

A AVALIAÇÃO E A MELHORIA DA EFICÁCIA DA APRENDIZAGEM

Rui Alberto Ferreira de Castro¹, José Matias Alves²

rfcastro8@gmail.com, jalves@porto.ucp.pt

¹ ² Centro de Investigação para o Desenvolvimento Humano, Faculdade de Educação e Psicologia, Universidade Católica Portuguesa, Porto, Portugal

Resumo

A avaliação, bem como as eternas discussões em torno do seu papel no processo de ensino-aprendizagem, é uma componente incontornável da vida de alunos e professores. Ao longo dos tempos, a forma como a avaliação é encarada tem evoluído desde uma primeira “geração da medida” centrada na ideia positivista de mensurabilidade e quantificação até uma terceira “geração da avaliação como apreciação do mérito” ou mesmo uma quarta “geração integrativa” em que cada vez mais se encara a avaliação como um processo de promoção da aprendizagem. Esta investigação, apoiando-se em resultados de investigações anteriores que indicam serem as avaliações frequentes de baixo impacto (*low-stake quiz*) uma das mais eficazes ferramentas de consolidação do processo ensino-aprendizagem, pretende contribuir para a compreensão da influência nas aprendizagens da utilização frequente e repetitiva de micro-avaliações (MA) formativas com feedback contextual num contexto escolar real. Mais concretamente, pretende-se avaliar o real impacto da introdução desta técnica avaliativa nos resultados dos alunos bem como a perceção por parte de alunos, pais e professores quanto ao seu real efeito na eficácia da aprendizagem. A investigação iniciou-se no ano letivo 2018/2019 e decorrerá durante três anos letivos numa escola do ensino básico envolvendo 20 turmas do 5º ao 9º ano. A população envolvida será de cerca de 450 alunos e 50 professores por ano. Serão triangulados os resultados da análise quantitativa das classificações das diferentes tipologias de provas bem como dos resultados finais e

das avaliações externas com os resultados de uma análise qualitativa da percepção por parte de alunos, pais e professores sobre o efeito da introdução das MA na eficácia da aprendizagem.

Palavras-Chave: avaliação, avaliação formativa, micro-avaliações, eficácia do ensino-aprendizagem.

1. Contextualização

A educação, nas suas inúmeras formas, é uma constante das sociedades humanas desde tempos imemoriais. A educação formal e mais particularmente a escola, sendo uma construção muito mais recente das sociedades modernas, sofreu fortes evoluções, pressões e constrangimentos tendo atualmente grande dificuldade em cumprir eficazmente as elevadas expectativas construídas à sua volta. Apesar da investigação em educação já existir há mais de um século, nas últimas dezenas de anos ela foi particularmente profícua no que diz respeito à construção de saber alicerçado em evidências científicas sobre o processo de aprender eficazmente. Citando Brown et al. no prefácio de “Make it Stick” (Brown et al., 2014)

... a investigação sobre o processo de aprender e de reter informação demonstra-nos que muito do senso comum sobre como estudar com sucesso conduz-nos, em grande parte, a um desperdício de esforço! ... recentes investigações permitiram construir um conjunto de conhecimentos na área da ciência da aprendizagem resultando em propostas de novas estratégias altamente eficazes e capazes de substituir com comprovado sucesso práticas antigas e enraizadas”¹

É neste contexto de crescente cientificação na constante procura da eficácia das aprendizagens que, interligando saberes das áreas da psicologia através das teorias da aprendizagem e das neurociências, foram sendo analisados e testados inúmeros fatores capazes de potenciar a aprendizagem. De entre esses fatores, a avaliação é um dos assuntos incontornáveis quando se tenta estudar e melhorar o processo de ensino-aprendizagem. Tal como muitos outros fatores, também a forma como a avaliação é encarada tem evoluído (Alves e Cabral, 2015) desde uma primeira “geração da medida” centrada na ideia positivista de mensurabilidade e quantificação, passando por uma segunda geração que, embora não abandonando a

