



uOttawa

L'Université canadienne
Canada's university

**FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES**



**FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES**

Serge Boulé

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A. (Éducation)

GRADE / DEGREE

Faculté d'éducation

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Utilisation du degré de certitude dans un contexte d'évaluation diagnostique critériée

TITRE DE LA THÈSE / TITLE OF THESIS

Dany Laveault

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Brad Cousins

Marielle Simon

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

SERGE BOULÉ

Utilisation du degré de certitude dans un contexte d'évaluation diagnostique critériée

Thèse

présentée comme exigence partielle à la Faculté des

études supérieures et postdoctorales de

l'Université d'Ottawa

pour l'obtention du grade de

Maîtrise ès arts en éducation (M.A.[Éd.]

FACULTÉ DES SCIENCES DE L'ÉDUCATION

UNIVERSITÉ D'OTTAWA

OTTAWA

© Serge Boulé, Ottawa, Canada, 2007



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-49171-3
Our file *Notre référence*
ISBN: 978-0-494-49171-3

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■ ■ ■
Canada

Résumé

À l'origine, la recherche sur le degré de certitude (DC) et son utilisation se sont développées dans le cadre d'évaluations certificatives normées. Dans un contexte d'évaluation pour l'apprentissage, des indicateurs et des profils de réalisme sont développés pour vérifier à quel point un étudiant peut tirer profit d'un tel feedback. Dans un contexte diagnostique le degré de certitude sert à identifier les items pour lesquels l'étudiant n'a pas répondu au hasard et a surestimé ses habiletés sans le savoir. Ces « événements critiques » nécessitent une remédiation. Cette recherche étudie le degré de réalisme en tenant compte des différences individuelles, des caractéristiques métriques et des demandes cognitives provenant des items.

Les questions de recherche se formulent ainsi : Comment et dans quelle mesure l'expression du degré de certitude en la réponse choisie, et par conséquent le réalisme des étudiants, varient-ils en fonction du niveau taxonomique des items et de leurs propriétés métriques? Comment et dans quelle mesure le sexe et le niveau de performance des étudiants sont-ils liés à l'expression du degré de certitude en la réponse choisie, et par conséquent au réalisme de l'étudiant?

L'échantillon visé se compose de 152 participants âgés de 18 ans et plus. Le test est de type questionnaire à choix multiples et comprend 40 items avec 4 choix de réponses. En plus des réponses aux items, l'étudiant doit exprimer son DC par rapport à ses réponses. Notre analyse permet de conclure que plus les étudiants sont performants, plus ils sont réalistes par rapport à la qualité de leurs réponses. Il ne semble pas y avoir de différences marquées entre les hommes et les femmes. Nous avons néanmoins remarqué des inégalités

dans la linéarité des distributions. Plus l'item est facile, plus les étudiants sont réalistes. Il existe une relation probable entre le coefficient de discrimination d'un item et le réalisme. Le lien entre les niveaux taxonomiques des items et le réalisme des étudiants n'est pas clair et mérite d'être étudié ultérieurement en utilisant un devis construit spécifiquement à cette fin.

Mots clés : Degré de certitude, régulation des apprentissages, test critérié, test diagnostique, métacognition, connaissance partielle.

Table des matières

Liste des tableaux et figures	vi
PROBLÉMATIQUE	1
REVUE DE LA LITTÉRATURE	6
Effet du sexe	18
Effet du niveau d'habileté	21
Effet des propriétés des items	22
LE CADRE CONCEPTUEL	24
Questions de recherche	33
LA MÉTHODE	35
Les participants	35
Les instruments	37
Le plan d'observation	40
La procédure.....	41
Le plan d'analyse des résultats	42
Le calcul du réalisme	43
Le calcul de la difficulté et de la discrimination des items	44
Les variables indépendantes et dépendantes	46

	v
RÉSULTATS	49
La cohérence interne du test	50
Les propriétés métriques des items	51
Distribution des résultats et statistiques inférentielles	52
Effet des différences individuelles	57
Analyse de régression linéaire	58
Analyse de la variance en plan factoriel	60
Effet des propriétés des items	66
DISCUSSION	81
LIMITES ET RECOMMANDATIONS	92
RÉFÉRENCES.....	95

Liste des tableaux et figures

Figure 1 - Exemple de l'utilisation du DC lors d'un test	9
Tableau 2 - Modèles de l'activité mentale d'une personne répondant à une QCM	10
Figure 3 - Relation entre la qualité de la réponse et le DC exprimé	13
Tableau 4 - Lien entre la réponse à l'item et le degré de certitude exprimé	14
Figure 5 - Algorithme portant sur les conditions de l'évaluation diagnostique	27
Tableau 6 - Statistiques descriptives sur l'échantillon	37
Tableau 7 - Tableau de spécification après les révisions	39
Tableau 8 - Plan d'observation	46
Tableau 9 - Sommaire des variables	48
Tableau 10 - Cohérence interne du test	51
Tableau 11 - Propriétés métriques des items.....	52
Tableau 12 - Statistiques sur les distributions	54
Figure 13 - Distribution des scores totaux au test	55
Figure 14 - Distribution de l'indice de réalisation des prédictions de l'item (transformée)	56
Figure 15 - Distribution de l'indice de réalisme des sujets	57
Tableau 16 - Résumé du modèle proposé par l'analyse de la régression pour les sujets.....	59

Tableau 17 - Tableau résumé de l'analyse de la variance	61
Figure 18 - Réalisme moyen des sujets selon le niveau de performance et le sexe	63
Figure 19 - Représentation graphique portant sur la relation entre score et réalisme	64
Figure 20 - Relation entre le score et le réalisme en fonction du sexe.....	65
Tableau 21 - Matrice des données sur les items	68
Tableau 22 - Résumé du modèle proposé par l'analyse de la régression pour les items	69
Figure 23 - Représentation graphique sur le lien avec le niveau de difficulté	70
Figure 24 - Représentation graphique sur le lien avec le coefficient de discrimination.....	71
Figure 25 - Représentation graphique sur le lien avec le niveau taxonomique des items ...	74
Tableau 26 - Matrice des pondérations factorielles	75
Tableau 27 - Racines (<i>Eigenvalues</i>) cumulatives	76
Tableau 28 - Résumé de l'analyse de la régression en fonction des regroupements proposés	77
Figure 29 - Lien entre degré de difficulté des items et regroupement factoriel	78
Tableau 30 - Propriétés métriques des items selon le regroupement factoriel.....	79
Annexe 1 - Instrument	I à XVIII

PROBLÉMATIQUE

Cette étude s'inscrit dans le cadre d'une recherche appliquée qui porte sur l'utilisation du degré de certitude (DC) en docimologie. Lors de la passation de tests, nous pouvons demander aux individus d'exprimer leur certitude par rapport à la qualité de leurs réponses. Il s'agit en quelque sorte d'ajouter une modalité de réponse supplémentaire à chaque item permettant ainsi à l'individu d'indiquer son degré de certitude pour chacune de ses réponses.

L'intérêt porté envers la certitude n'a rien de nouveau. La philosophie, l'épistémologie et l'ontologie ont vu leur part d'écrits sur le sujet. Descartes et Pascal, pour ne mentionner que ceux-là, ont contribué considérablement à ce questionnement. Des mots comme doute, certitude et incertitude ont pris tout leur sens dans les écrits de ces auteurs. Descartes (1637, Ed. 1983, p. 146) écrivait : « ... ce doit avoir été en confessant plus ingénument ce que j'ignorais que n'ont coutume de faire ceux qui ont peu étudié, et peut-être aussi en faisant voir les raisons que j'avais de douter de beaucoup de choses que les autres estiment certaines,... ». Au 20^e siècle, Bertrand Russell reprend : « Le problème du monde, c'est que les imbéciles sont présomptueux et les gens intelligents bourrés de doutes. »

Henmon (1911) introduit le concept de certitude des étudiants en leur réponse. Le degré de certitude d'un étudiant dans les réponses qu'il donne est un élément qui peut être ajouté au processus d'évaluation et qui tient compte des degrés variables de connaissance. Selon la Commission des communautés européennes (1991), l'utilisation du degré de certitude ouvre des perspectives de recherche très importantes

sur la capacité des étudiants à tirer parti des informations et à apprendre. Leclercq et Poumay (2005) définissent les degrés de certitude comme « Jugements, analyses et/ou régulations observables ou non observables effectués par l'apprenant sur ses propres performances (processus ou produits d'apprentissages), dans des situations de PRE, PER ou POST performance. »

Il existe plusieurs façons d'exprimer le degré de certitude. Shuford, Albert et Massengill (1966) proposent des procédures de mesure qu'ils appellent *Admissible Probability* (par intervalles de 20 répartis entre 0 et 100). Dans le cas d'une échelle utilisée par Reach, Zerrouki, Leclercq et D'Ivernois (2004), l'étudiant indique, en plus de sa réponse à l'item du test, son degré de certitude en utilisant une échelle de probabilités d'exactitude qui pourrait varier par échelons entre *pas sûr du tout* et *absolument sûr*. En quelque sorte, l'étudiant estime correctement ou non la qualité de ses réponses, ce qui est adapté à l'évaluation des apprentissages. L'évaluateur peut alors pondérer les réponses en tenant compte du degré de certitude exprimé pour chacune d'elles.

Il y a des erreurs pires que d'autres; surtout celles pour lesquelles nous n'avons aucun doute de nous tromper. Il serait utile de savoir si un étudiant doute qu'il se trompe lorsqu'il donne une mauvaise réponse à un item durant son test d'entrée dans sa profession. Surtout dans le cas où la mauvaise réponse pourrait avoir des conséquences en situation réelle. Hunt (1993) classe les états de la connaissance sous ces catégories : connaissance, ignorance et méconnaissance. « Ce n'est pas ce que nous ignorons qui nous nuit, c'est ce que nous savons et qui est faux. » (Twain, 1835-1910)

Le degré de certitude permet d'accroître la différenciation entre étudiants lorsque le score composite résulte de la combinaison entre la performance de l'étudiant et son degré de certitude exprimé. L'étudiant pourrait être pénalisé si son degré de certitude exprimé n'était pas en relation avec la qualité de sa réponse. Par exemple, deux étudiants pourraient avoir le même nombre de réponses correctes mais l'un d'eux est plus constant que l'autre dans l'expression de son degré de certitude. Il sera certain d'avoir une bonne réponse lorsqu'elle est en effet bonne. Inversement, il est incertain lorsqu'il rate des items. De deux étudiants ayant le même nombre de réponses exactes, celui qui exprime des degrés de certitude qui sont davantage en relation avec la qualité de ses réponses devrait obtenir le score le plus élevé. Le calcul de scores impliquant le degré de certitude est effectué à partir d'algorithmes sophistiqués. Ces derniers devraient pouvoir maximiser la variance en augmentant celle des scores vrais et devraient indirectement favoriser une augmentation de la fidélité des résultats au test. De plus, le système de pondération tend à diminuer l'intérêt à deviner les réponses, ce qui s'attaque à cette autre problématique de taille qui a fait l'objet de beaucoup de recherches sur les QCM.

À l'origine, la recherche sur le DC et son utilisation se sont développées dans le cadre d'évaluations certificatives normées. L'intérêt pour le DC était principalement lié à la correction pour divination dans le cas de questions à choix multiples (QCM) afin d'améliorer la fidélité des résultats. Récemment, la recherche sur le DC porte plutôt sur la dimension formative qui peut en découler. En comparant la certitude exprimée et la qualité des réponses, nous pouvons établir si l'individu testé est réaliste lorsqu'il juge la

qualité de ses réponses. La mesure du réalisme proposée s'obtient donc en vérifiant la concordance entre le DC exprimé et la qualité de la réponse. L'étudiant réaliste tend à anticiper une bonne réponse lorsque celle-ci est bonne et une réponse incorrecte lorsque celle-ci est incorrecte. L'utilisation du DC peut ainsi avoir une fonction formative à partir du moment où nous donnons un feedback à l'individu qui se concentrerait sur les éléments qui ont incité ce dernier à manquer de réalisme.

Cette étude se veut une première tentative d'utiliser le degré de certitude pour mieux différencier les étudiants et leurs habiletés. L'originalité de cette recherche tient à ce qu'elle étudie le degré de réalisme en tenant compte des différences individuelles et des caractéristiques métriques et cognitives des items, principalement le niveau taxonomique des questions. Le but est de parvenir à une évaluation diagnostique critériée individuelle qui peut orienter et aiguiller de manière plus valide et plus efficace l'étudiant vers les formations appropriées.

La certitude et par conséquent le réalisme sont-ils des traits ou des caractéristiques dépendantes du contexte? En mesure et en évaluation, il est important de savoir si le réalisme de l'étudiant est indépendant du test et quels sont les facteurs qui l'influencent. Nous devons comprendre l'effet de ces facteurs pour que l'utilisation du degré de certitude soit absente de biais et valide dans un contexte d'évaluation diagnostique. Il s'agit de déterminer les conditions d'utilisation du degré de certitude dans un tel contexte.

Pour répondre à ces questions, nous abordons d'abord les écrits antérieurs de manière à connaître l'état de la connaissance sur le sujet. Certains concepts

fondamentaux sont ensuite présentés et sont suivis des questions de recherche. Par après, nous discutons de la méthode utilisée pour recueillir les données qui servent ensuite à l'analyse des résultats. Finalement, nous discutons des résultats, des limites et des recommandations qui serviront à aiguiller les recherches ultérieures dans ce domaine.

REVUE DE LA LITTÉRATURE

Nous faisons un petit retour historique pour bien situer les préoccupations mentionnées antérieurement. Cette revue de la littérature évalue les orientations et la qualité de la connaissance accumulée jusqu'à aujourd'hui. L'orientation de cette recherche est déterminée par l'état de la connaissance actuelle sur l'utilisation des DC. Nous abordons d'abord certains concepts fondamentaux en docimologie. Nous introduisons ensuite la connaissance accumulée portant sur les thèmes de cette étude qui sont en l'occurrence l'effet du sexe, l'effet du niveau d'habileté et l'effet des propriétés des items.

Spearman (1907) jette les fondements de la théorie classique des scores (TCS). Dès lors, les tests consistent en plusieurs items et leur correction s'effectue de façon dichotomique. Les réponses de l'étudiant testé sont bonnes ou mauvaises. Cependant, Echternacht (1972) rapporte que la connaissance n'est pas un phénomène dichotomique ni polytomique comme les tests à choix multiples semblent le suggérer, mais un phénomène continu dans le sens où il y a des degrés variables de connaissance. En évaluation, une des problématiques que nous relevons concerne la question des crédits partiels. C'est le cas lorsque nous sommes dans une situation où nous avons une connaissance partielle d'un fait ou d'une solution. Il y a des situations de connaissance complète ou discrète comme dans le cas où seule une réponse exacte serait possible. Ce sont des questions de connaissance selon la taxonomie des objectifs cognitifs de Bloom. Il y a d'autres situations où nous pouvons déduire ou donner une approximation comme dans le cas où il faut estimer la distance entre Vancouver et Halifax. Il s'agit alors de

connaissance partielle où la réponse se situe sur une échelle continue. Plus l'estimation est précise, plus elle devrait être appréciée en proportion. Dans le cas des questionnaires à choix multiples (QCM), il y a des leurres qui sont plus erronés que d'autres et leur pondération devrait peut-être refléter cette réalité. Prenons l'exemple d'un item qui porterait sur la capitale de la province de l'Ontario. Les choix de réponses pourraient comprendre *Toronto*, *Ottawa*, *Montréal* et *Paris*. L'étudiant qui répondrait *Paris* serait plus loin de la bonne réponse que celui qui répondrait *Ottawa* qui est une erreur commune et attendue. De plus, le cas serait nettement aggravé s'il choisissait *Paris* en indiquant un degré de certitude élevé.

Pendant longtemps, les recherches sur la TCS se sont cristallisées autour de la fidélité des résultats aux tests et sur leur validité. L'intérêt est principalement lié à l'augmentation de la variance des scores vrais et, par conséquent, à l'accroissement de la capacité à discriminer les étudiants de niveaux d'habiletés différents. Dans cette perspective, il devient important de composer avec les problèmes causés par les réponses au hasard des étudiants testés, qui nuisent à la capacité de différencier les étudiants. On parle alors de correction pour divination.

La recension d'Echternacht (1972), malgré qu'elle remonte aux années 70, dresse un bilan des recherches dans ce domaine et de plusieurs indicateurs potentiellement pertinents. Cette recension remonte jusqu'à Henmon en 1911 et Hollingworth en 1913. Durant les années 30, les chercheurs s'intéressent à réduire l'erreur pour divination et ils cherchent ainsi à augmenter la fidélité des tests de type vrai ou faux. Les années 30 et 40 sont caractérisées par l'étude de la *témérité* chez les

répondants (*gambling*). Durant les trois décennies suivantes, des méthodes de pondération et des formats de présentation sont développés, et ceux-ci incluent les QCM. La correction pour divination et la fidélité sont aussi examinées. De Finetti (1965) élabore *l'approche des probabilités personnelles (Personal Probability Approaches)* en introduisant un haut degré de sophistication mathématique : « Seule une probabilité subjective peut donner une signification objective à toute réponse et toute méthode de notation et de mesure de gains d'apprentissage. » Avec cette approche, le répondant au test doit assigner sa probabilité personnelle pour chaque leurre. Cette approche a pris plusieurs formes durant cette période. Lord et Novick (1968), malgré qu'ils considèrent que les travaux de De Finetti sont attrayants sur le plan conceptuel, font état de plusieurs postulats qui sont pratiquement impossibles à satisfaire. Lord et Novick considèrent que le comportement des étudiants peut varier d'un test à l'autre selon le score qu'ils anticipent. Un étudiant qui se considère de performance moyenne pourrait ainsi prendre de grands risques de manière à maximiser ses chances de se distinguer de la norme. Ils reconnaissent cependant qu'il y a peu de recherches expérimentales qui portent sur le sujet. La figure 1 présente un exemple d'utilisation du DC dans la pratique.

Évaluation diagnostique mathématiques

Tu vas remarquer que, pour chaque question du test, il va falloir que tu dises si tu es certain de ta réponse. Ça ne change rien à ton résultat mais ça peut nous aider à mieux voir où tu es rendu. Il est important que tu répondes à toutes les questions et que tu indiques si tu es certain pour chacune des questions.

**6. Dispose les nombres ci-dessous en ordre croissant:
0,7 - 2/3 - 69%**

- a) 0,7 - 2/3 - 69%
- b) 2/3 - 0,7 - 69%
- c) 2/3 - 69% - 0,7
- d) 0,7 - 69% - 2/3

Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
- 10% - pas sûr
- 25% - un peu sûr
- 50% - ne sais pas
- 75% - plutôt sûr
- 90% - sûr
- 98% - complètement sûr

Figure 1. Exemple d'utilisation du DC lors d'un test.

Les recherches sur le degré de certitude ont culminé durant la période 1960-1970 pour ensuite ralentir. Durant cette période, les docimologues se sont concentrés sur la question de la correction pour divination. Choppin (1975) propose des Modèles de l'activité mentale d'une personne répondant à une QCM. Le tableau 2 traduit la pensée de Chopin.

Tableau 2

Modèles de l'activité mentale d'une personne répondant à une QCM

	Modèle 1	Modèle 2	Modèle 3
Conception d'une réponse à une QCM (Si l'étudiant sait)	L'étudiant connaît la réponse et il la choisit.	L'étudiant connaît la réponse et il la choisit.	L'étudiant ordonne les solutions en fonction de leur plausibilité.
Conception d'une réponse au hasard à une QCM (Si l'étudiant ne sait pas)	L'étudiant ne connaît pas la réponse et il choisit au hasard parmi k choix.	L'étudiant ne connaît pas la réponse et il procède par élimination avant de choisir parmi les k réponses qui n'ont pas été éliminées.	L'étudiant choisit la réponse la plus plausible.
Modalité de correction des probabilités de réussite pour divination	Le docimologue effectue une correction pour divination classique : $-1 / (k - 1)$.	Le docimologue effectue une correction pour divination : $-1 / (k - é - 1)$.	Les degrés de certitude permettent de vérifier le degré de certitude qu'a l'étudiant envers son choix.

Depuis, des systèmes de pondération sont développés pour résoudre le problème de la divination. Par exemple, en TCS on parle d'indice de difficulté et de correction pour l'effet du hasard chaque fois que l'on peut considérer que les leurres ont des chances à peu près égales d'être choisis. Nous notons également que les modèles de

réponses à l'item (MRI) utilisent un paramètre de pseudo-chance pour compenser dans le cas où l'asymptote la plus basse de la courbe caractéristique de l'item serait nettement supérieure à zéro, ce qui a l'avantage d'être estimé indépendamment des caractéristiques des individus.

Les recherches sur le degré de certitude ont été reprises par Leclercq (1993) qui expose des conditions méthodologiques fondamentales pour utiliser le degré de certitude de façon valide. Il utilise les degrés de certitude dans des contextes d'évaluations certificatives normées. Il tient également compte du degré de certitude exprimé dans la pondération du score total. Les conditions de Leclercq s'appliquent dans la mesure où les degrés de certitude exprimés sont compris dans la pondération. Les conditions sont :

1. Les directives doivent offrir une échelle métrique. Leclercq rapporte que des recherches ont montré que les échelles ordinales manquaient de stabilité.
2. La pondération doit être calculée en fonction des principes d'une théorie des décisions. En évaluation sommative, la pondération du degré de certitude n'a rien de simple. Leclercq a montré que, dans certains cas, des échelles de pondération sont inappropriées. Il en résulte que l'étudiant peut trouver une stratégie qui consiste à choisir un degré de certitude biaisé de manière à maximiser son gain de points.
3. Il doit y avoir une distinction entre mesure et récompense. Le score total n'est pas une mesure mais la combinaison de deux mesures, soit une mesure de performance avec le nombre de réponses correctes, soit une mesure de réalisme

selon la congruence de l'autoestimation de la qualité des réponses. Le concept de réalisme et ses indices sont discutés ultérieurement.

4. Un feedback explicite doit être fourni à propos du réalisme. Leclercq suggère de montrer des représentations graphiques du réalisme à l'étudiant. Il est clair que les étudiants doivent avoir un niveau de sophistication élevé pour bien comprendre ce type de feedback.
5. Les étudiants doivent être formés. Ils n'utilisent pas le degré de certitude de façon optimale la première fois qu'ils sont confrontés à ce genre de test.
6. La pondération doit s'adapter aux situations. Certaines approches seraient meilleures que d'autres.

Les recherches de Leclercq sont intéressantes puisqu'elles incorporent une dimension formative. Cette dernière prend la forme d'un feedback qui consiste en un profil de réalisme. Leclercq (1993) et Gilles (2002) postulent implicitement que l'indice de réalisme obtenu donne un feedback à l'étudiant et lui permet d'améliorer l'autoestimation de ses compétences. En d'autres mots, transmettre à l'étudiant de l'information sur son réalisme pourrait être formateur. Pour y arriver, Leclercq compare les réponses aux items et le degré de certitude exprimé de manière à créer un profil individuel de réalisme qui peut servir de feedback. Celui-ci est constitué essentiellement de représentations graphiques et de cotes liés au réalisme et d'une interprétation qui s'échelonne entre *excellent* et *minimal*. Cette interprétation est basée sur des valeurs repères qui sont essentiellement des normes locales établies en milieu universitaire. Ces valeurs repères ont un potentiel limité de généralisation puisque l'échantillon utilisé

pour les établir n'est pas représentatif de la population générale. La figure 3 présente conceptuellement la relation entre la qualité de la réponse et le DC exprimé.

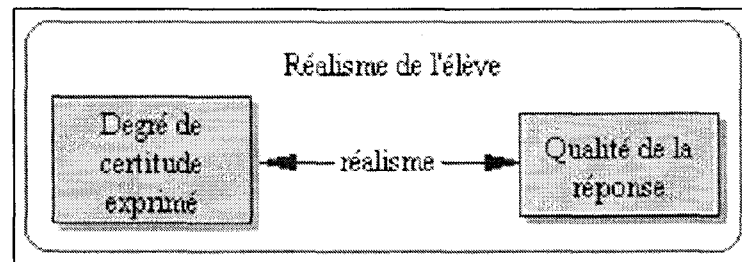


Figure 3. Relation entre la qualité de la réponse et le DC exprimé.

L'étudiant peut être réaliste ou non dans son jugement par rapport à sa performance. Sur le plan conceptuel, la perception correcte pour un étudiant de la qualité de ses réponses fait partie de ce que De Landsheere (1979) appelle une *image de soi réaliste, complète et stable*. Dans un contexte de testing, la mesure du réalisme est obtenue en vérifiant la concordance entre le degré de certitude exprimé en la réponse choisie et la qualité de cette réponse.

