

**Thématique**

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

Titre de la communication

**VALIDITE DES EVALUATIONS :**

**UNE APPROCHE EPISTEMO-DIDACTIQUE**

**Brigitte GRUGEON-ALLYS - Nadine GRAPIN - Université Paris Est Créteil (ESPE) -  
Laboratoire de Didactique André Revuz (LDAR)**

**Résumé court (300 mots)**

Déterminer si une évaluation est valide, c'est-à-dire si elle évalue bien ce qu'elle est sensée évaluer (et uniquement cela) est une question vive, partagée par tout concepteur d'évaluation. L'étude de la validité que nous présentons dans cette communication repose principalement sur une approche épistémologique, même si dans nos travaux (Grugeon & Grapin 2015, Grapin 2015), nous nous appuyons aussi sur des approches psychodidactiques et psychométriques (dans le cas des évaluations externes à grande échelle).

Il s'agit de considérer, dans un domaine mathématique donné, l'ensemble des tâches de l'évaluation au regard d'une organisation mathématique de référence définie à partir d'une étude épistémologique et didactique. La donnée d'une telle référence permet d'effectuer une sélection des tâches de l'évaluation (si l'analyse est réalisée en amont) ou d'analyser l'ensemble des items pour déterminer, par exemple, s'ils couvrent le domaine en étant représentatifs de ce dernier, s'ils sont cohérents avec leur objectif d'évaluation, si le codage de la réponse permet de repérer le raisonnement de l'élève, etc.

Nous considérons que l'analyse de validité se joue d'abord à un niveau local avec une analyse *a priori* de chacune des tâches de l'évaluation. Menée en lien avec la définition de l'organisation mathématique de référence, cette analyse conduit ensuite à une étude globale sur l'ensemble des items pour repérer d'éventuels manques ou redondances dans les types de tâches proposés ou encore une répartition peu équilibrée selon la complexité.

L'analyse de la validité menée à ces deux niveaux permet ensuite une réinterprétation des résultats obtenus à l'évaluation.

Après avoir décrit cet outil d'analyse, nous l'exploiterons pour étudier les bilans CEDRE école et collège, dans leur contenu et dans les résultats produits, dans deux domaines différents : les nombres entiers (incluant la numération, le calcul et la résolution de problèmes arithmétiques en fin d'école) et l'algèbre en fin de collège.

**Mots clés :** évaluation, évaluation externe, validité, psycho-didactique

**Résumé long (1000 mots)**

Une évaluation est considérée comme valide si évalue bien ce pour quoi elle est conçue, et uniquement cela. Différents aspects de la validité peuvent alors être définis (Goldstein 2015) et mis en lien avec la mesure. Nous retenons pour notre travail trois types de preuves de validité listés par Laveault & Grégoire (2014) :

- des preuves liées au contenu de l'évaluation : il s'agit par exemple de déterminer si les items sont représentatifs du domaine évalué, si le domaine est couvert ou non, si les techniques nécessaires à la résolution de la tâche sont bien celles attendues par les concepteurs...

- autre type de preuve de validité qui peut être recherchée, celle liée au processus de réponse : est-ce que les processus de réponse mis en jeu par les élèves sont effectivement

ceux envisagés et attendus par les concepteurs ?

- des types de preuves portant sur la structure interne du test en lien avec l'hypothèse d'unidimensionnalité ; ces questions étant plus spécifiques aux évaluations externes à grande échelle.

Si Laveault et Grégoire soulignent l'intérêt, pour analyser le contenu d'une évaluation, de suivre une méthodologie rigoureuse et que celle-ci se doit être déterminée par des experts de la discipline, ils ne donnent en revanche, pas de méthode pour réaliser cette analyse. Par ailleurs, en didactique des mathématiques, nous trouvons peu de travaux sur la validité du contenu des évaluations. Les travaux de Bodin menés dans les années 1990 à 2000 sont les premiers à interroger les résultats produits par des évaluations externes, en particulier PISA, à partir de leur contenu. D'autres travaux menés plus récemment en France, comme ceux de Roditi et Chesne (2012), Roditi & Salles (2015) ou de Sayac & Grapin (2015), toujours sur le contenu des évaluations externes à grande échelle, ont exploité des outils de la didactique pour analyser le contenu et interpréter les résultats, ces travaux restent peu nombreux et il y a un véritable enjeu à les développer.

