

ENSINO DE CONTEÚDOS ESCOLARES: A AVALIAÇÃO COMO FATOR ESTRUTURANTE¹

Leonor Santos; Jorge Pinto

Instituto de Educação da Universidade de Lisboa; Escola Superior de Educação do Instituto Politécnico de Setúbal

mlsantos@ie.ul.pt; jorge.pinto@ese.ips.pt

Sumário

Falar de avaliação das aprendizagens pode querer significar diversos entendimentos, pelo que este capítulo começa por clarificar o que se entende por avaliar, abordando, embora de forma não aprofundada, questões relativas aos fundamentos da avaliação. Dado o processo avaliativo conter uma tomada de decisão, é essencial ter em conta que esta resulta de uma comparação entre o produzido e o esperado, pelo que é indispensável a apropriação, por parte dos diversos envolvidos no processo avaliativo, dos critérios de avaliação. As secções seguintes abordam, respetivamente, aspetos relativos às duas modalidades de avaliação: a avaliação formativa e sumativa. Tendo por base resultados obtidos através de estudos empíricos, sobretudo desenvolvidos no trabalho quotidiano com os alunos na sala de aula, na segunda secção, são discutidas algumas estratégias avaliativas possíveis, como sejam a apropriação dos critérios de avaliação, o feedback e as questões de aula. Procura-se salientar, igualmente, as potencialidades e os desafios que se colocam, quer ao professor, quer aos alunos, no desenvolvimento dessas estratégias. Na terceira secção, são apresentadas algumas questões relativas a processos inerentes à avaliação enquanto medida. Pistas para a construção de instrumento de avaliação sumativa, bem como procedimentos para garantir a sua qualidade são discutidos. O capítulo termina com a proposta de algumas tarefas a realizar, preferencialmente em pequenos grupos de professores, que permitem pôr em uso alguns dos assuntos abordados, assim como são apresentadas algumas referências cuja leitura permite o seu aprofundamento a par com aquelas que foram sendo apresentadas ao longo do capítulo.

Palavras-chave: Avaliação, aprendizagem, funções da avaliação, avaliação sumativa, avaliação formativa, critérios de avaliação, feedback, questões de aula, prova, item, validade, fidelidade, objetividade, avaliação como medida, avaliação como construção social.

CONCEPTUALIZAÇÃO

1.1 O significado do conceito de avaliação

Desde o início do séc. XX até aos anos 60, a avaliação em educação estava praticamente circunscrita à avaliação das aprendizagens dos alunos. Na tradição da docimologia (Pierron, 1963), procurava-se essencialmente a construção de instrumentos de medida das aprendizagens que oferecessem um elevado grau de validade, medissem o que se queria medir, e de fidelidade, reduzissem a discrepância

¹ Santos, L., & Pinto, J. (2018). Ensino de conteúdos escolares: A avaliação como Fator estruturante. In F. Veiga (Coord.), *O Ensino como fator de envolvimento numa escola para todos* (pp. 503-539). Lisboa: Climepsi Editores.

entre avaliadores. Deste modo, a avaliação era tratada como uma medida ainda que desde há muito se lhe reconhecesse um baixo grau de rigor (Noizet & Caverni, 1985).

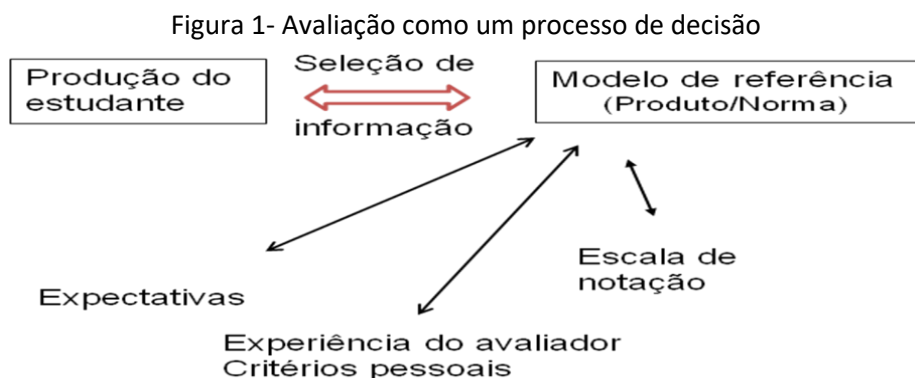
Não é assim de estranhar que, quando se fala de avaliação no contexto da Escola, existam duas ideias muito persistentes: uma que associa a avaliação à nota/classificação; outra aos testes ou exames, ou seja, aos instrumentos de recolha de informação. Estas traduzem uma visão muito limitada e redutora da avaliação pois, de acordo com os trabalhos de investigação da docimologia experimental (Noizet & Caverni, 1985), a avaliação não é uma medida, mas uma construção social sobre um desempenho (Pinto, 2002; Pinto & Santos, 2006) e os instrumentos e procedimentos utilizados, para além de nada medirem (Merle, 1996; Seibel & Levasseur, 2007), são muito redutores, pois excluem um conjunto de tarefas que os alunos têm de realizar no quotidiano, no cumprimento dos diversos programas. Uma outra consequência destas perspetivas é a separação da avaliação do currículo, concebendo-a como uma técnica neutra. Justapondo-se ao currículo, a avaliação tanto serve para fiscalizar à distância como decorre o processo de ensino- aprendizagem (muitas vezes designada de monitorização), como para medir os resultados em termos de aprendizagem de um currículo.

Assumida a avaliação como um processo neutro, que não interfere de forma direta com o ensino e a aprendizagem, ela pode ser objetiva se forem verificadas um conjunto de normas, onde as mais comuns parecem ser aquelas onde o avaliador não tem grande margem de interpretação. Tal pressuposto leva a configurar os testes ou exames como conjuntos de questões de resposta múltipla, ou ao desenvolvimento de normas de correção, procurando com estas anular as interpretações dos avaliadores e/ou dos erros existentes. Todos estes procedimentos são hoje rebatidos ao pôr em evidência os efeitos destas estratégias (Noizet & Caverni, 1985; Merle, 1996). Apesar de este debate sobre a objetividade/subjetividade ser antigo e poder ser levado até à exaustão, na medida em que nem tudo pode ser explicitado, também é claro que quanto mais explicitação interpessoal houver, mais objetivo se pode tornar o processo de avaliação. A objetividade nunca é uma garantia de partida, mas algo de constatável à chegada e será tanto mais quanto trabalhada for a intersubjetividade dos avaliadores, ou seja, a explicitação/consensualização das diferentes visões que cada

avaliador tem sobre os diversos aspetos do processo de avaliação, nomeadamente dos critérios de avaliação e da sua relação com as produções dos alunos.

A avaliação é uma forma particular de abordar, conhecer e compreender um determinado fenómeno, neste caso educacional (Mateo, 2000). É uma forma singular de relação com certos fenómenos em função de um determinado propósito, pessoal ou social. Esta relação passa essencialmente pela recolha, análise e interpretação de dados para uma tomada de decisão sobre o valor desses dados, tendo em conta a razão de ser da avaliação e das suas finalidades.

O desenvolvimento da psicologia cognitiva e do estudo dos processos de tomada de decisão permitiram encarar a natureza do processo avaliativo como uma tomada de decisão. A constatação de que, ao contrário de um processo de medida tal como acontece em Física, não temos, nem um objeto a medir estável, nem tão pouco um instrumento padronizado de medida (como o metro ou o kilo). Assim, passa a tornar-se insustentável falar de avaliação das aprendizagens como medida, muito embora o uso do termo tenha muitas vezes essa conotação. Mas se não é uma medida, então o que será? Ora é nesta perspetiva que Noizet e Caverni (1985) a estudam como um processo de tomada de decisão. De acordo com estes autores, a avaliação é um processo de vai e vem de seleção e de produção de informação, de evidências, entre a produção do estudante em resposta a um pedido e o modelo de referência do professor (Figura 1).



Este modelo é um conjunto interligado de conhecimentos de que o professor dispõe: (i) as normas a seguir para “essa” avaliação; (ii) os critérios de avaliação a usar “nessa” avaliação; e (iii) a sua representação sobre o que é avaliar aprendizagens. Para além disto, este modelo de referência é fortemente influenciado pelas expectativas que

todo o professor tem acerca dos seus alunos (de um “bom” ou “mau” aluno esperam-se desempenhos diferentes), pela sua experiência enquanto avaliador (o que considera certo ou errado, por exemplo varia com a sua experiência) e ainda pelo sistema de notação usado (uma escala discreta de níveis, por ex. 1 a 5, é diferente de uma escala contínua, por ex. 0 a 20). A decisão decorre justamente do confronto entre as evidências recolhidas e percebidas pelo professor com o seu próprio modelo de referência.

Assim, ao podermos afirmar que a avaliação é um processo de tomada de decisão, ela é também um processo de diálogo entre o produto do aluno (produtor) e o modelo de referência (avaliador). Desta forma, longe de ser uma medida rigorosa e neutra, à prova de pessoas, a avaliação é uma construção social localizada (Chevalard, 1990; Pinto, 2002; Vial, 2012) onde a comunicação desempenha um elemento fundamental.

1.2. A comunicação em avaliação e a relevância dos critérios de avaliação

Um processo de comunicação exige que ambos os intervenientes partilhem um mesmo código linguístico e sistema de representações de modo a que se possam entender. Ora, na educação, esta comunicação é muitas vezes assimétrica em termos de poder. O professor, ao deter o saber ou a responsabilidade institucional de ensinar, tem mais poder do que o estudante. Em termos de avaliação, este facto coloca-se com mais premência uma vez que o saber só se revela através de um desempenho. Assim, o professor quando quer avaliar o saber de um aluno pede-lhe que execute uma certa tarefa. Mas pode, ou não, explicitar o modo como deve ser realizada bem como os requisitos que deve ter para ser considerada bem realizada, isto é, pode revelar ou não as suas expectativas em relação à tarefa. Estas expectativas constituem o que se designa por critérios de avaliação. A sua explicitação e clareza são fatores decisivos para que o estudante perceba o pedido do professor e o execute da melhor forma que souber. Assim, quanto mais explícito for o pedido do professor e os seus critérios, mais fácil será interpretar o pedido e construir uma resposta adequada. Deste modo, os critérios de avaliação adquirem uma grande relevância no processo de avaliação (Perrenoud, 1999). A clarificação dos pedidos avaliativos, através da explicitação dos critérios, é uma peça fundamental na avaliação entendida como uma construção social.

