



D. Leclercq & M. Poumay (févr. 2005)

Taxonomie des Caractéristiques et Qualités de dispositifs d'évaluation pédagogique.

Université de Liège : LabSET

A. Caractéristiques des dispositifs

Dans les tableaux suivants,

-Les pôles de différentes caractéristiques ne sont pas automatiquement associés. Par exemple, contrairement à une représentation souvent véhiculée, le fait qu'une évaluation soit formative n'implique absolument pas l'absence de notation.

-Les deux pôles d'une même caractéristiques ne s'excluent pas non plus : dans un dispositif on peut trouver des évaluations formatives et d'autres sanctionnantes.

-Toutes les caractéristiques de l'évaluation ne sont pas toujours pertinentes pour tous les dispositifs.

B. Quelques définitions et remarques sur les caractéristiques :

(1) **Evaluation ponctuelle** : « Evaluation effectuée à un moment donné pour répondre à une question à propos d'un individu ou d'un programme (passage d'une classe, attribution d'une bourse,...). Ce type d'évaluation ne semble se justifier pleinement que dans les cas où une évaluation continue est impossible. » (G. de Landsheere, 1979, p.114)

(2) **Evaluation continue** : « Collecte systématique de scores ou d'appréciations...aboutissant à une note finale. Peuvent intervenir dans l'élaboration de cette note : les travaux faits en classe, les travaux à domicile, les résultats d'interrogations écrites, de tests, ainsi que le jugement subjectif de l'enseignant. L'évaluation continue est...un processus cumulatif, suivant le développement de l'élève et réfléchissant les changements qui interviennent dans ses réactions au cours. » (G. de Landsheere, 1979, p. 112).

(3) **L'objectivité** : elle n'existe pas ; on la définit comme « le consensus entre experts ». En évaluation, pour y parvenir, on s'efforce souvent d'obtenir des échantillons de comportements dont la mesure puisse être la même, quels que soient les juges, à condition qu'ils marquent leur accord sur les instruments et les critères (idéalement prédéfinis) et les appliquent. C'est le cas des QCM. Des échelles d'évaluation descriptives (c-à-d où les échelons ou niveaux sont décrits) sont d'autant plus objectives que la description de chaque échelon comporte moins d'ambiguïté (ex. un échelon « plus de deux fois par semaine » est plus objectif que « souvent »).

(4) **Observation participante** : « L'anthropologue Malinowski parlait d'observation participante qui « amène le chercheur à vivre la vie des groupes qu'il étudie, à partager le plus possible leurs activités pour mieux comprendre leur vision du monde » (G. de Landsheere, 1979,192).



C. Qualités : Un dispositif d'évaluation des performances cognitives a une...

...Validité Théorique.

- de **contenu** s'il assure une bonne couverture ou un échantillonnage représentatif de ce contenu. Par exemple, un test comportant des questions de tous les chapitres à étudier.
- de **construct** s'il est fondé sur un modèle crédible, scientifiquement fondé, des **Processus Mentaux**. Par exemple, si on pense, avec De Finetti (1965), que « *Only subjective probability can give an objective meaning to any system of questioning and scoring* », on recourt aux degrés de certitude et à la distribution spectrale des qualités des réponses. Autre exemple : un test comportant des questions des divers niveaux de la taxonomie des objectifs cognitifs de Bloom (1956).

...Validité Informative

s'il produit des données répondant à des questions pertinentes

- par étudiant . Par exemple, le degré d'aide demandé par le testé durant l'épreuve, des mesures distinctes de la cognition et de la métacognition, ou des divers niveaux de la taxonomie de Bloom.
- par matière. (Hunt, 1993 ; Baragibanie, 2003).

...Validité Conséquentielle...

...si il produit des effets sur le comportement des apprenants ou sur leurs processus mentaux ou les deux ou encore sur le dispositif de formation ou d'évaluation lui-même. Par exemple, un feedback et des dialogues métacognitif qui produisent des modifications de stratégies d'études. (RESSAC, 2003 ; Wislez, 2004).

