

EFEITO RETROATIVO DA AVALIAÇÃO NO ENSINO/APRENDIZAGEM DE LÍNGUAS: O ESTADO DA ARTE¹

MATILDE V. R. SCARAMUCCI
(UNICAMP)

ABSTRACT:

This article presents a state of the art review of the literature on the concept of washback in the context of language teaching. The concept, despite been frequently used in the international educational literature and in Applied Linguistics, is not well understood, as it does not seem to have been part of the research agenda for many years. After terminological and conceptual considerations, the article discusses recent empirical contributions, controversial aspects as well as those deserving more careful treatment in future research.

Keywords: washback, language testing/assessment, language teaching and learning.

INTRODUÇÃO

Observações sobre o impacto ou influência que exames ou testes² e avaliação em geral exerce potencialmente nos processos educacionais, seus participantes e produtos do ensino e aprendizagem — o que tem sido identificado na literatura como efeito retroativo (*backwash* ou *washback*) — têm sido numerosas no cenário internacional na área de Educação e, mais especificamente, de Linguística Aplicada (LA doravante) há muitos anos. Uma análise mais cuidadosa dessas observações, entretanto, revela um conceito que, embora popular e aparentemente poderoso, é pouco compreendido, usado por alguns para referir-se às relações entre avaliação e aprendizagem (Shohamy, Donitsa-Schmidt e Fernam, 1996) e, para outros, à influência ou impacto da avaliação no ensino/aprendizagem (Alderson & Wall, 1993; Hughes, 1989; Gates 1995; Messick 1996).

Essas controvérsias, dentre outras, no entanto, são previsíveis, dada a natureza complexa do conceito e o fato de não ter feito parte da agenda de pesquisa durante muitos anos. A publicação do artigo intitulado *Does washback exist?* por Alderson & Wall, em 1993, na revista *Applied Linguistics*³ expõe, de forma contundente, a fragilidade dos conhecimentos acerca do fenômeno e a necessidade urgente de comprovações empíricas:

¹ Este artigo foi elaborado no âmbito do projeto CNPq 303822/2002-3 (bolsa de produtividade em pesquisa).

² Os termos “exame” e “teste” são usados como sinônimos.

³ Apresentado pela primeira vez no *Language Testing Research Colloquium*, em 1992, e publicado nesse mesmo ano como um dos trabalhos da série *Working Papers* da Universidade de Lancaster, Inglaterra.

Is washback a concept to be taken seriously, or simply a metaphor which is useful in that it encourages us to explore the role of test in learning and the relationship between teaching and testing? We are not sure at present, but we suspect that if it is a metaphor, it needs to be articulated somewhat more precisely if it is to throw light on teaching and testing, or indeed on the nature of innovation and change. And if it is a concept to be taken seriously, then we need to examine critically, and see what evidence there might be that could help us in this examination (Alderson & Wall, 1992, p. 9).

A consideração dos impactos sociais, éticos e políticos da avaliação faz parte de uma tendência recente dessa pesquisa em contextos de ensino/aprendizagem de línguas, que reconhece o poder exercido pela avaliação nos participantes e sociedade em geral. O interesse por esses aspectos tem sido tão grande nos últimos anos a ponto de Davies (2003) considerá-los uma das três heresias da área de avaliação da linguagem de tradição anglo-americana. Para o autor, “os grandes desenvolvimentos nesse campo tendem a ser recebidos de forma tão entusiástica a ponto de levar a visões desequilibradas (ou heresias) em relação ao construto “visão de linguagem” (Davies, *op.cit.*, p.355).

No Brasil, não têm sido poucos os que têm responsabilizado os exames vestibulares, dada sua longa tradição e importância, pelas mazelas de nosso ensino médio, o que, na área educacional, convencionou-se denominar de “tirania do vestibular”. São comuns, portanto, na imprensa, afirmações tais como “O ensino médio teve um desvio muito grande por causa disso [vestibular]. Ficamos anos ensinando traquejos para se fazer uma prova” (Folha de S. Paulo, 30/8/98); ou ainda:

O mais grave efeito colateral dessa situação é que os principais vestibulares do país acabam se tornando o ponto de referência para os “curricula” do segundo grau. Em São Paulo, por exemplo, as escolas do ciclo médio ensinam aquilo que está no programa da Fuvest, que faz a seleção da USP. Essa situação é bastante indesejável. Afinal, a prova da Fuvest não foi concebida para converter-se em diretrizes curriculares, mas apenas para definir quais são os melhores alunos (Folha de S.Paulo, 18/07/2004).

Embora essas citações não façam referência explícita ao conceito de efeito retroativo, deixam transparecer a percepção de sua ação. A segunda afirmação sugere, entretanto, desconhecimento do potencial que grandes exames têm como instrumentos redirecionadores do ensino.

Apesar de freqüentes no Brasil, essas menções, muito menos do que no cenário internacional, têm sido acompanhadas de evidências empíricas. Constata-se nesse caso, muito freqüentemente, que exames são inseridos e descartados sem que uma análise e avaliação criteriosa de seus efeitos sejam realizadas⁴, o que é de se lamentar, considerando-se os recursos de natureza diversa mobilizados para seu desenvolvimento e implementação.

Entender melhor o conceito de efeito retroativo pressupõe compreender os mecanismos operantes na relação entre ensino/aprendizagem e avaliação que, de acordo com Bachman and Palmer (1996), é muito mais complexa do que apenas a influência de um teste no ensino e na aprendizagem. O impacto teria, nesse caso, de ser avaliado com referência às

⁴ Um exemplo é a recente substituição do Exame Nacional de Cursos (Provão) pelo ENADE (Exame Nacional de Desempenho dos Estudantes) como parte do Sistema Nacional de Avaliação da Educação Superior (SINAES).

variáveis, objetivos e valores da sociedade, do sistema educacional em que é usado, assim como dos resultados potenciais de seu uso.

Neste artigo, organizado em seis seções, apresento uma revisão da literatura acerca do conceito de efeito retroativo, colocando em uma perspectiva atual as discussões a seu respeito, trazendo definições e conceituações distintas, expondo aspectos que ainda merecem ser investigados. Na Seção I, faço considerações terminológicas e conceituais. Na Seção II, apresento as contribuições produzidas até a publicação do artigo de Alderson e Wall acima mencionado e nas Seções III e IV, os desenvolvimentos que o sucederam. Na seção V, discuto resultados dos estudos conduzidos e já publicados no Brasil assim como outros em andamento, encerrando o artigo com algumas considerações finais.

1. CONSIDERAÇÕES TERMINOLÓGICAS E CONCEITUAIS

Em alguns dicionários não técnicos, os termos em inglês *backwash* ou *washback effects* ou apenas *backwash* ou *washback* têm sido considerados sinônimos (Funk and Wagnalls, p. 1420), embora, como salienta Alderson & Wall (1993), *backwash* seja preferido na literatura da área de educação e *washback* na área de educação lingüística ou LA (pelo menos naquela de tradição inglesa). Ainda nesse dicionário (p. 105), o termo é definido como “any condition resulting from some previous act, remark”. Por outro lado, é interessante ressaltar que apenas *backwash* consta do dicionário Collins Cobuild (p. 93), definido como “the backwash of an event or situation usually unpleasant, that exists after it and as a result of it”. As conotações negativas de conceito também são ressaltadas por autores como Spolsky (1994), que se refere, principalmente, a algo “não esperado e geralmente indesejável, a um resultado subsidiário ou reação”, o que não seria difícil de se explicar dada a tradição de exames de grande escala, geralmente de múltipla escolha, considerados “maus exames”.

Em português, o termo usado para denominar tecnicamente o conceito tem sido *efeito retroativo* (vide, por exemplo, Almeida Filho & Schmitz, 1998). No Novo Aurélio, p. 1231, a palavra é definida como “retro + movimento para trás. De retro mais ativo; relativo ao passado; que modifica o que está feito; que afeta o passado, que retroage”. É interessante notar que, na designação em português, a ação no passado (no caso, da avaliação, que geralmente ocorre no final do processo, no ensino) parece ser mais enfatizada, além de não se observar a conotação negativa nem de efeitos colaterais do termo em inglês.

Para Alderson e Wall (1993), o conceito refere-se aos efeitos de testes no ensino e na aprendizagem e sugerem que talvez estivesse restrito ao que ocorre dentro da sala de aula por influência de exames e avaliações. Impacto, por outro lado, refere-se aos efeitos que testes têm nas pessoas, políticas e práticas, dentro da sala de aula, da escola, do sistema educacional ou sociedade. Esses autores fazem ainda uma outra distinção com “influência”, que associam às atitudes, percepções, opiniões dos alunos e professores a respeito de determinado exame. Essa visão, entretanto, não é consensual, uma vez que outros autores (Cheng 2004; Scaramucci 1999c, 2000/2001 e 2002, por exemplo) consideram efeito retroativo como sinônimo de “impacto”. Para outros ainda (Andrews 2004), seria a forma

técnica de tratar os efeitos dos exames no ensino, na aprendizagem e no sistema educacional em geral, enquanto “impacto” seria o termo não técnico, sinônimo de “efeito”.

Na literatura em avaliação, outros termos têm sido associados ao conceito de efeito retroativo, revelando as inúmeras controvérsias relacionadas a sua conceituação e abrangência: impacto do teste (*test impact* - Bachman & Palmer 1996); instrução guiada por medidas (*measurement driven instruction* - Popham 1987); alinhamento curricular (*curriculum alignment* - Shepard 1993); retorno do teste (*test feedback*); ensinar para o teste (*teach to the test*); validade sistêmica (*systemic validity* - Frederiksen and Collins 1989); validade consequencial (*consequential validity* - Messick 1989); validade retroativa (*washback validity* - Morrow 1991) e ainda alavancas para mudança (*levers for change* - Pearson 1988). Algumas dessas distinções mais importantes serão retomadas mais adiante.