L'approche épistémologique-didactique dans laquelle nous nous situons pour étudier la validité se situe principalement dans la Théorie Anthropologique du Didactique (Chevallard, 1994). À partir des tâches proposées dans les évaluations, il s'agit de caractériser les connaissances apprises en perspective du savoir à enseigner et enseigné : au regard des items proposés, et en prenant en compte la transposition du savoir savant en savoir enseigné, quelles sont précisément les inférences qui peuvent être faites à partir des scores ? De ce fait, au-delà d'une étude cognitive, nous fondons l'analyse des connaissances apprises par les élèves, dans un domaine mathématique donné, sur la caractérisation d'une praxéologie épistémologique de référence (Bosch et Gascon, 2005). La définition de cette dernière repose sur une analyse épistémologique des « des savoirs savants » complétée par un point de vue épistémologique ; elle conduit, relativement au domaine étudié, à une description des types de tâches, des techniques permettant de les traiter et des éléments technologiques et théoriques justifiant les techniques.

Nous considérons ensuite que l'analyse de validité se joue d'abord à un niveau local avec une analyse *a priori* de chacun des items ; il s'agit de préciser le type de tâches auquel il correspond, les techniques de résolution possibles, les éléments technico-théoriques sous-jacents. Nous prenons aussi en compte, dans l'analyse des tâches, le niveau de convocation des organisations mathématiques (Castela, 2008) relativement au niveau d'enseignement auquel l'évaluation a lieu ainsi que la congruence sémantique ou non des registres de représentation sémiotique (Duval, 1996).

Cette analyse locale conduit ensuite à une étude globale sur l'ensemble des items pour étudier le recouvrement du domaine par les tâches proposées dans l'évaluation, la représentativité des types de tâches et la variété de leurs niveaux de convocation ; cela permet ainsi de repérer d'éventuels manques ou redondances dans les types de tâches et d'observer s'il existe un certain équilibre selon la complexité, les registres de représentation en jeu, etc. Au delà de la prise en compte des résultats de l'évaluation en termes de réussite ou d'analyse de productions d'élèves pour étudier les techniques employées, il s'agit aussi de nous intéresser plus particulièrement à des items qui présentent des caractéristiques psychométriques spécifiques, comme par exemple, un faible pouvoir de discrimination, un indice de difficulté très élevé (ou au contraire très bas) ou encore, dans le cadre de comparaison temporelle entre 2008 et 2014, d'items présentant des fonctionnements différents différentiels (items biaisés). Pour ces items, une analyse didactique relativement aux savoirs en jeu et à leur enseignement peut permettre d'expliquer leurs caractéristiques statistiques (Roditi & Chesne, 2012) ; et réciproquement, les indicateurs statistiques correspondant à ces items peuvent nous conduire à interroger les pratiques enseignantes (praxéologies enseignées) et les savoirs évalués.

Après avoir exposé la façon dont nous étudions, dans une approche épistémologique-didactique, la validité des évaluations nous montrerons la complémentarité qui existe avec une approche psycho-didactique. La présentation du cadre de conception des évaluations bilans CEDRE et de certains de leurs résultats nous conduira d'abord à présenter et illustrer notre questionnement relativement au domaine de l'algèbre (fin de collège) et à celui des nombres entiers (fin d'école). À partir de ces évaluations et de leurs résultats, nous montrerons comment l'analyse de tâches qui a été effectuée a permis de dégager des éléments de validité dans les deux domaines mathématiques considérés. En conclusion, nous mettrons

en perspective les approches didactiques et psychométriques afin de montrer comment elles peuvent se révéler complémentaires à la fois pour la détermination du contenu de l'évaluation et pour l'analyse des résultats puis nous dégagerons des conditions à proposer pour "renforcer" la validité dans les dispositifs d'évaluation externe.