Nous pouvons distinguer chez Leclercq (1993) des indices individuels et collectifs qui permettent d'apprécier différents aspects de la manifestation du degré de certitude. Les indices individuels sont : le réalisme, la cohérence personnelle, l'acuité et la centration. L'indice de réalisme est obtenu en vérifiant la concordance entre l'exactitude des réponses et la valeur de la certitude obtenue. Prenons l'exemple suivant qui est présenté au tableau 4 :

Tableau 4

Lien entre la réponse à l'item et le degré de certitude exprimé

Lien entre la réponse à l'item et le degré de certitude exprimé		Qualité de la réponse à l'item (0/1)	
		Bonne (1)	Mauvaise (0)
Qualité de la réponse anticipée (Certitude)	Élevée	A Réaliste	B Surestime (requiert feedback et remédiation)
	Faible	C Sous-estime	D Réaliste

L'étudiant réaliste a tendance à anticiper une bonne réponse lorsque la réponse est en effet bonne (A). Il anticipe également une mauvaise réponse lorsque celle-ci est mauvaise (D). L'étudiant peu réaliste qui sous-estime la qualité de ses réponses (C) a tendance à avoir une faible certitude par rapport à ses bonnes réponses. L'autre type d'étudiant peu réaliste est celui qui surestime généralement la qualité de ses réponses (B). Il exprime une grande certitude pour des réponses incorrectes. Cette catégorie d'étudiants est particulièrement préoccupante. Ces étudiants méritent une attention particulière et une intervention qui favorise la régulation des apprentissages.

Leclercq (1993) signale qu'en évaluation sommative les résultats peuvent être plus valides si l'étudiant est réaliste. Henmon (1911) remarquait déjà qu'il semblait y avoir des différences individuelles par rapport à l'expression du DC. Selon Henmon, ces

différences individuelles étaient liées considérablement au DC exprimé sans pour autant montrer de tendances générales. Les résultats rapportés par Henmon sont supportés par Jonsson et Allwood (2003) qui étudient la surestimation de la qualité des réponses lors de tests d'intelligence cristallisée et d'intelligence fluide. Leurs résultats suggèrent que les participants surestiment moins la qualité de leurs réponses lors du test d'intelligence cristallisée que lors du test d'intelligence fluide. Les résultats ne rapportent pas de différences systématiques de réalisme en fonction des variables individuelles contrôlées. Ces résultats semblent compatibles avec les résultats obtenus antérieurement par Henmon.

Comme Henmon (1911), Jonsson et Allwood (2003) concluent que plusieurs variables semblent affecter le réalisme et qu'il est difficile de comprendre leur fonctionnement. L'identification des participants sur l'échelle de style cognitif (Need-for-Cognition) se fait à partir d'une version abrégée mais calibrée du NfC-test. Les auteurs anticipent que le score à ce test, qui offre une mesure de la perception qu'a un individu de son degré de capacité intellectuelle et de recherche de défis de nature cognitive, serait corrélé au réalisme métacognitif. Il est à noter que Jonsson et Allwood n'ont pas distingué les étudiants universitaires selon leur domaine de spécialisation. Ils ne tiennent pas compte de l'effet de sélection qui ferait en sorte que les étudiants inscrits dans un champ de spécialisation non traditionnel tendent à mieux réussir. Si le groupe testé était composé de plusieurs individus de ce type, les résultats pourraient être biaisés. Il est néanmoins vraisemblable que le réalisme n'est pas égal d'un étudiant à l'autre ou d'un groupe d'étudiants à l'autre.

Une recherche de Clack et Head (1999) étudie entre autres la capacité pour un étudiant en médecine d'évaluer ses limites et sa capacité de réviser son travail. Il s'agit de deux attributs personnels qui sont pertinents à nos recherches. Il se trouve que la recherche de Clack et Head ne rapporte aucune différence en fonction du sexe pour ces deux attributs personnels en particulier. Leur devis présente cependant plusieurs limites. Les auteurs ne montrent pas le questionnaire ainsi que les questions utilisées. Ceci a pour conséquence d'enlever au lecteur la chance d'apprécier davantage la qualité de l'instrument utilisé. De plus, la méthode du sondage postal est une façon de recueillir des observations et de sonder le terrain afin d'obtenir de l'information et des opinions. Puisque les étudiants ne sont pas dans un environnement contrôlé qui assure la constance de plusieurs variables environnementales et émotionnelles, les résultats ont un potentiel de généralisation limité et doivent être interprétés avec réserves. Leur devis est ainsi limité et l'importance de leurs conclusions en est affaiblie.

Néanmoins, au même titre que nous cherchons à bien comprendre l'effet du sexe sur la validité d'une évaluation diagnostique, les auteurs questionnent le lien entre le sexe et plusieurs qualités désirables que nous cherchons à retrouver chez les médecins. L'expression de la certitude est intimement liée à la prise de décision et aux règles de la conduite de l'esprit dans le sens où Descartes l'entendait. Il y a des erreurs pires que d'autres; surtout celles pour lesquelles nous n'avons aucun doute de nous tromper. Il serait utile de savoir si un étudiant en médecine doute qu'il se trompe lorsqu'il donne une mauvaise réponse à un item durant son test d'entrée dans sa profession. Surtout dans le cas où la mauvaise réponse pourrait avoir des conséquences fatales en situation réelle.

La prise de décision étant une activité importante et parfois lourde de conséquences en médecine, il est important de comprendre l'impact des différences individuelles sur celle-ci.

S'il était démontré que le réalisme pouvait varier en fonction des groupes d'individus, l'utilisation des méthodes de pondération du score fondées sur le réalisme des individus pourrait avoir un impact différencié sur le score total et, par conséquent, sur l'évaluation d'étudiants ou de groupes d'étudiants de réalisme différent. La fidélité et la validité d'un test seraient considérablement réduites puisque nous serions en présence d'une modalité de présentation du stimulus pouvant causer un biais dans l'interprétation des résultats. Il y aurait ainsi un risque accru de fonctionnement différentiel au niveau du test. Laveault et Grégoire (2002) font état des conséquences d'un tel problème. Ce biais dans l'interprétation peut se produire au niveau des résultats à un test. Il s'agit du potentiel d'impact différencié de certaines caractéristiques des individus sur les résultats à un test. Un test pourrait favoriser ou défavoriser des individus en fonction de leur groupe d'appartenance. Par exemple, prenons un groupe d'hommes et un groupe de femmes d'habiletés moyennes équivalentes en mathématiques. Si un groupe est constitué de membres qui sont en moyenne plus réalistes que ceux de l'autre groupe, ils seront avantagés par la méthode de pondération et obtiendront un score plus élevé. Les méthodes de pondération pénalisent les étudiants lorsque le DC exprimé n'est pas en relation avec la qualité des réponses. Tel qu'expliqué par Laveault et Grégoire, il est possible de vérifier la présence de ce biais

causé par le sexe ou la culture à l'aide de méthodes empiriques, de méthodes graphiques et de méthodes faisant appel aux jugements d'experts.

Or, les méthodes de pondération du score composite résultant de l'utilisation du DC tendent à favoriser les étudiants réalistes. Ceci suggère que l'utilisation du degré de certitude peut avoir un impact différencié sur le score total et, par conséquent, sur l'évaluation d'étudiants ou de groupes d'étudiants de réalisme différent. Il est également concevable que les caractéristiques des items du test puissent avoir une influence sur les degrés de certitude exprimés. La recension des écrits traite également de ces questions.

Effet du sexe

Il est probable que le comportement du répondant en rapport avec le DC soit lié à des caractéristiques comme le sexe. Lors de l'analyse de résultats à des tests, Swineford (1938) a conclu que les garçons ont tendance à prendre plus de risques que les filles. Sieber (1974) et Stankov (1998) soutiennent que les garçons surestiment légèrement plus la probabilité d'exactitude de leurs choix. Le résultat rapporté par Beyer (2002) soutient que lors du test en mathématiques, les hommes surestiment davantage la qualité de leurs réponses. Ce phénomène serait particulièrement présent chez les étudiants ayant des attentes faibles et pourrait interagir avec le niveau de performance de l'étudiant. Les résultats ne seraient pas du même ordre pour des matières comme l'anglais, l'histoire et la géographie. Les résultats rapportés par Pallier (2003) mettent également en lumière un impact différencié du sexe sur la propension qu'ont les individus à surestimer la qualité de leurs réponses. Les hommes auraient plus tendance à surestimer la qualité de leurs réponses, et ce, sans égard à la nature de la tâche et sans égard à l'âge. En

conformité avec ce qui avait été exprimé par Leclercq (1993), Pallier conclut que de la formation préalable pourrait contribuer à l'amélioration chez les individus de la précision du DC exprimé.

Le devis utilisé par Pallier (2003) consistait en des tests d'aptitudes informatisés administrés à des étudiants universitaires. Certaines autres limites sont observées par rapport à l'échantillonnage de l'étude portant sur l'âge. L'identification des 303 participants d'âges variés semble discutable. L'auteur rapporte avoir fait appel aux membres de la famille des étudiants et à des amis. L'échantillon devrait normalement être constitué d'individus choisis aléatoirement dans le bassin de population cible. L'auteur a aussi utilisé des tests informatisés. Il ne discute cependant pas des biais potentiels liés à l'effet de la modalité de présentation du stimulus. Ce biais peut considérablement brouiller les résultats des hommes et des femmes au point d'en arriver à des conclusions hâtives.

Une analyse de l'Office de la qualité et de la responsabilité en éducation (OQRE) rapporte qu'il y a un lien appréciable entre le rendement, la perception que les étudiants ont de leur niveau de réussite et la valorisation de la réussite dans ces matières (OQRE, 1999). D'après l'OQRE, en général les filles se disent davantage bonnes en lecture et écriture. Les garçons s'estiment bons en mathématiques. Bressoux et Pansu (2003) font état que les garçons ont une perception d'eux-mêmes systématiquement supérieure à celle des filles. Pallier (2003) met les résultats qu'il a obtenus en relation avec des notions d'identité. Il constate que les femmes, qui ont une faible perception d'elles-mêmes envers certaines tâches traditionnellement masculines, ont des résultats plus

faibles que ceux des hommes. Il est concevable qu'à mesure que les différences d'opportunités d'accès à l'éducation entre les hommes et les femmes s'estompent, ces écarts de perception puissent se rapprocher proportionnellement. Comme Auger, Dalley et Roy (2005) en font état, le stéréotype est le figement d'une représentation ou d'une pratique qui peut évoluer.

Koivula et Hassmén (2001) tentent de vérifier les liens potentiels entre le sexe des individus et leur propension à agir de façon typique ou stéréotypée. Près de 550 individus sont catégorisés selon leur sexe, la tendance de leur comportement à être stéréotypé et un index de désirabilité sociale. Les résultats rapportés mettent en lumière un impact différencié du sexe dans certaines conditions sur la surestimation de la qualité de leurs réponses. Les femmes ayant des comportements généralement associés aux hommes ont plutôt tendance à surestimer la qualité de leurs réponses. Selon les résultats de cette recherche, ce n'est pas le sexe comme l'attitude stéréotypée qui a un impact.

Même si ces recherches n'étudient pas le réalisme directement, cette tendance à prendre des risques et à surestimer leur performance peut potentiellement résulter en une baisse du score de réalisme chez une catégorie d'individus (sexe masculin ou féminin). À cause de l'impact potentiel du sexe sur le degré de réalisme des étudiants et du risque de fonctionnement différentiel au niveau du test, nous croyons important de contrôler le sexe et d'examiner son effet dans toute étude sur le DC de manière à éviter de confondre certains effets.

Si les différences individuelles ont un impact sur le score de réalisme, la prise en compte de ces variables lors de l'interprétation des résultats au test est essentielle pour

l'orienteur, l'enseignant et l'étudiant. L'interprétation des résultats au test a plus de chances d'avoir une validité prédictive acceptable si les intervenants concernés peuvent faire la part des choses advenant que les différences individuelles aient un impact sur le réalisme des individus.

Effet du niveau d'habileté

Le niveau de performance d'un étudiant à un test nous informe sur son niveau d'habileté par rapport à ce qui est évalué. Il peut s'agir par exemple d'un test portant spécifiquement sur une matière enseignée. Dans cette recherche, le niveau d'habileté ne fait pas référence à un niveau général d'aptitude académique ou à une définition de quotient intellectuel.

Fabre (1980) a conclu que la certitude est liée aux connaissances des étudiants testés. Lorsque le niveau de performance est bas, l'étudiant peut apprécier sa performance incorrectement sur la foi d'indices non pertinents. Il semble raisonnable de croire que la certitude exprimée par l'étudiant est liée à son niveau d'habileté et peut varier selon les situations et dans le temps. Selon les données de Stankov (1998) et de Kleitman et Stankov (2001), plus l'étudiant a de bonnes réponses, plus il a confiance en ses choix.

Nous ne connaissons cependant pas la nature du lien entre le niveau d'habileté et le réalisme des étudiants. Les étudiants moins performants dans une matière pourraient très bien exceller dans une autre. Ils pourraient être en mesure d'avouer leur ignorance et d'être réalistes quant à l'estimation de la qualité de leurs réponses dans un domaine, mais pas dans un autre.

Effet des propriétés des items

Nelson et Narens (1994) font la distinction entre plusieurs instances de jugements métacognitifs : jugement de la facilité d'apprendre « *ease-of-learning judgement* », jugement de l'apprentissage « *judgement of learning* », jugement de penser savoir « *feeling-of-knowing judgement* », confiance dans les réponses récupérées « *confidence in retrieved answers* ». Jans et Leclercq (1997) suggèrent qu'une taxonomie du réalisme métacognitif clarifierait le sujet et faciliterait les recherches appliquées. Pintrich (2002) décrit la connaissance de soi (« *Self-knowledge* ») comme l'un des trois types de connaissance métacognitive (« *Metacognitive knowledge* »). L'habileté à évaluer sa propre connaissance est un aspect important de la connaissance de soi. Nous pouvons étendre cette définition en présumant que le réalisme de l'étudiant est une forme de connaissance métacognitive. Le réalisme de l'étudiant cadrerait ainsi dans la taxonomie révisée d'Anderson et al. (2001) sous la rubrique « *Connaissance métacognitive* » des dimensions de la connaissance (« *Knowledge dimension* »). Cette connaissance métacognitive inclut la régulation du comportement de l'étudiant en situation d'examen.

Il est concevable que, sur le plan cognitif, le degré de certitude exprimé puisse être relié au niveau taxonomique des items. Nous soupçonnons également que l'indice de difficulté et le coefficient de discrimination de l'item ont un rôle à jouer. Jacobs (1974) a conclu que l'expression du degré de certitude pouvait varier en fonction de la difficulté de l'item. De Finetti (1965) a également souligné que les items difficiles entraînent un degré d'incertitude. Enfin, un item défectueux qui discriminerait peu

entraînerait de l'incertitude et l'interprétation des résultats doit prendre en considération de telles exceptions.

L'ensemble des recherches précédentes ne porte pas spécifiquement sur les variations de degré de certitude causées par les niveaux taxonomiques des items. Cependant, il semble y avoir suffisamment d'évidences pour justifier l'intérêt à prendre en compte cette variable. Nous pourrions nous attendre à ce que le degré de réalisme ne présente pas le même intérêt au niveau de l'évaluation pour des items de haut niveau taxonomique. Par exemple, il n'y a pas beaucoup d'ambiguïté dans les réponses à un item du niveau taxonomique *connaissance*. L'étudiant aura tendance à être totalement certain ou incertain. Cependant, pour des items du niveau de l'analyse ou de la synthèse, l'énoncé et les réponses ne sont généralement pas nets et suscitent l'incertitude. L'étudiant risque d'être moins absolu dans l'expression de son degré de certitude. L'information supplémentaire apportée par la prise en compte du degré de certitude à des items de niveau taxonomique élevé peut avoir une utilité non pas seulement pour différencier des étudiants, mais aussi différents niveaux d'habiletés. Ceci cadre bien dans un contexte d'évaluation critériée.

La section suivante présente le cadre conceptuel de cette recherche. Nous reprenons ainsi l'ensemble de la connaissance accumulée pour l'appliquer au contexte spécifique à cette recherche. Nous cherchons aussi à nous doter d'un schème d'interprétation des résultats.

LE CADRE CONCEPTUEL

Cette recherche reprend les idées de Leclercq. Elle s'attarde cependant au rôle du degré de certitude dans la différenciation des capacités des individus. Elle revêt un caractère original car elle cible l'utilisation du degré de certitude dans un contexte d'évaluation diagnostique. Scallon (2000) fait état d'une certaine confusion qui entoure la définition de l'évaluation diagnostique et sur ce qui la démarque de l'évaluation formative. Même si l'appellation « diagnostique » n'est pas universelle, Scallon considère que ce type d'évaluation peut avoir une fonction autre que pédagogique comme dans le cas du reclassement des étudiants. Ainsi, dans le cadre de cette recherche, l'objectif n'est pas de calculer un score total qui sert à différencier des étudiants en fonction de leur degré de certitude. Le score aux items et le degré de certitude exprimé sont cependant utilisés conjointement dans la création d'un feedback individuel portant sur le réalisme qui sert d'appui à l'aiguillage de l'étudiant et qui lui est remise à des fins diagnostiques.

La question est de savoir si le réalisme a une utilité diagnostique et peut contribuer à une évaluation diagnostique plus valide, fidèle et utile. Pour être utile, la mesure de réalisme doit être pertinente peu importe l'étudiant testé et le niveau taxonomique des items. Le but est de vérifier la pertinence des conditions d'utilisation dans un contexte diagnostique. Si nous examinons les conditions énoncées par Leclercq, nous constatons que les conditions 2, 3 et 6 liées à la pondération ne sont pas pertinentes dans un contexte diagnostique puisque la pondération du degré de certitude ne fait pas

partie du score composite de l'étudiant. Néanmoins, les conditions 1, 4 et 5 demeurent pertinentes.

1. Les instructions doivent offrir une échelle métrique.
4. Un feedback explicite doit être fourni à propos du réalisme.
5. Les étudiants doivent être formés.

Nous proposons un feedback enrichi tenant compte de la nature des erreurs. Pour aider l'étudiant, il semble important de mettre en évidence les items échoués que l'étudiant croyait avoir réussis. Selon Tourneur, Duquesne et Ledoux (1986), le feedback serait plus efficace lorsque son contenu n'est pas en conformité avec les attentes de l'étudiant. (Voir tableau 4, catégorie B)

Un tel feedback qui tient compte du réalisme de l'étudiant peut faire partie du processus de régulation des apprentissages. Selon Laveault (2005), le feedback est l'une des trois caractéristiques essentielles qui définissent la régulation : des buts ou des attentes, un feedback et une action ou une forme de remédiation visant un ajustement. Pour l'étudiant, une remédiation possible serait de corriger sa perception de ses habiletés et de changer ses habitudes de travail. Selon Laveault, l'autorégulation prend en considération les croyances sur soi, telles que les doutes et les craintes associés à différents contextes de performance. Selon nous, le feedback doit favoriser une prise de conscience de l'étudiant par rapport à ses croyances sur soi de manière à développer son autonomie. Laveault (2004) indique que l'étudiant autonome est celui qui « porte lui-même son jugement » (du grec *auto*, « soi-même » et *nomos*, « juge »). Cette position cadre également dans le contexte d'évaluation pour l'apprentissage « *assessment for*

learning » tel que proposé par Stiggins et Chappuis (2005). Stiggins et Chappuis soutiennent que l'étudiant, comme les enseignants, prend des décisions fondées sur des données.

Tel qu'utilisé antérieurement, le profil de réalisme est constitué d'indices et est remis à l'étudiant qui est en mesure d'apprécier et de comprendre l'utilité de ceux-ci. Ce profil de réalisme représente un feedback et favorise la remédiation des stratégies d'apprentissage par l'étudiant. Dans le cas d'une évaluation sommative, des indicateurs collectifs s'ajoutent aux indicateurs individuels. Les indices collectifs se calculent à partir des formules qui sont essentiellement des variations des formules utilisées pour les indices individuels. Les travaux récents de Gilles (2002) partent de la prémisse que le degré de certitude exprimé devrait être plus élevé pour les réponses correctes. Gilles parle alors de *cohérence spectrale*. Dans le même ordre d'idées, il y aurait *incohérence spectrale* lorsque le degré de certitude ne serait pas en relation avec l'exactitude des réponses. Contrairement à l'évaluation sommative qui utilise des indicateurs collectifs et individuels, dans un contexte d'évaluation diagnostique, on favorise les indicateurs individuels qui sont communiqués à l'étudiant. Pour leur part, les indicateurs collectifs demeurent nécessaires en ce sens qu'ils permettent d'établir l'amplitude respective des indicateurs individuels et permettent de vérifier si les difficultés sont propres à l'étudiant, au groupe ou aux items (voir figure 5). Nous considérons cependant que les indices individuels peuvent être utiles à des étudiants adultes. Pour des étudiants plus jeunes, nous soupçonnons que le surcroît d'information provenant de ces indices pourrait entraîner de la confusion. Baumeister, Heatherton et Tice (1994) parlent de

Algorithme permettant d'établir les conditions nécessaires pour une évaluation diagnostique au niveau de chaque item

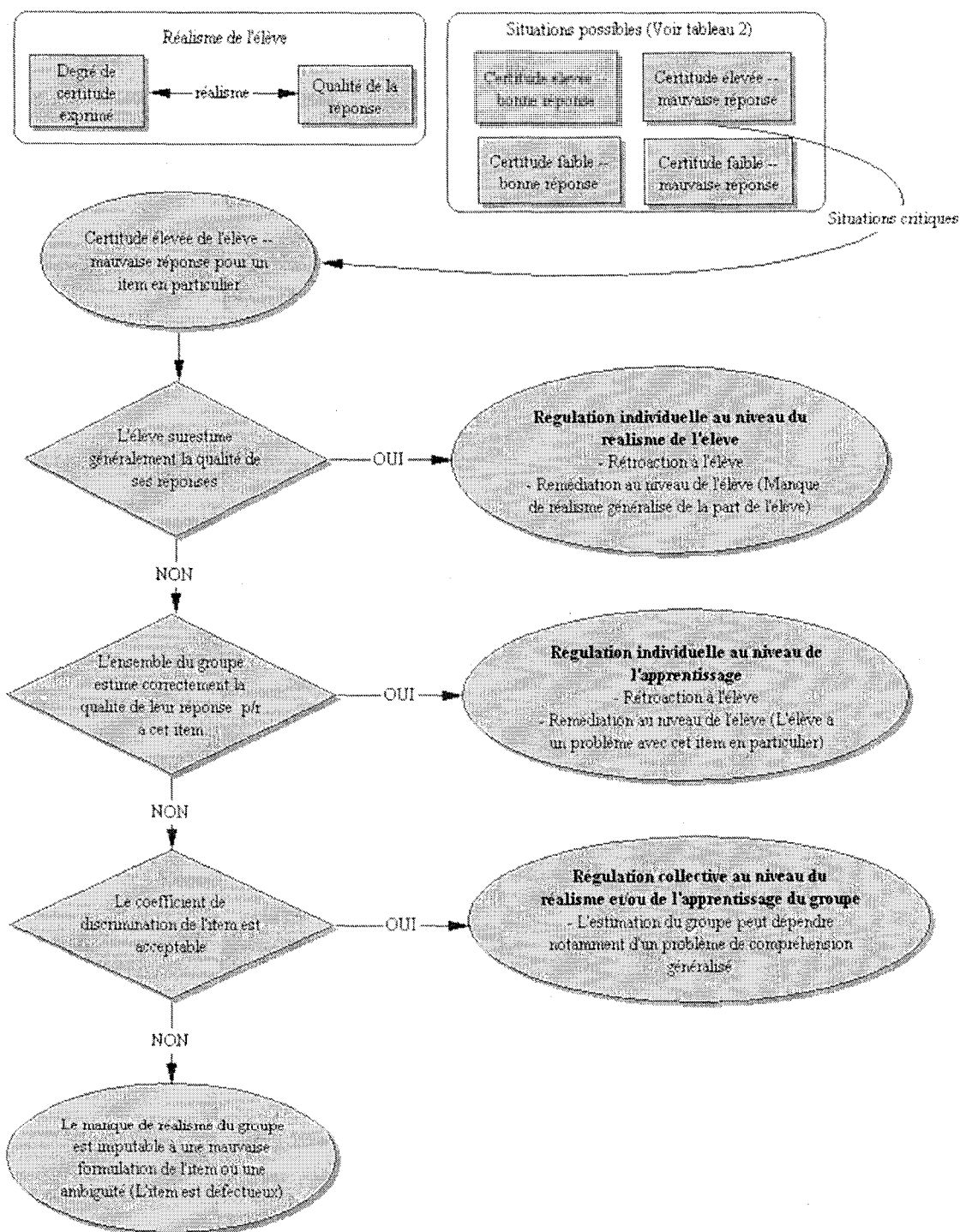


Figure 5. Algorithme portant sur les conditions de l'évaluation diagnostique.

régulation excessive (*overregulation*) qui consiste en l'exercice d'un contrôle excessif en vue d'atteindre un but. À ce sujet, Laveault (2005) fait état de plusieurs éléments de la problématique liée aux aspects dysfonctionnels de la régulation.