Várias também são as dimensões do conceito presentes na literatura (vide especialmente Watanabe 1997 e 2004): a) *especificidade* (geral e específico) - efeito retroativo geral é aquele produzido por um exame sem levar em conta suas características ou conteúdo; um exame que leva as pessoas a estudarem mais, independentemente de ser na direção dos conteúdos do exame, tem, portanto, efeito retroativo geral. Efeito retroativo específico, por outro lado, é definido como aquele que leva em conta o conteúdo do teste ou um aspecto específico desse conteúdo; b) *intensidade* (forte ou fraco) - o efeito retroativo pode ser forte ou fraco, dependendo da intensidade com que se manifesta na sala de aula. É forte quando determina praticamente todas as atividades de sala de aula ou todos os participantes; é fraco quando afeta parte da aula e alguns professores e alunos, mostrando a força de outros elementos; c) *extensão* (longo ou curto) - o efeito retroativo pode durar um período longo ou curto de tempo. Por exemplo, no caso de um exame vestibular, se os efeitos persistirem até os alunos entrarem em uma instituição, por exemplo, o efeito é curto; se persistirem após a entrada dos alunos, pode ser considerado longo. Essa dimensão pode ainda ser observada de uma outra perspectiva: um efeito que se observa logo após a introdução de um exame, mas que diminui com o tempo, é curto; e longo, por outro lado, quando esse efeito permanece por um longo tempo depois de o exame ter sido implementado; d) *intencionalidade* (intencional e não intencional): as conseqüências sociais intencionais e também as não intencionais de um exame são vistas como efeito retroativo (Messick, 1989; McNamara, 1996); e) *valor* (positivo e negativo): tem sido uma das dimensões mais discutidas, relacionada com intencionalidade — efeito retroativo positivo tem sido associado com efeito intencional (aquele desejado quando o exame é implementado) e efeito retroativo negativo (e positivo) com não intencional. Essa discussão é complexa por envolver julgamento de valor: um exame pode ser visto como positivo por professores e negativo por diretores, por exemplo. Segundo Watanabe (2004), esse julgamento deve ser feito sempre em relação aos candidatos ao teste.

2. PRINCIPAIS CONTRIBUIÇÕES DE ALDERSON E WALL (1993)

O artigo de Alderson e Wall (*op.cit*) foi elaborado tendo como pano de fundo um projeto de avaliação do impacto de um exame nacional de inglês no Sri Lanka. Além de um mapeamento das evidências empíricas disponíveis até então, expondo suas limitações

teóricas e metodológicas, o artigo é importante por mostrar qual era o estágio de entendimento do conceito até então.

Segundo os autores, o estudo considerado “clássico” ou mais conhecido foi conduzido na área de educação por Kellaghan, Madaus e Airaisan, publicado em 1982, concebido nos anos 60 e executado entre os anos de 1974 e 1977. Foi um projeto longitudinal irlandês-americano que se propunha a examinar o impacto da introdução de testes padronizados nas escolas irlandesas, que não tinham essa tradição. Os resultados complexos de serem interpretados mostram um efeito limitado desses testes nas práticas avaliativas ou organização escolar, servindo apenas para confirmar as avaliações que os professores fizeram dos alunos e, como indicam Alderson & Wall (*op.cit.*), sem incluir evidências suficientes para confirmar a crença de que o fato de submeter regularmente os alunos a exames possa melhorar seu desempenho. Os testes foram vistos como instrumentos justos e precisos, constituindo-se estímulos à aprendizagem, sem causar aumento da ansiedade, normalmente vista como um efeito negativo. As reações dos professores e alunos a esses instrumentos também foram consideradas positivas. Uma crítica feita a esse estudo, entretanto, é que se tratava de uma situação artificial: os testes foram introduzidos apenas como parte do experimento e, dessa forma, não tinham significado e conseqüências para os alunos dentro do sistema educacional irlandês — os alunos não passaram ou deixavam de passar por causa do exame, que não lhes negava acesso a outro nível educacional — daí não causarem efeitos negativos.

Uma outra limitação do estudo de Kellaghan et al. identificada pelos autores é que a informação coletada independentemente do que aconteceu nas classes de controle e experimental era muito restrita, resumindo-se ao que era relatado pelos professores e alunos (percepções), sendo mínimas as evidências do que realmente ocorreu em sala de aula (ações). Estudos como Paris, Lawton, Turner e Roth (1991), Haladyna, Nolen e Hass (1991) e Frederiksen (1984) também são apontados por terem essa limitação. A única exceção parece ser o estudo de Smith (1991), em que observações de sala de aula mostram que os programas de avaliação reduzem o tempo que deveria ser devotado à instrução efetiva e conseqüentemente, estreitavam o currículo e os modos de instrução.

Alderson e Wall (*op.cit.*) identificam a falta de evidências de sala de aula como uma limitação metodológica dessas pesquisas, salientando a necessidade de estudos de natureza etnográfica, que triangulem os métodos de coleta e análise dos dados e identifiquem as variáveis explicativas para o fenômeno. Dados de entrevistas e questionários permitiriam, portanto, a expressão das percepções dos alunos e professores, enquanto observações em sala de aula a análise na perspectiva do pesquisador. Em artigos mais recentes, a necessidade de uma triangulação de instrumentos tem sido constantemente reafirmada (vide, por exemplo, Bailey, 1999).

Na área específica de ensino de línguas, os poucos estudos publicados incluem Forbes (1973), que faz uma crítica à introdução de exames de múltipla escolha na Etiópia, e a resposta de Madsen (1976) a essa crítica. Nem um num outro, entretanto, fornece evidências para apoiar suas afirmações, que podem, portanto, ser consideradas impressionistas. Os únicos estudos que investigam sistematicamente a questão foram Wesdorp (1982), que examinou a introdução de testes de múltipla escolha nos Países Baixos (sem, todavia, fornecer evidência de impacto em relação à mudança de hábitos de estudo em decorrência

da introdução desse método de avaliar); Pearson (1988), Alderson e Wall (1990 e 1991), Wall & Alderson (1993), no Sri Lanka; Khaniya, (1990), no Nepal; e Hughes (1988), na Turquia.

Ao ser implementado em um contexto de ensino tradicional de base estruturalista, o exame no Sri Lanka tinha como meta causar mudanças no currículo e na metodologia de ensino, tornando-os mais alinhados às tendências comunicativas. Para surpresa dos autores, entretanto, seus resultados foram muito menos significativos do que os esperados: o exame teve um impacto considerável no conteúdo das aulas de inglês e na maneira como os professores elaboravam seus testes (alguns positivos e outros negativos), mas pouco ou nenhum impacto na metodologia que usavam em sala de aula ou na maneira como avaliavam o desempenho de seus alunos. Possíveis explicações para esses resultados são a falta de informação dos professores sobre o exame e entendimento de como preparar os alunos (vide também Wall 1996).

Hughes (*op.cit*), por sua vez, mostra as mudanças causadas no currículo, nos livros-texto e no ensino de inglês, com conseqüentes melhorias nos níveis de proficiência dos alunos envolvidos, pela introdução de um teste de proficiência no currículo da Universidade de Bogazici, em Istambul, Turquia. O estudo, entretanto, como outros, deixa de apresentar dados do que efetivamente ocorria nas salas de aula, levando o autor a reconhecer suas limitações e a necessidade de pesquisas futuras.

Além da caracterização do estado da arte no que se refere a efeito retroativo, o estudo de Alderson e Wall (*op. cit.*) também oferece reflexões decorrentes da análise dessa literatura assim como propostas de encaminhamentos para pesquisas futuras. Embora os autores reconheçam, assim como outros, uma associação entre efeito retroativo e resultados negativos, também salientam a possibilidade de esse efeito se manifestar de forma positiva ou benéfica, levando os professores a uma preparação mais cuidadosa de suas aulas e a um trabalho mais eficiente com os conteúdos; e, da mesma forma, incentivando os alunos a estudarem mais e se preocuparem com suas lições. Maus testes, assim, poderiam ter maus efeitos e bons testes, bons efeitos. Ou ainda maus testes e bons testes poderiam ter efeitos positivos e negativos: a qualidade do efeito parece ser independente da qualidade do teste, pois outras forças existentes na sociedade parecem interagir com as características do exame na determinação de seu efeito. Essas forças podem ser as crenças dos professores ou características de sua formação, requisitos das instituições em que atuam, ou ainda assuntos políticos, sociais e econômicos que prevaleciam no período em que os dados foram coletados, e que poderiam impedir que o exame funcionasse como uma “alavanca para a mudança” (Pearson, 1988). A noção determinista de efeito retroativo é, portanto, questionada por esses autores, que a identificam como a limitação teórica mais importante dessa fase da pesquisa.

