

# **Quand plusieurs correcteurs évaluent les même tâches complexes : regards sur la concordance inter-juges et point de vue des évaluateurs**

Pierre-François Coen et Sheila Pellegrini

Haute école pédagogique de Fribourg

La plupart des Hautes écoles pédagogiques (HEP) de Suisse romande ont embrassé des logiques de formation à partir de référentiels de compétences. Outre les changements curriculaires et structurels que cette approche nécessite, les institutions ont dû intégrer de nouvelles formes d'évaluation notamment l'évaluation de compétences par tâches complexes. Cette recherche s'attache à la manière dont les formateurs produisent leur jugement sur ces tâches et examine les quatre questions suivantes : 1) Quelle variabilité y a-t-il entre les examinateurs et à quoi est-elle due ? 2) Quelle est l'importance qu'ils accordent aux différents critères utilisés ? 3) Quelle est la confiance qu'ils accordent à leur jugement ? 4) Quelle procédure mettent-ils en place pour le produire ? Dans une perspective exploratoire, cette étude se penche sur deux tâches complexes dispensées dans la HEP de Fribourg et soumises au regard de 8 correcteurs issus de deux groupes de formateurs : des professeurs de l'institution et des formateurs praticiens (enseignants titulaires engagés à temps partiel par l'institution et associés à l'évaluation des tâches complexes). Les résultats nous permettront de voir dans quelle mesure apparaissent des différences entre les deux types de correcteurs et nous donnera également l'occasion de faire des hypothèses en lien avec le niveau d'expertise des examinateurs et le niveau de confiance qu'ils accordent à leur jugement.

Mots-clés : évaluation des compétences, concordance inter-juges, point de vue de l'évaluateur, tâches complexes

La formation des enseignants a subi en Suisse, ces dernières années, des transformations importantes tant au niveau structurel que curriculaire. Dans ce contexte, l'adoption de référentiels de compétences constitue un élément important qui a produit des incidences indéniables sur les procédures d'évaluation. En dépit du caractère instable de la notion de compétence (Crahay, 2006), la plupart des Hautes écoles pédagogiques (HEP) de Suisse romande ont embrassé cette logique et certaines d'entre elles ont choisi de mettre sur pied des évaluations par tâches complexes pour certifier les modules de formation. C'est notamment le cas de la HEP de Fribourg dans laquelle cette recherche a été conduite.

Selon nombre d'auteurs (Roegiers, 2004 ; Scallon, 2004 ; Tardif, 2006, Gérard, 2009), il semble que le recours à des tâches complexes convienne particulièrement bien à l'évaluation de compétences. En effet, de par son caractère intégrateur, une tâche complexe nécessite la mobilisation de ressources diversifiées que l'apprenant doit utiliser de manière adéquate en tenant compte des spécificités de la situation - plus ou moins inédite - qu'on lui propose. La production est singulière et son évaluation repose sur l'utilisation d'un nombre de critères limités (Roegiers, 2004). L'évaluation revêt ainsi un caractère relativement global qui exige de l'évaluateur une expertise du champ dans lequel s'inscrit la tâche. Scallon (2004) souligne cependant que ce type de production est relativement difficile à évaluer et, avec d'autres, pose la question de la fidélité des examinateurs (concordance inter-juges).

Dans une précédente recherche, nous nous étions attaché à la notion de validité des tâches complexes (Coen et al. 2008) en ce qu'elles représentent bien les compétences visées. Il nous semblait cohérent de poursuivre nos investigations en abordant la question à partir du regard des correcteurs (le point de vue des évalués est examiné par une autre équipe de notre

institution). Même si elle n'est pas récente et qu'elle occupe les docimologues et les statisticiens depuis de nombreuses années (voir Abernot, 1996), la fidélité entre évaluateurs revêt à nos yeux un caractère particulièrement important dans des institutions de formation où l'évaluation des compétences par tâches complexes a presque totalement remplacé les tests classiques, objets des études effectuées durant ces dernières décennies.