C'est pour cette raison que nous considérons une variation de la quatrième condition de Leclercq qui stipule qu'un feedback explicite doit être fourni à propos du réalisme. Dans un contexte diagnostique, l'évaluateur considère le profil de score et le profil de réalisme. Les résultats sont interprétés sur une base individuelle. Il peut consulter les erreurs qui étaient accompagnées d'un degré de certitude élevé et les réponses incertaines. Le jugement s'inspire de l'ensemble de ces résultats. L'évaluateur se charge de communiquer les résultats concernant le réalisme à l'étudiant. L'évaluateur donne alors le feedback conformément à la situation d'approche en donnant un feedback enrichi comprenant des pistes de rectification, des indices et une démystification de la nature des résultats.

La recherche sur le degré de certitude et son utilisation s'est concentrée sur des évaluations certificatives. Les enjeux étaient clairs pour les étudiants et la motivation d'exprimer avec rigueur leur degré de certitude envers leurs réponses était assurée par des barèmes communiqués à l'avance. Ces derniers faisaient état de la pondération et des récompenses associées à l'usage des degrés de certitude.

Cette recherche cible l'utilisation du degré de certitude dans un contexte diagnostique et critérié. Dans un contexte diagnostique, les étudiants sont motivés de façon différente à répondre aux items et à indiquer leur degré de certitude. Même si les enjeux ne sont pas les mêmes qu'en évaluation certificative avec la pondération des

degrés de certitude, les étudiants qui participent à une évaluation diagnostique devraient coopérer puisqu'ils n'ont pas intérêt à risquer une erreur d'aiguillage. De plus, puisque les scores ne sont pas pondérés par le degré de certitude exprimé, l'étudiant n'a pas besoin de chercher à développer une stratégie optimale qui favorise le gain de points.

Cette recherche se distingue également par son caractère méthodologique. Contrairement à l'usage traditionnel qui consiste en des évaluations normées, nous choisissons d'utiliser un instrument de mesure critérié en fonction d'un programme d'étude. Il ne s'agit pas d'un test formulé selon la théorie des facettes ou une autre méthode de spécification de domaine. Il s'agit plutôt d'une mesure fondée sur des objectifs d'apprentissage. Nous ne connaissons pas l'utilité du degré de certitude et la pertinence d'un feedback sur le réalisme dans un contexte d'évaluation par objectifs d'apprentissage.

Dans le contexte d'une évaluation diagnostique critériée, le degré de certitude ne sert donc pas à pondérer les résultats de l'étudiant en fonction de son degré de réalisme. Celui-ci sert plutôt à identifier les items pour lesquels il s'est surestimé, c'est-à-dire pour lesquels il croyait que sa réponse était juste avec un degré de certitude élevé, alors qu'elle ne l'était pas. La prise en compte du réalisme dans ce nouveau contexte permet d'identifier les domaines pour lesquels une remédiation devrait être entreprise de façon prioritaire. En effet, non seulement les réponses de l'étudiant sont incorrectes, mais celui-ci ne réalise pas encore qu'elles le sont. C'est ce que nous appelons les « situations critiques » telles qu'illustrées à la figure 5.

Dans un test de 40 items avec 4 choix de réponses, l'étudiant peut théoriquement réussir 25 % des items en répondant au hasard. Il pourrait même éliminer certains leurres qui sont probablement erronés et par conséquent réussir plus de 25 % des items avec une connaissance partielle. Laveault et Grégoire (2002) comparent plusieurs méthodes qui permettent de calculer un score minimal en tenant compte de la probabilité de réussite de chaque item. Or, il devient difficile d'interpréter de façon critériée les résultats à un test de type QCM de 40 items dont 25 % ou plus d'entre eux pourraient être réussis sans connaissance complète de la situation. Les choix au hasard par l'étudiant produisent des distorsions dans ses scores observés et masquent sa vraie performance. L'expression du degré de certitude en la réponse choisie permet d'identifier les items répondus avec peu de certitude ou au hasard. L'interprétation critériée des résultats en est grandement facilitée et probablement plus juste. L'évaluateur peut ainsi dénombrer les items répondus avec succès et certitude pour chacun des critères spécifiés dans le tableau de spécification (voir tableau 7) dans la mesure où l'étudiant ne surestime pas systématiquement la qualité de la plupart de ses réponses.

En évaluation diagnostique, l'interprétation critériée des résultats pourrait varier en fonction de l'ambiguïté et du niveau taxonomique d'un item en particulier. La figure 5 présente un algorithme permettant d'établir les conditions nécessaires pour une évaluation diagnostique au niveau de chaque item. Au niveau de l'étudiant, l'interprétation pourrait varier en fonction de son manque de réalisme. Par exemple, le feedback constitué des situations critiques de la catégorie B du tableau 4 devrait tenir

compte des propriétés métriques et taxonomiques de ces items. L'étudiant qui a une propension à surestimer la qualité de ses réponses pourrait bénéficier d'une rétroaction qui favoriserait une remédiation au niveau individuel. Si par contre la situation critique se produit alors que l'étudiant peut être considéré réaliste en général, il faut alors vérifier si le groupe est réaliste par rapport à l'item. Si tel est le cas, il faut favoriser la régulation individuelle au niveau de l'apprentissage cette fois-ci. Dans le cas où tout le groupe manquerait de réalisme par rapport à un item, le manque de réalisme de l'étudiant pourrait bien être attribuable à la pauvre discrimination d'un item. Il est concevable et acceptable que l'étudiant montre de l'incertitude devant un item ambigu. Dans le cas contraire, on doit favoriser la régulation à un niveau collectif et viser l'apprentissage ou le réalisme.

L'étudiant pourrait également avoir une propension à surestimer la qualité de ses réponses de façon généralisée. La littérature suggère que le degré de certitude exprimé, et par conséquent le réalisme, puisse être influencé par plusieurs caractéristiques individuelles et par les propriétés des items du test. Nous devons bien comprendre l'effet des différences individuelles et les processus cognitifs qui influencent le réalisme. Les interprétations qui seront inférées à partir des résultats ne seront valides que si les mécanismes sous-jacents sont bien compris. Le jugement diagnostique qui ne prendrait pas en considération les caractéristiques des items de certaines sections du test et qui ne tiendrait pas compte de certaines particularités des individus testés risquerait d'être erroné et peu pertinent. Nous soupçonnons que plusieurs facteurs peuvent influencer le réalisme de l'étudiant. La figure 5 illustre toute la complexité de l'interprétation critériée

des résultats à un test diagnostique prenant en compte le degré de certitude exprimé. La présence plus ou moins importante de différences individuelles et d'effets des niveaux taxonomiques des items est de nature à influencer la complexité de la tâche d'évaluation diagnostique. L'intervention de l'enseignant doit être guidée par le type de régulation requis.

Le feedback donné à l'étudiant n'a pas pour objet exclusif de favoriser la régulation de ses apprentissages. Lorsqu'il prend connaissance d'un manque de réalisme par rapport à l'ensemble du test ou quelques items, cette prise de conscience porte sur la régulation de sa perception. Carver et Scheier (2001) élaborent sur le sujet de la divergence de l'engagement de l'étudiant basé sur les principes de la théorie de la catastrophe où le degré de certitude représente une des variables. La certitude et le doute qu'a un étudiant envers l'accomplissement d'une tâche avec succès peuvent avoir un impact sur la persistance des efforts ou le désengagement par rapport à la tâche.

Cette recherche explore des facteurs tels le sexe, le niveau d'habileté de l'étudiant et le niveau taxonomique des items. Nous voulons vérifier si le manque de réalisme varie en fréquence et en importance selon le sexe et le niveau taxonomique des items. Si le réalisme des étudiants était lié au sexe et variait considérablement selon les niveaux taxonomiques des items, il est possible que nous soyons en présence de stratégies différentes de régulation. Nous sommes également intéressés à certaines caractéristiques des items telles que leur difficulté et leur discrimination.

Les buts de cette recherche sont de mieux connaître les facteurs qui affectent le manque de réalisme chez les étudiants. En fait, nous cherchons à déterminer si le

réalisme de l'étudiant est indépendant du test et quels sont les facteurs qui l'influencent. Les écrits antérieurs nous conduisent à considérer cette possibilité. Nous devons comprendre l'effet de ces facteurs pour que l'utilisation du degré de certitude soit valide dans un contexte d'évaluation diagnostique. En évaluation sommative ou certificative, ces facteurs n'auraient pas vraiment le même rôle à jouer et la même importance puisque l'intention de ce type d'évaluation n'est pas la même. Pour pouvoir effectuer un diagnostique approprié qui mène vers des pistes d'intervention, la compréhension des facteurs pouvant influencer et atténuer le réalisme est indispensable. Nous cherchons à assurer la justesse de l'aiguillage qui en découle. Tel qu'illustré à la figure 5, plusieurs indices individuels et de groupe liés au réalisme de l'étudiant sont donc pris en considération. La section portant sur le plan d'analyse des résultats présente les détails concernant le calcul de ces indices.

Questions de recherche

La revue de la littérature rapporte un certain nombre de recherches qui suggèrent que le réalisme de l'étudiant pourrait être affecté par des facteurs individuels et des facteurs liés aux propriétés des items du test. Dans le but de construire des tests diagnostiques critériés, nous cherchons à mieux connaître les facteurs qui affectent le manque de réalisme chez les étudiants. Les questions de recherche s'imposent donc de la façon suivante :

1. Il est concevable qu'un item qui serait difficile, ambigu ou qui s'adresse à des habiletés cognitives élevées engendre de l'incertitude. La question est de savoir si le réalisme de l'individu, qui est établi partiellement en fonction de sa

certitude exprimée, en est affecté. La question est de savoir comment et dans quelle mesure l'expression du degré de certitude en la réponse choisie, et par conséquent le réalisme des étudiants, varient-ils en fonction du niveau taxonomique des items et de leurs propriétés métriques.

2. Il est également probable que la propension qu'ont les individus à surestimer leurs habiletés puisse varier en fonction de certaines différences individuelles comme le sexe et le niveau d'habileté. Nous anticipons observer qu'un étudiant ayant un plus haut niveau d'habileté soit plus réaliste qu'un étudiant moins performant. En ce qui concerne le sexe, la littérature rapporte des résultats parfois divergents. Il est plus difficile d'anticiper la direction que prendront les résultats et la nature de l'interaction avec le niveau de performance. Il s'agit donc de déterminer comment et dans quelle mesure le sexe et le niveau de performance des étudiants sont-ils liés à l'expression du degré de certitude en la réponse choisie, et par conséquent au réalisme de l'étudiant?

LA MÉTHODE

Dans cette section, nous expliquons la méthode et les précautions utilisées lors de la cueillette de données. Nous présentons d'abord les participants et les instruments de mesure utilisés. Nous discutons ensuite de notre plan d'observation et de la procédure mis en place pour favoriser une cueillette appropriée des données. Finalement, nous expliquons les formules utilisées lors des calculs portant sur les variables de cette étude.

Les participants

L'échantillon visé est composé de 252 participants francophones ou francophiles inscrits à l'École des adultes Le Carrefour à Ottawa. Cette école est sous l'égide du Conseil des écoles publiques de l'Est de l'Ontario. Il s'agit d'un échantillon rendu disponible par les circonstances. Nous utilisons l'ensemble des participants qui se présentent à un test de classement en mathématiques rédigé en fonction des exigences du curriculum du secondaire du ministère de l'Éducation de l'Ontario. Les participants sont tous recrutés dans l'établissement même où se déroule la recherche. Le choix du site est motivé essentiellement par des raisons de faisabilité et par la disponibilité d'un centre d'évaluation supervisé. De plus, l'achalandage est considérable et permet de recueillir un échantillon de participants important dans un seul site tout en respectant les pratiques normales de l'école. Ceci nous permet de contrôler plusieurs variables.

152 copies complètes sont retenues pour les fins d'analyses pertinentes à cette recherche. Pour les besoins de cette recherche, nous ne pouvons pas traiter les données manquantes par imputation sans postuler des patrons de réponses et d'expressions du

degré de certitude. Après vérification avec Leclercq, nous concluons que ce genre de phénomène est fréquent. Les copies incomplètes seront utiles lors d'analyses ultérieures.

À la suite d'un constat de la population existante de l'école, nous avons anticipé une distribution relativement égale entre hommes et femmes. Comme le présente le tableau 6, nous avons retenu 83 copies chez les femmes contre 69 copies chez les hommes. Nous notons que l'échantillon comprend plusieurs participants nouvellement arrivés au Canada ou qui sont des immigrants de deuxième génération. Le niveau socio-économique des participants est hétérogène. L'échantillon est composé de participants francophones ou parlant le français à titre de langue seconde. Plusieurs participants détiennent des diplômes obtenus à l'étranger, alors que les autres cherchent à compléter leur 12^e année ou cherchent à développer des habiletés supplémentaires. Malgré que le français ne soit pas nécessairement la langue maternelle des participants, ils ont cependant tous une connaissance fonctionnelle du français. Cet échantillon de participants adultes semble représentatif de la population qui fréquente les écoles publiques destinées aux adultes en Ontario. Les niveaux de performance ont été catégorisés en utilisant 27 % des scores plus faibles et 27 % des scores supérieurs pour établir les trois catégories. Cette valeur limite a été choisie puisqu'elle correspond aux deux points d'inflexion de la courbe normale. Nous pouvons ainsi décrire plus facilement notre échantillon.

Tableau 6

Statistiques descriptives sur l'échantillon

Sexe / Performance	Faible	Moyenne	Supérieure	N total	Moyenne des scores	Écart type des scores
Femmes	26	40	17	83	18,89	8,839
Hommes	15	30	24	69	21,91	9,329
Total	41	70	41	152	20,26	9,159

Les instruments

L'instrument a été rédigé en fonction des exigences en mathématiques du curriculum du secondaire du ministère de l'Éducation de l'Ontario. Le test est construit à partir d'items développés par des enseignants qualifiés en mathématiques et qui connaissent bien la clientèle testée. Un système optique automatisé de correction et d'enregistrement des résultats est utilisé pour la correction des copies. La responsable du Centre d'évaluation de l'école inspecte d'abord les copies pour s'assurer que les participants les ont complétées. Ensuite, elle vérifie l'exactitude de la correction. Les bases méthodologiques et statistiques utilisées dans la construction du test suivent les directives puisées dans Laveault et Grégoire (2002).

Le test est de type QCM et comprend 40 items avec 4 choix de réponses. En plus des réponses aux items, le participant doit exprimer son degré de certitude par rapport à ses réponses. Par un choix, entre *pas sûr du tout*, *moyennement sûr* et *absolument sûr*, il

détermine ses chances de réussir l'item en termes de probabilité d'exactitude. Les items du test varient en matière de niveaux taxonomiques. Nous utilisons comme référence la taxonomie révisée des objectifs éducationnels de Bloom par Anderson et al. (2001). Nous retenons les quatre premières dimensions des processus cognitifs (« *Cognitive Process Dimension* ») de la taxonomie qui propose une hiérarchie d'objectifs cognitifs en 6 niveaux : 1) se souvenir, 2) comprendre, 3) appliquer, 4) analyser, 5) évaluer et 6) créer. La taxonomie originale de Bloom était pertinente. Nous considérons que la version révisée répond encore mieux aux besoins actuels. Quatre juges déterminent tour à tour le classement des items. La concordance inter-juges est calculée et favorise la validité du classement des items. Les items ne sont pas des cas typiques de chaque niveau taxonomique et ils n'ont pas été rédigés de manière à assurer une représentativité optimale des différents niveaux. Nous constatons qu'il est parfois difficile d'obtenir un consensus et que certains items sont des cas frontières. Devant cette difficulté, deux autres stratégies de classement sont également utilisées : un classement par domaine et un autre par extraction des composantes principales. La section portant sur les résultats présente l'analyse de l'effet de ces variables.

Selon les activités normales de l'école, un test pilote a été mis à l'essai en janvier 2004. L'instrument utilisé dans cette recherche est en quelque sorte la version révisée de ce dernier. Le test comprend 40 items avec 4 choix de réponses et 7 choix de degrés de certitude. Le tableau 7 présente les spécifications du test révisé. Nous remarquons que les items sont distribués selon les niveaux scolaires, les filières d'études et leur niveau taxonomique.

Tableau 7

Tableau de spécification après les révisions

Niveaux taxonomiques	Niveaux scolaires et filières d'études							Nombre d'items
	9 ^e - Appl.	9 ^e - Théo.	10 ^e - Appl.	10 ^e - Théo.	11 ^e - Pré-C	11 ^e - Pré-C/U	11 ^e - Pré-U	
Connaissance	1			1			1	3
Compréhension	1	3	1	2	2	1		10
Application	4	4	4	2		1	2	17
Analyse	1	2	3	3			1	10
Nombre d'items	7	9	8	8	2	2	4	40

Légende :

Niveaux scolaires 9^e, 10^e et 11^e : correspondent aux niveaux scolaires tels que décrits par le curriculum du secondaire de l'Ontario. En Ontario, l'école secondaire comprend les niveaux entre 9^e année et 12^e année. Après la 12^e année, l'étudiant peut accéder à l'emploi, le collège ou l'université.

Appl. : Filière d'études appliquée qui mène à des métiers

Théo. : Filière d'études théorique qui mène à des études postsecondaires

Pré-C : Filière d'études pré-collégiale qui mène à des métiers

Pré-U : Filière d'études pré-universitaire

Pré-C/U : Filière mixte qui peut être utilisée en Pré-C ou en Pré-U

Après plusieurs vérifications, nous arrivons à la conclusion qu'un test diagnostique en mathématiques fondamentales, qui cherche à aider à reconnaître les acquis académiques et qui est destiné à des adultes immigrants, doit contenir entre autres des items du niveau taxonomique application. Il ne serait pas raisonnable de poser des

questions qui font appel à la mémoire. Les étudiants visés sont adultes et plusieurs d'entre eux n'ont pas fréquenté l'école depuis plusieurs années. Un de ces étudiants pourrait savoir calculer l'hypoténuse d'un triangle rectangle sans se souvenir du nom de l'auteur du théorème. Historiquement, les tests de type QCM ont été largement critiqués pour avoir abusé des items de type connaissance. Cependant, cette critique ne s'applique généralement pas aux items utilisés dans le cadre de tests en mathématiques. Les items sont habituellement du type application. Les concepts de type compréhension peuvent facilement être évalués à l'aide d'items de type application. Il faut forcément avoir compris les concepts pour pouvoir les appliquer. De plus, les items de type application sont faciles à élaborer en mathématiques et sont particulièrement adaptés aux QCM qui font bon usage des problèmes à solution unique. Nous avons donc un test qui utilise majoritairement des items du niveau application.

Le plan d'observation

Cette recherche utilise une approche par questionnaire où le participant fournit deux réponses pour chaque question. La première réponse consiste à tenter de répondre correctement à l'item, alors que la deuxième consiste à fournir une autoestimation de la qualité de la réponse fournie. Nous proposons de tenir compte du sexe du participant, de son niveau de performance et du niveau taxonomique des items. Les données concernant le sexe et le niveau taxonomique des items sont disponibles *a priori*. Le niveau taxonomique des items est contrôlé durant la construction de notre instrument de mesure (test). Les participants sont catégorisés par le sexe. Une estimation du niveau de performance des participants est obtenue *a posteriori* par le score total au test

diagnostique. Nos variables dépendantes sont toutes enregistrées sur leurs échelles respectives durant la phase de testing.

La procédure

La procédure d'évaluation se déroule en cinq étapes principales durant une période d'environ deux mois.

- Étape 1 - Une séance d'information et de formation individuelle a lieu avec les orienteurs, l'enseignant et la responsable du Centre d'évaluation. Cette séance a pour but de revoir les étapes à suivre, les rôles respectifs des intervenants et l'interprétation des profils.
- Étape 2 - Le formulaire de consentement pour participer à la recherche et pour recueillir les données est remis au participant par l'orienteur lors de la rencontre initiale ou par la personne responsable du Centre d'évaluation. Ceux qui choisissent de ne pas participer doivent tout de même passer le test puisqu'il s'agit d'un test de classement obligatoire. Les données ne seront pas transmises au chercheur pour ceux et celles qui choisissent de ne pas participer.
- Étape 3 - Le test est administré dans un endroit réservé à cet effet.
- Étape 4 - La personne responsable du Centre d'évaluation effectue la correction et remet un profil de scores ainsi que le profil de réalisme à l'orienteur ou l'enseignant ayant recommandé le participant.

Étape 5 - L'orienteur ou l'enseignant portent le jugement diagnostique et détermine le classement du participant. L'orienteur ou l'enseignant se charge de communiquer les résultats au participant en mettant l'accent sur les items erronés que le participant estime avoir réussis.

Dans la même optique que Leclercq, nous établissons un profil de réalisme de l'étudiant. Le profil de réalisme de l'étudiant qui lui est communiqué dans un contexte d'évaluation formative et diagnostique peut constituer un feedback pertinent. Selon le point de vue opérationnel, le profil de réalisme et le degré d'atteinte des objectifs selon les critères du curriculum sont communiqués en personne par l'orienteur du programme ou par l'enseignant. Ce même profil permet à l'étudiant de valider l'autoestimation de ses connaissances. Ce même profil a des fonctions utiles pour l'enseignant et l'orienteur en leur permettant de réorienter leurs approches.

Les tests ont tous été administrés au même endroit sur une période de quelques mois. Les risques de diffusion sont minimes puisqu'il s'agit d'un test diagnostique pour des fins de classement. Il n'y a pas de conséquence irréversible ou d'intérêt à tricher.

Le plan d'analyse des résultats

Plusieurs analyses statistiques sont effectuées sur les mesures obtenues. Des calculs de la cohérence interne de l'instrument et des propriétés métriques des items permettent d'établir et de vérifier la qualité des items et de l'instrument. Nous vérifions ensuite la conformité des distributions aux postulats de l'analyse multivariée. Par la suite, nous comparons les effets des niveaux de performance sur le réalisme en fonction

du sexe de manière à vérifier notre première hypothèse. Finalement, la relation entre le coefficient de discrimination, le niveau de difficulté des items et le niveau taxonomique des items est analysée de manière à vérifier notre seconde hypothèse.

La prochaine section présente des formules permettant d'effectuer les calculs liés aux variables construites ou déduites de cette étude. Nous présentons ainsi les formules pour les indices de réalisme ainsi que pour les propriétés métriques des items. En ce qui concerne le réalisme, nous remarquons dans les travaux récents de Leclercq et Poumay (2005) un certain abandon de plusieurs indicateurs tels que la centration, l'indice de cohérence personnelle et l'acuité au profit d'interprétations graphiques plus conviviales.

De manière à répondre aux questions de recherche et à faciliter les analyses, les données sont organisées en deux matrices. La première matrice de données est organisée en fonction des réponses individuelles (R pour réponses) et permet les calculs liés au réalisme des étudiants R_s (voir prochaine formule) et à leur niveau de performance. Les étudiants sont aussi catégorisés selon leur sexe. La deuxième matrice de données met en évidence les propriétés métriques des items (Q pour questions), leur niveau taxonomique ainsi que le réalisme associé à chaque item, ce que Gilles (2003) appelle indice de réalisation des prédictions Rq (voir prochaine formule).

Le calcul du réalisme

Leclercq (1993) explique que le calcul de l'indice de réalisme est basé sur la concordance entre l'exactitude des réponses et le degré de certitude exprimé. Il s'agit de calculer l'erreur moyenne absolue de certitude (\bar{e}). Cette erreur de certitude est déterminée par le calcul de la différence entre le taux d'exactitude (T) et la valeur de la

certitude (C). Cette erreur doit être calculée pour chaque degré de certitude et sa moyenne représente le non-réalisme. Cette formule est utilisée dans cette recherche pour calculer le réalisme des étudiants (R_s) et l'indice de réalisation des prédictions des items (R_q).

$$\bar{e} = \frac{\sum_{i=1}^n (|C_i - T_i|) \cdot U_i}{J}$$

- i = indice des degrés de certitude
- n = nombre des degrés de certitude
- C_i = valeur de la Certitude i (l'échelle doit permettre de l'exprimer en pourcentages)
- R_i = nombre de réponses correctes pour la certitude i
- U_i = nombre d'utilisation de la certitude i
- T_i = Taux d'exactitude de la certitude i (en pourcentages) = $(R_i / U_i) \times 100$
- J = Nombre total de réponses = $\sum_i U_i$

Et l'indice de réalisme (sur une échelle de 1 % à 100 %) correspond à :

$$RÉALISME = 1 - 100 \cdot (\bar{e})$$

Le calcul de la difficulté et de la discrimination des items

L'indice p de difficulté de l'item représente la proportion des répondants ayant réussi l'item. Cet indice prend une valeur située entre 0 et 1. Plus la valeur est faible,

plus l'item est considéré difficile. Pour sa part, le coefficient de Discrimination D nous informe sur la capacité qu'a un item à départager les répondants qui ont eu un score élevé au test de ceux qui ont eu un score faible. Cet indice prend une valeur située entre 0 et 1. Plus la valeur est élevée ($D > 0,3$), plus l'item discrimine bien. Le concepteur de test recherche généralement des items ayant un coefficient de discrimination élevé. Les équations permettant d'obtenir ces valeurs sont les suivantes :

Indice p de difficulté

$$p = \frac{\sum x}{n}$$

x = nombre de répondants ayant réussi l'item

n = nombre de répondants

Coefficient de discrimination D

$$D = p_+ - p_-$$

p_+ = indice de difficulté p pour le groupe constitué des répondants ayant obtenu les scores les plus élevés (27 % des répondants)

p_- = indice de difficulté p pour le groupe constitué des répondants ayant obtenu les scores les moins élevés (27 % des répondants)

Les variables indépendantes et dépendantes

Le tableau 8 présente le plan d'observation qui est constitué des variables et des liens étudiés. Ce plan ne comporte que deux variables dépendantes construites, soit l'indice de réalisme des sujets (Rs) et l'indice de réalisation des prédictions de l'item (Rq). Elles sont traitées indépendamment selon la matrice analysée (soit R ou Q). La première matrice présente les données en fonction des items et met l'accent sur les propriétés métriques des items. La deuxième matrice de données est organisée selon les sujets et met en évidence les variables individuelles associées aux étudiants.

