

**Le dépistage des étudiantes et des étudiants qui cherchent à obtenir un résultat faible au test de classement en anglais, langue seconde, au collégial**

**Gilles Raïche, Université de Moncton  
Jean-Guy Blais, Université de Montréal**

**Texte proposé à la revue Pédagogie collégiale**

**6 août 2002**

---

Le personnel enseignant des départements de langues des cégeps et collèges du réseau collégial québécois est à la recherche, depuis longtemps, de moyens simples et efficaces pour, sinon contrer, du moins amenuiser le comportement de sous-classement aux tests de classement en anglais langue seconde de la part des étudiants. Plusieurs enseignants ont cru que la solution au problème du sous-classement résidait dans l'élaboration d'un test de classement plus efficace dans l'appréciation du niveau d'habileté de l'étudiant. Malheureusement, comme le démontrent les résultats consécutifs à l'élaboration par Laurier, Froio, Pearo et Fournier (1998) d'un test de classement en anglais langue seconde amélioré, le TCALS II, les étudiants du réseau collégial continuent à tenter de se sous-classer à ce test. D'autres solutions doivent donc être envisagées. Nous sommes à leur quête depuis plusieurs années (Raïche, 2000, 2002a).

Nous présentons ici les étapes, ainsi que les résultats, d'un projet de recherche qui propose une solution au problème (Raïche, 2002b). Le projet a été soutenu financièrement par le programme d'aide à la recherche sur l'enseignement et l'apprentissage du ministère de l'Éducation du Québec et a été réalisé au Collège de l'Outaouais.

**Objectif spécifique du projet**

Tenant compte du grand nombre d'administrations de test de classement dans chacun des collèges, soit plus de mille par an à l'intérieur d'un collège de taille moyenne, il est

nécessaire d'utiliser une formule simple pour dépister les tentatives de sous-classement. Selon nous, l'analyse de la cohérence des réponses à chacune des questions permettrait de détecter des patrons de réponses plutôt étranges et ainsi peu probables. Certains auteurs se sont intéressés à cette problématique et ont proposé des indices numériques de scores plutôt suspects à un test (Cizek, 1999; Reise et Flannery, 1996). C'est l'une des piste des plus prometteuse à envisager pour pallier ce problème. Maintes mesures d'ajustement inadéquat ont été proposées. Ces mesures ne s'adressent toutefois pas au problème spécifique du sous-classement. Habituellement, elles visent plutôt à déterminer de façon globale le mauvais ajustement des réponses au test sans se soucier de la nature du problème : copie, réponse au hasard, mauvaise utilisation de la feuille de réponses, etc. Il est donc nécessaire de développer un indice, ou des indices, d'ajustement inadéquat spécifiques au dépistage du sous-classement. C'était l'objectif spécifique de ce projet de recherche.

Pour nous permettre de cet objectif, nous devions, au départ, identifier les stratégies utilisées par les étudiants pour tenter de se sous-classer à au test de classement en anglais, langue seconde, utilisé à l'intérieur du réseau collégial québécois. À cette fin, nous avons organisé une rencontre avec des étudiants du Collège de l'Outaouais. Par la suite, nous avons sélectionnés et développés des indices de détection de patrons de réponses suspects qui seront mis à l'essai à partir d'une simulation par ordinateur. Enfin, ces indices seront appliqués à l'identification des étudiants des deux premiers tours du SRAM inscrits au Collège de l'Outaouais en 2002-2003 dont le patron de réponses au test de classement en anglais, langue seconde, est plutôt suspect.

### **Les stratégies de sous-classement des élèves**

Seize étudiants inscrits à un cours d'anglais langue seconde de niveau 102 au Collège de l'Outaouais (604-EWG-03) du même groupe-classe ont participé à la rencontre. Il s'agissait d'un cours adapté à la famille des programmes rattachés aux sciences sociales et aux techniques humaines, regroupement, bien sûr, propre au Collège de l'Outaouais.

