

2

Comprender que avaliar não é medir, mas confrontar em um processo de negociação

(Balanço dos saberes produzidos pela pesquisa)

A AVALIAÇÃO NÃO É UMA MEDIDA. A PROVA PELA NOTAÇÃO

Do caráter aleatório na notação

O que é uma medida? Segundo J.-P. Guilford, citado por Gilbert De Landsheere (1976, p. 70), medir significa atribuir um número a um acontecimento ou a um objeto, de acordo com uma regra logicamente aceitável. Isso implica que o objeto, ou o acontecimento, possa ser apreendido sob uma única dimensão, isolável, capaz de receber uma escala numérica (Bonniol, 1976). A medida é assim uma operação de descrição quantitativa da realidade. Mas a avaliação, pelo menos em sua forma dominante de prática de notação, não equivale precisamente a atribuir números a coisas? Aí está a origem da ilusão: aparentemente, há identidade formal entre as operações de medida e de notação. E a idéia de que a avaliação é uma medida dos desempenhos dos alunos está, como já vimos, solidamente enraizada na mente dos professores... e, frequentemente, na dos alunos. Mas observemos as coisas mais de perto. Uma medida é objetiva no sentido de que, uma vez definida a unidade, deve-se ter sempre a mesma medida do mesmo fenômeno. Certamente, um erro é sempre possível, devido às imperfeições da instrumentação, pois ele resulta então das condições de operacionalização dos instrumentos. Ele provém da própria operação de medida. Por essa razão, pode-se calculá-lo e, portanto, neutralizá-lo. O que acontece no domínio das notações? Vejamos o caso de um objeto a “medir”: um trabalho de aluno. Onde está o instrumento? Só pode ser a pessoa do corretor. Ora, é claro que esse instrumento não é confiável. Buscando várias medidas de um mesmo trabalho, verificou-se que, com frequência, as notas divergiam muito amplamente. Os primeiros trabalhos de docimologia forneceram exemplos abundantes disso. No sentido etimológico, a docimologia é a ciência (*logos*) dos exames, ou da medida por exame (*dokimê*). Isso foi, portanto, percebido pela docimologia, de saída, como um ato de medida. Mas os traba-

Ihns empreendidos pela docimologia iam justamente demonstrar que a idéia de partida, segundo a qual o exame é uma medida, revela-se no mínimo muito discutível. Como lembra Maurice Reuchlin (1971, p. 213), “os resultados obtidos no domínio da crítica experimental dos procedimentos de avaliação são totalmente convergentes e foram muitas vezes verificados” desde as primeiras pesquisas realizadas por Henri Piéron (1963).

Assim, deu-se atenção às notas atribuídas no exame do *baccalauréat*.^{*} Uma comparação das médias de notas atribuídas em 1955, em uma mesma matéria, por 17 bancas de um lado (habilitação em filosofia) e 13 de outro (habilitação em ciências matemáticas) revela “largas diferenças” (Reuchlin, 1971). Essas médias flutuam, por exemplo, de 5,81 a 9,06 na prova escrita de matemática, e de 8,30 a 13 na prova oral de física. Conseqüência: as porcentagens de candidatos aprovados são muito diferentes de uma banca à outra. Em ciências matemáticas, essas porcentagens variavam, naquele ano, de 31 a 53%!¹

Esses resultados são antigos. O que ocorre 40 anos mais tarde? A consideração de uma série de dados, no âmbito de uma pesquisa sobre a questão da pertinência da especialização professoral (enquanto produtora de julgamentos professorais) leva Pierre Merle (1996, p. 217) a concluir que a incerteza do julgamento professoral relativa, assim, à avaliação do conjunto das provas do *baccalauréat*, concerne a mais de um terço dos candidatos, o que corresponde mais ou menos à porcentagem de alunos que faz os exames orais de recuperação nessa prova final. Em outras palavras, para 30% dos candidatos, a aprovação depende do acaso da distribuição das bancas. O problema foi abordado de frente pelo reitor da Academia de Lille em 1984. O índice de aprovação no *baccalauréat* em sua academia^{**} era, com efeito, inferior à média nacional. Isso significava que os alunos eram efetivamente menos bons, ou que os corretores eram mais severos em Lille do que no resto da França? Essa segunda hipótese choca, de uma certa maneira, o senso comum: como imaginar que as divergências professorais sobre o valor dos alunos caminham na mesma direção? Elas não deveriam se anular? Pode-se conceber um “comportamento coletivo que tome a forma de uma severidade ou de uma indulgência comum a cada corretor?” (Merle, 1996, p. 218).

* N. de T. O *baccalauréat* é um exame realizado em nível nacional e que sanciona os estudos secundários, composto por provas orais e escritas feitas em duas etapas: ao final do penúltimo e do último ano do secundário.

** N. de T. A Academia é uma circunscrição administrativa do ensino na França; contam-se 25 na França metropolitana.

Essa hipótese questiona a competência avaliada individual dos professores e, ao mesmo tempo, a equidade do exame em nível nacional.

Como a questão foi levantada por um reitor, decidiu-se todavia testá-la, principalmente no que diz respeito à prova de ciências econômicas e sociais do *baccalauréat B*, submetendo ao julgamento de corretores de outras academias trabalhos de ciências econômicas e sociais corrigidos pela primeira vez em Lille. Dentre os trabalhos de seis bancas representativas da academia de Lille, sortearam-se 120 trabalhos, para submetê-los a quatro bancas de Lille, de Nancy e de Poitiers, segundo modalidades que permitiriam um estudo comparativo. As duas bancas compostas apenas por examinadores de Lille dão uma média sensivelmente idêntica (7,74 para um, 7,80 para o outro). Pode-se então falar efetivamente de uma “norma acadêmica implícita”. Os corretores de Nancy chegaram a uma média de 8,81, os de Poitiers, 9,05. Então, são bem mais severos em Lille. A variação máxima (9,05 – 7,74) é de 1,31 pontos, o que leva a 5,24 pontos, considerando o coeficiente da prova (4). É suficiente para não ser “passado” após o primeiro grupo de provas. Para candidatos de mesma competência, a probabilidade de aprovação no *baccalauréat* é maior em Poitiers do que em Lille. Chega-se à mesma conclusão de 1955: se as dissertações de ciências econômicas e sociais dos candidatos de Lille tivessem sido corrigidas em Poitiers, “a porcentagem de aprovados definitivos depois das provas escritas teria passado de 46,21 a 51,26%” (Merle, 1996, p. 221), ou seja, a probabilidade de aprovação dos candidatos varia de acordo com as regiões.

Um outro fato pode dar uma idéia da dimensão aleatória da aprovação introduzida pela existência de normas implícitas nas bancas; agora no caso dos concursos de recrutamento de professores.

Em 1989, a banca do CAPES* de ciências matemáticas admitiu 1.111 candidatos para 1.343 postos, provendo assim 82,7% das vagas oferecidas. Em letras modernas, a banca admitiu 708 candidatos para 1.060 postos, provendo assim 66,7% das vagas. E, em espanhol, 245 candidatos foram aprovados para 488 postos oferecidos, o que corresponde a 50,2% das vagas. Segundo o ministro da época, a banca de espanhol foi “excessivamente seletcionista” (*Le Monde*, 6 de setembro de 1989). Trata-se certamente de disciplinas diferentes. O nível dos candidatos pode, por outro lado, flutuar “objetivamente” de aluno para aluno. Mas existe “objetivamente” uma diferença de nível e de competência entre os

* N. de T. O CAPES é um exame que expede o Certificado de Aptidão à Docência do Ensino Secundário.

candidatos de espanhol e de ciências matemáticas de modo que se preencha 50,2% das vagas em um caso e 82,7% no outro? O ministro não será sem dúvida o único a falar, no caso do espanhol, de uma banca severa demais...

