

A avaliação formativa do desempenho docente¹

Dylan Wiliam

Universidade de Londres - Reino Unido

Resumo

A argumentação deste trabalho é de que os esforços para aumentar o desempenho educacional devem se concentrar principalmente na qualidade dos professores já em atividade nas escolas, por meio de uma ênfase sustentada no desenvolvimento profissional docente focado na prática de sala de aula. Este trabalho primeiro faz uma breve revisão da razão pela qual um melhor desempenho educacional é importante, tanto para os indivíduos quanto para a sociedade, e em seguida explora as influências do rendimento estudantil, chegando à conclusão de que a qualidade dos professores é uma das, senão a mais importante variável para a maioria dos sistemas educativos. Esforços no sentido de implementar a qualidade docente, seja por meio da melhora dos professores recém-chegados à profissão, ou da remoção dos ineficazes, demonstram ter pouco, se algum, impacto, levando à conclusão de que quaisquer esforços sérios para melhorar o desempenho educacional precisam focar no aumento da qualidade dos professores existentes, através do desempenho profissional, e o restante do artigo descreve como isso pode ser feito, lançando mão da avaliação formativa do desempenho docente. Em particular, o artigo revisa a pesquisa existente sobre feedback, mostrando que, para ser eficaz, precisa levar em consideração o contexto em que é dado, a relação entre os envolvidos na intervenção, e, o mais importante, a reação do professor ao receber esse *feedback*. Em seguida argumenta-se que o *feedback* pode ser mais efetivo quando é incorporado a um quadro teórico mais amplo (começando com uma visão clara da prática efetiva, evocando evidências de desempenho, fornecendo *feedback*, e ativando os docentes como recursos de aprendizagem, tanto para eles mesmos quanto para seus pares) e apoiados de modo correto. De maneira específica, a melhoria na prática é mais provável e mais plausível de ser sustentada quando os professores têm a escolha sobre quais aspectos da prática devem priorizar, a flexibilidade acerca de como implementar ideias nas salas de aula, o tempo para a realização de mudanças incrementais na prática, a responsabilidade por melhorá-la, e apoio para a sua realização.

1 Traduzido por Ana Cristina Vieira Lopes Romeiro

Introdução

A argumentação deste artigo se baseia em quatro propostas. Em primeiro lugar, fazem-se necessárias maiores realizações educacionais tanto para os indivíduos quanto para a sociedade. Em segundo lugar, para conseguir maiores realizações educacionais se requer maior qualidade nos professores. Em terceiro lugar, uma maior qualidade nos professores demanda investir nos professores que já se encontram trabalhando em nossas escolas. Em quarto lugar, esse investimento precisa tomar um formato radicalmente distinto do desenvolvimento profissional que comumente os professores receberam. A seção 1 revisa brevemente a investigação sobre os benefícios da educação para os indivíduos e à sociedade, e as seções 2 e 3 mostram que a qualidade do ensino nas escolas é um dos fatores mais importantes na hora de determinar quanto estão aprendendo os estudantes na escola. As seções 4 e 5 mostram que a evidência disponível sugere que melhorar a qualidade daqueles que ingressam na profissão docente e destituir os professores ineficientes é difícil por natureza, e pouco provável que assegure uma melhora na qualidade do ensino que se faz necessária, o que conduz à premissa central da seção 6, que consiste em que a chave da melhora nos resultados educativos é o investimento nos professores que já se encontram trabalhando em nossas escolas. Os leitores que estejam dispostos a aceitar esta premissa podem, portanto, ir direto para a seção 7, que mostra que o *feedback*, apesar de ter um alto potencial, costuma ser contraproducente, e requer que se tenha atenção cuidadosa com o contexto dado, sendo explorado na seção 8. A seção 9 propõe um modelo de avaliação formativa que identifica cinco “estratégias chave” de avaliação formativa; e a seção 10 delimita como podemos começar a pensar sobre a validade das avaliações formativas de desempenho docente. A seção 11 conclui o artigo com alguns princípios para a implantação da avaliação formativa do desempenho docente.

1. A educação é importante para os indivíduos e para a sociedade

Na realidade, a educação é importante tanto para os indivíduos quanto para a sociedade. Para os indivíduos, os benefícios incluem melhores rendimentos ao longo da vida, melhor saúde e uma vida mais longa. Ademais, os indivíduos mais educados têm menos possibilidades de se tornar pais adolescentes, ou de se envolver com o sistema judicial penal (Wiliam, 2011). Para a sociedade, os benefícios são, sem dúvida, maiores. As pessoas mais educadas são mais tolerantes e mais propensas a se comprometer em diversas atividades a favor da sociedade (Feinstein Et Al., 2008), e também contribuem mais com o crescimento econômico. Por exemplo, Hanushek e Woessman (2010) estimaram que um aumento nas notas alcançadas pelos jovens mexicanos de 15 anos nas provas PISA administradas pela OCDE teria um valor corrente líquido de 5 bilhões de dólares. Mas o mais surpreendente é que, se todos os estudantes mexicanos atingissem uma pontuação de 400 no PISA – o cálculo da OCDE do nível de habilidade requerido para funcionar de maneira efetiva em uma sociedade complexa – teria um valor corrente líquido de 26 bilhões de dólares para

a economia mexicana.

2. A qualidade do ensino é a variável crucial

Em várias jurisdições (por exemplo, Estados Unidos, Inglaterra) tem-se dado ênfase em melhorar a qualidade das escolas. Isto apresenta um atrativo intuitivo – todos os pais querem que seus filhos frequentem boas escolas -, mas o surpreendente é que a maioria dos países, contanto que se vá à escola, não importa muito a qual escola seja. Nos termos das provas PISA da OCDE, cerca de 18% da variação nas pontuações alcançadas pelos estudantes no México podem ser atribuídas à escola que o aluno frequenta (McGaw, 2008); o restante é constituído por variações dentro das escolas (61%), variações entre escolas que se explicam pelo contexto social da escola (16%) e variações entre escolas que se explicam pelo contexto social dos estudantes que frequentam a alguma delas (5 por cento).

Desde já, isso não deve ser interpretado como se significasse que não existem más escolas. Sem sombra de dúvidas existem más escolas, e bem poderia ser que para algumas escolas a melhor opção seria reconstruí-las substituindo os dirigentes, e inclusive a maioria dos professores, possivelmente (Bryk Et Al, 2010). O que acho importante compreender é que as diferenças no progresso dos estudantes em diferentes escolas no México são pequenas, e muito menores do que o tipo de melhora necessária para enfrentar os desafios do século XXI. Dito de outra maneira, ainda que cada escola fosse tão boa quanto a melhor de todas, indubitavelmente a educação no México melhoraria, mas isto por si só não produziria o tipo de melhorias necessárias.

O motivo para que as diferenças entre escolas sejam tão pequenas (uma vez que se têm em conta as diferenças dos contextos sociais dos estudantes que as frequentam) é que o fator decisivo do progresso dos estudantes na escola é, ainda que óbvio, a qualidade da instrução que recebem. O menos óbvio é que a qualidade da instrução que os estudantes recebem conforme avançam na escola é muito variável. Em todas as escolas existe bom ensino e ensino que não é tão bom, mas conforme os estudantes avançam nos estudos, a qualidade mediana da instrução que recebem durante suas carreiras não varia muito de uma escola para outra.

A qualidade do ensino depende de muitos fatores:

- O tempo que os professores dedicam para planejar o ensino.
- O tamanho das turmas.
- Os recursos disponíveis.
- As capacidades do professor.

Todos esses fatores são importantes, no entanto a qualidade do professor parece ser especialmente importante.

3. A qualidade do ensino não é o mesmo que a qualidade dos professores

A qualidade do ensino depende de certas variáveis, como a quantidade de tempo que os professores dedicam para preparar as aulas, o tipo de recursos disponíveis, o número de estudantes por turma, as habilidades do professor etc. Em alguns sistemas (por exemplo, Japão, Finlândia), o número de horas que os professores gastam ensinando realmente está abaixo das 700 anuais, enquanto em outros (por exemplo, Estados Unidos, Chile) está acima das 1000 (OCDE, 2013a, p. 396). Portanto, é importante se dar conta de que a qualidade do ensino é melhor do que a qualidade dos professores. Assim sendo, ao que parece, a qualidade dos professores em um sistema é uma variável crucial. Se colocamos os professores em três grupos de igual tamanho, abaixo da média, média, e acima da média, então os estudantes que são ensinados por um professor acima da média terão 50% mais êxito, e aqueles ensinados por um professor abaixo da média terão 50% menos do que os estudantes ensinados por um professor mediano (Hanushek, 2011). Os professores mais eficazes são, assim sendo, ao menos três vezes mais eficazes do que o menos eficaz. De fato, as diferenças na qualidade dos mestres podem ser ainda maiores, pois as crianças progridem, em particular no desenvolvimento da linguagem, simplesmente como resultado do amadurecimento. Na verdade, um estudo (Fitzpatrick, Grissmer & Hastedt, 2011) estimou que um terço do progresso realizado por crianças de sete anos de idade foi resultado da maturidade, então é possível que os professores mais eficientes sejam ao menos cinco vezes mais eficazes do que os menos eficientes. Além do mais, tanto nas escolas primárias (Hamre & Pianta, 2005) quanto nas secundárias (Slater, Davies & Burgess, 2008)² descobriu-se que os melhores professores beneficiam mais àqueles que têm mais deficiências, uma vez que o aumento da qualidade dos professores reduz a brecha de êxitos acadêmicos. Isto não significa que devemos nos concentrar somente na qualidade dos professores. É importante assegurar que os professores tenham os recursos necessários para efetuar seu trabalho. Necessitam de tempo, recursos materiais e de apoio dos gestores e dos colegas para fazer seu trabalho melhor. Mas a magnitude das diferenças entre os professores em seus impactos sobre a aprendizagem dos estudantes significa que melhorar a qualidade dos professores deve ser uma prioridade para qualquer sistema educativo.

Em vários países, isso tem tido como resultado esforços de melhora da qualidade dos professores por meio da substituição dos já existentes por outros melhores, mediante uma combinação entre a melhora da qualidade daqueles que ingressam à profissão (Barber & Mourshed, 2007) e a destituição dos professores ineficientes

² Nota do tradutor: no Brasil os respectivos níveis de ensino são identificados como Ensino Fundamental e Ensino Médio.

(Hanushek, 2010), o qual se discute nas próximas duas seções.

Se tomarmos um grupo de 50 professores, todos ensinando a mesma matéria:

- No grupo com o melhor professor, os estudantes aprendem em seis meses o que os estudantes com um professor mediano aprendem em um ano.
- No grupo com o professor menos eficientes, os estudantes vão levar dois anos aprender para aprender o mesmo (Hanushek & Rivkin, 2006).
- Nos grupos com os melhores professores, os estudantes com histórico de desvantagem aprendem o mesmo que os demais.

