

Comment réduire la subjectivité de l'évaluation ?

À l'université comme aux autres niveaux d'enseignement, l'évaluation fait partie intégrante des tâches quotidiennes des enseignants. De nombreuses épreuves jalonnent l'année académique et surtout la clôturent : rapports de laboratoire, travaux, examens oraux ou écrits... Deux qualités essentielles sont attendues de toute évaluation : la validité (l'épreuve permet-elle de mesurer l'atteinte de tous les objectifs d'enseignement et ne mesure-t-elle que cela ?) et la fidélité (l'épreuve permet-elle une mesure fiable et stable, quelle que soit la personne qui l'organise et quel que soit le moment et le contexte de l'évaluation ?). Cette seconde qualité est particulièrement cruciale lorsque diverses personnes sont impliquées dans une même évaluation. Souhaitant améliorer la fidélité des évaluations des rapports de laboratoire, le département de chimie a mené, en collaboration avec le Service de Pédagogie Universitaire, un travail de réflexion sur les procédures mises en œuvre et sur les moyens d'en accroître la qualité. Ce 78^e numéro de RESEAU propose un compte rendu de ce travail réalisé avec l'ensemble des assistants concernés par l'évaluation des rapports de travaux pratiques de chimie. Il a été rédigé en collaboration avec Isabelle Ravet et Diane Baillieul, membres de la Cellule didactique du département de chimie et à ce titre responsables de l'organisation des travaux pratiques de chimie, en première année des programmes des facultés des Sciences et de Médecine.

Mais qu'est-ce qu'évaluer ?

Si l'on se réfère à la définition classique de Stufflebeam (Madaus, Scriven & Stufflebeam, 1986), évaluer consiste à mesurer puis à apprécier, à l'aide de critères, l'atteinte des objectifs d'enseignement, en trois étapes. La première étape réside dans le recueil systématique, valide et fidèle d'informations appropriées aux objectifs d'enseignement. C'est la phase d'observation et de recueil de données. La deuxième étape consiste à interpréter les informations recueillies à l'aide de critères. C'est la phase d'analyse et d'interprétation. Celle-ci débouche sur la troisième étape, à savoir l'établissement de conclusions ou la prise de décisions. C'est la phase du jugement à proprement parler.

Lorsqu'un enseignant évalue les acquis de ses étudiants, il procède d'abord à un recueil standardisé d'informations. Par exemple, il soumet, à une même date et dans les mêmes conditions (temps de réponse standardisé, questions identiques ou équivalentes...) tous les étudiants à un même examen. Mais les informations ainsi recueillies (les réponses des étudiants) ne parlent pas d'elles-mêmes. Lors de la correction, les productions des étudiants sont analysées selon un certain nombre de critères congruents par rapport aux objectifs. Même analysées de manière rigoureuse au regard de critères, les productions des étudiants n'aboutissent pas non plus mécaniquement à un jugement, une décision ou une note finale. Il faut encore, dans un troisième temps, que l'enseignant construise un arbre de décisions, explicite ou implicite, lui permettant d'établir ses conclusions.

Évaluer n'équivaut donc pas à enregistrer de manière externe et objective des acquis, comme le thermomètre indique la température, sans l'intervention de l'homme. Au contraire, évaluer des acquis revient à construire un point de vue sur des performances censées représenter des acquis. La subjectivité intervient donc aux trois étapes de l'évaluation décrites ci-dessus ; elle est même inhérente au processus d'évaluation. C'est ce qui explique que la fidélité de l'évaluation n'est pas toujours assurée : quelle que soit la discipline, l'évaluation n'aboutit pas toujours à des résultats similaires si elle est reproduite à un autre moment ou si elle est réalisée par une autre personne.

Éradiquer toute subjectivité dans l'évaluation des acquis n'est ni possible ni souhaitable. La subjectivité n'est pas à éradiquer, mais à encadrer : il s'agit de développer des procédures garantissant qu'elle se déploie dans des proportions acceptables. Les manières de contenir la subjectivité sont multiples et diverses. La plus fondamentale réside sans doute dans l'explicitation collective des procédures, modalités et critères d'évaluation. C'est l'option qui a été prise dans l'expérience rapportée dans ce numéro.

