

Février 2011

Choisir ses stratégies d'évaluation

Amaury Daele – Denis Berthiaume

L'objectif de ce document est de présenter différents aspects à considérer lorsque l'on conçoit un dispositif d'évaluation des apprentissages des étudiant-e-s à l'université. Plusieurs questions sont abordées ici :

- Quels sont les apprentissages des étudiant-e-s à évaluer ?
- Comment obtenir une « trace » ou un « témoignage » de leurs apprentissages ? Quelles stratégies d'évaluation peuvent être mises en place pour y parvenir (comme un examen oral, un questionnaire, la réalisation d'un travail écrit, etc.) ?
- Comment interpréter ces traces des apprentissages des étudiant-e-s et leur accorder une note ?

Derrière ces questions apparaissent en filigrane des éléments importants liés à l'évaluation auxquels les enseignant-e-s universitaires sont souvent confronté-e-s : la subjectivité de leur évaluation, la validité et la fiabilité des stratégies d'évaluation choisies, ou encore les outils à mettre en œuvre pour mener ce processus à bien.

Objectifs de l'évaluation des apprentissages

Le processus d'évaluation des apprentissages a pour but d'obtenir de la part des étudiant-e-s une trace ou un témoignage qu'ils/elles ont réalisé les apprentissages ou développé les compétences visés par le cours (ou séminaire, TP, etc.). Dans ce cadre, on distingue généralement deux types d'évaluation :

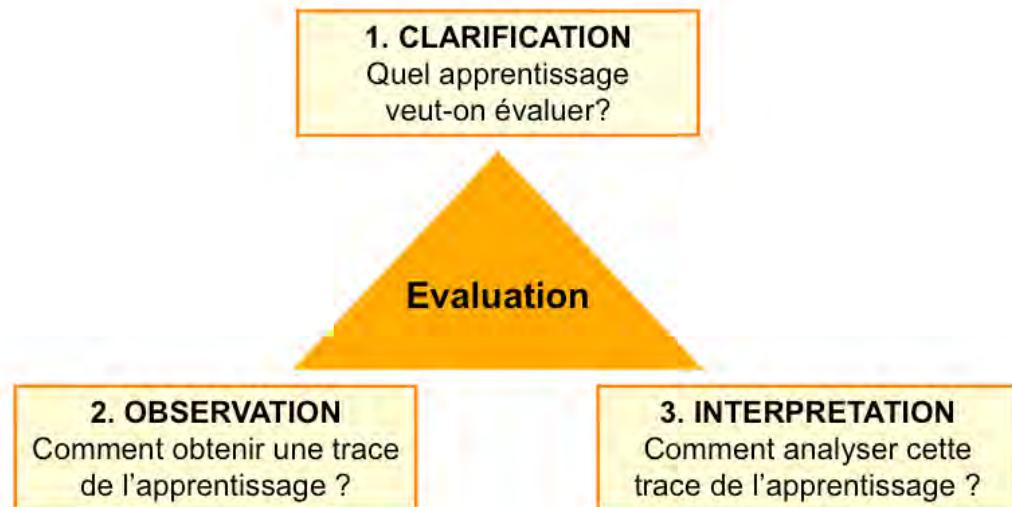
- L'*évaluation sommative* (ou certificative) qui intervient le plus souvent à la fin d'un apprentissage et qui vise à une prise de décision quant à la réussite d'un enseignement par l'étudiant-e ou pour l'attribution d'un diplôme. Il s'agit de faire la « somme » des apprentissages de l'étudiant-e et cette évaluation est le plus souvent accompagnée d'une note chiffrée ou alphanumérique.
- L'*évaluation formative* intervient en cours d'apprentissage et vise à donner des indications à l'étudiant-e à propos du chemin parcouru et du chemin qu'il lui reste à parcourir pour atteindre les apprentissages visés. Ceci se fait le plus souvent sous la forme d'un feed-back oral ou écrit, par exemple suite à un examen blanc ou à la réalisation d'une première version d'un travail écrit. Généralement, cette évaluation n'est pas accompagnée d'une note ou alors compte pour un faible pourcentage dans la note finale.