Os pesquisadores concluem o artigo chamando a atenção para a importância de se determinar as condições em que o conceito opera — o que o afeta, como alunos e professores preparam-se para os testes, a natureza da relação entre avaliação e ensino. Somente depois de entender e descrever sua natureza é que estaríamos em condição de explorar porque ele ocorre, ou quais são as causas desses efeitos. Um conjunto de hipóteses é sugerido como proposta de encaminhamento para futuras pesquisas:

1. Um teste tem influência no ensino.
2. Um teste tem influência na aprendizagem.
3. Um teste tem influência na forma como os professores ensinam.
4. Um teste tem influência no que os professores ensinam.
5. Um teste tem influência no que os alunos aprendem.
6. Um teste tem influência na forma como os alunos aprendem.
7. Um teste tem influência na taxa e seqüência da aprendizagem.
8. Um teste tem influência na taxa e na seqüência do ensino.
9. Um teste tem influência no grau e profundidade do ensino.
10. Um teste tem influência no grau e profundidade da aprendizagem.
11. Um teste tem influência nas atitudes, conteúdo, método, etc., de ensino e de aprendizagem.
12. Testes que têm conseqüências importantes terão efeito retroativo.
13. Testes que não têm conseqüências importantes não terão efeito retroativo.
14. Testes terão efeito retroativo em todos os alunos e professores.
15. Testes terão efeito retroativo em alguns alunos e em alguns professores, mas não em outros. (Alderson & Wall, 1992, pg 8-9)

Burrows (2004) denomina essa fase da pesquisa que precede o artigo de Alderson e Wall (*op.cit.*) de “tradicional”, fundamentada em um modelo de estímulo-resposta ou behaviorista, caracterizado pelo determinismo e inevitabilidade relacionados à qualidade do teste e não à variabilidade dos envolvidos, conforme Figura 1 a seguir:



Figura 1. Teoria tradicional de efeito retroativo: modelo de estímulo-resposta (Burrows, 2004:126)

3. OUTRAS CONTRIBUIÇÕES

O impacto do artigo de Alderson e Wall (*op.cit.*) foi grande, intensificando a busca por evidências empíricas e discussões teóricas mais aprofundadas acerca do tema, a ponto de considerá-lo, em artigo anterior (Scaramucci, 1999c), um verdadeiro divisor de águas dessa literatura. Inaugura, assim, uma segunda fase na pesquisa sobre o conceito de efeito retroativo, fase essa denominada por Burrows (2004) de modelo “caixa preta” (*black box*), reconhecendo que os professores⁵ podem reagir de forma distinta quando preparando alunos para exames — fazendo com que “uma única resposta em relação a efeito retroativo não seja inevitável”. A partir de suas reflexões, o autor propõe, ainda, um terceiro modelo, que denomina de “inovação curricular”, conforme mostram as Figuras 2 e 3 abaixo:

⁵ Seu modelo restringe-se aos professores, porque seu estudo, assim como muitos outros, considera os efeitos de testes no ensino. Entretanto, poderíamos expandir o modelo para considerar os efeitos em outros participantes (alunos, por exemplo).

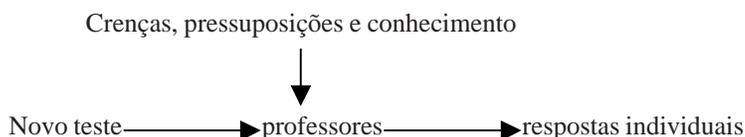


Figura 2. Visão de efeito retroativo dos anos 90: modelo da “caixa preta” (Burrows, 2004:126)

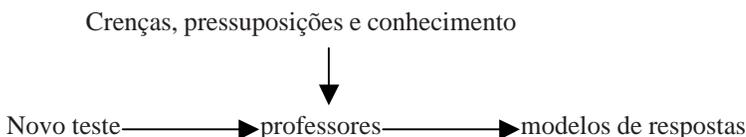


Figura 3. Visão proposta de efeito retroativo: modelo de inovação curricular (Burrows, 2004:126)

Diferentemente de outras áreas, em que as publicações são esparsas, os estudos sobre efeito retroativo têm sido publicados de forma agrupada: seis artigos, sendo duas reflexões teóricas e quatro relatos de pesquisa compuseram o volume 13 da revista *Language Testing*, organizado por Alderson e Wall, comprovando as várias dimensões do conceito através de estudos com focos distintos.

O artigo de Messick (1996) faz considerações de natureza teórica, apresentando o efeito retroativo como uma das várias conseqüências que testes podem ter, decorrentes de fatores a ele inerentes e ao seu contexto. Para esse autor, a necessidade de validação de um teste deve incluir a consideração das conseqüências de seu uso – daí o termo *validade de conseqüência* (*consequential validity*) por ele proposto. O conceito é, assim, equiparado ao de *validade sistêmica*, usado na área educacional mais geral (Frederiksen e Collins, 1989): um teste é sistemicamente válido se conseguir induzir, no sistema curricular e instrucional, mudanças que levem ao desenvolvimento das habilidades/conhecimentos/competências que o teste pretende avaliar. Da mesma forma, portanto, um teste teria *validade de conseqüência* ou ainda *validade retroativa* (um outro conceito semelhante proposto por Morrow, 1986) se influenciar as pessoas a fazerem coisas que elas não fariam necessariamente sem o exame.

Torna-se desnecessário dizer, entretanto, que essa visão não é consensual. Embora seja desejável que elaboradores de testes se preocupem com os possíveis efeitos de seu exame no ensino e na aprendizagem, de forma a encontrar maneiras de potencializar os efeitos positivos, há limites com relação ao que podem fazer devido exatamente às outras forças operantes nesse contexto. Aceitar que todas as conseqüências de um teste possam ser controladas pelos seus elaboradores pressupõe uma visão determinista e ingênua de seus efeitos (Alderson e Wall, *op. cit.*).

Para Messick (*op. cit.*) ainda, efeito retroativo deve ser evidência de impacto no ensino e na aprendizagem (práticas boas ou ruins) apenas se essa evidência puder ser relacionada à introdução e uso do teste, o que nem sempre é possível, uma vez que em muitos contextos em que testes são introduzidos, não se tem dados sobre as condições do

ensino (*baseline data*) antes da introdução do exame (Messick, 1994). Essa causalidade, portanto, teria de ser demonstrada. Na falta desses dados, uma abordagem sugerida por Qi (2004) — usada em Scaramucci (1999c e 2002), assim como em Souza (2002), Correia (2003) e Bartholomeu (2002) — é comparar as intenções (em termos de conteúdos e teoria subjacente) dos elaboradores do teste, explicitadas em materiais de divulgação, com o teste para determinar *se* e *como* este operacionaliza essas intenções. As práticas dos professores são, então, observadas e analisadas, para se determinar *o quê* e *como* esses professores ensinam. As intenções e as práticas são comparadas para se identificar pontos de contato e de conflito.

O autor relaciona, ainda, efeito retroativo positivo às avaliações autênticas e diretas e, mais basicamente, à necessidade de se minimizar sub-representações e dificuldades irrelevantes ao construto (Messick, 1996). Para o autor, portanto, as chances de efeito retroativo positivo são maiores em testes de desempenho, que trazem amostras representativas de situações reais de comunicação ou tarefas autênticas, ou aquelas que reproduzem uma situação de vida real para examinar a capacidade do aprendiz em lidar com ela (Carroll, 1980)⁶. Quanto mais direto for o exame e quanto menos depender de “macetes”, mais chances haverá de efeito retroativo positivo ocorrer (vide Scaramucci 1995, 1999a; Scaramucci & Rodrigues, no prelo, dentre outros, para discussões acerca do exame Celpe-Bras)⁷. Portanto, a autenticidade, também discutida por autores como Morrow (1991) e Wesche (1983), é retomada em Bailey (1996 e 1999) como fundamental para a potencialização do efeito retroativo de exames.

A autora, partindo de uma revisão de estudos recentes, traz contribuições importantes para um maior entendimento do fenômeno e de seus mecanismos de funcionamento, maneiras de potencializá-lo e de investigá-lo. Com base nas hipóteses de Alderson e Wall, idéias de Morrow (1991) e de um artigo de Hughes (1993), não publicado, a autora propõe um modelo (vide Figura 4) em que participantes, processos e produtos influenciam e ao mesmo tempo são influenciados por testes. Os participantes são os professores, os alunos, diretores, elaboradores de materiais didáticos, editores e demais pessoas cujas atitudes e percepções podem ser alteradas em função de um teste ou avaliação. O processo são as ações tomadas pelos participantes que podem contribuir para o processo de aprendizagem, o que inclui o desenvolvimento de materiais, elaboração do currículo, mudanças na metodologia, uso de estratégias de aprendizagem e de como se sair bem em testes, dentre outros. Produto, por sua vez, refere-se ao que é aprendido (fatos, habilidades) e à qualidade da aprendizagem (fluência, por exemplo). A partir desses elementos, a autora constrói um modelo: a natureza do teste pode afetar, primeiramente, as percepções e atitudes dos

⁶ Os autores referem-se à autenticidade da tarefa e do texto usado.

⁷ O *Certificado de Proficiência em Língua Portuguesa para Estrangeiros*, em uso desde 1998, é desenvolvido e outorgado pelo Ministério da Educação, aplicado em 17 centros credenciados no Brasil e 25 no exterior duas vezes por ano. No Brasil, é obrigatório para os alunos de graduação e de pós-graduação dos convênios do MEC que desejam ingressar em universidades brasileiras, assim como para médicos estrangeiros que queiram revalidar seus diplomas para atuarem no Brasil. A competência do candidato é avaliada diretamente através de seu desempenho em tarefas que se assemelham a situações de uso na vida real. Maiores informações poderão ser obtidas no Manual do Exame: <http://mec.gov.br/sesu>.

participantes em relação às tarefas de ensino e de aprendizagem. Essas percepções e atitudes afetam o que os participantes fazem (processo) que, por sua vez, afeta os resultados, ou o produto. Salienta Bailey (*op. cit.*), entretanto, que não necessariamente um processo leva diretamente à aprendizagem, mas a outros produtos que podem contribuir e promover a aprendizagem – novos materiais, melhoria no ensino, etc. As linhas pontilhadas no modelo indicam possíveis influências dos participantes no teste — o que van Lier (1989) denomina efeito proativo (*washforward*).