...Validité Prédictive ou Concurrente

...si les mesures qu'il produit sont prédictives (ou corrélées avec) des mesures réputées valides. Par exemple, lors du testing MOHICAN en octobre 1999, les scores ayant une composante métacognitive se sont avérés plus prédictifs de la réussite de l'année que les scores simples (Leclercq, 2003). Autre exemple : Les QCM à SGI à une épreuve de mi-parcours se sont avérées plus prédictives que les QCM simples de la réussite d'étudiants en médecine (Gilles et al., 2001).

...Validité de Réplicabilité (Reliability) ou de Fidélité

...si les mesures produites sont stables dans le temps (test / retest) pour les caractéristiques n'évoluant pas ou ne sont pas affectées par des variations de conditions (formes parallèles, lieu, temps, etc.). Par exemple, l'augmentation des questions (formule de Spearman-Brown) augmente la fidélité des tests (dont le minimum attendu, en pédagogie est 0,8).

...Validité d'Acceptability – Practicability

...si elle est acceptée et jugée commode en durée, matériel requis (ordinateur ou seulement papier ?), concentration, précautions (antifraude par ex.), de quel lieu, à quels moments..

- par le professeur ou l'institution:
- par l'étudiant

...Validité Déontologique ou Ethique

...si il respecte des critères

- d'équité entre les personnes évaluées (Ex : objectivité des corrections par automatisation, ou par critères, ou anonymat lors de jugements pour éviter les « effets de correction » par juges, ou doubles corrections)
- de transparence dans les procédures (ex : annonce des critères) et de contrôlabilité - imputabilité (Ex : recalculer les scores pour vérification à la demande, droit de consulter sa copie a-posteriori)

bref : **TIC PRAD**



D. Taxonomie :

Dimensions	Pôles extrêmes	définition	Qualité
à référence	critérielle	juge la performance évaluée avec des critères (Ex : Réussir 60% des Q, citer 3 ex).	Validité écologique des critères
	normative	classe la performance évaluée (ou son score) par rapport aux autres performances (ou scores) évalué(e)s (Ex : « Vous êtes 1° ou 10° ou au-dessus de la moyenne, ou à -1,26 écart-type de la moyenne de votre cohorte »).	Validité théorique : Spécificité de la population de référence et des dates, nuance des repères, fiabilité des normes
à feedback	sommatif	donne à l'évalué un score sommant ses points à diverses questions ou divers critères. (Ex : 12/20 ou « TB »)..	Replicabilité (fidélité : Faible Erreur de Mesure)
	Détaillé ou par profil diagnostique	permet à l'évalué de distinguer l'origine de ses forces et de ses faiblesses, soit sur les contenus (ex. scores séparés pour chaque chapitre), soit sur les démarches (ex : oublié de simplifier la fraction) soit sur les processus mentaux (ex : scores séparés pour la connaissance, la compréhension, l'analyse, le jugement métacognitif, etc.).	Précision, crédibilité (preuves, raisonnements fondateurs), acceptabilité (utilisabilité) pour l'enseignant, pour l'apprenant
à visée	sanctionnante	débouche sur une sanction soit sélective (ex : l'étudiant est admis) soit certificative (ex. l'étudiant obtient ou non son diplôme).	Replicabilité (Fiabilité ou fidélité) de coupure ou de classement).
	formative	a pour but de nourrir le processus d'apprentissage de l'étudiant.	Validité théorique : lien entre information et performance. Predictivité : Efficacité (Pré / Post) : Ampleur de l'Effet, sur le produit ou sur le processus (d'auto-évaluation).
à centration sur	les processus d'apprentissage	mesure et juge des données relatives au déroulement de l'apprentissage de l'étudiant, à son parcours de formation (ex : apprentissage « par cœur », « en profondeur », « massé vs distribué », « seul vs en groupe », « avec support verbal vs iconique », « type d'aide demandée », etc.).	Validité Informative : Précision (diagnosticité)
	les produits d'apprentissage	mesure et juge les résultats d'apprentissage en termes de productions (ex : exactitude et certitude de réponses à des QCM, qualités d'une performance complexe comme la rédaction d'un rapport).	Validité théorique des critères (notamment sensibilité des données, définition de la « connaissance ou de la compétence).
	Les objets de l'apprentissage	Fournit pour chaque question (ou matière) les statistiques de choix, de pouvoir discriminatif (Etudiants forts / faibles), les certitudes moyennes (une erreur avec certitude élevée indiquant les « misconceptions »	Validité théorique (séparation des composantes d'une question).
à dimension(s)	unique (ou unidimensionnelle)	prend en compte UNE SEULE dimension du travail, le plus souvent le nombre de bonnes réponses, parfois une seule autre qualité (ex. originalité de la réponse, délai de réponse, logique de raisonnement,...) de ce travail.	Validité théorique et Reproductibilité par Cohérence interne (unidimensionnalité).



