenvolvimento de critérios de referência tem enfrentado muitas complicações. A dificuldade comum é a de que, se os critérios são vastos e vagos, o valor formativo deles pode ser perdido e, por outro lado, se eles forem feitos de forma muito específica, os professores se afogarão em seus números arrebatadores, enquanto o ensino será fragmentado, fazendo com que as características gerais e relacionais se percam. (POPHAM, 1987).

Se a avaliação deve guiar o aprendizado durante o desenvolvimento de um programa, então deve refletir os critérios que são estabelecidos em uma sequência de aprendizagem. Em princípio, isso pode não ter valor como guia para acompanhar o progresso dos sujeitos individualmente e para localizar obstáculos ao seu aprendizado. Entretanto, isso não pode ser feito sem conhecimento bem fundamentado das sequências nas quais os alunos aprendem. (SIMON *et al.*, 1994). Onde tal conhecimento não está disponível, alguns autores tentam especificar a progressão por meio de afirmações que são vagas ao ponto de não servirem para orientar os professores. Por exemplo: eles podem dizer que uma etapa

da aprendizagem de um tópico é “entender o tópico minuciosamente” e o próximo é “entender o tópico mais minuciosamente”.

### Validade e confiabilidade

Testes externos gozam de maior confiança pública do que deveriam. Os limites de sua validade e confiabilidade não são compreendidos. Um conceito amplo de validade é ao mesmo tempo necessário e persuasivo em qualquer tentativa de melhorar a qualidade das avaliações. A afirmação inicial da revisão feita por Messick (1989) deu uma definição legítima: “validade é um julgamento avaliativo integrado do grau no qual a evidência empírica e as justificativas teóricas fundamentam a adequação e a propriedade de inferências e ações baseadas nos escores de testes ou outros modos de avaliação.” (p. 13)

O escopo das *inferências* e *ações* tem sido expandido em estudos recentes. Por exemplo, Moss (1992) destacou que os usos a serem feitos e os efeitos negativos no ensino são dois aspectos que devem ser considerados na avaliação da validade. Assim, se uma avaliação poderia ser usada para reforçar estilos de ensino impróprios para os objetivos da educação, sua validade poderia ser prejudicada apenas nesses termos. Quanto mais próxima a atividade avaliativa estiver da atividade real na qual seus resultados podem ser considerados relevantes, mais provável será a satisfação dos critérios válidos. À luz disso, a avaliação na sala de aula tem mais chance de sucesso do que testes formais, cronometrados e escritos.

A confiabilidade das avaliações na sala de aula é inerentemente difícil de ser explorada em razão da falta de qualquer medida independente das “reais” obtenções dos alunos. Hoge e Coladarc (1989) dão um amplo espectro da literatura que trata do julgamento dos professores sobre o desempenho acadêmico. Eles concluíram que as correlações internas podem exceder as validades convergentes e concorrentes dos testes psicológicos.

Igualmente, a confiabilidade limitada dos testes formais não é apreciada e, em muitos sistemas nacionais, tampouco medida. A consistência interna é facilmente pesquisada, mas, por uma variedade de razões, os alunos muito frequentemente têm resultados muito fracos nos testes formais como um todo. Para alguns testes de Ciências, Gauld (1980) descobriu que os alunos: a) frequentemente interpretam errado o enunciado da questão; b) parecem incompetentes por causa de uma simples falha em um processo complexo, uma vez que falham em usar o que eles sabem por que julgam o enunciado irrelevante e c) podem ser subavaliados em razão de o avaliador fracassar em compreender o pensamento

que está por trás de respostas fora do padrão.

Entretanto, para avaliações formativas, os critérios de validade e confiabilidade necessitam ser reformulados. A validade da avaliação formativa deve ser examinada em relação à sua efetividade na melhoria do ensino.