É importante salientar que a ação de um exame nas percepções e atitudes das pessoas depende, em grande escala, de suas crenças. Nem sempre essas percepções se baseiam em informações oficiais (materiais de divulgação ou de preparação) sobre o exame antes de o candidato submeter-se a ele. Importantes também são as informações não oficiais, e de certa forma até mesmo as consideradas “folclóricas”, muitas vezes obtidas a partir de outras pessoas que já fizeram o exame. Ainda uma terceira fonte são as informações de quem fez o exame, a partir de suas percepções acerca do que é fácil, difícil, justo ou injusto, tipos de itens não esperados, instruções não familiares, etc, assim como o retorno dos professores sobre as notas dos alunos e daqueles que ministraram o exame localmente, dentre outros. Observa-se, dessa maneira, que a informação pode ser oficialmente dada, inferida ou ainda imaginada, tornando fundamental, portanto, para a determinação de efeito retroativo, o conceito de “validade aparente” (*face validity*) do exame.

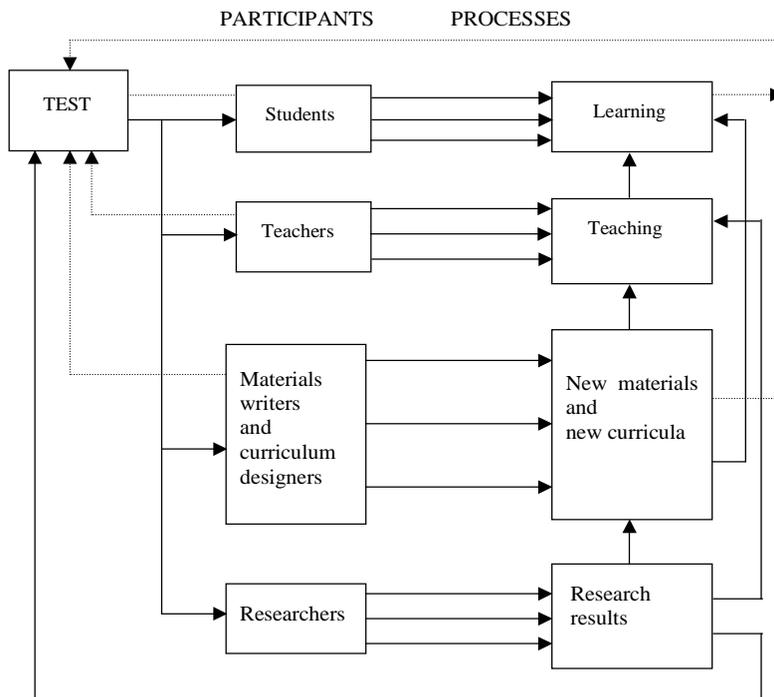


Figura 4: Um modelo básico de efeito retroativo (Bailey, 1996, p. 264)

A autonomia do aprendiz e auto-avaliação são também outros elementos vistos pela autora como importantes para potencializar o efeito de exames: aprendizes autônomos desenvolvem seus próprios valores em relação ao julgamento de seu progresso nos materiais ou habilidades a serem aprendidos. Essa “apropriação” e “auto-regulação” levam a um controle maior e mais profundo do material ensinado e avaliado, o que estaria relacionado à hipótese número 10 de Alderson & Wall: “um teste tem influência no grau e profundidade da aprendizagem”.

Em avaliações de rendimento conduzidas em sala de aula, isso significaria envolver os alunos na preparação dos exames e provas, e coletar informações sobre suas percepções a respeito de avaliações e tarefas (vide Avelar, 2001, Seção V). Se os alunos perceberem os exames como interativos e motivadores, têm mais chance de se prepararem e terem um desempenho melhor. Em testes externos, a auto-avaliação pode ser incluída através de questionários. Apresentar aos candidatos relatos diagnósticos detalhados, mostrando as várias dimensões do que está sendo avaliado, em vez de apenas um escore geral, também tem sido um outro fator potencializador de efeito retroativo positivo (vide Scaramucci, 2000/2001).

Shohamy, Donitsa e Ferman (1996) relatam resultados de um estudo que procurou avaliar os efeitos de longo prazo de dois exames nacionais em Israel, sendo um de inglês e o outro de árabe. Essa pesquisa traz uma outra dimensão importante do conceito, já mencionada por outros autores, mas que, nesse caso, vem acompanhada de evidências empíricas: a relação entre o status do teste (alta relevância ou baixa relevância), o status da língua e a intensidade do efeito retroativo. Quanto maior o status e prestígio da língua e do exame, maior o efeito retroativo por ele exercido. Uma outra conclusão importante desse estudo é que a intensidade do efeito pode aumentar ou diminuir depois de o exame ter sido implementado por um certo período, dependendo de fatores como status do teste e da língua avaliada, seu propósito, formato e habilidades.

Um outro estudo que traz elementos que permitem uma expansão da hipótese 15 de Alderson & Wall é relatado por Alderson e Hamp-Lyons (1996), conduzido para avaliar, no ensino de dois professores de uma mesma instituição, os efeitos do TOEFL — um exame de base estruturalista, reconhecido por exercer um efeito negativo no ensino, voltado exclusivamente para aumentar as notas sem levar necessariamente a melhorias no nível de proficiência (*test score pollution*). Apesar da popularidade do exame, é um dos poucos estudos até agora realizados para avaliar seu impacto. Os dados foram coletados através de entrevistas com alunos e professores, assim como observações em sala de aula regulares e de preparação específica para o exame⁸, ambas ensinadas pelos mesmos professores. Os resultados reforçam a complexidade do fenômeno ao concluir que “Testes têm efeitos retroativos de tipos e intensidades distintas em alguns professores e em alguns alunos”, uma constatação semelhante àquela feita por Cheng (1998) e Scaramucci (1999) e que foi

⁸ Mehrens and Kaminsky (1989) observam que quanto mais relevante for o exame maior é a necessidade que as pessoas sentem de se envolverem em práticas específicas de preparação. No Brasil, um bom exemplo é a institucionalização dos cursos pré-vestibulares. Situação semelhante ocorre no Japão, que também possui uma longa tradição de exames vestibulares.

denominada de “intensidade do efeito retroativo” ou “*washback intensity*”. Como afirmam os autores:

It is tempting to conclude that the TOEFL alone does not cause washback, but it is the administrators (who decree large classes), material writers (who provide no guidance to teachers on how to teach) and teachers themselves (who give little sign of thinking about how best to teach TOEFL) who cause the washback we have observed. (Alderson e Hamp-Lyons, 1996:295).

O estudo de Watanabe (1996) também apresenta dados de um estudo de caso conduzido no Japão para avaliar uma possível relação entre exames vestibulares e o ensino de inglês através do método de gramática-tradução, prevalente naquele país. Os resultados das observações em sala de aula mostram que fatores relacionados ao professor, sua formação, crenças e experiência interferem de forma positiva ou negativa com as características do exame na determinação de seu efeito (vide Seção V, para estudos conduzidos em contexto de exames vestibulares no Brasil).

O artigo de Wall (1996), trazendo mais detalhes sobre a implementação do exame no Sri Lanka, destaca a importância da teoria da inovação educacional (Fullan, 1991; Markee 1993), já apresentada em Alderson e Wall (1993), e de estudos sobre motivação e atitudes. Encerrando o volume, confirma a importância do contexto em que o exame é implementado e porque nem sempre testes têm o efeito que desejamos ou tememos.

Também relevantes para caracterizar essa fase de pesquisa são os estudos de Cheng (1998 e 1999), que buscou investigar de que maneira a introdução de um exame de inglês como segunda língua em escolas secundárias de Hong Kong⁹ foi percebida pelos professores e qual foi seu impacto em suas práticas de sala de aula. Os resultados mostram efeitos do exame distintos nas percepções de alguns professores e não de outros, conforme indicam os dados dos questionários. Na prática desses professores, entretanto, os efeitos foram menos intensos (vide Scaramucci 1998a, 1999b, 1999c e 2002, para resultados semelhantes).

Como essa revisão pode atestar, muitas dimensões importantes do conceito puderam ser adicionadas através das evidências empíricas dessa fase. Resultados semelhantes obtidos em contextos e exames distintos mostraram regularidades e padrões, importantes para a elaboração de uma teoria.

4. A FASE ATUAL DA PESQUISA

Uma terceira geração de estudos sobre efeito retroativo é apresentada em uma publicação recente, lançada em abril deste ano por ocasião do *Language Testing Research Colloquium* em Temecula, Califórnia. Organizado por Liying Cheng e Yoshinori Watanabe, o volume intitulado *Washback in Language Testing – Research Contexts and Methods* compreende onze artigos, dos quais três abordam aspectos conceituais e metodológicos; os demais são relatos de estudos conduzidos em diferentes partes do mundo (Estados Unidos, Inglaterra, Nova Zelândia, Austrália, Japão, Hong Kong e Israel). Um dos aspectos que chama a atenção no volume é o fato de todos os estudos ali relatados terem sido conduzidos

⁹ Vide também Lam (1994) e Andrews (2004).

em cenários em que o inglês é a língua avaliada, o que, de certa forma, pode ser visto como uma limitação, considerando-se a importância que o contexto e o status da língua têm em estudos dessa natureza. O relato de pesquisas realizadas em outros cenários ofereceria, sem dúvida alguma, um contraponto importante no estágio atual de elaboração de uma teoria.

