

2m11. 2529.10

Université de Montréal

**DÉVELOPPEMENT D'UNE STRATÉGIE DE TESTING INFORMATISÉE
ET INTERACTIVE, CENTRÉE SUR LE DIAGNOSTIC PÉDAGOGIQUE
DANS LE CONTEXTE DE LA CLASSE.**

par

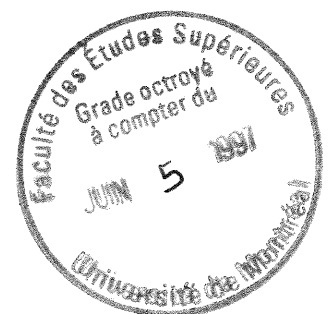
Michèle Perron

**Département d'études en éducation et
d'administration de l'éducation
Faculté des sciences de l'éducation**

**Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès arts (M.A.)
en mesure et évaluation en éducation**

Avril 1997

© Michèle Perron, 1997



LB
5
U57
1997
V.031

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

**DÉVELOPPEMENT D'UNE STRATÉGIE DE TESTING
INFORMATISÉE ET INTERACTIVE, CENTRÉE SUR LE DIAGNOSTIC
PÉDAGOGIQUE DANS LE CONTEXTE DE LA CLASSE.**

Présenté par

Michèle Perron

A été évalué par un jury composé des personnes suivantes :

Michel Laurier président rapporteur

Clement Jassa directeur de recherche

Djavid Azar membre du jury

Mémoire accepté le *06.05.97*

SOMMAIRE

Ce mémoire présente le **développement d'une stratégie de testing informatisée et interactive, centrée sur le diagnostic pédagogique dans le contexte de la classe**. La nouvelle stratégie d'administration d'un test a été intégrée à un système de tests adaptatifs existant, conçu pour répondre à un besoin d'outils pour l'évaluation formative qui respectent les contraintes organisationnelles du milieu scolaire québécois. Le système développé a pour but de renseigner rapidement l'enseignant et l'élève de l'état des apprentissages de celui-ci sans délai, afin que l'enseignant puisse entreprendre des interventions correctives efficaces auprès des élèves.

Le système de tests réalisé administre un test en deux phases. La première phase fournit un diagnostic structurel ainsi qu'une hypothèse du diagnostic causal reflétant les apprentissages d'un élève. La seconde phase du test permet d'établir un niveau de cohérence d'un élève présentant une difficulté cognitive préalablement identifiée dans la première phase et de communiquer ce diagnostic causal sous la forme d'un message qualitatif pour l'enseignant.

Pour répondre aux objectifs de la première phase du test, nous avons modifié la programmation de la stratégie "stradaptative" existante pour la rendre plus flexible et indépendante des banques d'items. Nous avons procédé à la vérification de l'algorithme de sélection des items "stradaptatif" par une étude de vérification simulant différents profils d'élèves.

Pour répondre aux objectifs de la seconde phase, nous avons convenu d'utiliser la catégorie de leurres la plus fréquemment choisie par l'élève lors de la première phase du test comme élément d'information pour la seconde phase. Nous avons ajouté au système de tests diagnostiques existant, une stratégie de sélection adaptative des items selon un modèle Rasch nécessaire pour estimer le niveau de difficulté cognitive de l'élève face à une catégorie de leurres donnée.

Nous avons établi, à partir des résultats des études de simulations, des intervalles de confiance pour les θ moyens estimés correspondant à des niveaux de difficulté cognitive et des profils d'élèves différents. À partir de ces intervalles, nous avons défini trois niveaux de messages diagnostiques exprimant le niveau de cohérence de l'élève à choisir une catégorie de leurres donnée : la difficulté n'est pas cohérente, la difficulté présente un certain degré de cohérence, la difficulté est cohérente.

Le diagnostic structurel et le diagnostic causal sont accessibles immédiatement après la passation du test par l'entremise de la fiche diagnostique individuelle de l'élève.

Cette recherche a permis d'intégrer dans un même système de tests, la stratégie "stradaptative" dont la programmation a été rendue indépendante du contenu des banques et une stratégie complémentaire de sélection des items utilisant le modèle de Rasch qui permet de déterminer un niveau de cohérence de l'élève à choisir une certaine catégorie de leurres.

Mots clés : Évaluation formative informatisée, mesure diagnostique des apprentissages, stratégie de tests adaptatifs, modèle de Rasch, mesure de difficulté cognitive.

TABLE DES MATIÈRES

SOMMAIRE	iii
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX	viii
LISTE DES FIGURES	x
REMERCIEMENTS	xii
INTRODUCTION	1
Chapitre premier : LA PROBLÉMATIQUE GÉNÉRALE	4
1.1 L'évaluation formative	5
1.1.1 L'origine de l'évaluation formative	5
1.1.2 Une définition de l'évaluation formative	6
1.1.3 La pédagogie de maîtrise	9
1.1.4 Le contexte pédagogique au Québec	11
1.1.5 L'évaluation formative à fonction diagnostique	12
1.2 Les pratiques d'évaluation	17
1.2.1 Les problèmes reliés aux pratiques d'évaluation	17
1.2.2 Les besoins d'outils pédagogiques adéquats pour l'évaluation	20
1.3 Le testing informatique	22
1.3.1 Le développement des outils informatiques	22
1.3.2 Les courants de développement du testing informatique	23
1.3.3 Des considérations liées au problème de recherche ..	31
Chapitre second : LE PROBLÈME DE RECHERCHE	34

Chapitre troisième : LA MÉTHODOLOGIE	58
3.1 La première phase du test	59
3.1.1 Le choix de la banque d'items	59
3.1.2 L'algorithme "stradaptatif"	63
3.1.3 Les contraintes et limites	68
3.1.4 La vérification de la phase I du test	69
3.1.5 La communication des résultats de la phase I du test	70
3.2 La seconde phase du test	71
3.2.1 La constitution des banques d'items spécifiques	
à une catégorie de leurres donnée	71
3.2.2 La calibration des items	72
3.2.3 La stratégie adaptative de sélection des items	74
3.2.4 L'étude de simulations de la phase II du test	78
3.2.5 La communication des résultats de la phase II du test	81
3.3 Le système de tests diagnostiques intégrant les phases I et II	82
3.3.1 Le fonctionnement général du système de tests	82
3.3.2 L'information diagnostique produite	82
3.3.3 Les contraintes du système de tests diagnostiques . .	83
Chapitre quatrième : LES RÉSULTATS	84
4.1 La première phase du test	85
4.1.1 Les objectifs	85
4.1.2 Les résultats de l'étude de vérification	85
4.1.3 Une interprétation des résultats	97
4.2 La seconde phase du test	99
4.2.1 Les objectifs	99
4.2.2 Les résultats de l'étude de simulations	99
4.2.3 Une interprétation des résultats	106
4.3 Les limites du système développé	109
CONCLUSION ET DISCUSSIONS	111