Comment améliorer la fidélité ?

Des notes différentes pour un même rapport

Dans le courant de l'année académique 2009-2010, Isabelle Ravet et Diane Baillieul ont été interpellées par plusieurs étudiants à propos de la note reçue pour leur rapport de laboratoire. Ils contestaient leur note, surtout en comparaison de celle obtenue par un autre étudiant. Autrement dit, à plusieurs reprises, deux correcteurs semblaient avoir octroyé des points différents pour une réponse similaire, aux yeux des étudiants du moins. Comme expliqué ci-dessus, ce constat classique relève d'un manque de fidélité de l'évaluation. Ce manque apparaît *a fortiori* ▲

lorsque diverses personnes évaluent un même travail, ce qui est le cas des rapports de laboratoire puisque 15 assistants sont impliqués dans leur correction.

L'objectif d'améliorer la fidélité de l'évaluation des rapports de laboratoire en vue d'augmenter l'équité entre les étudiants s'est concrétisé dans un travail réalisé en plusieurs temps.

Un travail collectif de réduction de la subjectivité

Une rencontre entre un membre du Service de Pédagogie Universitaire et les membres de la Cellule didactique en chimie a permis de cerner précisément la demande, son contexte et les ressources disponibles notamment en ce qui concerne le temps à consacrer au projet.

Ensuite, trois réunions rassemblant toutes les personnes impliquées dans la correction des rapports de laboratoires ont été mises sur pied pour construire ensemble un outil d'évaluation en lien avec les objectifs d'apprentissage, pour définir des critères observables de maîtrise de ces objectifs et pour les pondérer.

Réunion 1 : explicitation des objectifs d'enseignement

Trois rapports de laboratoire réels, mais anonymisés ont été distribués avec l'invitation à la première rencontre. Il a été demandé aux assistants de les corriger comme ils le font habituellement en cours d'année.

Critères observables : description qualitative

Présentation du rapport

- L'écriture est lisible et sans rature
- Une structure est présente (les titres sont notés et mis en évidence : couleur, soulignement, majuscules)
- Le rapport est propre (pas de tâches, non chiffonné...)

Chiffres significatifs des mesures et des résultats

- Les mesures sont fournies avec un nombre de chiffres après la virgule significativement correct lié à l'appareil de mesure
- Les résultats sont fournis avec un nombre de chiffres après la virgule significativement correct lié aux calculs effectués et aux mesures réalisées

Unités des mesures et des résultats

- La valeur des mesures et des résultats est assortie d'une unité correcte

Qualité des mesures effectuées

- Les mesures sont effectuées soigneusement et lues correctement (pas d'erreurs de mesure et de lecture inappropriées)

Équations des réactions

- Les équations des réactions sont inscrites correctement (réactifs, produits, coefficients stœchiométriques, flèche de réaction)

Calculs effectués

- Les démarches pour trouver les résultats à partir des mesures sont effectuées pas à pas pour mettre en évidence le raisonnement

Compréhension

- La réponse aux questions à propos d'étapes de manipulation

Figure 1 : Description des critères observables

Lors de cette première séance réunissant toute l'équipe, les objectifs du projet ont été explicités et un premier travail sur les représentations à propos de l'évaluation a été réalisé, en faisant associer par chacun un terme au processus de correction des rapports de laboratoire.

Ensuite les points attribués par chacun aux trois rapports ont été comparés. Les notes attribuées variaient considérablement entre

les différents correcteurs pour un même rapport (écarts compris entre 0 et 3,5 ou 0 et 5,5 points selon le rapport sur un total de 20 points). Pour un des rapports, cet écart faisait osciller la décision entre une réussite au travail et un échec, selon le correcteur. La nécessité de s'atteler à améliorer la fidélité de l'évaluation des rapports était ainsi démontrée.

