

# INFORMATISER L'ÉVALUATION : QUEL INTÉRÊT ET QUELLES LIMITES ?

Laura Weiss, Natacha Juda, Emiel Reith

Université de Genève, OFPC, DCEO

## LE CONTEXTE : L'ENTRÉE EN APPRENTISSAGE

Les procédures de recrutement à l'entrée en apprentissage en entreprise (formation duale, partiellement en entreprise et partiellement à école) sont déterminées par les employeurs et, dans certains métiers, par leurs associations professionnelles. Pour choisir leurs futurs apprentis, les patrons se basent sur les dossiers de candidature (CV, notes scolaires), sur les rapports de stages, sur des entretiens de motivation, ainsi que sur des tests d'entrée selon les métiers. Ces tests ont avant tout comme objectif de s'assurer que le jeune ait de bonnes chances de réussir la partie scolaire de sa formation. Comme les bulletins scolaires ne sont pas toujours explicites pour les patrons, certaines entreprises à Genève se sont d'abord tournées vers des tests privés payants, comme cela se fait dans d'autres cantons. Cette pratique pouvait cependant impliquer pour un jeune de déboursier plusieurs fois une centaine de francs pour chaque filière d'apprentissage potentiel, même si certaines entreprises acceptaient de rembourser le test aux jeunes qu'elles engageaient finalement. Pour contrer cette situation et offrir une alternative aux tests développés par des sociétés privées, le directeur général de l'Office pour l'orientation et la formation professionnelle et continue (OFPC) a mandaté un groupe de travail pour la mise sur pied de tests des connaissances scolaires en français et mathématiques pour l'entrée en apprentissage. Ces tests devaient être développés en collaboration avec le Cycle

d'orientation genevois (CO)<sup>1</sup> dans le cadre des travaux du Conseil interprofessionnel pour la formation (CIF). Les passations des tests seront gérées par l'OFPC, les tests seront gratuits pour les jeunes et assureront aux entreprises des résultats explicites et fiables, sur la base des connaissances fondamentales du CO. Toutefois, une autre condition devait être remplie : il n'était pas envisageable d'engager des correcteurs pour les quelques 1500 passations potentielles chaque année...

## LA CONCEPTION DES TESTS

Le groupe de pilotage chargé de concevoir ces tests baptisés EVA (Evaluations informatisées à l'entrée en apprentissage) n'avait donc qu'une solution : s'appuyer sur l'informatique pour la correction du test, ce qui impliquait de le concevoir en fonction de cette contrainte. A noter toutefois que les tests payants, mis en place par des sociétés privées, utilisent aussi cet outil, mais que leur fiabilité, les questions posées et les critères de correction restent opaques.

## LE CHAMP DU TEST DE MATHÉMATIQUES

La première étape était de concevoir un test de mathématiques<sup>2</sup> qui permette d'évaluer les connaissances et les compétences en mathématiques correspondant à celles attendues à la fin du CO, mais qui soit progressif, c'est-à-dire testant aussi les acquis des degrés scolaires précédents, de façon à ce que des candidats sortis du système scolaire depuis plusieurs années puissent aussi démontrer leurs compétences. Il ne pouvait donc pas être question de calquer entièrement EVA sur l'EVACOM (évaluation commune du CO) de mathématiques de 11<sup>e</sup> année qui survient à l'aboutissement de la scolarité obligatoire et qui a été précédée, pour les élèves du CO, d'autres EVA-COM en 9<sup>e</sup> et 10<sup>e</sup> années et d'épreuves cantonales à l'école primaire.

Les exercices d'EVA devaient donc recouvrir une plus large palette de connaissances et de compétences que l'EVACOM, tout en

<sup>1</sup> Élèves de 12-15 ans.

<sup>2</sup> Les contraintes étaient un peu différentes pour le test de français et le test d'anglais qui ne font pas l'objet de cet article.

restant compatibles avec le plan d'études romand (PER). Pour élaborer les questions, les concepteurs ont donc relu le PER de façon longitudinale en partant du 2<sup>e</sup> cycle<sup>3</sup> et aboutissant à la fin du 3<sup>e</sup> cycle<sup>4</sup> (y compris ses niveaux 2 et 3). Cette contrainte de balayer un champ plus vaste devait toutefois être mise en concurrence avec celle du temps de passation du test, qui ne pouvait pas dépasser une heure environ.

Cette contrainte temporelle a imposé des choix et occasionné des discussions dans le groupe de concepteurs. En effet, il y a eu de grandes hésitations sur l'indispensable, le nécessaire, l'utile, l'intéressant, le secondaire à tester pour évaluer les compétences mathématiques d'un jeune visant un apprentissage. Dans quelle mesure certains savoirs de mathématiques sont-ils nécessaires pour continuer une formation en dual ? Les jeunes concernés doivent-ils savoir énoncer le théorème de Pythagore et/ou savoir l'appliquer à une situation concrète, savoir additionner des fractions en toute généralité avec des dénominateurs ayant n'importe quel plus petit multiple commun (PPMC) ou seulement des fractions usuelles telles que des quarts et des tiers, savoir transformer des  $\text{mm}^3$  en  $\text{dam}^3$  ou seulement les unités les plus usitées, savoir résoudre des équations ou des problèmes à l'aide d'équations ? Le test doit-il porter aussi sur la maîtrise d'un vocabulaire précis ou plutôt sur des connaissances en acte ? Par exemple la compétence d'estimation a donné lieu à une discussion : en fait, on n'en trouve pas une mention explicite dans le PER, et pourtant elle semble indispensable pour des mathématiques « citoyennes », c'est-à-dire utiles au citoyen dans sa vie quotidienne.