RÉFÉRENCES BIBLIOGRAPHIQUES	119
LES ANNEXES	127
A. Liste des objectifs	128
B. Liste des catégories de leurres par objectif	137
C. Figures 2,3,4 et 5 tirées du document Lenman (1994)	149
D. Distributions des indices de difficulté générés	153
E. Liste des items et de leur indice de difficulté	157
F. Compilation de l'étude de vérification de l'algorithme "stradaptatif" de la PHASE I du test	159
G. Fiches diagnostiques individuelles à la PHASE I	160
H. Profils des élèves à la PHASE II	169

LISTE DES TABLEAUX

Tableau I:	Cube diagnostique pour l'objectif de géométrie 5.5.2	38
Tableau II:	Répartition des items selon les objectifs et la modalité d'administration des tests informatisés	40
Tableau III:	Fiche de diagnostic individuel résultant de l'administration du test correspondant à l'objectif de géométrie 5.5.2 . .	41
Tableau IV:	Fiches de diagnostic collectif résultant de l'administration du test correspondant à l'objectif de géométrie 5.5.2 . .	43
Tableau V:	Navigation ad hoc selon la stratégie "stradaptative" pour le test de modalité "laboratoire" de l'objectif géométrie 5.5.2	49
Tableau VI:	Répartition des catégories de leurres selon l'objectif visé	53
Tableau VII:	Liste des catégories de leurres pour le test correspondant à l'objectif de géométrie 5.5.2	54
Tableau VIII:	Répartition des catégories de leurres pour l'objectif de géométrie 5.5.2 dans la banque d'items et les tests de modalités "laboratoire" et "récupération"	55
Tableau IX:	Regroupement des dimensions majeures de l'objectif de géométrie 5.5.2	62

Tableau X:	Distribution des items parmi les banques spécifiques aux catégories de leurres retenues	72
Tableau XI:	Distribution et description des catégories de leurres admissibles à la PHASE II du test	80
Tableau XII:	Parcours de l'élève no.1 de la catégorie "Maîtrise parfaite"	91
Tableau XIII:	Parcours de l'élève no.1 de la catégorie "Maîtrise élevée"	92
Tableau XIV:	Parcours de l'élève no.1 de la catégorie "Maîtrise"	93
Tableau XV:	Parcours de l'élève no.1 de la catégorie "Absence de maîtrise" avec un score d'environ 50%	94
Tableau XVI:	Parcours de l'élève no.1 de la catégorie "Absence de maîtrise" avec un score d'environ 25%	95
Tableau XVII:	Parcours de l'élève no.1 de la catégorie "Absence de maîtrise" avec un score de 0%	96
Tableau XVIII:	Compilation des valeurs moyennes de l'étude de simulations PHASE II pour les banques d'items spécifiques aux catégories de leurres 2 et 8	101
Tableau XIX:	Intervalle de confiance pour le vrai score θ moyen pour chaque catégorie de profils d'élèves à 90%	103
Tableau XX:	Points de césure définis pour chaque banque d'items spécifique à une catégorie de leurres	105

LISTE DES FIGURES

- figure 1 - Processus de régulation des apprentissages intégrés à l'enseignement 13
- figure 2 - Navigation ad hoc selon la stratégie "stradaptative" pour le test de modalité "laboratoire" correspondant à l'objectif de géométrie 5.5.2. 51
- figure 3 - Ordinogramme illustrant la PHASE I du test 64
- figure 4 - Ordinogramme illustrant la PHASE II du test 75
- figure 5 - Intervalle de confiance à 90% pour le vrai score thêta moyen pour les différentes catégories d'élèves avec la banque d'items spécifique à la catégorie de leurres 2 104
- figure 6 - Intervalle de confiance à 90% pour le vrai score thêta moyen pour les différentes catégories d'élèves avec la banque d'items spécifique à la catégorie de leurres 8 104
- figure 7 - Distribution des thêta estimés pour les différentes catégories d'élèves avec la catégorie d'items spécifique à la catégorie de leurres 8 108
- figure 8 - Distribution des thêta estimés pour les différentes catégories d'élèves avec la catégorie d'items spécifique à la catégorie de leurres 2 108

REMERCIEMENTS

Tout d'abord, je tiens à remercier l'équipe de chercheurs composée de messieurs D.Ajar, R.Auger, C.Dassa, S.Ph.Séguin et de mesdames S.René de Cotret et C.Parent qui, dans le cadre d'un projet de recherche subventionné par le CRSH (1988-92) dirigé par monsieur C.Dassa, ont conçu et réalisé le système de tests adaptatifs initial. Les travaux et l'expertise de cette équipe de recherche m'ont servi de point de départ pour ce mémoire.

Je tiens à exprimer ma grande reconnaissance à mon directeur de recherche, monsieur Clément Dassa, professeur au département d'études en éducation et d'administration de l'éducation, pour ses encouragements et ses conseils judicieux tout au cours de cette recherche.

A mes proches, parents et amis, un merci tout spécial, pour leur soutien et leur patience.

INTRODUCTION

INTRODUCTION

Ce mémoire porte sur le développement d'une stratégie de testing informatisée et interactive, centrée sur le diagnostic pédagogique dans le contexte de la classe. Il s'inscrit dans la continuité des travaux de recherche relatifs à un système de tests adaptatifs informatisés existant, conçu pour l'évaluation formative en classe et respectant les contraintes organisationnelles du milieu scolaire québécois. Ce système a pour but de renseigner rapidement et précisément l'enseignant sur l'état des apprentissages des élèves par le biais de fiches diagnostiques afin que celui-ci puisse procéder à des interventions correctives efficaces auprès des élèves en difficulté. Or, certaines limites inhérentes à la programmation du système existant le rendent difficilement modifiable. Notamment, le fait que l'algorithme de sélection des items d'un test donné ne soit pas indépendant de la banque d'items. Dans un contexte pédagogique favorisant l'autonomie de l'enseignant, il devient important de lui permettre de gérer facilement le système en ajoutant, modifiant ou retirant des items de la banque d'items utilisée sans avoir à reprogrammer le système.

Nous avons modifié le système de tests existant de façon à rendre l'algorithme "stradaptatif" de sélection des items indépendant du contenu de la banque d'items. De cette façon, le diagnostic structurel résultant permet de mieux refléter les pratiques pédagogiques de l'enseignant à ce jour. De plus, nous avons ajouté une deuxième phase au système existant qui permet de mesurer précisément la propension de l'élève à commettre certaines erreurs. Pour ce faire, nous avons utilisé la catégorie de leurres, préalablement choisie, comme élément d'information pour la stratégie de sélection des items et nous avons déterminé, à l'aide d'un modèle de Rasch relevant de la TRI (Théorie des réponses aux items), une mesure de difficulté cognitive liée à cette catégorie de leurres. Dans un dernier temps, nous avons établi à partir d'études de simulations, des normes d'utilisation du système proposé. Le diagnostic causal établi à la suite de cette nouvelle phase de test est communiqué à l'enseignant en terme de message diagnostique par le biais d'une fiche individuelle. Cette information permet à

l'enseignant de différencier ses interventions correctives en se fondant sur un profil individuel précis des difficultés cognitives.