4. Prever quem será um bom professor é quase impossível

O fato de que a qualidade dos professores seja o ingrediente mais importante de um sistema educativo efetivo não indica, por si mesmo, o tipo de políticas que possam assegurar professores de alta qualidade.

Aumentar os níveis para ingressar à profissão docente parece ser uma opção política atrativa, especialmente porque as jurisdições de alto rendimento tendem a contratar professores da terça parte mais alta dos graduados universitários, e dados recentes da sondagem de habilidades em adultos da OCDE sugerem que a correlação entre as pontuações de conhecimentos matemáticos dos professores em um país e a pontuação na prova PISA é aproximadamente de 0.5 (OCDE, 2013b). Em alguns países de alto rendimento, como Finlândia e Singapura, há de dez a vinte candidatos qualificados para cada vaga nos programas de capacitação docente, e para além das qualificações acadêmicas de alto nível, os candidatos precisam ter boas habilidades de comunicação e qualidades pessoais necessárias para ser praticantes eficazes. Para os países que se encontram em posição privilegiada parece ser como se nada mais importasse; se tem-se a sorte de contar com as pessoas mais inteligentes de seu país aspirando a ser professores, então outras partes do quebra-cabeça podem estar mal e ainda ter um sistema educativo de alto rendimento. Não obstante, vale a pena observar que um processo de admissão à capacidade docente muito seletivo não garante um bom sistema educativo. Na República da Irlanda, a admissão à capacidade docente permanece do mesmo jeito há anos, extremamente seletiva (*INTERNATIONAL REVIEW PANEL ON THE STRUCTURE OF INITIAL TEACHER EDUCATION PROVISION IN IRELAND, 2010*), e mesmo assim o desempenho do país na prova PISA 2009 foi indistinguível do Reino Unido. Por outro lado, em Xangai, os professores não costumam ter qualificações educativas elevadas, todavia é proporcionada a eles uma capacitação de alta qualidade antes do início da carreira e durante a mesma. A seleção de professores entre os mais aptos academicamente pode, assim, não ser condição necessária nem suficiente para assegurar a alta qualidade dos professores.

Com efeito, parece que é extraordinariamente difícil identificar quem serão bons

professores até que estejam frente ao grupo (veja-se em Gladwell, 2008, um resumo do argumento e da evidência). Existe certa evidência de que os estudantes que são ensinados por professores com êxitos acadêmicos ou QI altos progridem mais (Slater Et Al, 2008; Hanushek, 1971), contudo a correlação é modesta, e outros estudos (por exemplo, Harris & Sass, 2007) efetivamente não encontram nenhuma relação entre o êxito dos estudantes e a educação prévia ao serviço ou os antecedentes acadêmicos do professor.

Há evidências de que as entrevistas bem estruturadas possuem alguma utilidade (veja-se, por exemplo, Dobbie, 2012), no entanto a correlação é, novamente, modesta, e por conseguinte existe o risco real de rejeitar aqueles que poderiam ser muito bons professores e aceitar aqueles que não serão. Mais importante mesmo é que melhorar a qualidade dos professores aumentando os padrões para ingressar à profissão requer demasiado tempo. Se os critérios para ingressar na profissão fossem levantados, demoraria ao menos 30 anos para que o último daqueles que entraram na profissão pare de ensinar, antes que os padrões fossem elevados.

Uma estratégia para aumentar a qualidade dos professores que é particularmente popular na atualidade é por intermédio de programas de elite, como *Teach for America* e *Teach First*, nos quais os graduados de alto rendimento empreendem o ensino em áreas com dificuldades socioeconômicas durante um tempo específico. As avaliações desses modelos não demonstraram ainda evidência clara de que sejam mais elevados do que os percursos tradicionais em direção ao ensino, apesar de que tendem a ser mais caros do que os programas de capacitação docente tradicionais. Tais esquemas podem aumentar o status do ensino como um trabalho que é digno dos mais capazes, mas a própria concepção desses programas, junto com o fato de que são explicitamente programas de “elite”, significam que a proporção de professores a adentrar nessas vagas por essas rotas é pouco provável que exceda o 1% da força docente, incluídos os prognósticos mais otimistas (gráfico VII.1).

5. Avaliar a qualidade dos professores é difícil por natureza

Se modificar a qualidade dos que ingressam à profissão é difícil, então uma alternativa óbvia poderia ser destituir os professores ineficientes, entretanto identificar os professores ineficientes é mais difícil do que parece, em primeiro lugar. Os protocolos de observação, como o *Framework for Teaching* (Marco para o Ensino) desenvolvido por Charlotte Danielson (1996), “funcionam” no sentido de que aprendem mais os estudantes que acessam o ensino por meio de professores que têm melhor classificação no esquema, contudo esses sistemas não conseguem identificar todos os aspectos do ensino eficaz.

Por exemplo, Sartain et al. (2001) encontraram uma clara relação positiva entre as qualificações dos professores no *Framework for Teaching* e o progresso realizado por seus estudantes. Os estudantes que receberam ensinamentos dos professores que

se classificaram como “distintos” (o nível mais alto no *Framework*) progrediram 30% mais do que os estudantes que aprenderam por meio de professores classificados como “insatisfatórios” (a classificação mais baixa). Esta é uma descoberta relevante. Vários intentos prévios fracassaram no momento de estabelecer um vínculo claro entre as condutas observáveis dos professores e o progresso de seus estudantes, por isso o fato de que agora possamos (ao menos para os professores dos Estados Unidos) capacitar o pessoal para classificar os professores de maneira que proporcionem nomenclaturas precisas da qualidade docente é um passo importante em frente. No entanto, como se observou acima, os melhores professores são 300% mais produtivos do que os menos eficientes (já que os melhores professores produzem 18 meses de progresso no mesmo lapso temporal que os menos eficientes, que produzem 6 meses de progresso). Isto indica que o *Framework for Teaching* captura somente algo em torno de 10% da qualidade dos mestres. Alguns estudos concluíram que, visto que a qualidade dos professores não é a mesma qualidade do ensino (dito de outra maneira, porque o desempenho do ensino é variável por natureza), é necessário um grande número de observações independentes para produzir estimativas acerca da qualidade dos professores que seja suficientemente confiável para respaldar decisões de alto impacto, como a rescisão do contrato de trabalho. Por exemplo, Hill et al. (2012) descobriram que somente para obter alguma confiabilidade das observações de até 0.90 (pode ser considerado um limite baixo para decisões de alto impacto) seria preciso observar um professor dando aulas para seis turmas distintas e cinco observadores independentes que dariam nota para cada uma.

Portanto, enquanto os esquemas de observação como *Framework for Teaching* identificam de forma confiável aspectos da qualidade do professor, quando os ditos esquemas são utilizados com a finalidade de avaliar o professor, e devido a considerar muito pouco da variância da qualidade do professor, existe o perigo de que os professores se tornem menos eficientes embora aumentem sua classificação no sistema.

Dada a pouca confiabilidade inerente às observações dos professores, diversos autores argumentaram que as evidências dessas observações devem ser complementadas com outras fontes de informação, como as avaliações dos estudantes em torno do ensino e as medidas do desempenho acadêmico dos estudantes (Kane & Staiger, 2012).

As medidas do desempenho acadêmico (chamadas com frequência “de valor agregado”) parecem identificar de maneira confiável distintos aspectos da qualidade dos professores a partir de observações (Rockoff & Speroni, 2011), contudo estimar o valor agregado de um professor é extraordinariamente difícil, inclusive quando levamos em conta os desempenhos prévios dos estudantes, já que a maioria das avaliações não representam completamente os resultados importantes da educação. Por exemplo, os bons professores continuam trazendo benefícios para os estudantes durante ao menos dois anos depois que param de dar-lhes aulas (Rothstein, 2010). Em outras palavras, os bons professores fazem com que os professores que ensinam

para seus estudantes em anos posteriores pareçam melhores do que realmente são. Um segundo problema com os modelos de valor agregado é que as diferenças nas suposições estatísticas realizadas no processo de modelagem produzem grandes variações nas classificações dos professores. Por exemplo, Goldhaber, Goldschmidt e Tseng (2013) descobriram que 9% dos professores que se classificaram entre os *top* 20% de valor agregado em um modelo (um modelo de efeitos aleatórios nos estudantes) foram classificados entre os 20% mais baixos dentro de um modelo tradicional de valor agregado. Considerando que também as classificações de valor agregado de um professor variam consideravelmente a cada ano (McCaffrey Et Al., 2008), as medidas de valor agregado do desenvolvimento dos estudantes são indicadores pouco confiáveis da qualidade dos professores.

A conclusão principal que deveria ser obtida de todos os intentos de mensurar a qualidade dos professores é que, visto que todas as medidas são pouco confiáveis, temos de emitir um juízo em torno do “peso do teste” requerido para identificar os professores inadequados. Se estabelecermos o peso demasiadamente alto, então deverão ser identificados poucos professores de baixa qualidade. Por exemplo, Winters e Cowen (2013) acharam em um estudo de professores de alfabetização na Flórida que, se o critério para a destituição de professores fosse estabelecido como pertencente aos 5% mais baixos de valor agregado durante dois anos consecutivos, somente 1 em cada 500 professores seria identificado para a demissão. Naturalmente, o peso do teste pode ser relaxado, o que teria como consequência mais professores destituídos, mas isso levaria à demissão de um maior número de professores altamente eficientes. Neste contexto, igualmente vale a pena as observações de Atteberry et al. (2013), que descobriram que os professores que foram identificados como altamente eficientes em seu primeiro ano não melhoraram (medida pelo valor agregado) durante os primeiros cinco anos de sua carreira docente, enquanto que aqueles que se classificaram como os menos eficientes em seu primeiro ano de ensino melhoraram constantemente.

O efeito cumulativo de todas as prescrições de política listadas anteriormente, até mesmo as implantadas de forma fiel e efetiva, seria demasiado pequeno, e demoraria muitos anos para se concretizar. Por esta razão, a melhora no desempenho dos professores em serviço terá de ser o componente principal da estratégia de cada país para a melhora da qualidade dos professores.

- O aumento do desempenho educativo dos estudantes é prioritário.
- Para o aumento do desempenho educativo dos estudantes se faz necessário melhorar a qualidade dos professores.
- Para melhorar a qualidade dos professores exige-se investir mais nos professores.

6. O desenvolvimento profissional é a chave para a qualidade dos professores

Isto poderia parecer um sonho inalcançável, no entanto a investigação sobre a experiência em áreas distintas indica que os professores que já se encontram em nossas escolas poderiam ser muito mais eficientes do que são na atualidade. Existe cada vez mais evidências de que medidas gerais de habilidades são bons indicadores do quão bem alguém realiza algo apenas nos estágios iniciais. Por exemplo, aqueles com QI mais alto são melhores jogadores de xadrez quando começam, mas depois de uns anos de prática a relação se torna muito menos perceptível. Um estudo estimou que somente cerca de 12% das variações nas classificações de jogadores de xadrez poderiam ser atribuídas ao QI (Grabner Et Al., 2007). De fato, as medidas das habilidades gerais compõem apenas 4% da variação no desempenho dos especialistas (Ericsson Et Al., 2006). David Berliner (1994) mostrou que a experiência no ensino parece ser muito similar à experiência em outros campos, ainda assim, conforme se observou anteriormente, a estratégia de obter “os melhores e os mais brilhantes” no ensino não somente não se mostra suficiente para construir uma força de trabalho de professores notáveis; nem é necessária.