Les trois phases du processus d'évaluation

Le processus d'évaluation pourrait être présenté en trois phases successives évoquées dans le schéma ci-dessous.



Adapté de Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Chaque phase correspond à une question centrale à se poser :

1. Phase de *clarification* des apprentissages ou des compétences à évaluer ;
2. Phase d'*observation*, ou d'obtention d'une trace d'apprentissage de la part des étudiant-e-s en recourant à une stratégie d'évaluation (par ex. : examen oral, exercices écrits, rédaction d'un texte, réalisation d'un projet, etc.) ;
3. Phase d'*interprétation* de la performance des étudiant-e-s pour déterminer si celle-ci correspond aux apprentissages attendus et pour y affecter une note.

Ces trois phases pourraient être apparentées à un processus de recherche qualitative au cours duquel des questions de recherche sont posées (phase de clarification), des instruments de collecte de données sont mis en œuvre (phase d'observation) puis des méthodes d'analyse qualitative sont utilisées pour répondre aux questions de recherche (phase d'interprétation). Dans les pages qui suivent, nous précisons en quoi chacune des trois phases concourt à la cohérence du processus d'évaluation des apprentissages.





Phase 1 : Clarification

Dans la phase de clarification, il s'agit de définir quels apprentissages devront être évalués. Pour ce faire, un outil utile est le tableau ci-dessous qui reprend différents types d'objectifs pédagogiques selon le domaine de compétence (cognitif, affectif ou psychomoteur) et le niveau de compétence visé (en surface, intermédiaire ou en profondeur). Dans ce tableau, chaque type d'objectif est décrit avec quelques exemples de verbes qui permettent de se faire une représentation du type d'action (intellectuelle ou psychomotrice) qu'il peut être demandé aux étudiant-e-s pour vérifier leur maîtrise des objectifs¹.

Apprentissage	Cognitif	Affectif	Psychomoteur
En surface	Rétention : identifier, définir, reconnaître, lister, nommer, etc.	Réception : reconnaître, développer une conscience de, etc.	Perception : distinguer, reconnaître, observer, lister, etc.
Intermédiaire	Compréhension : décrire, résumer, appliquer, utiliser, comparer, expliquer, etc.	Valorisation : développer une attitude, participer, faire des choix, etc.	Reproduction : utiliser un matériel, reproduire, appliquer, réaliser une performance, etc.
En profondeur	Réflexion : interpréter, évaluer, justifier, analyser, critiquer, etc.	Adoption : appliquer des principes, prendre des responsabilités, mettre en pratique, etc.	Perfectionnement : développer son savoir-faire, inventer, adapter, etc.

Par exemple, lorsque l'évaluation vise à vérifier si les étudiant-e-s maîtrisent des objectifs cognitifs en surface, les questions qui leurs seront posées vont chercher à ce qu'ils/elles identifient ou définissent des concepts, listent des éléments d'un domaine, reconnaissent des indices ou des concepts, etc. Si l'évaluation vise des objectifs affectifs intermédiaires, les étudiant-e-s seront amenés à opérer ou valoriser des choix personnels, à développer certaines attitudes, etc. Lorsque le but est de vérifier des objectifs psychomoteurs de haut niveau, il s'agira d'observer les étudiant-e-s développer un savoir-faire personnel, leur créativité, etc.

Cette phase permet d'assurer la validité de l'évaluation en focalisant l'attention sur ce qu'il faut exactement évaluer en tenant compte éventuellement du poids relatif de chaque objectif du cours. La clarification des objectifs d'apprentissage au début du processus d'évaluation permet ensuite de choisir plus aisément une stratégie adéquate pour obtenir de la part des étudiant-e-s une trace de leur apprentissage.

¹ Pour aller plus loin dans la conception des objectifs d'apprentissage, on peut se référer au document « L'identification et la rédaction des objectifs pédagogiques » (Amaury Daele et Denis Berthiaume) : <https://enseigner.unil.ch/ressources/document/lidentification-et-la-redaction-des-objectifs-pedagogiques/>.