Ce mémoire présente, au chapitre premier, la problématique générale de la recherche. Le chapitre second définit le problème de recherche alors que le chapitre troisième aborde la méthodologie suivie pour la réalisation de cette étude. Le quatrième chapitre traite des résultats tandis que le dernier chapitre présente les conclusions tirées de l'étude et aborde quelques pistes d'éventuelles recherches.

Chapitre premier :
LA PROBLÉMATIQUE GÉNÉRALE

Ce premier chapitre présente la problématique général du mémoire intitulé "Développement d'une stratégie de testing informatisée et interactive, centrée sur le diagnostic pédagogique dans le contexte de la classe". Nous examinerons et adopterons dans un premier temps une définition de l'évaluation formative, plus spécifiquement liée à la fonction diagnostique de l'évaluation dans le contexte du système scolaire québécois. Dans un second temps, nous exposerons différentes études traitant des pratiques de l'évaluation en classe pour mieux cerner les problèmes liés à celles-ci. La dernière section de ce chapitre sera consacrée à l'évolution du testing informatisé et aux principales considérations retenues pour le problème de la recherche.

1.1 L'évaluation formative.

Cette section présente un bref historique de l'évaluation formative, propose une définition générale de l'évaluation formative, traite de deux tendances dans l'application de la pédagogie de maîtrise et introduit le contexte pédagogique au Québec et l'application de l'évaluation formative à fonction diagnostique intégrée à l'enseignement.

1.1.1 L'origine de l'évaluation formative

L'évaluation formative émerge du courant de changement et d'ouverture de la société occidentale vis-à-vis la démocratisation de l'éducation au cours des années 50. Ainsi que mentionné dans un extrait de l'ouvrage de Tyler "Changing Concepts of Educational Evaluation" (Dassa *et al.*, 1987), certains facteurs sociaux tels les changements économiques, l'universalisation de l'accès, l'importance donnée à la résolution de problèmes, les nouvelles conceptions de l'apprentissage et les développements technologiques ont contribué à l'élargissement du rôle de l'évaluation en éducation. L'acte d'évaluer n'est plus restreint à des fins de sélection et de classement pour le système scolaire mais sert aussi d'outil de gestion pédagogique en classe.

L'origine de l'expression "évaluation formative" revient à Scriven en 1967. Scriven distingua l'évaluation formative et l'évaluation sommative par leurs rôles, leurs buts et les personnes responsables de leur pratique (Scallon, 1988a) en ce qui a trait aux programmes d'études et aux moyens d'enseignement (Bloom *et al.*, 1981).

Avant l'instauration de l'évaluation formative, l'évaluation des apprentissages n'avait qu'un rôle de sanction, de classement, qui s'exerçait à la fin d'un programme, d'une année scolaire ou d'une période donnée. Cette évaluation dite "sommative" répondait principalement aux exigences du système scolaire.

Des modifications apportées au système éducatif ont entraîné un élargissement du rôle de l'évaluation formative en éducation. Comme le mentionne Scriven, l'évaluation formative a pour but d'améliorer en cours d'élaboration un programme ou des moyens d'enseignement. Par la suite, plusieurs chercheurs, dont Bloom et ses collaborateurs, ont transposé l'évaluation formative définie par Scriven, à l'évaluation de l'enseignement et des apprentissages des élèves en cours de formation (Bloom *et al.*, 1971, 1981; Bloom 1984).

1.1.2 Définition de l'évaluation formative

Fondamentalement, évaluer ou procéder à une évaluation c'est poser un jugement (MEQ, 1981; Tousignant & Morrisette, 1990; Morrisette, 1984). La démarche évaluative fait partie d'un processus qui encadre le couple enseignement-apprentissage; à partir d'une intention, les apprentissages réalisés à la suite de l'enseignement font l'objet d'une prise d'informations et d'un jugement exprimé en vue d'une action pédagogique.

Dans le cadre de cette recherche, nous ne retiendrons que l'aspect de l'évaluation formative orientée vers les apprentissages.

À cet effet, nous retrouvons plusieurs définitions de l'évaluation formative dans la littérature. Pour n'en citer que quelques-unes :

«Formative evaluation is for us the use of systematic evaluation in the process of curriculum construction, teaching and learning for the purpose of improving any of these processes.» (Bloom *et al.*, 1981, p.155)

«l'évaluation formative n'est en fin de compte qu'une des composantes d'un dispositif d'individualisation des apprentissages et de différenciations des interventions et des encadrements pédagogiques.»(Perrenoud, 1993, p.110)

«l'évaluation formative est censée de contribuer à la régulation des apprentissages en cours dans le sens des maîtrises visées.» (Perrenoud, 1986, p.202)

«elle doit être considérée comme une activité pédagogique qui a pour but d'assister les élèves dans leur cheminement personnel, et d'aider le maître et la maîtresse à mieux jouer leur rôle de personnes-ressources auprès de ces derniers et de ces dernières.» (Tousignant & Morissette, 1990, p.193)

«un processus d'évaluation continue ayant pour objet d'assurer la progression de chaque individu dans une démarche d'apprentissage, avec l'intention de modifier la situation d'apprentissage ou le rythme de cette progression, pour apporter (s'il y a lieu) des améliorations ou des correctifs appropriés.» (Scallon, 1988a, p.155)

«mode d'évaluation à fonction diagnostique, où des évaluations effectuées en cours d'apprentissage fondées sur des interprétations tantôt critériées tantôt normatives, informent l'agent et le sujet en regard des objectifs atteints, ou non par ce dernier en vue de lui assurer le renforcement nécessaire, de déceler, s'il y a lieu, ses difficultés et de lui apporter l'aide utile le cas échéant.» (Legendre, 1988, p.260)

«Évaluation intervenant en principe, au terme de chaque tâche d'apprentissage et ayant pour objet d'informer élève et maître du degré de maîtrise atteint et éventuellement de découvrir où et en quoi un élève éprouve des difficultés d'apprentissage, en vue de lui proposer ou de lui faire découvrir des stratégies qui lui permettent de progresser.» (De Landsheere, 1979, p.125)

«Elle (l'évaluation formative) permet donc de déceler où et en quoi l'élève éprouve des difficultés afin de lui suggérer ou de lui faire découvrir des moyens de progresser. (...) est orientée vers une aide pédagogique immédiate auprès de l'élève.» (MEQ, 1981, p.7)

Nous retenons de ces définitions que l'évaluation formative se déroule en cours d'apprentissage, que c'est un processus continu et dynamique, interactif qui a pour but, entre autres, de régulariser les apprentissages par des activités pédagogiques correctives, des interventions pédagogiques ou des modifications dans la planification de l'enseignement, du choix des méthodes utilisées, de l'environnement ou des attitudes afin d'améliorer les activités d'apprentissage.

