

La validation des épreuves d'évaluation selon l'approche par les compétences

Georgina Clark, Ariane Gardi, Nadia Lutz

Présentation

Ce travail aborde la problématique « Comment mettre en place une méthode de validation des tests d'évaluation de compétences ? ». Le sujet est présenté sous la forme d'un mode d'emploi basé sur l'article de De Ketele & Gérard (2005) dont les affirmations sont reprises sans autres indications.

Notre conclusion présente les sujets, qui, de notre point de vue, nécessiteraient des éclaircissements supplémentaires avant que ce mode d'emploi devienne totalement opérationnel.

Le mode d'emploi

Quelle approche ?

Tout d'abord, il faut savoir dans quelle approche pédagogique on se situe :

- par contenu
- par objectif
- par compétences

... et adapter les modalités d'évaluation en conséquence.

Dans toute démarche de validation, il est essentiel de tenir compte de trois critères : pertinence, validité¹, fiabilité. Ces trois critères doivent être remplis en même temps. La plupart des exemples connus ne remplissent pas les trois critères:

Les épreuves classiques (TIMMS et PISA) sont valides, fiables mais pas pertinentes car les tests (sous forme d'items) sont issus d'une approche par contenus et/ou objectifs, alors qu'ils prétendent évaluer les compétences de résolution de situations complexes.

Par ailleurs, les épreuves actuelles pour l'évaluation des compétences sont pertinentes car elles proposent des situations complexes, mais il est difficile d'en attester la validité, parce que les méthodes pour assurer cette validation ne sont pas évidentes ou tout simplement pas connues.²

L'approche par compétence

L'approche par compétences cherche à développer la possibilité par les apprenants de mobiliser un ensemble intégré de ressources pour résoudre une situation-problème appartenant à une famille de situations. Dans l'optique des compétences, évaluer consiste à proposer une ou des situations complexes, appartenant à la famille de situations définie par la compétence, qui nécessitera(ont) de la part de l'élève une production elle-même complexe

¹ Le terme validité est ambigu. Il s'agit ici d'une validité de type « interne ». La pertinence correspond à une validité « externe ». C'est à cette validité que s'adresse le titre.

² Malheureusement l'article de propose aucun exemple concret illustrant ce point. On peut toutefois facilement imaginer qu'une épreuve du Rallye mathématique transalpin (<http://www.rmt-sr.ch/>) répond à ce critère.

pour résoudre la situation. Dans l'article, l'idée de compétences de base est largement sous-entendue³.

Si l'on se centre sur une telle approche, il faut faire attention aux deux points suivants.

Limites de la théorie classique des scores

L'utilisation de techniques de la psychométrie (scores) dans l'évaluation pédagogique des savoirs et des compétences n'est pas pertinent. En effet, en psychométrie on se réfère à la courbe de Gauss et non à la courbe en J. Or, dans une perspective pédagogique de la réussite, il semble logique que la majorité des élèves acquièrent les compétences de base qui sont visées et ceci s'exprime par la courbe en J. En conséquence, l'usage (très répandu) des indices de discrimination (indices qui permettent de faire une distinction entre élèves forts et élèves faibles) n'est pas approprié.

Les postulats de la théorie classique des scores ne sont donc pas applicables à la constitution d'une épreuve d'évaluation selon l'approche par compétences. De nouvelles techniques restent à inventer.

Limites des épreuves selon l'approche par compétences

Pour qu'un échantillon soit représentatif, il faut qu'il soit constitué de manière aléatoire : cela devrait être de même dans la sélection des items pour une épreuve évaluant les compétences. De plus, pour éviter l'erreur de mesure (donc, pour pas qu'il y ait des élèves qui se trouvent face à des questions plus difficiles que d'autres) il faudrait qu'il y ait un grand nombre d'items. Pourtant, le temps à disposition pour ces épreuves est limité, étant donné qu'il faut proposer des situations complexes, demandant de la part des élèves un certain temps (et de l'énergie) pour les résoudre. En outre, une épreuve par compétences ne peut pas être décomposée en items (sous-unités), car ceci ne permettrait plus d'évaluer la compétence qui consiste bien à gérer la complexité. De conséquence, le problème de la représentativité des situations est augmenté.

Problème de validation : le cas de l'approche par compétence

Problèmes de la validation a priori par recours à des juges

Il faut créer des situations comparables, car les élèves devraient pouvoir travailler sur plusieurs situations équivalentes. Cependant, il est difficile pour les juges d'évaluer l'équivalence des situations complexes. Il est donc indispensable de définir des paramètres garantissant l'équivalence des situations.

Problèmes de la validation empirique a posteriori

Validation empirique interne

La difficulté de la validation empirique interne d'une épreuve élaborée selon l'approche par les compétences et constituée d'une ou plusieurs situations complexes est qu'il n'existe ni de dimension unique, ni un ensemble dénombré d'items censés la mesurer puisqu'il s'agit d'une production complexe.