### Melhorando a avaliação formativa: as necessidades do professor

A tecnologia da coleta de dados sobre o progresso dos alunos está apenas começando a ser desenvolvida. Muitos professores têm usado uma variedade de fontes de maneira informal – mas isso pode ser enganoso. Um exemplo comum é o do professor que usa as respostas de apenas alguns alunos em um diálogo na sala de aula como reafirmação de que o plano de aula pode prosseguir. O que se necessita é aperfeiçoar a prática com uma visão de se obter dados mais abrangentes. Os livros de Fairbrother *et al.* (1993, 1995) dão exemplos de formulários de autoavaliação que colocam em prática objetivos de aprendizado para o trabalho na sala de aula nos quais cada aluno deve relatar seu julgamento sobre se esse objetivo foi atingido. Tais formulários dão ao professor o critério informativo essencial para a avaliação formativa. Além disso, por oferecerem evidência escrita de forma sistemática, eles liberam o professor da pressão de anotar e registrar inteiramente a evidência efêmera dos eventos da sala de aula. As observações pelos professores podem ter valor único: alguns consideraram surpreendentemente útil suspender suas intervenções ativas de ensino por um tempo – deixando claro para uma turma o que ela está fazendo e por que – e se concentrar apenas na observação e na escuta com alguns poucos alunos. (ver CAVENDISH *et al.*, 1990; CONNOR, 1991).

Entretanto, muito mais aspectos estão envolvidos além coleta de evidências sobre “erros”. Como Bonniol (1991) aponta, o professor necessita entender suas causas, que são complexas mesmo para os melhores especialistas em ensino. Além disso, a avaliação formativa não pode ser somente adicionada aos esquemas de trabalho existentes; ela precisa ser construída dentro de tais esquemas, e seu uso para orientar a aprendizagem dos alunos de acordo com suas diferentes necessidades pode somente ocorrer se o plano de ensino permitir a organização flexível que isso exige. Esse é o aspecto mais desafiador do trabalho formativo.

Uma maneira de lidar com as diferentes necessidades de forma macroscópica e a longo prazo é dividir os alunos em classes diferentes de acordo com os desempenhos anteriores. Essa é uma maneira tosca e não pode lidar com necessidades diversas e de curto prazo. Respostas mais refinadas

podem ser adotadas dentro de cursos organizados em módulos. Para dois terços do tempo do módulo, métodos corriqueiros e informais de avaliação são usados, mas há uma avaliação mais formal ao final dos dois terços para determinar um ensino diferenciado para a última parte: aqueles que compreenderam as ideias principais seguem para um trabalho mais extensivo, enquanto aqueles com dificuldades específicas trabalham com material projetado para ajudar em seus obstáculos. Dessa forma, alguns atingirão sucesso dentro de uma extensão maior do que outros, mas espera-se que todos atinjam o mínimo necessário para fazer sentido num trabalho futuro. (ver HARRY BLACK, 1993). Há outras abordagens menos formais e mais flexíveis, como, por exemplo, dar oportunidades para os alunos revisarem seu aprendizado com o objetivo de identificar suas dificuldades em conteúdos específicos ou, em atividades futuras, quando eles tiverem de usar essas ideias em uma nova atividade, dar tempo e ajudá-los a repetir o aprendizado original antes de começar um novo trabalho.

Algo que influencia essa última questão e muitas outras é em que medida os programas de ensino são flexíveis ao invés de rígidos. Os trabalhos menos rígidos, seja em investigações práticas, projetos de biblioteca ou atividades de ensino de ciências cuidadosamente pensados sobre a seleção e definição de tarefas, podem alcançar a diferenciação ao sintonizar a tarefa à capacidade e ao progresso do aluno. Entretanto, aparentemente os alunos mais fortes são os que mais se beneficiam de um trabalho mais flexível. Àqueles que progridem mais lentamente falta confiança e uma visão clara de suas necessidades, exigida para um autodirecionamento efetivo. Assim, para eles, um plano de trabalho mais rígido pode ser necessário.