Tableau 8

Plan d'observation

Matrice Q portant sur les items		Indice de réalisation des prédictions (Rq) calculé à partir de la qualité de la réponse et le DC
Propriétés métriques de l'item (Difficulté p et discrimination D de l'item)		Régression multiple, ANOVA et linéarité
Niveau taxonomique de l'item (Connaissance, compréhension, application et analyse)		Régression multiple
Matrice R portant sur les sujets		
		Réalisme de l'étudiant (Rs) calculé à partir de la qualité de la réponse et le DC
Sexe	Niveaux d'habileté	Régression multiple, ANOVA et linéarité

Le sommaire des variables présenté au tableau 9 permet de distinguer les variables selon leurs caractéristiques et leur type. Selon la colonne, les variables sont organisées en fonction de leurs caractéristiques. Il peut s'agir des caractéristiques individuelles des participants comme le sexe ou le niveau d'habileté. Les caractéristiques de la tâche incluent le niveau taxonomique des items et leurs propriétés métriques. Les caractéristiques de la réponse réfèrent à la qualité des réponses, la certitude exprimée envers ces dernières et le réalisme calculé du sujet ou des items.

Nous considérons qu'une telle classification permet de mieux situer chacune des variables. Les variables sont également classifiées selon le type d'échelle. L'échelle peut être catégorielle ou continue. Les catégories possibles des variables catégorielles sont spécifiées. Les deux variables construites sont de type continu. Il s'agit d'abord du réalisme des sujets obtenu en établissant une relation entre la certitude exprimée des sujets et la qualité de leurs réponses. Nous calculons ensuite le réalisme des sujets par rapport aux items, ce que Gilles (2002) appelle *Indice de réalisation des prédictions de l'item (Rq)*.

Tableau 9

Sommaire des variables

Caractéristiques /Type des variables	individuelles	de la tâche	de la réponse
Catégorielle	- Sexe	- Niveau taxonomique de l'item (connaissance; compréhension; application; analyse)	- Exactitude de la réponse (1; 0)
Construite/ Continue	- Niveau d'habileté du sujet (somme de tous les items réussis)		- Réalisme du sujet (Rs) - Indice de réalisation des prédictions de l'item (Rq)
Continue		- Difficulté de l'item (degré p) - Discrimination de l'item (coefficient D)	- Degré de certitude (DC) (variable continue dichotomisée selon les probabilités = 0,02; 0,1; 0,25; 0,5; 0,75; 0,9; 0,98)

RÉSULTATS

Les mesures obtenues lors de l'étude sont soumises à plusieurs analyses statistiques. Nous procédons d'abord avec le calcul de la cohérence interne de l'instrument et au calcul des propriétés métriques des items. Cette vérification nous permet d'établir et de vérifier la qualité des items et de l'instrument. Plusieurs analyses subséquentes font référence à la difficulté des items et à leur coefficient de discrimination. Les propriétés métriques des items ainsi établies nous permettent d'analyser en partie leurs effets sur le réalisme calculé. Après la vérification de la conformité des distributions aux postulats de l'analyse multivariée, nous analysons ensuite l'effet potentiel des différences individuelles. Nous comparons les effets des niveaux de performance sur le réalisme en fonction du sexe de manière à vérifier notre première hypothèse. Finalement, l'effet des propriétés des items est évalué. La relation entre le coefficient de discrimination, le niveau de difficulté des items et le niveau taxonomique des items est ainsi analysée pour permettre de vérifier notre seconde hypothèse.

Notre plan d'observation, quoique complexe, ne comporte que deux variables dépendantes construites à partir de plusieurs autres variables dépendantes. Les deux variables dépendantes construites, soit l'indice de réalisme des sujets (R_s) et l'indice de réalisation des prédictions de l'item (R_q), sont traitées indépendamment selon la matrice analysée (soit R ou Q). Nous employons ainsi deux matrices de données. La première matrice pour sa part présente les données en fonction des items et met l'accent sur les propriétés métriques calculées et recueillies sur les items. La deuxième matrice de

données est organisée selon les sujets et met en évidence les variables individuelles associées aux étudiants. Les données manquantes ne sont pas traitables puisque nous n'avons pas de points de référence sur lesquels nous baser pour effectuer des imputations. De plus, étant une recherche exploratoire qui se penche sur un aspect docimologique fondamental, il nous apparaît risqué, sur le plan conceptuel, d'estimer des réponses manquantes.

Nous avons retenu, à des fins exploratoires, un seuil de signification de $\alpha = 0,05$ pour une erreur de type I. Des calculs permettent d'obtenir des valeurs pour des variables construites telles que le réalisme de l'étudiant, l'indice de réalisation des prédictions de l'item et le niveau de performance de l'étudiant. Les formules de Leclercq (1993) et de Gilles (2002) sont utilisées (voir méthodologie).

La cohérence interne du test

Le calcul de la valeur alpha de Cronbach de la cohérence interne du test est effectué afin d'obtenir une estimation conservatrice de la fidélité des mesures effectuées et de nous donner des indications de la qualité générale de l'instrument. Le calcul de cette valeur alpha porte sur le score aux items et sur le score total. Le calcul est également effectué de façon itérative après le retrait de certains items peu corrélés avec le score total et apparemment défectueux.

Le tableau 10 présente les valeurs obtenues pour la cohérence interne du test au moyen du calcul de la valeur de l'Alpha de Cronbach. Nous remarquons que la valeur du coefficient alpha (0,913) pour les 40 items avec 152 sujets est excellente considérant que ce test est récent et qu'il n'y a eu que deux cycles de mise à l'essai jusqu'à

maintenant. Deux items retiennent cependant notre attention : les items n^{os} 35 et 41 montrent une faible corrélation item/total et ne contribuent pas à la cohérence interne de l'ensemble du test. Le concepteur du test a été avisé de ces résultats. Nous retenons tout de même ces deux items lors de nos analyses ultérieures puisque l'information produite par ces items peut contribuer à l'amélioration de notre compréhension sur l'effet d'items qui discriminent peu sur le réalisme des sujets.

Tableau 10

Cohérence interne du test

Valeurs du coefficient alpha de Cronbach

CONDITIONS	VALEUR ALPHA	N ITEMS	N SUJETS
Score total	0,913	40	152
Sans item n ^o 41	0,919	39	152
Sans items n ^{os} 35, 41	0,922	38	152

Les propriétés métriques des items

Les calculs des coefficients de difficulté et de discrimination des items permettent de juger des propriétés métriques des items et de vérifier leur pertinence par rapport au type d'outil d'évaluation utilisé. Le tableau 11 présente le niveau moyen de difficulté des items ainsi que le coefficient moyen de discrimination. Le niveau de difficulté recherché dans un contexte d'évaluation diagnostique est habituellement de 0,50. Dans le cas du test, le niveau est de 0,504 donc le niveau de difficulté moyen est

idéal. Les coefficients de discrimination doivent être supérieurs à 0,3 selon les valeurs repères d'Ebel (1965). La discrimination moyenne des items du test est de 0,58. Deux items retiennent notre attention par leur niveau élevé de difficulté et conséquemment leur très faible coefficient de discrimination. Nous remarquons qu'il s'agit des mêmes items qui ont une faible corrélation item/total et qui ne contribuent peu ou pas à la fidélité du test. Globalement, nous pouvons conclure que la majorité des items (38/40) permettent de bien discriminer les étudiants selon leur niveau de performance.

Tableau 11

Propriétés métriques des items

Difficulté et discrimination des items

PROPRIÉTÉS	Difficulté p	Discrimination D
Moyenne des items	0,504	0,58
Item n° 35	0,17	-0,04
Item n° 41	0,34	-0,23

Distribution des résultats et statistiques inférentielles

Nous effectuons entre autres une analyse de régression multiple. Ce type d'analyse postule que les distributions soient conformes à certaines caractéristiques concernant la normalité, l'homoscédasticité et la multicollinéarité. De plus, une attention particulière doit être portée envers les cas extrêmes. Après avoir examiné les

distributions des variables ayant des données de type continu, nous retirons trois cas qui montrent des indices de réalisme extrêmes et qui n'ont vraisemblablement pas compris les conditions de la tâche, ce qui porte le nombre de sujets retenus à 149, soit 82 femmes et 67 hommes. Les postulats de normalité, de linéarité et d'homoscédasticité ont été vérifiés sur les variables suivantes : score total, réalisme des sujets et indice de réalisation des prédictions de l'item.

Nous avons déterminé, à l'aide de l'analyse Kolmogorov-Smirnov, que dans l'ensemble les distributions sont plutôt conformes aux postulats. Le tableau 12 résume les statistiques pertinentes aux distributions. L'objectif de cette recherche n'étant pas d'étudier en profondeur les propriétés métriques du test ainsi que les techniques d'échantillonnages, il semble néanmoins pertinent d'effectuer une transformation sur certaines distributions de manière à assurer la conformité absolue au postulat de normalité. Le tableau 12 ainsi que les figures 13, 14 et 15 permettent de constater la conformité des distributions. Le test z constitue en quelque sorte un test d'hypothèse sur les valeurs d'asymétrie et de kurtose en établissant le rapport entre leur valeur et leur erreur type respective. En plus d'une valeur z élevée, une simple inspection visuelle des distributions révèle à la figure 13 une distribution platykurtique voire même presque bimodale dans le cas du score des sujets. Le coefficient de voussure s'éloigne significativement mais modérément de la normalité ($z = 2,94$; $p \leq 0,05$) ce qui suggère que la population testée est relativement hétérogène. Il semble probable qu'il s'agisse de deux populations : une qui serait particulièrement instruite et une autre qui le serait moins. Cette observation cadre en apparence avec la réalité observée par les

pourvoyeurs de formation aux nouveaux arrivants. La clientèle est essentiellement composée de diplômés nouvellement arrivés au pays en quête de reconnaissance de leurs acquis et d'immigrants réfugiés ou qui rejoignent leur famille. Puisque cette variable est dichotomisée en trois groupes de performance (égaux en nombre) pour les fins de calculs, il ne semble pas pertinent de transformer la distribution. Nous devons néanmoins garder à l'esprit qu'une distribution bimodale tend à favoriser à la hausse les coefficients de discrimination des items.

Tableau 12

Statistiques sur les distributions

Statistiques sur les distributions des scores Delta

	N	Écart		Erreur		z	Erreur		Z
		Moy.	type	Asymétrie	type		Kurtose	type	
Score du sujet	149	13	4	0,176	0,199	0,887	-1,161	0,395	-2,940 *
Réalisme du sujet (Rs) - transformé	149	13	4	0,210	0,199	1,055	0,348	0,395	0,881
Réalisme P/R item (Rq) - transformé	40	13	4	-0,046	0,374	-0,122	0,471	0,733	0,643

Une transformation linéaire en scores Delta est également effectuée de manière à faciliter l'interprétation des résultats obtenus. En utilisant des variables ayant des distributions caractérisées par les mêmes paramètres de moyenne et d'écart type, il est

plus facile de comprendre les figures qui utilisent ces variables. Nous pouvons comparer des distributions ayant une échelle commune. La distribution (g) en scores Delta (Δ) est caractérisée par sa moyenne de 13 et par son écart type de 4. La transformation linéaire s'obtient en transformant les distributions en score z et en utilisant l'équation qui suit (Crocker et Algina, 1986; Laveault et Grégoire, 2002).

$$\Delta_g = 4 \cdot z_g + 13$$

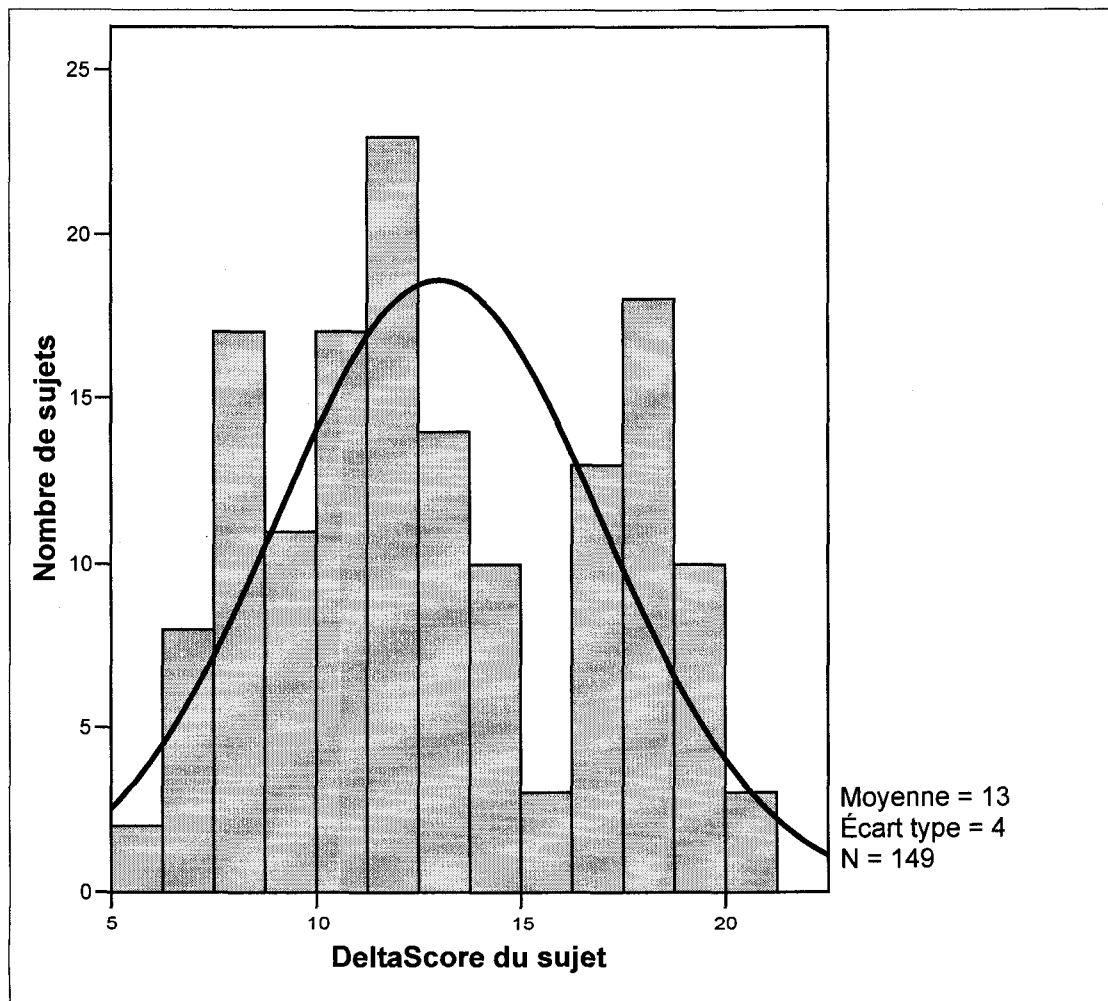


Figure 13. Distribution des scores totaux au test.

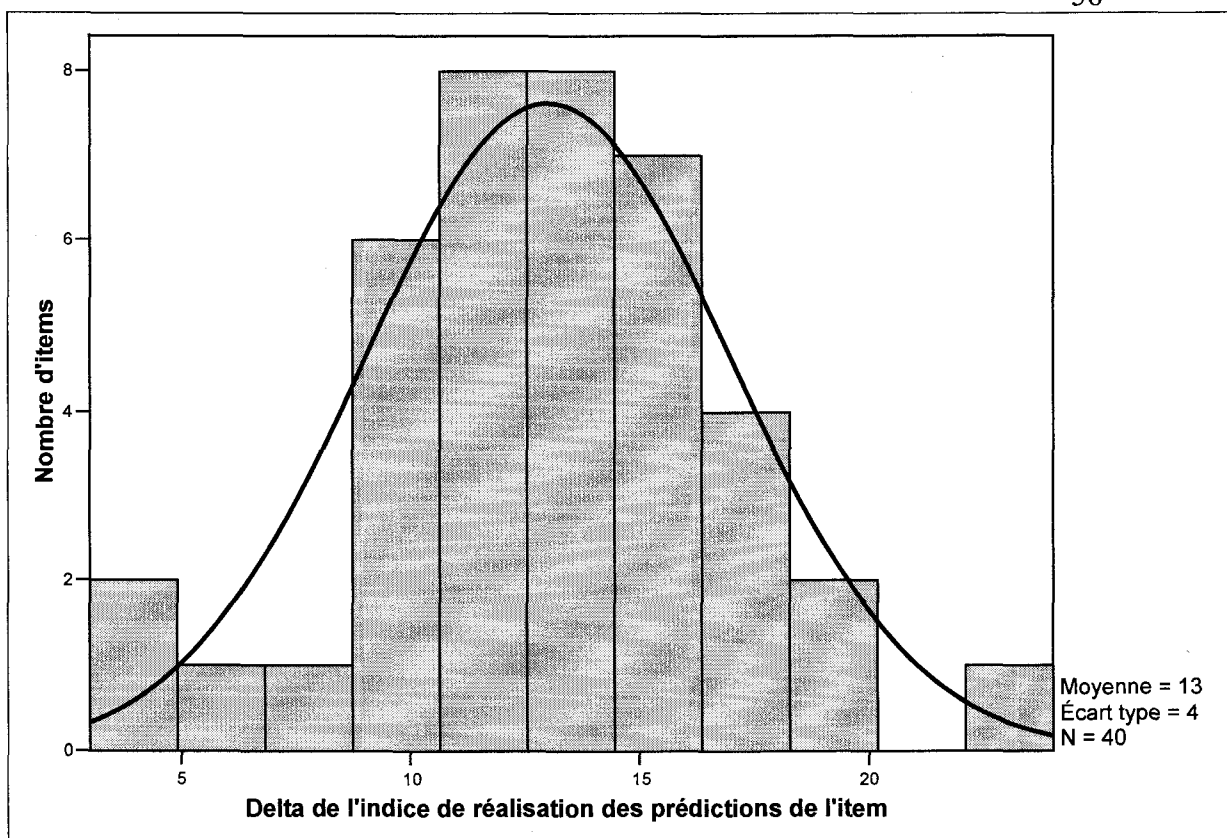


Figure 14. Distribution de l'indice de réalisation des prédictions de l'item (après transformation).

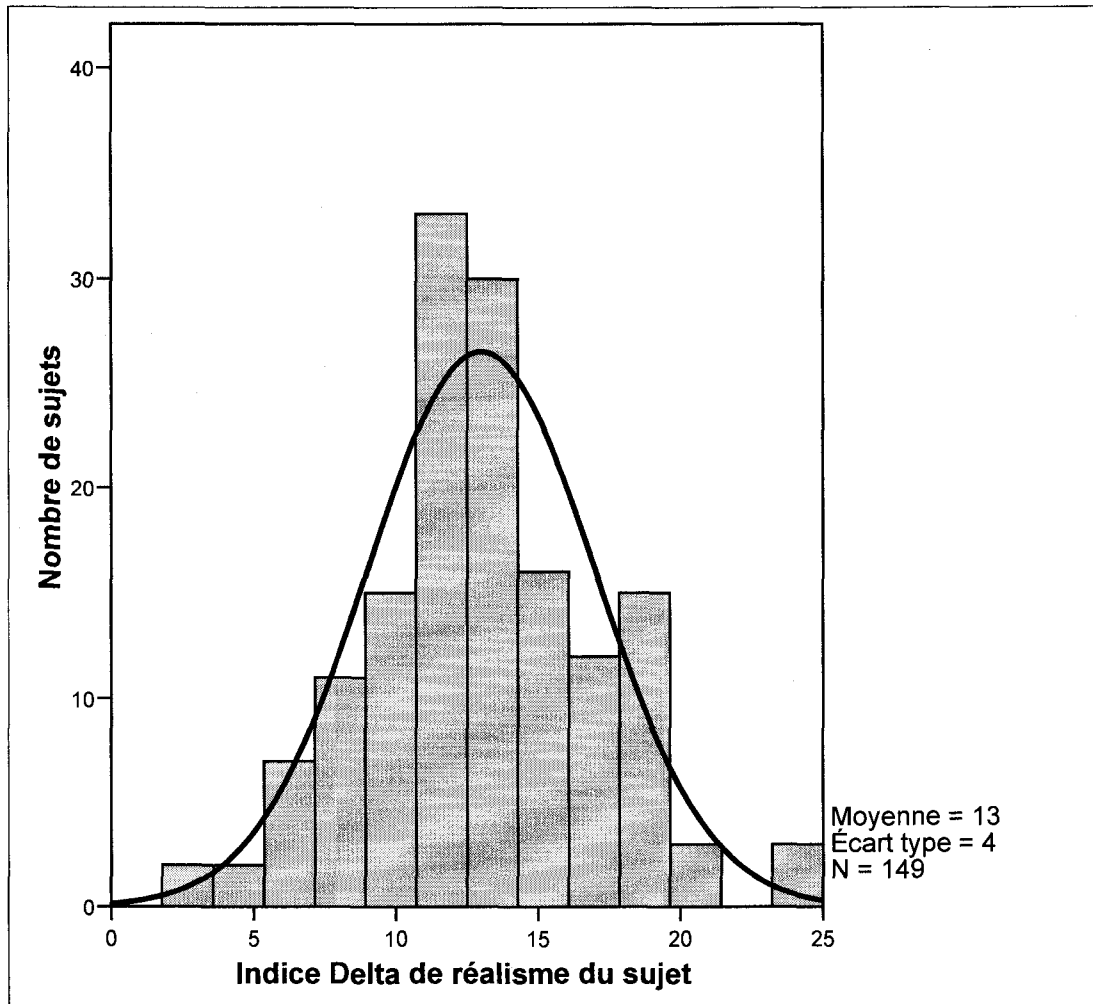


Figure 15. Distribution de l'indice de réalisme des sujets (après transformation).

Effet des différences individuelles

Notre première série d'analyses se concentre donc sur les données relatives aux sujets. Nous cherchons ainsi à répondre à notre première question de recherche qui porte sur le lien entre les différences individuelles et le degré de certitude exprimé, et par conséquent le réalisme de l'étudiant. Il s'agit de déterminer si le sexe et le niveau de performance de l'étudiant sont liés à son réalisme.

Deux types d'analyses sont effectués. D'abord, une analyse de régression multiple est effectuée pour déterminer le potentiel prédictif de nos deux variables indépendantes (sexe et niveaux de performance). Nous cherchons aussi à déterminer l'influence potentielle des variables individuelles et le poids respectif de celles-ci. Ensuite, de manière à bien comprendre la nature de la relation et des interactions entre les variables indépendantes sur le réalisme, une analyse de la variance en plan factoriel est effectuée. Il s'agit d'un plan d'expérience où les sujets ne sont testés que dans leur combinaison de conditions respective. L'analyse appropriée à ces données est l'analyse de la variance à deux dimensions pour échantillons indépendants. Il s'agit d'un plan factoriel à deux critères de classification 2 x 3, considérant que nous avons le sexe et 3 niveaux de performance.

Analyse de régression linéaire

Le test diagnostique de multicolinéarité de l'analyse de régression linéaire à deux variables indépendantes ne révèle pas d'évidence de colinéarité entre les variables. Les variables indépendantes ne sont donc pas redondantes et sont donc conservées. Nous effectuons une analyse de régression multiple en introduisant les variables indépendantes selon la méthode pas à pas « *stepwise* ». Le résumé du modèle présenté au tableau 16 indique que seul le score du sujet est retenu pour obtenir un coefficient de détermination (R^2) significatif de 0,372 ($F = 87,215$; $p = 0,000$). Le score du sujet explique 37,2 % de la variabilité totale du réalisme exprimé par le sujet. Le score du sujet est donc lié de façon significative à son réalisme. La mesure d'association révèle un Êta-carré de 0,625 qui constitue la corrélation qu'il est possible d'obtenir entre le

score et le réalisme de l'étudiant. Le test de linéarité pour sa part indique un écart de la linéarité significatif ($F = 2,420$; $p = 0,000$) ce qui nous porte à conclure qu'il y a présence de contributions non linéaires dans les données recueillies. La figure 19 permet également d'observer cet écart de la linéarité du lien entre le score obtenu et le réalisme exprimé par les sujets.