O que produz um especialista é ao menos dez anos de prática deliberada – uma abordagem que se empenha em melhorar o desempenho (Ericsson, 2002). A maioria dos estudos sobre os efeitos da experiência na produtividade dos professores descobriu que estes melhoram nos primeiros dois ou três anos, embora depois desse período a maioria o faça lentamente, e muitos continuam a melhorar (Rivkin, Hanushek & Kain, 2005). Isto indica que muitos deles tocam a superfície dos tipos de melhorias que são possíveis.

Se quisermos ajudar os professores a se tornarem especialistas, como a pesquisa sugere que é possível, então temos de recrutar aqueles que tenham paixão por esse trabalho. A prática deliberada não é desfrutável por si mesma – no entanto é fundamental para melhorar o desempenho -, e apenas aqueles que possuem paixão por contribuir a que todos os seus estudantes tenham desempenhos de alto nível estarão dispostos a investir a energia necessária.

Dessa maneira, precisamos criar ambientes em que todos os professores concordem com a ideia da melhora contínua. Essa não é a ideia estereotipada de “acompanhar os novos desenvolvimentos” – é, pelo contrário, a aceitação de que o impacto da educação nas vidas dos jovens gera um imperativo moral, inclusive para os melhores professores, de continuar melhorando -. A evidência dos estudos sobre os esforços focalizados em melhorar o desempenho dos professores em serviço (William Et Al., 2004; Allen Et Al., 2011) é que os efeitos podem ser duas ou três vezes maiores do que o efeito combinado de todos os intentos de melhorar o ensino por meio da substituição dos professores descrita acima.

Uma vez que se aceita os benefícios da educação, tanto para os indivíduos como

para a sociedade, criam um imperativo moral de melhorar para todos os professores, tornando o passo seguinte decidir como isso será alcançado. Obviamente, podemos deixar que os professores melhorem por seus próprios meios, todavia dada a relevância do êxito educativo, parece importante assegurar que os professores devam ser apoiados na melhoria de sua prática, e a forma mais óbvia de fazê-lo é mediante o abastecimento de *feedback*.

7. O feedback é mais complexo do que normalmente se pensa

O termo “*feedback*” tem suas origens na engenharia da computação (Wiener, 1948) e foi definido como o “controle de uma máquina com base em seu desempenho real, ao invés de seu desempenho esperado” (Wiener, 1950/1989, p. 24). Wiener e seus colegas identificaram dois tipos de ciclos de *feedback*: positivos e negativos, contudo esses termos foram empregados em um sentido técnico que não se relaciona diretamente com a forma na qual utilizamos hoje em dia. Um exemplo de um ciclo de *feedback* positivo é quando um gravador registra o som de um amplificador, que torna a amplificar-se, e por sua vez o gravador novamente registra, criando um barulho familiar do *feedback* acústico. Outro exemplo é quando, em tempos de escassez, as pessoas guardam as poucas reservas que têm, o que faz com que se tornem ainda mais escassas, provocando um efeito cascata, no qual as pessoas acumulam ainda mais reservas, e assim sucessivamente. A questão é que, em engenharia, o *feedback* positivo não é proveitoso, pois leva a um crescimento explosivo ou ao colapso. Um exemplo de um ciclo de *feedback* negativo é o termostato de uma casa que monitora a temperatura, e se esta se desvia demais da desejada, o termostato ativa um sistema de aquecimento ou esfriamento que restabelece a temperatura desejada na casa. Em engenharia, desse modo, somente o *feedback* negativo é útil, porque retoma um sistema ao estado que se deseja.

Quando foi desenvolvida a ideia do *feedback* em psicologia nas décadas de 1960 e 1970, o campo estava dominado pelo behaviorismo e, portanto, não surpreende que tenha sido decidido que o uso mais apropriado do *feedback* tenha sido para reforçar o comportamento desejado:

A máquina, como qualquer professor particular, reforça ao estudante por cada resposta correta, utilizando o *feedback* imediato não somente para dar forma à conduta da maneira mais efetiva, como também para mantê-lo forte de uma maneira que poderia ser descrita em termos coloquiais como “manter o interesse do estudante” (Skinner, 1968, p. 39).

Em resposta a isto, Kulhavy (1977) sugeriu:

Com tais declarações tão confiantes disponíveis, não surpreende que os acadêmicos tenham trabalhado demais para colocar o pino redondo em um orifício quadrado do reforço. Infelizmente, esta fé estoica no *feedback* como reforço tem levado com frequência os pesquisadores a omitir ou descartar explicações alternativas para seus dados. Um não tem que olhar muito para longe para encontrar artigos que se dediquem a explicar por que

seus dados não atingiram as expectativas de trabalho ao invés de tentar dar sentido ao que encontraram (p. 213).

Durante os anos seguintes, apareceram várias revisões da pesquisa sobre os efeitos do *feedback*, as quais forneceram maior evidência de que o *feedback* que apenas proporcionava reforço não era particularmente efetivo (veja-se por exemplo Bangert-Drowns et al., 1991).

Várias das revisões da pesquisa sobre os efeitos do *feedback* utilizaram a meta-análise para sintetizar os resultados de estudos diferentes. A meta-análise é uma técnica que exprime a força de uma descoberta em forma de uma medida padrão, como o *tamanho do efeito padronizado* (COHEN, 1988), e pode ser muito útil para sintetizar resultados de estudos diferentes. Não obstante, existem determinadas precauções que devem ser tomadas na hora de ver os resultados da meta-análise, em particular no campo da pesquisa educacional.

a) O problema do arquivo. Existe uma tendência bem documentada a favor da publicação de estudos que têm resultados significativos, em lugar de não significativos. Isto é compreensível, evidentemente, mas o que não se tem consciência de modo amplo é que a maioria dos estudos de pesquisa em educação e psicologia possui uma representatividade estatística baixa, geralmente porque os experimentos são muito pequenos para gerar de modo consistente resultados estatísticos significativos, inclusive se os fenômenos pesquisados são reais. Uma pesquisa considerou que a representatividade estatística mediana do experimento psicológico típico foi de 0.4 aproximadamente, o que significa que um experimento tem apenas 40% de probabilidade de provocar um resultado estatisticamente significativo, inclusive se o efeito submetido a estudo foi verdadeiro. Por essa razão, apenas os estudos que por casualidade revelam um efeito maior do que o normal têm a possibilidade de ser publicados, dado que as dimensões do efeito desses estudos são uma sobrestimação do efeito real.

b) Variação na variabilidade. Enquanto os estudos psicológicos observam com frequência fenômenos relativamente estáveis, como a personalidade, os estudos educativos em geral estão mais interessados na mudança. Esse é um problema particular para a meta-análise, uma vez que o efeito de uma intervenção dependerá da variabilidade da população. A medida mais comum do tamanho do efeito, o tamanho do efeito padronizado mencionado anteriormente, é calculado através da divisão da média do grupo de tratamento e do grupo de controle em um experimento entre o desvio-padrão da população submetida ao estudo. Por isso, se a população submetida ao estudo é um subconjunto de uma população mais ampla (estudantes portadores de necessidades especiais, ou estudantes superdotados, por exemplo), então reduz-se o denominador da fração no cálculo do tamanho do efeito, o que aumenta a estimativa resultante do tamanho do efeito. Esse é um problema particular quando observamos estudantes de diferentes idades. Bloom et al. (2008) descobriram que o crescimento em um ano para uma criança de 6 anos foi equivalente a 1.5 desvios-padrão, todavia para um jovem de 15 anos, um ano de crescimento foi de somente 0.2 desvios-padrão. Seria de esperar, nessa altura, que uma pesquisa que compreendeu uma intervenção que incrementou a taxa de aprendizagem em 50% tivesse um tamanho de efeito de 0.75, caso tivesse sido realizado com crianças de 6 anos, mas tão somente um efeito de 0.2 se ocorrida com jovens de 15 anos de idade.

c) Seleção dos estudos. A seleção de estudos para sua inclusão em uma meta-análise

inclui um número de decisões de diversos graus de subjetividade. Ruiz Primo e Li (2013) revisaram mais de 9.000 artigos potencialmente relevantes para a efetividade do *feedback* na aprendizagem de matemática, ciências e tecnologia. Dos 238 artigos selecionados, 95 tiveram descobertas quantitativas específicas sobre seus efeitos na aprendizagem de matemática e ciências, mas destes 95 artigos, 76 representaram um acontecimento de *feedback* individual que durou poucos minutos. Embora tais descobertas possam ser de interesse para os investigadores, é muito pouco provável que seus resultados generalizem de maneira direta os efeitos contínuos do *feedback* durante semanas ou meses.

Conforme se observou anteriormente, isso não significa que a meta-análise seja inútil como técnica para agregar descobertas de investigação de múltiplas fontes, mas prova que são necessárias precauções consideráveis na hora de tirar conclusões sobre os tamanhos do efeito, particularmente quanto à magnitude dos efeitos das diferentes intervenções.

Kluger e DeNisi (1996) realizaram uma meta-análise especialmente importante de estudos sobre *feedback* nos quais revisaram todos os estudos realizados sobre os seus efeitos desde 1905 até 1995.

Começaram por definir as intervenções de *feedback* como “ações tomadas por (um) agente(s) externo(s) para proporcionar informação sobre algum(s) aspecto(s) do desempenho pessoal em uma tarefa” (p. 255). Eles identificaram mais de 3.000 estudos relevantes publicados entre 1905 e 1995, porém observaram que vários estudos eram muito pequenos (em muitos casos tinham somente um participante), e foram registrados de maneira pouco detalhada para permitir o cálculo do tamanho do efeito para a intervenção. A fim de garantir que não fossem incluídos estudos de baixa qualidade, Kluger e DeNisi convencionaram três critérios de inclusão em sua revisão:

1. Os participantes tinham que se dividir em dois grupos, sendo que a única diferença entre ambos, tanto quanto possível, é que tivessem ou não recebido o *feedback*.
2. O estudo cobriu ao menos dez participantes.
3. Foi incluída uma medida de desempenho suficientemente detalhada para o cálculo do tamanho do impacto do *feedback* no desempenho.