O uso efetivo do *feedback* requer o julgamento pelo professor, a confiança e a flexibilidade no gerenciamento de um planejamento curricular que só pode resultar da propriedade de tal plano. Assim, ao que tudo indica, idealmente, qualquer esquema para incorporar boas oportunidades formativas deve ser construído pelos professores para si mesmos. Para garantir a consistência do auxílio aos alunos, os professores devem partilhar informações e práticas de avaliação. As dinâmicas interativas entre os alunos podem servir de base para as ações que podem ser desenvolvidas com vistas à formação da equipe da escola.

A incorporação da avaliação formativa na ação docente vai além da aquisição, pelos professores, de habilidades necessárias, pois tal incorporação envolve mudanças profundas de papel. Em particular, uma vez que muitos alunos podem adquirir hábitos de fazer apenas o necessário para ir em frente ou deixar de acreditar que podem ser compe-

tentes quanto ao conteúdo, o contrato entre professor e aluno tem de ser reformulado. Deslocar mais responsabilidade aos alunos é um meio de reduzir o fardo e existem outras razões mais importantes para tal deslocamento, conforme se discutirá adiante.

### **Melhorando as avaliações formativas – as necessidades dos alunos**

Tanto Perrenoud (1991) quanto Bonniol (1991), em seus artigos sobre os princípios da avaliação formativa, enfatizam que o desenvolvimento deste tipo de avaliação implica mudanças de papel tanto dos professores quanto dos alunos. Os alunos devem entender o quadro de referência do professor e o modelo de ensino que dá sentido aos critérios refletidos na avaliação. Tal compreensão pode construir a confiança dos alunos em sua própria capacidade de que podem fazer algo em relação às suas dificuldades.

Lacunas na compreensão entre estudantes e professores permitem o surgimento de fraquezas em quaisquer avaliações informais – ponto ilustrado em uma análise de Perrin (1991) sobre um estudo de alunos primários no Cantão de Genebra. Esses alunos acreditavam que as avaliações se destinavam à escola e a seus pais e não a eles mesmos. Os alunos mais fracos acreditavam que o propósito era o de fazê-los trabalhar mais. Uma vez que a avaliação não era usada para lhes dizer como trabalhar de forma diferente, eles a viam como uma forma de pressão, o que os deixava ansiosos. Como consequência de tal evidência, o Cantão decidiu reduzir seus testes sumativos e melhorar o papel formativo da avaliação. A carga de trabalho de avaliações sumativas freqüentes sobre os estudantes pode não ser vista dessa forma onde há informação e oportunidade para melhoria do desempenho, como demonstrado pelo estudo de Iredale (1990) sobre o desempenho dos alunos em um esquema de avaliação de Ciências.

Os alunos não podem exercer uma parte efetiva em suas próprias avaliações, exceto em um programa de longo prazo projetado para ajudá-los a adquirir e sustentar uma visão geral dos objetivos de seu aprendizado e aplicar os critérios aos seus próprios progressos. Os alunos apenas adquirirão tal visão se essa for explicitamente ensinada e tal ensino exige tempo para que se alcance sucesso. (BAIRD; NORTHFIELD, 1992; FAIRBROTHER *et al.*, 1995). Um elemento necessário é a tradução dos objetivos do currículo numa linguagem que todos os alunos possam entender e em tal nível de detalhe que os ajude a relacioná-los diretamente aos seus esforços de aprendizado. Segue-se ainda que os objetivos devam ser igualmente alcançáveis em curto prazo

e adequadamente modestos em relação às possibilidades de sucesso dos alunos.

A capacidade dos alunos em julgar seu próprio trabalho é de importância fundamental na aprendizagem. Zessoules e Gardner (1991), descrevendo o Harvard Project Zero, enfatizaram que a autoavaliação no processo de aprendizado é um componente crucial para o desenvolvimento de compreensões complexas através de hábitos mentais reflexivos. Assim, esse projeto tentou preparar condições apropriadas a garantir que os alunos seriam ativos e atenciosos em sua própria avaliação. Outros argumentaram que a metacognição, no sentido de consciência e autodireção sobre a natureza de seu trabalho de aprendizado, é essencial para o desenvolvimento dos alunos na aprendizagem conceitual. (ver BROWN, 1987; WHITE; GUNSTONE, 1989).