De ces 16 étudiants, six étaient de sexe masculin et dix, de sexe féminin. Six étaient inscrits dans un programme technique, soit Techniques administratives (410.12), tandis que les dix autres étaient inscrits dans un programme pré-universitaire, soit Sciences humaines (300.12 et 300.13). Enfin, quelques étudiants ont indiqué qu'ils n'ont pas participé à l'administration du TCALS en 2001-2002.

La rencontre, d'une heure, s'est déroulée à la fin de la période d'un cours de niveau 102 le jeudi 14 mars 2002, la semaine suivant la semaine de relâche au Collège de l'Outaouais. Nous avons pris le temps de bien expliquer le contexte et les objectifs de la recherche aux étudiants. Nous nous sommes assuré que les étudiants se souvenaient bien du type de questions dont est composé le TCALS et nous avons répondu aux questions des étudiants pour ensuite présenter le déroulement de la cueillette des informations. La cueillette consistait à obtenir par écrit, dans un premier temps, la réponse à la question suivante : *si vous désiriez vous sous-classer au test de classement en anglais, langue seconde, quelle stratégie, ou quelles stratégies, utiliseriez-vous?* On a bien insisté sur le caractère confidentiel des commentaires écrits. Les étudiants avaient au plus 10 minutes pour répondre par écrit à cette question. On a ensuite recueilli les commentaires des étudiants et, par la suite, on a continué la discussion avec les étudiants, principalement quant aux stratégies éventuelles de sous-classement volontaire

Considérant la nature exploratoire de cette étape de la recherche, l'analyse des résultats se limite au recueil et à l'interprétation des commentaires reçus de la part des étudiants. Par 14 fois, les étudiants ont indiqué qu'ils répondraient au hasard au test. Choisir la mauvaise réponse, avec diverses variantes, a été proposé par dix étudiants : c'est ce que nous nommons, plus loin, la stratégie de réponses inversées. Plusieurs ont souligné qu'ils utiliseraient cette dernière stratégie seulement s'il connaissait la bonne réponse. Malheureusement, puisque nous avons recueilli ces informations par écrit, nous n'avons pas pu découvrir si ces étudiants adopteraient une autre stratégie lorsqu'ils ne connaissent pas la bonne réponse. Six étudiants opteraient pour omettre de répondre à certaines questions, tandis que trois choisiraient de donner la bonne réponse aux questions faciles et la mauvaise réponse aux questions difficiles. Dans ce dernier cas, on peut se demander

comment les étudiants peuvent juger du niveau de difficulté des questions. Certains ont indiqué qu'ils donneraient le même choix de réponse à plusieurs questions du test.

Certaines stratégies proposées ne nécessitent pas l'application d'une procédure bien complexe pour dépister le comportement de sous-classement. C'est le cas notamment des stratégies où l'étudiant omet sciemment de répondre à des questions. Seules quelques recommandations aux collègues, faciles à appliquer, sont alors nécessaires. Toutefois, les stratégies qui ont été le plus fréquemment identifiées, telles que de répondre au hasard ou de choisir la mauvaise réponse, nécessitent la mise en œuvre de procédures de dépistage plus sophistiquées.

Les suites à cette étude ont d'ailleurs consisté à élaborer des procédures de dépistage de ces deux dernières stratégies. Ces procédures sont basées sur les modélisations issues de la théorie de la réponse à l'item (Hambleton, Swaminathan et Rogers, 1991). Des indices de détection de patrons de réponses aberrants non spécifiques à un comportement de sous-classement ont été identifiés à travers la littérature. Les indices retenus sont  $I_z$ ,  $W$  et  $Zeta$ . Les auteurs les ont généralement considérés plutôt efficaces à détecter des individus dont la qualité des réponses à des tests ou à des questionnaires est douteuse (Meijer et Sijtsma, 2001). Trois indices spécifiques au comportement de sous-classement ont aussi été élaborés selon l'approche d'identification optimale préconisée par Drasgow, Levine et Zickar (1996). Le premier de ces indices,  $I_{hasard}$ , permettrait de détecter un étudiant qui répond au hasard au test, tandis que le second,  $I_{inverse}$ , permettrait d'identifier un étudiant qui chercherait à donner des mauvaises réponses, soit des réponses inversées. Un troisième indice,  $I_{sous}$ , combine en un seul les propriétés des deux indices précédents. Ces indices, leur interprétation ainsi que leur formulation mathématique, sont décrits en détail par Raïche (2002b).