1.3 A avaliação segundo os seus propósitos

No contexto educativo é frequente falar-se de avaliação sumativa e de avaliação formativa como sendo dois tipos muito diferentes de avaliação. Normalmente, esta ideia está baseada no tipo de instrumentos usados, bem como, no rigor da sumativa face à formativa, pois na primeira utilizam-se processos formais, normalmente testes, que assumem muitas vezes outras designações, que permitem obter classificações. Já a formativa recorre frequentemente a processos informais e o professor não dispõe de informação igual para todos os alunos. Em contraponto, podemos encontrar outra perspetiva que considera que tudo é mais ou menos a mesma coisa, pois os testes também dão indicações ao professor e ao aluno sobre as suas aquisições e podem constituir um fator de motivação e de aprendizagem para o aluno. Ora ambas as posições não contribuem para uma clarificação, quer conceptual, quer das práticas de avaliação.

O núcleo central da distinção entre a avaliação formativa e sumativa está no seu propósito, no uso a dar às evidências recolhidas:

A mesma informação, recolhida do mesmo modo, chamar-se-á formativa se for usada para apoiar a aprendizagem e o ensino, ou sumativa se não for utilizada deste modo, mas apenas para registar e reportar. (Harlen, 2005, p. 208)

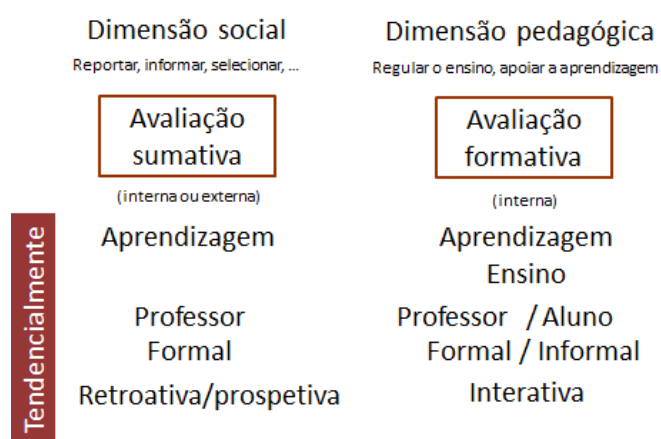
Qualquer avaliação é meramente um meio de recolher evidência sobre a aprendizagem. É o uso dado a essa evidência que distingue o formativo do sumativo, embora naturalmente os métodos usados e as interpretações dos resultados possam diferir de acordo com os usos pretendidos. (Black, 2013, p. 170)

Por outras palavras, as finalidades da avaliação são diversas, mas podem ser agrupadas em dois grandes tipos (Weiss, 1996). Um deles contempla finalidades de natureza administrativa, isto é, de organização e regulação administrativa do sistema educativo, o outro inclui funções de natureza pedagógica, ou seja, uma função de regulação pedagógica do sistema educativo, do currículo e das aprendizagens. O primeiro caracteriza-se essencialmente por funções de controlo do ensino e da aprendizagem, através dos desempenhos escolares e da decisão se tais desempenhos são os suficientes em termos de transição de ano ou da opção por outros percursos educativos. O segundo engloba funções de regulação dos suportes/apoios ao ensino e

à aprendizagem dos alunos. Em síntese, numa primeira leitura, podemos dizer que o processo nuclear da avaliação é o mesmo, contudo, as funções a que a avaliação se destina levam a que as suas práticas sejam distintas (Figura 2). Sendo o tipo de decisões diferentes, a informação que se recolhe, como se recolhe e a sua interpretação, e a ação subsequente tendem também a ser diferentes. Estamos, assim, perante dois grandes modos não só de conceber, mas também de desenvolver a avaliação. Embora o processo em termos cognitivos seja o mesmo, as suas metodologias na ação são diferentes.

Assim, entende-se por avaliação sumativa, também por vezes designada por avaliação das aprendizagens (Black, Harrison, Lee, Marshall, & Wiliam, 2003), um processo no qual as evidências são usadas para inventariar as aquisições dos estudantes num certo momento da sua aprendizagem (reportar, informar, ...) e tomar decisões em conformidade com o estabelecido administrativamente (hierarquizar, selecionar, ...). A avaliação formativa, ou avaliação para as aprendizagens, embora seja uma designação abrangente, pois cobre um leque variado de práticas, usa as evidências para perceber onde o aluno está em termos de aprendizagem para tomar decisões no sentido de providenciar mais e melhores aprendizagens e para regular o ensino (Figura 2).

Figura 2. Características da avaliação sumativa e formativa



É ao professor que cabe a responsabilidade da avaliação sumativa, enquanto a avaliação formativa pode ser da responsabilidade do professor, do aluno ou de uma responsabilidade partilhada. Tendencialmente, o momento em que é desenvolvida também é distinto. Enquanto a avaliação sumativa acontece no final de um período de

ensino e aprendizagem, podendo prever as possibilidades de futuro escolar do aluno, a formativa acontece preferencialmente durante e a par com o processo de ensino e aprendizagem.

A avaliação diagnóstica, por vezes considerada como uma outra modalidade de avaliação, enquadra-se no universo da avaliação formativa, uma vez que tem os mesmos propósitos e as mesmas finalidades que a avaliação formativa. Como o próprio nome indica e fazendo o paralelo com a medicina, o diagnóstico serve, no caso de confirmação de doença, para permitir decidir uma terapêutica que resolva o que foi diagnosticado. Na educação, a função é semelhante: procura-se perceber o estado dos alunos em função de um certo caminho de aprendizagem a percorrer para adequar as melhores estratégias para assegurar esse percurso sem sobressaltos. Todavia, toda a avaliação formativa é de diagnóstico pois o seu propósito é ajudar a perceber o ponto em que cada aluno está relativamente ao ponto em que era suposto estar e encontrar as melhores estratégias para superar essas dificuldades e entrar de novo no tilho da aprendizagem (William, 2007).

Também em relação ao estatuto do erro ou dificuldades do aluno existem interpretações diferentes, conforme se está num ou noutro registo. Na avaliação sumativa, o erro é visto como uma falta de saber. Chega-se às classificações muitas vezes através do somatório dos erros dados. Estes transformam-se facilmente num mero elemento contabilístico (Santos, 2008). Num registo formativo, o erro é visto como a manifestação de uma dificuldade na construção de um saber, como um elemento que permite ao estudante e professor identificar essa dificuldade e encontrar o melhor caminho para a superar (Jorro, 2000).

PRÁTICAS DE AVALIAÇÃO FORMATIVA

Existe um largo leque de estratégias de avaliação formativa que podem ser desenvolvidas de forma que a avaliação esteja ao serviço da aprendizagem. Nesta secção iremos abordar, a título de exemplo, três destas estratégias: o uso partilhado de critérios de avaliação; o feedback dado pelo professor; e as questões de aula. A razão da nossa escolha fundamenta-se nos três processos chave do ensino e aprendizagem referidos por Black e Wiliam (2009):

- Estabelecer onde o aluno se situa na sua aprendizagem;
- Estabelecer onde se dirigir;
- Estabelecer o que é necessário fazer para lá chegar.

É de fazer notar que, contrariando uma visão de ensino mais tradicional, estes autores apelam a que estes processos chave devem contar com o papel ativo dos diversos intervenientes em presença: o professor, os pares e o aprendiz. No entanto, é o professor que tem a principal responsabilidade de orientar todo o processo, monitorizando o que se faz e como se faz para o desenvolvimento destes processos chave.

2.1 Critérios de avaliação

Num registo de avaliação formativa faz sobretudo sentido falar nos critérios de realização (Bonniol & Vial, 1997). São os critérios que pretendem orientar o aluno indicando-lhe aquilo que é necessário desenvolver para a realização de uma dada tarefa. Geralmente definidos pelo professor e dependentes do conhecimento que este tem do currículo e das suas conceções sobre o ensino e a aprendizagem (Semana & Santos, 2013), indicam os atos concretos que esperam que os alunos façam a partir do que lhes é pedido para fazer.

Quanto melhor os alunos compreenderem o que se espera deles, mais predispostos estarão para aprender e, conseqüentemente, aumentará a probabilidade de acontecer aprendizagem. Um estudo desenvolvido, em Portugal, com alunos do 1º ano de escolaridade, evidenciou que os alunos foram aperfeiçoando a sua compreensão e capacidade de resolver problemas em Matemática à medida que construíram e usaram os critérios de avaliação (Peres, 2012). A forma como os alunos se envolveram desde o primeiro momento no processo de definição dos critérios de avaliação parece ser uma razão forte para explicar o sucesso desta experiência pedagógica (Pinto & Santos, 2012).

Um outro estudo, desenvolvido no 2º ciclo de escolaridade, aponta para uma evolução positiva por parte dos alunos da compreensão do que é resolver tarefas exploratórias ou de investigação em Matemática. Os alunos foram-se apropriando de nova terminologia em Matemática e realizando atividades matemáticas novas para si

(Beirão, 2012). Um outro estudo, com uma turma do 8º ano ainda a Matemática (Semana & Santos, 2013), evidenciou que o investimento do professor feito através do uso de critérios de avaliação, acompanhados por uma tabela de descritores, embora levante desafios ao professor, revela-se contributivo para a sua apropriação por parte dos alunos. Ajuda-os a clarificar conceitos e procedimentos matemáticos, estratégias e processos de raciocínio e a desenvolver a capacidade de comunicação matemática (Santos & Semana, 2015).

Contudo, o processo de apropriação (conhecimento profundo) dos critérios de avaliação não é simples nem linear. Muito embora todo o professor tenha os seus critérios de avaliação, muitas vezes não é para si fácil explicitá-los. Mas não basta informar os alunos dos critérios de avaliação para que de imediato eles passem a ser por si apropriados. Os critérios valorizados, ainda que explicitados, não têm necessariamente um significado igual para todos aqueles a quem são apresentados (Pinto, 2002). Estudos evidenciam a existência de padrões autoimpostos, que regulam a atividade dos alunos, constituindo-se gestores de referência para a atividade em curso e para o desenvolvimento de estratégias de verificação e de correção (Santos & Gomes, 2006). Os alunos podem retroceder quando influenciados por outros aspetos, como seja aquilo que o professor acabou de tratar, levando-os a pensar que a tarefa seguinte só pode ser de aplicação da matéria acabada de dar (Küchemann & Hoyles, 2006). É no trabalho continuado com esses critérios que a compreensão sobre eles se vai desenvolvendo. Este facto é tão mais importante quanto se tem presente que existe uma forte relação entre a compreensão e a valorização dos critérios (Bruno, 2013).