³ Le thème des compétences est actuellement l'objet d'un large débat. Un cahier des Sciences de l'éducation (anciennement du Service de pédagogie expérimentale) de l'université de Liège est consacré au sujet (21-22, 2005, Les compétences : concepts et enjeux). Ce cahier débute par l'article très critique de M. Crahay repris dans la Revue Française de pédagogie (Crahay, 2006). Le site <http://educmath.inrp.fr> consacre une rubrique « En débat » à ce sujet qui donne la parole à Y. Matheron, B. Schneuwly et A. Sierpiska à propos de la citation de M. Crahay : « les compétences : une mauvaise réponse à un vrai problème ». L'Éducateur consacre également un dossier au sujet (4/2007, mars).

Pour chaque compétence évaluée, il importe donc de déterminer les critères qui sont utilisés pour évaluer les productions des élèves. Chacun de ces critères pourra être opérationnalisé dans une situation à travers des indicateurs qui sont des éléments directement observables, la plupart du temps spécifiques à la situation. Attention: on ne s'intéresse pas à la justesse de la réponse de l'élève mais on essaye plutôt de repérer la présence d'indicateurs / d'éléments observables relatifs au critère en question.

Dans l'approche par compétences, toutes les corrélations inter-situations à propos d'un même critère devraient être proches de 1 et supérieures à celles appartenant à des critères différents.

L'apprentissage dans le cadre de l'approche par compétences implique de présenter aux élèves des situations progressivement plus complexes : l'élève mobilise dans une situation X+1 des ressources déjà mobilisées dans une situation X. Les situations peuvent être complexes à deux niveaux :

- soit parce qu'il y a un rajout de critères de même niveau de difficulté à la situation (X+1) (dans ce cas, le niveau d'exigence reste le même dans les situations X et les situations X +1).
- soit parce que le niveau d'exigence augmente par rapport à un même critère. (p.ex. : si dans une première étape, pour un critère « correction orthographique », on prend en compte des règles simples, dans une deuxième, on tiendra compte de règles plus complexes).

Validation empirique externe

L'évaluation de ces épreuves pourrait être complétée par des études comparatives entre un groupe expérimental et un groupe contrôle équivalent.

Il existe deux cas de figure pour constituer le groupe contrôle :

- Tous les enfants apprennent grâce à une approche par compétences mais le groupe contrôle n'a pas encore eu l'occasion de mobiliser les ressources en question, contrairement au groupe expérimental.
- Les enfants du groupe contrôle n'ont pas appris grâce à une approche par compétence contrairement au groupe expérimental.

La maîtrise de chacun des critères qui évaluent la compétence devrait être supérieure dans le groupe expérimental. Pour approfondir les comparaisons, il faudrait construire une épreuve en 2 parties : la 1^{ère} évaluerait les ressources et la 2^{ème} évaluerait les compétences. Au niveau de la maîtrise des ressources, le groupe expérimental devrait obtenir de meilleurs résultats que le groupe contrôle équivalent.

On peut également recourir à une étude corrélationnelle : on évalue le même groupe d'élèves plusieurs fois. Une cohérence au niveau des productions personnelles de chaque élève devrait être observée, tant dans la durée que par rapport à chaque critère.

Problème de la certification par la preuve d'une note chiffrée

Problème de la « preuve » : on ne sait toujours pas si la maîtrise des différents critères est la preuve de la maîtrise de la compétence visée. (Ce n'est pas parce que l'élève a rempli les différents critères, qu'il maîtrise la compétence dans sa globalité).

- La maîtrise de la compétence est complexe : le degré d'interrelation entre les différents critères n'est pas connu.
- L'évaluation de la compétence est relative aux critères utilisés, à leur nombre, leur opérationnalisation, leur statut.

Lorsque l'on veut traduire la maîtrise d'une compétence en note chiffrée (comme le veulent habituellement les organismes institutionnels), on se rend compte que la note unique ne correspond pas à la performance observée, car elle est trop simpliste en comparaison à la complexité de la compétence en question. De plus, la classification des élèves sur une échelle unidimensionnelle en fonction des notes n'a pas d'importance car il s'agit ici d'une performance multidimensionnelle.

Faut-il donc abandonner la note chiffrée ? Faut-il recourir à une stratégie sommative puisée dans l'évaluation certificative ? L'enjeu est important car il s'agit de déterminer l'échec ou la réussite de l'élève ! Il serait intéressant de développer la notion de profil en rapport aux notions de réussite et d'échec en explorant d'autres recherches.

L'évaluation certificative pour décider la réussite ou l'échec

Dans l'évaluation certificative la décision de réussite est :

- pertinente si l'élève maîtrise les compétences minimales nécessaires à la tâche scolaire de niveau supérieur ou professionnelle.
- valide si l'élève maîtrise effectivement ce qu'il est censé maîtriser.
- fiable si la décision ne dépend pas des circonstances ou de l'évaluateur.