Un autre point d'achoppement a été la question de la calculatrice. Au niveau des EVACOM du CO, différentes politiques ont été appliquées au cours des années en fonction des années scolaires mais aussi de la position des enseignants : calculatrice autorisée, permise seulement pour une partie de l'épreuve ou totalement interdite.

<sup>3</sup> Élèves de 8-12 ans.

<sup>4</sup> Élèves de 12-15 ans.

Finalement l'option choisie pour EVA a été celle d'un test en deux parties, la première sans calculatrice, puis la seconde avec autorisation de l'utilisation de l'outil.

## L'INFORMATISATION DES QUESTIONS

Une fois que la table de spécification du test (tableau listant les différents objectifs et sous-objectifs du test en regard desquels sont placés les différents exercices) a été établie, la seconde étape a consisté à informatiser les exercices. La plateforme Moodle choisie comme support pour la passation du test et largement utilisée par le Département de L'instruction Publique genevois (DIP), a l'avantage de permettre plusieurs formats de questions. Pour certains exercices, le choix s'est porté sans hésitation sur le format « réponse courte » qui valide toute réponse écrite correspondant à un élément de la liste des réponses acceptées établie préalablement. Ainsi, en étant attentif aux différentes possibilités de noter un résultat, par exemple lorsqu'il s'agit d'un calcul avec des nombres décimaux à une décimale en validant 4 ; 4,0 ; 4.0 ; +4 ; +4.0 et +4,0, (pratiquement) toutes les réponses correctes possibles sont couvertes. C'est évidemment plus simple quand le résultat attendu est une fraction irréductible à l'écriture univoque. L'autre format qui a été privilégié est naturellement celui des questions à choix multiple (QCM), c'est-à-dire où le répondant doit choisir parmi une liste de réponses possibles. On sait cependant que ce format induit d'autres démarches de résolution : par exemple, au lieu de construire la réponse, le candidat peut tester l'une après l'autre les réponses proposées. Enfin, d'autres formats ont aussi été utilisés : l'« appareillement » qui demande d'associer un élément d'une première liste avec un autre d'une seconde liste, par exemple pour interroger les testés sur les équations équivalentes et le format « glisser-déposer-sur-une-image » qui demande de déplacer des éléments (textes, nombres, figures) sur un schéma, un tableau ou une image, ce qui permet de présenter des tâches nécessitant plusieurs réponses ainsi que de faire appel à des compétences plus visuelles.

Toutefois, certains exercices ont dû être

transformés pour rentrer dans ce cadre assez contraignant. Il reste pour le moment difficile d'évaluer des démarches ou des réponses partielles. En effet, l'évaluation d'une réponse intermédiaire d'un problème en fait un exercice dit « à tiroirs », c'est-à-dire que la réponse ultérieure va dépendre de la précédente. Si la correction par une personne humaine permet de prendre en compte ces situations, le correcteur pouvant suivre le raisonnement du testé et valider une démarche correcte même si la réponse finale est fautive, le logiciel informatique utilisé pour la correction ne le peut pas (encore).

Il persiste d'ailleurs des problèmes à tiroirs dans certaines EVACOM et dans plusieurs épreuves d'enseignants, internes à un établissement, semestrielles et/ou de maturité. En effet, pour rendre le problème plus accessible, les concepteurs ajoutent une question intermédiaire, le transformant ainsi de fait à problème à tiroirs. Dans ces épreuves papier crayon, les concepteurs peuvent alors utiliser une solution un peu bancal pour éviter l'interdépendance des résultats : une réponse intermédiaire, qui n'est pas le résultat correct mais qui est plausible, est donnée en indiquant à l'élève : « si vous n'avez pas trouvé la solution à cette première partie du problème, prenez cette valeur pour résoudre la suite du problème ». Avec la plateforme Moodle, cela est impossible et le problème doit nécessairement être coupé en deux items indépendants. Le groupe de travail d'EVA s'est donc fixé actuellement le projet de trouver un moyen pour évaluer informatiquement des problèmes complexes.

