

L'ÉVALUATION DYNAMIQUE DES PUBLICS PEU QUALIFIÉS

Daniel Pasquier

Psychologue industriel, docteur en sciences de l'éducation, consultant, Avenir & Entreprise,
Saint Jean de Braye ; laboratoire PRIS, Rouen, France

BPS et DRTEFP, Toulouse

23 mars 2006

THEMES ABORDES

- La situation évaluative
- Evaluation pédagogique et docimologie
- Evaluation psychométrique et biais de réponse
- Evaluation dynamique
- Applications aux publics peu qualifiés

Ces thèmes partent du plus général au plus spécifique, ce qui est dit sur le plan général s'appliquant aux plans plus spécifiques.

La situation évaluative implicite et spontanée

Dans toute situation de face à face, lorsqu'on rencontre une personne pour la première fois, avant tout échange verbal, un processus évaluatif se met en route, plutôt sur un mode non-conscient. Par ce processus continu, en arrière-plan, on se forme une impression évaluative sur la personne, impression positive, « C'est quelqu'un de bien ! », ou bien impression négative, « Il -elle- est bizarre ! ».

Cette impression se construit par la synthèse d'un ensemble de perceptions visuelles, olfactives, auditives..., par le faisceau des impressions convergentes (Bourdieu), organisées progressivement en forme de scripts suite aux différentes expériences évaluatives vécues positivement ou négativement, tout du long de l'existence. Ces schémas évaluatifs classent l'autre en termes de distance à soi (familier vs étranger), de distance chronologique (jeune vs vieux), de distance hiérarchique...

Dans ce processus évaluatif spontané s'immiscent le jeu des stéréotypes (schémas évaluatifs partagés par une même groupe) et les risques de stigmatisation encourus par ce jeu.

On transmet à l'autre de manière discrète la façon dont on le classe via les diverses modalités du paralangage (mimiques, attitudes, postures, regard, intonation...). Cette transmission peut prendre un tour plus brutal, par le moyen d'énoncés verbaux plus ou moins agressifs