Phase 2 : Observation

Dans cette seconde phase, il s'agit de choisir et de mettre en œuvre une stratégie adéquate d'évaluation. Plusieurs questions peuvent ici se poser :

- Le moment de l'évaluation : celle-ci peut en effet se dérouler *avant* un enseignement (pour tester des prérequis par exemple), *pendant* l'enseignement (pour évaluer la compréhension pendant une séance de cours par exemple) ou *après* un enseignement (pour une évaluation sommative).
- L'objet de l'évaluation : l'attention de l'évaluation peut porter sur un produit fini (la qualité d'un texte par exemple) ou sur un processus de pensée (la qualité de l'argumentation d'une opinion par exemple). Elle peut porter aussi sur une performance ponctuelle (une présentation orale par exemple) ou sur les progrès réalisés sur une certaine période de temps.
- Le choix de la stratégie d'évaluation : certaines stratégies requièrent peu d'interprétation de la part de l'enseignant-e pour corriger les réponses fournies par les étudiant-e-s. C'est le cas notamment des QCM, des questions vrai/faux, des questions d'appariement (*matching*) ou des questions appelant des réponses courtes comme un mot ou une phrase. Avec ces stratégies, la correction est réputée plus « *objective* » car les réponses fournies par les étudiant-e-s sont moins sujettes à interprétation. A l'opposé, certaines stratégies requièrent davantage de travail d'interprétation des réponses des étudiant-e-s comme les observations avec ou sans échelle², les questions qui demandent des réponses écrites développées, les productions écrites longues ou les exposés oraux. Dans ces cas, la correction est souvent plus « *subjective* », mais nous verrons plus loin comment il est possible de limiter cette subjectivité.

DOMAINE	NIVEAU	STRATEGIES D'EVALUATION
Cognitif	Rétention	QCM, Vrai/Faux, Matching, Réponses brèves
	Compréhension	Réponses à développement, Travail long, Exposé oral
	Réflexion	Travail long, Exposé oral
Affectif	Réception	QCM, Réponses brèves, Observation (sans échelle)
	Valorisation	Travail long, Exposé oral, Observation (avec échelle)
	Adoption	Travail long, Exposé oral, Observation (avec échelle)
Psychomoteur	Perception	Observation (sans échelle)
	Reproduction	Observation (avec échelle)
	Perfectionnement	Observation (avec échelle), Travail long, Exposé oral

Pour choisir une stratégie d'évaluation adéquate par rapport aux domaines et niveaux d'apprentissage visés, on peut se référer au tableau ci-contre qui propose différentes stratégies d'évaluation.

Nous présentons en outre brièvement dans le tableau qui suit huit stratégies différentes en mentionnant leurs avantages, leurs inconvénients possibles et

les éléments particuliers à retenir si on veut les appliquer avec ses étudiant-e-s.

² L'observation, par exemple de l'expression orale ou d'une performance psychomotrice, se fait généralement avec une check-list de comportements à observer chez la personne qu'on évalue. « *Avec échelle* » signifie que pour chaque élément de la check-list, on note en plus si le geste est « *très bien* », « *bien* » ou « *insuffisamment* » accompli.