Surpreendentemente, somente 131 dos 3.000 estudos relevantes correspondiam aos critérios, sendo que os estudos selecionados reportaram 607 tamanhos dos efeitos e compreenderam 23.663 observações de 12.652 participantes. Em todos eles o tamanho do efeito mediano para o *feedback* foi de 0.41 desvios-padrão, sem embargo os efeitos variaram de maneira considerável entre os diferentes estudos. Vale a pena ressaltar que 38% dos 607 tamanhos do efeito foram negativos. Em outras palavras, em quase dois de cada cinco casos de *feedback* o desempenho médio diminuiu, na realidade. Com a finalidade de compreender como isso aconteceu, buscaram “moderadores” dos efeitos do *feedback* (variáveis que pudessem explicar as diferenças nos efeitos dos diferentes estudos) e descobriram que as intervenções

foram menos efetivas quando centraram sua atenção em si mesma, mais efetivas quando focaram na tarefa central, e tiveram a maior efetividade quando tiveram como foco os detalhes da tarefa central e quando abarcaram o estabelecimento de metas.

Não obstante, concluíram que se o *feedback* “funciona” ou não, e se esse fosse o caso, por vezes, não eram as perguntas corretas que precisavam ser feitas:

Antes de concluir, devemos refletir na implicação empregada em nosso estudo. A identificação de vários moderadores sugere que em certas situações o F (*Feedback*) pode provocar um efeito amplo e positivo no desempenho. Especificamente, um F para uma tarefa familiar, que contenha pistas que apoiem a aprendizagem, centrando a atenção em dar *feedback* sobre discrepâncias padrão ao nível da tarefa (velocidade do F e estabelecimento de metas), e esteja desprovido de pistas para o nível de meta tarefa (por exemplo, pistas que dirijam a atenção a si mesmo), tem a possibilidade de provocar ganhos impressionantes no desempenho, superando possivelmente F. Mesmo que um F nem sempre seja uma intervenção eficiente. Inclusive quando o F tem efeitos positivos consideráveis no desempenho, sua utilidade por meio de um aumento da motivação em uma tarefa, o efeito pode depender de um F contínuo. A eliminação de tal F pode dar como resultado um revés, como demonstraram alguns estudos de campo (Komaki & Cols., 1980). Portanto, o custo de manter uma intervenção contínua deve ser considerado na hora de avaliar tal intervenção. Mesmo assim, se o F afeta o desempenho mediante processos de aprendizagem de tarefas, o efeito pode propiciar a aprendizagem superficial somente se interferir com a aprendizagem mais elaborada. A falta de aprendizagem elaborada afeta a capacidade de utilizar o material aprendido em tarefas de transferência onde a tarefa é similar, mas não idêntica (por exemplo, Carroll e Kay, 1988). Ademais, a evidência de algum efeito de aprendizagem aqui foi, no melhor dos casos, mínima. Certamente, na bibliografia sobre paradigma de aprendizagem de probabilidade de pistas múltiplas (MCPL, por suas siglas em inglês), diversos revisores duvidam se os F têm algum valor na aprendizagem (Balzer & Cols., 1989; Brehmer, 1980) e sugerem alternativas aos F para aumentar a aprendizagem, como também proporcionar ao aluno mais informações sobre as tarefas (Balzer & Cols., 1989). Outra alternativa aos F é elaborar ambientes de trabalho ou aprendizagem que incentivem o teste ou o erro, maximizando assim a aprendizagem a partir do *feedback* da tarefa sem intervenção direta (Frese & Zapf, 1994). Essas considerações sobre a utilidade e as intervenções alternativas indicam que mesmo o F com efeitos positivos demonstrados sobre o desempenho não deveria ser administrado sempre que possível. Pelo contrário, é preciso um maior desenvolvimento dos F a fim de estabelecer as circunstâncias mediante os quais os efeitos positivos do F sobre o desempenho também sejam duradouros e eficientes, e quando esses efeitos são transitórios e têm utilidade questionável. Essa investigação deve focar nos processos induzidos pelo F e não na pergunta geral sobre se os F melhoram o desempenho – basta ver o quão pouco houve progressos em 90 anos de intentos de responder esta última pergunta (p. 278).

Quadro 1. Respostas às intervenções de feedback

(Kluger e DeNisi, 1996)

Tipo de resposta	O feedback indica que o desempenho ultrapassou a meta.	O feedback indica que o desempenho não alcançou a meta.
Mudar comportamento	Desprender menos esforço	<i>Aumentar o esforço</i>
Mudar a meta	<i>Aumentar a aspiração</i>	Reduzir a aspiração
Abandonar a meta	Decidir que a meta é demasiada sensível	Decidir que a meta é demasiado difícil
Rechaçar o <i>feedback</i>	Ignora-se o <i>feedback</i>	Ignora-se o <i>feedback</i>

Em outras palavras, qualquer intento de compreender os efeitos do *feedback* sem considerar a reação de quem o recebe está condenado ao fracasso, ainda que possa ser eficaz dar o mesmo *feedback*, mas pode não ser para outro indivíduo muito similar, pela maneira que os indivíduos participantes reagem a ele.

Para abordar essa questão, Kluger e DeNisi propuseram uma “teoria preliminar da intervenção de *feedback*” baseada na observação de que existem duas situações em que se pode fornecer o *feedback* (aquelas nas quais o desempenho real cai abaixo do desempenho desejado e aquelas nas quais o desempenho real supera o desempenho desejado) e existem quatro respostas que um indivíduo pode ter ante o *feedback* (mudança de comportamento, mudança na meta, abandono da meta ou rejeição ao *feedback*); o que conduz aos oito efeitos possíveis das intervenções de *feedback* mostrados no quadro 1.

Isto é, existem oito possíveis respostas a uma intervenção de *feedback*, e é possível que seis delas sejam ineficazes ou piores. Somente duas respostas, enfatizadas em itálico no quadro 1, têm a possibilidade de proporcionar resultados positivos. Crucialmente, os efeitos do *feedback* dependem do contexto em que ocorre. Portanto, com o intuito de compreender como o *feedback* pode melhorar o rendimento dos professores, faz-se necessária a sua incorporação em um esquema teórico mais amplo que inclua tanto o papel daqueles que proporcionam o *feedback*, como daqueles que o recebem.

Existem diversas formas pelas quais isso pode ser alcançado, mas a discussão de alguns deles vai além do alcance deste artigo. No que vem a seguir, se explora em detalhes um modelo para a melhora do ensino no contexto da avaliação formativa do desempenho docente.

Conforme salienta um relatório da OCDE sobre a avaliação dos professores (Santiago & Benavides, 2009), essa detém, comumente, dois objetivos principais: melhorar o desempenho dos professores e proporcionar evidências que permitam aos docentes e às instituições educativas (por exemplo, escolas, municípios, estados) prestar contas. O relatório também assinala que há uma tensão fundamental entre esses dois propósitos. Por exemplo, onde as classificações do desempenho no ensino se vinculam com as decisões acerca da permanência no emprego ou as recompensas financeiras, é pouco provável que os professores tentem enfoques inovadores, e organizando suas lições de forma a minimizar as possibilidades de que sejam reveladas as áreas fracas da sua prática. O menos óbvio é que, assim como se encontram em tensão em termos de como atuam, as diferentes funções da avaliação de desempenho docente necessitam ser validadas de distintas formas. Nas próximas duas seções será explorada em detalhes a natureza da avaliação formativa, enquanto na seção 10 analisa-se a validade das avaliações formativas de desempenho docente. O artigo termina com algumas recomendações sobre formas de implementar a avaliação formativa de desempenho docente.

8. A avaliação formativa do desempenho docente

Não existe uma definição consensual sobre o que constitui, exatamente, a avaliação formativa. A distinção entre a avaliação formativa e a somativa foi feita pela primeira vez por Michael Scriven, no contexto da avaliação do currículo. Por um lado, assinalou que a avaliação “pode ter um papel na melhora contínua do currículo” (Scriven, 1967, p. 41), enquanto por outro, a avaliação “pode servir para permitir aos administradores decidir se o currículo concluído em sua totalidade, refinado mediante o uso do processo de avaliação em seu papel primeiro, representa um avanço suficientemente significativo relativamente às alternativas disponíveis para justificar a despesa de adoção por parte de um sistema escolar” (p. 41-42). Então propôs “utilizar os termos de avaliação ‘formativa e somativa’ para qualificar a avaliação nesses papéis” (p. 43). A mesma distinção implementou depois Benjamin Bloom à avaliação de estudantes individualmente:

Contrasta bastante o uso da “avaliação formativa” para proporcionar *feedback* e ações corretivas em cada etapa do processo de ensino-aprendizagem. Entende-se por avaliação formativa a avaliação por meio de provas curtas utilizadas por professores e alunos como auxílio no processo de aprendizagem. Enquanto tal evidência possa ser qualificada e usada como parte das funções do teste e da classificação da avaliação, vemos um uso muito mais efetivo da avaliação formativa se estiver separada do processo de qualificação e sendo utilizada principalmente como suporte para o ensino (Bloom, 1969, p. 48).

Desde então, diversos autores têm proporcionado um número de definições para a expressão “avaliação formativa” (consulte em Wiliam, 2011, uma análise ampliada a respeito). As fontes principais de variação entre essas definições são:

- a) A duração do intervalo entre a coleta das evidências dos êxitos e seu uso.

- b) Se é essencial que os estudantes, de quem a evidência foi obtida, se beneficiem do processo.
- c) Se a avaliação tem de alterar as atividades de avaliação previstas.
- d) Se os estudantes têm que participar ativamente no processo.

Em uma tentativa de proporcionar uma definição ampla da avaliação formativa, Black e Wiliam (2009) propuseram uma abrangente que abarcou todas as questões identificadas anteriormente como variações de uma ideia central. Parafraseada em relação à avaliação formativa do desempenho docente, a definição dos autores é a seguinte:

A avaliação de desempenho docente funciona formativamente na medida em que a evidência do desempenho do professor obtida por meio da avaliação é interpretada por dirigentes, professores ou seus pares para a tomada de decisões acerca do desenvolvimento profissional docente que possivelmente sejam melhores, ou melhor fundamentadas, do que aquelas tomadas na ausência da dita evidência.

Diversas características desta definição merecem uma explicação. A primeira é que a definição de avaliação formativa repousa sobre o papel que desempenha a avaliação, ao invés da natureza da avaliação por si mesma. Posto que qualquer avaliação pode funcionar formativa ou cumulativamente, não existe “uma avaliação formativa”, mas sim uma avaliação em que há evidências de que é utilizada formativamente. Em segundo lugar, o enfoque da definição se encontra nas decisões, em detrimento das intenções por trás da coleta da dados. Tal noção se presta para assegurar que as situações em que as evidências são coletadas com a intenção de melhorar o desempenho docente, mas onde a evidência não é realmente usada, não sejam consideradas formativas. Dito de outra forma, o foco está na coleta de dados guiada por decisão, em vez de tomada de decisão baseada em dados. Em terceiro lugar, a definição não fala sobre quem (quer dizer, dirigentes, os próprios professores, ou seus pares) toma as decisões (o termo “líder” é utilizado aqui para qualquer responsável profissionalmente pelo desenvolvimento profissional dos professores, quer sejam chamados de líderes, mentores, instrutores etc.). Em quarto lugar, a definição não requer que o processo realmente melhore o desenvolvimento profissional do professor – dada a complexidade da aprendizagem humana, não pode haver tais garantias -. No entanto, a definição requer que as decisões resultantes tenham a possibilidade de melhorar a aprendizagem desse professor, e este é o quinto ponto, que a definição permita situações nas quais as evidências corroborem que as ações tomadas na ausência dessas foram, de fato, as corretas. Nessa situação, as decisões tomadas não são as melhores decisões (posto que são precisamente as mesmas), mas estão melhor fundamentadas, baseadas em evidências mais sólidas.