Cabe ao professor desenvolver um conjunto de práticas intencionalmente pensadas para facilitar a apropriação, por parte dos alunos dos critérios de avaliação (Santos, 2008). Naturalmente, em primeiro lugar, o professor deve explicitá-los para si. Em seguida, estes devem ser comunicados e/ou discutidos e negociados com os alunos. Os alunos devem ser encorajados a discuti-los entre si e com o professor. Para a construção de um significado comum, é vantajoso proporcionar aos alunos a possibilidade de confrontarem estes critérios no contexto do seu próprio trabalho. Deste modo, o recurso a exemplos ilustrativos de qualidade variável de trabalhos

realizados por outros alunos poderá ser uma possível estratégia facilitadora para a compreensão do que se está a discutir. A clarificação de uma ideia passa muitas vezes pela sua concretização, através de um exemplo concreto. Posteriormente, dar feedback aos trabalhos produzidos numa primeira fase, tendo em conta os critérios acordados, é outra oportunidade a não perder. Recorrer à coavaliação entre pares e à autoavaliação, confrontando as apreciações feitas por outros alunos com as suas próprias poderão ser outras estratégias possíveis para ajudar à apropriação dos critérios de avaliação por parte dos alunos. É assim um processo que se inicia antes da ação e se vai progressivamente construindo ao longo do tempo e a partir de diversas experiências de aprendizagem. A aprendizagem e a compreensão dos critérios de avaliação vão-se desenvolvendo par a par, contribuindo cada um deles para o desenvolvimento do outro (Santos & Gomes, 2006).

Assim, estratégias que facilitem a apropriação dos critérios de avaliação por parte dos alunos podem passar pelo desenvolvimento de coavaliações dentro do grupo, pelos feedbacks do professor, pela confrontação efetuada na sala de aula, pelas oportunidades de melhoramento de produções realizadas e pelas próprias autoavaliações. A realização de forma continuada de relatórios escritos, apoiados por um guião e critérios de avaliação negociados com os alunos de uma turma do 8.º ano em Matemática, levou-os a ajustarem as suas representações iniciais e a autocontrolarem os seus padrões iniciais (Semana & Santos, 2009). Do exposto emerge que a apropriação, por parte dos alunos, dos critérios de avaliação mais do que depender de uma ou outra estratégia pedagógica desenvolvida pelo professor, resulta de uma ação combinada de diversas estratégias (Bruno, 2013).

2.2. Feedback

O que se entende por feedback não é consensual entre os diversos autores. Se o foco for o conteúdo do feedback, poder-se-á dizer que o feedback é todo o comentário avaliativo que informa sobre a qualidade daquilo que foi feito. Ou, se for adicionada uma intencionalidade formativa, é toda a informação que permite ainda ao próprio identificar o que lhe falta fazer para atingir o que era esperado que fizesse. Caso o foco sejam os efeitos do feedback, também aqui podemos considerar duas perspetivas: ou o feedback é entendido como todo o comentário que procura que o próprio melhore

ou o comentário que leva efetivamente a reduzir a diferença entre o que se fez e o que se esperava que fosse feito. Vamos considerar, neste texto que o feedback é toda a informação produzida de forma intencional para ajudar o aluno a melhorar o seu desempenho (mesmo que efetivamente não consiga fazê-lo).

O feedback é um elemento chave na avaliação formativa (Sadler, 1998), talvez mesmo o mais poderoso mediador para melhorar o desempenho. Dado que pode ser produzido oralmente ou por escrito, iremos nas páginas seguintes tratar separadamente estas duas situações.

2.2.1 Questionamento oral

A comunicação na sala de aula concretizada através da interação que se estabelece entre os diversos sujeitos presentes, professor e alunos, acontece independentemente do método de ensino usado. Olhamos com naturalidade para essa comunicação, dada a sua existência quase permanente. Contudo, raramente é analisada do ponto de vista de uma estratégia avaliativa contributiva para a aprendizagem.

Muitos têm sido os referenciais teóricos criados para analisar a comunicação na sala de aula. Por exemplo, Gipps (1999) foca-se na natureza das questões colocadas pelo professor (perguntas fechadas, perguntas abertas) e alerta-nos para o risco das perguntas fechadas, nomeadamente as específicas de diagnóstico, quando repetidas poderem levar os alunos a mudar rapidamente de opinião. Os alunos vão lançando respostas, mais ao acaso do que baseadas em algum tipo de raciocínio, esperando acertar. Segundo esta autora, perguntas que incentivem a reflexão são potenciadoras de aprendizagem, como sejam, O que fizeste?; Por que pensaste assim?; De onde te surgiu a ideia?; Em que outras situações esta estratégia também se poderia aplicar?; Se quisesses convencer alguém, o que dirias?

Naturalmente que a comunicação não passa apenas pela formulação de questões. Outras dimensões podem ser acrescentadas na análise da comunicação na sala de aula. Santos & Pinto (2008) criaram uma grelha de análise das interações avaliativas na sala de aula que inclui três dimensões: a dinâmica da interação (quem a produz e a quem se dirige), o foco (a que zona da atividade se centra), e o significado (sentido pedagógico contextualizado) (Figura 3).

Fig. 3. - Grelha de análise da comunicação na sala de aula (Santos & Pinto, 2008)

Dinâmica			
Quem a produz	Professor (P)	Aluno (A)	Grupo de alunos (As)
A quem se dirige	Professor (P)	Aluno (A)	Grupo de alunos (As)

Foco	Conceptualização (C)	Processo (Pr)	Produto (Pd)	Gestão da sala de aula (G)
-------------	----------------------	---------------	--------------	----------------------------

Significado	Questionar (Q)	Responder (R)	Explicar (E)
--------------------	----------------	---------------	--------------

O questionar compreende várias subcategorias: pedir um resultado (Qres), pedir uma justificação (Qjust), colocar uma questão que reorienta a linha de raciocínio (Qrac), e remeter a validação para outros (Qval).

O responder compreende várias subcategorias: repetir (Rrep), resolver (Rsol), corrigir (Rc), validar (Rval), e justificar (Rjust).

O explicar integra uma descrição total (Et) ou parcial (Ep).

Num estudo realizado com uma professora do 8º ano de Matemática (Semana & Santos, 2012), os resultados obtidos a partir do uso desta grelha de análise evidenciam que foi possível identificar um padrão geral nas intervenções da professora por cada fase de exploração de uma tarefa. Na fase de lançamento, a intervenção dominante é o da professora e o seu foco recai essencialmente na gestão da sala de aula. Na fase de desenvolvimento, as intervenções com foco no processo e na conceptualização remetem principalmente para a justificação/explicação e as intervenções focadas no produto tendem a estar associadas ao resolver. Por último, na fase de discussão, a professora reassume um papel mais interventivo na interação com os alunos. Assim, é possível afirmar-se que o uso da grelha permitiu evidenciar que o papel da professora não é o mesmo ao longo da exploração da tarefa. Este resultado vem na linha do resultado obtido num outro estudo, desenvolvido em Inglaterra (Hodgen, 2007). Uma professora do 8º ano, também de Matemática, considerada pelos seus pares como competente na prática de avaliação formativa, na fase de desenvolvimento da tarefa, escuta os alunos e estimula a colaboração e partilha entre eles, já na fase de discussão deixa pouco espaço de liberdade para os alunos comentarem e desenvolverem as suas próprias ideias matemáticas.

Trabalhar o questionamento oral de modo a torná-lo potenciador de aprendizagem é uma tarefa desafiante para o professor. Num projeto da responsabilidade do Assessment Group do King's College de Londres (Black *et al.*, 2003), o questionamento foi objeto de atenção junto de dois professores dos 7º, 8º e 10º anos de Matemática e dois de Ciências de cada uma das seis escolas participantes no projeto. Aumentar o

tempo que permeia a colocação de uma pergunta e a resposta que o próprio professor dá foi um problema difícil de ultrapassar. Contudo, ao fim de um ano de trabalho, os professores aumentaram o seu tempo de espera, muito embora considerassem que esta mudança de prática tenha sido “dolorosa” e a existência de “momentos mortos” lhes ter parecido antinaturais. Estes resultados vão na mesma linha dos obtidos por Pinto & Santos (2010) que contabilizaram 642 intervenções do professor numa aula de 90 minutos, o que corresponde a uma média de 7,1 intervenções por minuto, sendo a duração média de cada uma de 8,4 segundos. Aceder à transcrição da aula, permitiu ao professor tomar consciência da extensão do seu tempo de intervenção, o que o levou a sentir a necessidade de pensar melhor nas perguntas a colocar aos alunos e no modo como o fazer.

A concluir este ponto não podemos deixar de reafirmar que o questionamento, para além de ser a prática de ensino mais frequentemente realizada na sala de aula, é potenciador de uma avaliação formativa, uma vez que (i) acontece a par com as experiências de aprendizagem, permitindo uma regulação no momento; (ii) recorre à forma mais habitual de comunicação entre professor e alunos - a forma oral, e (iii) a sua responsabilidade pode deslocar-se do professor para o aluno. Quando acontece de forma adequada, envolve os alunos nas discussões e, através de questões de natureza aberta, leva-os a desenvolver o raciocínio (Black et al., 2003). Mas colocar boas questões no contexto da sala de aula não é simples (Gipps, 1999).

2.2.2 Feedback escrito ou escrita avaliativa

A principal questão que se coloca sobre o feedback é a sua eficácia. É neste sentido que Nicol (2010) nos alerta para a importância da qualidade da interação provocada pelo feedback: “O feedback é um processo dialógico e não o envio de uma mensagem” (p. 503). A partir de um feedback externo, o aluno estabelece um diálogo consigo próprio. O que é decisivo é a forma como o aluno adapta e integra o feedback.

Assim, a eficácia do feedback é uma questão complexa, razão pela qual se compreende por que o feedback tenha merecido a atenção da investigação ao longo de várias décadas (Black & Wiliam, 1998) até aos dias de hoje. Todo o feedback ocorre num contexto específico onde estão presentes diversos tipos de variáveis suscetíveis de intervirem na sua eficácia. Passaremos, de seguida, a discutir a eficácia do feedback

segundo: (i) as suas características; (ii) os fatores associados aos alunos; e (iii) as variáveis de contexto.