O que vale para as bancas vale, *a fortiori*, para os examinadores. Quando se comparam as notas distribuídas por corretores diferentes, constata-se primeiramente que o espaço de dispersão é diferente (alguns notam de 1 a 19, outros de 6 a 13, por exemplo*) e que as médias, assim como para as bancas, também são diferentes. Pode-se distinguir desse ponto de vista examinadores severos (média baixa) e indulgentes (média mais elevada).

No *baccalauréat* de 1996, uma examinadora de inglês foi destituída e substituída durante o exame devido à “notação severa demais”. A média das primeiras notas que havia atribuído na prova oral de inglês a alunos de série tecnológica era 8, com notas de 1 a 13 sobre 20 (*Le Monde*, 7-8 de julho de 1996). E não bastará, para corrigir isso (pois foi preciso se decidir a corrigir as divergências dos corretores!), ajustar as distribuições de notas buscando, por exemplo, as notas dos candidatos julgados por examinadores ou bancas muito severos, de maneira a alcançar uma harmonização das médias, visto que isso não corrige as diferenças nas classificações operadas pelos diferentes corretores. Na experiência lembrada por Pierre Merle, “o desacordo entre os corretores sobre o ‘valor’ do trabalho é, em quase um trabalho entre dois, igual ou superior a três pontos” (1996, p. 222). Seria então necessário não somente harmonizar as médias das bancas, mas também aquelas de todos os corretores de cada disciplina. E isso não seria absolutamente uma garantia de maior equidade, já que o obstáculo mais sério permanece: o da medida “objetiva”, por um corretor, do valor de cada trabalho.

As pesquisas sobre os resultados obtidos no *baccalauréat* mostram, pois, que a aprovação deve-se, em parte, ao acaso da atribuição a uma banca. Desse primeiro ponto de vista, o exame não é uma ciência exata. Mas isso revela, como mostra o último exemplo já apresentado, que o que se poderia chamar de “exame” individual, de um trabalho-produto, por um examinador que quisesse (quem poderia duvidar disso?) ser o instrumento de uma justa medida, não é confiável. Efetivamente, as variações de notas, para um mesmo produto, de um examinador a outro, vão bem além do que seria apenas uma incerteza normal devido às condições “locais” da tomada de medida. Todas as experiências de multicorrecção — e elas foram numerosas — demonstram isso abundantemente. Nós mesmos (Hadjji,

1989) citamos inúmeros exemplos. Ver também os trabalhos de Piéron (1963), de De Landsheere (1976), Noizet e Caverini (1978). Contentemo-nos com dois exemplos, mais recentes.

Primeiro exemplo: o mesmo trabalho de ciências econômicas e sociais, n.º 19064, recebe 15/20 em Lille e 5/20 em Nancy. Estima-se em Lille que ele fez “uma utilização bastante boa dos documentos difíceis” e, em Nancy, que “os documentos são mal explorados” (Merle, 1996, p. 222-223). Inversamente, um mesmo trabalho pode receber 8/20 em Lille e 16/20 em Nancy.

Segundo exemplo: nas provas de 1994 do *baccalauréat*, uma jovem, aluna do último ano do secundário, da escola Henri V em Paris, tira na prova escrita de filosofia a nota 1 sobre 20. Ora, naquele ano, ela fora premiada no concurso geral de filosofia (séries B, C, D, E) tirando... o segundo prêmio! O jornalista que relatou esse fato conclui: “A roda da loteria... continua girando” (*Le Monde*, 10-11 de julho de 1994).

Quer sejam disciplinas literárias (o que, para o senso comum, poderá parecer menos surpreendente) ou científicas (o que parecerá mais chocante aos meios!), os resultados das experimentações vão sempre na mesma direção. O que concluir disso, senão que:

a) se é realmente uma operação de medida, o instrumento de medida carece totalmente de confiabilidade. Será então muito difícil, até mesmo impossível, reverter a situação;

b) trata-se provavelmente de algo bem diferente que de uma operação de medida.

A impossível reforma do instrumento-avaliador

Foi buscando, no eixo da primeira hipótese, tornar o instrumento (isto é, não esses queçamos, o corretor no caso da notação) mais confiável, que se compreendeu como a segunda hipótese era pertinente. Com efeito, para melhorar o instrumento, é preciso corrigir seus defeitos. Quais são eles?

— A primeira resposta resume-se a uma palavra: a subjetividade do corretor. Pode-se pensar aqui em suas mudanças de humor, em suas preferências imediatas, em suas “paixões” (no oral), em suas liberalidades mais ou menos repressíveis (*Le Monde*, 5 de julho de 1990: um professor acusado de ter “dado uma nota mais alta” a um aluno, cujo estilo e caligrafia reconheceu em um dos trabalhos do exame de técnico superior), mas também em suas dúvidas, em seus momentos de cansaço, de “melancolia”. Querer neutralizar tudo isso constitui certamente um vasto programa...

*N. de T. No sistema educacional francês, as notas são calculadas sobre 20 pontos.

– Acreditar (ou esperar) em uma possível neutralização desse fator evidente de erro só tem sentido quando se acredita na objetividade de algum modo natural do julgamento professoral que se expressa durante a operação de avaliação. Deplora-se ainda mais as derivas da subjetividade quando vêm perturbar e perverter o jogo de um julgamento professoral que, liberado desse fator de perturbação, deveria poder desenvolver-se de modo satisfatório. Pelo menos quando se aposta na “certeza avaliativa” (Merle, 1996, p. 214), isto é, quando se acredita que o corretor, em sua essência de examinador puro, liberto do que corrrompe essa essência, é uma máquina de medir objetivamente os trabalhos. Que peso tem essa convicção, essa crença na existência de uma especialização professoral (dissemos “natural”, mas que poderia ser construída por uma formação profissional) tal que o professor-notador-especialista poderia quase infalivelmente (com a condição de dar provas de senso moral e de não se deixar perturbar por nada de fora: volta-te para ti mesmo, examinador...) dizer a verdadeira nota de cada trabalho?

Acabamos de ver que “a especialização professoral é um julgamento em parte aleatório” (Merle, 1996, p. 217). Os trabalhos de Pierre Merle mostram que esse julgamento é sempre infiltrado por elementos provenientes do contexto escolar e social, desde a carga afetiva e a dimensão emocional devido à presença efetiva dos alunos (em aula, ou na prova oral), até as representações dos alunos (o modelo do bom aluno) ou do sistema escolar durante a correção de um trabalho anônimo, ou de um conselho de classe. O julgamento professoral ignora, em geral, que baseia-se em parte em uma representação construída do aluno e em convicções íntimas que nada têm de científicas! As práticas avaliativas dos professores são orientadas por “uma história escolar e social singular” (p. 208). E conforme as ponderações sociais, as estratégias de precaução, o jogo duplo do professor (treinador-guia e árbitro-selecionador), a manutenção da disciplina no dia-a-dia, a avaliação torna-se o “produto de interações múltiplas” (p. 143), que expressa e traduz contradições. O julgamento professoral inscreve-se, pois, em “uma construção social em movimento” (p. 233) do duplo ponto de vista da história do professor e daquela da classe e depende muito do contexto escolar e social. Ele implica sempre uma parcela de interpretação, socialmente condicionada. Assim, fortemente dependente do contexto social, está sempre às voltas com “vieses sociais”.