A partir da definição anterior, pode-se observar que a avaliação formativa se preocupa com a criação de “momentos de contingência” na aprendizagem dos professores e no aproveitamento dos mesmos, com o propósito de regulagem desses processos

de aprendizagem docente. Isso pode parecer uma abordagem muito estreita, no entanto ajuda a diferenciar uma teoria da avaliação formativa de uma teoria geral do ensino e da aprendizagem. Não obstante, embora esse enfoque seja acanhado, seu impacto é amplo, uma vez que a forma na qual os professores, alunos e seus pares criam e aproveitam esses momentos de contingência envolve considerações sobre o design, o currículo, a pedagogia, a psicologia e epistemologia do ensino. Na seção 9 é analisada mais detalhadamente a ideia central da avaliação formativa.

Avaliação para a aprendizagem é qualquer avaliação projetada e aplicada com a finalidade de promover a aprendizagem dos estudantes. Difere da avaliação idealizada principalmente com o escopo de responsabilização, *ranking* ou para certificar competências. Uma avaliação pode ser usada para aprender se fornece informações das quais os professores e os estudantes possam usar como *feedback* que valorize a si mesmo e para modificar as atividades de ensino e aprendizagem em que estão comprometidas. A referida avaliação se transfigura em *avaliação formativa* quando a evidência é realmente usada para adaptar o ensino para satisfazer as necessidades de aprendizagem (Black Et Al., 2004).

Quadro 2. Aspectos da Avaliação Formativa³

	Onde vai o professor	Onde está o professor no momento	Como chegar
Gestor	Clarificar, compreender e compartilhar as intenções de aprendizagem e os critérios de desempenho	Desenhar situações, atividades e tarefas efetivas que produzam evidências de desenvolvimento.	Proporcionar <i>feedback</i> que faça os professores avançarem.
Colega		Ativar os professores como autores de sua própria aprendizagem para eles e seus pares.	
Professor		Ativar os professores como autores de sua própria aprendizagem.	

³ Baseado em Wiliam e Thompson, 2008.

9. Estratégias de avaliação formativa do desempenho docente

Sendo que a avaliação formativa do desempenho docente lida essencialmente com a regulação do processo de aprendizagem dos professores, uma forma de pensar sobre isso é que ela lida com três processos principais:

- a) A meta para a aprendizagem docente.
- b) Seu nível de desempenho atual.
- c) Os passos necessários para alcançar a meta.

Se considerarmos esses três processos juntos aos papéis dos diferentes indivíduos participantes nesse processo – o professor, os seus pares e aqueles que são responsáveis por sua aprendizagem, denominados “dirigentes” para fins práticos desse artigo -, o cruzamento dessas duas dimensões leva a uma matriz de 3 x 3 células. O conteúdo de cada uma das nove células pode ser discutido individualmente, porém o modelo é simplificado consideravelmente se agruparmos algumas das células, conforme mostrado no quadro 2 que foi modificado por Wiliam e Thompson (2008). Cada uma das cinco estratégias é analisada a seguir.

Talvez o aspecto mais problemático da avaliação formativa do desempenho docente relacione os pontos estabelecidos na seção 5. Como temos ideia das características da prática docente efetiva, é difícil certificar-se de que a avaliação formativa dos professores seja direcionada corretamente. Em termos simples, se sabemos como é o bom ensino, como podemos melhorar os professores? O mais importante é que, como nossos esquemas de observação somente capturam uma pequena proporção da variação na qualidade dos professores, há um perigo real de se estabelecer metas para o desenvolvimento docente que tornem os docentes menos eficientes.

Em anos recentes, foi dada atenção considerável ao uso de rubricas para comunicar padrões aos aprendentes, tanto para os estudantes como para o desempenho docente. Todos os principais modelos de avaliação docente, como o *Framework for Teaching* de Denielson e o modelo de avaliação docente desenvolvido por Marzano e Toth (2013) apresentam níveis de desempenho dos professores em forma de rubricas que identificam distintos nivelamentos na área em questão. Sem sombra de dúvida pode ser útil apresentar os níveis de desempenho em forma de rubricas, mas é importante apontar que podem não ser efetivos na hora de comunicar os níveis de desempenho àqueles que ainda não são capazes de prová-los. As rubricas com frequência são tratadas como se fossem instruções sobre como melhorar o desempenho, entretanto tendem a ser descrições *post hoc* de qualidade. Mais importante é que, enquanto os especialistas geralmente são capazes de identificar o que estão fazendo com as descrições contidas nas rubricas, *o conteúdo destas não é utilizado por esses especialistas em seu desempenho*. Conforme escreveu Michael Polanyi:

As máximas são regras cuja aplicação correta é parte da arte que governam. As verdadeiras

máximas do golfe ou da poesia amplificam nossa compreensão acerca do golfe ou da poesia e podem até fornecer orientação valiosa aos golfistas e aos poetas; no entanto essas máximas se condenariam ao absurdo caso tentassem substituir a habilidade do golfista ou a arte do poeta. *As máximas não podem ser compreendidas, muito menos postas em prática por quem ainda não tem um bom conhecimento prático da arte em questão. Seu interesse é derivado de nossa apreciação da arte e não pode substituir ou estabelecer tal avaliação por conta própria* (POLANYI, 1958, p. 31-32; grifos do autor).

Portanto, as rubricas podem proporcionar um ponto de partida valioso para as conversas entre os professores e seus dirigentes, porém é pouco provável que o apego servil ao texto propriamente dito melhore o ensino. Pela relevância do contexto, há a possibilidade de que exemplos da prática real, de preferência junto com comentários que enfatizem características significativas, sejam muito mais eficazes no momento de comunicar aos professores sobre os aspectos atinentes ao desempenho de alta qualidade.

Configuração de situações, atividades e tarefas efetivas que obtenham evidências de desenvolvimento

A investigação sobre a capacidade de generalizar as classificações do desempenho dos professores tratadas na seção 5 sugere que é pouco provável que qualquer observação individual do desempenho docente produza evidência sólida das capacidades de um mestre. Isso é, obviamente, um problema significativo para a função de prestação de contas da avaliação, já que o desempenho observado em algum momento individual não é um indicador confiável do desempenho docente em outro momento. Apesar disso, para a função de melhora da avaliação a variabilidade do desempenho docente pode ser útil, uma vez que as observações sobre o ensino podem ser programadas para ocasiões específicas no momento em que a observação sobre a prática do ensino possa ser mais benéfica ao desenvolvimento docente. Geralmente, desta maneira, isso sugere que o professor que é observado deveria ser também o que escolhe a classe a ser observada. Um ponto importante que deve ser levado em conta é que todas as observações dependem da teoria. Inclusive em física, conforme observou Werner Heisenberg, “o que aprendemos não é próximo da natureza em si, mas da natureza exposta a nossos métodos de indagação” (Johnson, 1996, p. 147). Para o observador da prática do ensino pode ser útil ter a oportunidade de conhecer o professor antes da aula para compreender o que procura realizar, e ter um período significativo de tempo após a aula para tentar entender a própria compreensão do professor sobre o que aconteceu na aula. Como destaca David Ausubel há vinte anos: “Se tivesse de reduzir toda a psicologia educativa a um princípio, diria o seguinte: O fator individual mais importante que influi na aprendizagem é o que o aluno já sabe. Considere e ensine-o de acordo” (AUSUBEL, 1968, p. 6). Parece que isso pode ser tão correto para a aprendizagem dos professores como é para a aprendizagem dos alunos.

A ideia de que o professor observado deveria escolher a classe a ser observada

foi um recurso particularmente significativo do sistema de treinamento *My Teaching Partner* (Meu Parceiro de Ensino) (ALLEN ET AL., 2011), o qual centrou sua atenção em três aspectos do ensino: o apoio emocional aos estudantes (relações positivas, sensibilidade do professor e consideração das perspectivas adolescentes), a organização da sala de aula (gestão da conduta, maximização do tempo de aprendizagem e formatos de ensino efetivos) e apoio no ensino (compreensão de conteúdos, análise e resolução de problemas e qualidade do *feedback*). A cada duas semanas, os professores participantes gravavam uma aula e enviavam a gravação para um servidor seguro, onde o treinador podia revisar o vídeo e selecionar um pequeno número de segmentos curtos (de um ou dois minutos de duração) para uma discussão telefônica detalhada. Após dois anos, os estudantes que receberam ensino dos professores que participaram do sistema *My Coaching Partner* aprenderam 50% mais do que aqueles ensinados por professores não participantes do programa.

Proporcionar feedback que faça avançar os professores

Como ficou claro na extensa discussão sobre o *feedback* na seção 7, a qualidade da relação entre quem dá o *feedback* e quem o recebe é crucial para determinar se tem um efeito positivo. Os gestores precisam conhecer seus professores para que possam saber o momento de ser críticos e quando proporcionar-lhes apoio. É igualmente importante que os professores confiem em seus gestores, porque a menos que o docente acredite que o gestor leva em consideração seus interesses e credibilidade como instrutor, é pouco provável que o docente invista algum esforço para aperfeiçoar a sua prática. Isto quer dizer que não pode existir uma receita simples para o *feedback* efetivo para os professores sobre seu desempenho docente, devendo existir um par de princípios derivados de outras pesquisas que podem ser úteis neste caso. A primeira dita que o *feedback* deve causar o processo de pensamento. O *feedback* que provoca uma reação emocional, como geralmente acontece quando o *feedback* compara o desempenho de um professor em particular com o de outros docentes, é improvável que seja útil. Ao que parece, tem maior utilidade fazer a comparação do desempenho de um professor com seu próprio desempenho prévio (em outras palavras, para o docente, esse foi um dos seus “melhores níveis?”), o que provavelmente pode auxiliar a que o professor adote uma “mentalidade de crescimento” (Dweck, 2006). O segundo princípio é que o *feedback* deve ser traduzido como mais trabalho para quem o recebe do que para quem dá. A atividade de *feedback* por si só pode ser relativamente insignificante na hora de aprimorar o desempenho dos docentes; o que importa é a consequência da ação que toma o professor.