## Une simulation d'étudiantes et d'étudiants qui tentent de se sous-classer

Puisque à partir des données réelles d'une cohorte étudiante nous ne pouvons pas connaître pas à l'avance quels sont les étudiants qui ont cherché à se sous-classer, il est impossible de mettre à l'épreuve les indices de détection directement avec ces étudiants. Seulement une simulation des patrons de réponses normaux et aberrants nous permet de vérifier le pouvoir de détection des différents indices.

Par ordinateur, nous avons simulé 1000 patrons de réponses dans 85 conditions différentes, soit 85 000 patrons de réponses différents. Plus précisément, nous avons produit au hasard 1000 patrons de réponses à 17 niveaux d'habileté, variant, sur une échelle de score  $z$ , entre -2,00 et 2,00 par saut de 0,25, et selon cinq conditions : 1) des patrons de réponses totalement normaux; 2) des patrons de réponses où 10 % des premières questions reçoivent des réponses produites au hasard; 3) des patrons de réponses où 20 % des premières questions reçoivent des réponses produites au hasard; 4) des patrons de réponses où 10 % des premières questions reçoivent des réponses inversées; 5) des patrons de réponses où 20 % des premières questions reçoivent des réponses inversées.

Puisqu'il est nécessaire de détecter un étudiant qui cherche à se sous-classer en ne répondant au hasard ou en ne donnant des mauvaises réponses qu'à quelques questions, seulement 10 % et 20 % des réponses au test sont simulées de cette façon. Les autres questions sont simulées selon un patron de réponse normal. De plus, nous croyons que l'étudiant, en cherchant à contrôler le nombre de mauvaises réponses, n'applique la stratégie de réponses au hasard ou la stratégie de réponses inversées qu'au début du test. Ce sont ainsi les premières questions du test qui sont produites au hasard.

Dans les simulations, de tous les indices de détection évalués, quelle que soit la stratégie de sous-classement adoptée et peu importe le pourcentage de réponses aberrantes, c'est l'indice  $l_z$  qui s'est montré le plus efficace à identifier les cas de sous-classement. Pour être en mesure de bien évaluer l'efficacité de l'indice  $l_z$ , on peut observer les résultats de

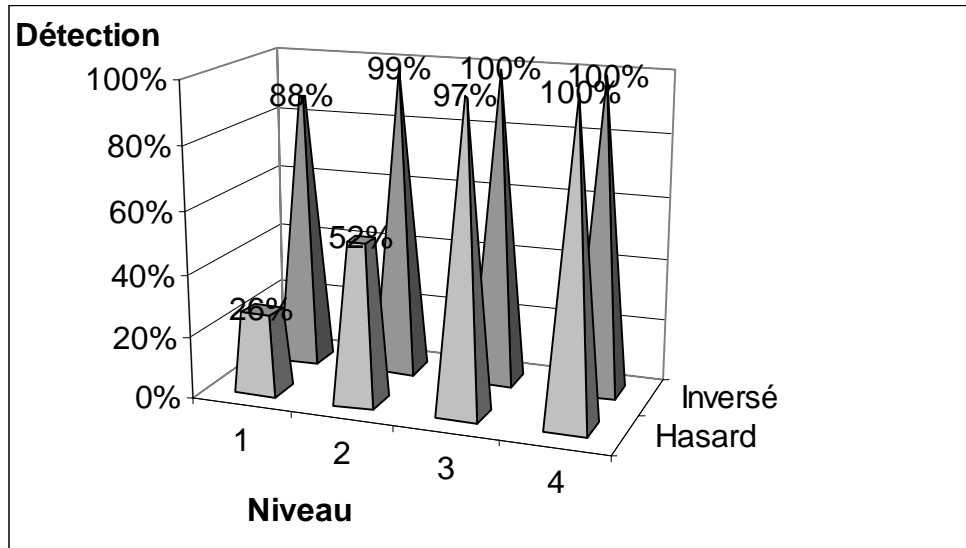
détection aux figures 1 et 2. La figure 1 nous renseigne sur l'efficacité de l'indice  $I_z$  à identifier les cas de sous-classement lorsque seulement 10 % des réponses sont, soit au hasard, soit inversées.