Stratégies	Avantages	Inconvénients	A retenir
Questionnaire à choix multiple	<ul style="list-style-type: none"> Rapide à corriger Interprétation des réponses généralement peu subjective Permet de diagnostiquer les difficultés ou erreurs d'apprentissage 	<ul style="list-style-type: none"> Construction d'un QCM valide et fiable assez complexe (travail de spécialiste) Construction d'un QCM prend du temps (validité des énoncés et des réponses) 	<ul style="list-style-type: none"> Ne permet d'évaluer généralement que les apprentissages de surface Evalue principalement les apprentissages du domaine cognitif et quelques fois ceux du domaine affectif
Vrai/Faux	<ul style="list-style-type: none"> Rapide à corriger Interprétation des réponses généralement peu subjective 	<ul style="list-style-type: none"> Possibilité pour les étudiant-e-s de deviner ou déduire la bonne réponse Difficile de rédiger un énoncé qui soit absolument et toujours vrai 	<ul style="list-style-type: none"> Ne permet d'évaluer que les apprentissages de surface Evalue principalement les apprentissages du domaine cognitif et quelques fois ceux du domaine affectif
Matching (association/appariement)	<ul style="list-style-type: none"> Rapide à corriger et assez rapide à construire Interprétation des réponses généralement peu subjective 	<ul style="list-style-type: none"> Possibilité pour les étudiant-e-s de deviner ou déduire la bonne réponse Evalue parfois la capacité à rechercher et à éliminer les mauvaises réponses 	<ul style="list-style-type: none"> Ne permet d'évaluer que les apprentissages de surface Evalue principalement les apprentissages du domaine cognitif
Réponses à compléter (texte à trous)	<ul style="list-style-type: none"> Assez rapide à construire et à corriger Interprétation des réponses généralement peu subjective (mais davantage que les évaluations à réponse choisie) 	<ul style="list-style-type: none"> Peut encourager les étudiant-e-s à apprendre le contenu du cours par cœur Énoncé peut être difficile à comprendre si le « trou » est au début de l'énoncé 	<ul style="list-style-type: none"> Ne permet d'évaluer que les apprentissages de surface Evalue principalement les apprentissages du domaine cognitif
Réponses brèves	<ul style="list-style-type: none"> Assez rapide à construire Interprétation des réponses généralement peu subjective (mais davantage que les évaluations à réponse choisie) 	<ul style="list-style-type: none"> Doit chercher les éléments de réponse dans la construction par l'étudiant-e Essentiel de préparer une grille de correction (par souci de fiabilité) 	<ul style="list-style-type: none"> Permet d'évaluer les apprentissages de surface et parfois intermédiaire Evalue principalement les apprentissages du domaine cognitif mais aussi affectif
Réponses à développement	<ul style="list-style-type: none"> Rapide à construire Interprétation subjective des réponses Peut cibler un contenu assez vaste 	<ul style="list-style-type: none"> Essentiel d'identifier des critères d'évaluation (par souci de fiabilité) Obstacle potentiel provenant de l'expression écrite de l'étudiant-e 	<ul style="list-style-type: none"> Permet surtout d'évaluer les apprentissages intermédiaires Permet d'évaluer les apprentissages des domaines cognitif et affectif
Travail long	<ul style="list-style-type: none"> Rapide à construire Interprétation subjective des réponses Laisse davantage d'autonomie aux étudiant-e-s (selon la situation) 	<ul style="list-style-type: none"> Essentiel d'identifier des critères d'évaluation (par souci de fiabilité) Obstacle potentiel provenant de l'expression écrite de l'étudiant-e 	<ul style="list-style-type: none"> Permet d'évaluer les apprentissages intermédiaires ou en profondeur Evalue les apprentissages des domaines cognitif et affectif (et parfois psychomoteur)
Exposé oral	<ul style="list-style-type: none"> Rapide à construire Interprétation subjective des réponses Laisse davantage d'autonomie aux étudiant-e-s (selon la situation) 	<ul style="list-style-type: none"> Essentiel d'identifier des critères d'évaluation (par souci de fiabilité) Obstacle potentiel provenant de l'expression orale de l'étudiant-e 	<ul style="list-style-type: none"> Permet d'évaluer les apprentissages intermédiaires ou en profondeur Evalue les apprentissages des domaines cognitif et affectif (et parfois psychomoteur)





Phase 3 : Interprétation

Dans la phase d'interprétation, il s'agit d'analyser la production des étudiant-e-s et d'attribuer une valeur aux traces d'apprentissage qu'ils/elles ont fournies. Plusieurs éléments sont à déterminer ici :

- La référence : il s'agit de déterminer par rapport à quoi les apprentissages des étudiant-e-s vont être évalués. On peut d'une part se référer à la *moyenne générale du groupe* d'étudiant-e-s, déterminée dans une courbe normale. Il s'agit donc de comparer les résultats des étudiant-e-s entre eux. Cette courbe peut constituer la référence dans le cas d'un examen d'entrée par exemple ou si le nombre d'étudiant-e-s pouvant être admis-e-s dans l'année suivante doit être limité. Dans ces cas, ne sont admis que les 20, 30 ou 40 premier-ère-s d'une volée. On peut d'autre part se référer à un *critère absolu de réussite* où les étudiant-e-s doivent faire la preuve de la maîtrise d'une compétence spécifique ou minimale pour réussir l'évaluation ou être admis-e-s dans l'année suivante.
- La personne qui évalue : cette personne n'est pas toujours l'enseignant-e qui a dispensé le cours. Cela peut être un jury externe (pour les mémoires ou les thèses par exemple), l'étudiant-e lui/elle-même (auto-évaluation formative par exemple) ou les autres étudiant-e-s (évaluation par les pairs).