Ativar professores como recursos de aprendizagem para si e seus pares

Porque, como já foi mencionado anteriormente, a questão da confiança entre quem dá e quem recebe o *feedback* é crucial para seu possível êxito (veja também Santiago & Benavides, 2009), pode ser útil envolver colegas, em vez daqueles com uma função de liderança formal dentro da escola ou distrito, no momento de proporcionar *feedback* aos professores. Isso é particularmente importante quando os gestores têm um papel formal na função de prestação de contas, uma vez que pode ser-lhes difícil separar os dois papéis e, por exemplo, ignorar evidências que possam ser relevantes à função de responsabilização caso tenham que se concentrar na função de melhoria. Mesmo que os líderes sejam capazes de fazer isso, em última análise o comportamento do professor não dependerá de o gestor ser capaz de separar esses papéis claramente, mas se o docente acredita que está pronto para fazê-lo. Se o professor acredita que a evidência de fraqueza na práxis revelada em uma observação que aparentemente visa a melhorá-la pode afetar o julgamento emitido sobre a eficácia docente, então é mais provável que isso “seja seguro”, assim sendo, o potencial de observação é reduzido para melhorar a práxis. Quando os colegas se comprometem a observar a sala de aula, pode ser particularmente útil ter um protocolo claro de observação de classe que explicita que:

- a) O professor observado especifica o centro de atenção da observação.
- b) O professor observado especifica a evidência a ser colhida.
- c) O professor observado é dono de qualquer anotação que o observador venha a fazer durante a aula.

Ao salientar que o professor observado “é o dono” do processo, tal observação se distingue claramente das observações com fins de responsabilização e, desta feita, torna mais fácil desenvolver uma relação de confiança.

Ativar professores como proprietários de sua própria aprendizagem

Em última instância, a quantidade de tempo para que os dirigentes e os colegas observem a prática será limitada, então, se as melhorias vão acontecer, a maioria deve ser gerada por meio dos próprios esforços do professor para melhorar. Alguns argumentaram que isso se alcança de uma maneira melhor mediante sistemas de incentivos, em particular os incentivos financeiros para os professores, no entanto a evidência, tanto na profissão docente, como em um sentido mais amplo, sugere que a remuneração vinculada ao desempenho não é bem sucedida quando se trata de melhorar o mesmo (Pfeffer, 1998; Springer Et Al., 2010). Um caminho mais apropriado para o aprimoramento dos professores advém do compromisso com o imperativo moral identificado na seção 1 e de estarem conscientes de que, conforme foi mostrado na seção 2, os docentes podem fazer a diferença. Enquanto as estimativas acerca da magnitudes relativas das diferentes influências na aprendizagem dos alunos estão repletas de dificuldades, e podem variar consideravelmente de cultura em

cultura e de um país a outro, agora existe evidência substancial de que o impacto dos efeitos do professor pelo menos rivaliza com o impacto do contexto familiar e das condições socioeconômicas, podendo até mesmo superá-los (Rowe, 2003). Quando os professores executam melhor o seu trabalho, seus estudantes o experienciam mais, são mais saudáveis e contribuem mais com a sociedade. Atualmente, em vários sistemas, o apoio profissional parece algo de que somente os praticantes mais deficientes necessitam, assim sendo, a oferta desse suporte profissional pode ser percebida como indicador de baixo desempenho. Em contrapartida, quando todos os professores aceitam a ideia de que podem melhorar, não porque não sejam bons o suficiente, mas sim porque podem melhorar, isto cria uma rede que apoia todos os docentes a aceitar a necessidade de melhoria contínua. De acordo com as revisões empreendidas na seção 6, a maioria dos professores melhoram lentamente, e vários deixam de fazê-lo depois de dois ou três anos de trabalho, assim a pesquisa acerca da experiência indica que os aprimoramentos consideráveis são possíveis se todos os docentes, em lugar dos deficientes, estão comprometidos com um aperfeiçoamento profissional contínuo.

10. A validade da avaliação formativa do desempenho docente

Uma vez que os sistemas adotam a avaliação formativa de desempenho dos professores, surge uma preocupação imediata em torno da qualidade dessa avaliação. Tradicionalmente, na avaliação, as preocupações sobre a qualidade foram abordadas a partir do conceito de validade. Originalmente conceitualizada como uma das propriedades de um teste (ou seja, um teste é válido até certo ponto, ou não), foi aceito que a validade só faz sentido como propriedade de inferências com base nos resultados das avaliações. Por exemplo, um teste de matemática com alta demanda de leitura poderia apoiar conclusões sobre as habilidades matemáticas dos bons leitores, porém quanto aos leitores medíocres não poderíamos saber se seu baixo desempenho foi devido a uma capacidade matemática deficiente ou a uma leitura medíocre. O teste apoiaria algumas inferências (como em relação à capacidade matemática de bons leitores), mas outras não (em relação à dos leitores medíocres, por exemplo). Conforme sublinha Lee Cronbach: *“Uma não valida um teste, mas uma interpretação de dados que surgem de um procedimento específico”* (Cronbach, 1971, p. 447; grifos do autor).

Para a função de prestação de contas da avaliação, as inferências que desejamos destacar em geral se referem à qualidade observada do ensino, e um aspecto particularmente importante é que estão livres de subjetividade – em outras palavras, queremos ter certeza de que, apesar da pouca capacidade de generalização das qualificações dos professores observada previamente, a qualificação que é atribuída a um professor não depende da pessoa que a designou -. Dada a complexidade da práxis, é improvável que tal qualificação possa de fato se basear em um critério (isto é, exigindo somente a aplicação de critérios específicos), não obstante tais julgamentos podem se livrar da subjetividade mediante a capacidade dos observadores. Podemos

dizer que a avaliação é baseada em um princípio quando os diferentes avaliadores concordam com a qualidade do que observam (William, 1994), nos apoiando no princípio compartilhado de qualidade nas mentes daqueles que julgam. Em outras palavras, os significados dos resultados da avaliação devem ser compartilhados pelos diversos avaliadores. E por outro lado, a consistência dos significados entre os intérpretes é muito menos importante quando o objetivo principal é a melhoria. Se os diferentes avaliadores interpretam um desempenho docente particular de maneira diversa, e sugerem distintas atividades de desenvolvimento profissional que poderiam ser igualmente bem sucedidas quando se trata de fazer progressos na aprendizagem docente, então, de acordo com a definição da avaliação formativa adotada na seção 8, seriam da mesma maneira válidas em termos de impacto na aprendizagem do professor. Ou seja, adotando a distinção utilizada por Samuel Messick (1988), se as funções cumulativas da avaliação são validadas por seus *significados*, então as funções formativas da avaliação são validadas por suas *consequências* (William & Black, 1996).

Antes de deixar o assunto da validade da avaliação formativa, vale a pena destacar mais um comentário a respeito. Como foi observado, para que a avaliação funcione formativamente, os professores precisam ser membros da mesma comunidade de prática da qual seus dirigentes já são membros – necessitam compartilhar os conceitos implícitos de qualidade que os qualificadores compartilham quando concordam sobre a qualidade -. Conforme enunciado por Royce Sadler (1989), no contexto de estudantes e mestres:

As condições indispensáveis à melhora são que o estudante se apegue a um conceito de qualidade similar ao que tenha o professor, que seja capaz de monitorar de maneira contínua a qualidade do que é produzido durante o ato de produção em si, e que tenha um repertório de movimentos ou estratégias alternativos para onde ir a qualquer momento. Em outras palavras, os estudantes devem ser capazes de julgar a qualidade do que estão produzindo e de regular o que fazem enquanto estão fazendo (p. 121).

Embora possa ser um exagero afirmar que essas condições sejam indispensáveis, parece possível que sejam indicadores muito potentes de uma aprendizagem efetiva dos professores.

11. Avaliação formativa do desempenho docente

O compromisso com a avaliação formativa de desempenho dos professores não proporciona nenhum indício de como deveria realizar-se. Existem diferentes modelos que poderiam ser adotados, e cada um terá pontos fortes e fracos em relação às instituições nas quais serão usados. Porém, como resultado de um amplo trabalho realizado durante um período de mais de 15 anos com professores desenvolvendo a práxis na aula, parecem ser particularmente relevantes cinco princípios da aprendizagem docente (William, 2012). São eles: escolha, flexibilidade, pequenos passos, responsabilidade e suporte. A seguir, cada um deles é analisado.

Escolha

Quando os resultados das observações de sua práxis são relatados aos professores, é comum o aparecimento de informações tais como “pontos fortes” e “áreas por desenvolver”. O uso da expressão “áreas por desenvolver” supostamente deve fazer com que a crítica seja algo mais aceitável, contudo o efeito consiste em criar uma implicação infeliz de que as fraquezas são, necessariamente, áreas a serem desenvolvidas. É claro que, para alguns docentes, as fraquezas podem ser na realidade áreas por desenvolver, no entanto para outros podem ser somente pontos fracos. O ponto importante para a avaliação formativa do desempenho dos professores não é o perfil dos pontos fortes e fracos, mas quais áreas do desempenho docente, se desenvolvidas, teriam maiores repercussões na aprendizagem dos estudantes. No mundo dos negócios, ao longo de aproximadamente 30 anos, aumentou a conscientização de que as organizações podem se beneficiar mais se tiverem indivíduos que realmente se destacam em coisas nas quais já são bons, ao invés de se preocupar demais com as fraquezas (Belbin, 1981; Buckingham, 2007). Da mesma forma, o objetivo do desenvolvimento profissional para os professores não deveria ser tornar cada docente um clone de outros docentes, mas sim ajudar cada um a ser o melhor professor que possa ser. Para alguns deles, isso pode demandar centrar sua atenção em seus pontos fracos, contudo, para a maioria, manter o foco nos pontos fortes será mais exitoso.

Flexibilidade

Além de ter voz no que vão desenvolver, também é salutar que os professores tenham flexibilidade na hora de “transformar” ideias que encontram para fazê-las funcionar em suas próprias aulas. O problema reside no fato de que quando os professores adaptam ideias derivadas dos resultados de uma pesquisa para fazê-las funcionar em suas próprias aulas, com frequência as distorcem tanto que já não são mais eficazes em aprimorar o desempenho dos estudantes. Se pretendemos dar liberdade aos professores para adaptar as ideias que encontram para fazê-las funcionar em suas próprias aulas, também temos que proporcionar-lhes esquemas fortes que assegurem que as mudanças feitas não se tornem ineficientes. Por exemplo, está consolidado (veja-se, por exemplo, Slavin, 1995) que a aprendizagem colaborativa é uma forma bem sucedida de aumentar os êxitos dos estudantes, sempre e quando a maneira como se projeta o ensino requer metas de grupo (a fim de que os alunos como grupo ao invés de somente em um grupo) e responsabilização individual (para que o fracasso de um indivíduo tenha impacto sobre o grupo todo). Uma pesquisa de professores primários descobriu que 93% dos docentes disseram que usaram a aprendizagem colaborativa. Contudo, as entrevistas feitas posteriormente com 21 deles revelaram que só cinco implantavam-na de tal forma que os objetivos do grupo e a responsabilidade individual foram criados (Antil, Jenkins, Wayne & Vadasy, 1998).