Quand aussi peu que 10 % des réponses sont données au hasard, le taux de détection des cas réels de sous-classement est égal ou supérieur à 97 % pour les étudiants des niveaux 3 et 4 et ce quelle que soit la stratégie adoptée. L'indice  $I_z$  se montre donc très efficace en ce qui a trait aux étudiants dont le niveau d'habileté en anglais, langue seconde, est élevé. C'est justement ces étudiants que nous espérons pouvoir identifier par cette procédure. Le résultat est supérieur à ce que nous avions espéré.

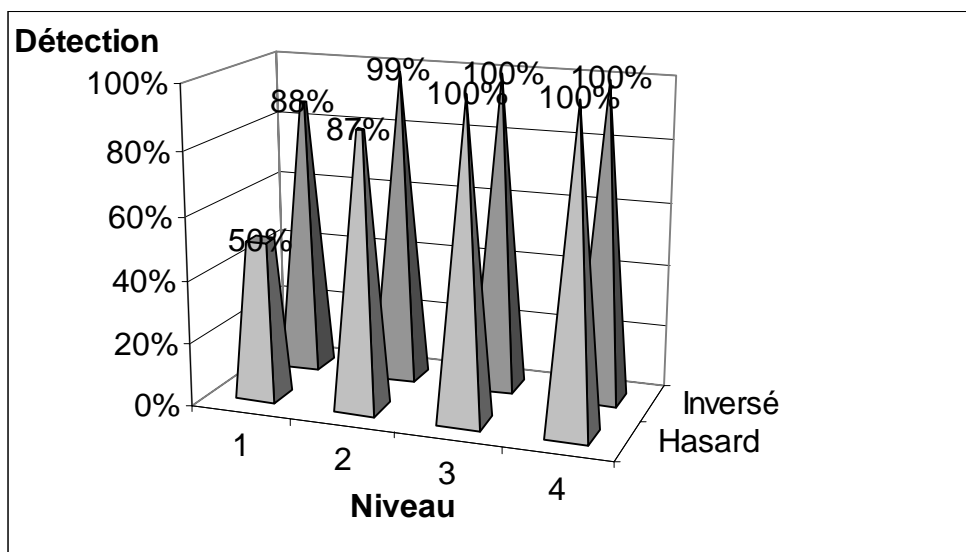
Lorsque la stratégie de réponses au hasard est utilisée, au niveau 2, on peut tout de même détecter un peu plus de la moitié des cas (52 %), tandis que seulement 26 % des cas peuvent être dépistés au niveau 1. Cependant, lorsque la stratégie de réponses inversées est considérée, les taux de détection augmentent considérablement, soit respectivement 88 % et 99 % : à ce moment, avec aussi peu que 10 % de réponses aberrantes, l'efficacité de la procédure d'identification est encore très importante.

Lorsque 20 % des réponses sont données au hasard ou inversées, ces pourcentages augmentent considérablement. C'est ce qu'illustre la figure 2. Même au niveau 1, au moins la moitié des cas de sous-classement sont détectés, tandis qu'aux niveaux 2, 3 et 4 le taux de détection est toujours supérieur ou égal à 87 % quelle que soit la stratégie adoptée.

L'indice  $I_z$  constitue donc un indicateur très efficace pour nous permettre de détecter l'utilisation d'une stratégie de réponses au hasard ou inversée. De plus, son efficacité est à son maximum chez les étudiants qui, selon nous, risquent le plus souvent de désirer se sous-classer, soit ceux dont le niveau d'habileté en anglais est élevé.



**Figure 1.** Taux de détection des patrons de réponses aberrants lorsque 10 % des questions sont affectées en fonction du niveau de classement réel et de la stratégie adoptée



**Figure 2.** Taux de détection des patrons de réponses aberrants lorsque 20 % des questions sont affectées en fonction du niveau de classement réel et de la stratégie adoptée

## **Une application aux 1<sup>e</sup> et 2<sup>e</sup> tours du SRAM de la cohorte 2002-2003 au Collège de l'Outaouais**

Forts de ces résultats, nous avons appliqué l'indice  $l_z$  à la détection des cas potentiels de sous-classement chez les étudiants de la cohorte 2002-2003 admis et inscrits au Collège de l'Outaouais au premier et au second tours du Service d'admission du Montréal métropolitain (SRAM).