¹ Tradução adaptada (Brown et al., 2014)

base psicométrica da avaliação, procura antes comparar o desempenho dos alunos face a objetivos previamente definidos criando-se primeira vez a distinção entre os domínios cognitivo, sócio-afetivo e psicomotor (Bloom, 1956) até uma terceira “geração da avaliação como apreciação do mérito” em que pela primeira vez com Scriven (1966 p. 6:8) se discute a avaliação formativa versus sumativa bem como a utilização da avaliação como um processo de promoção da aprendizagem, terminando numa quarta “geração integrativa” em que cada vez mais se encara a avaliação como um processo de promoção da aprendizagem “adotando uma perspectiva construtivista, participativa e negocial do processo avaliativo” (Alves e Cabral, 2015 p. 636). Citando Fernandes (2014 p. 2)

... a avaliação das, e para as, aprendizagens dos alunos é um domínio de inquestionável relevância no processo pedagógico, intrinsecamente associado aos processos de ensino e de aprendizagem. Por isso, é necessário, por exemplo, investir na clarificação da sua conceptualização, dos seus propósitos, das suas utilizações assim como na sua forma e nos conteúdos e tarefas que a devem acompanhar.

Chegados aqui, parece-nos consensual que a variável *avaliação* é certamente, de entre as variáveis passíveis de serem controladas e manipuladas, uma das que adquire maior relevância quando se trata de potenciar a eficácia do processo ensino-aprendizagem. Citando novamente Fernandes (2014, p. 3:4),

através de uma meta-análise realizada a partir de 250 investigações, (Black e Wiliam, 1998) concluíram que: a) práticas sistemáticas de avaliação formativa estão claramente associadas a melhorias muito significativas das aprendizagens de todos os alunos; (...) O seminal trabalho destes autores mostra que muito se poderá fazer ao nível das avaliações internas para que os alunos aprendam mais e melhor e para que, conseqüentemente, os sistemas educativos integrem plenamente....

No referido trabalho de (Black e Wiliam, 1998) em que são analisadas inúmeras experiências científicas, emergem dois conceitos extremamente importantes para a eficácia da utilização da avaliação como ferramenta para a melhoria da aprendizagem: a frequência com que a avaliação é realizada e o feedback dado aos alunos. Ao longo do artigo e reportando-se a diferentes e variadas publicações, os autores demonstram o impacto da frequência das avaliações, citando Black e Wiliam (1998, p. 35)

(...) meta-análise de 40 estudos relevantes demonstrou que o desempenho melhorou com provas frequentes aumentando com o aumento da frequência até um certo nível, mas acima desse nível (cerca de 1 a 2 provas por semana) o desempenho voltou a diminuir.²

Em relação ao feedback, ele é apontado como o fator complementar da frequência para a eficácia da melhoria das aprendizagens. Salienta-se, no entanto, que a qualidade do feedback é apresentada como determinante no sucesso desta técnica para a prossecução dos referidos objetivos e citando novamente Black e Wiliam (1998, p. 36:37)

(...) a qualidade do feedback fornecido é uma característica fundamental para qualquer procedimento de avaliação formativa. (...) usando uma meta-análise de 58 experiências obtidas de 40 relatórios. Os efeitos do feedback foram reduzidos se os alunos tivessem acesso às respostas antes que o feedback fosse transmitido. Quando esse efeito foi permitido, foi a qualidade do feedback que foi a maior influência no desempenho. A instrução programada e os itens de avaliação de conclusão simples foram associados aos menores efeitos. O feedback foi mais eficaz quando foi concebido para estimular a correção de erros através de uma abordagem cuidadosa em relação à aprendizagem original relevante para a tarefa.³

Investigações mais recentes corroboram a linha de pensamento apresentada acrescentando algumas características interessantes. Assim, em Karpicke (2009) os autores reforçam a ideia de que a repetição da avaliação como forma de estudo tem um efeito bastante superior à repetição da leitura. Em McDaniel (2011) os autores conduziram duas experiências para avaliar se os “*low-stake quiz*” são passíveis de ser usados para promover a aprendizagem dos alunos, concluindo que este tipo de testes produz benefícios significativos na aprendizagem, com ganhos entre 13% e 25% no desempenho em futuros testes sumativos. Em Huelser (2012), é investigado

² (...) meta-analysis of 40 relevant studies showed that performance improved with frequent testing and increased with increased frequency up to a certain level, but that beyond that (somewhere beyond 1 and 2 tests pe week) it could decline again.