Pequenos passos

Diante o imperativo moral de melhorar o ensino descrito nas seções 1 e 6, não surpreende que aqueles que elaboram as políticas – os políticos e os administradores – queiram que os professores desenvolvam sua prática na aula o mais rápido possível. Não obstante, a pesquisa sobre o impacto do desenvolvimento profissional sugere que os benefícios foram decepcionantes – o que Michael Fullan disse há mais de 20 anos parece ter a mesma vigência de então: “Nada prometeu tanto nem tem sido tão frustrantemente ineficiente quanto os milhares de *workshops* e conferências que não levaram a nenhuma mudança significativa na prática quando os professores voltaram às suas salas de aula” (Fullan & Stiegelbauer, 1991, p. 315).

Há quem atribua a lentidão da mudança nos professores à sua resistência – que os docentes se agarram a um conjunto de hábitos profissionais que representam uma parte nuclear da identidade profissional de cada professor, razão pela qual não estão dispostos a mudar. Tais crenças proporcionam a justificativa para os esquemas de incentivo – a ideia é que os professores adotarão novas ideias se forem pagos para fazê-lo. Entretanto, conforme visto na seção 6, não há indícios de que os incentivos financeiros tenham sido efetivos no momento de melhorar o desempenho docente. Uma explicação plausível acerca da lentidão na mudança dos professores advém da pesquisa sobre a experiência tratada na seção 6. A experiência é o resultado da práxis deliberada ampla e seu desenvolvimento não pode ser facilitado dizendo-lhes “o que fazer”. O desenvolvimento profissional implica na aquisição de novos aspectos da experiência, o que leva tempo.

Responsabilidade

Conforme exposto na seção 2, os resultados educativos dependem de certo número de fatores, vários dos quais estão além do controle das escolas e dos professores. Tornar escolas e professores responsáveis por coisas sobre as quais eles não podem influenciar parece contrário aos princípios da justiça natural. Contudo, o que cada docente controla é se eles melhoram ou não. Assim sendo, a função de melhoria na avaliação é mais importante do que a de responsabilização. Podemos investir muito tempo e energia tentando medir a qualidade dos professores, mas mesmo se pudéssemos fazê-lo bem e se fosse possível destituir os menos eficientes, os benefícios ainda assim seriam modestos. Um enfoque de responsabilização para melhoria terá um impacto muito mais significativo (posto que todos os professores estão melhorando), e no meu trabalho com os docentes durante os últimos 15 anos, achei útil envolvê-los em uma abordagem de planejamento bem estruturada que enfatize quatro processos; embora, é claro, existam inúmeros protocolos que poderiam ser adotados para o planejamento de ações, mas nossa experiência no trabalho com os professores no desenvolvimento de sua prática na avaliação formativa indica que as seguintes características são particularmente importantes:

1. *O planejamento deve identificar um pequeno número de mudanças que o professor realizará em sua docência.* Quando os professores tratam de transformar mais do que duas ou três coisas em sua prática simultaneamente, o resultado costuma ser que suas rotinas de classe se deterioram de maneira significativa, e então tendem a recorrer àquelas com as quais se sentem cómodos e “seguros”.
2. *O plano deve ser escrito.* Escrever um plano faz com que seja mais provável que o professor pense nele enquanto o escreve, tornando as ideias mais concretas e igualmente crie um registro que signifique que os professores têm menos probabilidade de esquecer o que planejaram fazer.
3. *O plano deve se concentrar nas áreas da prática que possam beneficiar os estudantes.* Nem todas as mudanças feitas pelos professores em sua prática beneficiarão os estudantes. Dado o imperativo moral de melhorar a educação, os docentes devem se concentrar nos aspectos que tenham mais possibilidades de aprimorar os resultados dos alunos. Devido a que as evidências das pesquisas raramente proporcionem sinais de que possam ser aplicadas em todos os contextos, os professores precisam utilizar de seu julgamento profissional na hora de decidir onde concentrar seus esforços, no entanto devem ser capazes de proporcionar alguma evidência de que a parte que escolhem para trabalhar se trata, ao menos, de um caso *prima facie*, ou seja, que à primeira vista surte efeitos nos resultados de seus estudantes.
4. *O plano deve identificar o que professor planeja reduzir, ou deixar de lado, para dar tempo às mudanças.* A maioria dos professores trabalham tão duro quanto podem, por isso, se essas mudanças são interpretadas como fardos extras à carga já existente, é pouco provável que alguma vez sejam postas em prática. Para dar tempo a essas mudanças, o plano de ação deve identificar algo que os docentes estão fazendo, que cessem ou reduzam parte disso a fim de ter tempo disponível às transformações. Quando as pessoas são solicitadas a estabelecer essas prioridades claras, geralmente esperam que as mudanças necessárias possam ser feitas por serem mais eficientes no uso do tempo, entretanto comumente trata-se de um otimismo sem esperança. A única maneira de ter tempo disponível para coisas novas é reduzir, ou deixar de fazer completamente as coisas que estão sendo feitas, com a finalidade de abrir espaço à inovação.

Suporte

O último elemento do processo – suporte – está intimamente relacionado à responsabilidade. Certamente, alguns autores têm descrito essa atividade como uma característica individual aos ambientes de aprendizagem efetivos para os professores: apoiar a responsabilidade (Ciofalo & Leahy, 2006). A ideia central se baseia na criação de estruturas que, ao mesmo tempo em que responsabilizam os professores pelo desenvolvimento de sua prática, também fornecem o suporte para alcançá-la. Portanto, pode ser concebido como apoio e responsabilidade, dois lados da mesma moeda. Conforme foi observado anteriormente, o papel do professor é comprometer-se a melhorar a sua prática, e concentrar-se nas mudanças que possam beneficiar seus estudantes. Os dirigentes podem criar ambientes de aprendizagem efetivos para seus professores ao engendrar expectativas de uma prática que melhore continuamente, mantendo no centro das atenções coisas que façam a diferença para os alunos, proporcionando-lhes tempo, espaço, isenção e apoio à inovação e,

finalmente, apoiando os professores na hora de correr riscos

Conclusão

Esse artigo revisou uma pesquisa acerca de várias estratégias para a melhoria dos resultados educativos dos estudantes, e concluiu que o investimento em servir os professores deve ser a estratégia central de qualquer tentativa séria de aprimorar a educação. Para obter o efeito máximo, o investimento em professores deve ser construído sobre as evidências em torno de que tipo de mudanças na prática docente tem o maior impacto no desempenho dos estudantes, mas igualmente tenha em mente como aprendem e se desenvolvem os docentes. A evidência aqui apresentada indica que as melhorias significativas nos resultados educativos para os jovens, com os benefícios concomitantes para os indivíduos e à sociedade são possíveis se nos concentrarmos no poder da avaliação para melhorar, ao invés de medir o desempenho dos professores.

Referências

- Allen, J. P., R. C. Pianta, A. Gregory, A. Y. Mikami y J. Lun (2011), "An Interaction-based Approach to Enhancing Secondary School Instruction and Student Achievement", *Science*, 333 (6045): 1034-1037.
- Amabile, T. M., y S. J. Kramer (2011), *The Progress Principle: Using Small Wins to Ignite Joy, Engagement and Creativity at Work*, Cambridge, MA, Harvard Business Review Press.
- Antil, L. R., J. R. Jenkins, S. K. Wayne y P. F. Vadasy (1998), "Cooperative Learning: Prevalence, Conceptualization and the Relation between Research and Practice", *American Educational Research Journal*, 35 (3): 419-454. 196
- Ausubel, D. P. (1968), *Educational Psychology: A Cognitive View*, Nueva York.
- Holt, Rinehart & Winston. Atteberry, A. et al. (2013), "Do First Impressions Matter? Improvement in Early Career Teacher Effectiveness", Panel III: New Research on Training, Growing and Evaluating Teachers, Center for Analysis of Longitudinal Data in Education Research (CALDER), Conference, Washington, D. C.
- Bangert-Drowns, R. L., C.-L. C. Kulik, J. A. Kulik y M. Margan (1991), "The Instructional Effect of Feedback in Test-Like Events", *Review of Educational Research*, 61 (2): 213-238.
- Barber, M. y M. Mourshed (2007), *How the World's Best-Performing School Systems Come Out on Top*, Londres, McKinsey & Co.

- Belbin, R. M. (1981), *Management Teams: Why They Succeed or Fail*, Oxford, Heinemann.
- Berliner, D. C. (1994), "Expertise: The Wonder of Exemplary Performances", in J. N. Mangieri y C. C. Block (comps.), *Creating Powerful Thinking in Teachers and Students: Diverse Perspectives*, Fort Worth, TX, Harcourt Brace College, pp. 161-186.
- Black, P.J. y D. Wiliam (2009), "Developing the Theory of Formative Assessment", *Educational Assessment, Evaluation and Accountability*, 21 (1): 5-31.
- Bloom, B. S. (1969), "Some Theoretical Issues Relating to Educational Evaluation", en R. W. Tyler (ed.), *Educational Evaluation: New Roles, New Means: The 68th Yearbook of the National Society for the Study of Education (Part II)*, 68 (2): 26-50, Chicago, IL, University of Chicago Press.
- Bloom, H. S., C. J. Hill, A. R. Black y M. W. Lipsey (2008), "Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions", *Journal of Research on Educational Effectiveness*, 1 (4): 289-328.
- Bryk, A. S., P. B. Sebring, E. Allensworth y S. Luppescu (2010), *Organizing Schools for Improvement: Lessons from Chicago*, Chicago, IL, University of Chicago Press.
- Buckingham, M. (2007), *Now Go Put Your Strengths to Work*, Nueva York, Simon & Schuster.
- Ciofalo, J. y S. Leahy (2006), *Personal Action Plans: Helping to Adapt and Modify Techniques*", documento apresentado no Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Cronbach, L. J. (1971), "Test Validation", in R. L. Thomdike (ed.), *Educational Measurement*, 2a ed., Washington, D. C., American Council on Education, pp. 443-507.
- Danielson, C. (1996), *Enhancing Professional Practice: A Framework for Teaching*, Alexandria, ASCo.
- Dobbie, W. (2012), *Teacher Characteristics and Student Achievement: Evidence from Teach for America*, Cambridge, MA, Harvard University Faculty of Arts and Sciences.
- Dweck, C. S. (2006), *Mindset: The New Psychology of Success*, Nueva York, Random House.
- Ericsson, K. A. (2002), "Attaining Excellence through Deliberate Practice: Insights