En fait, avant même d'avoir eu le temps d'informatiser les indices de détection traités plus haut, nous avons prévu, conjointement avec le personnel enseignant du département des langues secondes du Collège de l'Outaouais, d'utiliser un indice destiné à l'évaluation de l'ajustement des patrons de réponses intégré au logiciel Bilog pour détecter les étudiants qui auraient éventuellement tenté de se sous-classer au TCALS II. Ceux-ci seraient ensuite rappelés pour rédiger une composition écrite. 108 étudiants ont été détectés par Bilog et ont été rappelés. Les résultats de cette démarche, toutefois, ne nous sont pas encore totalement disponibles. Ils le seront uniquement suite à la rentrée scolaire de l'automne 2002. Cependant, pour le moment, nous avons jugé pertinent de vérifier si les étudiants détectés à partir de Bilog sont les mêmes que nous pouvons maintenant détecter à partir de l'indice  $l_z$ . À cette fin, nous avons appliqué l'indice  $l_z$  aux résultats des étudiants admis et inscrits au 1<sup>e</sup> tour du SRAM au TCALS II et avons noté quels étaient les étudiants qui avaient aussi été détectés par Bilog. Nous avons remarqué que plusieurs des étudiants détectés par l'indice  $l_z$  étaient les mêmes que ceux identifiés par Bilog, mais toutefois, pas tous. Des 69 étudiants identifiés par  $l_z$  au 1<sup>er</sup> tour du SRAM, seulement 42, soit 60,90 %, ont aussi été identifiés par Bilog.

La date d'obtention des résultats des étudiants admis au 2<sup>e</sup> tour du SRAM ne nous a pas permis de les inclure à l'exercice de rappel pour rédiger une composition écrite et nous n'avons donc pas appliqué l'indice intégré à Bilog à ceux-ci. Toutefois, nous avons tout de même appliqué l'indice  $l_z$  aux résultats des ces étudiantes au TCALS II. Une fois regroupés, les étudiants des premier et second tours du SRAM sont au nombre 1361. De ces 1361, l'indice  $l_z$  a permis d'identifier 141 étudiants dont le patron de réponses est

douteux, soit 10,36 % de la cohorte. Cela nous donne une idée du pourcentage d'étudiants que cette démarche nous permettra de détecter à l'avenir.

## **Recommandations**

Quelques recommandations s'imposent. Premièrement, il est clair que nous jugons approprié d'utiliser l'indice  $I_z$  pour détecter les cas de sous-classement intentionnels de la part des étudiants. Il ne faut toutefois pas oublier que celui-ci ne détecte pas spécifiquement les cas de sous-classement. Il ne faut donc pas prendre pour acquis que l'identification d'un étudiant par cet indice signifie nécessairement que celle-ci ou celui-ci a tenté de sous-classer. Le patron de réponses peut provenir d'un étudiant qui n'était tout simplement pas disposé à répondre au test ce jour-là. Il se peut que l'administration du test se soit déroulée dans des conditions peu usuelles. Cependant, à chaque fois que l'indice  $I_z$  présente une valeur suspecte, il est totalement justifié de mettre en doute le résultat au test et d'investiguer plus à fond ce qui s'est passé. Des procédures appropriées, qui selon nos observations au Collège de l'Outaouais ne devraient toucher au plus que 10 % des nouvelles admissions, adaptées aux réalités de l'institution d'enseignement, doivent être planifiées.

En second lieu, Il faut prévoir un processus de détection des omissions et des répétitions consécutives du même choix de réponse aux questions du TCALS II. Un étudiant qui omet de répondre ou répète consécutivement un même choix de réponse à plus de 10 % des questions du test, devrait retenir l'attention des intervenants.