Tableau 16

Résumé du modèle proposé par l'analyse de la régression pour les sujets

Résumé du modèle de la régression multiple

Modèle	R	R carré	F	Sig.	Êta carré
Score du sujet	0,610	0,372	87,215	0,000	0,625

Modèle retenu et coefficients

	Coefficients bruts		Standardisés		
	B	Erreur standard	Beta	t	Sig.
Ordonnée à l'origine	5,067	0,888		5,703	0,000
Score du sujet	0,610	0,065	0,610	9,339	0,000

Analyse de la variance en plan factoriel

Nous effectuons une analyse de la variance en plan factoriel de manière à explorer si le sexe peut avoir une importance dans certaines conditions. Le tableau 17 nous présente les résultats de l'analyse de la variance effectuée selon un plan factoriel à deux critères de classification 2 x 3. Les valeurs F pour le sexe et la performance sont respectivement : 1,701 et 47,048. La valeur F pour le sexe indique que la différence n'est pas significative. Le devis ainsi que le nombre de sujets ne nous permettent pas d'avoir suffisamment de puissance avec 25,4 % de probabilité de rejeter l'hypothèse nulle lorsqu'elle est fausse. Nous soupçonnons que la curvilinearité du lien entre le niveau de performance des participants et leur réalisme peut réduire la puissance de notre analyse pour ce qui est des effets du sexe et de l'interaction sur le réalisme. Dans le cas des niveaux de performance, la valeur F est élevée et nous pouvons conclure qu'il y a une différence significative. Nous notons une valeur d'êta-carré assez élevée pour les niveaux de performance des étudiants (0,397), ce qui indique qu'une bonne proportion de la variance est expliquée par les différents niveaux de cette condition. Le R carré global nous indique aussi que 41,4 % de la variance de réalisme est expliquée par le modèle à deux critères de classification.

Tableau 17

Tableau résumé de l'analyse de la variance

Variable dépendante : Indice du réalisme						Éta-	
Source	SC	DI	CM	F	Sig.	partiel	Puissance
Sexe	16,512	1	16,512	1,701	0,194	0,012	0,254
Performance	913,188	2	456,594	47,048	0,000	0,397	1,000
Sexe * performance	22,093	2	11,047	1,138	0,323	0,016	0,247
Erreur	1387,799	143	9,705				
Total	27549,000	149					
Alpha = 0,05	R carré = 0,414						

L'analyse de la variance en plan factoriel permet de constater qu'il n'y a pas de différences significatives entre les moyennes des hommes et des femmes. À la lumière de ces résultats, nous ne pouvons pas rejeter l'hypothèse nulle concernant l'effet principal lié au sexe de l'étudiant. Il semblerait que le réalisme exprimé n'est pas lié directement au sexe de l'individu. Néanmoins, nous rejetons l'hypothèse nulle concernant l'effet principal lié aux niveaux de performance de l'étudiant. Les résultats soutiennent que le réalisme de l'étudiant varie en fonction de son niveau de performance. Par ailleurs, la valeur F pour l'interaction entre les deux conditions étudiées est de 1,138 et n'est pas significative. Tel qu'expliqué antérieurement, il n'y

aurait donc pas d'évidences d'interaction entre les deux variables indépendantes. La faiblesse du devis et la faible puissance à 0,247 sèment néanmoins des doutes quant aux probabilités de rejet de l'hypothèse nulle.

La figure 18 nous permet de mieux visualiser la relation entre le réalisme et le score. Elle nous permet de constater que plus les étudiants sont performants, plus ils ont un score de réalisme élevé, et ce, sans égard au sexe. La progression du réalisme semble s'accélérer chez les sujets plus performants. Le lien entre le réalisme des étudiants en fonction de leur niveau de performance semble de prime abord s'opérer de la même façon chez les hommes que chez les femmes. Malgré qu'il ne semble pas y avoir de différences significatives entre les moyennes des hommes et des femmes pour les différents groupes de performance, la variabilité des pentes en est pour le moins remarquable. Les figures (voir figure 20), qui portent sur la relation entre le score et le réalisme en fonction du sexe, montrent clairement que les droites de régression les mieux adaptés sont curvilinéaires et résultent de fonctions quadratiques. Les valeurs de R carré quadratique sont élevées à 0,563 pour les hommes et 0,529 pour les femmes. L'ajustement aux courbes quadratiques entraîne des résultats supérieurs aux droites suggérées par le modèle linéaire. Cette curvilinéarité, ou absence de linéarité, du lien semble plus évidente chez les femmes avec une différence importante entre le R carré linéaire et le R carré quadratique passant de 0,268 à 0,529. Cette différence est moins drastique chez les hommes avec un R carré linéaire de 0,474 et un R carré quadratique de 0,529. L'analyse graphique, bien que spéculative, nous permet d'observer un changement de tendance chez les participants moins performants. Le degré de réalisme

semble se distribuer de façon parallèle entre les hommes et les femmes qui sont de niveaux de performance moyens et forts. Cette tendance est beaucoup moins claire chez les participants de performance modérée et faible. Puisque cette tendance n'est pas significative et que l'erreur type des moyennes fait en sorte que les moyennes se situent vraisemblablement à l'intérieur de l'intervalle de confiance, il pourrait bien entendu s'agir d'un simple phénomène statistique.

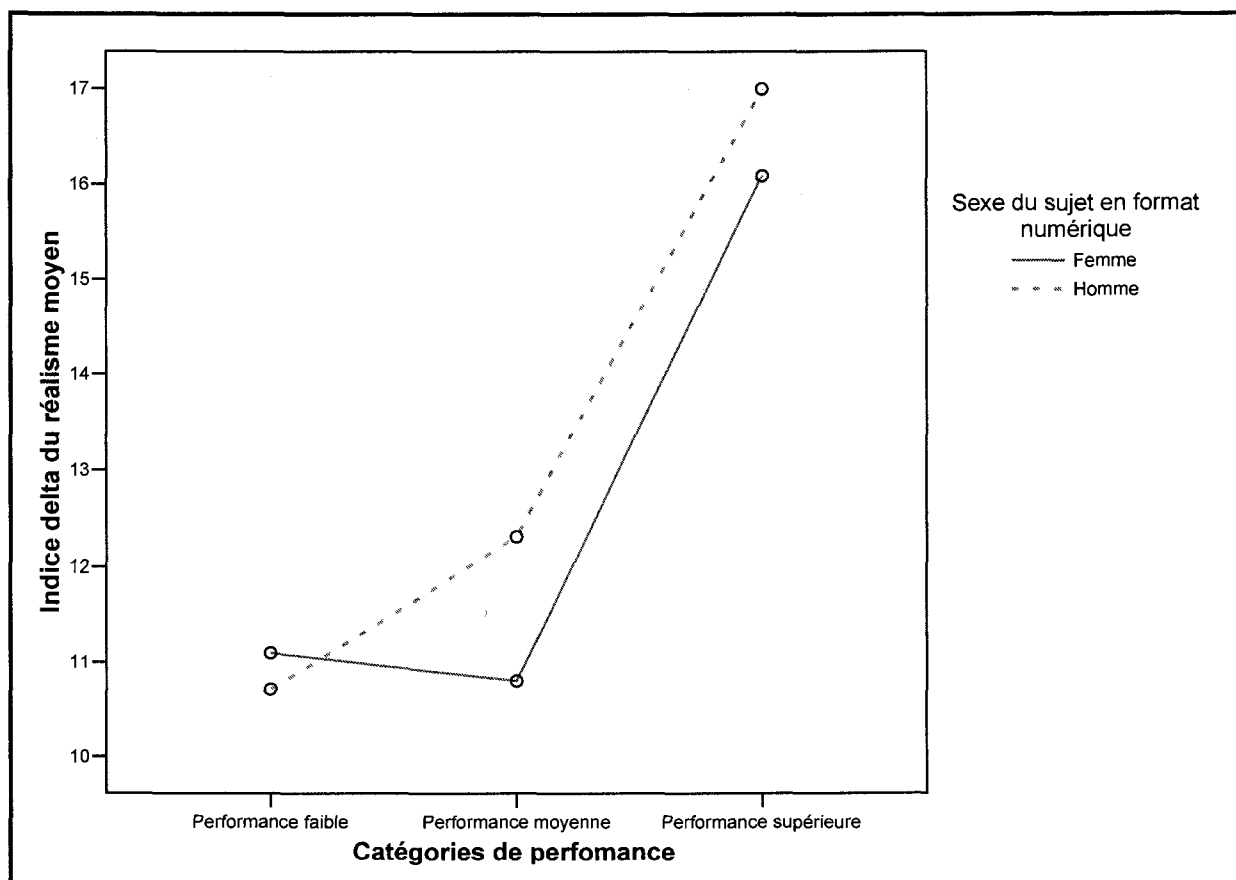


Figure 18. Réalisme moyen des sujets selon le niveau de performance et le sexe.

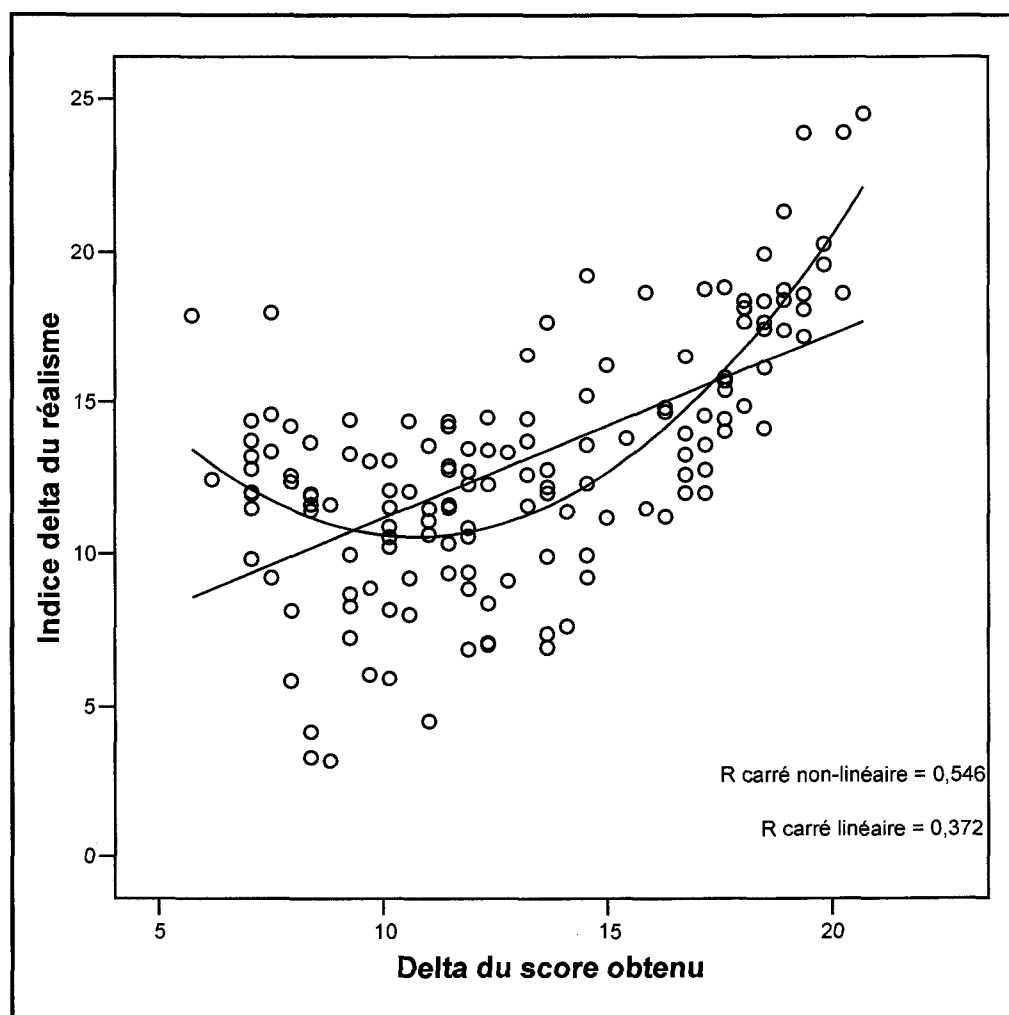


Figure 19. Représentation graphique portant sur la relation entre score et réalisme.

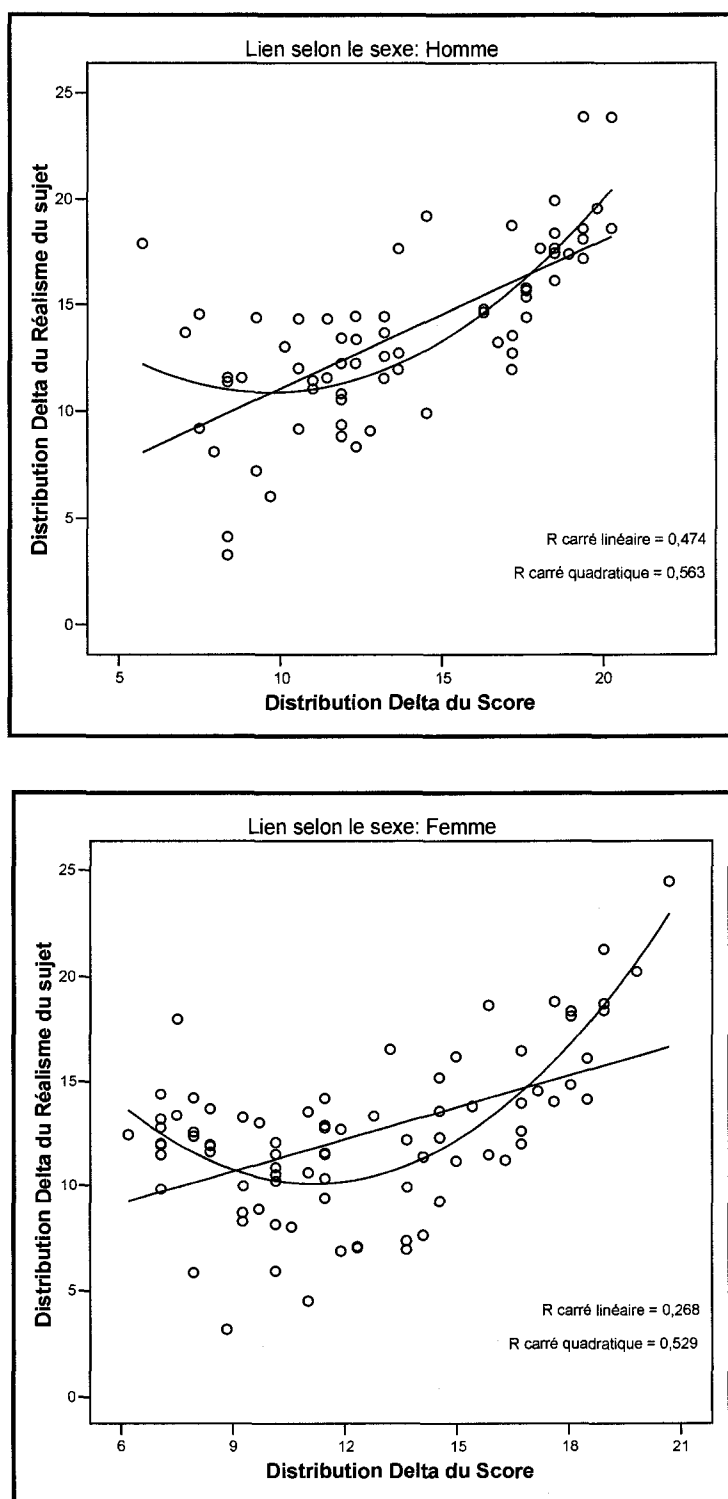


Figure 20. Relation entre le score et le réalisme en fonction du sexe.

Effet des propriétés des items

Après avoir examiné les liens entre les différences individuelles et le réalisme (Rs), notre deuxième série d'analyses se concentre sur les données relatives aux items. Nous cherchons ainsi à répondre à notre deuxième question de recherche qui porte sur le lien entre les caractéristiques des items, leur niveau taxonomique et le degré de certitude exprimé. Il s'agit de déterminer si le niveau de difficulté de l'item, son coefficient de discrimination et son niveau taxonomique sont liés au réalisme des étudiants. L'analyse appropriée à ces données est l'analyse de régression multiple qui permet de déterminer le potentiel prédictif de nos variables indépendantes.

Le tableau 21 présente la matrice de données recueillies. Les items sont triés selon le niveau de difficulté de ceux-ci de façon ascendante. La variable dépendante correspond à l'indice de réalisation des prédictions de l'item (Rq). Les variables indépendantes principales sont : degré de certitude moyen, coefficient de discrimination, degré de difficulté et classification taxonomique. Nous avons aussi choisi d'inclure d'autres classifications provenant de deux analyses factorielles et d'un spécialiste en mathématiques. Le rationnel derrière ces classifications est discuté ultérieurement.

Le test diagnostique de multicollinéarité de l'analyse de régression linéaire sur les variables principales ne révèle pas d'évidence de colinéarité entre les variables. Les variables indépendantes ne sont donc pas redondantes et sont donc conservées à l'exception du degré de certitude moyen. Le degré de certitude moyen exprimé par rapport à un item est de fait un excellent prédicteur du réalisme. Cependant, cette variable peut difficilement être considérée comme étant indépendante et n'ajoute rien à

nos analyses. Nous avons effectué une analyse de régression multiple en introduisant les variables indépendantes selon la méthode pas à pas « *stepwise* ». Le résumé du modèle présenté au tableau 22 indique que le niveau de difficulté et le coefficient de discrimination sont retenus pour obtenir un coefficient de détermination (R^2) significatif de 0,649 ($F = 34,188$; $p = 0,000$). Ces deux variables indépendantes ont donc une influence significative sur l'indice de réalisation des prédictions de l'item. Ces variables expliquent environ les 2/3 de la variation du réalisme associé à un item en particulier. Des analyses ultérieures démontreront que la contribution au modèle par le coefficient de discrimination est difficile à expliquer pour le moment et montre des faiblesses.

Comme le rapporte le tableau 22, le coefficient de discrimination contribue peu au R carré (0,046) du modèle à deux variables indépendantes. La mesure d'association entre le niveau de difficulté et l'indice de réalisation des prédictions de l'item révèle un Êta-carré de 0,902. Le test de linéarité entre le niveau de difficulté de l'item et l'indice de réalisation des prédictions de l'item ne détecte pas la présence de contributions non linéaires ($F = 0,816$; $p = 0,681$), ce qui suggère que la relation linéaire observée soit vraisemblablement adéquate. Nous remarquons cependant la présence potentielle d'une relation non linéaire ($F = 7,550$; $p = 0,031$) entre le coefficient de discrimination de l'item et la variable dépendante avec un Êta carré de 0,989 qui constitue une augmentation potentielle considérable de la corrélation actuelle.

Tableau 21. *Matrice des données sur les items*

Numéro des items	Analyse Factorielle à 3 composantes	Analyse Factorielle à 4 composantes	Classification par spécialiste en mathématiques	Classification taxonomique par experts	Niveau croissant de difficulté p	Coefficient de discrimination D	Degré de certitude moyen	Indice de réalisation des prédictions Rq
8	3	1	3	2	0,81	0,43	0,75	90,25
17	1	1		3	0,79	0,55	0,78	94,34
15	3	3		3	0,75	0,6	0,76	87,44
7		1	8	4	0,73	0,67	0,79	83,99
14	1	1		2	0,69	0,67	0,63	85,48
37	2		1	3	0,68	0,47	0,63	76,21
13	2			3	0,67	0,67	0,72	88,99
35		3	9	2	0,66	0,57	0,78	84,51
2	1	1	2	3	0,65	0,54	0,72	82,22
10	1	1		3	0,59	0,86	0,75	84,56
34	1	1	7	2	0,59	0,76	0,54	83,49
3		1	3	1	0,59	0,74	0,82	72,71
4	1	1		3	0,56	0,62	0,66	77,72
1	1	3		2	0,56	0,64	0,75	80,59
22	1	1		4	0,55	0,78	0,64	86,94
33	3		7	2	0,52	0,74	0,54	80,77
19	2	2	3	3	0,51	0,64	0,58	79,35
6		3		3	0,51	0,76	0,79	68,5
23			2	4	0,5	0,57	0,77	68,17
24	1	1	2	3	0,5	0,88	0,64	85,25
9			2	3	0,5	0,76	0,76	70,24
5	1	1		3	0,49	0,86	0,75	72,6
21	1			3	0,48	0,76	0,58	87,29
38	1	1	2	3	0,48	0,57	0,66	71,97
25	1			4	0,46	0,76	0,6	84,2
18	2	2		4	0,43	0,54	0,57	72,74
20	2	2	2	2	0,43	0,64	0,69	66,57
32	2	2	7	2	0,42	0,61	0,51	80,46
11	1	3	3	4	0,41	0,54	0,53	70,4
39	1	1		1	0,4	0,61	0,52	77,84
31		4	7	4	0,39	0,52	0,5	77,54
29	1			3	0,39	0,49	0,49	81,37
12	2		3	2	0,38	0,49	0,6	67,21
27	2	2	7	2	0,38	0,47	0,51	73,8
36				3	0,34	-0,23	0,57	56,26
26		4		4	0,32	0,68	0,5	74,04
28		4	7	1	0,31	0,2	0,6	61,05
40	2	2	8	4	0,3	0,42	0,53	73,29
16	2	2		4	0,3	0,42	0,74	50,76
30		4	5	3	0,17	-0,05	0,53	46,26

Légende: Classification taxonomique par experts et par consensus (1 = Connaissance; 2 = Compréhension; 3 = Application; 4 = Analyse)
 Classification par spécialiste en mathématiques (utilise 8 catégories comme critères qui correspondent à des domaines spécifiques du curriculum)
 Analyse factorielle à 3 ou 4 composantes (résume les patrons de corrélations entre les variables observées)

Tableau 22

Résumé du modèle proposé par l'analyse de la régression pour les items

Modèle retenu	R	R carré	Influence sur R carré	F	Sig.	Éta carré
Degré de difficulté et coefficient de discrimination	0,777	0,603	0,603	57,755	0,000	0,902
	0,806	0,649	0,046	34,188	0,000	0,989

Modèle retenu et coefficients

	Coefficients				
	bruts		Standardisés		
	B	Erreur standard	Beta	T	Sig.
(Constante)	1,133	1,550		0,731	0,000
Niveau de difficulté p	0,677	0,107	0,677	6,298	0,000
Coefficient de discrimination D	0,236	0,107	0,236	2,195	0,000

Variable dépendante: Indice de réalisation des prédictions de l'item (Rq)

Tests de contributions non linéaires

	F	Sig.
Niveau de difficulté p	0,816	0,681
Coefficient de discrimination D	7,550	0,031

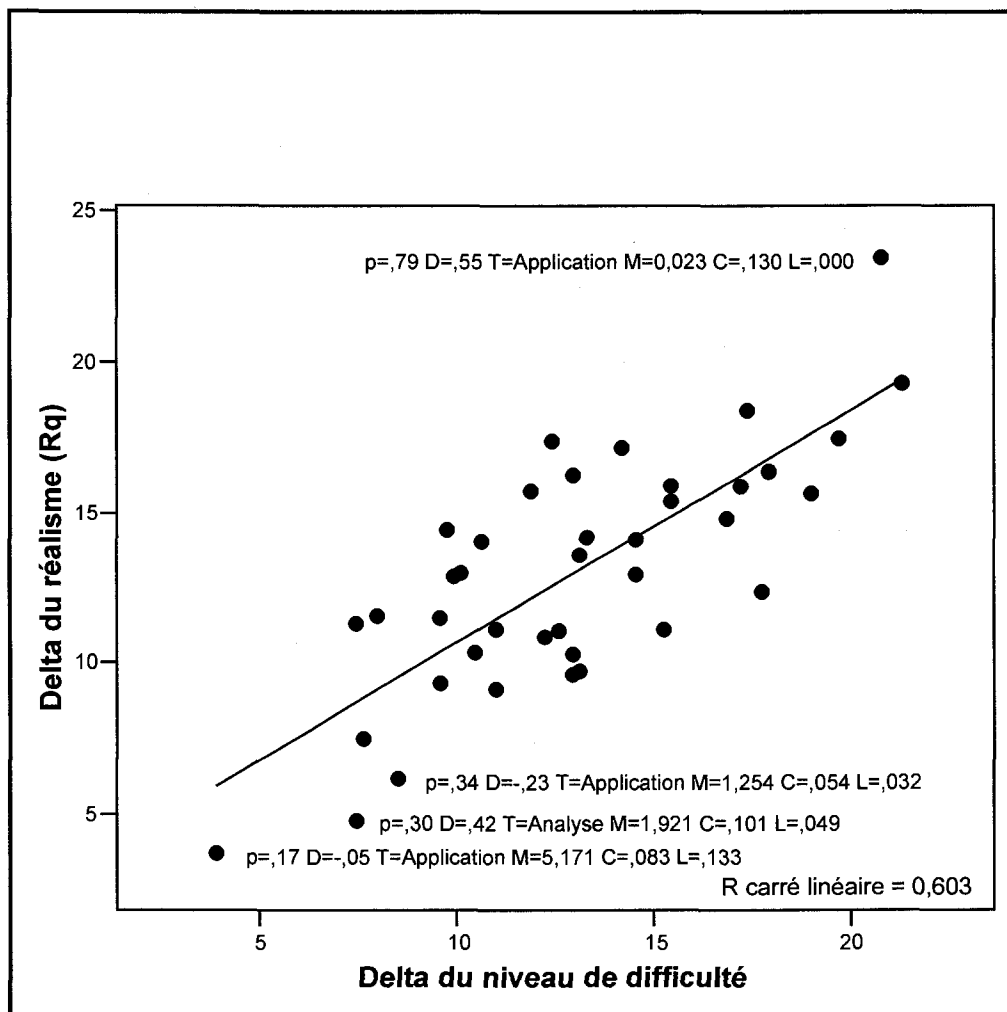


Figure 23. Représentation graphique sur le lien avec le niveau de difficulté.