O que pensar de um instrumento de medida sensível ao contexto social da medida? Fica claro que é inútil insistir em tornar a avaliação tão objetiva quanto uma medida. É por essa razão que todos os procedimentos de redução de divergências das notas revelam-se pouco eficazes em alcançar seu objetivo. A maior parte deles não passa de procedimentos de moderação estatística, que conseguem apenas atenuar os efeitos mais visíveis da incerteza da correção, sem combater a própria incerteza. Tratam-se os sintomas sem atacar o mal, quando se ajustam a

posteriori médias pela modificação de um conjunto de notas próprias aos diferentes corretores por referência a uma variação padrão. Neste duplo caso, fica-se, de algum modo, na superfície das coisas. Os procedimentos *a priori* (acordo entre corretores, construção e/ou utilização de uma tabela comum) também não são muito mais eficazes, o que é particularmente decepcionante, quando se trata da operacionalização de uma tabela, da qual se podia esperar muito. Mas a maioria das pesquisas mostra que isso não muda fundamentalmente as coisas. Alguns estudos até mesmo revelaram que em ciências matemáticas, por exemplo, é sem tabela imposta que as variações de nota são menores! O que compreender? Pierre Merle, que confirma a constatação (“a utilização de uma tabela de notação por ponto, até por meio ponto, não constitui uma garantia de precisão da correção”, 1996, p. 225), analisa o mecanismo produtor desse fenômeno tão decepcionante. Se as variações de notação podem ser muito limitadas sobre questões que valeram três pontos, uma grande incerteza pode, paradoxalmente, afetar a notação de uma questão que vale um único ponto. O que, no total, produzirá variações consideráveis. Em seguida, a tabela é sempre objeto de interpretação e de aprendizagem. Em terceiro lugar, as imposições que pesam sobre a correção, particularmente o tempo, levam os professores a tomar liberdades acerca da tabela. Enfim, o refinamento da tabela não aumentará a precisão da notação a não ser que haja consenso entre os corretores sobre o que é exatamente esperado do aluno. (p. 226 e 248). É a incerteza sobre o objeto da medida que é a causa essencial das distorções persistentes. A tabela continua ainda na superfície das coisas. Para progredir realmente, será necessário questionar a natureza exata da relação avaliador/avaliado. Não se pode mais considerá-la como aquela de um instrumento de medida com um objeto medido. O instrumento é demasiado incerto, e o objeto, demasiado vago.

Com efeito, o que se quer “medir” exatamente? O trabalho? Mas o que há a “medir” no trabalho, objeto multidimensional? Se o corretor não pode ser entendido e corrigido, isso talvez ocorra também e sobretudo porque o objeto sobre o qual se debruça e que gostaria de apreender não é realmente apreensível. O caráter vago que cerca esse objeto chamou a atenção da maioria dos observadores. Que o leitor se interrogue sobre o que se “mede”, exatamente, por meio de um exercício de matemática, ou de uma redação. Um saber? Um *savoir-faire*? Uma aptidão? Uma competência? Uma capacidade? Efetivamente, o trabalho não é o verdadeiro objeto da avaliação. Ele é apenas um pretexto, uma ocasião, para apreender algo que é próprio ao aluno-produtor, e que se espera ter construído e/ou desenvolvido nas seqüências de ensino que precederam. Designar esse “algo” não é evidente. No entanto, será preciso, se quisermos tornar a avaliação menos incerta, tentar fazê-lo. Vê-se que a melhoria passará aqui por uma melhor preparação

da prova... por parte do professor. Será necessário fazer o esforço de designar o objeto, de modo a poder, como escreve Pierre Merle (1996, p. 251), "construir normas de competências dos candidatos que não sejam intrinsecamente objeto de dúvida quanto à sua significação avaliativa".

Assim, a progressão para a objetividade depende, primeiramente, de uma especificação do objeto da avaliação, de maneira a relacionar coerentemente o objeto avaliado e o exercício de avaliação. Para isso, será preciso construir, no ato de avaliação, o objeto de avaliação. Tentaremos mostrar como fazê-lo, na segunda parte. Registremos aqui o fato de que hoje se sabe que a avaliação não é uma medida pelo simples fato de que o avaliador não é um instrumento, e porque o que é avaliado não é um objeto no sentido imediato do termo. Todos os professores-avaliadores deveriam, portanto, ter compreendido definitivamente que a noção de "nota verdadeira" quase não tem sentido. Henri Piéron forneceu dados sobre isso: se pudéssemos considerar a avaliação como uma medida, a fim de neutralizar os erros de cálculo, cada trabalho de matemática deveria ser corrigido por 13 corretores, e cada dissertação de filosofia, por 127 pessoas! (Piéron, 1963, p. 23). Jean-Jacques Bonniol (1976) retomou e corrigiu esses cálculos: seriam necessários 78 corretores em matemática e 762 em filosofia. Demonstração pelo absurdo: é ridículo e vão-se desgastar para melhorar a objetividade da notação aumentando, até a obsessão, o número dos corretores. A multicorreção não é o caminho correto. Um tratamento apenas quantitativo em nada muda o fundo do problema. Deve-se requestionar a natureza e o sentido da atividade de avaliação. Ela não é uma medida. Mas, então, em que consiste?

A avaliação é um ato que se inscreve em um processo geral de comunicação/negociação

Poder-se-ia dizer que a avaliação está para a medida assim como o *Canada Dry* está para o álcool, referindo-se à publicidade dessa bebida. * A nota citrada tem a aparência de um resultado de medida. Ela se apresenta sob a mesma forma (quantitativa). Tem seu odor, sua cor. Todavia, o avaliador não é um instrumento de medida, mas o ator de uma comunicação social. A avaliação é primeiramente "problema de comunicação" (Weiss, 1991). Hoje em dia, sabe-se qual a "incidência das situações e dos contextos sociais sobre a avaliação". Compreendeu-se que a avaliação é "uma interação, uma troca, uma negociação entre um avaliador e um

avaliado, sobre um objeto particular e em um ambiente social dado" (Weiss, 1991, p. 6). Jean Cardinet fala, a esse respeito, no âmbito de uma "abordagem sociocognitiva", de um "modelo qualitativo", que se oporia ao "modelo da medida", característico da "abordagem psicométrica" (Cardinet em Weiss, 1991, p. 199-200). Mas "qualitativo" não significa que esse modelo resulte de uma análise não científica da realidade. Bem ao contrário; ele se encontra na confluência de análises provenientes da psicologia cognitiva, da psicologia social e da teoria dos sistemas, que chegam a uma "extraordinária convergência de visões" (p. 208) no que tange à descrição do processo avaliativo e, a partir daí, à identificação das condições de possibilidade de uma avaliação formativa. O tema dominante de todos os pesquisadores é que a avaliação é socialmente condicionada, o que pode ser visto quando se corta o cordão umbilical que liga os dois principais atores do processo (apenas para as necessidades da análise pois, como mostrará Yves Chevallard, deve-se necessariamente levar em conta os dois comportamentos ao mesmo tempo, o do professor e o do aluno, e manter juntas as duas pontas da cadeira), tanto no âmbito do aluno quanto naquele do professor-avaliador.

Para o aluno, o desempenho depende do contexto

Analisando os trabalhos de pesquisadores dos campos científicos anteriormente mencionados, Jean Cardinet ressalta que, contrariamente ao que postula a psicometria, o desempenho no exame não é um fato bruto, um dado imediato, que forneceria o valor escolar do aluno. Se é assim impossível atribuir um valor, ou um nível, a um aluno da mesma maneira que se atribui um número a um objeto, é porque o desempenho concreto é, em grande parte, resultado da interação com o professor (em situação de aula), com o examinador (em situação de prova cumulativa), ou, de uma maneira mais geral, com uma situação social que exige que se mostre as capacidades no exame, de acordo com uma fórmula que certos candidatos no *baccalauréat* conhecem bem. O desempenho é, portanto, tanto função das circunstâncias quanto do próprio aluno (de seu valor escolar).