Uma outra característica dessa coletânea é a retomada e reafirmação de aspectos já discutidos, sugerindo uma fase de consolidação de conhecimentos (vide capítulos de Watanabe e Cheng, discutindo resultados de seus projetos no Japão e Hong Kong, respectivamente). O que ainda se observa nesses estudos é a tendência da fase anterior de focalizarem a influência de exames *externos* no *ensino* e nas *ações* de professores em detrimento de estudos que consideram os efeitos de *avaliações de rendimento, internas* aos processos de ensino/aprendizagem (geralmente conduzidas pelo professor em sua sala de aula) em outros participantes do processo (*alunos, pais, etc*) ou ainda na *aprendizagem*. A única exceção é o artigo de Burrows, que examina o conceito no contexto de um exame de rendimento na Austrália e propõe um modelo (vide Figura 4 acima) que coloca o professor, com suas crenças, pressuposições e conhecimentos, no centro do efeito retroativo.

Na apresentação do livro, Charles Alderson afirma que, embora não se possa negar, pelas evidências empíricas acumuladas nesses anos, que efeito retroativo existe, sabemos hoje tratar-se de um conceito altamente complexo, muito distante da noção que se tinha um tempo atrás de que exames causam efeito negativo no ensino. A questão que hoje se coloca não é mais se efeito retroativo existe, mas como podemos descrevê-lo, porque existe e quais são suas causas.

Cheng e Curtis abrem o volume, fazendo uma revisão da literatura acerca do conceito, abordando sua origem, definições e complexidade do fenômeno, suas funções e mecanismos. Os autores apontam duas áreas importantes de estudos sobre efeito retroativo: aqueles relativos a exames de grande escala, geralmente tradicionais, de múltipla escolha, percebidos por suas influências negativas, e aqueles relativos a exames que foram modificados ou até mesmo elaborados (como, por exemplo, o Celpe-Bras) de forma a exercerem um efeito positivo ou redirecionador do ensino, o que tem sido denominado de “*washback by design*”, e que têm mostrado efeitos positivos, negativos ou até mesmo nenhuma influência.

O artigo de Andrews, por sua vez, explora a relação entre efeito retroativo e inovação curricular também a partir de uma revisão de estudos e da teoria atual sobre efeito retroativo no cenário educacional geral e em LA. O autor chega a três conclusões que, apesar de importantes, são reafirmações de outros autores: abordagens coercitivas não são eficientes para promover, de forma eficaz, inovações que sejam duradouras e auto-sustentáveis; que o contexto em que a inovação está sendo implementada é importante, uma vez que os resultados da implementação de um exame podem não ser antecipados; a necessidade de pesquisas futuras.

O artigo de Watanabe, apesar de retomar aspectos já mencionados anteriormente, é importante pela sistematização que apresenta das várias dimensões do conceito de efeito retroativo. O autor assina ainda um outro capítulo, com dados de sua pesquisa de doutorado sobre a influência de exames vestibulares no ensino secundário no Japão. Da mesma forma, Cheng também retoma, em outro capítulo, resultados de seus estudos em Hong Kong.

Dois capítulos (Hayes & Read e Saville & Hawkey) são devotados a estudos relacionados ao IELTS (International English Language Testing System): enquanto o primeiro investiga efeitos desse exame em cursos preparatórios na Nova Zelândia, o segundo discute o desenvolvimento de instrumentos para coleta de dados em estudos sobre o impacto desse exame em livros e outros materiais.

A partir dos aspectos discutidos na literatura internacional, apresento, na próxima seção, os estudos em andamento e concluídos, conduzidos no cenário brasileiro.

5. EFEITO RETROATIVO EM CONTEXTO BRASILEIRO

Os estudos pioneiros sobre o conceito de efeito retroativo em cenários de ensino de Inglês como língua estrangeira foram realizados por Scaramucci (1998b, 1999c e 2002) e Gimenez (1999). Os dois estudos, de natureza qualitativa e exploratória, foram idealizados conjuntamente e, apesar de terem metodologias e procedimentos semelhantes de análise e coleta, foram realizados em contextos distintos: enquanto o de Scaramucci (*op. cit.*) investigou o efeito retroativo dos exames vestibulares de Inglês da Universidade Estadual de Campinas nas atitudes, comportamentos e ações de três professores em escolas de nível médio (particular, pública e curso preparatório particular) daquela cidade, o de Gimenez (*op. cit.*) focalizou o efeito do exame de inglês da Universidade Estadual de Londrina em contextos semelhantes na cidade de Londrina.

Para suprir as limitações metodológicas identificadas em estudos anteriores, os estudos usaram uma triangulação de informações/pontos de vista/métodos de pesquisa: dados de observação em sala de aula (anotações e gravações em áudio) e de questionários/entrevistas com professores e alunos foram triangulados com dados das análises da proposta/materiais do exame e da Proposta Curricular de Língua Estrangeira da Secretaria de Educação do Estado de S. Paulo (no estudo de Scaramucci) e da proposta equivalente do Estado do Paraná (no de Gimenez *op. cit.*), tendo em vista também sua função direcionadora.

Comparando as práticas de sala de aula e dados de questionário e entrevistas com professores nos três contextos com a proposta do exame de inglês da Unicamp, Scaramucci (1998b, 1999c e 2002) concluiu que o efeito, além de parcial — na medida em que refletiu apenas alguns aspectos da proposta, percebidos como salientes ou inovadores pelos professores — tinha intensidades diferentes, constituindo-se evidência de que sua ação não é determinista. Apesar de o exame ter sido orientado teoricamente por uma visão de leitura como construção de sentidos, de estar em uso por mais de uma década, não conseguiu, principalmente levando-se em conta os dados de observação em sala de aula, mudar as crenças dos professores em relação à leitura e ao ensino da leitura, ainda vistos como um processo passivo de decodificação, seguindo a tradição do ensino de leitura em língua materna predominante.

Os aspectos do exame que parecem ter tido um impacto nos professores e sua prática foram o conteúdo, ou seja, a leitura, que passou a ser trabalhada nos três contextos; os textos autênticos de gêneros e assuntos diversos, que foram utilizados esporadicamente em dois dos contextos; as perguntas de compreensão em português, que passaram a ser utilizadas juntamente com as perguntas em inglês que acompanhavam os materiais didáticos

utilizados em dois dos contextos; e, finalmente, o uso de respostas abertas juntamente com as de múltipla escolha que já utilizavam, para avaliar os textos em sala de aula.

O efeito de aspectos mais implícitos do exame — sua visão de linguagem interativa e de leitura como construção de sentidos, que pressupõe o desenvolvimento de sub-habilidades diversas e conscientização a respeito de níveis de compreensão distintos, dentre outros — foi praticamente nulo, principalmente se forem levados em conta os dados de sala de aula.

Esse trabalho nos permite concluir, portanto, da mesma forma que muitos outros, que inovações educacionais não são conseguidas, automática e unicamente, através da implementação de propostas direcionadoras e exames externos. Observamos, portanto, que o efeito potencial inovador de exames é limitado, principalmente, pelas crenças e formação do professor, sugerindo — como mais promissoras — propostas de formação que levem o professor a refletir e teorizar sua prática a partir da explicitação de suas crenças, concepções e pressupostos teóricos (Scaramucci 1999c, 2000/2001; Burrows 2004).

Gimenez (1999), a partir de dados de aulas observadas em um cursinho preparatório, mostra que somente algumas partes do exame de inglês da Universidade Estadual de Londrina tiveram efeito retroativo nesse contexto. Conclui seu artigo também chamando a atenção para a importância das crenças do professor (em relação ao ensino de língua estrangeira num contexto de preparação para um exame e em relação às chances de aprovação de seus alunos) como fatores mediadores dos efeitos do exame nas práticas de ensino da professora envolvida.

Ainda no contexto do exame vestibular da Unicamp, foram conduzidos mais dois estudos¹⁰, com focos distintos: Bartholomeu (2002), que investigou o efeito retroativo das provas de Inglês dos vestibulares das três universidades paulistas (USP, Unicamp e Unesp) nas percepções, atitudes e motivações de alunos do terceiro ano de uma escola particular no interior paulista. E Correia (2003), que, em uma pesquisa de metodologia semelhante, procurou levantar as percepções dos alunos de um curso preparatório comunitário na cidade de Campinas sobre a prova de LE do vestibular da Unicamp e o efeito retroativo por ela exercido na preparação desses alunos para o exame. Ao centrar seu foco nos alunos, a primeira investigação tinha por objetivo preencher uma das lacunas identificadas nos estudos sobre o efeito retroativo e já mencionada neste artigo, ou seja, de focalizar o professor e seu ensino. Os resultados mostram um efeito retroativo não determinista, que se manifesta em intensidades diferentes, dependendo das características individuais dos alunos envolvidos. Oferece evidências também para se concluir que o ensino pode potencializar esse efeito, uma vez que as atitudes e as motivações a respeito das provas e de sua preparação parecem ser dependentes do professor e do seu ensino, dada a relação entre a cultura de aprender dos alunos e a de ensinar do professor.