Claramente, o envolvimento dos alunos pode ser mais viável para os professores promoverem um programa de avaliação formativa. Entretanto, esse envolvimento também modifica tanto o papel do aluno como estudante e a natureza da relação entre professor e aluno, fazendo este último mais responsável pelo aprendizado. Assim, a avaliação formativa melhorada pode levar a mudanças de significância mais ampla – mudanças que seriam de uma grande ajuda ao desenvolvimento pessoal dos alunos e que poderiam ser parte de qualquer programa para ajudá-los a serem aprendizes mais efetivos.

## Conclusões

Quando os professores estão dispostos a experimentar melhorias, uma instituição de formação pode oferecer um fórum em um terreno neutro onde eles podem receber sugestões e ideias que a literatura tem a oferecer e podem trocar planos e experiências com outros trabalhando com os mesmos fins.

As questões indicadas nesse artigo poderiam fornecer uma base para parte de tal programa. No entanto, uma característica essencial seria que o trabalho dos professores em planejar as mudanças em suas próprias salas de aula necessitaria de um acompanhamento regular de encontros onde sucessos e fracassos pudessem ser relatados e comparados.

Em suas responsabilidades avaliativas, os professores devem conciliar as necessidades de aprendizagem de seus alunos, o que deve ser a primeira preocupação a cumprir, apesar das pressões para obter bons resultados em testes nacionais obrigatórios e expectativas que eles irão funcionar no limite das políticas escolares e cumprir as expectativas dos pais. Essas exigências são em geral inconsistentes.

Se eles devem estar aptos para fazer as mudanças radicais que são propostas neste artigo, os professores necessitarão trabalhar juntos para apoiar uns aos outros de forma a enfrentar as pressões e pedir ajuda e compreensão solidária, que pode ser obtida daqueles que estão fora da escola.

Finalmente, a evidência do impacto de todo o contexto sistêmico de testes e avaliações das práticas dos professores mostra que a ação pública e política é essencial para estabelecer e apoiar tanto os contextos sistêmicos quanto a compreensão pública necessários à sobrevivências desses novos desenvolvimentos.

## Referências

- BAIRD, J. R.; NORTHFIELD, J. R. (Eds.) **Learning from the PEEL experience**. Monash University, 1992.
- BLACK, H. Assessment for learning. In: NUTTALL, D. L. (Ed.) **Assessing Educational Achievement**. Falmer Press, 1986.
- BLACK, H. Assessment: a Scottish model. In: FAIRBROTHER, R.; BLACK, P. J.; GILL, P. (Eds.) **TAPAS: Teacher Assessment of Pupils: Active Support**. King's Education Papers, n.3. London: C.E.S. King's College, 1993.
- BLACK, P. J. Formative and summative assessment by teachers. **Studies in Science Education**, n.21, p. 49-97, 1993a.
- BLACK P. J. Assessment policy and public confidence: comments on the BERA policy task group's article "Assessment and the improvement of education". **The Curriculum Journal**, v.4, n.3, p. 421-427, 1993b.
- BONNIOL, J. J. The mechanisms regulating the learning process of pupils: contribution to a theory of formative assessment. In: WESTON, P. (Ed.) **Assessment of pupils achievement: motivation and school success**. Swets and Zeitlinger, 1991.
- BROWN, A. Metacognition, executive control, self-regulation and other more mysterious mechanisms. In: WEINERT, F. E.; KLUWE, R. H. **Metacognition, motivation and understanding**. Lawrence Erlbaum, 1987.
- BUTLER, J. E.; BEASLEY, W. F. The impact of assessment changes on the science curriculum. **Research in Science Education**, n.17, p. 236-243, 1987.
- CAVENDISH, S.; GALTON, M.; HARGREAVES, L.; HARLEN, W. **Observing activities**. Paul Chapman, 1990.
- CONNOR, C. **Assessment and testing in the primary school**. Falmer Press, 1991.
- CROOKS, T. J. The impact of classroom evaluation practices on students. **Review of Educational Research**, v.58, n.4, p. 438-481, 1988.
- DES. **National curriculum: task group on assessment and testing: a report**. Department of Education and Science and Welsh Office, 1987.
- FAIRBROTHER, R. W.; BLACK, P. J.; GILL, P. (Eds.) **Tea-**