Pour les fins de cette recherche, nous retenons la définition de l'évaluation formative la plus générale, soit celle de Perrenoud (1986):

«l'évaluation formative est censée de contribuer à la régulation des apprentissages en cours dans le sens des maîtrises visées.» Perrenoud (1986, p.202)

Dans le courant de recherche traitant de l'évaluation formative des apprentissages, notons les travaux de Bloom et de ses collaborateurs (Bloom *et al.*, 1971). À partir d'une taxonomisation de l'apprentissage des divers domaines dont notamment le domaine cognitif (Bloom *et al.*, 1956), Bloom et ses collaborateurs ont développé une approche pédagogique reposant sur la structuration des unités d'apprentissage en objectifs. Ainsi l'évaluation formative des apprentissages, telle que définie par Bloom, se base sur la formulation préalable d'unités d'apprentissage structurées et ordonnées. Leurs travaux ont aussi porté sur l'élaboration d'instruments pour la planification des cours et la

construction de tests, tels que les tableaux de spécifications, en vue d'évaluer les apprentissages.

Les travaux de Bloom constituent une application pédagogique du modèle temporel d'apprentissage de Carroll (Birzée, 1982) qui postule que l'apprentissage est une fonction du temps et qui stipule que tout élève peut apprendre dans une période de temps qui lui est propre. Plusieurs autres stratégies d'application de l'évaluation formative ont été développées à partir du modèle temporel de Carroll. La pédagogie corrective de Bonboir, de même que les diverses formes de la pédagogie de maîtrise sont autant de stratégies s'inspirant de ce modèle.

Étant donné le contexte de notre système éducatif au Québec, nous retiendrons pour cette recherche, la stratégie de la pédagogie de maîtrise dans l'application de l'évaluation formative des apprentissages. Le terme "Mastery Learning" fut d'abord introduit par Bloom en 1968 et vise à "faire réaliser un minimum de performances par un grande majorité des élèves". Le terme équivalent adopté en français "la pédagogie de maîtrise", revient à De Landsheere en 1971 (Birzée, 1982).

1.1.3 La pédagogie de maîtrise

Au cours des années 70 et 80, et jusqu'au début des années 90, la stratégie de la pédagogie de maîtrise a occupé une place importante dans la recherche en évaluation, aussi bien en Amérique qu'en Europe.

Tel que mentionné par Birzée, la stratégie de pédagogie de maîtrise qui intègre l'évaluation formative demande de façon générale (Birzée, 1982, p.24):

- 1) la définition précise de la maîtrise en termes de tâches à accomplir;
- 2) l'évaluation continue des performances après chaque unité d'apprentissage;

3) la correction individuelle des erreurs.

Par conséquent, pour Birzée, trois conditions sont obligatoires pour toutes les stratégies de la pédagogie de maîtrise, soit l'enseignement par objectifs, l'intégration de l'évaluation formative et du traitement correctif dans l'enseignement et la généralisation du critère de maîtrise (Birzée, 1982).

Nous pouvons distinguer deux tendances dans l'application de la stratégie de la pédagogie de maîtrise. L'approche développée aux États-Unis met l'emphase sur la structuration des interventions, "une approche instrumentée fondée sur une formulation spécifique des objectifs d'apprentissage, qui structure les interventions et oriente vers un plus petit nombre d'objectifs un plus grand nombre d'élèves" (Dassa, Vazquez & Ajar 1993). Par contre, l'approche développée par les européens comme Allal, Cardinet, Perrenoud (Perrenoud, 1993) est centrée sur l'individualisation des apprentissages et la différenciation des interventions pédagogiques. Cette dernière vision proposée de la pédagogie de maîtrise est fondée sur des pratiques de l'évaluation des apprentissages en classe qui permettent à l'enseignant d'adapter ses actions pédagogiques en fonction des forces et des faiblesses des apprentissages de chaque élève (Allal, Cardinet, Perrenoud, 1979).

L'approche européenne est plus qualitative, moins instrumentée, moins quantitative et considère les aspects affectifs, environnementaux et sociaux.

Cependant ces deux approches, américaine et européenne, ne sont pas incompatibles mais assujetties à des contraintes, une culture et des besoins pédagogiques différents. Notamment, il existe des différences en ce qui concerne la mesure. L'instrument privilégié par les Américains est le "test" tandis que la forme "examen" ainsi que d'autres moyens informels s'utilisent fréquemment en Europe lors de l'évaluation de la maîtrise des apprentissages (Cardinet, 1988).

Au Québec, plusieurs chercheurs, entre autres Scallon, ont travaillé dans le sens d'une approche mixte alliant "la recherche d'informations qualitatives à des informations quantitatives" (Dassa, Vazquez & Ajar 1993).

1.1.4 Le contexte pédagogique au Québec.

Le Québec suivra le courant de démocratisation de l'éducation et s'engagera dans le mouvement d'optimisation du rendement scolaire (CSE, 1990).

Ainsi, la réforme du système scolaire au Québec de 1979 a amené la redéfinition systématique des programmes d'études au primaire et au secondaire, sous la forme de finalités et objectifs, par l'application des propositions de Tyler. Suite à cette réforme et aux besoins exprimés par le milieu, le Ministère de l'Éducation du Québec (MEQ) a adopté une politique générale d'évaluation pédagogique en 1981. Depuis la mise en application de cette politique, les modes d'évaluation sommatif et formatif coexistent et se complètent. Cependant, le ministère privilégiera, à partir de cette époque, l'évaluation formative orientée vers une aide pédagogique immédiate dans le but de favoriser et encadrer l'application des mesures pédagogiques correctives. L'évaluation sera désormais considérée comme partie intégrante de l'enseignement et des apprentissages (MEQ, 1979, 1981).

Le premier responsable de l'évaluation formative est l'enseignant en classe car il est en "relation directe et continue avec l'élève" (MEQ, 1981). L'enseignant peut suivre le cheminement de l'élève et intervenir immédiatement, tel que prescrit par le MEQ, dans la régulation des apprentissages par des actions correctives.

1.1.5 L'évaluation formative à fonction diagnostique

Comme nous avons vu, le MEQ privilégie la pratique de l'évaluation des apprentissages en classe par l'enseignant. De plus, le MEQ oriente cette pratique vers une approche mixte de la stratégie de maîtrise, c'est-à-dire une stratégie à la fois instrumentée et qualitative. La combinaison des deux approches (américaine et européenne) résulte des diverses contraintes imposées par chacune d'elles.

