

L'évaluation par les pairs dans les Moocs : quelques repères

▲ www.adjectif.net/spip/spip.php



Pour citer cet article :

Khaneboubi, Mehdi (2014). L'évaluation par les pairs dans les Moocs, quelques repères. *Adjectif.net* [En ligne], mis en ligne le 25 novembre 2014. URL : <http://www.adjectif.net/spip/spip.php?article323>

Résumé :

Ce texte fait le point sur la correction de devoirs par les pairs au sein de Moocs.

Mots clés :

Évaluations, MOOC



par Mehdi Khaneboubi

Le nombre important de participants dans les cours en ligne massivement ouverts (c'est à dire les MOOC) rend strictement impossible une correction par des enseignants de chaque production rendue. C'est pourquoi lorsqu'une évaluation automatique est impossible, les apprenants se notent entre eux. Comme pour les pédagogies mutuelles du 19e siècle (Gréard, 1911), le système didactique peut être influencé par un critère économique.

Ce système de notation est-il apparu avec les Mooc ? Quelle est la fiabilité de cette évaluation ? Quels sont les écarts entre la notation d'un enseignant et celle des élèves ? Dans ce texte nous allons essayer de présenter des éléments de bibliographie qui répondent à ces questions.

Tout d'abord, l'évaluation par les pairs est une activité antérieure à l'arrivée des MOOC. Auparavant, cette méthode était utilisée principalement pour des raisons pédagogiques (Sadler & Good, 2006 ; Thomson, Smith, & Annesley, 2014 ; Topping, 1998). On peut dire, d'un point de vue statistique, que les notes d'enseignants et d'apprenants évaluant les mêmes travaux sont très fortement corrélées (Kulkarni et al., 2013 ; Sadler & Good, 2006) mais pas suffisamment pour que ce soit jugé satisfaisant d'un point de vue pédagogique.

Dans le cas d'une évaluation par les pairs réalisée par des élèves de sciences dans un lycée des USA au début des années 2000, Sadler & Good (2006) ont remarqué un phénomène plutôt normatif de la notation des élèves : les « bons » ont de moins bonnes notes et les « mauvais » de meilleures que celle donnée par le professeur. La proximité entre les évaluations des élèves est celle de l'enseignant est très variable selon les tâches à réaliser, les critères d'évaluation, l'expérience des élèves, etc. En étudiant les évaluations par les pairs réalisées dans un cours d'algorithmique, Chinn (2005) affirme que les étudiants évaluent de mieux en mieux à mesure que le cours avance et qu'il existe un lien fort entre la qualité de leur évaluation et leur performance dans les évaluations ordinaires.

Dans une étude qui concerne l'enseignement supérieur, Falchikov & Goldfinch (2000) estiment que cette proximité est plus grande lorsque l'évaluation repose sur des critères généraux bien compris par les élèves

plutôt que par une série d'items uniques. De plus, les évaluations des élèves et des enseignants se ressemblent davantage pour des procédures et des méthodes plutôt que des pratiques professionnelles.

Enfin, les auteurs ne trouvent pas de meilleurs résultats en sciences et sciences de l'ingénieur que dans les autres disciplines ni entre les étudiants de premier et de deuxième cycles. Dans le cas d'un enseignement professionnel en Anglais langue étrangère au Japon, Saito & Fujita (2009) rapportent une similarité globale entre la notation des élèves et celle des enseignants avec des différences notables selon la difficulté des critères d'évaluation.

Statistiquement, la corrélation est importante entre notation des apprenants et de l'enseignant, mais d'un point de vue pédagogique il n'est pas possible de considérer les notes d'élèves comme valides sans une (ré)vision de ces notes par un enseignant (Salder & Good, 2006). C'est pourquoi différentes méthodes existent pour pondérer, corriger ou assister un enseignant dans la validation des notations d'apprenants de MOOC.

Par exemple la Peer rank method (Walsh, 2014) est directement inspirée de l'algorithme présumé du moteur de recherche Google. Il s'agit de donner un poids équivalent aux notes données et aux notes reçues : la notation d'un apprenant ayant de bonnes notes a plus d'importance que celle d'un apprenant avec de mauvaises.

L'Automated Essay Scoring utilisé dans un plug in d'Edx et le Calibrated Peer Review utilisé dans Coursera, correspond à des systèmes d'évaluation par les pairs de textes simples et courts avec une intervention automatique permettant de réduire l'intervention des enseignants ou de les aider dans la validation. Cela se base sur des méthodes statistiques, des algorithmes de machine learning et du traitement automatique de texte (Balfour, 2013). D'après Piech et al. (2013) ces procédures peuvent être encore améliorées par des méthodes algorithmiques testées dans Coursera avec des améliorations significatives.

Nous en sommes en tout cas à une période de transition, où il est difficile de prédire ce qui va se développer et être durable. On peut faire l'hypothèse que les évolutions ne se produiront que très progressivement tout ce qui est en contradiction avec une organisation traditionnelle est probablement condamné à n'avoir qu'un essor temporaire.

Références

- Balfour, S. P. (2013). Assessing Writing in MOOCs : Automated Scoring and Calibrated Peer Review. *Research and Practice in Assessment*, 8(Summer), 40–48. <http://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF4.pdf>
- Chinn, D. (2005). Peer Assessment in the Algorithms Course. In *Proceedings of the 10th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education* (p. 69–73). New York, NY, USA : ACM.
- Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education : A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, 70(3), 287-322.
- Gréard, O. (1911). Mutuel (enseignement). In F. Buisson (Éd.), *Nouveau dictionnaire de pédagogie et d'instruction primaire*. Lyon : INRP. <http://www.inrp.fr/edition-electronique/lodel/dictionnaire-ferdinand-buisson/document.php?id=3249>
- Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Klemmer, S. R. (2013). Peer and Self Assessment in Massive Online Classes. *ACM Trans. Comput.-Hum. Interact.*, 20(6), 33:1–33:31.
- Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned models of peer assessment in MOOCs. arXiv preprint arXiv:1307.2579. <http://arxiv.org/abs/1307.2579>
- Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational assessment*, 11(1), 1–31. http://www.tandfonline.com/doi/abs/10.1207/s15326977ea1101_1
- Saito, H., & Fujita, T. (2009). Peer-Assessing Peers' Contribution to EFL Group Presentations. *RELC Journal*,

40(2), 149-171.

Thomson, P., Smith, A., & Annesley, S. (2014). Exploration of the effects of peer teaching of research on students in an undergraduate nursing programme. *Journal of Research in Nursing*.

Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, 68(3), 249-276.

Walsh, T. (2014). The PeerRank Method for Peer Assessment. <http://arxiv.org/abs/1405.7192>