	plurielle (ou pluri-dimensionnelle)	prend en compte PLUSIEURS dimensions du travail de l'évalué.(Ex : l'exactitude de ses réponses, leur originalité, leur rapidité ou délai de réponse, le réalisme de l'autoestimation d'exactitude, la logique de raisonnement, la justification, etc.)	Validité théorique : Pertinence du poids relatif de chaque composante.
à cible	individuelle	évalue UN apprenant sur ses connaissances, ses compétences, ses conduites, etc.	Validité théorique ; ici « modèle de l'apprenant sous-jascent » : « <i>person free tests</i> » ? Justice distributive.
	de groupe	porte sur les performances d'un GROUPE d'évalués, qui seront solidaires de certaines composantes de la note attribuée au groupe.	Validité déontologique ou éthique : équité de l'attribution (la répartition) des notes.
à destination	privée	garde certaines données destinées au seul évalué confidentielles, pour lui seul ou ses parents s'il est mineur.	Validité déontologique ou éthique : garantie de la confidentialité.
	publique	rend accessible au public certaines données relatives à la performance évaluée d'une personne. (Ex : Les résultats sont affichés aux valves).	Validité déontologique ou éthique : justification de la publicité.
à opérateur(s)	allo (allo-opératrice)	recourt à autrui (le professeur, un autre expert, un tuteur, un pair, etc.) pour mesurer et/ou juger et/ou fournir le feedback.	Validité déontologique ou éthique : légitimité de l'expert. Par ex les pairs ne peuvent sanctionner leurs condisciples.
	auto (auto-évaluation)	ne dépend pour la mesure et/ou le jugement et/ou le feedback ni du professeur, ni d'un autre expert (tuteur), ni de ses pairs mais de l'évalué lui-même.	Validité théorique (pertinence de ce choix) et d' acceptabilité (faisabilité, commodité).
	Composite (allo et auto)	recourt à l'allo-évaluateur (Ex : pour le choix des questions et des critères de correction) et à l'auto-évaluateur (Ex : pour du moment du test et l'application des critères de mesure, etc.).	Validité théorique : pertinence de la répartition des tâches.
à périodicité	Ponctuelle (1)	n'est pas combinée avec d'autres, ne tient pas compte d'autres.	Validité théorique (choix du moment, préparation, annonce, etc.)
	Continue (2)	est combinée avec d'autres pour fournir une appréciation d'ensemble tenant compte de la succession des évaluations (par ex. des progrès).	Validité théorique (justification de la décomposition)
	Répétée	est une évaluation continue où l'épreuve par laquelle on mesure les progrès de l'évalué est toujours la même.	Validité théorique, de contenu (Ex : dans les Progress Tests de médecine à Maastricht, les rubriques) ou de construct (Ex : idem : les cas + VF.
à durée de vie	définitive	fournit des notes non modifiables.	Validité théorique (des critères), de reproductibilité .
	améliorable	permet à l'étudiant de proposer plusieurs versions de son travail ou de son test, versions qui seront prises en compte dans la note finale.	Validité conséquentielle : les premières notes doivent compter, sinon le premier travail est bâclé.
à source	objective	est menée selon des procédures faisant l'accord des experts (3) et aboutissant donc à des notes identiques quels que soient les experts.	Validité théorique (des critères), de reproductibilité .
	subjective	est donnée par des juges et ne repose pas sur un consensus d'experts. (Ex : évaluation de travaux complexes, de rapports, de portfolios, etc.)	Validité théorique (des critères), de reproductibilité .