Jean Cardinet propõe que se considere que o desempenho observado é uma função com muitas variáveis (Cardinet em Weiss 1991, p. 210). Sem dúvida, mas em uma proporção difícil de avaliar, função desse valor escolar "objetivo". Mas igualmente função da história das interações do aluno com o professor, e até mesmo, em última hipótese, de toda a história escolar do indivíduo (o mesmo para o professor!); e função da formulação do problema que lhe é proposto no exame.

De fato, segundo uma interrogação feita por Jean-François Perret e Martine Wirthner, "Por que o aluno pensaria que uma questão pode esconder uma outra?"

* N. de T. Bebida sem álcool, vendida na França.

(em Weiss, 1991, p. 137). As práticas avaliativas apresentaram-se fundamentalmente como trocas de questões e de respostas, no decorrer das quais se instaurou um certo número de mal-entendidos sobre, no que diz respeito ao aluno, o sentido das questões e sobre o que o professor espera. Para o aluno, pode ser difícil distinguir os momentos de aprendizagem e os momentos de avaliação. Dessa forma, o erro é permitido no primeiro caso. Sancionado no segundo.

Além das ambigüidades inerentes à apresentação dos problemas e à formulação das questões (o que torna necessário um grande esforço de explicitação e de simplificação por parte dos professores-avaliadores), a criança frequentemente se equivoca sobre a própria intenção do adulto. Ele quer que se faça cuidadosamente um desenho geométrico ou que se contente em raciocinar? Ele quer respostas: mas por que então não as deu inicialmente? Quer que se extraia de um texto, como idéia principal, a idéia mais manifestamente presente, a mais original, ou a mais rica em informações? O aluno deve então adivinhar o que o professor espera e decodificar suas expectativas implícitas, o que exige “*savoir-faire* sociais” (Cardinet em Weiss, 1991 p. 203). Assim, “a criança, ator inconsciente de um roteiro que não conhece, esforça-se para dar sentido à situação, baseando-se em suas experiências anteriores. Para ela, ser interrogado é bem mais do que resolver um problema. É ser confrontado com uma situação social que deve ser bem compreendida primeiramente” (p. 203).

Os processos psicossociais complexos postos em jogo pela avaliação dependem não somente dos conteúdos em jogo, mas também das condições sociais da própria prova de avaliação. Por essa razão, com o mesmo “equipamento intelectual”, (p. 205), um aluno poderá produzir outros desempenhos se o contexto social for diferente. Jean-Marc Monteil mostrou muito bem, nesse sentido, qual o peso, sobre o desempenho, das inserções sociais e das atribuições de valor. Os alunos têm comportamentos diferentes em situação de anonimato ou de visibilidade; e seus desempenhos também mudam se estiverem em situação de comparação social (quando a existência de diferentes níveis é publicamente proclamada) ou não (quando declara-se aos indivíduos que todos têm o mesmo nível).

Na condição com comparação social, os bons alunos dão-se melhor em situação de visibilidade do que de anonimato: nesse caso, eles se “superam”. Quanto aos maus alunos, é o inverso, como se, em situação de visibilidade, eles fracassassem conforme o hábito “social”; ao passo que em situação de anonimato, ainda que saiba-se publicamente que são maus alunos, são capazes de ter êxito, o que revela a existência das competências necessárias ao êxito, ainda que normalmente fracassem!

Na condição sem comparação social, de fato, seja qual for a situação (anonimato, visibilidade), os bons alunos têm êxito e os maus, desempenhos fracos. A esses “resultados perturbadores” (Monteil, 1989, p. 180), pelo menos no que diz respeito à condição de comparação, acrescentam-se outros.

Bons alunos podem obter, na mesma prova, resultados diferentes conforme aprovados ou reprovados, publicamente, mas ao acaso, em uma prova anterior. Pode-se então “fabricar” êxito ou fracasso e produzir, a partir disso, resultados tão “perturbadores”. Os alunos a quem se disse terem sido reprovados têm êxito em situação de anonimato, mas fracassam em situação de visibilidade: ao passo que aqueles a quem se disse terem sido aprovados... fracassam em situação de anonimato (e, no entanto, são bons alunos!), mas têm êxito em situação de visibilidade.

O que concluir disso? Em situação de visibilidade social, os desempenhos estão de acordo com as atribuições sociais: o sujeito inscreve então sua conduta “no sistema de expectativas engendradas pela escola” (Cardinet em Weiss, 1991, p. 182). Em situação de anonimato, eles vão no sentido oposto às sanções atribuídas, como se o anonimato “corrigisse” o que havia sido socialmente fabricado. A conclusão que J.-M. Monteil tira disso é clara: “os desempenhos cognitivos do sujeito não são [...] independentes das condições sociais nas quais eles são realizados” (p. 183). De um lado, as atribuições sociais de valor – que correspondem a um fenômeno de categorização – determinam, em parte, o êxito ou o fracasso. De outro, as situações de visibilidade social acentuam o peso dos fatores sociais. Assim, são evidenciados dois fatos fundamentais. A importância, para o aluno, do contexto social da avaliação, mas também o impacto desta sobre o futuro do aluno. Se podemos gerar fracasso escolar por construção experimental assim tão facilmente, compreende-se qual a “importância considerável das categorizações iniciais e o impacto dos lugares e posições sociais atribuídos a um sujeito sobre seus comportamentos posteriores” (p. 192). Vamos ver que análises de Noizet e de J.-P. Caverni colocaram paralelamente em evidência a importância desse impacto sobre os avaliadores. Como um fenômeno reforça o outro, é muito fácil, pelo jogo das primeiras avaliações, treinar o aluno em uma espiral de fracasso... ou de êxito! Para pôr as avaliações a serviço das aprendizagens, uma regra essencial seria, portanto, jamais se pronunciar levianamente e contar até 10 antes de fazer um julgamento... sobretudo se for negativo!

Portanto, como lembra Jean Cardinet, se o desempenho observado pelo avaliador jamais constitui “uma base quase científica para fazer um julgamento sobre as capacidades das pessoas” (Cardinet, em Weiss, 1991, p. 205), isso acontece, entre outras razões, porque o aluno deve interpretar a situação de avaliação

para imaginar a intenção provável daquele que o interroga. Seu desempenho depende “estritamente” dessa interpretação (p. 204). Como essa interpretação exige *savoir-faire* sociais, um mau desempenho pode ser devido, não a uma falta de competência cognitiva ou de aptidão, mas a uma falta de *savoir-faire* social. O que confirma o caráter particularmente vago, já salientado, do objeto de avaliação. Quando se pensava apreender uma competência pessoal (competência que, além disso, seria quase natural aos olhos de certos professores), apreende-se frequentemente um construto social, cujas manifestações flutuam conforme o contexto social!