Le reste de la réunion a été consacré à la recherche d'un consensus sur les objectifs d'enseignement pour l'ensemble des séances de laboratoire. De manière individuelle, chaque assistant a d'abord inscrit sur des fiches ses objectifs en matière d'acquis des étudiants (par exemple : être critique par rapport aux résultats). Ensuite, les objectifs ont été partagés et ont fait l'objet de regroupement et de classement. Enfin, les objectifs d'une même catégorie ont été reformulés de manière à tenir compte des points de vue de chacun.

Au terme de la séance, une évaluation a été effectuée. Elle a révélé qu'une majorité d'assistants étaient intéressés par le travail (9 participants), qu'ils attendaient de voir à quoi il pouvait aboutir (8), mais la moitié était optimiste (6) et certains étaient même ravis et heureux (3) de pouvoir échanger à propos d'une

tâche importante de leur métier, alors que d'autres se sentaient « effrayés » par le résultat de l'évaluation des trois rapports (2).

Dans le but de préparer la deuxième réunion et d'accélérer quelque peu le processus, une description qualitative de critères observables dans un rapport de laboratoire a été rédigée par le Service de Pédagogie Universitaire à partir des grilles de correction des deux premiers laboratoires, fournies par la Cellule didactique (figure 1).

Réunion 2 : vers des critères observables communs

Lors de cette réunion, un premier temps individuel a été consacré à la lecture de la grille de critères (figure 1). Dans un deuxième temps, les participants ont comparé, en sous-groupes, cette grille de critères aux objectifs énoncés lors de la première réunion et regroupés par thème. Chaque sous-groupe a produit des transparents dont le contenu consistait à corriger et adapter la grille de critères en se référant aux objectifs identifiés. Une technique d'animation (le carrousel romain) a créé une dynamique de travail entre les groupes. Cette technique consiste à travailler par groupe de 4 (un rapporteur, deux secrétaires et un gestionnaire du temps) pendant 30 minutes. Au terme des 30 minutes, le rapporteur et un des deux secrétaires de chaque groupe partent dans un autre groupe.

UN COUP D'ŒIL DANS LE RÉTROVISEUR...

La validité et plus prosaïquement le sérieux des examens universitaires ont constitué un sujet fréquent de railleries, comme dans le passage suivant tiré des Mémoires de ma vie de Charles Perrault, publié en 1759.

« Un valet qui vint nous parler à la fenêtre, ayant su ce que nous souhaitions [une licence] nous demanda si notre argent était prêt. Sur quoi, ayant répondu que nous l'avions sur nous, il nous fit entrer et alla réveiller les docteurs, au nombre de trois, qui vinrent nous interroger avec leur bonnet de nuit sous leur bonnet carré. (...) Un de nous, à qui l'on fit une question dont il ne me souvient pas répondit hardiment: *Matrimonium est legitima maris et foemince conjunctio, individuam vitae consuetudinem continens, et dit sur ce sujet une infinité de belles choses qu'il avait apprises par cœur. On lui fit ensuite une autre question sur laquelle il ne répondit rien qui vaille. Les deux autres furent ensuite interrogés et ne firent pas beaucoup mieux que le premier. Cependant, ces trois docteurs nous dirent qu'il y avait plus de deux ans qu'ils n'en avaient interrogés de si habiles et qui en sussent autant que nous. Je crois que le son de notre argent, que l'on comptait derrière nous pendant que l'on nous interrogeait, servit de quelque chose à leur faire trouver nos réponses meilleures qu'elles n'étaient. »*

Thélot Cl. (2001), *Les écrivains français racontent l'école*. Paris : Delagrave, p. 302.

Ils y présentent le résultat de la réflexion de leur groupe (ici le document adapté) et les deux autres membres du groupe ont pour mission de poser des questions pour améliorer le document et de s'inspirer des « bonnes idées » pour améliorer leur propre document. Après deux tours de carrousel, les groupes initiaux sont à nouveau formés. Chaque groupe exploite alors les idées glanées auprès des autres afin de finaliser leur document.