L'approche instrumentée impose à l'enseignant une planification rigoureuse en ce qui a trait au développement des instruments de mesure qui permettent la cueillette d'informations (Allal, 1988); l'enseignant devra donc posséder une formation adéquate pour entreprendre le développement des instruments de mesure. Des auteurs (Dumoulin, 1991; Burelle *et al.*, 1991a) ont constaté le manque de formation des enseignants relativement à l'élaboration d'instruments de mesure et par conséquent leur obligation d'utiliser des instruments de mesure conçus par d'autres intervenants scolaires tels que les conseillers pédagogiques ou les docimologues. Or, ces instruments ne répondent pas nécessairement aux besoins particuliers de la classe de l'enseignant.

Certains chercheurs (Scallon, 1988a; Perrenoud, 1993) constatent que l'approche qualitative présente des lacunes notamment au chapitre de la pertinence et de la complétude des enseignants envers certains élèves. Qui plus est, la collecte d'informations est rarement effectuée en fonction de visées précises et préalablement établies; il s'ensuit un problème de congruence entre l'information colligée et l'évaluation réalisée.

La combinaison des deux approches, quantitative et qualitative, permet à l'enseignant de choisir les modalités en fonction des contraintes personnelles (formation, croyances), pédagogiques, administratives, organisationnelles et institutionnelles (Burelle *et al.*, 1991b).

Nous pouvons représenter la pratique de l'évaluation formative intégrée à l'enseignement par une séquence de boucles (figure 1) et d'unités d'interactions entre l'apprenant (l'élève) et l'enseignant. Chacune de ces boucles symbolise un cycle de régulation des apprentissages: à une période d'enseignement succède une évaluation des apprentissages puis un ensemble d'interventions correctives si nécessaire et, finalement, un retour à une période d'enseignement portant soit sur les mêmes objectifs soit sur d'autres objectifs. Les interactions entre les intervenants élèves-enseignant jouent un rôle prépondérant dans le processus général de régulation des apprentissages.

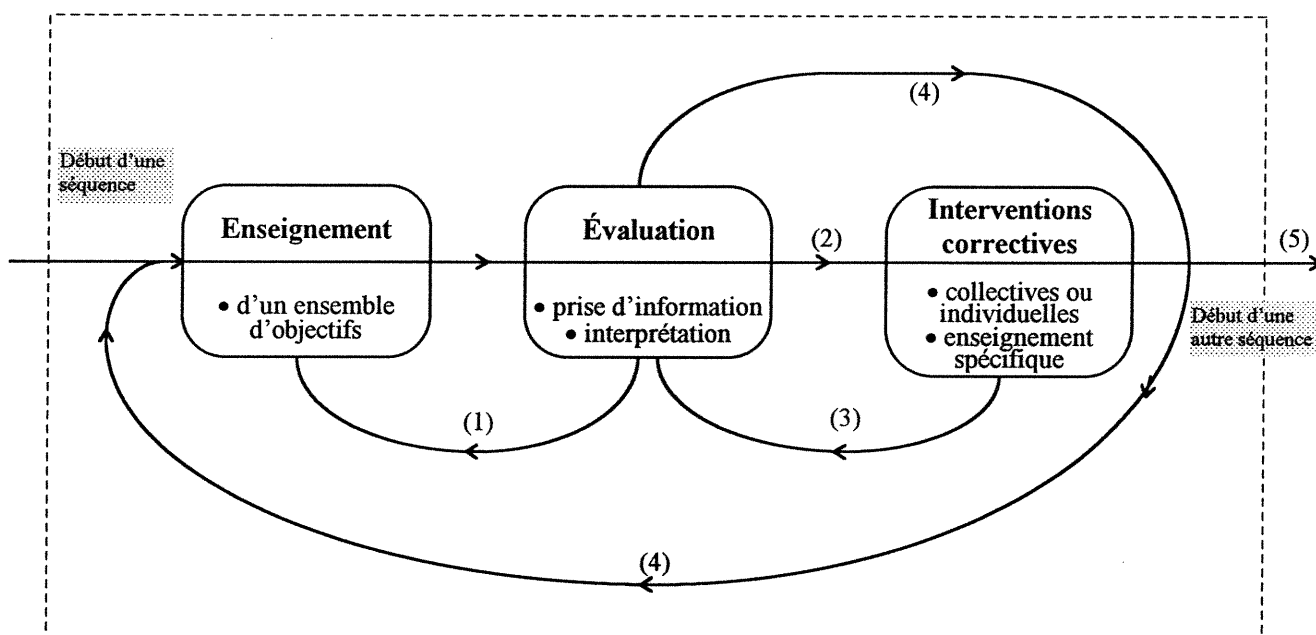


figure 1 - Processus général de régulation des apprentissages intégrés à l'enseignement.

La période d'enseignement porte sur un ou des objectifs planifiés préalablement à l'intérieur de laquelle l'élève acquiert ou approfondit des connaissances, des habiletés selon le niveau des objectifs enseignés. L'évaluation

comprend une prise d'informations et l'interprétation de celles-ci, selon leur nature, par l'enseignant. L'évaluation sert à l'enseignant à planifier le déroulement de ses prochaines interventions et à l'élève à prendre conscience du niveau de ses apprentissages. Les interventions correctives s'effectuent de façon individuelle ou collective auprès des élèves par l'enseignant.

Selon le contexte de la classe, l'enseignant effectuera la reprise de l'enseignement d'un objectif ou d'un ensemble d'objectifs (boucle 1). L'enseignant effectuera une régulation collective ou individuelle par la reprise de l'enseignement en se centrant sur les difficultés ponctuelles (boucle 2). Alternativement, une régulation collective pourrait être suivie d'une reprise de l'évaluation sur les objectifs de base ou sous forme de synthèse (boucle 3). Finalement, l'enseignement se poursuivra avec d'autres objectifs : l'enseignant élargira les objectifs initiaux par une mise en perspective par rapport à la matière, procédera à un retour synthèse (boucle 4) ou poursuivra son enseignement en abordant une nouvelle séquence d'objectifs (boucle 5). Rappelons que selon les programmes officiels, au niveau de l'évaluation et des interventions correctives, l'enseignant est le premier responsable du processus de régulation des apprentissages intégré à l'enseignement dans la classe et de ce fait, le maître d'oeuvre de ce processus de régulation (MEQ, 1981).

Les deux approches, instrumentée et qualitative, présentées précédemment diffèrent au sens de la mesure.

L'approche instrumentée comprend une mesure formalisée par un "test" portant sur des objectifs préalablement définis. L'interprétation de la mesure permet de s'assurer que les objectifs visés sont maîtrisés ou non. Traditionnellement, dans le contexte de la classe, les résultats de la mesure s'expriment en terme de moyennes individuelles et collectives. Aucune information n'est rapportée en terme de nature des erreurs ou, a fortiori, de cause. L'action pédagogique régulatrice des apprentissages et subséquente à

l'interprétation de la mesure n'est pas formalisée.