En dernier lieu, comme nous l'annonçons plus haut, il est nécessaire d'appliquer un mécanisme de suivi à l'identification des comportements suspects au test. On pourrait, par exemple, rappeler les étudiantes et les étudiants détectés et leur administrer une composition écrite ou une entrevue orale. D'ailleurs, le seul fait de rappeler des étudiants pour une seconde procédure de classement envoie un message assez clair à ceux-ci : il est possible de détecter les tentatives de sous-classement intentionnel. Selon les échos que

nous avons reçus en Outaouais, l'application de cette procédure a fait jaser beaucoup dans la région. À elle seule, cette pratique pourrait éventuellement être suffisamment dissuasive pour diminuer la fréquence des tentatives de sous-classement.

### **Les suites au projet**

Pour assurer la réalisation de nos travaux, nous avons dû élaborer un outil informatique spécialisé. Toutefois, le programme actuel est actuellement un peu lourd à utiliser et nécessitera quelques améliorations pour permettre son utilisation dans la plupart des collèges. Nous espérons, d'ici peu, développer une version adaptée du logiciel pour les intervenants des services informatiques du réseau collégial. Entre temps, dans certains cas, nous pourrions accepter de traiter les données du TCALS II en provenance des collèges.

Nous désirons aussi continuer les travaux suites à l'administration au Collège de l'Outaouais d'une composition écrite et vérifier quels ont été les cas de sous-classement détectés par l'administration de cette composition écrite. Il nous sera alors possible d'identifier formellement des étudiants sous-performants, mais surtout d'étudier les patrons de réponses de celles-ci et de ceux-ci. Nous croyons pouvoir ainsi percer plus à fond les stratégies utilisées par eux.

Conjugué à nos recherches antérieures sur le testing adaptatif par ordinateur, ce projet de recherche prépare aussi le terrain à des travaux d'élaboration d'une procédure adaptative qui permettrait d'administrer un test dont le niveau de difficulté des items varierait non seulement avec l'estimateur du niveau d'habileté de l'étudiant, mais aussi avec le niveau d'aberrance du patron de réponses. Actuellement, à notre connaissance, aucun travail en ce sens n'a été entrepris. Nous espérons, alors, appliquer ces résultats au développement d'un environnement informatique destiné à soutenir une version adaptative du TCALS II. Cette version adaptative pourrait alors intégrer des mécanismes de dépistage des tentatives de sous classement au test.

## Références

- Cizek, G.J. (1999), *Cheating on test: how to do it, detect it, and prevent it*. Mahwah : Laurence Erlbaum Associates.
- Drasgow, F., Levine, M.V. et Zickar, M.J. (1996). Optimal identification of mismeasured individuals. *Applied Measurement in Education*, 9(1), 47-64.
- Fournier, P. (1992). *Pour un tests incontestable : rapport de recherche sur les tests de classement en anglais (langue seconde) au collégial*. Québec : ministère de l'Éducation, Direction générale de l'enseignement collégial.
- Hambleton, R.K., Swaminathan, H. et Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park : Sage.
- Laurier, M., Froio, L., Paero, C. et Fournier, M. (1998). *Test de classement d'anglais langue seconde au collégial*. Montréal : Collège de Maisonneuve.
- Meijer, R.R. et Sijtsma, K. (2001). Methodology review : Evaluating person fit. *Applied psychological Measurement*, 25(2), 107-135.
- Raïche, G. (2000). *Le sous-classement des élèves aux tests de classement en anglais langue seconde au Collège de l'Outaouais*. Gatineau : Collège de l'Outaouais.
- Raïche, G. (2002a). Pour dépister les étudiants qui cherchent à sous-performer au test de classement en anglais langue seconde dans le réseau collégial québécois. *Bulletin de l'Association pour la recherche au collégial*, 15(3), 8-9.
- Raïche, G. (2002b). *Rater un examen, c'est payant! Le dépistage du sous-classement au test de classement en anglais, langue seconde, au collégial*. Gatineau : Collège de l'Outaouais.

Reise, S.P. et Flannerey, W.P. (1996). Assessing person-fit on measures of typical performance. *Applied Measurement in Education*, 9(1), 9-26.