Correia (2003), por sua vez, complementa os dados de Bartholomeu (*op. cit.*), ao fazer um levantamento de como os alunos de um curso pré-vestibular preparavam-se para o exame de inglês da Unicamp. Como o curso preparatório em questão era mantido pela universidade e ensinado por alunos e ex-alunos, acreditava-se que a preparação seria, nesse

¹⁰ Os estudos em questão foram conduzidos sob minha orientação no Programa de Pós-graduação em Lingüística Aplicada, Unicamp como parte do projeto CNPq 301151/96-2 (Scaramucci, 2001).

caso, mais específica para o exame da Unicamp do que para outros cursos preparatórios, e, conseqüentemente, esperava-se um efeito retroativo mais intenso. Os resultados mostraram, entretanto, da mesma forma que no estudo anterior, que a preparação nesse curso também visava outros vestibulares e não era específica para a proposta de exame da Unicamp. Além disso, apesar de muitos alunos terem demonstrado conhecer a prova de inglês da Unicamp, declararam preparar-se da mesma maneira que o fariam para outros vestibulares de universidades públicas do país, que priorizam a gramática em suas provas, mostrando que o exame não conseguiu mudar sua concepção de leitura. Da mesma forma que no estudo de Bartholomeu, pode-se observar que a maioria dos alunos parecia contentar-se com as informações sobre a prova supridas pelas professoras, limitando-se às atividades e orientações propostas em sala de aula, o que pode ser indicativo, mais uma vez, de que pelo menos em nossa cultura de aprender, o ensino é um fator importante por interagir com as características do exame na determinação de seu efeito.

Os dados de observação em sala de aula evidenciaram que, embora as professoras das turmas observadas tivessem declarado conhecer a prova de inglês da Unicamp, reconhecendo seu efeito no conteúdo das aulas e material didático utilizado, deixaram de focalizar aspectos que realmente poderiam instrumentalizar os alunos para se sair bem numa prova fundamentada por uma visão de leitura como construção de sentidos, fazendo uso constante de práticas de tradução e leitura mais condizentes com visões tradicionais.

Avelar (2001), ao investigar o efeito da implementação de uma proposta de avaliação negociada em uma classe de inglês de uma escola de ensino médio e técnico de uma cidade de Minas Gerais, nas percepções do professor e dos alunos sobre o processo de ensino/aprendizagem, buscou suprir uma outra lacuna nos estudos sobre efeito retroativo — foco em exames externos em detrimento de avaliações de rendimento. De natureza etnográfica, a pesquisa contou com duas fases distintas, sendo a primeira diagnóstica e a segunda interventiva, com dados obtidos através de gravação de aulas em áudio e vídeo, diário e anotações de campo da professora-pesquisadora e questionários respondidos pelos alunos.

Os resultados mostram que a avaliação, com sua potencialidade reflexiva, pode ampliar a percepção da professora e alunos sobre o processo de ensino/aprendizagem de LE. Os alunos mostraram-se capazes de relacionar suas dificuldades na aprendizagem com os problemas vivenciados na prática, percebendo a contribuição ou não das atividades propostas para sua aprendizagem, discutindo e questionando a maneira com que foram avaliados pelo professor. A implementação da proposta também impôs à professora-pesquisadora, entretanto, a tarefa de confrontar-se com suas próprias crenças, sucessos e fracassos.

Souza (2002) investigou o ensino da escrita em curso particular de inglês como LE em uma cidade do interior paulista, procurando identificar as forças determinantes desse ensino — em outras palavras, se podia ser decorrente da concepção de escrita do professor ou daquela que fundamenta seu material didático, ou se tal ensino era influenciado pela concepção de linguagem presente em um dos exames externos de proficiência a que os alunos dessa escola geralmente se submetiam, ou ainda à visão de linguagem que fundamentava as provas finais elaboradas pela escola. Para a coleta de dados foram utilizados um questionário, respondido por professores da escola em questão, e um outro pelos alunos

de uma das turmas escolhida para observação. Uma entrevista também foi conduzida com o professor da turma selecionada, assim como análises do material didático utilizado, das provas finais e do exame em questão.

Os dados revelaram que o exame de proficiência e as provas finais do curso em questão apresentaram uma concepção inovadora de linguagem escrita. De acordo com essa concepção, a escrita é vista como processo, tem uma função social e está integrada com outras habilidades. Entretanto, embora o professor da turma reconhecesse a importância de trabalhar a escrita por causa da avaliação (exame externo e provas finais) não foram observados momentos de trabalho efetivo de ensino da escrita, que era constantemente avaliada mas não efetivamente ensinada. A análise do material didático corrobora essa visão de escrita como atividade auxiliar, utilizada como reforço para as outras habilidades. Dessa forma, pode-se concluir que a avaliação, nesse contexto, embora tenha despertado o interesse e a preocupação do professor pela escrita, não foi capaz de mudar a concepção que fundamenta sua prática, que continuou a ser orientada por concepções de base estruturalista, presentes no material didático.

Lanzoni (2004) busca investigar, através de dois estudos, sendo um qualitativo e um outro quantitativo, o efeito retroativo de um exame de proficiência em leitura de textos acadêmicos em inglês — *TEAP* (Test of English for Academic Purposes) — nas atitudes, preparação e aprendizagem de candidatos a programas de mestrado na área de saúde de uma renomada universidade pública paulista. O exame, por ser pré-requisito para participação no processo seletivo, pode ser considerado de alta-relevância e, como tal, capaz de exercer efeito retroativo. Dados de noventa sujeitos, obtidos através de um questionário, foram analisados quantitativamente e contrastados com análises de seu desempenho nos exames. Dados de três sujeitos que prestaram o exame *TEAP* mais de uma vez, antes de obterem a pontuação mínima necessária, foram analisados qualitativamente. Os resultados mostram que, embora o exame pareça afetar a maioria dos candidatos, levando-os a fazerem coisas que não fariam se não fosse pela necessidade de se submeterem a ele, os efeitos que causa não têm a mesma intensidade. Em alguns dos sujeitos, o efeito foi até mesmo nulo, o que o autor atribui a fatores de ordem pessoal ou psicológica (auto-estima, chances de aprovação, por exemplo, dentre outros).

Uma outra conclusão do estudo é a constatação de uma relação inversamente proporcional entre nível de proficiência em leitura em inglês do candidato, percepção de relevância e, conseqüentemente, efeito retroativo, que parece ser menor em sujeitos com maior domínio do idioma e maior em sujeitos com baixa proficiência.

Por outro lado, sujeitos com níveis de proficiência mais altos, mas com uma auto-estima baixa parecem ter mais afetados pelo exame do que aqueles com auto-estima alta.

Esses resultados corroboram evidências obtidas a partir da primeira fase de um estudo abrangente (Scaramucci 2001, 2003, 2004a e 2004b) cujo objetivo é investigar o efeito retroativo do exame *Celpe-Bras* nas percepções e ações do professor de Português LE, assim como das percepções e atitudes dos candidatos. Para a coleta dos dados foram elaborados dois questionários: um para professores, enviado eletronicamente para os centros aplicadores do exame no Brasil e no exterior, e um para os candidatos, preenchido após cada aplicação do exame.

As respostas às perguntas referentes às percepções que os professores tinham de seu ensino antes da implementação do exame assim como ao impacto que o exame causou nessa prática indicam duas tendências: a) reconhecimento de que essa prática era distinta daquela que hoje se observa (geralmente com foco na gramática, em exercícios mecânicos e descontextualizados, em alguns contextos com foco excessivo na oralidade em detrimento da produção escrita, falta de tratamento referente a aspectos culturais, e, portanto, com materiais de ensino voltados a esses aspectos) e que o exame teve impacto nessa prática; e b) reconhecimento de que o ensino já era fundamentado em visão semelhante à do exame e que este veio apenas corroborar essa visão e os materiais didáticos já utilizados. Embora as duas tendências possam ser observadas nos dois grupos, há, sem dúvida alguma, um número maior de professores atuantes no Brasil e em um dos centros no exterior nessa segunda categoria. O diferencial, nesse caso, parece ser a formação do professor, uma vez que os contextos de aplicação do exame no Brasil são geralmente universidades e os professores mais bem informados das tendências contemporâneas na área de ensino de língua estrangeira.

As respostas de muitos professores às outras perguntas que buscam aprofundar o entendimento da prática antes e depois do exame, entretanto, sugerem que, na realidade, o exame não parece ter *transformado* sua visão de linguagem, mas apenas funcionado no sentido de os fazer incorporar outros elementos a uma prática já existente (mais exercícios orais em alguns contextos, mais textos para leitura e escrita em outros). Esses resultados são limitados, pois se baseiam em depoimentos dos professores e serão, numa etapa futura do estudo, complementados por dados de observação em sala de aula.

Esse estudo também procurou investigar, ainda nesse mesmo contexto, quais as atitudes e percepções dos candidatos a respeito do exame e do impacto que causa em sua preparação. Os dados do questionário respondido pelos candidatos na aplicação de abril de 2004 confirmam a atitude positiva crescente dos mesmos em relação ao exame. A grande maioria, com algumas poucas exceções percebem o exame como “diferente em relação aos demais”, “divertido”, “avalia o que deveria avaliar (leitura, redação)”, “uso da língua em situações quase reais (escrever e-mails, cartas, etc.)”; “um estímulo para estudar o português”, dentre outros. As respostas indicam, portanto, que o exame tem valor aparente para os candidatos, o que é um aspecto fundamental para sua aceitação e motivação para preparar-se. Um número muito pequeno dos que o responderam menciona não gostar do exame por achá-lo subjetivo, por incluir muita redação; o exame deveria, na opinião desses candidatos, ser de múltipla escolha e avaliar a gramática explicitamente, a exemplo de outros exames como DELE e TOEFL.

Além desses estudos, podemos citar ainda o de Pessoa (2002 e neste volume), conduzido para investigar o efeito retroativo do PAS da Universidade de Brasília no ensino de Inglês como língua estrangeira de escolas daquela cidade.