à procédure	standardisée	met, partiellement ou totalement les évalués dans les mêmes conditions (mêmes jour, lieu, questions, surveillance, critères, procédures, etc.).	Acceptabilité, déontologique ou éthique par équité (ex : les étrangers : avec un dictionnaire traductif ?)
	adaptative	ne standardise pas toutes ses composantes, mais les adapte à l'évalué, soit par l'évalué lui-même (ex. il choisit ses questions) soit par le testeur (ex. examen oral), voire même du système (ex. certains tests informatisés).	Acceptabilité, déontologique ou éthique (questions plus faciles que d'autres)
	Composite (standardisée / adaptative)	Est standardisée par certains aspects (règles, délais) et adaptative par d'autres (choix des questions, ordre de réponse)	Acceptabilité, déontologique ou éthique (égalité des chances)
à implication	interne	ou participante (4) est menée par des évaluateurs qui ont été impliqués dans la formation de l'évalué.	Validité théorique (la compétence et rien que la compétence) et de contenu (toute la compétence) et déontologique ou Ethique (juge et partie ; éviter les effets de halo, d'inertie)
	externe	est menée par des évaluateurs qui n'ont pas été impliqués dans la formation des évalués.	Validité déontologique ou éthique (équation personnelle des correcteurs ; préparation au type d'examen)
à contrat	imposé	« vient d'en haut » sans place pour une discussion entre l'évaluateur et l'évalué.	Acceptabilité (légitimité, justice distributive)
	négocié	laisse une place à la discussion entre l'évaluateur et l'évalué sur certains points (ex. le thème du projet personnel), en imposant d'autres.	Acceptabilité (légitimité, justice distributive)
à repères/ échelle	mobiles	change ses critères avec la cohorte (avec le niveau) des évalués. Il s'agit du système le plus fréquent. (Ex : les questions de 4e année sont plus compliquées que celles de 3e année).	Validité théorique (Ajustement au niveau des apprenants) et déontologique ou éthique (préparation des apprenants)
	fixes	présente des questions ou tâches identiques (ou équivalentes) indépendamment du niveau (d'études) de la cohorte. Seule change la proportion d'items que l'étudiant y est supposé réussir selon son niveau d'études (ex. 10% en 1ere année, 20% en seconde, 60% en 4e, etc).	Validité conséquentielle (minimise le bachottage) et théorique de contenu : cela nécessite un grand nombre de questions (ex : Les Progress tests de la faculté de médecine de Maastricht) .
À visibilité des critères	annoncés	précise dès le départ tous les critères d'évaluation de la performance et éventuellement leur pondération.	Validité conséquentielle : préciser les pondérations encourage les productions « stratégiques », soignant les seuls critères « lourds ».
	cachés	ne précise pas à l'étudiant à l'avance les critères d'évaluation de sa performance.	Acceptabilité, déontologique ou éthique (trop de place au hasard ?).
à contexte	écologique ¹	est proche de la future situation professionnelle des participants, en contenus ou en conditions de passation.	Acceptabilité (faisabilité pour l'évaluateur)
	scolaire	met en œuvre des conditions de passation et des objectifs éloignés d'un contexte professionnel réel.	Validité théorique (lien entre le contexte scolaire et le réel (ex : les « leçons publiques »).