La subjectivité de l'évaluation

La procédure en trois phases décrite ci-dessus vise en fait non pas à supprimer toute subjectivité dans l'évaluation mais plutôt à tendre vers une objectivité « procédurale », c'est-à-dire visant à atténuer les effets de la subjectivité en standardisant la façon dont l'évaluation est appliquée. Comme dans toute recherche, une attention particulière est portée à la validité et à la fiabilité de la méthode : évalue-t-on bien ce que l'on prétend évaluer et est-on sûr-e que l'on évalue tou-te-s les étudiant-e-s de la même façon et selon les mêmes critères ? Deux outils spécifiques permettent de rencontrer ces exigences et de rendre le processus d'interprétation moins subjectif, c'est-à-dire dépendant de circonstances externes à l'évaluation elle-même.

1. Les critères d'évaluation

Le premier de ces outils consiste en l'établissement de critères d'évaluation, c'est-à-dire d'indices observables de la qualité des réponses ou des prestations des étudiant-e-s lors d'une évaluation. Ces critères, directement liés aux objectifs pédagogiques, permettent de clarifier ce qui est évalué en fournissant une description détaillée de ce qui doit être acquis, appris ou maîtrisé par les étudiant-e-s. Quelques exemples : « les explications fournies par l'étudiant-e sont pertinentes et concises », « l'étudiant-e utilise les concepts vus au cours en leur donnant une définition exacte », « l'étudiant-e collabore avec





ses pairs de façon efficace », « l'étudiant-e formule son opinion de façon concise et ses arguments sont pertinents et justifiés », etc. Les critères permettent en définitive, lorsqu'on examine la prestation d'un-e étudiant-e, de focaliser son attention sur ce qu'il y a lieu d'évaluer en faisant abstraction le plus possible de ce qui constitue seulement le contexte de l'évaluation (la propreté d'une copie par exemple ou la façon dont l'étudiant-e est habillé-e...).

2. Les niveaux (ou indicateurs) de performance

Le second outil évoqué consiste en l'identification de niveaux de performances qui permettent de clarifier les différents niveaux de compétence des étudiant-e-s et donc de faire la distinction entre les étudiant-e-s qui ont atteint les apprentissages attendus de façon « excellente », « suffisante », « insuffisante », etc. On distingue en général deux ou trois niveaux de compétence et ceux-ci peuvent être quantitativement ou qualitativement différents. Quelques exemples :

- Deux niveaux. « Succès » : l'étudiant-e rencontre toutes les exigences requises ; « Echec » : l'étudiant-e ne rencontre pas toutes les exigences requises.
- Trois niveaux qualitatifs. « Excellent » : dépasse les exigences minimales ; « Acceptable » : rencontre les exigences minimales ; « Inacceptable » : ne rencontre pas les exigences minimales.
- Trois niveaux quantitatifs. « Maîtrise complète » : peut poser plus de 25 gestes correctement ; « Maîtrise partielle » : peut poser entre 10 et 25 gestes correctement ; « Maîtrise insuffisante » : peut poser moins de 10 gestes correctement.





3. Les grilles d'évaluation critériées

Si l'on combine ces deux outils dans un tableau, on obtient ce qu'on appelle une *grille d'évaluation critériée* (ou *rubrique*) dont voici un exemple :

Niveaux Critères	Excellent	Acceptable	Inacceptable
Professionalisme Se soucier ou s'assurer de la qualité des soins offerts	Va au-delà des attentes minimales en s'assurant, de façon constante, du bien-être du/de la patient-e	Rencontre les attentes minimales en s'assurant, la majeure partie du temps, du bien-être du/de la patient-e	Ne rencontre pas les attentes minimales en ne s'assurant pas, la majeure partie du temps, du bien-être du/de la patient-e
Travail en équipe Collaborer avec des collègues dans le but d'accroître l'efficacité du groupe	Dépasse les attentes minimales en recherchant les occasions de collaboration	Rencontre les attentes minimales en collaborant lorsque nécessaire	Ne rencontre pas les attentes minimales en évitant les situations de collaboration