³ (...) the quality of the feedback provided is a key feature in any procedure for formative assessment. (...) using a meta-analysis of 58 experiments taken from 40 reports. Effects of feedback were reduced if students had access to the answers before the feedback was conveyed. When this effect had been allowed for, it was then the quality of the feedback which was the largest influence on performance. Programmed instruction and simple completion assessment items were associated with the smallest effects. Feedback was most effective when it was designed to stimulate correction of errors through a thoughtful approach to them in relation to the original learning relevant to the task.

o impacto do erro na aprendizagem concluindo que ser colocado perante um erro, desde que seja seguido por um feedback corretivo, resulta numa melhor retenção das respostas corretas do que simplesmente estudar as respostas corretas desde o início. Em Taylor (2010) é investigada a prática letiva intercalada⁴ caracterizada pelo princípio de, após uma dada aula, serem realizados um conjunto de problemas práticos retirados de aulas anteriores, de modo a que dois problemas do mesmo tipo não apareçam consecutivamente, exigindo assim que os alunos escolham a forma de resolução apropriada com base no próprio problema e não na forma de resolução que estão a trabalhar. Este tipo de prática letiva apresenta melhores resultados que a prática letiva em blocos em que cada assunto é ensinado e depois testado. Por último, num excelente artigo de Dunlosky et al. (2013), um conjunto de investigadores dos departamentos de psicologia e de neurociências de diferentes universidades norte-americanas elabora uma extensa investigação sobre a eficácia de diferentes técnicas de estudo. Os investigadores escolheram analisar profundamente 10 técnicas diferentes que foram selecionadas por serem as que mais vezes eram referidas pelos estudantes entrevistados como sendo as técnicas de estudo que usavam frequentemente e/ou porque eram técnicas de estudo já muito estudadas cientificamente em inúmeros estudos científicos anteriores. Para além de muitos e interessantes parâmetros de análise, os investigadores classificam as 10 técnicas num ranking global de “overall utility” (Dunlosky et al., 2013 p. 45) tendo apenas duas dessas técnicas atingido a classificação máxima de “High” e sendo uma delas a “Practice Testing”, aqui definida como testes “low-stake or no-stake” (Dunlosky et al., 2013 p. 29).

2. Problema de investigação, questões e objetivos

Os resultados de diferentes investigações apontam consistentemente para um afastamento entre a eficácia percebida pelos estudantes (e pelos próprios professores) sobre as técnicas de estudo que utilizam e a eficácia medida experimentalmente (Dunlosky et al., 2013 p. 45). Esta conclusão abre portas, em teoria, à real possibilidade de melhoria da eficácia do ensino-aprendizagem através da introdução de novos mecanismos pedagógicos. Por outro lado, a clássica avaliação sumativa de caráter eminentemente classificativo embora cumpra o

⁴ Interleaved Practice

referido papel classificativo é muito pouco eficaz para o processo ensino-aprendizagem.

Citando Dunlosky et al. (2013 p. 45:46),

poucas técnicas foram avaliadas em contextos educativos representativos. Essa avaliação (...) sugere duas direções para investigações futuras (...) Em primeiro lugar (...) Em segundo lugar, o benefício da maioria das técnicas em ambientes educacionais representativos precisa ser mais amplamente explorado. Versões fáceis de usar das técnicas mais promissoras devem ser desenvolvidas e avaliadas em investigações controladas conduzidas em contextos educacionalmente representativos. Idealmente, as medidas de aferição incluiriam testes de alta relevância, como desempenho em exames nacionais e em testes sumativos. Percebemos que tais esforços de investigação podem ser demorados e dispendiosos, mas realizá-los será crucial para recomendar mudanças educacionais que terão uma probabilidade razoável de melhorar a eficácia do processo ensino-aprendizagem do aluno.⁵

Procurando responder ao desafio acima enunciado, o problema central desta investigação será: Qual o real impacto nas aprendizagens dos alunos da introdução de micro-avaliações (*low-stake quiz with frequent feedback*) no processo pedagógico de uma escola.

