

De nombreux enseignants, confrontés à des questions multiples et complexes, demandent aujourd'hui des textes pour penser leurs pratiques d'évaluation sommative. Cette chronique poursuit ce but. Quatre questions récurrentes issues du terrain seront traitées tour à tour dans les prochains numéros. Chacune d'entre elles donnera à voir la collaboration entre un formateur et des enseignants travaillant au développement de leurs pratiques évaluatives sommatives, à l'aide d'outils d'analyse. Une synthèse clôturera la réflexion. Les approches pour penser cette problématique étant variées, ces chroniques viseront moins à dire comment évaluer qu'à donner quelques pistes pour réfléchir à ses pratiques, dans une perspective critique et compréhensive.

Première question vive de la pratique

Raphaël Pasquini, UER EN, HEP Vaud – raphael.pasquini@hepl.ch

Les résultats de mes évaluations ne correspondent souvent pas à ce qu'a produit l'élève: qu'est-ce qu'une évaluation sommative «juste»?

L'histoire commence dans une formation continue dispensée à un groupe d'enseignant-e-s de 3-4P. Deux

d'entre elles souhaitent soumettre à l'analyse du groupe une évaluation en arts visuels en 4e. Le duo a travaillé le dessin d'animaux selon un modèle. Le jour de l'évaluation, les élèves ont dû dessiner un hibou. Ils ont reçu le guide de travail et la feuille de critères suivants:

DESSIN DU HIBOU : GUIDE DE TRAVAIL

1° Je dessine la tête sur 1/3 et le corps sur les 2/3 restants de la feuille.	<table border="1"> <tr> <td>Tête</td> <td></td> </tr> <tr> <td>Corps</td> <td></td> </tr> </table>	Tête		Corps		
Tête						
Corps						
2° Je dessine 3 ronds les uns dans les autres pour chaque oeil.						
3° Je dessine 1 triangle pour chaque oreille et 1 autre pour le bec.						
4° Je dessine 1 ovale pour le ventre et 3 petits ovales pour chaque patte.						
5° Dans le ventre, je trace entre 5 et 8 lignes de pré-écriture différentes.	<pre> e e e e e e e e e e x x x x x x x x x x </pre>					
6° Je passe sur tous mes traits avec un stylo						
<p style="text-align: center;">Attention !</p> <p style="text-align: center;">Lorsque tu peindras en suivant la consigne n°7, sois attentif à :</p> <ul style="list-style-type: none"> *ne pas laisser de blanc *ne pas dépasser *ne pas faire de taches 						
7° Je peins chaque plage d'une seule couleur. Attention !! Deux plages de même couleur ne se touchent pas.						

Date : _____ Prénom : _____

Evaluation significative d'art visuel

DESSIN DU HIBOU

Critères d'évaluation :	OUI	NON
La tête occupe 1/3 de la feuille.		
Le corps occupe 2/3 de la feuille.		
Chaque oeil est composé de 3 ronds (les uns dans les autres).		
Chaque oreille est dessinée avec un triangle.		
Le bec est dessiné avec un triangle.		
Le ventre est dessiné avec un ovale.		
Chaque patte est composée de 3 petits ovales.		
Le ventre est décoré avec différentes lignes de pré-écriture. (5 à 8 lignes)		
Passer sur tous les traits de construction avec un feutre noir permanent de manière continue.		
Peindre 1 couleur par plage.		
2 plages de la même couleur ne se touchent pas.		
Peindre sans laisser de blanc.		
Peindre sans dépasser.		
Le dessin est soigné, il n'a pas de taches.		

Nombre de critères réussis :

NA	PA	A	AA	LA
0-6	7-8	9-10	11-12	13-14

Les deux enseignantes sont très satisfaites de leurs documents. Elles ont passé un temps considérable à les élaborer. De plus, les résultats sont très bons, les dessins fidèles aux modèles, ce qui les conforte dans leurs choix. Cette évaluation sera soumise à l'analyse du groupe la séance suivante, car elle pourrait inspirer d'autres collègues. De retour chez moi, je ne résiste pas: je me mets au travail, et en référence aux documents, je dessine «mon» hibou, en cherchant à être fidèle aux critères.

Voici ce que je produis:

Je présente à toute l'équipe mon oiseau en introduction de notre deuxième rencontre. C'est la stupéfaction. Le duo n'y croit pas. Comment est-ce possible? Il doit y avoir une erreur.

Pourtant, le groupe, suite à son analyse, constate que chaque critère a été respecté, sauf un, qui posait problème. Le résultat est donc maximal. Interloquées, les conceptrices me demandent comment, sur la base de leurs documents, il a été possible de dessiner «ça»? En d'autres termes, elles cherchent à comprendre comment, malgré des critères et des consignes détaillés, un élève pourrait obtenir une excellente appréciation¹ tout en passant complètement à côté du produit attendu et des apprentissages évalués.