La figure 23 présente la relation entre les données recueillies du niveau de difficulté des items et l'indice de réalisation des prédictions associé respectivement à ces derniers. Nous observons la présence d'un lien linéaire fort entre les deux variables (R carré = 0,601) et une tendance observable de l'augmentation de l'indice de réalisation des prédictions (réalisme associé à un item) en fonction de l'augmentation du niveau de difficulté de l'item. Trois valeurs au bas de la figure retiennent notre attention. Nous

avons indiqué leur niveau de difficulté (p), leur coefficient de discrimination (D), leur niveau taxonomique (T) et des mesures de distances : Mahalanobis (M), Cook (C) et levier (L). Nous notons que la valeur représentant l'item le plus difficile peut être considérée comme une donnée extrême ($M > 5$).

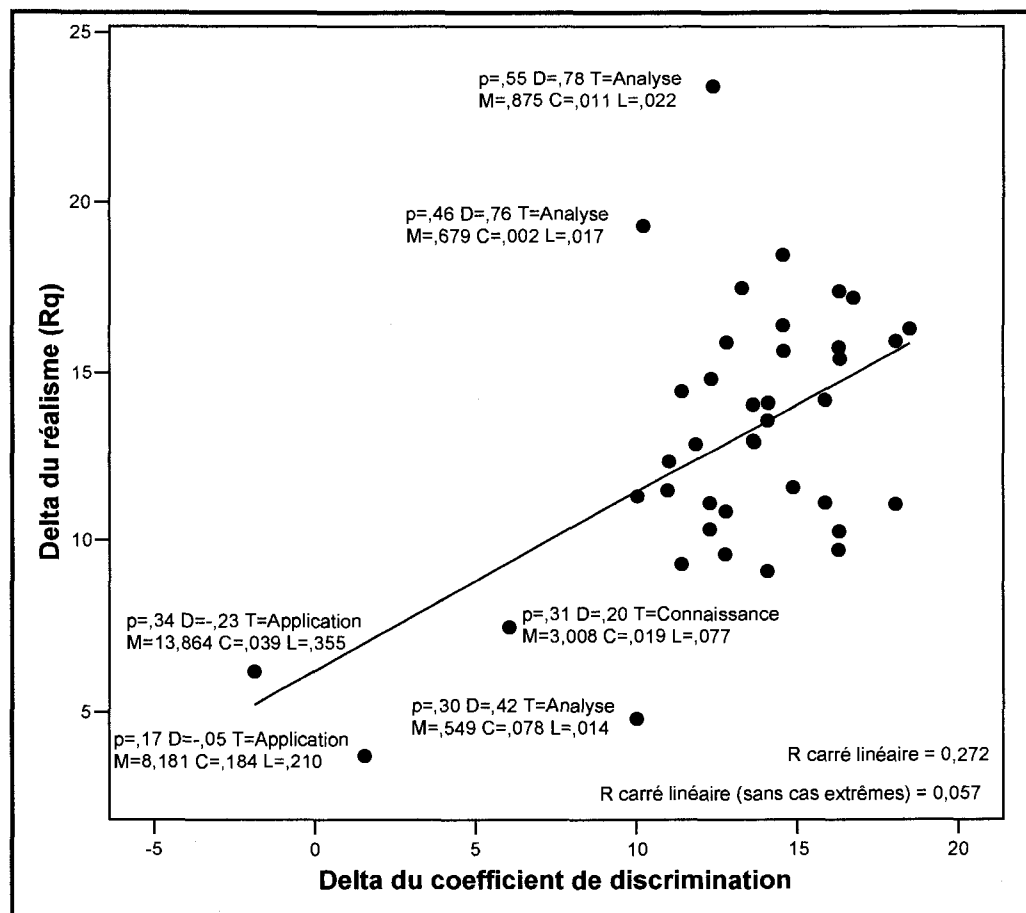


Figure 24. Représentation graphique sur le lien avec le coefficient de discrimination.

La figure 24 présente le lien entre le coefficient de discrimination de l'item et l'indice de réalisation des prédictions associé à celui-ci (R carré = 0,272). Nous

observons que la corrélation est fortement influencée par un petit nombre de valeurs extrêmes, qui sont d'ailleurs les mêmes observées à la figure précédente (figure 23). Les mesures de distances révèlent que les deux items qui discriminent le moins, en plus d'être considérées extrêmes ($M = 13,864; 8,181$), ont suffisamment d'effet de levier ($L = 0,355; 0,210$) pour influencer les résultats de l'analyse de régression. Deux des trois items ont un coefficient de discrimination D très faible et sont du niveau taxonomique application. Les mesures de distance nous informent que le retrait de ces items altérerait la pente de la droite de régression entre ces deux variables ainsi que la nature du lien linéaire observé. En supprimant les trois items en question, le R carré linéaire passe de $0,272$ à $0,057$. L'instrument utilisé lors de cette expérience est essentiellement composé d'items qui discriminent bien et il est difficile d'observer clairement ce que serait l'impact d'items qui discriminaient modérément ou peu. Paradoxalement, un instrument constitué d'un plus grand nombre d'items ambigus aurait été nécessaire afin de mieux observer la relation entre nos deux variables.

L'analyse portant sur l'impact du niveau taxonomique sur le réalisme ou l'indice de réalisation des prédictions pose des problèmes. La figure 25, qui est une représentation graphique du lien entre l'indice de réalisation des prédictions et le niveau taxonomique des items, montre la difficulté rencontrée. La position de chaque point par rapport à l'ordonnée nous indique pour chaque item son indice de réalisation des prédictions selon l'échelle delta. Chaque point ou item est aussi positionné selon sa catégorie taxonomique. La droite de régression linéaire la mieux ajustée montre une absence apparente de relation avec un r carré de $0,031$.

Certaines difficultés par rapport aux niveaux taxonomiques des items incluent le petit nombre d'items dans certaines catégories ainsi que la présence de plusieurs cas frontières qui sont difficiles à catégoriser. Nous n'analysons que les 3 derniers niveaux taxonomiques puisqu'ils ont suffisamment de données pour permettre une analyse. Même en faisant abstraction du premier niveau taxonomique, les résultats ne montrent pas vraiment de tendance générale. Devant cette difficulté et pour des fins exploratoires, nous avons tenté de catégoriser les items de plusieurs façons telles que présentées au tableau 21. Nous avons aussi obtenu une classification des items par un spécialiste en mathématiques qui correspond essentiellement aux domaines tels que spécifiés dans le curriculum. Lorsqu'il n'était pas possible de placer un item dans une seule catégorie, nous avons choisi de ne pas l'indiquer. Ces cas sont ainsi traités au même titre que des données manquantes. Au tableau 26, nous rapportons les pondérations factorielles de deux analyses factorielles à trois et à quatre composantes. À l'aide de la méthode graphique (*scree plot*) et des racines (*Eigenvalues*) du tableau 27, nous avons déterminé que les matrices à 3 et à 4 composantes peuvent avoir du potentiel. Les coefficients de saturation ainsi obtenus permettent d'identifier les items appartenant à une composante ou une autre. Plus le coefficient de saturation d'un item est élevée, plus sa contribution à la composante principale est élevée. Nous pouvons ainsi examiner quels sont les items qui semblent appartenir à une catégorie en particulier sans nécessairement savoir ce que signifie cette catégorie. Après avoir consigné les données recueillies selon ces nouvelles catégories, nous effectuons une analyse de régression multiple sur les variables portant sur ces catégorisations d'items.

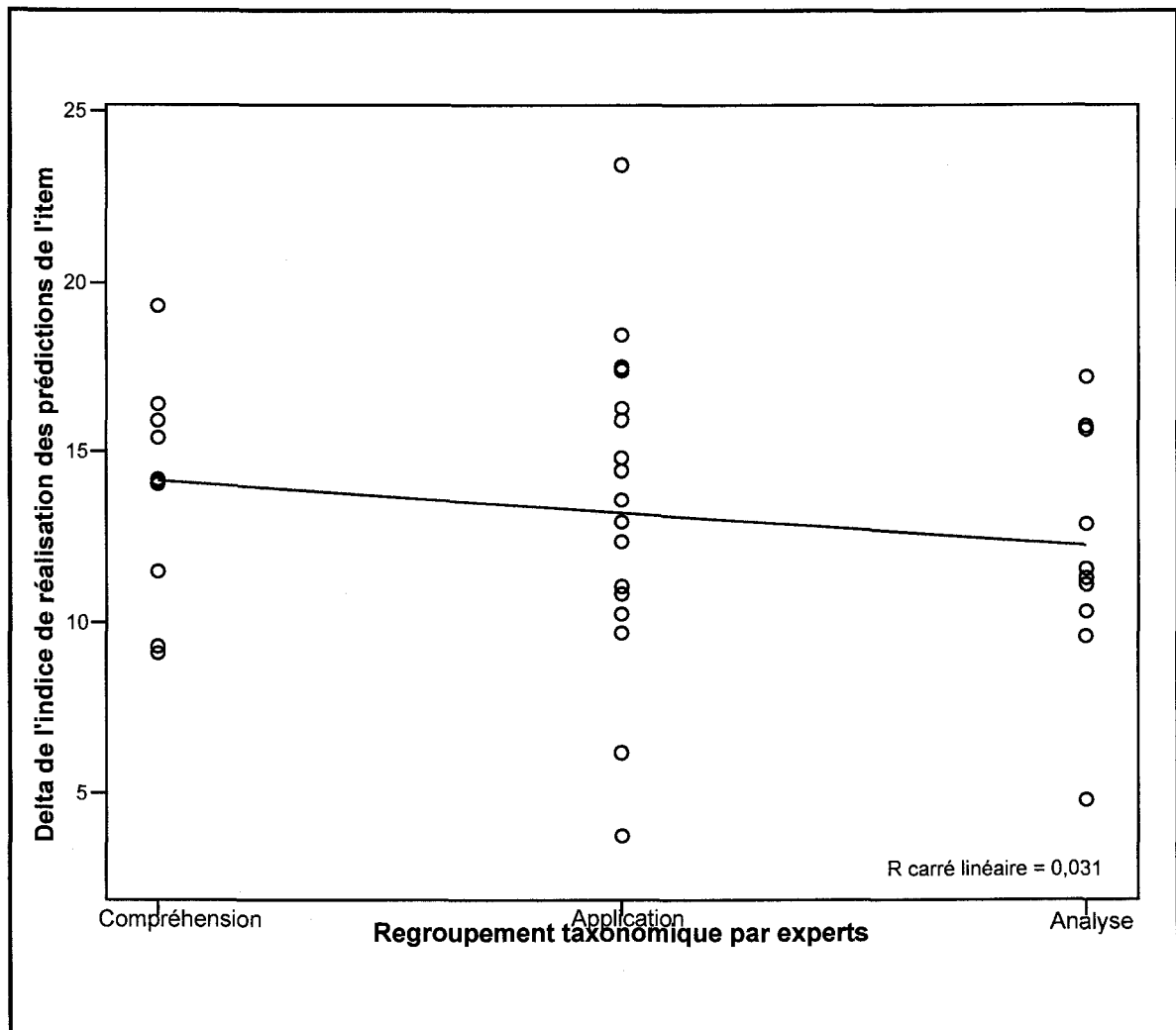


Figure 25. Représentation graphique sur le lien avec le niveau taxonomique des items.

Tableau 26

Matrices des pondérations factorielles

item	Composantes (3)			Composantes (4)			
	1	2	3	1	2	3	4
Q1	0,450	0,217	0,125	0,295	0,164	0,459	0,037
Q2	0,442	0,055	0,154	0,502	0,025	0,062	-0,111
Q3	0,316	0,283	0,381	0,391	0,267	0,143	-0,324
Q4	0,556	-0,056	0,078	0,600	-0,099	0,063	-0,029
Q5	0,643	0,189	0,017	0,623	0,129	0,216	0,093
Q6	0,330	0,311	0,214	0,096	0,262	0,625	-0,012
Q7	0,484	-0,026	0,446	0,397	-0,066	0,434	-0,308
Q8	0,255	0,021	0,585	0,309	0,016	0,201	-0,532
Q9	0,174	0,386	0,386	0,140	0,372	0,325	-0,284
Q10	0,531	0,082	0,101	0,520	0,036	0,189	-0,013
Q11	0,414	0,212	-0,139	0,249	0,156	0,362	0,282
Q12	0,267	0,350	-0,341	0,105	0,301	0,269	0,460
Q13	0,075	0,464	0,135	0,353	0,480	-0,333	-0,214
Q14	0,366	0,283	0,132	0,416	0,255	0,095	-0,076
Q15	0,255	0,037	0,612	0,193	0,022	0,419	-0,498
Q16	0,112	0,439	-0,060	0,041	0,417	0,204	0,142
Q17	0,518	-0,158	0,066	0,433	-0,207	0,259	0,034
Q18	0,136	0,437	0,198	0,203	0,430	0,072	-0,161
Q19	0,221	0,456	0,092	0,274	0,438	0,074	-0,044
Q20	0,026	0,549	0,110	-0,014	0,539	0,221	-0,031
Q21	0,472	0,268	0,084	0,404	0,219	0,302	0,039
Q22	0,523	0,259	-0,006	0,509	0,210	0,182	0,101
Q23	0,211	0,174	0,132	0,261	0,161	0,043	-0,105
Q24	0,453	0,221	0,024	0,538	0,187	-0,002	0,012
Q25	0,613	0,095	0,022	0,517	0,033	0,326	0,114
Q26	0,294	0,270	-0,300	0,214	0,227	0,132	0,377
Q27	0,089	0,542	-0,035	0,040	0,523	0,192	0,115
Q28	0,252	-0,005	-0,503	0,193	-0,043	-0,047	0,524
Q29	0,309	-0,005	0,103	0,218	-0,038	0,267	-0,011
Q30	-0,008	-0,050	-0,433	-0,036	-0,061	-0,138	0,408
Q31	0,240	0,252	-0,453	0,105	0,206	0,153	0,537
Q32	0,020	0,574	-0,161	0,131	0,572	-0,152	0,144
Q33	0,260	0,186	0,390	0,239	0,168	0,286	-0,300
Q34	0,494	0,101	0,055	0,456	0,054	0,214	0,040
Q35	0,292	-0,021	0,261	0,042	-0,063	0,604	-0,080
Q36	-0,155	-0,065	-0,205	-0,109	-0,051	-0,208	0,142
Q37	0,219	0,355	0,239	0,136	0,330	0,356	-0,120
Q38	0,416	0,291	0,112	0,556	0,266	-0,057	-0,095
Q39	0,488	0,128	-0,145	0,518	0,084	0,017	0,192
Q40	-0,124	0,445	-0,153	-0,053	0,454	-0,132	0,128

Méthode d'extraction : Analyse des composantes principales.

Méthode de rotation : Varimax.

Tableau 27

Racines (Eigenvalues) cumulatives

Variance totale expliquée Racines (<i>Eigenvalues</i>) initiales			
Composantes	Total	% de la variance	% cumulatif
1	6,624	16,559	16,559
2	2,535	6,339	22,898
3	1,904	4,761	27,658
4	1,635	4,088	31,747

Comme le rapporte de tableau 28, les quatre composantes identifiées par l'analyse factorielle constituent des prédicteurs acceptables du degré de difficulté, du coefficient de discrimination ainsi que de l'indice de réalisation des prédictions des items. Par ailleurs, le modèle à trois composantes constitue un prédicteur acceptable dans le cas du coefficient de discrimination de l'item. Nous avons analysé l'influence des autres variables, qui sont en l'occurrence la classification des items par un spécialiste de la matière et par un groupe d'experts cherchant à déterminer le niveau taxonomique, pour voir si elles avaient un lien avec le réalisme, l'indice de réalisation des prédictions et les propriétés métriques des items. Ces analyses ne rapportent rien de significatif et l'analyse de régression multiple ne produit pas de résultat.

Tableau 28

Résumé de l'analyse de la régression en fonction des regroupements proposés

Modèle à trois composantes :

Variables dépendantes (ou mises en relation)	R	R carré	F	Sig.	B	Constante
Coefficient de discrimination	0,408	0,166	5,388	0,028	-1,462	16,025

Modèle à quatre composantes :

Variables dépendantes (ou mises en relation)	R	R carré	F	Sig.	B	Constante
Degré de difficulté	0,522	0,272	10,468	0,003	-2,046	17,139
Coefficient de discrimination	0,517	0,267	10,199	0,003	-1,632	16,408
Indice de réalisation des prédictions	0,473	0,223	8,049	0,008	-1,759	16,527

Tel qu'illustré à la figure 29, le regroupement par la méthode à quatre composantes concorde avec la tendance générale au regroupement d'items de niveaux de difficultés comparables. En effet, comme il existe un lien entre le regroupement en composantes principales d'items dichotomiques et le niveau de difficulté des items, il est normal d'observer des patrons d'items tels que ceux présentés à la figure 29. Les items sont ainsi catégorisés selon leur regroupement factoriel obtenu par leur degré de corrélation avec les composantes indiquées sur une échelle de 1 à 4. Les composantes sont numérotées puisque nous ne connaissons pas les traits auxquels ils se réfèrent. Chaque point ou items sont placés en fonction de leur correspondance au niveau de difficulté tel qu'indiqué sur l'axe des Y. Le tableau 30 permet aussi de constater

certaines tendances en termes de propriétés métriques des items. Nous remarquons une nette corrélation entre le degré de difficulté des items, leur coefficient de discrimination et le degré de réalisme, ce qui est consistant avec les observations effectuées antérieurement. Après avoir effectué ces constats, les composantes identifiées ne semblent pas être liées à une forme de classification taxonomique quelconque. Des analyses ultérieures pourraient prendre ces composantes en considération.

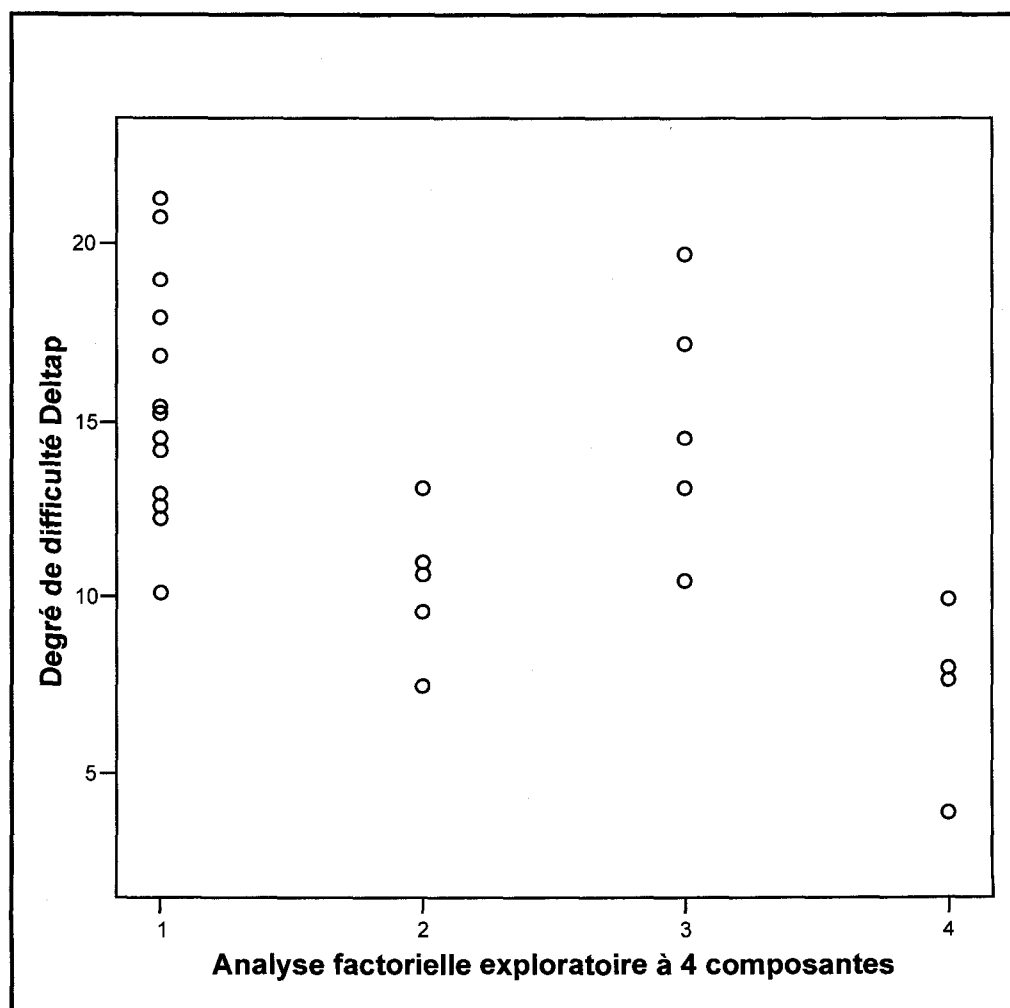


Figure 29. Lien entre degré de difficulté des items et regroupement factoriel.

Tableau 30

Propriétés métriques des items selon le regroupement factoriel

Composantes		Réalisme (échelle Delta *)	Difficulté (échelle Delta)	Discrimination (échelle Delta)
1	Moyenne	15,24	15,62	14,84
	Écart type	3,45	3,26	2,50
2	Moyenne	10,78	10,04	12,16
	Écart type	3,11	2,06	1,82
3	Moyenne	13,51	15,01	13,75
	Écart type	3,41	3,58	1,56
4	Moyenne	8,91	7,36	8,59
	Écart type	4,15	2,52	5,94
Total	Moyenne	13,07	13,11	13,20
	Écart type	4,09	4,31	3,47

* Échelle Delta : moyenne = 13; écart type = 4

En somme, les analyses effectuées permettent de constater qu'il y a des différences significatives entre les moyennes du réalisme associé aux items dans le cas de la difficulté et de la discrimination des items. À la lumière de ces résultats, nous pouvons rejeter l'hypothèse nulle concernant l'effet principal lié à la difficulté des

items. Il semblerait que le réalisme exprimé par rapport à un item est lié à ces propriétés des items. Nous ne pouvons pas rejeter l'hypothèse nulle concernant l'effet principal lié aux niveaux taxonomiques des items. Les résultats soutiennent que le réalisme de l'étudiant et sa certitude exprimée varient en fonction de la difficulté et de la discrimination des items.

DISCUSSION

Cette recherche proposait de vérifier, dans un contexte d'évaluation diagnostique et critériée, l'effet des propriétés métriques et cognitives des items sur le degré de certitude exprimé par les étudiants et par conséquent, sur le réalisme associé à un item en particulier. Également, nous avons fait l'hypothèse que les différences individuelles pouvaient aussi avoir une influence sur le réalisme des étudiants. Dans cette section, nous allons jeter un dernier regard sur les résultats obtenus et sur les pistes de recherche à considérer.

La revue de la littérature nous permettait d'anticiper que le réalisme pouvait varier en fonction de certaines différences individuelles. Il était ainsi raisonnable de croire qu'un étudiant très performant soit plus réaliste dans son appréciation de la qualité de ses réponses. Nous pouvions aussi présumer que les hommes soient plus téméraires et qu'ils aient une propension à surestimer leurs habiletés en mathématiques. Les différences individuelles que nous avons examinées sont le sexe de l'étudiant et son niveau de performance obtenu *a posteriori*.