Scaramucci & Retorta (2004), por sua vez, relatam os resultados parciais de um estudo em andamento¹¹ cujo objetivo é investigar o efeito retroativo da prova de língua inglesa do

¹¹ Tese de doutorado de Miriam Sester Retorta no Programa de Pós-graduação em LA, Unicamp sob minha orientação.

vestibular da Universidade Federal do Paraná (UFPR) no ensino, aprendizagem e avaliação de ensino médio público e privado e cursos pré-vestibulares. O projeto de pesquisa, como os demais, triangula dados de diversas fontes. Os resultados obtidos até agora indicam uma forte influência do vestibular da UFPR em escolas particulares e cursos-pré-vestibulares não somente perceptível no ensino, mas principalmente, nas avaliações conduzidas nessas escolas. A maioria dos professores e alunos das escolas particulares e cursos pré-vestibulares parece mudar seu comportamento e hábitos em relação ao ensino, aprendizagem e avaliação, principalmente no último ano do ensino médio. A influência do Vestibular da UFPR é praticamente inexistente nas escolas públicas investigadas.

Em andamento¹² estão, ainda, mais dois trabalhos que, embora não investiguem diretamente o conceito de efeito retroativo, o utilizam como potencial para implementação de mudanças. O objetivo do primeiro é oferecer aos professores de Português LE, a partir da explicitação da fundamentação teórica do exame Celpe-Bras e de seus princípios norteadores, subsídios teórico-didáticos que lhes permitam preparar melhor seus alunos para o exame e, conseqüentemente, potencializar as chances de efeito retroativo benéfico ou positivo do exame no ensino de Português LE/L2 no Brasil e no exterior. A crença de que sem cursos preparatórios específicos — aqueles que transformam a sala de aula em espaço para treinamento de exames passados — não é possível uma boa preparação é compartilhada por muitos professores e alunos, apesar de muitos pesquisadores considerarem esse tipo de preparação uma espécie de efeito retroativo negativo. O exame Celpe-Bras oferece um contexto particularmente interessante para investigar essas questões por ser um exame inovador, fundamentado em uma visão de linguagem sócio-interativa e, portanto, não condizente com aquela que fundamenta a maioria das aulas de Português LE. Essa iniciativa de explicitar os conceitos norteadores do exame, sugerindo subsídios teórico-didáticos para que os professores possam não apenas preparar bem seus alunos para o exame, mas redirecionarem sua forma de ensinar Português LE encontra respaldo nos resultados das inúmeras pesquisas aqui relatadas que têm mostrado as limitações da implementação de inovações unicamente através da implementação de propostas ou exames inovadores.

O segundo estudo também parte do pressuposto de que exames podem ser “alavancas para a mudança” e, como tal, visam a oferecer subsídios teóricos para a elaboração de um exame de proficiência em inglês como língua estrangeira para o professor de inglês não nativo, atuante em contexto brasileiro. Ao caracterizar um perfil de proficiência desejável para esse professor, espera-se sejam definidas diretrizes para sua formação. Um exame de proficiência para professores de inglês elaborado nesses moldes poderia fornecer elementos importantes, portanto, para a melhoria da qualidade da formação desse profissional e levar, conseqüentemente, a um re-direcionamento da prática de ensino de inglês no Brasil.

¹² Meirélen Saviano Rodrigues e Teresa Martins, respectivamente, são mestrandas no Programa de Pós-graduação em LA na Unicamp, trabalhando sob minha orientação.

VI CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi mostrar o estado da arte das pesquisas sobre o conceito de efeito retroativo no contexto de ensino/aprendizagem de línguas. Para isso, foram apresentados as contribuições mais importantes, os aspectos controvertidos, as lacunas e limitações na pesquisa.

Apesar de os estudos, nos últimos anos, terem oferecido contribuições importantes, a ponto, como salienta Alderson (2004), de não termos mais dúvida de que efeito retroativo realmente existe, podemos dizer que o debate ainda não está resolvido, pois não há até o momento uma teoria satisfatória para explicar seus mecanismos. O que parece diferente, todavia, na literatura recente, é a consciência de que a única maneira de se produzir essa teoria é através de evidências que possam contrapor-se à especulação que imperava anos atrás. O maior valor desses estudos, entretanto, não está nas tentativas individuais desse empreendimento, mas em descrições que, compostas a partir de estudos diferentes em contextos diferentes, mostre padrões e regularidades que possam ser documentados (Scaramucci, 2002), de forma a nos orientar no sentido de fazer com que testes sejam instrumentos educacionalmente benéficos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALDERSON, J. C., & WALL, D. (1990). The Sri Lankan O-Level Evaluation Project: Second Interim Report. *Lancaster University*.
- _____. (1991). The Sri Lankan O-Level Evaluation Project: Third Interim Report. *Lancaster University*.
- _____. (1992). Does washback exist? *Working Paper Series II*. CRILE, Lancaster University Republicado em *Applied Linguistics* 14, p. 115-129, 1993.
- ALDERSON, J. C. (2004). Foreword. Em Cheng, L., Watanabe, Y & Curtis, A. (eds), *Washback in Language Testing – Research Contexts and Methods*. Lawrence Erlbaum Associates, New Jersey.
- ALDERSON, J. C. & L. HAMP-LYONS (1996). TOEFL preparation courses: a study of washback. *Language Testing* 13 (3), p. 281-297.
- ALMEIDA FILHO, J. C. & SCHMITZ, J. R. (1998). *Glossário de Lingüística Aplicada*. Pontes Editores, Campinas.
- ANDREWS, S. (2004). Washback and curriculum innovation. Em Cheng, L., Watanabe, Y & Curtis, A. (eds), *Washback in Language Testing – Research Contexts and Methods*. Lawrence Erlbaum Associates, New Jersey, p. 137-50.
- AVELAR, S. L.T. (2001). *Mudanças na concepção e prática da avaliação e seu efeito no ensino/aprendizagem de língua estrangeira (Inglês) em uma escola de ensino médio e técnico*. Dissertação de mestrado. Programa de Pós-graduação em Lingüística Aplicada, IEL, Unicamp.
- BACHMAN, L. F & PALMER, A. S. (1996). *Language Testing in Practice*. Oxford, England: Oxford University Press.

- BAILEY, K. (1996). Working for washback: *A review of the washback concept in language testing*. *Language Testing*, 13, p. 257-279.
- _____. (1999). Washback in language testing. TOEFL Monograph Series 15, Educational Testing Service, <http://www.ets.org>.
- BARTHOLOMEU, M. A. N. (2002). *Prova de língua estrangeira (Inglês) dos vestibulares e sua influência nas percepções, atitudes e motivações de alunos do terceiro ano do ensino médio*. Dissertação de mestrado, Programa de Pós-graduação em Lingüística Aplicada, IEL, Unicamp.
- BURROWS, C. (2004). Washback in classroom-based assessment: a study of the washback effect in the Australian adult migrant English program. Em Cheng, L., Watanabe, Y & Curtis, A. *Washback in Language Testing – Research Contexts and Methods*. Lawrence Erlbaum Associates, New Jersey, p. 113-128.
- CARROLL, B. J. (1980). *Testing communicative performance*. Oxford: Pergamon Press.
- CHENG, L. (1998). *The washback effect of public examination change on classroom teaching: an impact study of the 1996 Hong Kong certificate of education in English on the classroom teaching of English in Hong Kong secondary schools*. Tese de doutorado inédita. Universidade de Hong Kong, Hong Kong.
- _____. (1999). Changing assessment: washback on teacher perspectives and actions. *Teaching and teacher education*, 15 (3), p. 253-271.
- _____. (2004). The washback effect of a public examination change on teachers' perceptions towards their classroom teaching. Em Cheng, L., Watanabe, Y & Curtis, A. (eds) *Washback in Language Testing – Research Contexts and Methods*. Lawrence Erlbaum Associates, New Jersey, p. 147-170.
- _____. WATANABE, Y & CURTIS, A. (eds) (2004). *Washback in Language Testing – Research Contexts and Methods*. Lawrence Erlbaum Associates, New Jersey.
- CORREIA, R. M. D. (2003). *O efeito retroativo da prova de inglês do vestibular da Unicamp na preparação de alunos em um curso preparatório comunitário*. Dissertação de mestrado. Programa de Pós-graduação em Lingüística Aplicada, IEL, Unicamp.
- DAVIES, A. (2003). Three heresies of language testing research. *Language Testing* 20 (4), p. 355-368.
- FORBES, D. Selling English short. (1973). *English Language Teaching Journal* 27, p. 132-137.
- FREDERIKSEN, J.R. The real test bias. (1984). *American Psychologist*, March, p. 193-202.
- FREDERIKSEN, J.R. & COLLINS, A. (1989). A systems approach to educational testing. *Educational Researcher* 18 (9), p. 27-32.
- FULLAN, M.G. & STIEGELBAUER, S. (1991). *The new meaning of educational change*. Segunda edição. Londres: Cassell.
- GATES, S. (1995). Exploiting washback from standardized tests. Em J.D. Brown & S.O. Yamashita (eds) *Language Testing in Japan*, p. 101-106. Tokyo: Japanese Association for Language Teaching.
- GIMENEZ, T. (1999). Concepções de linguagem e ensino na preparação de alunos para o vestibular. *Trabalhos em Lingüística Aplicada*, 34, p. 21-37.
- HALADYNA, T.H., NOLEN, S.B. & HASS, N.S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher* 20 (5), p. 2-7.
- HAYES, B. & READ, J. Test preparation in New Zealand: preparing students for the IELTS academic module. Em Cheng, L., Watanabe, Y & Curtis, A. (eds), *Washback in Language Testing – Research Contexts and Methods*. Lawrence Erlbaum Associates, New Jersey, p. 97-111.