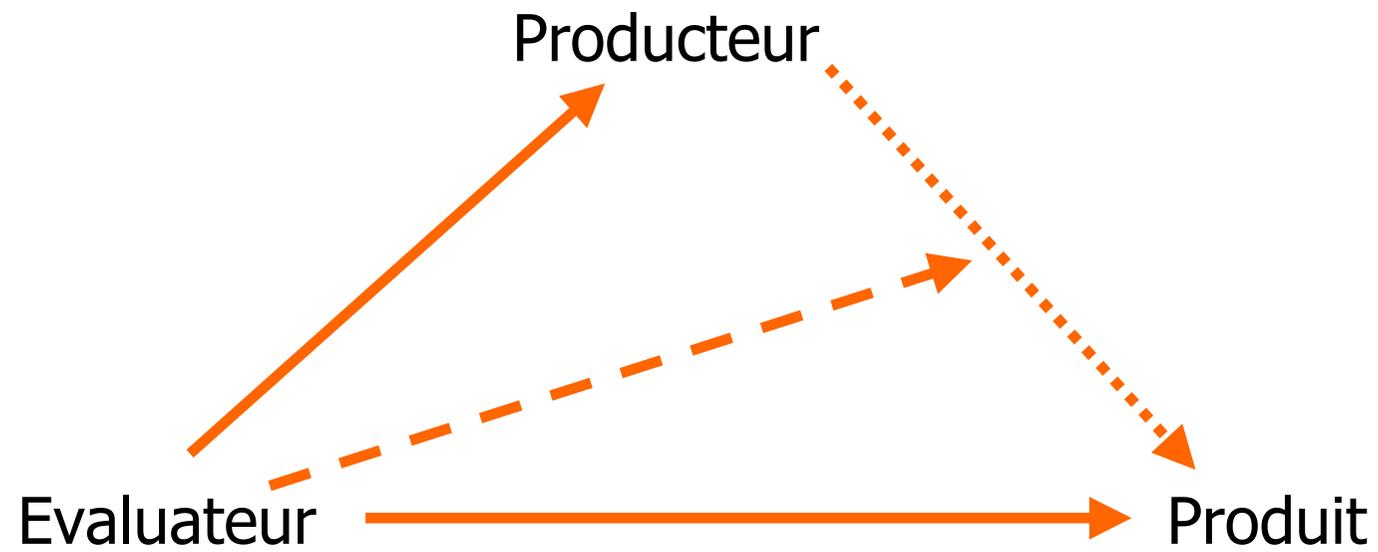
La situation évaluative professionnelle

Ce qui est dit de la rencontre au quotidien s'applique également dans le cas de la rencontre dans un contexte professionnel, d'où l'importance pour le professionnel de bien connaître ses propres cadres de références s'il veut commencer à maîtriser ce qui relève de la méta-communication dans le processus relationnel

Sur un plan technique, on peut dire qu'évaluer c'est formuler un jugement de valeur... Ce jugement peut porter sur trois entités : sur le produit (jugement technique), sur le producteur (jugement personologique), sur le rapport entre le producteur et sa production (jugement relatif).

Par exemple, on note une dictée (8 sur 20), on juge le producteur (« Peut mieux faire. ») ; on relative la note par rapport au niveau habituel (jugement positif si en moyenne l'élève a 6 sur 20 ; jugement négatif si en moyenne il a 12 sur 20).

Le triangle évaluatif



Evaluation pédagogique et docimologie

L'évaluation académique a été étudiée à partir des années 1920. On lui donne le nom savant de docimologie : « étude systématique des examens » (Piéron). L'approche peut être centrée sur les techniques d'évaluation (docimastique) ou sur les effets psychologiques de l'évaluation (doxologie).

Les constats de la docimologie

Depuis la première étude docimologique publiée en 1922 jusqu'à nos jours, les constats des analyses docimologiques vont toujours dans le même sens : l'évaluation académique n'est pas fiable.

On donnera deux exemples de ce manque de fiabilité parmi des dizaines possibles. On a donné 3 copies de mathématiques à corriger à 150 professeurs de mathématiques.

copie	note la plus basse	note la plus haute	note moyenne
A	0,5	11,5	5,7
B	11,5	20	16
C	3,5	11,5	8

Les mathématiques étant supposées une « science exacte », on devait s'attendre à ce qu'une même copie obtienne la même note quelque soit le correcteur. Il n'en est rien et on obtient une variation maximale de 11 points pour la première copie, de 8,5 points pour la seconde et de 8 points pour la troisième.

matière	nombre de correcteurs
mathématiques	13
physique	16
version latine	19
anglais	28
composition française	78
philosophie	127

On a pu calculer le nombre de correcteurs qu'il faudrait mobiliser pour la même copie afin d'approcher sa « valeur vraie » en faisant la moyenne des notes attribuées par les différents correcteurs. Le nombre de correcteurs augmente en fonction du caractère littéraire de la matière considérée pour atteindre 127 pour la philosophie. L'attribution de la note se rapproche alors franchement d'une loterie.

Quelques facteurs de variation

Les études docimologiques ont mis en évidence un certain nombre de critères de variations entre correcteurs, et aussi chez un même correcteur. Parmi ceux-ci, on peut citer des critères d'appréciation personnels relevant de la subjectivité. L'ordre des copies joue également et par un effet de contraste, une copie moyenne sera moins bien notée si elle suit une copie excellente et mieux évaluée si elle suit une copie médiocre. La tradition veut que selon les disciplines, les correcteurs utilisent des échelles de notation divergentes : plus serrées autour de la moyenne chez les littéraires et plus étendue chez les scientifiques.

Dans un autre d'idée, on montre régulièrement l'interférence produite par les informations préalables sur les évalués. On retrouve là le classique *effet Pygmalion*, ou réalisation automatique des attentes, qui fait que le correcteur va noter en partie en fonction de sa représentation positive ou négative de l'évalué. Jouent également un rôle dans ces processus de réduction de la dissonance cognitive liée aux stéréotypies le prestige du statut et du cursus scolaire suivi, le sexe et la catégorie socioprofessionnelle. Enfin, il faut dire deux mots sur le risque d'ethnisation qui entraîne le correcteur à juger la production non pas sur sa valeur intrinsèque mais plutôt selon la valeur qu'il attribue à l'appartenance ethno-culturelle du producteur.