Les résultats soutiennent comme prévu notre hypothèse que les étudiants plus performants sont plus réalistes par rapport à l'estimation de la qualité de leurs réponses que ceux qui sont moins performants, ce qui est conforme aux observations de Beyer (2002), de Fabre (1980), de Stankov (1998) et de Kleitman et Stankov (2001). Étant donné que cette modalité de réponse, qu'est l'expression du degré de certitude en la réponse choisie, est souvent utilisée en conjonction avec le score à l'item pour créer un score composite total, les concepteurs et les utilisateurs de ces tests doivent s'attendre

qu'une telle pondération des scores risque de favoriser davantage les étudiants performants que ceux qui le sont moins et permet d'augmenter la variabilité des scores. L'utilisation de cette modalité de réponse risque d'être plus pertinente et utile dans un contexte universitaire qu'elle le serait dans une école secondaire où la majorité des étudiants ne sont pas nécessairement connus pour leur haut niveau de performance scolaire. Dans un contexte diagnostique, où le résultat n'est pas pondéré, l'attention de l'orienteur sera davantage portée sur les items où l'étudiant généralement performant indiquera un degré de certitude moins élevé qu'à l'habitude et où l'étudiant moins performant fera preuve de manque flagrant de réalisme en indiquant un fort degré de certitude envers une réponse erronée. Les étudiants très performants et réalistes ne sont pas nécessairement ciblés par l'orienteur.

Par ailleurs, nous ne pouvons pas rejeter l'hypothèse nulle sur l'effet principal qui stipulait que le sexe de l'individu pouvait être lié à son réalisme, ce qui corrobore les constats de Henmon (1911), de Jonsson et Allwood (2003) et de Clack et Head (1999). Il semblerait que les hommes et les femmes de notre échantillon expriment leur certitude par rapport à leurs réponses avec des niveaux de réalisme relativement équivalents. Nous observons que plus les étudiants sont performants, plus ils ont un score de réalisme élevé, et ce, sans égard au sexe. La progression du réalisme semble s'accélérer chez les sujets plus performants. Le lien entre le réalisme des étudiants en fonction de leur niveau de performance semble de prime abord s'opérer de la même façon chez les hommes que chez les femmes. Un des constats inattendus de cette recherche fait état de l'absence apparente de linéarité de la relation entre le niveau de performance et le

réalisme exprimé. Le test de linéarité indique un écart de la linéarité significatif ($F = 2,420$; $p = 0,000$). Malgré qu'il ne semble pas y avoir de différences significatives entre les moyennes des différents groupes de performance, la variabilité des pentes en demeure pour le moins remarquable. Le lien entre le réalisme exprimé et le niveau de performance des étudiants semble linéaire chez les étudiants de performance moyenne et de performance élevée. Le sexe ne semble pas avoir d'impact différencié chez les étudiants de ces deux groupes. Ceci suggère que le réalisme chez les hommes et les femmes serait sensiblement équivalent chez les individus de performances moyenne et supérieure.

Nous remarquons cependant un écart de linéarité observable chez les étudiants de performance faible. L'inspection visuelle révèle que cet écart ne semble pas indépendant du sexe. Les différences de moyennes des hommes et des femmes ne sont cependant pas significatives. Elles se trouvent à l'intérieur de l'intervalle de confiance déterminé par l'erreur type. Beyer (2002) rapporte des résultats similaires en montrant des différences chez les hommes qui ont des attentes faibles envers la qualité de leur performance. Si de telles différences existaient, elles auraient eu pour effet d'affaiblir les analyses antérieures qui ont porté sur le sujet. Ces observations sont bien entendu spéculatives et nous fournissent tout au plus des indications d'une tendance qui pourrait être vérifiée ultérieurement. Par ailleurs, cette observation supporte un constat de Laveault et Grégoire (2002) qui fait état de la présence de liens curvilinéaires entre des variables liées au niveau de performance des individus.

La valeur F pour le sexe ($F = 1,701$) indique que cette variable ne semble pas avoir d'impact sur le réalisme exprimé par l'ensemble des étudiants. Le devis ainsi que le nombre de sujets ne nous permettent pas d'avoir suffisamment de puissance avec 25,4 % de probabilité de rejeter l'hypothèse nulle lorsqu'elle est fausse. La quasi-absence de relation entre le sexe de l'individu et son réalisme peut aussi être imputable à plusieurs autres facteurs. Les recherches consultées portaient principalement sur des enfants. La situation chez les adultes n'est peut-être pas la même que chez les enfants et mérite sans doute d'être explorée davantage.

Somme toute, ces résultats suggèrent que le réalisme des étudiants serait indépendant du sexe chez les étudiants plus performants. Du même coup, ces résultats semblent suffisamment de doutes pour justifier des analyses ultérieures sur des échantillons plus volumineux de sujets de performance moyenne. Ces analyses prendraient en considération le manque de linéarité potentiel entre les variables de performance et le réalisme exprimé de manière à ne pas sous-estimer le degré de relation entre ces variables. Des méthodes d'analyses plus robustes pourraient ainsi être utilisées sans aller à l'encontre des postulats de l'analyse multivariée.

Un autre aspect de notre recherche consistait à vérifier l'effet des propriétés métriques et cognitives des items sur le degré de certitude exprimé par les étudiants et par conséquent sur le réalisme associé à un item en particulier. Le niveau de difficulté de l'item constitue la première variable portant sur les propriétés des items analysés dans cette recherche. Les résultats appuient la prémisse que plus les items sont faciles, plus la certitude et le réalisme de l'individu seront élevés. Ceci concorde d'ailleurs avec les

éléments répertoriés dans notre revue de la littérature (Jacobs, 1974; De Finetti, 1965).

Cette conclusion n'est pas particulièrement surprenante mais confirme néanmoins le potentiel d'impact de la qualité métrique d'un item sur le réalisme exprimé par rapport à cet item. Un item mal réussi et difficile qui aurait un indice de réalisation des prédictions faible est en quelque sorte un item pouvant être considéré comme défectueux. Le degré de certitude exprimé par rapport à un item constitue un autre élément pouvant s'ajouter aux propriétés métriques de l'item et qui peut servir à juger de sa qualité intrinsèque, de sa pertinence et de son utilité pratique. Dans un contexte diagnostique, il est évident qu'une attention particulière doit être portée aux items là où l'étudiant a exprimé un degré de certitude élevé à un item échoué. Ce sont ces situations critiques, telles que présentées au tableau 4 sous la catégorie B, qui doivent retenir l'attention de l'orienteur, du professeur et de l'étudiant.

Nous pouvons conclure de la même façon par rapport au coefficient de discrimination de l'item puisque le réalisme de l'étudiant semble varier en fonction des différents niveaux de discrimination de l'item. Plus un item discrimine bien, plus l'étudiant a une propension à être réaliste en la qualité de sa réponse. Un peu comme pour le niveau de difficulté d'un item, un item qui discrimine peu ou pas peut être défectueux de fait ou entraîner une confusion chez l'étudiant qui n'est pas souhaitée. Ces items sont à éviter dans à peu près tout contexte d'évaluation puisqu'ils fournissent peu d'information utile. Après avoir analysé les données, il nous est apparu que pour fins de recherches ultérieures, le chercheur pourrait trouver intéressant d'utiliser un instrument qui est constitué d'items moins performants et de degrés variables de

discrimination. Nous soupçonnons que la relation entre le coefficient de discrimination de l'item et le réalisme des étudiants associé à un item mérite d'être étudiés et documentés en profondeur.

Pour leur part, les niveaux taxonomiques posent plusieurs défis de taille. Même s'il est intuitivement évident que les items de niveaux taxonomiques d'ordre supérieur entraînent une baisse du degré de certitude exprimé, la classification même de ces items est complexe et vient quelque peu brouiller les résultats. Les items du niveau application sont communs dans les instruments d'évaluation en mathématiques. L'instrument choisi comporte une majorité d'items de ce niveau. Ceux qui ne l'étaient pas peuvent être considérés comme des cas frontières en ce sens qu'ils ne s'éloignent que rarement des habiletés cognitives de niveau application. Or, ayant ainsi un échantillon d'items relativement petit, les conditions n'étaient pas favorables à une étude en profondeur des effets des différents niveaux taxonomiques sur le réalisme des étudiants associé aux items. Les tentatives de regroupement factoriel ont révélé certains patrons. Les résultats de ces analyses ne permettent pas d'identifier une forme de classification taxonomique quelconque.

Comme Gilles (2002) dans le cadre de ses travaux portant sur l'analyse des items, nous avons pu observer que les DC peuvent présenter des avantages en ce qui a trait au repérage des items défectueux au même titre que le degré de difficulté d'un item, de son coefficient de discrimination ou de sa corrélation item total. Cet avantage rend son utilisation pertinente lors d'évaluation diagnostique ou sommative, avec ou sans pondération. De plus, l'interprétation des scores par les professeurs et les orienteurs a

l'avantage de bénéficier de plus d'information et permet de mettre l'accent sur les situations critiques. L'expression « situations critiques », rappelons-le, est l'appellation que nous donnons aux situations caractérisées par une dissonance critique entre la certitude exprimée et la qualité de la réponse telle qu'illustrée par la catégorie B du tableau 4. Cette situation se produit lorsqu'un étudiant choisit une réponse erronée avec un degré de certitude élevé. Nous considérons que l'étudiant ne sait pas qu'il commet une erreur et que nous avons le devoir de l'aviser. Ce nouveau protocole s'inscrit dans la régulation des apprentissages en ce sens qu'il offre un feedback pertinent à l'orienteur et à l'étudiant qui auront intérêt à porter attention à ces situations critiques de façon prioritaire.

Nous cherchons à mieux comprendre les facteurs qui affectent le réalisme. Le réalisme de l'étudiant fait partie du processus d'autoévaluation et de régulation des apprentissages. Le profil de réalisme de l'étudiant peut contribuer à une évaluation diagnostique plus valide, fidèle et utile. Nous pouvons accroître l'utilité de l'évaluation diagnostique si nous pouvons mieux comprendre comment les différents niveaux taxonomiques des items et les propriétés métriques des items peuvent faire varier le réalisme des étudiants. Cette connaissance est d'autant plus nécessaire que l'autoévaluation est encouragée dans les programmes de formation.

Nous considérons que les degrés de certitude ont des rôles qui varient fondamentalement selon qu'il s'agit d'une évaluation certificative (bilan), évaluation avec feedback (formative), ou encore diagnostique. La fonction utile des DC est ainsi

liée à son contexte d'utilisation. L'utilisation traditionnelle des DC est liée au contexte certifiatif. Nous devons explorer davantage leur utilisation dans les autres contextes.

Suite au constat de la forte proportion de données manquantes ou de copies incomplètes, nous considérons que l'utilisation des DC nécessite beaucoup de formation ou de pratique. Leclercq (1993) en était déjà parvenu à cette conclusion. De plus, les DC représentent une modalité de réponse supplémentaire qui n'est pas habituelle et qui revêt un certain degré de complexité. Cette modalité ajoute un fardeau qui n'est pas négligeable et qui peut à la limite constituer une source de confusion pour l'étudiant, et plus particulièrement, pour l'étudiant peu expérimenté par rapport à la passation de tests. Un étudiant universitaire habitué à être évalué et à réguler son apprentissage risque de saisir plus facilement les conditions de la tâche et à tirer parti des avantages éventuels d'indiquer son choix à toutes les modalités de réponse.

Nous considérons aussi qu'un rapport coûts/bénéfices doit être calculé avant de s'engager à élaborer un test ayant les DC comme modalité supplémentaire de réponse. Cette modalité de réponse augmente considérablement la durée de la passation du test. De plus, la certitude exprimée par rapport aux réponses n'est pas nécessairement liée de façon linéaire au trait mesuré par les items du test. La figure 19 montre explicitement un lien apparemment curvilinéaire entre le score et le réalisme des individus. De plus, il n'y a aucune évidence que le réalisme soit une partie intégrante du trait mesuré. Nous nous retrouvons ainsi en présence de difficultés d'ordre conceptuel dès que nous cherchons à combiner ces deux caractéristiques.

Si l'objectif du concepteur du test était d'utiliser un score composite de manière à maximiser l'indice de fidélité du test, il serait probablement préférable d'augmenter le nombre d'items plutôt que de pondérer en fonction du DC. L'augmentation du nombre d'items aurait probablement le même effet sur la fidélité du test, et ce, sans compromettre la validité conceptuelle de ce dernier. En utilisant la qualité de la réponse ainsi que le DC exprimé par rapport à ces réponses à des fins formatives et diagnostiques plutôt que sommatives, le concepteur du test demeure ainsi dans une zone confortable et évite plusieurs considérations d'ordre conceptuel.

De plus, puisqu'il s'agit d'utiliser les DC dans un contexte diagnostique et critérié, le niveau de performance de l'étudiant est établi en fonction de l'atteinte du seuil de maîtrise de la compétence évaluée. Le seul intérêt pour l'enseignant à comparer le réalisme de l'étudiant à celui du groupe consiste à déterminer la stratégie optimale qui favorise la régulation des apprentissages au niveau individuel ou au niveau du groupe (voir figure 5). Ces comparaisons font aussi partie du processus de détection des items ambigus et défectueux. Si un item a un vice de construction, l'item doit ainsi être retiré et aucune régulation des apprentissages n'est requise.

Cette recherche montre qu'il y a des liens potentiels entre les propriétés des items et le réalisme exprimé par les étudiants. Des tests destinés à utiliser les DC comme modalité de réponse devraient être calibrés en conséquence en prenant en considération le réalisme exprimé par rapport à un item comme indicateur supplémentaire de qualité des items. L'interprétation critériée des résultats devrait aussi tenir compte des items ayant des faiblesses de manière à mener à des jugements plus valides.

La redevabilité et la responsabilité en matière d'éducation sont devenues des priorités. Des tests sont administrés à grande échelle dans le but d'identifier les lacunes individuelles et les carences dans les institutions d'éducation elles-mêmes. Sur le plan pratique, des tests standardisés, rapides à administrer et efficaces dans ce qu'ils peuvent nous apprendre sont donc nécessaires. De nombreux tests diagnostiques sont utilisés en éducation pour faciliter l'aiguillage des étudiants vers des cours ou des programmes de formation. Les écoles comptent de plus en plus d'étudiants qui ont poursuivi une formation et obtenu une diplomation à l'étranger. En plus des tests certificatifs qui sont nécessaires à la reconnaissance des acquis, les corps professionnels et les institutions d'enseignement doivent entre autres établir le niveau de compétence linguistique et aiguiller les candidats vers des cours spécifiques et des programmes de formation sur mesure. L'élaboration de ces instruments doit tenir compte des situations de connaissances partielles et des facteurs qui peuvent avoir un impact sur leur fidélité et leur validité.

L'étude du réalisme peut également contribuer à jeter un éclairage nouveau sur les performances différentielles de différents groupes d'étudiants. Des étudiants trop optimistes peuvent négliger de réviser leurs réponses ou prendre moins de temps pour remplir un questionnaire d'examen, entraînant des résultats différents. Cette notion est potentiellement prometteuse pour tenter d'expliquer les différences entre garçons et filles à des tests tels que ceux de l'OQRE.

Nous considérons que notre cadre conceptuel présenté à la figure 5 constitue la plus grande contribution originale de cette recherche. Il suggère un processus

décisionnel qui prend en considération le réalisme des étudiants (Rs) sur une base individuelle et en groupe. L'intervention de régulation peut ainsi être guidée en comparant le réalisme exprimé par l'individu avec celui exprimé par le groupe. Il est clair que l'interprétation des résultats d'un étudiant ou d'un groupe doit tenir compte et même faire abstraction des items défectueux. La démonstration de cette relation potentielle entre le coefficient de discrimination et le réalisme constitue probablement l'une des contributions de cette recherche en ce sens qu'elle représente une valeur ajoutée en matière de cadre conceptuel (voir figure 5) et d'analyses des qualités métriques et métrologiques des items dans un contexte diagnostique. Les effets potentiels du coefficient de discrimination des items et leur degré de difficulté sur le réalisme apportent des précisions à l'approche diagnostique en permettant à l'orienteur de faire la part des choses entre ce qui relève de l'item et ce qui relève du sujet. Il peut ainsi orienter son intervention visant à favoriser la régulation des apprentissages tant au niveau individuel que collectif.

LIMITES ET RECOMMANDATIONS

Cette recherche, de par sa nature exploratoire, avance quelques idées originales quant à l'utilisation des DC dans un contexte diagnostique. En utilisant cette approche, nous avons éliminé d'emblé plusieurs difficultés liées à la pondération des items en fonction des DC exprimés. Nous reconnaissons que, dans un contexte d'évaluation certificative, la pondération des items contribue à assurer la pleine participation des étudiants et limite entre autres la quantité de données manquantes. Les réserves portant sur la validité de l'utilisation des DC et la pondération des items exprimées par Lord et Novick (1968), pour ne mentionner que ceux-là, demeurent entières et requièrent encore l'appui de recherches expérimentales.

Le devis utilisé dans cette recherche présente quelques difficultés qui limitent les généralisations que nous pouvons tirer des résultats. D'une part, l'instrument est constitué d'items limités en nombre et qui ne se distribuent pas également parmi les catégories taxonomiques. D'autre part, le coefficient de discrimination moyen des items est particulièrement élevé et la distribution entre les items du coefficient de discrimination est fortement asymétrique. Ces deux conditions limitent l'interprétation des résultats. Des recherches ultérieures pourraient vérifier nos hypothèses en utilisant des séries de données d'envergure obtenues à partir d'instruments qui présentent des caractéristiques souhaitables en termes de niveaux taxonomiques et de coefficients de discrimination des items. La distribution des coefficients de discrimination devrait avoir une étendue plus large et devrait se distribuer normalement. De cette manière, nous pourrions analyser les items ayant un faible coefficient de discrimination sans les

considérer extrêmes. Les items devraient aussi être des cas typiques de chaque niveau taxonomique. Ils devraient être en nombre suffisamment représentatif de chaque niveau pour pouvoir effectuer les analyses désirées. Il serait aussi préférable d'éviter la présence de cas frontières.

Dans un contexte d'évaluation critériée utilisant une mesure fondée sur les objectifs, il serait probablement plus approprié d'examiner l'impact des facettes de résolution d'un problème sur le DC. L'interprétation de l'effet du niveau taxonomique global sur le DC pose certaines difficultés. Nous pourrions éventuellement démontrer un certain impact du niveau taxonomique d'un item sur le DC. Cependant, l'importance de cet impact n'est pas claire dans un contexte d'évaluation critériée. Des recherches ultérieures portant sur le DC pourraient considérer une approche alternative.

Cette recherche met en lumière un effet potentiel des différences individuelles sur le réalisme des étudiants dans l'expression de leur certitude par rapport à leurs réponses. Nous considérons que la différence du niveau de performance des étudiants a un impact différencié sur l'expression du degré de certitude en la réponse choisie et sur le réalisme tel que calculé. La présence d'un lien curvilinéaire entre le niveau de performance et le réalisme exprimé soulève également plusieurs questions par rapport aux résultats de recherches antérieures à la nôtre. Nous constatons que l'impact du sexe sur le réalisme exprimé peut être différent selon le niveau de performance des étudiants de sexe différent. Des études ultérieures sur des séries de données d'envergure permettraient de confirmer et de vérifier cette tendance observée.

Les recherches ultérieures pourraient tenter de valider sur le terrain la qualité de la prise de décision qui découle de l'utilisation de l'algorithme tel que suggéré par notre cadre conceptuel présenté à la figure 5. Il faudrait ainsi mesurer la justesse de l'aiguillage. La pertinence et la qualité des interventions au niveau de la régulation des apprentissages qui découlent de la prise de décision par les orienteurs et le personnel enseignant devraient aussi être évaluées. Cette validation pourrait contribuer à une pratique appropriée de l'évaluation utilisant des instruments faisant usage des DC.

Cette recherche exploratoire contribue plus par les questions qu'elle soulève que par les réponses qu'elle apporte. L'impact potentiel du coefficient de discrimination d'un item sur le réalisme exprimé par rapport à cet item reste à vérifier à partir de séries de données appropriées. Les analyses ne seront possibles qu'avec des résultats à des tests où le coefficient de discrimination des items se distribue normalement. La prise en compte de la discrimination d'un item pourrait contribuer à une interprétation plus valide des résultats lors de l'utilisation des DC.

L'observation d'un lien potentiellement curvilinéaire entre le niveau d'habileté et le réalisme en fonction du sexe n'était pas attendue. Nous considérons que la nature de ce lien potentiel doit être réexaminée. Les recherches et les analyses qui portent sur l'impact du sexe sur le réalisme et sur la surestimation postulent une relation linéaire entre ces variables. Nous considérons que la linéarité de ces relations doit être testée lors d'analyses mettant ces variables en relation.

RÉFÉRENCES

- Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. et Wittrock, M.C. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.
- Auger, N., Dalley, P. et Roy, S. (2005). La sociolinguistique du changement : Le cas du bilinguisme stéréotypé en classe de français langue seconde et en milieux minoritaires. Acte de colloque au Colloque : Le stéréotype, Montpellier, juin 2005.
- Baumeister, R.F., Heatherton, T.F., Tice, D.M. (1994). Losing Control: How and Why People Fail at Self-Regulation. San Diego: Academic Press.
- Beyer, S. (2002). The effects of gender, dysphoria, and performance feedback on the accuracy of self-evaluations. Sex Roles, 47, (9-10), 453-464.
- Bressoux, P. et Pansu, P. (2003). Quand les enseignants jugent leurs élèves. France : PUF.
- Commissions des communautés européennes. (1991). Principes communs pour évaluer les résultats cognitifs de la formation. Bruxelles : Author.
- Carver, C.S. et Scheier, M.F. (2001). Optimism, Pessimism, and Self-Regulation: In Chang, E.C. (Eds.), Optimism et Pessimism: Implications for theory, research, and practice: Washington: APA.
- Choppin, B. (1975) Guessing the answer on objective tests. British Journal of Educational Psychology, 45, 206-213.

- Clack, G. B., et Head, J. O. (1999). Gender differences in medical graduates' assessment of their personal attributes. Medical Education, 33, (2), 101-105.
- Crocker, L. et Algina, J. (1986). Introduction to classical et modern test theory. Florida: Harcourt Brace Jonanovich.
- De Landsheere, G. (1979). Dictionnaire de l'évaluation et de la recherche en éducation. Paris : PUF.
- Descartes, R. (1637). Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences. In Bridoux, A. (Eds.), Descartes : Œuvres et lettres, Bibliothèque de la Pléiade. Paris : Gallimard.
- De Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item. British Journal of Mathematical and Statistical Psychologie, 13, 87-123.
- Ebel, R.E. (1965). Measuring educational acheivement. Anglet Cliffe, N.J.: Prentice-Hall.
- Echternacht, G. J. (1972). The use of Confidence Testing in Objective Tests. Review of Educational Research, 42, (2), 217-236.
- Fabre, J.-M. (1980). Jugement et certitude. Berne-Francfort : Lang.
- Gilles, J.-L. (2002). Qualité spectrale des tests standardisés universitaires. Thèse de doctorat non publiée, Université de Liège : Belgique.
- Henmon, V.A.C. (1911). The relation of the time of a judgement to its accuracy. Psychological Review, 18, 186-201.

- Hunt (1993). Validity, Reliability and Acuity of Self-Assessment in Educational Testing, In Leclercq, D., et Bruno, J. (Eds.), Item Banking: Interactive Testing and Self-Assessment, NATO ASI Series. Heidelberg: Springer Verlag, 113-131.
- Jans, V. et Leclercq, D. (1997). Metacognitive Réalisme: A Cognitive Style or a Learning Stratège?, Educational Psychologie, 17, (1-2), 101-110.
- Jacobs, S.S. (1974). Behavior on Objective Tests Under Theoretically Adequate, Inadequate and Unspecified Scoring Rules. ERIC document. Accession Number : ED090312.
- Jonsson, A., et Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. Personality and Individual Differences, 34 (4), 559-574.
- Kleitman, S. et Stankov, L. (2001). Ecological and Person-Oriented Aspects of Metacognitive Processes in Test-Taking. Applied Cognitive Psychology, 15, 321-341.
- Koivula, N., Hassmén, P. et Hunt, D. P. (2001). Performance on the swedish scholastic aptitude test: Effects of self-assessment and gender. Sex Roles, 44, (11-12), 629-645.
- Laveault, D. et Grégoire, J. (2002). Introduction aux théories des tests en psychologie et en sciences de l'éducation. Bruxelles : De Boeck Université.

Laveault, D. (2004). L'évaluation des apprentissages, un jeu de serpents et échelles.

Conférence de clôture du 24e colloque annuel de l'Association québécoise de pédagogie collégiale. St-Hyacinthe : 10 juin 2004.

Laveault, D. (2007). De la régulation au réglage : étude des dispositifs d'évaluation

favorisant l'autorégulation des apprentissages. In L. Allal et L. Mottier Lopez

(Eds.), Régulations des apprentissages en situations scolaire et en formation.

Bruxelles: De Boeck (pages 207-234).

Leclercq, D. (1993). Validity, Reliability and Acuity of Self-Assessment in Educational

Testing, In Leclercq, D., et Bruno, J. (Eds.), Item Banking: Interactive Testing and

Self-Assessment, NATO ASI Series. Heidelberg: Springer Verlag, 113-131.