En fin de séance, les groupes ont présenté le fruit de leur travail et une courte évaluation a été réalisée. Parmi les aspects appréciés, les participants ont relevé surtout le fait de pouvoir échanger sur des questions pédagogiques importantes. Au contraire, le manque de temps, le temps nécessaire pour réaliser ce travail ainsi que les différences significatives de vues au sein de l'équipe ont été les éléments principalement pointés comme aspects dérangeants.

Réunion 3 : pondération des critères et échelle de notation

Les participants ont été invités à répartir 100 points entre les différents critères observables repris sur de grands panneaux, à l'aide de

post-it, selon l'importance et le poids qu'ils souhaitaient attribuer à chacun de ces critères. Ensuite, une mise en commun a été effectuée en trio et un consensus sur la pondération a été recherché. Enfin, une mise en commun de la pondération en grand groupe a été réalisée au tableau et un échange a été mené dans le but d'aboutir à un accord sur la pondération.

En grand groupe, un exemple d'échelle de notation a été présenté. Ensuite des duos d'assistants ont été invités à élaborer une pareille échelle pour deux critères.

Pour clôturer la séance, une mise en forme de la grille de notation par le Service de Pédagogie Universitaire, sur la base des propositions de chaque duo d'assistants, a été proposée et une dernière évaluation a été organisée.

Enfin, une dernière entrevue entre le Service de Pédagogie Universitaire et la Cellule didactique a été planifiée pour peaufiner la grille de notation.

Afin de valider cette grille et de boucler la boucle, les assistants ont à nouveau corrigé les trois rapports anonymisés en exploitant

UNE PUBLICATION RÉCENTE...

PLATTEAUX H. & SCHAUB M. (2012), « Innover dans l'évaluation des apprentissages : pourquoi et comment ? », *Revue internationale de pédagogie de l'enseignement supérieur* [En ligne], 27-2.

Ce numéro thématique de la revue de l'AIPU vise à donner une vision de quelques-unes des innovations qui sont mises en œuvre aujourd'hui en matière d'évaluation des apprentissages. Les analyses et les réflexions faites par les auteurs des articles dépassent le cadre de l'expérience pratique et de l'étude de cas et invitent le lecteur à une réflexion plus générale sur les principes mêmes qui fondent l'évaluation des apprentissages et sur les outils au service de l'évaluation comme, par exemple, la grille d'évaluation à base de critères.

Les différentes contributions font réapparaître une tension structurante forte au sein de l'enseignement supérieur en matière d'évaluation des apprentissages, le rapport entre objectivité et subjectivité. La lecture de ce numéro constitue dès lors un excellent prolongement de ce RESEAU.

<http://ripes.revues.org/495>

la nouvelle grille de notation. Les membres de la Cellule didactique ont ensuite recensé les points attribués pour ces rapports par les assistants et le résultat a été communiqué à toute l'équipe.

Les écarts de notes entre les différents correcteurs ont été fortement réduits (écarts compris entre 0 et 2 au maximum) et en tout cas, les correcteurs étaient unanimes sur le verdict de réussite ou d'échec pour chaque rapport.

Lors de cette correction des trois rapports « témoins », les écarts subsistant entre les notes étaient liés à des interprétations différentes de la grille par les différents correcteurs. La grille avait été conçue pour être applicable à tous les rapports de l'année, mais il est apparu nécessaire de préciser comment lire la grille pour chaque manipulation. Pour chaque TP, Isabelle Ravet et Diane Baillieux ont clairement identifié, en fonction du contenu attendu du rapport, la portée de chaque critère en précisant pour chacun d'eux l'échelle quantitative adaptée.