Le fait qu'un système évaluatif aussi peu fiable ne soit jamais remis sérieusement en cause sur le fond interroge sur les fonctions de ce système.

Les fonctions de l'évaluation

Les partisans du « tout évaluatif » en souligne le caractère stimulant des dynamismes intellectuels et motivationnels, la réussite amplifiant le sentiment de compétence, l'échec prenant valeur de défi à relever. Mais on peut souligner en contrepartie le fait que toute situation évaluative constitue un un générateur de stress. L'excès d'évaluation conduit également au renforcement de l'individualisme et au bachotage dont la conséquence se manifeste en premier lieu en termes de déculturation, l'apprenant négligeant tout ce qui ne rentre pas dans le champ de l'évaluation.

La reproduction sociale

Sur un plan plus sociologique, Bourdieu et Passeron ont souligné que dans son ensemble le système scolaire fonctionnait comme un moyen de reproduction sociale, sur la base de l'évaluation d'un implicite jamais enseigné (les manières

d'être et de s'exprimer). Lurçat a pu pointer l'ambiguïté fondamentale d'un système compétitif à la fois juge et partie qui est amené à sélectionner sur ce qu'il doit transmettre. De toute évidence l'égalité des chances renvoie à l'égalité des malchances dont on rendra l'acteur responsable dans le cadre d'un renforcement de l'erreur fondamentale d'attribution (Ross).

Au niveau individuel, l'exaspération du système évaluatif peut conduire à des conséquences graves.

Les traumas évaluatifs

On peut lister différents types de traumas scolaires et/ou évaluatifs. Par exemple, la dévalorisation de la culture d'origine s'est manifesté par l'invention du bonnet d'âne pour ceux qui s'obstinaient à parler leur langue maternelle, patois ou créole.

Le jugement péjoratif de la personne, surtout lorsqu'il se répète, génère un processus d'auto-dévalorisation de soi aux manifestations plus ou moins dramatiques. La baisse de l'estime de soi peut initialiser une dépression, le manque de confiance en soi conduit à l'isolement social progressif, à la peur d'affronter des situations nouvelles, à l'évitement ou au refus des apprentissages.

On observe également des sorties des traumas évaluatifs par le biais de la construction d'une identité négative alimentée par toutes les formes de bénéfices secondaires.

Pourtant, une alternative existe : l'évaluation formative.

Vers l'évaluation formative

Alors que l'évaluation académique (ou sommative) fait l'impasse sur la « boîte noire », l'évaluation formative trouve son originalité dans la prise en compte du processus mental qui conduit le traitement de l'information, l'élaboration et la production de la réponse.

A partir de l'inférence de ce processus mental à l'aide de modèles théoriques appropriés aux objectifs pédagogiques visés, des stratégies de résolution peuvent être transmises à l'apprenant et/ou les objectifs peuvent être individualisés.

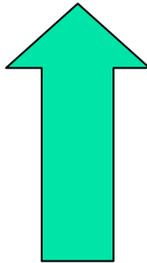
stimuli → réponse

boîte noire

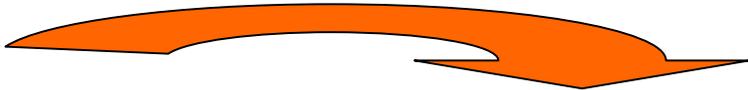
stimuli → processus mental → réponse



donner des stratégies...



inférence



définir des objectifs personnalisés...

La méthode psychométrique

Si l'évaluation formative marque un progrès qualitatif consistant, on peut également réduire les biais de mesure sur un plan quantitatif en adoptant le point de vue de la méthode psychométrique, méthode à la disposition des formateurs et des psychologues qui souhaitent s'y former.

Cette méthode met en œuvre différentes étapes : échantillonnage du contenu, soit de manière aléatoire, soit par sélection des contenus les plus représentatifs du domaine évalué ; standardisation détaillée de la procédure afin que l'évaluation puisse répondre au plus près du critère « toutes choses égales par ailleurs » ; échantillonnage de la population, par tirage au sort ou par quotas ; étalonnage. Cette dernière étape est l'opération qui fait passer de la notation du contenu au positionnement social de la performance en termes de classement.