Par ailleurs et en dépit de critères explicites, les évaluateurs interprètent bien souvent la production des apprenants à l'aune de plusieurs référentiels. En ce sens, ils opèrent un jugement professionnel (Lafortune et Allal, 2008) basé sur la collecte d'informations auxquelles tous n'accordent pas la même importance. Cet élément pourrait expliquer les éventuelles différences de jugements entre évaluateurs particulièrement lorsqu'on se situe aux abords du seuil de maîtrise. Notons encore que l'ancrage culturel des évaluateurs (dans notre cas il y a deux types d'évaluateurs : des formateurs praticiens et des professeurs institutionnels) pourrait avoir une incidence sur les postures d'évaluateur (Jorro, 2009) et interférer sur ce processus.

L'étude exploratoire que nous présentons se penche sur ces différents aspects au travers des questions suivantes : 1) Quelle variabilité y a-t-il entre différents examinateurs pour l'évaluation d'une même tâche complexe ? 2) Quelle importance accordent-ils aux différents critères pris en compte et comment s'opère cette pondération ? 3) Quelle est la confiance qu'ils accordent à leur jugement et de quoi dépend-elle ? 4) Quelle procédure mettent-ils en œuvre pour produire leur jugement final, sont-elles différentes et sur quoi se basent-elles ?

Pour répondre à ces questions nous avons sélectionné deux tâches complexes - faisant partie des épreuves soumises aux étudiants de la HEP de Fribourg - de 6 étudiants différents. Toutes ces tâches avaient déjà fait l'objet d'une précédente validation (dans un contexte officiel). La note obtenue lors de cette première évaluation nous a permis de choisir des épreuves qui se situaient dans des niveaux de réussites différents : bonne réussite, réussite proche du seuil de maîtrise et échec. Nous avons soumises ces différentes tâches au regard de 8 évaluateurs (professeurs de l'institution et formateurs praticiens – enseignants de terrain associés aux modules de formation) qui pour chacune d'elle ont produit un score. Pour compléter notre collecte d'information, nous leur avons demandé d'associer un niveau de confiance (sur une échelle allant de « Je suis tout à fait sûr de mon jugement » à « Je doute beaucoup de mon jugement »). Les évaluateurs devaient en outre estimer leur niveau d'expertise dans le champ de compétences évaluées. Ils ont ensuite été interrogés via des entretiens dirigés afin d'explicitier les processus mis en œuvre lors de l'évaluation (importance donnée à tel critère, choix des informations jugées importantes, etc.).

Les résultats nous permettront de mettre en évidence d'éventuelles différences de jugement entre nos évaluateurs (formateurs de terrain et formateurs institutionnels) mais il nous sera également possible de mettre ces constats en lien à la fois avec la confiance que les correcteurs associent aux scores donnés et leur niveau d'expertise dans le champs de compétences évaluées par la tâche complexe. Un questionnement sur les procédures mises en place dans la HEP ne manquera d'émerger notamment sur la nécessité d'envisager d'éventuelles corrections multiples, si ce n'est dans tous les cas au moins lorsque l'expertise du correcteur est douteuse ou que la confiance qu'il accorde à son jugement est insuffisante.

Abernot, Y. (1996). *Les Méthodes d'évaluation scolaire*. Paris: Dunod.

Allal, L. (2008). *Jugement professionnel en évaluation : pratiques enseignantes au Québec et à Genève*. Québec: Presses de l'Université du Québec.

Coen, P.-F, Galland, A., Monnard, I., Perrin, N., & Rouiller, Y. (2008). Pertinence et validité de tâches complexes dans l'évaluation de futurs enseignants. Dans G. Baillat, J. De Ketele, L. Paquay, &

- C. Thélot (Éd.), *Évaluer pour former. Outils, dispositifs et acteurs* (p. 159-175). Bruxelles: De Boeck.
- De Landseere, G. (1980). *Évaluation continue et examens, précis de docimologie*. Bruxelles: Labor-Nathan.
- Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Revue française de pédagogie*, 154, 97-110.
- Gérard, F.-M. (2009). *Évaluer des compétences, guide pratique*. Bruxelles: De Boeck.
- Jorro, A. (2009). L'évaluation comme savoir professionnel. Dans L. Mottier Lopez & M. Crahay (Éd.), *Évaluation en tension. Entre la régulation des apprentissages et le pilotage des systèmes* (p. 219-231). Bruxelles: De Boeck.
- Roegiers, X. (2004). *L'école et l'évaluation. Des situations pour évaluer les compétences des élèves*. Bruxelles : De Boeck.
- Scallon, G. (2004). *L'évaluation apprentissages dans une approche par compétences*. Bruxelles : De Boeck.
- Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal: La Chenelière.