Grille de notation des rapports des laboratoires

Un extrait de la grille finale de notation des rapports des laboratoires est présenté dans la figure 2. La grille complète est disponible sur simple demande. Dans cette grille, ▲

Grille de correction des rapports de laboratoire					
Critères observables et mesurables Total du rapport sur 20 points (jusqu'à -8 points de pénalité possible)	Echelle quantitative				
	+4 points	+3 points	+2 points	+1 point	0 point
I. Présenter un rapport correct au niveau du fond +13/20 points I.1. Présenter la valeur des mesures et des résultats avec des chiffres significativement corrects +4/20 points • Écrire la valeur des mesures avec un nombre de chiffres après la virgule significativement correct en fonction de la précision de(s) l'appareil(s) de mesure utilisé(s) • Écrire les résultats avec un nombre de chiffres après la virgule significativement correct lié aux calculs d'erreurs effectués et à la précision des mesures réalisées • Arrondir correctement (0-4) → on arrondit à l'unité inférieure ; (5-9) → on arrondit à l'unité supérieure • Garder tous les chiffres en mémoire dans la calculatrice lors des calculs	Toutes les mesures et tous les résultats sont fournis avec des chiffres significatifs corrects	1 ou 2 mesures ou résultats sont fournis avec un chiffre significatif incorrect	Le ¼ ou plus du ¼ des mesures et des résultats sont fournis avec un chiffre significatif incorrect	Le 1/3 ou plus du 1/3 des mesures et des résultats sont fournis avec un chiffre significatif incorrect	Le ½ ou plus de ½ des mesures et des résultats sont fournis avec un chiffre significatif incorrect
I.2. Présenter les mesures et les résultats avec des unités correctes +3/20 points • La valeur des mesures et des résultats est assortie d'une unité correcte	Toutes les mesures et tous les résultats sont fournis avec des unités correctes	1 mesure ou résultat est fourni avec une unité incorrecte ou sans unité	2 mesures ou résultats sont fournis avec une unité incorrecte ou sans unité	Le ¼ ou plus du ¼ des mesures et des résultats sont fournis avec une unité incorrecte ou sans unité	
I.3. Présenter les calculs et les graphiques correctement +6/20 points • Présenter les démarches pour trouver les résultats pas à pas à partir des mesures pour mettre en évidence le raisonnement <ul style="list-style-type: none"> Utiliser les « formules » adéquates Tenir compte des rapports stœchiométriques Utiliser les valeurs exactes données au laboratoire (inscrites sur le tableau du laboratoire) Effectuer correctement les calculs <ul style="list-style-type: none"> Utiliser correctement la calculatrice Estimer la validité de l'ordre de grandeur obtenu • Présenter les graphiques en respectant rigoureusement les règles (voir syllabus) <ul style="list-style-type: none"> Rédiger un titre Représenter correctement les deux axes (traits forts, flèches, grandeurs mesurées, unités) Sélectionner l'échelle appropriée et l'indiquer sur les axes Tracer correctement les courbes 	Les démarches pour trouver les résultats à partir des mesures sont effectuées pas à pas, les calculs sont effectués correctement et les graphiques sont présentés en respectant les règles	Les ¼ des démarches pour trouver les résultats à partir des mesures sont effectuées pas à pas, les calculs sont effectués correctement et ¼ des graphiques sont présentés en respectant les règles	La ½ des démarches pour trouver les résultats à partir des mesures sont effectuées pas à pas, les calculs sont effectués correctement et la ½ des graphiques sont présentés en respectant les règles	Moins de ¼ des démarches pour trouver les résultats à partir des mesures sont effectuées pas à pas, moins de ¼ des calculs sont effectués correctement et les graphiques ne sont pas présentés en respectant les règles	

Figure 2 : Extrait de la grille de notation finale

NOM :	TITRE DU LABORATOIRE	DATE :
PRENOM :		NOM DE L'ASSISTANT :
SECTION :		
N° DU MODULE :		TOTAL :

Forme (-2 points de pénalité)	Fond (+20 points et -4 points de pénalité)																		Compo- rtement (-2 points de pénalité)									
	Chiffres significatifs			Unités			Calculs et graphiques						Equations			Comprendre Données ad hoc				Comprendre Réponses aux questions								
0	-	-	0	1	2	3	4	0	1	2	3	0	1	2	3	4	0	1	2	3	4	5	0	1	2	0	1	2

Figure 3 : Feuille de rapport

la première colonne comprend les critères observables et les points attribués à chaque critère. Les colonnes suivantes sont dédiées à l'échelle correspondant aux différents critères. Cette grille se trouve dans le syllabus de TP des étudiants qui peuvent donc prendre connaissance des critères selon lesquels leurs rapports de laboratoires seront évalués.