Cette méthode offre les garanties de fiabilité métrologique propres à tout instrument de mesure : garantie de sensibilité soit le pouvoir de discriminer finement les performances ; garantie de fidélité en lien avec la stabilité des résultats en cas de mesure répétée ; garantie de validité de contenu -le test mesure effectivement ce qu'il est censé mesurer- et de validité critérielle -par rapport à des tests équivalents ; par rapport à des prédictions-.

Le choix d'un test passe par l'examen de ces différentes garanties. Malgré toutes ces précautions, il subsiste des biais qui affectent les résultats.

Quelques biais des tests

Parmi ces biais, on peut citer les biais d'items (items non discriminants par exemple), les biais différentiels liés au sexe (les femmes réussissent moins bien les épreuves spatiales...), à l'âge (les jeunes réussissent mieux les épreuves de raisonnement logique), liés à la catégorie socioprofessionnelle ou au niveau de qualification, liés au groupe ethnoculturel dans le cas de contenus spécifiques à une culture donnée.

On connaît également des biais liés à l'examineur, les examinateurs les plus expérimentés obtenant en moyenne des résultats meilleurs que les débutants, les premiers se montrant plus encourageants que les seconds.

Le biais de désirabilité sociale affecte essentiellement les questionnaires de personnalité, les répondants ayant tendance à choisir les réponses qui les présentent sous un meilleur jour plutôt que de choisir les réponses qui les décrivent le mieux.

Enfin, selon la nature du contexte, les résultats pourront fluctuer. Par exemple, dans une situation de recrutement, un stress positif va stimuler l'investissement dans l'épreuve et optimiser la performance, alors qu'un stress négatif va interférer avec le fonctionnement des capacités cognitives et affecter la performance.

L'évaluation dynamique permet de limiter ces biais.

L'évaluation dynamique

L'évaluation dynamique consiste à introduire une séquence d'apprentissage dans la procédure d'évaluation. Elle fonctionne un peu à la manière d'une microformation et elle permet d'observer la réactivité de la personne évaluée, qui tient ici le rôle d'apprenant, à cette mise en situation d'apprentissage. Sur un plan quantitatif, elle donne une mesure du potentiel d'apprentissage. L'aspect dynamique tient au fait que la mise en situation d'apprentissage suppose une modification des stratégies d'apprentissage au fur et à mesure du temps de la passation.

Elle se rapproche de l'évaluation formative par ses buts: mettre en évidence le processus cognitif ; réajuster les objectifs ; transmettre des stratégies. Mais elle s'en distingue par une double spécificité: elle s'inscrit dans le cadre de l'approche psychométrique ; elle place le répondant en situation d'apprentissage.

Cette approche a été préconisée dès les années trente du siècle passé.

Les précurseurs

Si on retient de Binet l'évaluation de l'intelligence à l'aide des tests, les psychologues ont curieusement fait l'impasse sur le postulat de l'éducabilité de l'intelligence énoncé par cet auteur. On peut se poser la question des déterminants qui ont conduit à cet « oubli » du facteur plasticité intellectuelle au profit des thèses fixistes privilégiant l'idéologie des dons.

Dans les années 1930, deux auteurs majeurs sont à prendre en considération. En U.R.S.S., Vygotski, s'inspirant du concept de médiation de l'outil proposé par Marx, a élaboré la notion de zone proximale de développement. Il part de l'exemple suivant. Si deux enfants obtiennent à un même test le même âge mental de 8 ans, on pourra conclure qu'ils sont identiques et préconiser le même traitement pédagogique. Par contre, imaginons qu'on prenne ce premier résultat comme un point de départ, qu'on apprenne à ces deux enfants à répondre

correctement aux questions du test et que suite à cet apprentissage, on donne à nouveau le test à passer. On observe alors que l'un des enfants atteint 9 ans d'âge mental et l'autre 12 ans. On ne peut plus dire qu'ils sont identiques : l'un a progressé d'un an et l'autre de quatre, cette différence concrétisant la zone proximale de développement. On ne peut alors préconiser le même traitement pédagogique pour ces deux enfants.

Dans la même période, Rey propose en 1934 un moyen de mesure de l'éducabilité, le *Test du labyrinthe manuel*. Il met en évidence des styles d'apprentissage (verbal, visuel, kinesthésique), des courbes de développement et aussi des profils pathologiques montrant que certains répondants ne parviennent pas à stabiliser leur apprentissage.