Características do feedback. No sentido de estudar a relação entre o conteúdo do feedback e os seus efeitos, diversas tipologias foram sendo criadas para categorizar esse conteúdo. A título de exemplo, refira-se Gipps (1999) que considera o feedback *avaliativo* e o *descritivo*. O primeiro traduz-se, sobretudo, num juízo de valor, com utilização implícita ou explícita de normas, centrado nas características/attitudes do aluno. O segundo incide na realização do aluno e na tarefa proposta. Na mesma linha, Jorro (2000) distingue dois tipos de escrita avaliativa. A *anotação como transmissão de informação*, que se traduz por juízos de valor ou por enunciados vagos e a *anotação como diálogo* que procura questionar, dar pistas e incentivar a reflexão por parte do aluno. Hattie e Timperley (2007) apresentam-nos outra categorização para o feedback, agora tendo por base o seu enfoque: a *tarefa*, o *processo*, a *autorregulação* e o *self*. Note-se que um mesmo professor pode usar, em momentos distintos, diferentes tipos de feedback. Nenhum tipo de feedback é característico de uma disciplina específica. Segundo Tunstall e Gipps (1996), todos os tipos de feedback foram observados em todas as áreas disciplinares estudadas. De acordo com estes autores, as tipologias de feedback constituem quadros de referência sobre os quais os professores podem refletir sobre a sua prática. Apresentamos, de seguida, um quadro que ilustra estas tipologias (Quadro 1).

Quadro 1. Exemplos de feedback e sua tipologia⁽¹⁾

Feedback	Tipo	Comentário
Pouca atenção! Não estudaste!	Avaliativo, reprovação Transmissão de informação Dirigido <i>ao self</i>	O professor recorre ao seu poder, fazendo uma inferência não assente em evidência
Tens de estudar mais	Avaliativo, castigo ou incentivo Transmissão de informação Dirigido <i>ao self</i>	Estudar o quê? Não fornece suficientes pistas para o aluno prosseguir
Bem feito!	Avaliativo, aprovação Transmissão de informação Dirigido à tarefa	Não preciso de fazer mais nada? O que está bem feito?
O teu trabalho está cheio de erros de cálculo! Corrige-os	Descritivo, aperfeiçoamento específico Transmissão de informação Dirigido à tarefa	Mas quais são? Não esclarece o suficiente para o aluno prosseguir
Se em vez destes valores tivesses outros chegarias à	Descritivo, construção do percurso a seguir	Dá pistas ao aluno como continuar/reformular

mesma conclusão? Experimenta e compara com os teus resultados. O que conclusis?	Anotação como diálogo Dirigido ao processo	
Relê o enunciado da tarefa. Vai anotando as diferentes informações. No final, compara-as com as que usaste. São as mesmas?	Descritivo, construção do percurso a seguir Anotação como diálogo Dirigido à regulação	Dá pistas ao aluno como continuar/reformular

⁽¹⁾ Adaptado de Santos et. al. (2010)

Embora diferentes, estas tipologias assentam no pressuposto de que o feedback que procura descrever, estabelecer um diálogo e não se centrar no *self*, é potencialmente mais eficaz. Contudo, no quadro anterior, os exemplos que verificam estas condições são bem mais longos, criando uma nova dificuldade: a exigência do aluno ter a capacidade de interpretar informação escrita.

Deste modo, procurar ser exaustivo no feedback, incluir informação em demasia, poderá pôr em risco a eficácia do feedback (Brookhart, 2007/08). A meta-análise de um conjunto de estudos desenvolvidos na aula de Matemática com alunos do 3º ciclo de escolaridade evidenciou que a dimensão do feedback é um fator que pode influenciar a sua eficácia (Santos & Pinto, 2009), resultados que vão na mesma linha dos obtidos noutra estudo desenvolvido no contexto da Física e Química (Bruno & Santos, 2010). Os feedbacks curtos parecem ser mais eficazes do que os longos, uma vez que ajuda os alunos a focarem-se em certos aspetos concretos da tarefa. Contudo, parece haver uma relação entre a dimensão do feedback e a natureza da tarefa. As tarefas com grau de desafio elevado tendem a requerer um feedback longo, relacionado com o grau de dificuldade da tarefa (Dias & Santos, 2010). Mas o feedback dado a tarefas abertas parece ser mais favorecedor de aprendizagem do que em tarefas mais estruturadas e guiadas (Bangert-Drowns, Kulick, & Morgan, 1991). Esta é certamente uma dimensão do feedback que requer mais investigação.

Também a forma sintática do feedback poderá relacionar-se com a sua eficácia. A forma interrogativa ou mista (interrogativa e afirmativa), em comparação com a afirmativa, parece facilitar a compreensão da mensagem do feedback por parte dos alunos e envolvê-los no prosseguimento do trabalho. Contudo, há o risco de os alunos responderem de forma direta à questão colocada, não prosseguindo no revisitar do processo desenvolvido e no seu questionamento (Santos & Pinto, 2009).

Em síntese, a investigação aponta para algumas características do feedback que parecem potenciar a sua eficácia, muito embora não a garantam:

- apontar pistas de ação futura, de forma que a partir dela o aluno saiba como prosseguir;
- incentivar o aluno a reanalisar a sua resposta;
- não incluir a correção do erro, no sentido de dar ao próprio a possibilidade de ser ele mesmo a identifica-lo e a alterá-lo de forma a permitir que aconteça uma aprendizagem mais duradoura;
- identificar o que já está bem feito, no sentido não só de dar autoconfiança, como igualmente permitir que aquele saber seja conscientemente reconhecido. (Santos, 2003, p. 19)
- utilizar uma linguagem acessível aos alunos, concreta, contextualizada e diretamente relacionada com a tarefa. (Bruno & Santos, 2010)

Fatores associados aos alunos. As crenças e motivações dos alunos são elementos que a investigação destaca como podendo estar presentes na forma como estes usam o feedback. “O feedback regula e é simultaneamente regulado por crenças motivacionais” (Nicol & Macfarlane-Dick, 2006, p. 201).

Também a perceção que os alunos têm sobre o feedback parece ter importância para a sua eficácia. Um estudo conduzido com alunos entre os 13 e os 15 anos de idade na Noruega (Gamlem & Smith, 2013) considerou quatro tipos de feedback a que os alunos atribuíam diferentes significados e potenciais efeitos. Os alunos sentiam que o feedback dado pelo professor que premeia, castiga e classifica, não os ajuda a aprender. Já o feedback de aprovação motiva-os a prosseguir o trabalho. O feedback, dado pelo professor ou pelos pares, que especifica a melhoria, é altamente apreciado pelos alunos. Quando fornecido pelo professor, este feedback torna-se útil se for dado tempo ao aluno para melhorar o seu trabalho. Por último, o feedback dialógico, interativo e que aponta pistas para prosseguir, dado pelo professor ou pelos pares, é raramente usado na sala de aula, mas, quando acontece, os alunos consideram-no muito útil.

O desempenho escolar é outra dimensão muito importante e que poderá, em certos casos, explicar por que feedbacks semelhantes fornecidos a produções, também elas

equivalentes, não produzem os mesmos efeitos: nalguns alunos ajuda-os a progredir, noutros não. Num estudo desenvolvido em duas turmas do 8.º ano de escolaridade em Físico-Química pode ler-se: “Os comentários que são eficazes para uns podem não o ser para outros. Em algumas situações fornecemos comentários iguais a produções muito semelhantes e verificámos que uns alunos conseguiam melhorar a sua produção, mas outros não” (Bruno, 2006, p. 200).

Um outro estudo, desenvolvido no 7º ano de escolaridade vem reforçar que o mesmo feedback escrito não serve da mesma forma todos os alunos. Este estudo parece indicar que “alunos com desempenho médio a Matemática necessitam de um feedback mais descritivo e menos simbólico” (Santos & Dias, 2006, p. 15). Alunos com desempenho fraco revelam dificuldade em compreender um feedback que faça apelo a conceitos matemáticos ou a ideias matemáticas abstratas. Esta situação é agravada pelo facto de apenas os alunos com melhor desempenho tenderem a recorrer ao professor para clarificação do que foi escrito (Semana, 2008).

Assinalar o erro através de uma simbologia parece ser suficiente para a compreensão de alunos com desempenho elevado. Já para alunos com desempenho fraco parece ser necessário acompanhar a sinalização do erro com uma pista explícita (Santos & Dias, 2006). Deste modo, o feedback pode contribuir para a aprendizagem, quando a escrita avaliativa atende às necessidades específicas de cada aluno (William, 1999).

Variáveis de contexto. O momento certo para dar feedback escrito parece ser um aspeto importante a ter em conta. Diversos estudos apontam que o feedback nunca deve surgir antes do aluno ter oportunidade para pensar e trabalhar sobre uma dada tarefa (William, 1999). Acresce que a eficácia do feedback dado a uma produção previamente classificada pode ser reduzida (William, 1999).

A natureza da relação que se estabelece entre alunos e professor é também referenciada como um aspeto que pode influenciar positiva ou negativamente a eficácia do feedback. Quando essa relação é evidente para os alunos, eles envolvem-se (Price, Handley, Millar, & O'Donovan, 2010). Atitude semelhante foi observada com alunos portugueses do 4º ano, que receberam feedback de forma continuada:

Ao longo das sessões, verificou-se que deixou de ser preciso questionar os alunos quanto à sua vontade para melhorar a sua aprendizagem

através da realização de novas tarefas, pois eram os próprios que solicitavam apoio. (Pimentel, 2013, p. 81)

Dar feedback é uma prática profissional exigente para o professor e consumidora de muito tempo (Bruno & Santos, 2010; Leal, 1992; Menino & Santos, 2004). Para além disso, não basta que os professores conheçam os fundamentos e orientações teóricas de uma escrita avaliativa formativa para que a sua prática seja conforme essas mesmas orientações (Bruno & Santos, 2010; Menino, 2004). Existe uma forte tendência, numa primeira fase, para uma escrita fortemente marcada por juízos de valor, de cariz simbólico, tendencialmente normativa e essencialmente afirmativa ao invés de interrogativa e favorável à reflexão por parte do aluno.