- from the Study of Expert Performance", in M. Ferrari (ed.), *The Pursuit of Excellence through Education*, Mahwah, NJ, Lawrence Erlbaum Associates, pp. 21-55.
- Ericsson, K. A., N. Charness, P. J. Feltovich y R. R. Hoffman (2006), *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge, Cambridge University Press.
- Feinstein, L., D. Budge, J. Vorhaus y K. Duckworth (2008), *The Social and Personal Benefits of Learning: A Summary of Key Research Findings*, Londres, Institute of Education, University of London.
- Fitzpatrick, M. D., D. Grissmer y S. Hastedt (2011), "What a Difference a Day Makes: Estimating Daily Learning Gains During Kindergarten And First Grade Using A Natural Experiment", *Economics of Education Review*, 30 (2): 269-279.
- Fullan, M. y S. Stiegelbauer (1991), *The New Meaning of Educational Change*, Londres, Cassell.
- Gladwell, M. (2008), "Most Likely to Succeed", *New Yorker*, 15 de dezembro, pp. 36-42.
- Goldhaber, D. D., P. Goldschmidt y F. Tseng (2013), "Teacher Value-added at the High School Level: Different Models, Different Answers?", *Educational Evaluation and Polietj Analysis*, 35 (2): 220-236.
- Goos, M. y A. Manning (2007), "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain", *Review of Economics and Statistics*, 89 (1): 118-133.
- Grabner, R. H., E. Stern y A. C. Neubauer (2007), "Individual Differences in Chess Expertise: A Psychometric Investigation", *Acta Psychologica*, 124 (3): 398-420.
- Hamre, B. K. y R. C. Pianta (2005), "Academic and Social Advantages for At-Risk Students Placed in High Quality First Grade Classrooms", *Child Development*, 76 (5): 949-967.
- Hanushek, E. A. (1971), "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data", *American Economic Review*, 61 (2): 280-288.
- Hanushek, E. A (2004), "Some Simple Analytics of School Quality", NBER Working Paper W10229, Washington, D. C., National Bureau of Economic Research.
- Hanushek, E. A (2010), "Teacher Deselection", en D. Goldhaber y J. Hannaway (comps.), *Creating a New Teaching Profession*, Washington, D. C., Urban Institute Press, pp. 165-180.
- Hanushek, E. A (2011), "The Economic Value of Higher Teacher Quality", *Economics*

- of Education Review, 30 (3): 466-479.
- Hanushek, E. A., y L. Woessmann (2010), The High Cost of Low Educational Performance: The Long-Run Impact of Improving PISA Outcomes, París, OCDE.
- Hanushek, Eric A. y Steven G. Rivkin (2006), "Teacher Quality", en Handbook of the Economics of Education, Elsevier.
- Harris, D. N. y T. R. Sass (2007), Teacher Training, Teacher Quality and Student Achievement, vol. 3, Washington, D. C., National Center for Analysis of Longitudinal Data in Education Research.
- International Review Panel on the Structure of Initial Teacher Education Provision in Ireland (2012), Report, Dublín, Department of Education and Skills.
- Johnson, G. (1996), Fire in the Mind: Science, Faith and the Search for Order, Londres, Viking.
- Kane, T. J. y D. O. Staiger (2012), Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains, Redmond, WA, Bill and Melinda Gates Foundation.
- Kluger, A. N. y A. DeNisi (1996), "The Effects of Feedback Interventions on Performance: A Historical Review, A Meta-Analysis, and a Preliminary Feedback Intervention Theory", Psychological Bulletin, 119 (2): 254-284.
- Kulhavy, R. W. (1977), "Feedback in Written Instruction", Review of Educational Research, 47 (2): 211-232.
- Marzano, R. J. y M. D. Toth (2012), Teacher Evaluation that Makes a Difference: A New Model for Teacher Growth and Student Achievement, Alexandria, VA, ASCD.
- McCaffrey, D., T. R. Sass y J. R. Lockwood (2008), "The Intertemporal Stability of Teacher Effect Estimates", Nashville, TN, NCPI Working Paper No. 2008-22.
- McGaw, B. (2008), "The Role of the OECD in International Comparative Studies Of Achievement", Assessment in Education: Principles Policy and Practice, 15 (3): 223-243.
- Messick, S. (1988), "The Once and Future Issues of Validity: Assessing the Meaning and Consequences of Measurement", en H. Wainer y H. I. Braun (comps.), Test Validity, Hillsdale, NJ, Lawrence Erlbaum Associates, pp. 33-45. OCDE
- Messick, S. (2013a), Education at a Glance, Paris, OCDE.
- Messick, S. (2013b), "What Teachers Know and How that Compares with College

- Graduates Around the World", *Education Today*. Disponível em <http://oecdeducationtoday.blogspot.com/2013/11/what-teachers-know-and-how-that.html>
- Pfeffer, J. (1998), "Seven Practices of Successful Organizations", *California Management Review*, 40 (2): 96-124.
- Rivkin, S. G., E. A. Hanushek y J. F. Kain (2005), "Teachers, Schools and Academic Achievement", *Econometrica*, 73 (2): 417-458.
- Rockoff, J. E. y C. Speroni (2011), "Subjective and Objective Evaluations of Teacher Effectiveness: Evidence from New York City", *Labour Economics*, 18 (5): 687-696.
- Rowe, K. J. (2003), "The Importance of Teacher Quality as a Key Determinant of Students' Experiences and Outcomes of Schooling", documento apresentado na ACER Research Conference, Melbourne, Australia. Acesso em novembro de 2013 de https://www.det.nsw.edu.au/proflearn/docs/pdf/qt_rowe2003.pdf
- Ruiz-Primo, M. A. y M. Li (2013), "Examining Formative Feedback in the Classroom-Context: New Research Perspectives", in J. H. McMillan (comp.), *Sage Handbook of Research on Classroom Assessment*, 2a ed., Thousand Oaks, CA, Sage, pp. 215-232.
- Sadler, D. R. (1989), "Formative Assessment and the Design of Instructional Systems", *Instructional Science*, 18: 119-144.
- Santiago, P. y F. Benavides (2009), *Teacher Evaluation: A Conceptual Framework and Examples of Country Practices*, París, OCDE.
- Sartain, L., S. R. Stoelinga, E. R. Brown, S. Luppescu, K. K. M. Matsko, F. K. Miller, C. E. Durwood, J. Y. Jiang y D. Glazer (2011), *Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation*, Chicago, IL, Consortium on Chicago School Research.
- Scriven, M. (1967), "The Methodology of Evaluation", en R. W. Tyler, R. M. Gagné y M. Scriven (comps.), *Perspectives of Curriculum Evaluation*, vol. 1, Chicago, IL, Rand McNally, pp. 39-83.
- Skinner, B. F. (1968), *The Technology of Teaching*, Nueva York, Appleton-Century-Crofts.
- Slater, H., N. Davies y S. Burgess (2008), *Do Teachers Matter? Measuring the Variation in Teacher Effectiveness in England*, vol. 09/212), Bristol, Inglaterra, University of Bristol, Institute of Public Affairs.
- Slavin, R. E. (1995), *Cooperative Learning: Theory, Research and Practice*, 2a ed., Boston, MA, Allyn & Bacon.

Springer, M. G., D. Ballou, L. Hamilton, V.-N. Le, J. R. Lockwood, D. McCaffrey, M. Pepper y B. M. Stecher (2010), *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching*, Nashville, TE, National Center on Performance Incentives at Vanderbilt University.

Wiener, N. (1948), *Cybernetics, or Control and Communication in the Animal and the Machine*, Nueva York, John Wiley & Sons.

Wiener, N. (1950/1989), *The Human Use of Human Beings: Cybernetics and Society*, Londres, Free Association Books.

Wiliam, D. (1994), "Assessing Authentic Tasks: Alternatives to Mark-Schemes", *Nordic Studies in Mathematics Education*, 2 (1): 48-68.

Wiliam, D. (2010), "Standardized Testing and School Accountability", *Educational Psychologist*, 45 (2): 107-122.

Wiliam, D. (2011a), "How Do We Prepare Students for a World We Cannot Imagine?", documento presentado en el Salzburg Global Seminar, Salzburgo, Austria. Acceso em junho de 2012 de www.dylanwiliam.net

Wiliam, D. (2011b), "What is Assessment for Learning?", *Studies in Educational Evaluation*, 37 (1): 2-14. ---(2012), *Sustaining Formative Assessment with Teacher Learning Communities*, Seattle, WA, Kindle Direct Publishing.

Wiliam, D. y P.J. Black (1996), "Meanings and Consequences: A Basis for Distinguishing Formative and Summative Functions of Assessment?", *British Educational Research Journal*, 22 (5): 537-548.

Wiliam, D. y M. Thompson (2008), "Integrating Assessment with Instruction: What Will It Take to Make It Work?", in C. A. Dwyer (ed.), *The Future of Assessment: Shaping Teaching and Learning*, Mahwah, NJ, Lawrence Erlbaum Associates, pp. 53-82.

Wiliam, D., C. Lee, C. Harrison y P. J. Black (2004), "Teachers Developing Assessment for Learning: impact on Student Achievement", *Assessment in Education: Principles Policy and Practice*, 11 (1): 49-65.

Winters, M. A. y J. M. Cowen (2013), "Who Would Stay, Who Would Be Dismissed? An Empirical Consideration of Value-Added Teacher Retention Policies", *Educational Researcher*, 42 (6): 330-337

Biografia

Dylan Wiliam

Professor Emérito de Avaliação Educacional na University College de Londres. Após a graduação em matemática e física, por um ano teve experiência como professor em uma escola particular, tendo após lecionado em escolas carentes de Londres por sete anos, período durante o qual obteve mais diplomas em matemática e em educação matemática.

Em 1984, juntou-se ao Chelsea College, Universidade de Londres, que mais tarde se fundiu, tornando-se parte do King's College London. Durante esse período, trabalhou no desenvolvimento de esquemas inovadores de avaliação matemática antes de assumir a liderança do programa de educação de professores de matemática no King's. Entre os anos de 1989 e 1991, foi Coordenador Acadêmico do Consórcio para Avaliação e Teste nas Escolas, no qual desenvolveu uma variedade de avaliações estatutárias e não estatutárias para o currículo nacional na Inglaterra e no País de Gales. Após o seu retorno ao King's, terminou o seu doutorado, onde abordou algumas das questões técnicas levantadas pela adoção de um sistema de níveis de realização referenciados pelo critério independente de idade no currículo nacional da Inglaterra e do País de Gales.

De 1996 a 2001, foi decano e Chefe da Escola de Educação no King's College de Londres, e de 2001 a 2003, atuou como diretor assistente do College. Em 2003, mudou-se para os Estados Unidos, para atuar como diretor sênior de pesquisa no Educational Testing Service, em Princeton, N.J. Em 2006, retornou ao Reino Unido como vice-diretor do Instituto de Educação da Universidade de Londres, tendo, em 2010, deixado o cargo para poder dedicar mais tempo à pesquisa e à docência.

Nos últimos 15 anos, seu trabalho acadêmico tem se concentrado no uso da avaliação como apoio à aprendizagem (também chamada de avaliação formativa). Foi co-autor, juntamente com Paul Black, de uma considerável revisão acerca das evidências de pesquisa em avaliação formativa publicadas, e desde então, tem se dedicado ao desenvolvimento de práticas de avaliação formativa com grupos de professores do mundo inteiro.