A partir des années 1950, les travaux de trois nouveaux auteurs vont émerger. Ombredane et ses collaborateurs, travaillant pour les compagnies forestières de l'ex-Congo belge, montre l'intérêt de l'évaluation dynamique dans le cadre d'une approche transculturelle : la simple réplique d'un test de matrices permet une élévation du niveau de performance, une meilleure discrimination des résultats et majore la prédiction de la réussite professionnelle (conduite des camions).

Hurtig va prolonger les travaux d'Ombredane dans le cadre de sa thèse centrée sur l'étude des marges d'éducabilité de l'intelligence. Ses travaux montrent que la dynamisation de la passation va réduire les écarts habituellement observés entre les niveaux des enfants en fonction de la C.S.P. de leurs parents. C'est en ce sens qu'il évoque un dispositif de la seconde chance.

Outre Atlantique Budoff va catégoriser les profils d'apprenants.

Les profils

Budoff distingue trois types de profils: les performants réussissent le test par eux-mêmes, sans avoir besoin d'aides extérieures ; les « gagnants » partent d'un niveau médiocre et, réactifs à la situation d'apprentissage, ils profitent un maximum des aides apportées ; enfin les « non-gagnants » restent insensibles aux aides apportées et ne parviennent pas à progresser.

Brown et Ferrara ont proposé deux dimensions additionnelles à ces profils : la vitesse (rapidité vs lenteur, hyperreflexivité) et l'empan de transfert (large vs étroit, limité au contexte) en fonction de la capacité du répondant à traiter des problèmes de plus en plus éloignés de la situation d'apprentissage de départ.

Budoff, par toute une série d'études mettant en relation potentiel d'éducabilité et d'autres variables ou indicateurs d'insertion, a pu établir un certain nombre de constats.

Les constats

Le constat fondamental rapporté par Budoff est la proximité des gagnants et des performants par rapport aux non-gagnants sur différents critères comme la rigidité cognitive, la gestion du stress et de la frustration, la valence de image de soi, l'anxiété, les réalisme des buts... Il faut prendre la mesure de l'importance des conséquences de ce constat fondamental : **l'interprétation d'un bas niveau obtenu à un test statique n'est pas synonyme d'incapacité. Il peut masquer un potentiel d'apprentissage élevé.**

On note également l'importance de la tenue dans le temps de l'apprentissage qui facilite la distinction entre des difficultés cognitives liées au handicap mental et celles liées aux difficultés sociales et/ou affectives.

La réduction des biais d'appartenance socio et/ou ethno culturelle rejoint la problématique des tests *culture fair* et on constate que les aides apportées génèrent un effet de reclassement et de rattrapage susceptibles de limiter, voire de combler les écarts liés à la C.S.P.

D'une façon générale, les différents auteurs notent une meilleure valeur prédictive des scores de potentiel d'apprentissage que des scores classiques par rapport à des critères de réussite scolaire, d'insertion sociale et professionnelle...

Concrètement, la dynamisation de la passation d'un test peut se mettre en œuvre de différentes façons.

Les paradigmes

La procédure test-entraînement-retest est appropriée pour évaluer l'effet de l'entraînement sur la performance initiale.

L'entraînement au cours du test, limité aux passations individuelles, permet de se centrer plus finement sur l'observation de la qualité et des modalités de l'intégration des aides apportées.

Le schéma apprentissage-test, le plus proche d'une microformation, est apte à mettre en évidence les stratégies d'apprentissage, de transfert et de contrôle.

Enfin, la séquence apprentissage-test-retest est nécessaire si on s'intéresse à la tenue de l'apprentissage dans le temps.

Mais dans tous les cas de figures, la passation se conclut par un entretien structuré de restitution afin de valider les différentes inférences du processus mental formalisées; il permet aussi d'apprécier le degré de généralisabilité des résultats enregistrés ainsi que la signification que le répondant donne à ces résultats.

Quelques outils

Le *Labyrinthe Manuel de Rey* est donc le premier test d'apprentissage opérationnel.