L'approche qualitative ne comprend pas de mesure formelle, elle prend la forme d'évaluations directes et individuelles des apprentissages de l'élève. L'évaluation peut être considérée comme partielle lorsqu'elle repose sur une information partielle. Idéalement, l'évaluation qualitative doit être systématique et rigoureuse (Miles & Huberman, 1984); par exemple, les interventions qualitatives auprès des élèves doivent être structurées à partir d'objectifs préalablement établis. L'action pédagogique, régulatrice des apprentissages, prend la forme d'interventions individuelles, ad hoc; elle dépend donc des interactions enseignant-élève, essentiellement informelles.

L'approche mixte, instrumentée et qualitative, nécessite de formuler des objectifs, de procéder à l'enseignement, à une prise d'informations instrumentée (ou pas), à l'interprétation des informations recueillies et finalement, à des actions pédagogiques régulatrices. Cette approche mixte permet de vérifier s'il y a maîtrise ou non, mais cherche en priorité, à partir des informations recueillies, à identifier la nature des erreurs et à déterminer leurs causes probables pour mieux cibler les actions pédagogiques correctives. Cette recherche d'identification et de détermination des erreurs constituent le diagnostic pédagogique qui s'inscrit dans le processus de l'évaluation formative.

Nous retenons à cet effet, la définition opérationnelle du diagnostic proposée par Dassa (1988, 1990) et reformulée par Dassa et René De Cotret (1993):

« "diagnostic pédagogique", partie intégrante de l'évaluation formative (...) Il consiste à repérer l'erreur, décrire la nature de l'erreur, décrire le contexte pédagogique où l'erreur se produit, tenter de relier la ou les causes de l'erreur à un paradigme de l'apprentissage ou de l'enseignement, proposer des interventions correctrices s'accordant à la didactique et à la technologie utilisées et s'inscrivant éventuellement dans un cadre théorique préalablement déterminé.» (Dassa, René De Cotret, 1993, p.8)

Les erreurs commises par un élève découlent de ses conceptions (représentations mentales) relatives à un concept donné et qui font état de ses connaissances fausses ou incomplètes. Une conception est dite juste, localement juste ou erronée. Une conception erronée peut provenir de la généralisation d'une conception localement juste de certains concepts ou dans une situation donnée mais non généralisable. Le système de conceptions d'un élève peut être influencé par ses connaissances antérieures, ses limitations personnelles ainsi que par l'enseignement qu'il a reçu (Dassa, René De Cotret, 1993).

Mentionnons que certaines "causes" de l'erreur, en dehors du paradigme d'apprentissage et de l'enseignement, telles la motivation, les attitudes de l'élève vis-à-vis le processus d'apprentissage et les approches pédagogiques, ne seront considérées dans les interventions correctives proposées.

1.2 Les pratiques d'évaluation.

Cette section présente les pratiques d'évaluation en classe et plus particulièrement les lacunes rapportées par des études américaines, canadiennes et québécoises. De ce constat, nous identifierons ensuite les éléments importants à prendre en compte pour développer des outils pédagogiques adéquats, c'est-à-dire propres à soutenir une pratique efficace de l'évaluation formative.

1.2.1 Les problèmes reliés aux pratiques d'évaluation

Des études américaines (Gullickson, 1982, 1985; Bloom, 1984; Kulik & Kulik, 1987; Crooks, 1988; Stiggins, 1988) portant sur des pratiques d'évaluation instrumentées en classe (Classroom Assessment), font ressortir les différents aspects, variables et impacts liés aux apprentissages.

Les recherches de Crooks (1988) relèvent le fait que l'évaluation peut avoir des effets positifs et négatifs. Certaines variables affectives comme le stress, l'anxiété, la motivation, l'attitude influencent de façon importante, selon les études mentionnées par Crooks, la capacité d'apprentissage de l'élève. De plus, d'autres facteurs directement associés à l'évaluation, tels que sa nature (formelle ou informelle), sa fréquence, le type d'interaction suite à sa réalisation, son orientation (centrée ou non sur la compétitivité), son individualisation, la présence (ou l'absence) de coopération ont un effet différent sur les apprentissages de l'élève.

Des études de Stiggins (1988, 1985) notent que l'évaluation est considérée comme une tâche complexe à effectuer par l'enseignant à cause des nombreux facteurs entrant en jeu reliés aux caractéristiques personnelles de l'enseignant et à celles du groupe d'élèves. La tâche est d'autant plus complexe que l'évaluation a un impact certain sur l'apprentissage des élèves.

L'étude de Kulik et Kulik (1987) avance que le principal avantage du processus d'évaluation dans le cadre de la pédagogie de maîtrise est lié à l'identification structurée des erreurs jumelée à l'interaction enseignant-élève en vue de la correction de ces erreurs. La rétroaction détaillée a donc des effets positifs sur l'apprentissage.

Pour ce qui est des pratiques d'évaluation des enseignants en classe, l'étude de Gullickson (1982), montre que l'approche informelle constitue une source importante d'informations pour les enseignants. Dans un article de Bloom (1984), l'étude d'examens conçus par les enseignants (les examens maison) révèle que le niveau taxonomique des questions posées correspond essentiellement au niveau taxonomique de la connaissance et par le fait même restreint le diagnostic et la remédiation à ce niveau. De plus, une étude de Gullickson (1985) note que les enseignants utilisent peu les techniques formelles de construction de tests ou les analyses statistiques permettant de les améliorer.

Les problèmes liés aux pratiques d'évaluation relevés dans ces études américaines (Gullickson, 1982, 1985; Bloom, 1984; Stiggins, 1988) ont trait au manque de formation des enseignants quant aux techniques de mesure spécifiques à l'éducation ainsi qu'à la construction d'instruments de mesure destinés à être utilisés en classe. D'autres lacunes ont trait à l'inadéquation des examens produits par des spécialistes autres que les enseignants.

Au Canada, selon l'étude d'Anderson (1989), le test objectif est la forme la plus commune des instruments de mesure utilisés; les enseignants ne sont pas intéressés aux tests provenant de l'extérieur; les enseignants combinent différentes sources d'informations provenant de différents instruments. En pratique, les évaluations réalisées dans le contexte de la classe n'intègrent guère les principes essentiels liés à la mesure tels l'identification de la nature de la mesure, la définition univoque d'attributs conceptuels, l'identification des comportements mesurables conformes aux objectifs d'enseignement prévus.

Des études en provenance du Québec (Scallon, 1988a; Dassa & Dumoulin, 1990; Fontaine, 1990; Dumoulin, 1991; Lanoue, 1993) notent que la cueillette d'informations dans le cadre de l'évaluation formative prend souvent la forme d'interrogations orales, d'interactions élève-élève ou enseignant-élève.