O uso do feedback exige uma aprendizagem por parte do professor através de uma constante reflexão sobre os resultados obtidos em termos dos seus efeitos na aprendizagem dos alunos. Esta reflexão exige uma intencionalidade comprometida com uma conceção de avaliação para a aprendizagem que leve a fazer escolhas nas tarefas que se usa, na forma como se analisa e, sobretudo, no modo como essa análise se devolve aos alunos. Com o evoluir do tempo, os professores podem adquirir plasticidade nos feedbacks ajustando-os, quer aos alunos concretos, quer às próprias tarefas (Santos & Pinto, 2010). A evolução de práticas avaliativas não se faz de uma forma linear (William, Lee, Harrison, & Black, 2004), mas acontece.

2.3 Questões de aula

Questões de aula são questões que se colocam no final de uma dada aula, de resposta breve (cerca de 10m a 15m), que abordam os assuntos tratados nessa aula. Esta prática avaliativa surgiu já há alguns anos nas escolas portuguesas. Pela sua natureza permite colocar diversas questões sobre ideias importantes trabalhadas na aula e identificar no momento ideias erróneas construídas pelos alunos (Glasgow University, s/d). Com estas informações, o professor pode regular o seu ensino para a aula seguinte. Naturalmente, uma vez que as questões de aula têm de ser respondidas em tempo muito limitado, as aprendizagens que testam limitam-se à verificação da aquisição de conhecimentos.

Muito embora não existam, em Portugal, pelo menos de forma significativa, estudos que descrevam práticas desenvolvidas com questões de aula, existe a ideia formada a

partir do que se ouve dizer que esta estratégia avaliativa é sobretudo usada, a partir do 2º ciclo, como um teste feito em diversos momentos. Por outras palavras, um conjunto de questões de aula substitui um teste escrito por período letivo. Há mesmo quem aponte o número dez para definir o conjunto de questões que perfazem um teste.

Numa lógica de avaliação sumativa, o seu uso parece ser justificado pela ideia que esta estratégia avaliativa ajuda os alunos a estarem atentos à aula, uma vez que sabem que, no seu final, serão sujeitos a uma verificação das aprendizagens realizadas nessa aula. Poder-se-á dizer que, pelo menos para os alunos em que as questões de aula produzem este efeito, esta estratégia avaliativa pode contribuir para a melhoria do desempenho escolar, tal como evidenciado num estudo desenvolvido no contexto da disciplina de Matemática, no tópico de Geometria (Shirvani, 2009). Participaram, neste estudo, quatro turmas de alunos espanhóis do ensino secundário com o mesmo professor. Foi formado o grupo experimental com duas turmas, a quem foram aplicadas questões de aula de 10m, ao longo de seis semanas, e o grupo de controlo com outras duas turmas, a quem foi aplicado um teste uma vez por semana. Os resultados evidenciaram que os alunos do grupo experimental melhoraram de forma significativa o seu desempenho a Matemática.

Coloca-se, contudo, a questão de saber se haverá outras formas de explorar com os alunos as questões de aula tornando-as mais profícuas para a aprendizagem. Um estudo recentemente desenvolvido nos EUA procurou compreender a perceção dos alunos face à exploração de questões de aula com uma intencionalidade formativa. Participaram 1408 alunos dos 2º e 3º ciclos e secundário. Os resultados obtidos evidenciaram que 92% dos alunos inquiridos consideraram que as questões de aula, quando não classificadas, os ajudou a aprender e 72% disseram sentir-se menos ansiosos do que quando realizavam testes (Agarwal, D'Antonio, Roediger III, McDermott, & McDaniel, 2014)

Em Portugal, numa turma de 26 alunos do 2º ano de escolaridade aplicaram-se questões de aula, percorrendo três etapas: 1) os alunos respondiam às questões de aula; 2) o professor levava as produções realizadas e na aula seguinte devolvia-as aos alunos com feedback, registado em *post-it* (crianças desta idade não gostam que

fiquem registados nos seus trabalhos comentários do professor (Santos & Pinto, 2011)); e 3) os alunos tinham de reformular as suas primeiras respostas consideradas incompletas ou erradas. Para os alunos que resolviam sem erros as questões de aula era-lhes colocado um novo desafio. Em nenhum momento, as questões de aula foram sujeitas a classificação. Os resultados evidenciam que as questões de aula promoveram “nos alunos o interesse e a motivação para aprender e para melhorar o seu desempenho” (Gomes, 2016, p. 67). Progressivamente foi aumentando o número de alunos que passou a corrigir todos os erros assinalados.

Dos estudos apresentados pode concluir-se que, quer do ponto de vista dos alunos, quer do desempenho que revelam, as questões de aula poderão contribuir para a aprendizagem, sobretudo se forem trabalhadas com essa intencionalidade, isto é, não lhes atribuindo uma natureza sumativa, mas antes formativa. As estratégias a levar à prática na sala de aula podem ser diversas: (i) o professor analisar as respostas dadas pelos alunos, permitindo-lhe planificar a aula seguinte de forma fundamentada; (ii) se necessário, retrabalhar o mesmo assunto, seguindo uma nova abordagem; (iii) dar oportunidade para os alunos retrabalharem as mesmas questões; (iv) propor tarefas de análise dos erros cometidos; (v) recorrer à coavaliação enquanto contexto para desenvolver algumas das anteriores estratégias. Mas estas estratégias não estão naturalmente desprovidas de dificuldades. Desde logo, há que contar com uma eventual resistência por parte dos alunos face a uma prática que não segue a cultura de avaliação que conhecem (Gomes, 2016). Acrescem desafios inerentes a possíveis mudanças de prática, como seja os professores sentirem falta de tempo, quer para dar feedback a todos as produções dos alunos, quer para desenvolverem, na sala de aula, as estratégias anteriormente elencadas.

PRÁTICAS DE AVALIAÇÃO SUMATIVA

Quando se fala em Portugal de instrumentos de avaliação sumativa destaca-se, desde logo, o teste escrito tradicional, por ser este o que predomina na prática avaliativa dos professores, independentemente do ciclo de escolaridade ou da disciplina. Estamos a entender por teste escrito tradicional toda a prova escrita, formada por um número de questões/itens, realizada individualmente e em tempo limitado. Esta prática, embora

habitual, carece de questionamento: Será que o teste escrito é a forma mais adequada para validar os conhecimentos dos alunos? Este instrumento é credível? Este é o instrumento que garante a qualidade da avaliação desejada?

Uma forma de alterar esta prática poderia ser mudar o ponto de partida do processo. Em vez de se assumir que garantidamente é o teste escrito que irá ser utilizado num momento formal de avaliação sumativa, poder-se-ia começar por responder às seguintes questões: O que se quer avaliar? Qual o tempo que o professor disponibiliza para esta tarefa? Quantos alunos serão abrangidos? De que materiais se dispõe? Que instrumentos já foram usados anteriormente? As respostas obtidas poderiam ser analisadas tendo em conta uma listagem de possíveis instrumentos de avaliação de forma a decidir qual o mais adequado naquele momento.

3.1. O teste escrito

Imaginemos que se pretendia construir um teste escrito. Como proceder? Recorrer a uma técnica antiga pode ser uma boa opção. Falamos da construção de uma matriz de conteúdos/capacidades (Quadro 2). Por outras palavras, construindo uma tabela onde devo identificar quais os conteúdos e capacidades que quero avaliar e os pesos relativos que vou atribuir a cada um. Naturalmente que a tabela será preenchida de acordo com o que se quer avaliar, com o que foi trabalhado na aula, e a importância relativa que foi sendo atribuída a cada conteúdo e capacidade durante o processo de ensino e aprendizagem. Só após a construção desta matriz se deve começar a selecionar/adaptar/construir as questões ou itens a incluir no teste.

Quadro 2. Exemplo de uma versão por completar de uma matriz de conteúdos e capacidades

Conteúdos	Conhecimentos	%	Capacidade 1	%	Capacidade 2	%	%
Conteúdo 1							
(...)							

Os itens a incluir num teste podem ser de *seleção* ou de *construção* (Neves & Ferreira, 2015). Entre os de seleção, podemos elencar os de:

- *Escolha múltipla* - responder implica selecionar a opção correspondente à resposta correta entre possíveis alternativas apresentadas;
- *Associação* - requer estabelecer relações entre dois conjuntos de expressões fornecidos;

- *Ordenação* - exige estabelecer uma sequência ordenada das expressões apresentadas;
- *Verdadeiro/falso* - exige decidir se a afirmação apresentada é “verdadeira” ou “falsa”;
- *De completamento* - requer o preenchimento de uma ou mais lacunas numa frase, a partir de informação fornecida.

Como itens de construção, podemos ter:

- *De completamento* - requer o preenchimento de uma ou mais lacunas numa frase, através da elaboração do próprio;
- *Resposta curta* - exige apresentar uma frase sucinta, uma palavra ou número;
- *De desenvolvimento* - requer a elaboração de um texto, que deve ou não respeitar um conjunto de indicações (resposta orientada ou não orientada).

Qualquer que seja a tipologia do item, ele é adequado para recolher dados sobre um conjunto de aspetos e desadequado para outros, isto é apresenta vantagens e limitações. Requer ainda que sejam decididos à partida quais os critérios de classificação a aplicar em cada situação. O quadro 3. sintetiza os critérios de classificação para cada tipo de item, de acordo com Neves e Ferreira (2015).

Quadro 3. Tipologia de itens e critérios de classificação

Tipo de item		Critérios de classificação
Itens de seleção	Escolha múltipla	Dicotómico (certo/errado)
	Associação	Dicotómico ou não (níveis de desempenho)
	Ordenação	Dicotómico (certo/errado)
	Verdadeiro/Falso	Dicotómico ou não (níveis de desempenho)
	De completamento	Dicotómico ou não (níveis de desempenho)
Itens de construção	Resposta curta	Dicotómico (certo/errado)
	De completamento	Dicotómico ou não (níveis de desempenho)
	De desenvolvimento	A indicar

Todo o item exige um conjunto de procedimentos para a sua elaboração, sendo uma tarefa bem mais complexa do que aquilo que aparentemente se possa pensar. Vejamos, a título de exemplo, o caso dos itens de escolha múltipla, por serem atualmente muito usados (possivelmente por existir pelo menos uma questão deste tipo em muitos exames). Existe um conjunto de cuidados a ter na sua elaboração:

- A base deve conter um único problema;
- A base deve ser formulada positivamente;

- A base não deve ser sobrecarregada com expressões irrelevantes;
- A base não deve incluir sugestões para a escolha da opção;
- Só uma das opções deve conter a resposta correta;
- Os distratores devem ser respostas incorretas, mas plausíveis;
- As opções devem conter os erros que os alunos habitualmente cometem;
- As opções devem ser curtas e com dimensão semelhante;
- As opções devem ser colocadas numa coluna vertical e em número de quatro;
- A posição da resposta correta deve ser definida aleatoriamente;
- Devem ser colocadas letras em vez de números antes das opções (Noizet & Caverni, 1985).