Ce phénomène, que j'ai appelé «effet du hibou» (Pasquini, 2013), traduit un manque de validité dans l'évaluation (Brookhart, 2005), c'est-à-dire un déficit de cohérence entre ce qui a été enseigné, ce que les enseignants pensent évaluer, ce qu'ils évaluent réellement et les résultats des élèves au regard des objectifs en jeu, et cela même en présence d'un guidage précis et de critères d'évaluation.

L'«effet du hibou» marque les esprits, de la 3P au post-obligatoire et dans toutes les disciplines, car il rend visible un décalage que de nombreux enseignants constatent spontanément dans leurs évaluations sommatives, entre ce qu'ils «ressentent» ou attendent des productions des élèves, les outils qu'ils utilisent pour permettre ces productions, et les notes qu'ils attribuent. C'est bien de cet effet dont il s'agit quand des enseignant-e-s de langue disent, par exemple, ne pas comprendre comment un texte perçu comme «pauvre» peut obtenir un 5, ou au contraire, un autre, estimé «intéressant», n'être sanctionné que par un 3,5. Et c'est encore lui qui est en jeu dans toutes les autres disciplines, lorsque les enseignant-e-s observent que des élèves peuvent obtenir un 4 alors que de nombreuses lacunes persistent, ou sont en échec alors qu'ils font preuve de maîtrises avérées.

Ce problème de validité, courant, très (trop) souvent relevé dans les pratiques d'évaluation sommative, a également été abondamment documenté par la recherche (e. g. Black, Harrison, Hodgen, Marshall et Serret, 2010). Il y est décrit comme encore plus pénalisant dans des systèmes scolaires où les trajectoires des élèves sont essentiellement déterminées par les notes et les moyennes, comme en Suisse romande. Il y a donc urgence à le juguler.



Retour en salle de travail. Pour éviter cet effet, nous avons analysé les documents produits par notre duo en nous basant sur quelques questions fondamentales lorsqu'il s'agit d'évaluation sommative:

- Quel-s contenu-s et quel-s objectif-s sont à évaluer? (pour rester centré sur les apprentissages)
- Comment ces contenus ont-ils été enseignés, ces objectifs travaillés avec les élèves? (afin d'éviter l'inédit, l'aléatoire)
- Quelle-s situation-s d'évaluation (tâches, consignes, items) permet-tent le mieux d'évaluer l'apprentissage en question? (parce que toutes les situations ne se valent pas, ne mettent pas les mêmes éléments de l'apprentissage en évidence)
- Comment pondérer ces situations dans l'épreuve en référence à l'objectif et aux contenus à maîtriser: des points, des critères, les deux? (pour mettre en évidence la valeur de l'apprentissage évalué)
- Quel type de barème conviendra le mieux pour rendre compte des performances des élèves et construire les notes? (partant de l'idée que le barème devrait être en rapport avec l'apprentissage évalué)

Ces questions, très parlantes pour l'équipe, ont permis d'identifier des gestes fondamentaux pour une validité de l'évaluation: expliciter les apprentissages évalués, formuler des consignes, des questions, des tâches et des critères référés à ces apprentissages, éviter tout système de pondération et de notation de type quantitatif et/ou normatif. Autant de caractéristiques, de fait, participant à une évaluation sommative «juste».

Notre réflexion continuera fidèlement à ce premier lot d'éléments réflexifs. Ainsi, en repartant de questions issues du terrain, nous nous demanderons le mois prochain «comment savoir que les tâches évaluent vraiment ce que nous pensons évaluer?». Puis, le mois d'après, nous aborderons la problématique «critères ou points?» et leur rôle respectif de mise en évidence de la valeur de l'apprentissage évalué. Ensuite, dans un continuum, nous réfléchirons à la construction de la note: avec quels barèmes, quelles échelles? Nous concluons notre série de chroniques en donnant des éléments de synthèse d'une évaluation sommative la plus valide possible, en espérant qu'ils permettront de poursuivre la réflexion au sein des équipes, voire des établissements. Pour tenter d'éviter, le plus possible, la reproduction de ce discriminant «effet du hibou», tout en tenant compte des contextes et des contraintes de chacun-e.

¹ Le problème est exactement le même avec des notes chiffrées

Pour approfondir

Black P., Harrison C., Hodgen J., Marshall B. & Serret N. (2010). *Validity in teachers' summative assessments. Assessment in Education Principles, Policy & Practice*, 17 (2), 215-232.

Brookhart S. M. (2005). *Assessment Theory for College Classrooms. New directions for teaching and learning*, 100, 5-14.

Pasquini R. (2013, mai). *Former à l'évaluation en référence au PER, ou comment amener des enseignants à passer d'une logique d'exécution à une logique de conception?* Conférence présentée à la Journée d'étude IRDP, Admée-Europe, AIRDF «L'évaluation en classe. Les pratiques d'évaluation des enseignants: que nous apprennent-elles?», Neuchâtel. Repéré à www.irdp.ch/data/secure/1537/document/presentation_2_2013.pdf