A percepção que o examinador tem do desempenho é igualmente dependente do contexto social

O efeito de categorização, de rotulação social, que vale para o aluno, também vale para o examinador. Isso foi muito bem evidenciado por notáveis pesquisas experimentais feitas por Noizet e Caverni. Partindo da constatação da imprecisão da notação tradicional, pesquisaram causas de divergência no que tange à psicologia dos avaliadores escolares, evidenciando o jogo de variáveis de situação. Para eles, as divergências se deveriam menos à subjetividade propriamente dita dos avaliadores (a qual, como já vimos, foi o primeiro inimigo da docimologia) do que às “condições” particulares nas quais eles trabalharam. Interessaram-se portanto pelos processos cognitivos especificamente postos em andamento pelo avaliador e capazes de sofrer o jogo de determinantes ou de mecanismos próprios à situação de ter de avaliar — determinantes e mecanismos independentes do avaliador. Mostraram assim efeitos de ordem ou de contraste devido à sequência dos trabalhos. Como uma infelicidade nunca vem desacompanhada, com efeito, os trabalhos em geral chegam aos corretores em pacotes inteiros... A ordem das correções revela-se importante: os primeiros trabalhos corrigidos são (salvo o primeiro) superavaliados, e aqueles corrigidos por último, subavaliados (efeito de ordem). Um mesmo trabalho é superavaliado ou subavaliado conforme segue imediatamente um trabalho muito ruim ou muito bom (Noizet e Caverni, 1978, p. 23). Poder-se-ia dizer que a existência de lotes acarreta efeitos de loteria.

Porém, o mais importante não está aí. Reside na revelação de efeitos de assimilação, fazendo com que variem as informações dadas a dois grupos de corretores equivalentes corrigindo o mesmo trabalho, constata-se:

— que o conhecimento das notas anteriormente obtidas pelo produtor do trabalho influencia a avaliação deste. Há assimilação de um desempenho aos desempenhos anteriores (Noizet e Caverni, 1978, p. 80);

— que, do mesmo modo, o conhecimento do estatuto escolar do aluno (nível forte, ou o melhor possível, fraco) influencia a correção. Os avaliadores têm tendência a assimilar a produção ao nível presumido de seu autor (p. 87);

— que a posse de informações referentes à origem socioeconômica do produtor do trabalho também influencia a correção, ainda que essa informação se reduza à origem de determinada escola. O mesmo trabalho recebe uma nota melhor quando é atribuído a um aluno de uma escola de excelência, recebendo uma nota mais baixa quando atribuído a um aluno de uma escola de periferia. Neste último caso, o corretor dá a nota em parte em função da idéia que tem do valor escolar do aluno, em razão de sua origem social, imaginada através da frequência a esta ou aquela escola!

Portanto, a avaliação é sempre influenciada pela consideração de informações *a priori*. Noizet e Caverni justificam esse fenômeno (sobre o qual o mínimo que se pode dizer, mais uma vez, é que é “perturbador”...) recorrendo à teoria da dissonância cognitiva. Há dissonância quando coexistem dois elementos cognitivos opostos (um acarretando o inverso do outro). Esse estado é penoso, e o sujeito que o conhece se esforçará consequentemente para restaurar a consonância. Quando há então conflito entre o que o corretor percebe “objetivamente” no trabalho e o que ficou sabendo do produtor, esse conflito deverá ser resolvido pelo triunfo de uma das duas fontes de informação (a percepção atual, ou então a informação socialmente atestada). Os efeitos de assimilação marcam a vitória (parcial) da informação socialmente atestada. O que revela muito sobre a força dos determinantes sociais. Mais uma vez, aparece a pregnância do que já foi socialmente julgado.

A avaliação escolar traduz arranjos em uma dinâmica de negociação

Será possível, porém, desfazer-se dessa influência social? Certamente, o conhecimento desse fenômeno pode constituir, para o corretor, um primeiro fator de liberação. Por esse motivo, era necessário lembrar aqui esses fatos. Todavia, as pesquisas mais recentes de Pierre Merle mostram que o julgamento professoral de avaliação, complexo e socialmente oblíquo, “não pode ser realizado fora de categorias professorais e de contextos sociais particulares necessitando a operacionalização de arranjos avaliativos específicos” (Merle, 1996, p. 102). Esses arranjos são o resultado de uma negociação, implícita ou explícita, entre um professor que quer manter sua turma, e alunos que querem alcançar seu objetivo, que é o de passar para o ano seguinte. Por outro lado, eles são o fruto de um confronto

com os julgamentos, reais ou supostos, produzidos pelos outros. No primeiro caso, o professor-avaliador arranja-se com sua turma; no segundo, o corretor arranja-se com os julgamentos socialmente atestados... e sua consciência. Pierre Merle mostrou muito bem como esses arranjos podem ter finalidade externa (a administração, os pais), interna (os alunos), ou podem ser em relação ao próprio avaliador (suas próprias exigências). São função da relação que o corretor mantém com a nota, de seu passado de aluno, da relação presente com o/ou os alunos e do nível médio da turma (Merle, 1996, p. 84-85).

Assim, aparece claramente uma dupla dimensão da operação de avaliação escolar. Ela é um ato de comunicação que se inscreve em um contexto social de negociação. Esta última dimensão já tinha sido bastante evidenciada por Yves Chevallard que, efetuando uma análise didática dos fatos de avaliação, conseguiu mostrar que o objeto adequado para compreender o funcionamento concreto da avaliação, no âmbito de um procedimento descritivo, não era o exame, mas a turma. Para Yves Chevallard, o modelo metrológico (a notação é assimilada a um ato de mensuração) deve ser substituído pelo modelo da transação (a notação faz parte de uma transação e inscreve-se em um processo de negociação didática). De fato, cada um em função de seu objetivo geral – para o professor, guiar a turma em uma progressão de saber e, para o aluno, modificar o percurso rumo a uma linha de menor dificuldade –, professores e alunos deverão negociar e transigir: as tácticas destes opoem-se à estratégia global daquele. Desse ponto de vista, o controle dos conhecimentos fixa os “cursos”. O professor não é uma máquina de dar notas, mas um ator que regateia, em função do valor essencial para ele, a média (que fixa o curso médio do saber na classe). As tácticas dos alunos desenvolvem-se em vista da construção de uma imagem correta, a imagem escolar na qual se fixará seu valor de aluno, ao menor custo. De modo que, se a prova de avaliação mede alguma coisa, é a “habilidade táctica do aluno” (Chevallard, 1986, p. 49) para construir a melhor imagem ao menor custo. O que recoloca muito claramente o problema do objeto “medido”, pois, nesse contexto didático, “a nota atribuída torna-se uma mensagem” (p. 55). Negociação e comunicação andam juntas. Por isso, o que a avaliação escolar precisa para progredir (para mais justiça e, ao mesmo tempo, mais objetividade) é, primeiramente, de um “contrato social” (p. 58).

Sem dúvida, definitivamente, o defeito principal do avaliador é sua demasiada sensibilidade aos fenômenos sociais. Mas essa sensibilidade não poderia, a rigor, ser neutralizada senão no âmbito de um “contrato social” que determinasse e fixasse as regras do jogo. Como diz Yves Chevallard, o importante é aprender a dominar as regras do jogo. A relação de ensino não escapa à lei de toda relação humana: ser moldada por regras do jogo (p. 49).

Mas então, devido à própria importância dessa dimensão social, é preciso, de uma vez por todas, libertar-se da alegoria da medida. A conclusão dessa segunda análise dos saberes agora construídos sobre o processo de avaliação reencontra aquela da primeira análise. O avaliador não é um instrumento de medida, mas um ator em um processo de comunicação social. Certamente, ele precisa dizer, apesar de tudo, o valor de um produto. Mas o próprio uso do termo *valor* adverte que não se está, propriamente falando, no domínio da medida. A exatidão da especialização professoral não passa de uma crença, desmentida pelos fatos. Isso nos leva a compreender que a avaliação, mesmo em sua forma mais rigorosa, aparentemente, de notação, diz menos sobre o “verdadeiro valor” de um objeto (expressão sem dúvida marcada por uma contradição interna) do que expressa outra coisa. Mas o quê?