Embora se tenham indicado no Quadro 3. apenas dois critérios de classificação, certo ou errado, há quem recorra ainda a métodos para reduzir a parte do acaso nos resultados, como sejam o método de correção do acaso ou a auto-ponderação (Noizet & Caverni, 1985).

3.2. Critérios de qualidade da avaliação sumativa

Associada às condições de construção dos itens está a ideia da garantia de qualidade, tanto de cada item, como naturalmente do instrumento de avaliação, no seu todo. Numa perspetiva de avaliação enquanto medida, existem três critérios de qualidade que usualmente se associam aos instrumentos de avaliação: a *aplicabilidade*, a *validade*, e a *fidelidade*. A aplicabilidade, tal como o seu nome indica, diz respeito à garantia de verificação das condições necessárias para que o instrumento seja aplicado.

Falar-se de validade significa que o instrumento permite medir aquilo para o qual foi pensado. Tradicionalmente são consideradas quatro dimensões da validade:

- *Validade de conteúdo* – se o instrumento de avaliação/prova constitui uma amostra representativa e relevante do conteúdo que se pretende medir;
- *Validade concorrente* – se existe correlação entre a prova e outra que avalia os mesmos conhecimentos, isto é, se as classificações obtidas nas duas provas são semelhantes;

- *Validade preditiva* – se existe correlação entre a classificação obtida na prova e a predição de um desempenho futuro dos alunos;
- *Validade de constructo* – avalia a validade de uma prova ou medida de um constructo que não lhe tem associado critérios quantificáveis (Gipps, 1994).

A investigação identificou diversos fatores que podem comprometer a validade, como sejam o grau de representatividade dos conhecimentos que se pretendem avaliar, de familiarização com os contextos, e da clareza das instruções (rigor da linguagem); a forma sintática das questões e a ordem pela qual são apresentadas, a pertinência, relevância, a clareza e qualidade gráfica dos materiais e, naturalmente, a disposição física e/ou psíquica do aluno (Noizet & Caverni, 1985).

Falar de fidelidade corresponde a saber até que ponto a medida obtida através do instrumento de avaliação é independente do avaliador. Esta questão está naturalmente associada à eterna problemática da objetividade versus subjetividade do ato de avaliar, discutida na primeira secção deste capítulo. Noizet e Caverni na investigação que desenvolveram por volta dos anos 70 do séc. XX, tendo por base a psicologia experimental, deram contributos importantes para a compreensão mais aprofundada desta problemática. Estes autores apresentam-nos dois tipos de fatores decorrentes da pessoa do avaliador que podem comprometer a fidelidade: os *efeitos de contraste* e os *efeitos de assimilação*.

Os efeitos de contraste dizem respeito à dinâmica de correção, isto é à reformulação dos critérios de avaliação que o avaliador vai produzindo à medida que os vai usando. Quando se elabora uma prova, antes de a aplicar, atribui-se uma classificação a cada questão, subdivide-se a classificação de cada situação e prevêem-se situações possíveis de respostas, de modo a decidir o que descontar em algumas situações antecipadas de erro. Ou seja, para classificarmos uma prova, partimos sempre de um modelo de referência, tão pormenorizado quanto possível. O efeito de contraste diz respeito exatamente à alteração involuntária desse modelo de referência à medida que o vamos usando. Como este processo não é semelhante à produção de uma medida, mas sim de seleção e tratamento de informação, a qualidade da prova anterior influencia o modo como apreciamos a seguinte. Quando nos confrontamos com uma produção muito boa, na prova seguinte alteramos o modelo de referência,

uma vez que nos confrontamos com uma alteração do produto esperado. A existência de variação acontece igualmente perante uma produção muito fraca. A resposta do avaliador “é função de uma interação entre a personalidade do avaliador e a situação” (Noizet & Caverni, 1985, p. 68). Se isto não fosse assim, não faria sentido existirem, por exemplo, os recursos a exames. O comportamento do avaliador de um recurso caracteriza-se por querer aproveitar o mais possível toda a informação suscetível de ser avaliada. Há, assim, uma mudança de comportamento do avaliador quando está nesta situação. E isto resulta noutra alteração do modelo de referência.

Nos efeitos de assimilação, o avaliador está condicionado ao conjunto de conhecimentos que tem sobre quem realizou as provas. Se for um aluno que conhecemos e sobre o qual temos uma representação positiva, mas que errou, temos a tendência para o desculpar, arranjando razões explicativas da situação (distraiu-se, estava mal disposto,...). Se for um aluno que consideramos de fraco aproveitamento, mas que apresenta alguma coisa que nos surpreende pela positiva, procuramos de igual modo encontrar explicações que a justifique (copiou pelo colega, ...). Mesmo na correção dos exames, que são anónimos, olhamos para uma prova e construímos uma imagem do seu produtor. Se está muito organizada, limpa e sem rasuras, tendemos a formar uma imagem positiva do seu autor. Essas representações que construímos a partir de informações que recolhemos do produto anónimo que estamos a analisar têm influência na classificação, na forma como aplicamos os referenciais. Há, assim, um conjunto de informações que são assimiladas ou construídas sobre o autor da produção que estamos a classificar e que influencia o que conseguimos e fazemos com o modelo de referência de classificação. Estas atitudes dos avaliadores estão associadas a razões psicológicas. Quando nos deparamos com algo que não está consonante com a ideia prévia que construímos, isto é em que existe uma dissonância cognitiva, temos necessidade de arranjar justificações e até eventualmente forçar situações, para anular essa dissonância, de modo a que a nossa ideia prévia não seja posta em causa, não seja questionada.

Existem algumas estratégias, alguns processos de moderação, no sentido de minimizar os efeitos referidos, como seja conhecer o teste de forma profunda, resolvendo-o detalhadamente antes de iniciar a sua classificação, bem como o modelo de referência

a usar, podendo este ser discutido entre pares de professores. Outra estratégia possível é classificar o teste item a item, e não cada teste na sua globalidade, antes de prosseguir para o seguinte. Ao terminar a classificação de um item, devemos rever os primeiros testes para verificar se mantemos a classificação inicialmente atribuída, verificando se houve grandes desvios do referencial, por outras palavras, usar a multicorreção, neste caso envolvendo o mesmo avaliador. Esta técnica pode ser igualmente aplicada recorrendo a avaliadores distintos. Uma outra estratégia será reduzir a escala de classificação.

Note-se que a validade e a fidelidade são critérios de qualidade dependentes entre si. Quando um aumenta, o outro em geral diminui (Harlen, 2010). Por exemplo, quando se reduz a diversidade da tipologia de itens aumenta-se a fidelidade, mas reduz-se a validade (Santos, 2011). Um caso extremo é o dos itens de escolha múltipla que surgiram exatamente para dar resposta à dificuldade de encontrar instrumentos de avaliação que garantissem a fidelidade:

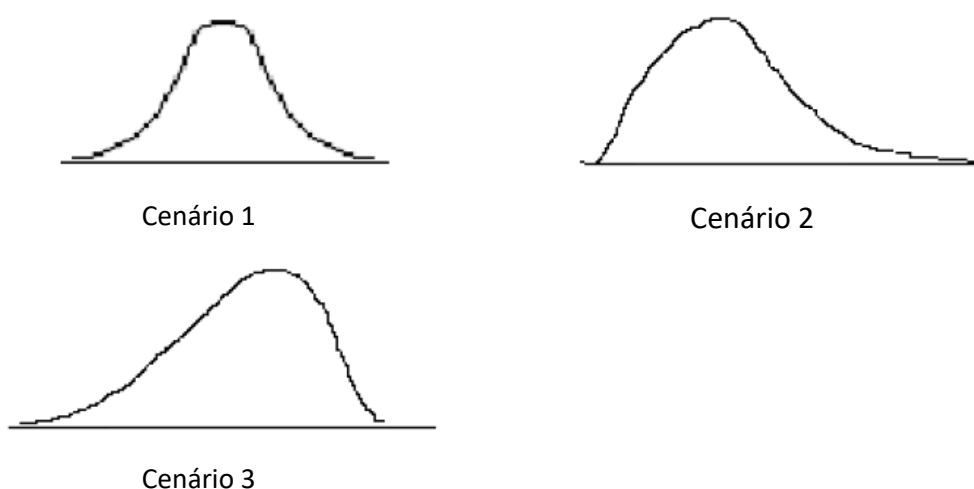
Perante a verificação das divergências na avaliação, perante a dificuldade de estabelecer métodos de moderação e em face dos seus resultados pouco satisfatórios, imaginou-se recusar formas tradicionais do exame e substituí-las por técnicas de avaliação que deixem de dar origem a desacordos entre avaliadores.

Foi assim que se conceberam exames por questionários chamados de escolha múltipla. (Noizet & Caverni, 1985, p. 159)

Contudo, embora a fidelidade seja totalmente atingida neste tipo de itens, dois avaliadores diferentes que apliquem a mesma grelha obtêm o mesmo resultado, a validade decresce. Por um lado, porque a construção deste tipo de itens é muito exigente tecnicamente, por outro, porque implica uma alteração do trabalho de quem responde: passa-se de um trabalho de produção para um de avaliação (Noizet & Caverni, 1985). O respondente tem de criar estratégias de resposta porque não está, em geral, num total estado de ignorância, onde optaria por uma lógica de responder ao acaso, nem total estado de saber, em que a resposta seria dada na base desse saber. Ora, são essas estratégias de resposta que podem mudar numa mesma pessoa, em momentos distintos de aplicação destes tipos de itens, pondo assim em risco a validade da prova.