¹ Cette expression est due à Egon Brunswick (1943).



E. Références

- Brunswick, E. (1943) Organismic achievement and Environment Probability, *Psychological Review*, 50, 255-272.
- Leclercq, D. (1999), *Edumétrie et Docimologie*, Service de Technologie de l'Éducation, Université de Liège.
- Poumay, M. (2003), *Évaluation formative et EAD - Définition, justifications théoriques et illustrations*, document de coformation interne, LabSET, Université de Liège.
- de Landsheere, G. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : PUF.

Annexes



Représentation synoptique des facettes et dimensions de dispositifs d'évaluation

T= Toutes M= Manque de données ABS = Absurdité, ne s'applique pas

	Facettes	Dimensions	Justif	Dimensions	Justif	C
Pourquoi ? (fonctions)	Référence	Critérielle		Normative		
	Résultat	Sommatif		Diagnostique		
	Visée	Sanctionnante		Formative		
Quoi ? (objets)	Centrage sur	Processus		Produits		
	Dimension	Uni-dimens		Pluri-dimens.		
Qui ? (sujets)	Cible	Individuelle		de groupe		
	Destinataires	Privée		Publique		
	Opérateurs	Allo-opératrice		Auto-opératrice		
Comment ? (méthodes, conditions)	Périodicité	Ponctuelle		Continue / Répét		
	Révisabil. des notes	Définitive		Améliorable		
	Source	Objective		Subjective		
	Procédure	Standardisée		Adaptative		
	Implication	Interne		Externe		
	Contrat	Imposé		Négocié		
	Repères/ échelle	Mobiles		Fixes		
	Critères	Annoncés		Cachés		
Validités	Contexte	écologique		scolaire		
	Théorique »					
	Informative					
	Conséquentielle					
	Prédictive					
	Replicabilité					
	Acceptable					
Déontologique						



Représentation synoptique des facettes et dimensions du dispositif d'évaluation

T= Toutes (les deux) M= Manque de données ou « Cela dépend » ABS = Absurdité ou Ne s'applique pas

17 Q :	Dimensions	Facettes	Dimensions	C
100 80 60 40 20 0	Critérielle	Référence	Normative	0 20 40 60 80 100
100 80 60 40 20 0	Sommatif	Résultat	Diagnostique	0 20 40 60 80 100
100 80 60 40 20 0	Sanctionnante	Visée	Formative	0 20 40 60 80 100
100 80 60 40 20 0	Processus	Centrage sur	Produits	0 20 40 60 80 100
100 80 60 40 20 0	Uni-dimens.	Dimension	Pluri-dimens.	0 20 40 60 80 100
100 80 60 40 20 0	Individuelle	Cible	de groupe	0 20 40 60 80 100
100 80 60 40 20 0	Privée	Destinataires	Publique	0 20 40 60 80 100
100 80 60 40 20 0	Allo-opératrice	Opérateurs	Auto-opératr.	0 20 40 60 80 100
100 80 60 40 20 0	Ponctuelle	Périodicité	Continue Répét	0 20 40 60 80 100
100 80 60 40 20 0	Définitive	Révisabil. des notes	Améliorable	0 20 40 60 80 100
100 80 60 40 20 0	Objective	Source	Subjective	0 20 40 60 80 100
100 80 60 40 20 0	Standardisée	Procédure	Adaptative	0 20 40 60 80 100
100 80 60 40 20 0	Interne	Implication	Externe	0 20 40 60 80 100
100 80 60 40 20 0	Imposé	Contrat	Négocié	0 20 40 60 80 100
100 80 60 40 20 0	Mobiles	Repères/ échelle	Fixes	0 20 40 60 80 100
100 80 60 40 20 0	Annoncés	Visibilité critères	Cachés	0 20 40 60 80 100
100 80 60 40 20 0	Ecologique	Contexte	Scolaire	0 20 40 60 80 100