A partir deste problema central, gostaríamos ainda de responder às três seguintes subquestões de investigação: Qual a percepção criada em alunos, pais e professores pela realização de muitas dezenas de micro-avaliações ao longo do ano; se o impacto na aprendizagem é dependente de fatores como o tipo de currículo da disciplina, o nível acadêmico do aluno e o perfil do aluno (idade e atitude perante o estudo); qual o impacto na aprendizagem de fatores característicos das micro-avaliações tais como frequência, tipo de feedback e nível na taxonomia de Bloom.

⁵ “(...) few techniques have been evaluated in representative educational contexts. This appraisal (...) suggests two directions for future research (...) First, (...) Second, the benefit of most of the techniques in representative educational settings needs to be more fully explored. Easy-to-use versions of the most promising techniques should be developed and evaluated in controlled investigations conducted in educationally representative contexts. Ideally, the criterion measures would include highstakes tests, such as performance on in-class exams and on achievement tests. We realize that such research efforts can be time-consuming and costly, but conducting them will be crucial for recommending educational changes that will have a reasonable likelihood of improving student learning and achievement.”

Procurar-se-á que esta investigação cumpra os critérios de “*ambiente educacional representativo*” envolvendo um elevado número de alunos em situação real de trabalho pedagógico e de avaliação. Os trabalhos decorrerão ao longo de três anos letivos numa escola do ensino básico privado envolvendo 20 turmas do 5º ao 9º ano, ou seja, uma população de cerca de 480 alunos (486 em 2018/19) e 50 professores por ano, isto é, um total de cerca de 1440 alunos e 150 professores no final dos 3 anos.

Se a hipótese colocada à partida se confirmar de que a influencia das micro-avaliações é claramente positiva nas reais aprendizagens dos alunos, um último e ambicioso objetivo será equacionado: tenter criar um conjunto de recomendações capazes de facilitar a introdução desta ferramenta (MA) noutros ambientes escolares.

3. Descrição detalhada do projeto

O projeto de investigação em curso tem como objetivo principal responder ao seguinte problema central desta investigação: Qual o real impacto nas aprendizagens dos alunos da introdução de micro-avaliações (low-stake quiz with frequent feedback) no processo pedagógico de uma escola. Recordando Dunlosky et al. (2013 p. 45:46) procuraremos criar um ecossistema de investigação que permita a designação de “*ambiente educacional representativo*” por forma a não excluir a possibilidade de, validadas as hipóteses, criar um conjunto de recomendações.

O modelo de investigação será misto, embora com uma preponderância claramente quantitativa.

A recolha de dados quantitativos e posterior tratamento estatístico ocorrerá ao longo de três anos letivos (até 2020/21), tendo já começado no final de 2018/19 com os dados finais do ano letivo 2018/19. A título de curiosidade, já foram recolhidos todos os momentos de avaliação nos diferentes departamentos curriculares bem como os respetivos resultados correspondentes a todos os alunos do 5º ano até ao 9º ano de escolaridade nas disciplinas em que se aplicam MA (português, matemática, inglês, francês, história, geografia, ciências da natureza e físico-química). Os números totais de dados para 2018/19 são os seguintes: número total de alunos são 486; Número total de elementos de avaliação diferentes excetuando

as MA é 529, ou seja cerca de (486 alunos /20 turmas * 529 avaliações) 12850 classificações; Número total de elementos de avaliação correspondentes à designação teste sumativo ou equivalente é 375, ou seja cerca de 9112 classificações; Número de total de MA diferentes efetuadas pelos alunos foram cerca de 1180, ou seja cerca de 28600 classificações.

Por forma a determinar o real impacto da varável MA iremos usar como provas de aferição os testes sumativos para os alunos do 5º ao 8º ano e os testes sumativos e as avaliações externas para os alunos do 9º ano. É assim possível estabelecer um termo de comparação entre o antes e o depois da introdução das MA.