As razões do erro resultantes do instrumento de avaliação podem dever-se, nomeadamente à dificuldade dos itens, à sua ambiguidade, ou a conteúdos inapropriados. Na prática corrente dos professores, a representação gráfica dos resultados dos alunos de uma turma a uma prova pode ser um método rápido que dá uma ideia aproximada da sua adequação ao grupo de alunos. Partindo do pressuposto que a avaliação enquanto medida é normativa, o confronto entre a curva de distribuição dos resultados e a curva normal permite-nos tirar primeiras conclusões (Figura 4).

Figura 4. Possíveis curvas de distribuição de resultados de um teste



O cenário 1 permite-nos afirmar que a distribuição dos resultados segue uma curva normal e como tal a prova foi adequada ao grupo de alunos. O cenário 2 traduz que a prova foi difícil demais, uma vez que houve uma deslocação dos resultados para a esquerda, querendo com isto dizer que o maior número de resultados se encontra abaixo do esperado. Já o cenário 3 ilustra uma situação oposta. Uma curva assimétrica para a direita representa uma concentração de resultados acima do valor médio esperado.

Existem, contudo, outras técnicas, mais precisas e finas que podem ser igualmente usadas pelos professores, mas um pouco mais trabalhosas, que poderão ser feitas de vez em quando. Enquanto a representação gráfica da distribuição dos resultados diz respeito aos resultados globais obtidos, as técnicas que a seguir se apresentam focam-se em cada item da prova (Borgatto & Andrade, 2012).

Para determinar o grau de dificuldade de um item procede-se de acordo com as seguintes etapas:

1. Ordenam-se os resultados obtidos no item;
2. Define-se o grupo dos 25% de alunos com piores resultados (GP) e o grupo dos 25% de alunos com melhores resultados (GM);
3. Calculam-se os números de respostas erradas de cada grupo construído: P do GP e M do GM;
4. Para calcular o grau de dificuldade (Gf), aplica-se a seguinte expressão aos valores de que se dispõe:

$$Gf = \frac{M+P}{N}, \text{ em que } N \text{ é o número total de alunos de ambos os grupos formados}$$

5. Gf varia entre 0 e 1. Representa um valor aceitável quando estiver entre 0,3 e 0,7.

Ao grau de dificuldade podemos acrescentar o grau de discriminação de um item, que nos permite saber até que ponto é que este é capaz de diferenciar os alunos, tendo em conta que não é demasiado difícil, ou seja, que o grau de dificuldade já foi avaliado e está nos níveis aceitáveis. O procedimento é em todo semelhante ao da determinação do grau de dificuldade:

1. Ordenam-se os resultados obtidos no item;
2. Define-se o grupo dos 25% de alunos com piores resultados (GP) e o grupo dos 25% de alunos com melhores resultados (GM);
3. Calculam-se os números de respostas corretas de cada grupo construído: Cp do GP e Cm do GM;
4. Para calcular o grau de discriminação (Gd), aplica-se a seguinte expressão aos valores de que se dispõe:

$$Gd = \frac{Cm-Cp}{N/2}, \text{ em que } N \text{ é o número total de alunos de ambos os grupos formados}$$

5. Gd varia entre -1 e 1. Se o grau de discriminação for negativa deve eliminar-se o item. É aceitável quando o valor obtido for igual ou superior a 0,35.

A terminar não podemos deixar de nos referir às possíveis escalas de classificação. A expressão da avaliação enquanto medida é feita através de uma notação. Estas escalas podem ser qualitativas ou quantitativas. Temos, como exemplo de uma escala qualitativa, a escala com cinco valores/categorias – Excelente, Muito bom, Bom,

Suficiente, Insuficiente. Este é um exemplo de uma escala ordinal, uma vez que ordena os diversos valores segundo um dado processo de comparação, mas não indica a medida das diferenças entre valores. Situação idêntica se poderia apresentar numa escala cujas categorias são as primeiras cinco letras do alfabeto, A, B, C, D e E, que embora pouco usada no nosso país tem expressão noutros.

Um outro exemplo de escala ordinal, mas agora quantitativa, é a escala de níveis de 1 a 5, usada na atribuição da classificação final de período dos alunos nos 2º e 3º ciclos do Ensino Básico em Portugal. Exatamente por se tratar de uma escala ordinal, os valores atribuídos serem discretos, não existe nada entre o 2 e o 3. Por exemplo, em rigor, atribuir 3⁺ ou 3,5 nesta escala não faz qualquer sentido.

Pode-se falar em escala com valores quantitativos contínuos no caso da escala de 0 a 20 valores ou de 0 a 100. Estas escalas não só permitem ordenar, mas também conhece-se a medida das diferenças entre valores, pelo que podem ser usadas com os valores expressos em unidades, em décimas, em centésimas, etc.

Note-se que, em nossos entender, a escolha da escala de classificação a usar não depende da área curricular a que corresponde (havia a ideia tradicional que, por exemplo seria mais adequado usar uma escala com menos valores a Filosofia do que a Matemática), mas sim da natureza da tarefa sobre a qual se pretende atribuir uma medida sobre a qualidade do desempenho. Quanto mais a tarefa for de construção e aberta, tendemos a dizer que uma escala ordinal com um número reduzido de categorias será mais adequada.

Relacionar o tipo de tarefa com a natureza da escala de classificação pode ser percebida por um outro ângulo. Em vez de atender à natureza das categorias, podemos distingui-la pela forma como apreciamos a resolução em apreço. Se usarmos uma lógica de divisão do todo em partes, seguimos uma “escala de classificação analítica”. Se usarmos um olhar global, técnica apenas recomendada a avaliadores muito experientes, estamos perante uma escala designada por “Impressão geral”. Por fim, uma “escala de classificação holística focada” é aquela que define o que se entende por cada uma das categorias que a constitui. Olha para a produção como um todo e não como um somatório de partes (Charles, Lester & O’Daffer, 1987).

Iniciámos a presente secção questionando se o teste escrito tradicional, instrumento de recolha de informação para fins de validação, seleção, hierarquização, e certificação, seria um meio válido, credível, e que garantisse a qualidade da avaliação pretendida. Ao longo destas linhas fomos alertando para a complexidade do processo de recolha de informação, sua conseqüente interpretação e definição de ações a desenvolver, nomeadamente pela sua dependência às questões de natureza técnica e à pessoa do avaliador. Esperamos que o tema aqui tratado contribua para uma prática avaliativa sumativa informada e adequada.

ATIVIDADES PRÁTICAS

Segue-se um conjunto de atividades que têm por finalidade uma aplicação de conceitos tratados ao longo deste capítulo.

Tarefa 1

Construa uma tabela com duas colunas. Numa delas liste todos os instrumentos de avaliação que considera adequados na sua disciplina. Noutra, selecione no **máximo 3 objetivos gerais**, retirados do Programa, que cada instrumento preferencialmente poderá cobrir. No final todos os objetivos deverão ter sido cobertos.

Compare e discuta, de forma fundamentada, com colegas as resoluções obtidas.

Que implicações retira para a sua prática avaliativa?

Tarefa 2

Identifique os objetivos a trabalhar com os seus alunos e planifique as estratégias avaliativas de acordo com a seguinte tabela:

O que quero saber?	Qual o tipo de saber? (declarativo; prático)	Através de que tarefa? (questão escrita, oral, trabalho?)	Como sei que o aluno aprendeu? (critérios)	Dificuldades previsíveis e formas de ajuda
...				
...				

Tarefa 3

Pense em situações/instrumentos de avaliação formativa que usa habitualmente. Escreva-as e explicita como as utiliza. Que benefícios consegue identificar para a aprendizagem dos alunos? Que razões podem explicar as mais-valias identificadas? Faça um balanço das dificuldades que sente e/ou que os seus alunos manifestam na utilização destas formas de avaliação e identifique estratégias para as minimizar.

LEITURAS SUGERIDAS PARA APROFUNDAMENTO

Para além das obras referenciadas no texto, sugerimos ainda como possíveis leituras para aprofundamento as seguintes:

Fernandes, D. (2005). *Avaliação das aprendizagens: Desafios às teorias, práticas e políticas*. Lisboa: Texto Editores.

J. Gardner, W. Harlen, L. Hayward, & G. Stobart (Eds.), (2010). *Developing teacher assessment*. London: McGraw-Hill Education & Open University Press.

J. H. McMillan (2013). *SAGE Handbook of Research on classroom assessment*. California, USA: SAGE Publications Inc.

Santiago, P.; Donaldson, G.; Looney, A. & Nusche, D. (2012). *OECD Reviews of evaluation and assessment in education: Portugal*. OECD (<http://www.oecd.org/edu/evaluationpolicy>)

Santos, L. (2002). Auto-avaliação regulada: porquê, o quê e como? In Paulo Abrantes e Filomena Araújo (Orgs.), *Avaliação das Aprendizagens. Das concepções às práticas* (pp. 75-84). Lisboa: Ministério da Educação, Departamento do Ensino Básico.

Santos, L. (2009). Diferenciação pedagógica: Um desafio a enfrentar. *Noésis*, 79, 52-57.

Santos, L., & Pinto, J. (2003) O que pensam os alunos sobre a avaliação? *Educação e Matemática*, 74, 2, 15, 21, 32, 73, 84 e 85.