SCARAMUCCI – Efeito retroativo da avaliação...

- HUGHES, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading.
- _____. (1988). *Testing English for University Study*. ELT documents 127, Modern English Publications.
- _____. (1989). *Testing for Language Teachers*. Cambridge University Press.
- KELLAGHAN, T, G. MADAUS & P. AIRASIAN. (1982). *The effects of standardized testing*. London: Kluwen, Nijhoff Publishing,
- KHANIYA, T.R. (1990). Examinations as instruments for educational change: investigating the washback effect of the Nepalese English exams. Tese de doutorado, University of Edinburgh, Scotland.
- LAM, H. P. (1994). Methodology washback – an insider’s view. Em Nunan, D., Berry, R. & Berry, V. Bringing about change in language education: *Proceedings of the International Language in Education Conference*. Hong Kong: University of Hong Kong.
- LANZONI, H. DE P. (2004). *Exame de proficiência em leitura de textos acadêmicos em Inglês: um estudo sobre efeito retroativo*. Tese de doutorado. Programa de Pós-graduação em Linguística Aplicada, Instituto de Estudos da Linguagem, Unicamp.
- MADSEN, H. (1976). New Alternatives in EFL exams or “How to avoid selling English short”. *English Language Teaching Journal* 30 (2), p. 135-144.
- MCNAMARA, T. F. (1996). Policy and social considerations in language assessments. *Annual Review of Applied Linguistics* 18, p. 304-19.
- MARKEE, N. (1993). The diffusion of innovation in language teaching. *Annual Review of Applied Linguistics* 13, p. 229-243.
- MEHRENS, W.A. & KAMINSKY, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: issues and practice* 8, p. 14-22.
- MESSICK, S. (1989). Validity. Em Linn, R(ed), *Educational measurement*, terceira edição, p. 13-103. New York: Macmillan.
- _____. (1996). Validity and washback in language testing. *Language Testing* 13 (3), p. 241-255.
- _____. (1994). The interplay of evidence and consequences in the validation of performance assessment. *Educational Researcher* 23 (1), p. 13-23.
- MORROW, K. (1986). The evaluation of tests of communicative performance. Em Portal, M (ed), *Innovations in language testing. Proceedings of the IUS/NFER conference*, p. 1-13. London: NFER/Nelson.
- _____. (1991). Evaluating communicative tests. Em S. Anivan (ed.), *Currents developments in language testing*, pg 111-118, Singapore: Regional Language Centre.
- PARIS, S.G, LAWTON, T.A, TURNER, J.C. & ROTH, J.L. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher* 20 (5), p. 12-19.
- PEARSON, I. (1988). Tests as levers for change. In D. Chamberlain & R. Baumgardner (eds), *ESP in the Classroom: Practice and Evaluation*.
- PESSOA, A. R. (2002). *O efeito retroativo do Programa de Avaliação Seriada da Universidade de Brasília no Ensino de língua estrangeira do Distrito Federal*. Dissertação de mestrado. Programa de Pós-graduação em Linguística Aplicada, Universidade de Brasília, DF.
- POPHAM (1987). The merits of measurement -driven instruction. *Phi Delta Kappa* 68, p. 679-682.

- QI, L. (2004). Has a high-stakes test produced the intended changes? Em Cheng, L., Watanabe, Y & Curtis, A. (eds) (2004). *Washback in Language Testing – Research Contexts and Methods*. Lawrence Erlbaum Associates, New Jersey, p. 171-190.
- SAVILLE, N. & HAWKEY, R. (2004). The IELTS impact study: investigating washback on teaching materials? Em Cheng, L., Watanabe, Y & Curtis, A. (eds) (2004). *Washback in Language Testing – Research Contexts and Methods*. Lawrence Erlbaum Associates, New Jersey, p.73-96.
- SCARAMUCCI, M.V.R. (1995). O projeto Celpe-Bras no âmbito do Mercosul: contribuições para uma definição de proficiência comunicativa. In Almeida Filho, J.C. (org.), *Português para Estrangeiros: Interface com o Espanhol*. Pontes, p. 77-90.
- _____. (1998a). Vestibular: instrumento direcionador do ensino de segundo grau? Trabalho apresentado no V Congresso Brasileiro de Linguística Aplicada (V CBLA), em Porto Alegre, Rio Grande do Sul.
- _____. (1998b). O efeito retroativo dos vestibulares de língua inglesa da Unicamp no ensino de segundo grau de escolas públicas e particulares de Campinas. Relatório final de pesquisa, FAPESP 95/06551-9.
- _____. (1999a). Celpe-Bras: um exame comunicativo. Em Santos, P. e M. J. Cunha (orgs.) *Ensino e Pesquisa em Português para Estrangeiros* Brasília: Editora Universidade de Brasília, p. 75-81.
- _____. (1999b). University entrance examinations and EFL teaching: a study of washback in a Brazilian context. Trabalho apresentado no 12th World Congress of Applied Linguistics (AILA99), em Tóquio, Japão.
- _____. (1999c). Vestibular e ensino de língua estrangeira (Inglês) na escola pública. Em *Trabalhos em Linguística Aplicada* 34, p. 7-20, Departamento de Linguística Aplicada, Unicamp.
- _____. (2000/2001). Propostas curriculares e exames vestibulares: potencializando o efeito retroativo benéfico no ensino de LE (Inglês). *Contexturas* 5, p. 97-109, APLIESP, Ibilce, Unesp, São José do Rio Preto, SP.
- _____. (2001). O efeito retroativo da avaliação no ensino/aprendizagem, nas atitudes de professores e alunos e na elaboração de material didático em língua estrangeira. Relatório de pesquisa CNPq 301151/96-2 (NV).
- _____. (2002). Entrance examinations and TEFL in Brazil: a case study. *Revista Brasileira de Linguística Aplicada* 2 (1), Universidade Federal de Minas Gerais, Belo Horizonte, MG, p. 61-81.
- _____. (2003). Evaluación de “proficiencia” em lengua extranjera: relaciones com la enseñanza y evaluación de rendimiento. Palestra apresentada no Sexto Ciclo Internacional de Enseñanza de Lenguas Extranjeras, Buenos Aires, Argentina.
- _____. & Rodrigues, M.S. Compreensão (oral e escrita) e produção escrita no exame Celpe-Bras: análise do desempenho de candidatos hispano-falantes. Em *Português para falantes de espanhol - artigos selecionados escritos em português e inglês* (orgs) Simões, A, Carvalho, A.M. & Wiedemann, L. Pontes Editores, SP, p. 126-144 (no prelo).
- _____. (2004). Celpe-Bras exam and teachers’ perceptions of impact on teaching/assessment of Portuguese as a FL/L2. Trabalho apresentado no *Language Testing Research Colloquium*, Temecula, Califórnia, USA.
- _____. (2004). Efeito retroativo da avaliação no ensino/aprendizagem em contextos diversos de língua estrangeira/segunda língua: percepções e ações de professores e alunos. Relatório de pesquisa CNPq 03822/2002-3.
- _____. & RETORTA, M.S. (2004). Washback effects of entrance examinations on high school teaching/learning/assessment in Parana, Brazil. Trabalho apresentado na Ontario Conference in Applied Linguistics (OCAL). Queen’s University, Ontário, Canadá.
- SHEPARD, L.A. (1993). The place of testing reform in educational reform: a reply to Cizek. *Educational Researcher* 22 (4), p. 10-14.
- SHOHAMY, E, DONITSA-SCHMIDT, S., & FERNAM, L. (1996). Test impact revisited: washback effect over time. *Language Testing* 13 (3), p. 298-317.
- SMITH, M.L. (1991). Put to the test: the effects of external testing on teachers. *Educational Researcher* 20 (5), p. 8-11.

- SOUZA, L.G. (2002). *Ensino da produção escrita em língua estrangeira (inglês) em um curso de línguas: influência da avaliação ou da concepção de escrita do professor?* Dissertação de mestrado. Programa de Pós-graduação em Linguística Aplicada, IEL, Unicamp.
- SPOLSKY, B. (1994). The examination-classroom backwash cycle: some historical cases. Em D. Nunan, R. Berry & V. Berry (eds) Bringing about changes in language education. *Proceedings of the International Language Education Conference*, p. 55-66, University of Hong Kong, Hong Kong.
- VAN LIER, L. (1989). Reeling, writhing, drawing, stretching and feinting in coils: oral proficiency interviews as conversation. *TESOL Quarterly* 23, p. 489-508.
- WALL, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing* 13 (3), p. 334-354.
- _____. (2000). The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System* 28(4), p. 499-509.
- _____. & J.C. ALDERSON. (1993). Examining washback: The Sri Lankan impact study. *Language Testing* 10 (1), 41-69.
- WATANABE, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing* 13 (3), p. 318-333.
- _____. (1997). *Washback effects of the Japanese university entrance examination: classroom –base observation*. Tese de doutorado inédita, Lancaster University, UK.
- _____. (2004). Methodology in washback studies. Em Cheng, L., Watanabe, Y & Curtis, A. (eds) (2004). *Washback in Language Testing – Research Contexts and Methods*. Lawrence Erlbaum Associates, New Jersey, p. 19-36.
- WESDORP, H. (1982). Backwash effects of language testing in primary and secondary education. Stichting Centrum voor onderwijsonderzoek van de Universiteit van Amsterdam.
- WESCHE, M..B. (1983) Communicative language testing in a second language. *Modern Language Journal* 67 (1), p. 41-55.