- cher assessment of pupils:** active support for teachers. London: Centre for Educational Studies, King's College, 1993.
- FAIRBROTHER, R. W.; BLACK, P. J.; GILL, P. (Eds.) **Teachers assessing pupils:** lessons from science classrooms. Association for Science Education, 1995.
- GAULD, C. F. Subject-oriented test construction. **Research in Science Education**, n.10, p. 77-82, 1980.
- HARLEN, W.; QUALTER, A. Issues in SAT development and the practice of teacher assessment. **Cambridge Journal of Education**, v.21, n.2, p. 141-152, 1991.
- HARLEN, W.; GIPPS, C.; BROADFOOT, P.; NUTTALL, D. Assessment and the improvement of education. **The Curriculum Journal**, v.3, n.3, p. 215-230, 1992.
- HOGUE, R. D.; COLADARCI, T. Teacher-based judgments of academic achievement: a review of literature. **Review of Educational Research**, v.59, n.3, p.297-313, 1989.
- IREDALE, C. Pupils' attitudes towards GASP. **School Science Review**, v.72, n. 258, p.133-137, 1990.
- LINN, R. L. Current perspectives and future directions. In: LINN, R. L. (Ed.) **Educational measurement**. 3rd edition. Collier Macmillan, 1989.
- MESSICK, S. Validity. In: LINN, R. L. (Ed.) **Educational measurement**. 3rd edition. Collier Macmillan, 1989.
- MOSS, P. A. Shifting conceptions of validity in educational measurement: implications for performance assessment. **Review of Educational Research**, v.62, n.3, p. 229-58, 1992.
- PERRENOUD, P. Towards a pragmatic approach to formative evaluation. In: WESTON, P. (Ed.) **Assessment of pupils achievement:** motivation and school success. Swets and Zeitlinger, 1991.
- PERRIN, M. Summative evaluation and pupil motivation. In: WESTON, P. (Ed.) **Assessment of pupils achievement:** motivation and school success. Swets and Zeitlinger, 1991.
- POPHAM, W. J. Two-plus decades of educational research. **International Journal of Education Research**, v.11, n.1, p. 31-41, 1987.
- RENSNICK, L. B.; RENSICK D. P. Assessing the thinking curriculum: new tools for educational reform. In: GIFFORD, B. R.; O'CONNOR, M. C. (Eds.) **Changing assessments:** alternative views of aptitude, achievement and instruction. Kluwer, 1992.
- SCOTT, D. Issues and themes: coursework and coursework assessment in the GCSE. **Research Papers in Education**, v.6, n.1, p.3-19, 1991.
- SIMON, S.; BLACK, P. J.; BROWN, M.; BLONDEL, E. Progression in understanding the equilibrium of forces. **Research Papers in Education**, v.9, n.2, p. 249-280, 1994.
- TORRANCE, H. Expanding school-based assessment: issues, problems and future possibilities. **Research Papers in Education**, v.1, n.2, p. 48-59, 1986.
- WHITE, R. T.; GUNSTONE, R. F. Meta-learning and conceptual change. **International Journal of Science Education**, n.11, p. 577-586, 1989.
- ZESSOULES, R.; GARDNER, H. Authentic assessment: beyond the buzzword and into the classroom. In: PER-RONE, V. (Ed.) **Expanding student assessment**. Alexandria, Virginia, USA: Association for Supervision and Curriculum Development, 1991.

Recebido em 10/10/2008

Autorizado para publicação em 27/03/2009