A AVALIAÇÃO É UM ATO DE CONFRONTO ENTRE UMA SITUAÇÃO REAL E EXPECTATIVAS REFERENTES A ESSA SITUAÇÃO

Fica claro, a partir do que precede, que não se encontrará mais justiça, rigor e objetividade em um hipotético aperfeiçoamento do avaliador – instrumento que faria dele uma melhor máquina de dar notas –, mas em um ajuste do ato de avaliação a seu “elemento estrutural constante” (Chevallard, 1990, p. 23). De fato, a primeira maneira de avaliar bem é... avaliar. Mas o que é, precisamente, esse elemento estrutural constante? Para Yves Chevallard, há avaliação cada vez que “um indivíduo, Y (o aluno, o professor), produz um enunciado avaliativo acerca de um “objeto” O (aqui, este professor ou este aluno)” (1990, p. 23). Isso é totalmente verdadeiro. Mas o problema é saber o que caracteriza, essencialmente, um enunciado avaliativo e o que o distingue de qualquer outro enunciado. Para saber se é realmente possível, e em que condições, pôr a avaliação a serviço das aprendizagens, precisamos saber em que consiste esse enunciado, ou, de um modo mais geral, o próprio ato de avaliação. Hoje em dia, já se sabe.

Os trabalhos de Noizet e Caverni mostraram o caminho, privilegiando a noção de expectativa. De fato, a teoria da dissonância cognitiva só consegue realmente dar conta do jogo do mecanismo de assimilação com a condição de estar atento à importância das expectativas do professor. A posse de informações *a priori* sobre o aluno-produtor orienta suas expectativas acerca dele (ou acerca de seu trabalho). Essas expectativas vão perturbar sua visão do objeto, influenciando, por exemplo, a leitura que fez do trabalho. As informações *a priori* vão provocar uma “modificação na coleta dos indícios” (Noizet e Caverni, 1978, p. 144). A recusa da dissonância cognitiva vai incitar o avaliador “a coletar nesse trabalho apenas os indícios congruentes com sua expectativa” (p. 117). Este termo – indí-

cio — é capital para a compreensão do próprio ato de avaliação. De fato, Noizet e Caverni evidenciaram por outro lado a existência de efeitos próprios à dinâmica da coleta de informações. O avaliador opera uma “leitura” de seu objeto. Esta, no caso de um trabalho, leva tempo. Essa dimensão temporal tem consequências importantes.

Quando, por exemplo, incorreções foram inseridas na primeira parte de um trabalho, este recebe uma nota pior do que quando as mesmas incorreções foram inseridas na segunda metade do mesmo trabalho (salvo por esse único aspecto). Poder-se-ia dizer que há, então, efeito de assimilação às primeiras informações retidas pelo corretor. Se a primeira impressão é boa, ele espera ver na sequência mais coisas boas e... será mais atento e sensível ao que é bom daquilo que é menos bom. E reciprocamente. Em suma, o preconceito instala-se bem depressa. As expectativas são moldadas pelo que o avaliador tem em mãos: informações *a priori*, ou as primeiras impressões produzidas pelo trabalho do aluno.

Assim, essas primeiras informações ou impressões provocam inferências que guiam a coleta posterior dos indícios. A busca de consonância pode então se operar. Nessas condições, a avaliação é uma operação de levantamento seletivo de indícios. Os comportamentos de avaliação pertencem à categoria cognitiva dos comportamentos em que um levantamento de indícios leva a uma decisão (Noizet e Caverni, 1978, p. 67). O estudo experimental era precisamente sobre esse levantamento de indícios, cujas condições eram variadas. Verifica-se que *a avaliação é uma leitura influenciada por expectativas específicas referentes à produção de um produtor particular, em função do que se sabe, ou do que se descobre, progressivamente, sobre ele*. Efeitivamente, o levantamento de indícios é seletivo quando se considera, de um lado, a espessura, ou a densidade, do objeto a avaliar, que ultrapassa sempre, de certa forma, as possibilidades de apreensão do avaliador. Várias leituras de uma mesma produção ou de um mesmo comportamento são sempre possíveis, o que constitui um outro aspecto daquilo que chamamos de caráter vago do objeto. E ele é seletivo quando se considera, de outro lado, a dimensão temporal da maioria dos atos de avaliação. Um trabalho é lido do início ao fim. Isso dá o tempo necessário para que os primeiros indícios recolhidos possam guiar a busca posterior dos outros.

Assim, a avaliação apresenta-se, de certo modo, como uma operação que articula expectativas e indícios. Falar-se-á, após os trabalhos de Marcel Lesne (1984) e de Jean-Marie Barbier (1985), de um confronto referente/referido. De fato, avaliar um trabalho consiste exatamente em dizer o que ele vale. Pode-se, para tanto, tentar estabelecer uma correspondência entre esse trabalho e uma es-

cala de notas. Já se sabe que o valor assim atribuído expressa tanto um certo estado do “mercado” didático quanto o valor “intrínseco” do produto e que flutua em função das histórias pessoais e sociais do avaliador e do avaliado. Mas deve-se observar, por outro lado, que essa operação de atribuição de valor pode assumir certamente a forma de uma nota, mas também de um discurso (é bom; ele é muito bom), de um sorriso... ou de uma careta. A nota cifrada não passa de uma maneira entre outras de dizer o valor.

Contudo, o que exige, fundamentalmente, essa operação: dizer o valor? O que ela coloca em jogo? O que ela implica necessariamente? Observemos primeiramente que é uma operação através da qual se toma distância de uma realidade dada a fim de poder pronunciar-se sobre ela. O avaliador, ainda que esteja no coração da ação, está, de um certo modo, sempre em derivação, em retirada. Ele se afasta para observar (levantar indícios...) e julgar. Yves Chevallard (1990, p. 25) descreve o ato de avaliação como “uma declaração apreciativa de um indivíduo Y acerca de um ‘objeto’ O (que pode ser um indivíduo)”. A declaração do avaliador tem a função de dizer a verdade (*vere dictum*); ela tem a significação de “veredito” (p. 26). Sobre o que é esse veredito? Sobre a relação, dentro de uma instituição I, de um sujeito X, com um objeto O. Quando X é um aluno, Y um professor e O um objeto de saber, a avaliação dá o “veredito de Y acerca da relação de X com O” (p. 27). Mas como pode-se chegar a tal veredito? Sobre o que ele repousa? Yves Chevallard fornece uma chave essencial. A instituição I definiu em seu seio diversas “posições”, entre outras, aquela do aluno (e), aquela do professor (E), etc. Para cada objeto institucional O, a instituição faz emergir uma relação institucional com O para p, que será notada Rp (O). Essa relação institucional representa a relação normal, poder-se-ia dizer prototípica, e consequentemente valorizada, do sujeito X com o objeto O: “Rp (O) é a relação que cada sujeito X, em posição p dentro de uma instituição I, deveria manter de modo ideal com o objeto O” (p. 27, o grifo é nosso). Essa relação define, quando X é um aluno e O um objeto de saber, a posição do “bom” aluno. O que é então avaliado é “a adequação da relação pessoal R (X,O) com a relação institucional Rp (O)”. Traduzindo: o veredito traz uma resposta à pergunta: o aluno está em conformidade com o que a instituição pode esperar dele?