### LES PRÉTESTS

Une fois les exercices conçus, choisis et informatisés, ils ont été largement prétestés pour s'assurer de leurs qualités métriques, dont leur difficulté et leur capacité à discriminer les candidats. Pour ce faire le DIP a accepté de mettre à disposition des heures de cours et des salles d'informatique pour que des classes d'élèves de 11<sup>e</sup> année du CO et des classes préparatoires de l'École de Culture Générale prétestent les exercices. Leurs réponses ont été ensuite analysées à

l'aide de tests statistiques, dont la méthode de la Théorie de la réponse aux items (TRI) utilisée par différents tests à grande échelle, par exemple les tests PISA et méthode de la Généralisabilité, pour s'assurer que les exercices étaient d'une difficulté convenable pour la population visée et qu'ils différenciaient adéquatement les élèves selon leur niveau de compétences : chaque exercice doit être mieux réussi par les élèves globalement les meilleurs et moins bien réussi par les élèves globalement les moins bons, sinon on considère qu'il dysfonctionne. Cette étape a permis de sélectionner les exercices à retenir dans une banque d'items de quelques 300 exercices préalablement sélectionnés comme pertinents du point de vue pédagogique.

Toutefois une deuxième série de prétests avec des élèves a été nécessaire pour mesurer les qualités intrinsèques du test dans son ensemble (et non plus les exercices indépendamment les uns des autres), telles sa consistance et celle de ses sous-domaines, (c'est-à-dire le fait que les exercices du même sous-domaine testent tous des compétences de même type) ainsi que sa difficulté. Cette étape a aussi permis de vérifier l'équivalence d'EVA mathématiques avec l'EVACOM en profitant du fait que le public du prétest avait passé l'EVACOM la même année. Il s'agissait d'assurer l'équité entre les deux évaluations des futurs apprentis, c'est-à-dire de s'assurer qu'EVA classe les élèves de la même façon que l'EVACOM, puisque cette épreuve est prise en compte pour l'entrée en apprentissage plein temps en école.

### UN TEST STABLE ET CONSISTANT

Après ces travaux préparatifs, EVA mathématiques a été mise en place conjointement à EVA français dès janvier 2010. Depuis novembre 2010, il existe trois versions équivalentes du test EVA mathématiques qui comprend 70 exercices portant sur les principaux domaines des mathématiques du PER : Nombres et opérations, Algèbre et fonctions, Grandeurs et mesures. Une attention particulière a été portée pour proposer des questions en lien avec les outils mathématiques utiles dans les cours pour les ap-

prentis, y compris les cours professionnels, comme la proportionnalité et les transformations d'unités.

La feuille des résultats transmise aux entreprises indique le nombre de points obtenus par la personne testée pour le test dans son ensemble, ainsi que pour chaque sous-domaine. Elle est complétée par la courbe de distribution des résultats de l'ensemble des candidats ayant passé le test l'année précédente. Aucune note, ni seuil de réussite ne sont donnés, sachant qu'EVA n'a qu'une valeur indicative et que la feuille de résultats doit être mise en lien avec les autres éléments du dossier de candidature.

Le nombre de jeunes passant EVA est allé croissant au cours des années, atteignant près de 1500 jeunes en 2016. Bien que le test donne entière satisfaction, il est ausculté de près chaque année : les résultats des candidats sont soumis à diverses analyses statistiques pour contrôler d'une part l'équivalence des trois versions du test et d'autre part la distribution des résultats du public en fonction de ses caractéristiques, comme la profession visée, le sexe et l'âge des candidats, le temps utilisé, etc. Ces analyses répétées année après année garantissent la grande stabilité du test.

A la fin du test, les testés sont interrogés sur la difficulté perçue de celui-ci, s'ils considèrent le temps à disposition suffisant ou non, et leur préférence entre un test sur l'ordinateur ou un test papier-crayon. Les réponses à cette dernière question montrent que les jeunes s'adaptent facilement à un test informatisé.

## CONCLUSION

EVA est un test des connaissances et compétences scolaires qui participe à la sélection par les entreprises des candidats à l'apprentissage dual. Pour des raisons de faisabilité pratique, EVA mathématiques est totalement informatisé : les élèves répondent aux questions directement sur un ordinateur et la correction est automatique. Un travail de contrôle pédagogique et docimologique montre sa compatibilité avec EVACOM qui est pris en compte pour l'entrée en formation plein temps.

L'inquiétude que soulève un test à correction informatisée, particulièrement chez les enseignants, semble pouvoir être calmée, car le test analysé en continu fait preuve de grandes qualités de stabilité et de cohérence. A la critique d'une évaluation des résultats et non des démarches, on peut opposer deux arguments : le premier est qu'il s'agit d'un test sommatif et non formatif, la prise en compte de démarches qui n'aboutissent pas perdant dans ce cas la plus grande partie de son intérêt, puisque la personne testée n'a pas l'occasion d'apprendre en partant de sa démarche incomplète. Le second est que dans les tests scolaires, les traces des démarches sont bien des résultats intermédiaires qui se trouvent ainsi crédités de points. EVA mathématiques le fait partiellement en posant séparément des questions simples et des questions plus difficiles qui demandent plusieurs étapes pour arriver au résultat, qui est seul évalué. Un travail reste cependant encore à faire pour affiner et mieux systématiser cette façon de faire.

## Référence

Test EVA : <http://www.citedesmetiers.ch/eva>