Critères de **validation a priori** pour une évaluation certificative (au niveau de la pertinence) :

1. la validation porte sur des compétences minimales
2. elle distingue les critères minimaux et ceux de perfectionnement
3. elle détermine adéquatement les seuils de maîtrise minimaux
4. elle est menée une fois les apprentissages achevés (règle souvent enfreinte dans le cadre de l'école !)

Une validation externe peut compléter la validation interne : par exemple si l'élève entame son apprentissage sans difficulté.

Critères d'évaluation de la validité : La décision qui est prise sur la base d'une situation X sera valide si la décision peut également être prise avec une situation Y ou Z, censée appartenir à la même famille de situations et évaluée dans les mêmes conditions.

Critères d'évaluation de la fiabilité : La décision prise sur la base d'une situation X sera fiable si des évaluateurs différents prennent la même décision. La fiabilité peut être vérifiée par un coefficient PHI de Guilford ou un coefficient k de Cohen.

L'évaluation certificative pour attribuer une note

Pour attribuer une note qui prenne en compte la pertinence, la validité et la fiabilité, il faut aussi tenir compte de trois critères supplémentaires :

5. 75% de l'échelle doit être réservée pour les critères minimaux ; 25% pour les critères de perfectionnement
6. $2/3$ = seuil de maîtrise minimal
7. seuil de réussite à 50% de l'échelle de notes (pour minimiser réussites et échecs abusifs)

Une fois les notes attribuées (**validation a posteriori**), il faudra veiller à remplir les critères suivants :

1. Distribution des effectifs sous la forme d'une courbe en J

2. Corrélation proche de +1 dans la performance d'un même élève à des tâches d'une même famille
3. Etudier la corrélation entre la distribution des résultats sur la situation X avec les résultats obtenus avec une ou plusieurs variables critère (validation empirique externe)
4. Calculer un coefficient de fiabilité (corrélation) entre notes attribuées par des évaluateurs deux à deux ou coefficient de Kendall.

Attention : Pour des évaluations intermédiaires, il faudrait déterminer les types d'élève : en très grande difficulté ou à réorienter (-25%), les élèves qui réussissent mais ont besoin d'une consolidation (entre 50% et 74%), les élèves sans problème (+75%).

Conclusion de l'article

La validation d'épreuves construites en cohérence avec l'approche par compétences est loin d'être un problème résolu de façon satisfaisante. Les éléments évoqués dans ce travail ne sont par conséquent que des pistes pour la construction d'un outil d'évaluation de tâches complexes (édumétrie).

Nous avons pu voir que la pertinence est centrale dans la validation et à prendre en compte contrairement à ce que font les méthodes d'évaluation actuelles qui ne sont que valides et fiables. Si on prend une métaphore, on pourrait dire que ces épreuves se centrent sur une cible statique, ne permettant donc pas de prendre en compte la capacité réelle de l'élève chasseur de centrer une cible amovible, telle qu'elle existe dans la réalité !

Notre conclusion

Comme nous avons pu le relever dans ce travail, les postulats de la théorie classique des scores ne sont pas applicables à la constitution d'une épreuve d'évaluation selon l'approche par compétences. Dans ce mode d'emploi nous avons essayé d'esquisser des pistes pour mettre en œuvre des épreuves d'évaluation adaptées à cette approche. Cependant, nous avons constaté la difficulté qu'implique cette démarche, et il est évident que des références supplémentaires sont nécessaires pour créer de nouvelles techniques qui soient opérationnelles.

En effet, quelque points restent obscurs et devraient être approfondis dans des recherches futures. Par exemple, les auteurs de l'article de référence évoquent la nécessité de mettre en place une épreuve en deux parties pour comparer un groupe expérimental et un groupe contrôle équivalent. Si pour eux il s'agit dans un premier temps de comparer les ressources des élèves lors de la résolution des problèmes, et dans un deuxième temps les compétences, ils n'expliquent pourtant pas comment ces épreuves peuvent être mises en place.

Un autre point qui mérite d'être approfondi est celui qui fait référence à la situation de réussite et d'échec de l'élève. Les auteurs proposent d'utiliser une sorte de « profil » pour déterminer la situation de l'élève en question. Il serait intéressant de développer la notion de profil en rapport aux notions de réussite et d'échec en explorant d'autres recherches. En effet, l'enjeu est important car il s'agit de déterminer l'échec ou la réussite de l'élève !

Finalement, un problème concerne la confusion qui peut naître entre une mesure des compétences et pédagogie par compétences. Les deux aspects sont-ils forcément liés ?

Bibliographie

Crahay, M. (2006). Dangers, incertitudes et incomplétudes de la logique de la compétence en éducation. *Revue française de pédagogie*, 154 janvier-février-mars, 97-110.

UNINE - Institut de psychologie et éducation

Raisonnements dans l'analyse de données en psychologie et en sciences de l'éducation, 07

De Ketele, J.-M & Gérard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et évaluation en éducation*, 28 (3), 1-26.