Pour chaque critère de la colonne de gauche correspondent trois niveaux de performance différents. A l'aide de cette grille, il s'agit alors de déterminer pour chaque critère d'évaluation à quel niveau de performance correspond la copie ou la prestation d'un-e étudiant-e. Une telle grille constitue un outil des plus utiles lorsque de nombreuses copies sont à corriger par exemple ou lorsque plusieurs personnes sont amenées à évaluer ensemble les prestations des étudiant-e-s. Cela permet de systématiser la correction et de se mettre d'accord à plusieurs sur la façon dont les étudiant-e-s seront évalué-e-s.

Quelques éléments à retenir

- L'évaluation des apprentissages est souvent assez complexe et demande beaucoup d'attention lors des trois phases du processus pour identifier les objectifs à évaluer, choisir une stratégie d'évaluation et interpréter les résultats pour attribuer une note aux étudiant-e-s.
- L'évaluation est un processus qui prend toujours beaucoup de temps, soit pour concevoir l'épreuve que l'on va soumettre aux étudiant-e-s, soit pour corriger les copies, soit pour donner du feed-back aux étudiant-e-s à propos de leurs performances, etc.
- La subjectivité souvent décriée de l'évaluation peut être atténuée grâce au respect d'une certaine procédure : identification d'objectifs, choix adéquat d'une stratégie, élaboration de critères et de niveaux de performance et conception d'une grille d'évaluation. Sans prétendre à l'objectivité totale, on se prévaut ainsi de certains biais liés à l'évaluation qualitative de performances parfois très complexes.





Références

Les livres listés ci-dessous sont disponibles aux Centre de Soutien à l'Enseignement :

- Bloxham, S., & Boyd, P. (2007). *Developing effective assessment in higher education: A practical guide*. Maidenhead, UK: Open University Press.
- Brown, S., & Glasner, A. (1999). *Assessment matters in higher education: Choosing and using diverse approaches*. SRHE/Open University Press.
- Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. London: Kogan Page.
- Cranton, P. (1989). *Planning instruction for adult learners*. Chapitre 6 : Evaluating Learning (pp. 136-179). Toronto : Wall & Emerson, Inc.
- Fenwick, T., & Parsons, J. (2000). *The art of evaluation: A handbook for educators and trainers*. Toronto, ON: Thompson Educational Publishing, Inc.
- Morissette, D. (1996). *Évaluation sommative*. Montréal, QC: Éditions du Renouveau Pédagogique.
- Prégent, R., Bernard, H., & Kozanitis, A. (2009). *Enseigner à l'université dans une approche-programme*. Chapitre 5 : Evaluer des compétences en situation authentique (pp. 137-189). Montréal : Presses Internationales Polytechnique.
- Scallon, G. (2000). *L'évaluation formative*. Montréal, QC: Éditions du Renouveau Pédagogique.
- Stevens, D. D., & Levi, A. J. (2005). *Introduction to rubrics*. Sterling, Virginia: Stylus.
- Svinicki, M. & McKeachie, W. J. (2011). *McKeachie's Teaching Tips. Strategies, Research, and Theory for College and University Teachers*. Chapitre 7 : Assessing, Testing, and Evaluating : Grading is Not the Most Important Function (pp. 72-82). Belmont, CA : Wadsworth, International Edition.
- Tardif, J. (2006). *L'évaluation des compétences: Documenter le parcours de développement*. Montréal: Chenelière Education.

Quelques sites généraux sur l'évaluation à l'université :

- Dossier « Assessment » sur le site de la London Metropolitan University : <https://www.londonmet.ac.uk/about/resources/assessment-compendium/>
- Dossier « Stratégie d'évaluation » sur le site Enseigner de L'Université Laval : <https://www.enseigner.ulaval.ca/pedagogie/strategie-devaluation>
- « Authentic Assessment Toolbox » par J. Mueller : <https://jonfmueller.com/toolbox/>

