



### Thématique

5. RECHERCHE : L'étude de la validité des dispositifs d'évaluation et de leur contenu.

### Titre de la communication

Du curriculum à l'évaluation en Suisse romande: comment caractériser des tâches d'évaluation en référence au Plan d'études romand

### Mots-clés (3)

- 1- Plan d'études
- 2- Validité
- 3- Accord interjuge

### Résumé court

Avec des pratiques évaluatives très hétérogènes (Wirthner&Ntamakiliro, 2008), la mise sur pied d'un système d'évaluation intercantonale demeure un sujet délicat en Suisse romande. Si un curriculum commun – le Plan d'études romand (PER) – a été adopté en vue d'harmoniser les pratiques, c'est bien celles d'enseignement apprentissage qui étaient visées, et la traduction de ces objets en objets d'évaluation constitue une étape qu'il s'agit encore d'investiguer. Cette communication s'inscrit dans le cadre d'une recherche visant à caractériser les pratiques évaluatives des cantons romands et nous aborderons ici une étape de ce projet : la mise en relation (appariement) de tâches d'évaluation issues des épreuves cantonales de mathématiques avec des éléments du PER (Voir Brown, 1996, sur le Content validity). Nous présenterons les résultats d'un exercice de caractérisation de tâches évaluatives confié à cinq enseignants, ayant tous suivi une formation sur le processus de codage, ainsi que sur les catégories utilisées. Force sera de constater que la consistance dans le codage de certaines tâches est faible. En partant des résultats de l'accord entre les participants (accord interjuge), nous explorerons la question suivante : sur la base de quels critères est-il possible d'attribuer un ou des élément(s) du PER à une tâche d'évaluation ou, à l'inverse, comment caractériser les éléments du PER afin de les traduire en activités d'évaluation?

### Résumé long

Traditionnellement décentralisé, le système éducatif suisse tend à s'homogénéiser depuis une dizaine d'années. Cette tendance s'est traduite notamment par l'adoption de standards nationaux de compétence (telles que définies par le Concordat pour l'harmonisation de la scolarité obligatoire – HarmoS) et de plans d'études pour les trois régions linguistiques : Suisse alémanique, italienne et romande. En Suisse romande, le Plan d'études romand (PER) est successivement entré en vigueur dans tous les cantons depuis 2010. Parallèlement, la Convention régissant la coopération entre les cantons romands pour les questions scolaires (Convention scolaire romande) prévoit la création de moyens d'enseignement et la mise en place d'évaluations (épreuves romandes communes – EpRoCom) "afin de mesurer l'atteinte des objectifs du Plan d'études romand". Cependant, et bien que de nombreuses études aient vu le jour ces dernières années sur les futures EpRoCom (Marc et Wirthner, 2012), les autorités politiques ont finalement décidé de retarder la création d'une épreuve commune à proprement parler et de mettre à disposition des enseignants, dans un premier temps, une banque d'items en lien avec le PER.

Etant chargés de conduire les travaux conduisant à la création de cette banque d'items et de mener des réflexions sur sa fonction et son contenu, nous nous sommes principalement interrogés sur la définition d'une ligne d'évaluation commune pour la Suisse romande, car les pratiques cantonales demeurent très hétérogènes. Mais la tâche s'est avérée complexe, car si le PER a été créé en vue d'harmoniser les pratiques, c'est bien celles d'enseignement-apprentissage qui étaient visées et la question de l'évaluation n'y est nullement abordée. Ainsi, la traduction des objets d'enseignement-apprentissage en objets d'évaluation constitue une étape indispensable dans la validation du dispositif d'évaluation romand, qu'il s'agit encore d'investiguer. Comme l'attestent notamment les travaux de Brown (1996) et Messick (1989), il s'agit de s'assurer que les items que l'on construira mesureront vraiment ce que l'on souhaite évaluer. Cette communication s'inscrit dans ce contexte et abordera une recherche visant à caractériser les pratiques romandes en matière d'évaluation en mathématiques. Nous inspirant de la méthodologie de profiling développée par Ruiz-Primo et Li (2015), nous avons développé un outil d'analyse et de classification des items de mathématiques (voir Carulla and al., 2015), qui permet de caractériser les items au niveau du contenu (par rapport au PER et à HarmoS), ainsi qu'au niveau de la forme (énoncés, types de réponses, particularités des contextes, etc.). Nous nous focaliserons sur une étape exploratoire du projet : la mise en correspondance entre des items extraits des épreuves cantonales et des éléments du PER. Cet exercice de classement des tâches évaluatives a été proposé à cinq enseignants de profils différents (enseignants généralistes, spécialisés et spécialistes des mathématiques) Les résultats du codage de certains items révèlent une faible consistance entre les participants (i.e. une même tâche n'étant pas classée de la même façon par tous les codeurs). Les exercices de codage et de mise en accord ont néanmoins permis d'affiner et parfois d'explicitier certaines règles de classification qui étaient implicites, notamment sur des éléments du PER ayant pour objet des tâches assez concrètes. Le travail de traduction des éléments du PER en éléments d'évaluation observables n'en est qu'à son début et nécessitera certainement la participation d'autres acteurs romands. Dans le cadre de cette communication, nous souhaiterions présenter et discuter des résultats du classement des tâches d'évaluation d'une part, et du travail de traduction des objets du PER en objets d'évaluation d'autre part. En partant des résultats de l'accord entre les participants (accord interjuge), nous explorerons par ailleurs la question suivante : sur la base de quels critères est-il possible d'attribuer un ou des élément(s) du PER à une tâche d'évaluation ou, à l'inverse, comment caractériser les éléments du PER afin de les traduire en activités d'évaluation?

**Références :**

Carulla et al. (2015). Outil pour la caractérisation d'items mathématiques d'épreuves externes. Actes de l'ADMEE

2015. [http://admee2015.sciencesconf.org/conference/admee2015/admee15\\_Actes\\_du\\_colloque4incp.pdf](http://admee2015.sciencesconf.org/conference/admee2015/admee15_Actes_du_colloque4incp.pdf) , consultée le 27 janvier 2015.

Brown, J. D. (2000). What is construct validity ?.Shiken: JALT Testing & Evaluation SIG Newsletter, 4 (2) Oct 2000 (p. 8 - 12) <http://jalt.org/test/PDF/Brown8.pdf>, consultée le 4 décembre 2015.

Marc, V. &Wirthner, M. (2013). Développement d'un modèle d'évaluation adapté au PER : rapport scientifique du projet d'épreuves romandes communes. Neuchâtel : Institut de recherche et de documentation pédagogique (IRDP).

Messick, S. ( 1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed.,pp. 13-103). New York: Macmillan.

Ruiz-Primo, M.A., & Li, M. (2015). The Relationship Between Item Context Characteristics and Student Performance. Teachers College Record n°117.