RECURSOS ONLINE

<http://area.fc.ul.pt>

<http://www.dgeec.mec.pt/np4/192.html>

<http://www.cne.pt>

REFERÊNCIAS

- Agarwal, P., D'Antonio, L., Roediger III, H., McDermott, K., & McDaniel, M. (2014). Classroom-based programs of retrieval practice reduce middle School and high School students' test anxiety. *Journal of Applied Research in Memory and Cognition*, 3(3), 131-139.
- Bangert-Drowns, R., Kulick, J., & Morgan, M. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, 85, 89-99.
- Black, P. (2013). Formative and summative aspects of assessment: Theoretical and research foundations in the context of pedagogy. In J. McMillan (Ed.), *SAGE Handbook of research on classroom assessment* (pp. 167-178). California, USA: SAGE Publications Inc.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning. Putting into practice*. London: Open University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1) 5-31.
- Beirão, E. (2012). *O desenvolvimento do raciocínio matemático apoiado pelo uso continuado de critérios de avaliação: Um estudo com alunos do 2.º ciclo de escolaridade* (Dissertação de mestrado, Universidade de Lisboa)
- Bonniol, J., & Vial, M. (1997). *Les modèles de l'évaluation*. Paris: De Boeck & Larcier s.a.
- Borgatto, A. & Andrade, D. (2012). Análise clássica de testes com diferentes graus de dificuldade. *Estudos em Avaliação Educacional*, 23(52), 146-156.
- Brookhart, S. (2007/2008). Feedback that fits. *Educational Leadership, Informative assessment*, 65(4), 54-59.
- Bruno, I. (2006). *Avaliação das aprendizagens: O processo de regulação através do feedback – um estudo em Físico-Química no 3º ciclo do ensino básico*. (Tese de Mestrado, Universidade de Lisboa).
- Bruno, I. (2013). *Os critérios de avaliação para o desenvolvimento da autorregulação das aprendizagens*. (Tese de doutoramento, Universidade de Lisboa)

- Bruno, I., & Santos, L. (2010). Evolução da escrita avaliativa num contexto de trabalho colaborativo. *Revista de Educação, XVII(2)*, 61-92.
- Charles, R., Lester, F., & O'Daffer, P. (1987). *How to evaluate progress in problem solving*. Reston, Virginia: National Council of Teachers of Mathematics.
- Chevalard, Y. (1990). Évaluation, véridiction, objectivation. In J. Colomb & J. Marsenach (Orgs.), *L' évaluateur en révolution* (pp. 13-36). Paris: INRP.
- Dias, S., & Santos, L. (2010). O feedback e os diferentes tipos de tarefas matemáticas. *XXI SIEM* (CD ROM) (pp. 126-136). Aveiro: APM.
- Gamlem, S., & Smith, K. (2013). Student perception of classroom feedback. *Assessment in Education: Principles, Policy & Practice, 20(2)*, 150–169.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*, London: The Falmer Press.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education, 24*, 355-392.
- Glasgow University (s/d). *Assessment for learning: Using Moodle Quizzes*. (<http://www.gla.ac.uk/services/senateoffice/academic/assessmentpolicies>)
- Gomes, S. (2016). *Práticas de questões-aula numa perspetiva de avaliação formativa em Matemática*. (Relatório de Mestrado, Instituto Politécnico de Setúbal)
- Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal, 16(2)*, 207-223.
- Harlen, W. (2010). What is quality teacher assessment?. In J. Gardner, W. Harlen, L. Hayward & G. Stobart (Eds.), *Developing teacher assessment* (pp. 29-52). London: McGraw-Hill Education & Open University Press.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77(1)*, 81-112.
- Hodgen, J. (2007). Formative assessment: Tools for transforming school mathematics towards a dialogic practice? *CERME 5* (CD-ROM) (pp. 1886-1895). Cyprus: University of Cyprus.
- Jorro, A. (2000). *L'enseignant et l'évaluation. Des gestes évaluatifs en question*. Bruxelles: De Boeck Université.

- Leal, L. C. (1992). *Avaliação da aprendizagem num contexto de inovação curricular*. (Tese de Mestrado, Universidade de Lisboa). Lisboa: APM.
- Küchemann, D., & Hoyles, C. (2006). Influences on students' mathematical reasoning and patterns in its development. *International Journal of Science and Mathematics Education*, 581-608.
- Mateo, J. (2000). *La evaluación educativa, su práctica y otras metáforas*. Barcelona: ICE – Universidad de Barcelona, cuadernos de educación.
- Menino, H. (2004). *O relatório escrito, o teste em duas fases e o portfolio como instrumentos de avaliação das aprendizagens em Matemática: Um estudo no 2º ciclo do Ensino Básico*. (Tese de Mestrado, Universidade de Lisboa). Lisboa: APM.
- Menino, H., & Santos, L. (2004). Instrumentos de avaliação das aprendizagens em matemática. O uso do relatório escrito, do teste em duas fases e do portefólio no 2º ciclo do ensino básico. *XV SIEM* (pp. 271-291). Lisboa: APM.
- Merle, P (1996). *L'évaluation des élèves - enquête sur le jugement professoral*. Paris: PUF.
- Neves, A. & Ferreira, A. (2015). *Avaliar é preciso? Guia prático de Avaliação para professores e formadores*. Lisboa: Guerra e Paz, Editores, S. A..
- Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517.
- Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Noizet, G., & Caverni, J. (1985). *Psicologia da avaliação escolar*. Coimbra: Coimbra Editora. (obra original em francês, publicada em 1978)
- Peres, A. (2012). *O uso de critérios de avaliação na resolução de problemas*. (Tese de Mestrado, Universidade de Lisboa)
- Perrenoud, Ph. (1999). *Avaliação. Da excelência à regulação das aprendizagens. Entre duas lógicas*. São Paulo: Artmed.
- Pierron, H. (1963). *Examens et docimologie*. Paris: PUF.

- Pimentel, I. (2013). *O contributo do portefólio para as aprendizagens dos alunos no 1.º ciclo do ensino básico*. (Relatório de Mestrado, Instituto Politécnico de Setúbal)
- Pinto, F., & Santos, L. (2010). A comunicação em sala de aula no desenvolvimento de uma tarefa de natureza exploratória. *EIEM 2010, Comunicação no Ensino e na Aprendizagem da Matemática* (pp. 87-101). Costa da Caparica: SPIEM. (<http://www.spiem.pt/publicacoes/arquivo/encontro-2010/>)
- Pinto, J. (2002). *A avaliação formal no 1º ciclo do ensino básico: Uma construção social*. (Tese de Doutoramento, Universidade do Minho).
- Pinto, J., & Santos, L. (2006). *Modelos de avaliação das aprendizagens*. Lisboa: Universidade Aberta.
- Pinto, J., & Santos, L. (2012). Avaliar para aprender nos primeiros anos. In Mª. J. Cardona, & C. Mª. Guimarães (Orgs.), *Avaliação na educação de infância* (pp. 334-352). Viseu: Editora Psicosoma.
- Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: all the effort, but what is the effect? *Assessment & Evaluation in Higher Education*, 35(3), 277-289.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77-84.
- Santos, L. (2003). Avaliar competências: uma tarefa impossível? *Educação e Matemática*, 74, 16-21.
- Santos, L. (2008). Dilemas e desafios da avaliação reguladora. In L. Menezes, L. Santos, H. Gomes & C. Rodrigues (Eds.), *Avaliação em Matemática: Problemas e desafios* (pp. 11-35). Viseu: Secção de Educação Matemática da SPCE. (http://www.esev.ipv.pt/mat1ciclo/avaliacao_files/MA_livro_Aval..pdf)
- Santos, L. (2011). Que critérios de qualidade para a avaliação formativa? In D. Fernandes (Org.), *Avaliação em educação: Dez olhares sobre uma prática social incontornável* (pp. 155-165). Curitiba: Editora Melo.
- Santos, L. & Dias, S. (2006). Como entendem os alunos o que lhes dizem os professores? A complexidade do feedback. *ProfMat2006*. (CD-ROM). Lisboa: APM.

- Santos, L., & Gomes, A. (2006). Apropriação de critérios de avaliação: um estudo com alunos do 7º ano de escolaridade. *Revista Portuguesa de Pedagogia*, 40(3), 11-48.
- Santos, L. & Pinto, J. (2008). The teacher's oral feedback and learning. *Proceedings ICME11*. Monterrey, México.
- Santos, L., & Pinto, J. (2009). Lights and shadows of feedback in mathematics learning. In M. Tzekaki, M. Kaldrimidou, & M. Sakonidis (Eds.), *PME 33* (vol. 5, pp. 49-56). Thessaloniki, Greece: PME.
- Santos, L., & Pinto, J. (2010). The evolution of feedback practice of a mathematics teacher. In M. M. Pinto, & T. F. Kawasaki (Eds.), *PME 34* (vol. 4, pp. 145-152). Belo Horizonte, Brazil: PME.
- Santos, L. & Pinto, J. (2011). Is assessment for learning possible in early school years? *Procedia Social and Behavioral Sciences*, 12, 283-289.
- Santos, L., Pinto, J., Rio, F., Pinto, F., Varandas, J., Moreirinha, O., Dias, P., Dias, S., & Bondoso, T. (2010). *Avaliar para aprender. Relatos de experiências de sala de aula do pré-escolar ao ensino secundário*. Porto: Porto Editora e Instituto de Educação, Universidade de Lisboa.
- Santos, L., & Semana, S. (2015). Developing mathematics written communication through expository writing supported by assessment strategies. *Educational Studies in Mathematics*, 88(1), 65-87.
- Seibel, C., & Levasseur, J. (2007). *Les effets nocifs du redoublement précoce*. Paris: Haut Conseil de l'Éducation.
- Semana, S. (2008). *O relatório escrito enquanto instrumento de avaliação reguladora das aprendizagens dos alunos do 8.º ano de escolaridade em Matemática*. (Tese de Mestrado, Universidade de Lisboa)
- Semana, S., & Santos, L. (2009). Estratégias de avaliação na regulação das aprendizagens em matemática. *XIX SIEM* (CD-ROM). Viana do Castelo: Associação de Professores de Matemática.
- Semana, S., & Santos, L. (2012). A comunicação oral numa discussão matemática em grupo-turma: O papel da professora. In A. P. Canavarro, L. Santos, A. M.

Boavida, H. Oliveira, L. Menezes, & S. Carreira (Eds.), *IEM 2012. Práticas de ensino da Matemática* (pp. 307-320). Lisboa: SPIEM (http://spiem.pt/DOCS/ATAS_ENCONTROS/2012/Atas_EIEM_2012.pdf)

Semana, S., & Santos, L. (2013). Teaching practices to enhance students' self-assessment in mathematics: Planning a focused intervention. In B. Ubuz, C. Haser & M. A. Mariotti (Eds.), *Proceedings of CERME 8* (pp. 3257-3266). Ankara: Middle East Technical University.

Shirvani, H. (2009). Examining an assessment strategy on high School Mathematics achievement: Daily quizzes vs. weekly tests. *American Secondary Education*, 38(1), 34-45.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: a typology. *British Educational Research Journal*, 22(4), 389-404.

Vial, M. (2012). *Se repérer dans les modèles de l' évaluation. Méthodes, dispositifs e outils*. Bruxelles: De Boeck.

Weiss, J. (1996). Évaluer plutôt que noter. *Revue Internationale d'Éducation*, 11, 19-31.

William, D. (1999). Formative assessment in mathematics. *Equals: Mathematics and Special Educational Needs*, 5(3), 8-11.

William, D. (2007). Keeping learning on track. In F. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1053-1098). Charlotte: Information Age Publishing.

William, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, 11(1), 49-65.