Nous retenons, la définition de Dumoulin (1991), à l'effet que les pratiques d'évaluation comprennent :

«toutes interactions entre l'enseignant et l'élève (ou les élèves) qui se présentent sous forme de communication orale, écrite ou gestuelle ayant pour but de recueillir, d'une façon formelle ou informelle, de l'information, de l'analyser, de l'interpréter, de poser un jugement et une décision-action nécessaire au cheminement ultérieur de l'élève». Dumoulin (1991, p.16)

L'évaluation formative est un processus dynamique (plutôt que statique) entre l'enseignant et l'élève qui permet, à partir de l'information recueillie, d'éclairer l'enseignant dans les actions régulatrices des apprentissages auprès de l'élève.

A l'instar des études américaines, celle de Burrelle *et al.* (1991a, 1991b) reconnaît l'influence de facteurs affectifs tels les croyances et les attitudes des enseignants à l'égard de la pratique de l'évaluation en classe.

Les principaux problèmes relevés dans les études en provenance du Québec (Scallon, 1988a; Dassa & Dumoulin, 1990; Fontaine, 1990; Dumoulin, 1991; Lanoue, 1993) et reliés à l'évaluation formative se résument comme suit : le manque de temps des enseignants pour la régulation individuelle des apprentissages, le manque de formation des enseignants en ce qui concerne la pratique de l'évaluation, le manque d'outils adaptés à la classe permettant de recueillir des informations pertinentes pour la remédiation des apprentissages.

A cet effet, Scallon (1988a, 1996) relève les réticences que les enseignants éprouvent face aux changements réels ou appréhendés vis-à-vis l'évaluation formative prônée par le Ministère de l'Éducation ainsi qu'à l'égard du

surplus de tâches quotidiennes qu'elle occasionnerait.

Les études de Dumoulin (1991) et Burelle *et al.*(1991b) montrent bien que les enseignants ne sont pas convaincus du bien fondé de ces changements pédagogiques touchant à la pratique de l'évaluation formative. Elles montrent aussi la difficulté qu'ils éprouvent à développer des instruments de mesure pour leur classe.

1.2.2 Besoins d'outils pédagogiques adéquats pour l'évaluation

Afin de répondre au besoin d'outillage pédagogique adapté à l'évaluation en classe, il convient d'en établir les caractéristiques désirables (Gullickson, 1982). En premier lieu, quelque soit le type d'outil pédagogique (démarche qualitative ou instrument de mesure) envisagé, il faut s'assurer que l'information recueillie soit conforme aux objectifs d'enseignement prévus (Allal, Perrenoud, Cardinet, 1979; Bloom *et al.*, 1981). Cette information doit être précise, pertinente et complète (Dassa, 1988; Perrenoud, 1991) pour que la prise de décision soit utile à l'enseignant et permette de mieux réaliser la régulation des apprentissages. D'autres contraintes relatives à l'organisation et à la gestion pédagogique de la classe, notamment le temps, devront être prises en compte pour que l'outil pédagogique puisse être intégré aux pratiques d'évaluation de la classe.

Il ressort des études canadiennes et québécoises citées précédemment, la nécessité de former les enseignants à la pratique de l'évaluation formative et notamment à la régulation des apprentissages. Il apparaît aussi nécessaire de les outiller en leur procurant des tests, des examens et différents types d'instruments de mesure ou d'évaluation qualitative qui soient adéquats et utilisables.

Pour que les outils pédagogiques destinés à la pratique de l'évaluation formative adaptée à la classe soit adéquats, ils doivent être fondés sur les

objectifs des programmes d'études du MEQ de façon à assurer l'efficacité de la régulation des apprentissages. De plus, le processus de développement des outils proprement destinés à l'évaluation devra être ouvert à la participation des enseignants. Cette participation devrait idéalement reposer sur une formation adéquate permettant de s'appropriier les outils élaborés. Pour que l'enseignement soit efficace, ces outils devraient renseigner l'enseignant sur les apprentissages de ses élèves d'une façon pertinente, complète et précise, tout en lui faisant gagner du temps.

La recherche de l'efficacité dans le contexte fortement contraint du couple enseignement-apprentissage en classe justifie l'utilisation de technologies informatiques de façon à répondre aux caractéristiques des outils préalablement définies (Dassa, Vazquez & Ajar, 1993). À cet effet, plusieurs études démontrent l'efficacité de l'informatique en terme de gain de temps et de ses capacités de gestion de processus complexes (Berthelot, 1987; Kapes & Vansickle, 1992). De plus, l'évolution rapide de ces technologies a permis aux systèmes informatiques développés d'être transférables aux applications pédagogiques de l'ordinateur.

Ces approches traitent de l'évaluation des apprentissages réalisés dans le contexte d'un enseignement collectif. Les considérations dominantes dans la littérature sont plus orientées vers l'enseignement que vers le processus d'apprentissage. Pourtant, les considérations de l'évaluation sont plus proches du processus même de l'apprentissage (Scallon, 1988a; 1996).

Certes, il existe potentiellement d'autres pistes pour atteindre une efficacité élevée pour soutenir la pratique de l'évaluation formative. Dans cette étude, nous choisissons d'exploiter les avantages liés aux technologies informatiques.

1.3 Le testing informatique.

Il existe un grand besoin au Québec et ailleurs, en instruments de mesure adaptés à la pratique de l'évaluation formative en classe. De tels outils doivent nécessairement porter sur les objectifs des programmes enseignés, fournir de l'information précise, complète et pertinente, afin d'assister l'enseignant dans la régulation des apprentissages tout en minimisant le temps dévolu à l'évaluation et en tenant compte des contraintes organisationnelles de la classe. À cet effet, nous considérons que l'apport des technologies informatiques au développement d'outils d'évaluation permet de gagner du temps, de gérer des processus complexes, de rendre l'outil efficace et modifiable facilement. Certaines études au Québec (Boulet *et al.*, 1990; Dassa, Vazquez & Ajar, 1993; Bordeleau, 1994a) ont prôné l'informatique comme un outil pédagogique. Si l'évaluation sommative sert d'abord le système éducatif, l'évaluation formative devrait servir davantage l'interaction enseignant-élèves en guidant les régulations ainsi que les démarches d'apprentissage des élèves.

Dans cette section, nous examinerons l'évolution des outils informatiques, du testing informatique et de ses stratégies de façon à mieux situer le problème de recherche.

1.3.1 Le développement des outils informatiques

L'utilisation de l'ordinateur à l'école trouve son origine aux États-unis dans les recherches et les développements appliqués réalisés dans le domaine militaire. Selon Bordeleau (1994b) l'ordinateur a d'abord servi à l'automatisation des cours; on l'intégra ensuite successivement aux systèmes de tuteurs, aux exercices et aux APO (Applications Pédagogiques de l'Ordinateur).