Yves Chevallard extrai duas consequências disso. Primeiramente, todo julgamento de avaliação é institucional. Para a avaliação, não há legitimidade senão institucional. “Todo ‘juiz’ fala e sustenta sua declaração enquanto sujeito envolvido com uma instituição” (p. 26). Neste caso: desejo de ver os alunos tornarem-se bons alunos com referência ao modelo institucional do bom aluno. Em segundo lugar, existe uma equivoquidade essencial dos vereditos. Yves Chevallard fala de caráter vago (eis novamente o termo...), de um

trêmto, de um não-determinismo dentro da própria instituição (p. 28), pois a relação institucional com a atividade de avaliação “não fornece um ‘algoritmo’ de comportamento, levando Y a um veredito univocamente determinado” (p.28). E, por outro lado, em uma outra instituição, outro veredito seria dado. De modo que todo veredito é marcado por uma “arbitrariedade que, aparentemente, nada pode reduzir totalmente” (p.28.).

Resumindo: a avaliação não é uma operação científica. A declaração do avaliador é sempre equívoca. Ela só tem legitimidade no seio de uma instituição. E – para nós, o fato fundamental – ela expressa a adequação (ou a não adequação) percebida entre a relação atual do aluno com o saber, objeto da avaliação, e a relação ideal do aluno com o saber, objeto do “desejo” institucional. É em nome dessa relação ideal que é declarado o valor do aluno.

A avaliação é uma operação de leitura da realidade

O termo “leitura” significa, aqui, que o avaliador aborda seu objeto como o leitor seu texto. Ele tem, na mente, conhecimentos e sobretudo questões correspondendo a expectativas, que determinam uma atitude “antes-da-leitura” (a atitude seletiva na coleta das informações extraídas do trabalho, descrita por Noizet e Caverini). Também tem, diante de si, um conjunto de sinais que deverá decifrar e dar sentido. Neste caso, declarar o valor é dizer que sentido se atribui à posição do aluno em relação à posição institucional ideal. O sentido não está presente no objeto antes do ato de avaliação. O objeto, ele sim, está presente. Mas será preciso pronunciar-se sobre uma relação com o que pode ser, nesta instituição, almejado, esperado. Uma leitura é sempre seletiva. O leitor levanta indícios na “estrutura de superfície” do objeto (o conjunto do que se mostra, da informação visual) para construir sentido em função de informações disponíveis em sua estrutura cognitiva. Leitura não é medida.

A avaliação, como toda leitura, é orientada

A avaliação, por essência, não pode ser objetiva. O imperativo de objetividade implica que se apreenda um objeto tal como ele é, e tudo o que for possível apreender desse objeto do ponto de vista da dimensão isolada. Não se trata de dizer a “verdade” do objeto, mas de dizer quanto (ele pesa, ele mede, etc.). Haveria para a avaliação um imperativo – impossível de satisfazer incondicionalmente, como mostra Yves Chevallard – de verdade. O discurso do avaliador deve ser “verda-

deiro” quanto à relação que ele enuncia de um aluno com um objeto de saber, porém sempre com referência a uma relação ideal entre um aluno ideal e o mesmo objeto de saber.

Em outras palavras: o avaliador tem sempre um pé fora do presente do ato de avaliação. Ele tem um pé no dever-ser. Ele sabe (ou deve saber) o que deve ser (deve saber fazer) o aluno. É em nome desse dever-ser, que representa o conteúdo de uma expectativa específica, que ele julga (aprecia) o desempenho atual do aluno. Avaliar não consiste pois, simplesmente, em medir esse desempenho, mas em dizer em que medida ele é adequado, ou não, ao desempenho que se podia esperar desse aluno. Em nome de quê? Do modelo ideal que orienta a leitura da realidade e que preside ao levantamento de indícios. Será em função desse modelo que se manifestarão expectativas precisas acerca do aluno.

A avaliação é uma leitura orientada por uma grade que expressa um sistema de expectativas julgadas legítimas, que constitui o referente da avaliação

Dizer que não pode haver avaliação sem critério significa que a leitura do objeto avaliado efetua-se necessariamente através de uma grade constituída pelas expectativas específicas concernentes a esse objeto. Não se poderia avaliar um objeto do qual não se esperasse nada. Neste sentido, a relação de avaliação é uma relação de não-indiferença com o objeto avaliado. As expectativas de quem? Daqueles a quem a instituição reconhece o direito de fixar a relação institucional do objeto-aluno com o objeto-saber. Mas como são vários a se atribuírem essa legitimidade, isso não ocorre sem problemas. É claro que cabe primeiramente a toda a comunidade falar, através das leis, instruções oficiais e programas, mas a palavra do inspetor também tem sua legitimidade. E, evidentemente, a do professor, que traça seu caminho nos programas e desenvolve estratégias próprias para conduzir bem a tarefa que lhe foi institucionalmente confiada: conduzir a turma em uma progressão de saber. Porém, também os pais e os alunos têm algo a dizer. Devemos então considerar várias coisas:

a) A avaliação exige a construção daquilo que foi designado como seu *referente*, ou seja, um conjunto de critérios especificando um sistema de expectativas. Cada critério define o que se julga poder esperar legitimamente do objeto avaliado.

b) O que percebemos como o caráter vago do objeto de avaliação tem sua origem, além da “espessura” e da opacidade do próprio “objeto” avaliado (objeto a decifrar, que sempre ultrapassa o que se pode dizer dele), na imprecisão e no caráter vago das expectativas. A leitura se faz graças aos critérios,

através deles. Se forem imprecisos, a leitura ficará necessariamente embaralhada (de onde vem o termo “trêmido” usado por Yves Chevallard). Uma necessidade absoluta, se assim pode-se dizer, para pôr a avaliação a serviço dos alunos é especificar seus critérios, seu sistema de expectativas. Ao menos para si mesmo, para saber o que se julga poder legitimamente esperar dos alunos que serão interrogados. Que eles saibam – o quê? – que tipo de saber (declarativo? procedural?). Que eles saibam fazer... no âmbito de que tarefa precisa? Com quais habilidades específicas? O que se vai poder designar em termos de capacidades, de competências, ou de outra coisa ainda? Vimos acima que o perigo das categorizações devia levar à prudência: jamais se pronunciar levanamente. Compreendemos agora que esse dever de prudência acompanha-se de um *dever de reflexão prévia*: levar o tempo necessário para identificar o mais claramente possível o que se julga precisamente poder esperar, e cuja “existência” vai ser precisamente “verificada” por essa prova de avaliação. Veremos mais adiante que pode ser muito sensato permitir ao próprio aluno que se aproprie desses critérios.

c) O ato de avaliar implica, neste sentido, bem mais do que objetividade, ou mesmo de verdade, um *imperativo de legitimidade*. Não realmente dizer o real (isso cabe à ciência), embora seja necessário pronunciar-se sobre um objeto real através dos indícios levantados; não realmente dizer o verdadeiro, embora o veredito deva traduzir o mais possível em sua verdade a adequação da relação pessoal à relação institucional com o objeto de saber; mas dizer o valor de um aluno ou de seu produto, mostrando como eles se situam em relação ao que é legítimo esperar deles.

d) Como o “se” (do “que se julga poder”) designa uma pluralidade de pesos (ou de instituições, dentro de uma instituição mais ampla), o referente de cada operação de avaliação deve ser construído a partir de um sistema plural de expectativas, selecionando, escolhendo (o porquê da negociação, implícita ou explícita) as expectativas a considerar prioritariamente para determinada avaliação. Isso acarreta duas consequências. Não pode haver espaço para subterfúgio. E como a triagem, a escolha, é sempre discutível, nenhum referente dispõe de uma legitimidade indiscutível. Dizer o valor é uma operação delicada devido a essa relatividade fundamental do referente. Desse ponto de vista, a avaliação levanta problemas de ordem ética e técnica. Sem que isso se torne paralisante, sempre há possibilidade de questionar a legitimidade de seu referente, de modo, por exemplo, a não exigir de um aluno o que manifestadamente ele não tem condições de produzir, em função dos saberes e do *savoir-faire* que domina no momento. Isso seria condená-lo a um fracasso e talvez começar a inscrevê-lo naquela espiral cuja força já vimos.