Feuerstein a repris différentes épreuves de Rey dans une batterie plus large, le *L.P.A.D. -learning potentiel assessment device-* destinée aux publics de préadolescents relevant du cadre de l'orthopédagogie.

Autre épreuve historique, le *Matrix Educabilité* d'Hurtig, applicable aux enfants d'âge scolaire (sous réserve d'un nouvel étalonnage).

Le *Test des carrés* de Mary & Mariel, mis au point pour les enfants d'école maternelle, dans le prolongement des travaux d'Hurtig.

Les Cubes de Kohs, depuis leur création en 1920, ont fait l'objet de nombreuses adaptations pour tous les publics, y compris en termes d'évaluation dynamique (par exemple procédure de Ionescu & col.)

Le *T.E.D.E.6¹ -Test d'Evaluation Dynamique de l'Educabilité-* de Pasquier a été réalisé à la demande d'entreprises industrielles dans le cadre de problématique de mobilité interne. Des étalonnages spécifiques ont été réalisés en réponse à la demande de Centres de Formation d'Apprentis.

Le *T.E.D.E.-Adapté²*, du même auteur, a été adapté du *T.E.D.E.6* pour un usage avec des publics spécifiques, identifiés comme fonctionnant à bas niveau (publics des établissements spécialisés, illettrés, traumatisés crâniens...).

Il reste à remettre en perspective ces différents éléments par rapport aux publics peu qualifiés.

¹ Disponible aux E.C.P.A.

² Disponible chez A & E.

Application aux publics peu qualifiés

En résumé, on peut souligner l'importance pour le professionnel de maîtriser au mieux les techniques d'évaluation qu'il met en œuvre, tant au niveau de ses stéréotypes plutôt négatifs qui accompagnent la représentation des publics peu qualifiés, de ses attentes implicites, de son comportement relationnel langagier et paralangagier que du point des qualités métrologiques des outils qu'il utilise.

Dans le champ pédagogique, l'évaluation formative, dans la mesure où elle se centre sur le processus d'élaboration de la réponse, offre le moyen de donner du sens à l'évaluation et de dépasser les jugements de valeur sommatifs plus ou moins englobants qui confondent production et producteur.

Le recours à une méthodologie psychométrique, y compris pour l'évaluation des savoirs dits de base, permet d'éviter les écueils les plus massifs du manque de fiabilité des évaluations académiques ordinaires. Toutefois, quelques biais subsistent qui interfèrent avec le résultat et l'évaluation dynamique permet de franchir une dernière étape pour optimiser les procédures d'évaluation.

Ce type d'évaluation permet de dépasser le simple constat d'un niveau difficilement exploitable sur un plan pédagogique et qui risque de renforcer les attentes ou d'alimenter les stéréotypies plus ou moins favorables envers les publics peu qualifiés.

L'évaluation dynamique limite les biais les plus courants, et plus spécifiquement, les biais liés au degré de familiarité avec la tâche, à l'appartenance socio-ethno-culturelle, au stress ou à l'émotivité.

Par rapport aux publics peu qualifiés, l'avantage essentiel que procure la dynamisation d'un test de niveau consiste à pouvoir différencier les potentiels d'apprentissage et on a vu que les potentiels élevés offraient les mêmes probabilités de réussite que les répondants d'emblée performants. En ce sens, on peut dire que les publics peu qualifiés constituent une mine de potentiels humains à détecter et à développer pour le bénéfice de tout un chacun et de la société en général.

Sur un plan qualitatif, la mise en situation d'apprentissage permet d'observer le mode de fonctionnement de l'apprenant dans les séquences d'apprentissage, de transfert et de contrôle ; cette mise en situation particulière stimule également l'image de soi en situation d'apprentissage, voie royale en direction de la problématique existentielle la plus large.

Enfin, sous l'angle des suites pédagogiques de l'évaluation, le recours à des procédures dynamiques facilite la préconisation fine d'objectifs de remédiation cognitive ou psycho-affective ainsi que la définition de cursus de formation mieux adapté au potentiel d'apprentissage et au mode de fonctionnement individuel.

En deux mots, on peut dire que l'évaluation dynamique optimise l'articulation des approches diagnostique et pronostique dans un contexte d'une meilleure fiabilité évaluative globale.