A implementação das MA neste primeiro ano letivo sofreu de um problema logístico sério já que o programa informático que deveria ter sido usado pelos professores não se revelou com todas as características necessárias (não fornecia feedback contextual, por exemplo) ao modelo conceptual das MA. Durante o ano de 2018/19 alguns professores usaram programas públicos complementando-os com feedback “manual” e outros optaram pelas micro-avaliações em papel. Está já a decorrer uma implementação informática capaz de resolver todos os problemas detetados, desenvolvida pelo autor que, caso seja satisfatoriamente concluída em tempo útil para a tese, será apresentada sobre a forma de anexo a esta tese.

Para que possam ser analisadas as questões de investigação 2 e 3, será naturalmente necessário cruzar dados de alunos, sempre de uma forma anónima, com os dados pessoais característicos do aluno. Será ainda necessário que a aplicação informática desenvolvida, permita a calibração dos parâmetros em causa no estudo.

Para a recolha de dados qualitativos, vamos empregar duas técnicas distintas: modelo de inquérito online para obter informação de todos os alunos e respetivos pais quanto à perceção sobre a validade desta medida pedagógica. A população será, portanto, todo o universo de alunos e suas famílias que frequentam a escola do 5º ao 9º ano. Para a recolha da perceção dos professores iremos conduzir igualmente inquéritos online, mas que depois serão complementados com entrevistas a uma amostra de 20 professores. Estas entrevistas serão alvo de uma análise de conteúdo utilizando o software MAXQDA.

A forma como no conselho pedagógico da escola ficou decidida a introdução das MA bem como a sua articulação com os restantes elementos avaliativos foi a seguinte:

Cada departamento curricular decidiu os momentos do ano bem como a sequência em que as MA eram realizadas com respeito às restantes avaliações.

Todas as disciplinas com MA deveriam realizar 12 MA durante o ano, ou seja, 6 em cada semestre.

Para cada conjunto de 6 MA, e por forma a acentuar o seu carácter formativo (low-stake) são eliminadas as duas MA com pior classificação. Para evitar que os alunos não atribuísem a devida importância a este instrumento de avaliação, foi ainda decidido que cada conjunto das restantes 4 MA entraria na equação final da nota com o peso equivalente de um teste sumativo.

4. Referências bibliográficas

- Alves, José & Cabral, Ilídia. (2015). Os demónios da avaliação: memórias de professores enquanto alunos. *Estudos em Avaliação Educacional*. 26. 630.10.18222/ae.v26i63.3690.
- Brown, P. C., Roediger, H. L. & McDaniel, M. A. (2014). *Make it Stick: The Science of Successful Learning*, The Belknap Press of Harvard University Press - ISBN 978-0-674-72901-8
- Black, Paul and Wiliam, Dylan (1998) 'Assessment and Classroom Learning', *Assessment in Education: Principles, Policy & Practice*, 5:1, 7 — 74
- Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. New York: David McKay Company.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Fernandes, D. (2014). Avaliação das Aprendizagens e Políticas Educativas: O difícil percurso da inclusão e da melhoria. In M. L. Rodrigues (Org.), *Quarenta anos de políticas de educação em Portugal: a construção do sistema democrático de ensino* (Volume I), pp. 231-268. Coimbra: Almedina.
- Huelser, B.J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, 40, 514-527.
- Karpicke, Jeffrey D., Butler, Andrew C. and Roediger III, Henry L.(2009) 'Metacognitive strategies in student learning: Do students practise retrieval when they study on their own?', *Memory*,17:4,471 — 479
- McDaniel, Mark A., Agarwal, Pooja K. and Huelser, Barbie J. (2011) " Test-Enhanced Learning in a Middle School Science Classroom: The Effects of Quiz Frequency and Placement" (2011)

Journal of Educational Psychology © 2011 American Psychological Association, Vol. 103, No. 2, 399–414

Nóvoa, António (2009) "Educação 2021: Para uma história do futuro", *Revista Iberoamericana de Educación*, n.º 49, Janeiro-Abril, 2009. pp.181-199) ISSN 1681-5653) Retirado de <http://hdl.handle.net/10451/670>

Scriven, Michael (1966). *The Methodology of Evaluation*. Lafayette, Ind.: Purdue University.

Taylor, Kelli e Rohrer, Doug (2010) The Effects of Interleaved Practice. *Applied Cognitive Psychology* 24: 837–848