Dans un souci de transparence pour l'étudiant et de facilité pour l'assistant, cette grille est reprise sur la feuille de rapport prédéfinie que l'étudiant doit remettre en fin de séance de laboratoire (cf. figure 3).

Conclusion

Le travail réalisé pas à pas avec l'équipe des assistants de chimie constitue une procédure possible pour concevoir une épreuve d'évaluation qui tende vers davantage d'objectivité dans l'appréciation du travail des étudiants. Cette procédure consiste à :

- partir de la liste des objectifs d'apprentissage qui font l'objet de l'évaluation;
- rédigier des critères observables pour chaque objectif;
- pondérer chacun des critères en fonction de leur importance dans l'apprentissage;
- concevoir une échelle quantitative pour chacun des critères afin de pouvoir les mesurer;

- intégrer tous les éléments (critères observables, pondération, échelle quantitative) dans une grille commune et opérationnelle.

“ La subjectivité n'est pas à éradiquer, mais à encadrer

Selon les participants, l'intérêt de cette expérience, via le produit obtenu (la grille de notation), est d'avoir permis d'augmenter la fidélité de la correction des rapports de laboratoire. De plus, la

démarche suivie par l'équipe et la succession des étapes franchies pour construire la grille de correction pourraient aisément se transférer à d'autres situations d'évaluation. Le Service de Pédagogie Universitaire reste d'ailleurs à la disposition de toute équipe pédagogique qui souhaiterait réaliser pareille démarche d'évaluation réalisée sur la base de critères explicites, au bénéfice d'une plus grande fidélité dans l'appréciation des acquis des étudiants.

Références

- De Landsheere, G. (1976) *Évaluation continue et examens. Précis de docimologie*. Paris : Fernand Nathan.
- Madaus, G.F., Scriven, M. & Stufflebeam, D. L. (1986). *Evaluation models. Viewpoints on educational and human services evaluation*. Hingham : Kluwer Academic Publishers.
- Prégent, R., Bernard, H. & Kozanitis, A. (2009). *Enseigner à l'université dans une approche programme*. Montréal : Presses Internationales Polytechniques.

UN SITE À CONSULTER...

Géré par un ancien assistant (Amaury Daele) du Service de Pédagogie Universitaire installé désormais à l'Université de Lausanne, ce site est dédié aux stratégies d'évaluation dans l'enseignement supérieur. Il comprend notamment un document de synthèse qui, sans prétendre à l'exhaustivité, propose plusieurs questions et conseils judicieux à garder en tête au moment de procéder à l'évaluation des apprentissages des étudiants. Par ailleurs, l'auteur a rassemblé plusieurs liens utiles à propos des pratiques d'évaluation dans l'enseignement supérieur.

<http://pedagogieuniversitaire.wordpress.com/2011/02/21/choisir-ses-strategies-devaluation/>

Romainville, M. (2002). *L'évaluation des acquis des étudiants dans l'enseignement universitaire*. Paris : Rapport établi à la demande du Haut Conseil de l'évaluation de l'école. (disponible sur le site : <http://cisad.adc.education.fr/hcee/publications-2003.html>)

Scallon, G. (1988) *L'évaluation formative des apprentissages. Tome 2: L'instrumentation*. Sainte-Foy : Les Presses de l'Université Laval.

Scallon, G. (2000) *L'évaluation formative*. Québec : Éditions du Renouveau pédagogique.

Scallon, G. (2004) *L'évaluation des apprentissages dans une approche par compétences*. Québec : Éditions du Renouveau pédagogique.

Tardif, J. (2006). *L'évaluation des compétences: Documenter le parcours de développement*. Montréal : Chenelière Éducation.



SERVICE DE PÉDAGOGIE UNIVERSITAIRE

14, Place Saint-Aubain - 5000 Namur

Responsable : Marc Romainville