A avaliação é uma leitura que implica construção de um “modelo reduzido” do objeto avaliado, que será o referente da avaliação

Os critérios reunidos no referente constituem assim uma grade de interrogação para o objeto avaliado. Somente graças a ela, logicamente, será possível consagrar-se tecnicamente à operação de busca de indícios. Fala-se de indícios ou de indicadores para designar os aspectos do objeto avaliado, os elementos colhidos em sua realidade, e sobre os quais vai basear-se para se pronunciar sobre o modo como as expectativas são satisfeitas.

No domínio da avaliação escolar, esses dados (é assim que se chama, em geral, o que seria melhor chamar de “coletados”) são na maioria das vezes produzidos por ocasião de uma tarefa proposta ao aluno. É quando o aluno enfrenta essa tarefa, ou no produto de sua atividade, que se levantam os elementos observáveis que vão constituir “o objeto de trabalho” (Lesne, 1984, p. 133) da operação pela qual, segundo a fórmula de Jean-Marie Barbier, passa-se de uma “representação factual” de um objeto a uma “representação normatizada” do mesmo objeto (Barbier, 1985, p. 64). A busca dos observáveis é, portanto, um processo de produção de informação para a avaliação, ao que a avaliação não se reduz em nenhum caso. Quando há uma preocupação com os instrumentos da avaliação, pensa-se de modo prioritário nos instrumentos que vão permitir recolher as informações para a avaliação. Esses instrumentos serão diversos, em função dos tipos de dados possíveis. Pode-se pensar (cf. Barbier, 1985, p. 70) em dados já existentes (os resultados de avaliações passadas, as informações contidas em um boletim escolar); em dados produzidos por observação, seja aquela do comportamento “natural” do aluno (na aula), ou aquela de seu comportamento no decorrer de um teste: este é, como já dissemos, o caso mais frequente, pois a tarefa permite observar a atividade do aluno, ou seu produto (há então observação indireta do sujeito produtor através dos traços de sua produção); ou então, enfim, em dados produzidos por meio de uma entrevista com o aluno. Em todos os casos, esses dados constituem (ou permitem extrair) indicadores, que só indicam algo em referência ao critério. Assim:

a) O indicador jamais é um dado imediato, que bastaria apanhar, abaixando-se. É sempre o resultado de um trabalho de elaboração, de construção, feito à luz dos critérios. Um indicador não indica nada que não esteja relacionado a um critério.

b) É a coerência critérios/indicadores que importa antes de tudo. Nisso reside toda a objetividade que se pode esperar de um julgamento de avaliação. Ele deve dizer o valor, baseando-se nos sinais (os indicadores) mais característicos da

realidade (é preciso que os indicadores desvelem o próprio objeto), mas com referência às expectativas. Em outras palavras, o indicador deve ser, ao mesmo tempo, *representativo* da realidade avaliada (objetividade externa), e *significante* em relação a uma expectativa precisa (“objetividade” interna, ou coerência). A significação não está contida no indicador; ela não é uma propriedade dele. Ela nasce da proximidade com o critério. O indicador só pisca quando a grade de leitura lhe é superposta. Teremos oportunidade de precisar isso mais adiante.

c) Sendo o referido – aquilo a partir do que se poderá fazer o julgamento de valor – assim construído por meio do levantamento de indícios ou de indicadores (designa-se, portanto, pelo termo de referido o conjunto da informação que foi possível produzir para a avaliação), o ato específico de avaliação, que leva à produção de um julgamento de valor, consiste em relacionar um referido e um referente (Lesne, 1984, p. 132) para dizer em que medida a realidade apreendida por meio do referido está de acordo com o ideal que transparece no referente.

Tomemos um exemplo: posso legitimamente esperar dos alunos do ciclo das aprendizagens fundamentais (grande seção, CP, CE1*), do ponto de vista da construção dos conceitos fundamentais de espaço e de tempo (expectativa legítima em função das instruções oficiais: cf. *Les cycles à l'école primaire*, CNDP, Hachette, 1991, p. 33), que se situem em relação, primeiramente, a um passado e a um futuro próximos. Eis o referente (a relação institucional Ro [O] adequada do aluno X com o objeto de saber O). A relação pessoal do aluno X com esse objeto O (R [XO]) é adequada à relação institucional? Eis a questão levantada pela avaliação, a partir da qual vai-se organizar a leitura da realidade X (o aluno do ciclo das aprendizagens fundamentais).

Para produzir observáveis, poderei, por exemplo, submeter o aluno à seguinte prova: completar um calendário, ordenando diferentes etapas que correspondem ao passado e ao futuro próximos do próprio aluno (cf. *Aide à l'évaluation des élèves. Cycle des apprentissages fondamentaux*, Ministério da Educação Nacional, Direção de Avaliação e Prospectiva, volume 2, p. 44). Esta é a tarefa que deve permitir recolher informações para a avaliação. Trata-se de completar etiquetas vazias correspondendo a etapas importantes da escolaridade durante três anos, em uma faixa vertical em que esses anos estão representados. As respostas do aluno (exata esperada; exata parcial

sem elemento errôneo, errônea; ausência de resposta) constituirão o referido... que deverá ser interpretado em relação ao referente para formular um julgamento de valor do tipo: tanto quanto se pode julgar através deste exercício, tal aluno realizou perfeitamente o que se podia esperar dele. Ele sabe situar-se em um passado ou em um futuro próximos. Ou então: já este outro não possui, ou não domina senão parcialmente a competência visada. Observemos que se poderá sempre, atribuindo pontos aos tipos de respostas, traduzir isso em uma nota. Mas essa nota só terá sentido se não se perder de vista o que ela “traduz”: o grau de adequação de um comportamento cognitivo real a um comportamento cognitivo desejado.

Desse ponto de vista, o avaliador colocará a avaliação... primeiramente a seu serviço (para servir à sua função de professor), tornando-a informativa, por meio de uma explicitação o mais precisa possível do referente, e, a partir daí, uma coleta o mais pertinente possível dos dados que poderão então ser interpretados, em relação ao referente.

Esta nos parece ser a terceira aquisição fundamental dos trabalhos sobre a avaliação e que, em certo sentido, é a mais importante. Efectivamente, compreende-se que a avaliação não é uma medida (aquisição nº 1); que se inscreve em um processo de comunicação/negociação (aquisição nº 2), por referência à sua característica essencial: é uma operação de confronto, de correlação, entre expectativas e uma realidade (ou, no outro sentido, entre “o existente e o desejado ou o esperado” Lesne, 1984, p. 132). Essas expectativas são essencialmente sociais. É preciso estabelecer expectativas legítimas, o que nem sempre é evidente... pode exigir negociações. O julgamento de valor produzido destina-se a atores sociais (o próprio professor, o aluno, a administração, os pais, etc.), para quem o que se diz, nessa comunicação significa muito. O lugar que se conseguirá na sociedade será, em parte, função do valor escolar, apreciado e proclamado na escola. Apreciado, e não medido no sentido estrito, já que não se trata de medir um objeto, mas de dizer em que medida esse objeto corresponde a expectativas específicas sobre ele. A primeira maneira de pôr a avaliação a serviço dos alunos é, para o professor-avaliador, compreender tudo isso.

* N. de T. A Escola Maternal, na França, compõe-se de três anos: *petite, moyenne e grande section*. O ensino primário, de cinco anos: CP (curso preparatório), CE1 (curso elementar 1), CE2 (curso elementar 2), CM1 (curso médio 1) e CM2 (curso médio 2).