Leclercq, D. et Poumay, M. (2005). Degrés de certitude : Épistémologie, méthodes et

conséquences. 18^e Colloque International de l'ADMÉE-Europe 2005, Reims.

Lord, F.M. et Novick, M.R. (1968). Statistical theories of mental test scores. Reading,

Mass. :Addison-Wesley.

Nelson, T.O. et Narens, L. (1994). Why investigate Metacognition?, In J. Metcalfe et

A.P. Shimamura (Eds.), Metacognition: knowing about knowing: Cambridge: MA,

London: A Bradford Book, MIT Press.

OQRE (1999). Étude numéro 4, Toronto : Author.

Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive

tasks. Sex Roles, 48, (5-6), 265-276.

- Pintrich, P.R. (2002). The Role of Metacognitive Knowledge in Learning, Teaching, and Assessing. Theory into Practice, 41, (4), 219-225.
- Reach, G., Zerrouki, A., Leclercq, D., et d'Ivernois, J.-F. (2005). Adjusting insulin doses: from knowledge to decision. Patient Education and Counselling, 56, 98-103.
- Russell, B. (1872-1970) Citation attribuée sans date de référence.
- Scallon, G. (2000). L'évaluation formative. Québec : Éditions du Renouveau Pédagogique.
- Shuford, E., Albert, A. et Massengill, N.E. (1966). Admissible probability measurement procedures. Psychometrika, 31, (2), 125-145.
- Sieber, J.E. (1974). Effects of decision importance on ability to generate warranted subjective uncertainty. Journal of Personality and Social Psychology, 30, 688-694.
- Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. American Journal of Psychology, 18, 161-169.
- Stankov, L. (1998). Calibration curves, scatterplots and the distinction between general knowledge and perceptual tasks. Learning and Individual Differences, 10, (1), 29-50.
- Stiggins, R., Chappuis, S. (2005). Putting testing in perspective: It's for learning. Principal Leadership, 6, (2), 16-20.
- Swineford, F. (1938). Measurement of a personality trait. Journal of Educational Psychology, 29, 295-300.

Tourneur, Y., Duquesne, F. et Ledoux, I. (1986). Étude de deux paramètres déterminant l'efficacité pédagogique du feedback : la qualité de la réponse et la confiance du sujet dans sa réponse. Éducation et recherche, 8, 7-24.

Twain, M. (1835-1910) Citation attribuée sans date de référence.

ANNEXE 1

Formulaire papier n°1 : 1

Évaluation diagnostique mathématiques



Conseil des écoles publiques de l'Est de
l'Ontario



Éducation permanente

Évaluation diagnostique mathématiques

De la 9^{ème} à 12^{ème} année

1. Date:

2. Nom:

3. Prénom:

4. Sexe:

5. Numéro d'identification:

Consignes pour l'élève:

- Cette évaluation n'est pas un test ordinaire. Elle nous aidera à mesurer tes connaissances en mathématiques acquises jusqu'à maintenant dans le but de t'offrir le meilleur cours possible pour ton avenir.
- Il n'y a pas de limite de temps. Relaxe et essaie de répondre au mieux de tes connaissances. Plus tu avanceras, plus les questions seront difficiles. C'est normal, le test est construit comme cela !
- Tu auras 40 questions avec des choix de réponses. Tu dois noircir sur ta feuille réponse une seule lettre, celle que tu crois être la bonne. Tu peux faire tes calculs sur une feuille brouillon.
- Tu as droit à la calculatrice.

Bonne chance et à plus tard !!

Formulaire papier n°1 : 2

Évaluation diagnostique mathématiques

Tu vas remarquer que, pour chaque question du test, il va falloir que tu dises si tu es certain de ta réponse. Ça ne change rien à ton résultat mais ça peut nous aider à mieux voir où tu es rendu. Il est important que tu répondes à toutes les questions et que tu indiques si tu es certain pour chacune des questions.

6. Dispose les nombres ci-dessous en ordre croissant:
0,7 - 2/3 - 69%

- a) 0,7 - 2/3 - 69%
 b) 2/3 - 0,7 - 69%
 c) 2/3 - 69% - 0,7
 d) 0,7 - 69% - 2/3

Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

7. Simplifie l'expression suivante:

- a) $3 + x$
 b) $3x$
 c) $7x + 3$
 d) $-2x^2 + 4x + 3$

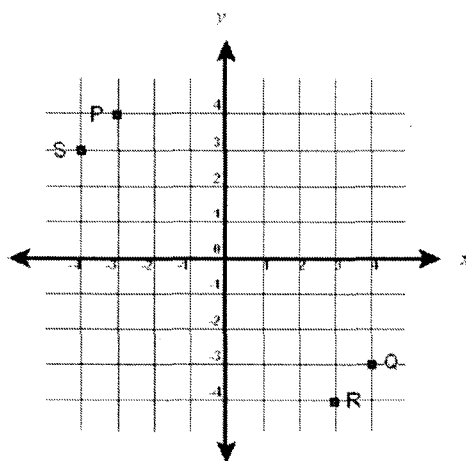
$$|x^2 + 4x + 3| + x|3 - x|$$

Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

8. Quelle lettre correspond au point de coordonnées (4, -3) représenté dans le graphique ci-dessous?

- a) P
 b) Q
 c) R
 d) S



Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

Formulaire papier n°1 : 3

Évaluation diagnostique mathématiques

9. Calcule la pente de la droite passant par les points (2, -3) et (-4, 9) si la formule de la pente est:

- a) $m = 2$
 b) $m = 1/2$
 c) $m = -2$
 d) $m = -1/2$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

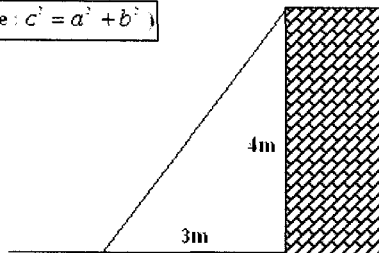
10. Une échelle est placée contre un mur de 4 m de hauteur. Le pied de l'échelle est à 3 m du mur. Calcule la longueur de l'échelle nécessaire pour rejoindre le haut du mur.

- a) 4,7 m
 b) 5 m
 c) 5,4 m
 d) 25 m

Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

(Indice : $c^2 = a^2 + b^2$)

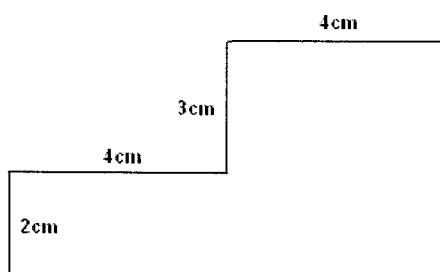


Formulaire papier n°1 : 4

Évaluation diagnostique mathématiques

11. Trouve l'aire de la figure ci-contre si la formule de l'aire d'un rectangle est $A = \text{Longueur} \times \text{largeur}$.

- a) 13 cm^2
- b) 20 cm^2
- c) 26 cm^2
- d) 28 cm^2



Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

12. Le coût total, C , en dollars, pour mettre une annonce dans le journal comprend un frais initial de 12\$, plus 5\$ pour chaque jour d'impression, n . Quelle est l'équation qui représente cette relation?

- a) $C = 12 + 5n$
- b) $C = 12n + 5$
- c) $C = 12n + 5n$
- d) $C = 12 + 5$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

Évaluation diagnostique mathématiques

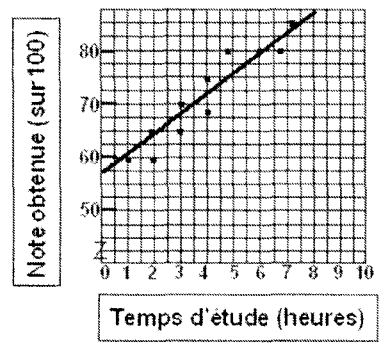
13. On a effectué un sondage auprès de 19 élèves pour connaître le temps qu'ils ont étudié pour un test et la note qu'ils ont obtenue sur 100. On a représenté les résultats obtenus à l'aide d'un nuage de points pour ensuite tracer la droite la mieux ajustée.

- a) 80 %
- b) 85 %
- c) 90 %
- d) 95 %

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

Si un élève a étudié six heures, quelle note a-t-il le plus de chances d'obtenir ?



14. En respectant l'ordre des opérations (PEDMAS), trouve la valeur de cette expression.

- a) 4
- b) 14
- c) 18
- d) 0

$$6 - 2(7 - 3) - 8 + 20 \div 2$$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

Évaluation diagnostique mathématiques

15. Quelle est la valeur de l'expression suivante?

- a) -36
- b) -12
- c) 12
- d) 36

$$2xy \text{ si } x = 2 \text{ et } y = -3$$

Tu es certain(e) de ta réponse à :

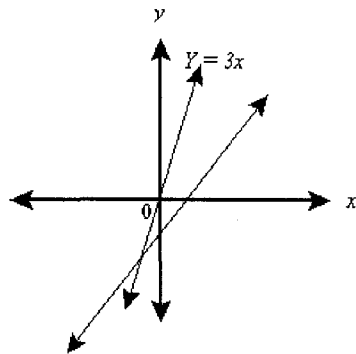
- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

16. Jacques a tracé la droite $y = 3x$. Il a tracé le croquis d'une deuxième droite, mais il n'a pas identifié cette dernière au moyen d'une équation. Par quelle équation pourrait-on identifier le mieux cette deuxième droite?

- a) $y = 2x + 5$
- b) $y = 2x - 3$
- c) $y = -3x + 4$
- d) $y = -2x - 6$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr



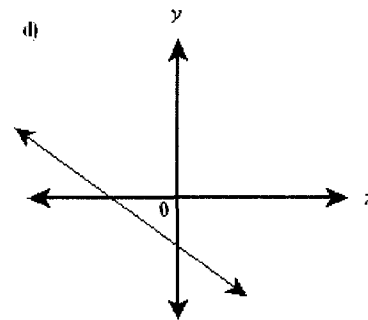
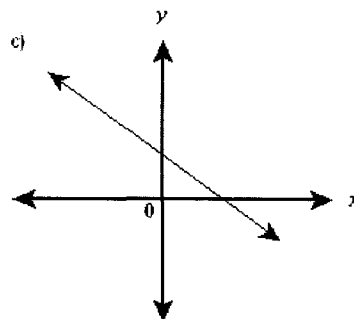
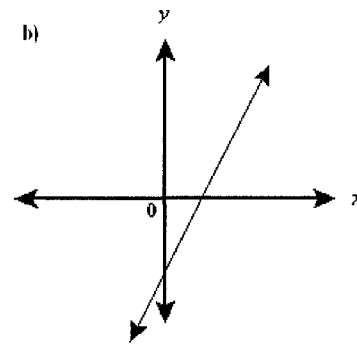
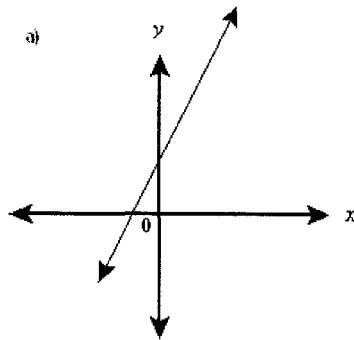
Évaluation diagnostique mathématiques

17. Étant donné l'équation selon le modèle $y = mx + b$ où m représente la pente et b l'ordonnée à l'origine, lequel de ces graphiques représente le plus fidèlement l'équation $y = 2x + 4$?

- a)
 b)
 c)
 d)

Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

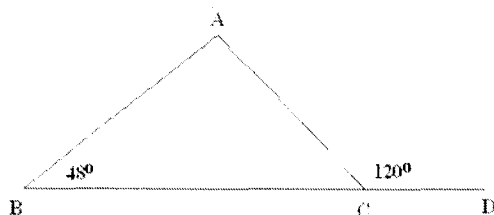


Formulaire papier n°1 : 8

Évaluation diagnostique mathématiques

18. Dans la figure suivante, la mesure de l'angle ABC est 48 degrés et celle de l'angle ACD est 120 degrés. Indique la mesure de l'angle BAC en degrés.

- a) 48°
 b) 60°
 c) 66°
 d) 72°

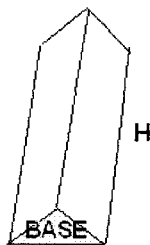


Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

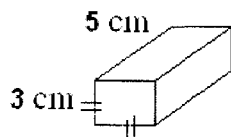
19. Pour calculer le volume d'un prisme, la formule est: Aire de la base x la hauteur. Pour la figure suivante, trouve le volume.

- a) 78 cm^2
 b) 60 cm^2
 c) 45 cm^2
 d) 11 cm^2



Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr



Évaluation diagnostique mathématiques

20. Le coût C , en dollars, pour un buffet dans un hôtel est représenté par l'équation $C = 20n + 500$, où n représente le nombre de personnes invitées. Combien coûte un banquet pour 450 personnes ?

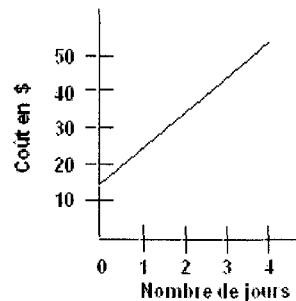
- a) 9 000\$
 b) 9 500\$
 c) 19 000\$
 d) 20 500\$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

21. Pierre veut passer une annonce dans le journal. Voici le graphique qui représente la relation entre le coût de la publication de son annonce en fonction du nombre de jours de parution: Combien coûterait-il pour faire passer l'annonce pendant 8 jours ?

- a) 85 \$
 b) 95 \$
 c) 100 \$
 d) 110 \$



Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

22. Trouve la valeur de x dans l'expression suivante:

- a) $x = -3$
 b) $x = 8$
 c) $x = 12$
 d) $x = 24$

$$\frac{8}{12} = \frac{x}{36}$$

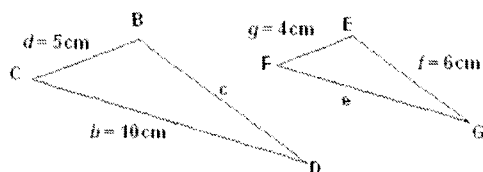
Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

Évaluation diagnostique mathématiques

23. Voici deux triangles semblables: Détermine la valeur de e et de c.

- a) $e = 10 \text{ cm}$, $c = 6 \text{ cm}$
 b) $e = 9 \text{ cm}$, $c = 8 \text{ cm}$
 c) $e = 8,5 \text{ cm}$, $c = 8 \text{ cm}$
 d) $e = 8 \text{ cm}$, $c = 7,5 \text{ cm}$



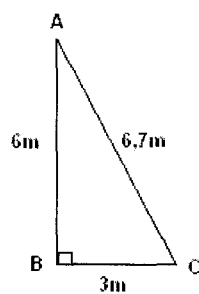
Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

24. Dans le triangle rectangle suivant, quelle est la valeur de $\sin A$?

- a) $\sin A = 3/6$
 b) $\sin A = 3/6,7$
 c) $\sin A = 6/6,7$
 d) $\sin A = 6,7/3$

Rappel : $\sin \theta = \frac{\text{mesure du côté opposé à l'angle } \theta}{\text{mesure de l'hypoténuse}}$



Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

Évaluation diagnostique mathématiques

25. Quelle équation permet d'isoler complètement la variable y pour ?

- a) $-2y = 4 - 2x$
- b) $2y = -2x + 4$
- c) $y = -x + 2$
- d) $y = x - 2$

$$2x - 2y = 4$$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

26. Dans la formule ci-dessous, v est la vitesse en mots par min., m le nombre de mots, e le nombre d'erreurs et t le temps en min. Une personne tape 72 mots par min. Elle tape 500 mots en 5 min. et fait quelques erreurs. Combien d'erreurs a-t-elle faites ?

- a) 25
- b) 18
- c) 14
- d) 10

$$v = \frac{m - 10e}{t}$$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

27. Quel point permet de satisfaire les deux équations suivantes ?

- a) (0, 4)
- b) (2, 2)
- c) (0, -4)
- d) (-2, -2)

$$y = 2x + 4 \quad \text{et} \quad y = -2x + 4$$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

28. Mets en facteur ou factorise complètement l'expression suivante:

- a) $2(3a + 9b)$
- b) $3(2a + 6b)$
- c) $6(a + 3b)$
- d) $6(a - 3b)$

$$6a + 18b$$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

Formulaire papier n°1 : 12

Évaluation diagnostique mathématiques

29. Quelle est la forme développée de l'équation suivante?

- a) $y = x^2 - 4x + 7$
 b) $y = x^2 + 4x + 7$
 c) $y = 2x^2 - 4x - 1$
 d) $y = 2x^2 - 8x + 11$

$$y = 2x^2 - 2x + 3$$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

30. Résous le système d'équations suivant à l'aide de la méthode de ton choix. (Graphique, substitution ou élimination)

- a) (0, 3)
 b) (5, -2)
 c) (-10, 28)
 d) (10, -12)

$$2x + y = 8 \quad \text{et} \quad 3x + 2y = 6$$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

31. Le diamètre d'un cercle relie les points de coordonnées C (-7, -4) et D (-1, 10). Quelles sont les coordonnées du centre du cercle?

- a) (-4, 3)
 b) (-3, 2)
 c) (-4, 4)
 d) (-5, 2)

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

32. Choisis la bonne formule qui permet de calculer la distance entre les points A et B si les coordonnées du point A sont (-2, 1) et du point B sont (2, -2).

- a) $d = \sqrt{(-2+2)^2 + (1-2)^2} = 1$
 b) $d = \sqrt{(-2-2)^2 - (1+2)^2} = \sqrt{7}$
 c) $d = \sqrt{(-2-2)^2 + (1+2)^2} = 5$
 d) $d = \sqrt{(-2+2)^2 - (1-2)^2} = \sqrt{-1}$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

Formulaire papier n°1 : 13

Évaluation diagnostique mathématiques

33. Quelle est la propriété pour que deux triangles soient qualifiés d'homologues (correspondants) en mathématiques?

- a) Les angles correspondants sont exactement égaux dans chacun des triangles.
- b) Les côtés correspondants mesurent la même chose dans les deux triangles.
- c) Les rapports entre les côtés correspondants et les angles correspondants sont les mêmes.
- d) Les trois angles ont une somme de 180 degrés dans chacun des triangles.

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

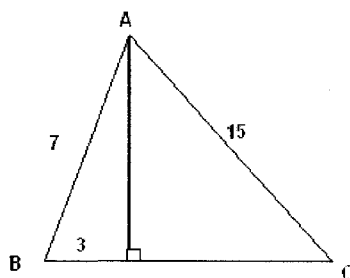
34. Dans la figure ci-contre, trouve au degré près la mesure de l'angle B.

- a) mesure de $\angle B = 23,0^\circ$
- b) mesure de $\angle B = 25,0^\circ$
- c) mesure de $\angle B = 63,0^\circ$
- d) mesure de $\angle B = 65,0^\circ$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

Rappel : $\sin \theta = \frac{c.\text{opposé}}{\text{hypoténuse}}$, $\cos \theta = \frac{c.\text{adjacent}}{\text{hypoténuse}}$, $\tan \theta = \frac{c.\text{opposé}}{c.\text{adjacent}}$



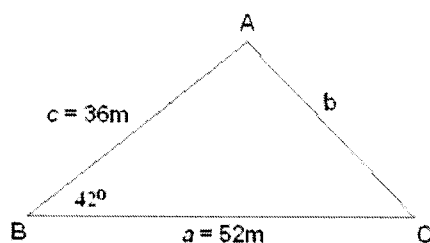
Formulaire papier n°1 : 14

Évaluation diagnostique mathématiques

35. Soit le triangle acutangle ABC, où la mesure de l'angle $\angle B = 42,0^\circ$
 $a = 52$ m et $c = 36$ m. Quel serait la bonne façon de calculer le côté b
 d'après la formule ci-jointe:

- a) $b^2 = 52^2 + 36^2 - 2 \times 52 + 36 \cos 42^\circ$
 b) $b^2 = 52^2 + 42^2 - 2 + 52 + 42 \cos 36^\circ$
 c) $b^2 = 52^2 + 36^2 - 52 \times 36 \cos 42^\circ$
 d) $b^2 = 52^2 + 42^2 - 2 \times 52 \times 42 \cos 36^\circ$

Loi du cosinus : $c^2 = a^2 + b^2 - 2ab \cos C$

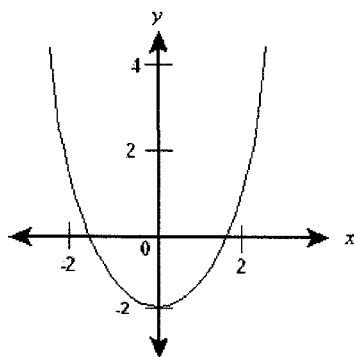


Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

36. Quelle est l'image de la fonction représentée
 ci-dessous ?

- a) $\{y \geq -2, y \in \mathbb{R}\}$
 b) $\{y \leq -2, x \in \mathbb{R}\}$
 c) $\{y > -2, y \in \mathbb{R}\}$
 d) $\{x \in \mathbb{R}\}$



Tu es certain(e) de ta réponse à :

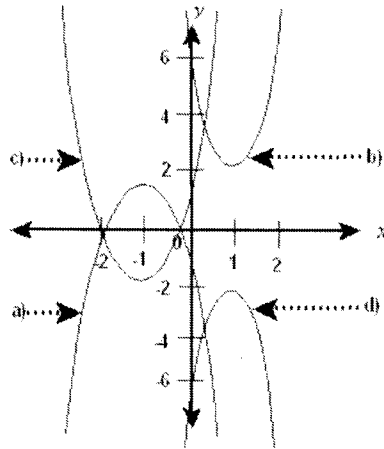
- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

Évaluation diagnostique mathématiques

37. Quelle est la bonne représentation graphique de l'équation suivante?

- a)
- b)
- c)
- d)

$$y = 3(x - 1)^2 + 2$$



Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

38. Laquelle de ces fonctions représente une fonction exponentielle?

- a) $y = mx + b$
- b) $y = x^2 + 3x + 4$
- c) $y = 6$
- d) $y = 2^x$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

39. Laquelle de ces fonctions a la croissance la plus rapide selon toi?

- a) $y = 5^x$
- b) $y = 10^x$
- c) $y = (1/3)^x$
- d) $y = (0,2)^x$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
- 10 % - pas sûr
- 25 % - un peu sûr
- 50 % - ne sais pas
- 75 % - plutôt sûr
- 90 % - sûr
- 98 % - complètement sûr

Évaluation diagnostique mathématiques

40. Soit les 5 premiers nombres de la suite de nombres suivante: (4, 11, 18, 25, 32). Prédit le 8e nombre.

- a) 46
 b) 53
 c) 55
 d) 60

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

41. À l'aide de la formule suivante, calcule ce que rapportera l'investissement de Karim s'il dépose à la banque un montant de 1000\$ à 12% d'intérêt par an capitalisé mensuellement pendant trois ans.

- a) 59 135,57\$
 b) 1 030,30\$
 c) 1 430,77\$
 d) 1 404,90\$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

Voici la formule de l'intérêt composé : $M = C(1 + i)^n$
 où M représente le montant accumulé, C le capital investi;
 i l'intérêt par période et n le nombre de périodes d'intérêt.

42. Trouve la valeur de x dans l'équation exponentielle suivante:

- a) 6
 b) -6
 c) 5
 d) -5

$$2^x = 64$$

Tu es certain(e) de ta réponse à :

- 2 % - pas sûr du tout
 10 % - pas sûr
 25 % - un peu sûr
 50 % - ne sais pas
 75 % - plutôt sûr
 90 % - sûr
 98 % - complètement sûr

Évaluation diagnostique mathématiques

43. Simplifie l'expression suivante:

- a)
 b)
 c)
 d)

$$\frac{2ab}{5c} + \frac{14a^2b^2}{15c^2}$$

- a) $\frac{28a^3b^3}{75c^3}$ c) $\frac{3c}{7ab}$
 b) $\frac{75a^3b^3}{3c^3}$ d) $\frac{7ab}{3c}$

Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

44. Représente l'expression trigonométrique suivante à l'aide d'un seul terme.

- a) $\sin x$
 b) $\cos x$
 c) $\tan x$
 d) $\cos^2 x$

$$\frac{1}{\cos x} - \sin x \tan x$$

Prends note que : $\tan x = \frac{\sin x}{\cos x}$ et $\sin^2 x + \cos^2 x = 1$

Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

Évaluation diagnostique mathématiques

45. Quelle est la période et l'amplitude de la fonction sinusoidale suivante?

- a) Période = 6π , Amplitude = -4
 b) Période = 3π , Amplitude = 4
 c) Période = 3π , Amplitude = -4
 d) Période = 6π , Amplitude = 4

Tu es certain(e) de ta réponse à :

- 2% - pas sûr du tout
 10% - pas sûr
 25% - un peu sûr
 50% - ne sais pas
 75% - plutôt sûr
 90% - sûr
 98% - complètement sûr