Au Québec, l'avènement de la micro-informatique au début des années 80 et la politique d'acquisition d'équipement du MEQ en 1985 ont permis aux écoles

d'augmenter considérablement leur parc informatique (Danvoye, 1987; Danvoye, 1990). Néanmoins, d'après l'étude de Danvoye (1989-1990), on ne dénombre en moyenne que 15 ordinateurs par école (secteurs primaire et secondaire), le plus souvent regroupés dans une salle à cette effet (laboratoire informatique). Le ratio est de un ordinateur pour deux élèves pour un groupe classe d'environ 30 élèves présent en même temps au laboratoire informatique.

A cette même époque, le gouvernement investit dans le financement du développement de logiciels éducatifs, d'un centre de recherche et de développement (*APO-QUÉBEC* de 1985-1990), dans la publication de répertoires des logiciels disponibles et dans la conception d'une banque d'items centralisée BIM destinée à l'évaluation des apprentissages. Par la suite, des centres régionaux d'excellence en micro-informatique scolaire (les CEMIS) furent créés. Ces événements ont stimulé le développement et l'intégration à l'école des A.P.O. axés sur le processus d'apprentissage (Delisle, 1992).

1.3.2 Les courants de développement du testing informatique

Comme dans le cas des autres outils informatiques, le testing informatique fut d'abord développé aux États-unis dans le domaine militaire. Bunderson *et al.* (1989) répartissent en quatre générations le testing informatique en éducation. Cette classification regroupe les tests informatisés (CT:*Computerized Testing*), les tests adaptatifs (CAT:*Computerized Adaptive Testing*), les systèmes ou tests informatisés à mesure continue (CM:*Continuous Measurement*) et les systèmes ou tests informatisés intelligents (IM:*Intelligent Measurement*). En général, les tests informatisés présupposent l'utilisation d'items à choix de réponse multiples.

La première génération des tests informatisés n'est en fait qu'une adaptation des tests conventionnels *papier-crayon* utilisant l'ordinateur. Par test conventionnel, on entend des tests élaborés selon la théorie classique des tests. Les instruments de mesure nécessitent l'administration d'un ensemble fermé

d'items présenté linéairement (Lord & Novick, 1968). La théorie classique des tests comporte plusieurs postulats qu'il faut respecter afin de contrôler l'erreur systématique imputable au score réel. Les tests conventionnels à interprétation normative présentent un certain nombre de problèmes fondamentaux bien documentés (voir, par exemple, Séguin & Auger, 1986a, 1986b). Il s'agit essentiellement de la dépendance des valeurs des caractéristiques psychométriques des items par rapport aux groupes de sujets testés, de la dépendance de l'estimation de l'habileté par rapport aux caractéristiques des items, de la référence des scores au groupe et non à l'individu et finalement, du fait que la fidélité relative d'un test s'exprime statistiquement en fonction d'un groupe et non de l'habileté individuelle.

Dans ce cas-ci, l'unique avantage de l'informatisation des tests réside dans ses capacités à emmagasiner l'information, à sa vitesse de traitement des données et de production des résultats. Cette génération de tests ne comporte pas de stratégie d'administration des items autre que celle définie par l'administration traditionnelle des tests papier-crayon, stratégie équivalente à une administration linéaire des items basée sur la théorie classique des tests.

La prise d'informations est la même pour tous les élèves mais elle peut se faire en différé selon l'organisation des laboratoires d'informatique et la disponibilité des ordinateurs. Les résultats s'expriment généralement en termes de scores individuels ou collectifs pour chacun des items et globalement pour l'ensemble du test. Aucune interaction entre le répondant et le système de testing n'est possible, la prise de mesure est qualifiée de statique "*static measurement*" (Bunderson *et al.*, 1989).

La seconde génération des tests informatisés se distingue de la précédente par l'adaptation du test en fonction des réponses de l'élève, ces tests sont dits "*adaptatifs*". Les stratégies d'administration des items des tests adaptatifs sont nombreuses : l'adaptativité basée sur le temps de réponse, le contenu, les

paramètres des items selon diverses approches (DoubleLevel, FlexiLevel, MultiLevel) et, plus récemment, selon la théorie des réponses aux items (TRI).

Un des avantages des tests adaptatifs est de rendre possible un cheminement particulier de l'élève selon la réponse fournie après chaque item. De plus, plusieurs auteurs (Laurier, 1992; Kapes & Vansickle, 1992; Wise & Plake, 1992, 1990; Auger, 1989) notent que les tests adaptatifs permettent de diminuer significativement le nombre d'items à administrer ainsi que la durée du test. Par ailleurs, dans la mesure où les items présentés ont un niveau de difficulté voisin du niveau d'habileté de l'élève, ils augmentent la précision du trait mesuré.

Le développement de tests adaptatifs requiert la construction de banques d'items, un mécanisme de sélection des items (incluant les conditions d'arrêt du test), d'estimation et d'interprétation des scores. L'élaboration d'une banque d'items nécessite la rédaction d'items pour un trait donné, l'expérimentation de ces items, leur sélection et leur paramétrisation finales. Le mécanisme de sélection des items consiste à présenter à l'élève l'item qui correspond le mieux à son habileté. L'estimation du score d'habileté de l'élève s'exprime mathématiquement selon le modèle choisi (modèle de Rasch, modèles à deux paramètres, à trois paramètres...) et vise à déterminer une valeur exprimant le niveau d'habileté de l'élève après chaque item. La condition d'arrêt d'un test adaptatif est déterminée habituellement par l'atteinte d'un seuil d'erreur prédéterminé et acceptable. Dans le cas de la TRI, c'est à partir de la fonction d'information de chaque item administré que l'on détermine la quantité d'information fournie par le test au fur et à mesure de son déroulement. Dans ce cas, l'erreur de mesure est une fonction inverse de l'information.

Le concept central des tests adaptatifs remonte aux recherches de Binet (1906) pour lesquels la mesure de l'intelligence était basée sur l'âge. D'autres travaux, notamment ceux de Lord (1952; 1980), de Birnbaum (1968) et de bien

d'autres (Rasch, 1980; Wright & Stone, 1979; Samejima, 1977; Weiss, 1974; Weiss & Kingsbury, 1984), ont contribué au développement et à la popularité de la TRI. Les développements de la TRI portent principalement sur le calcul du score d'habileté, sur le design des tests adaptatifs et leur utilisation à des fins d'évaluation (Hambleton & Swaminathan, 1985). La TRI comprend plusieurs modèles (unidimensionnel, multidimensionnel, linéaire et non-linéaire, réponse dichotomique, polytomique ou continue). Tous ces modèles expriment mathématiquement la relation entre le trait ou l'habileté sous-jacente à partir d'un ensemble d'items et la performance observable de l'élève (son score) (Hambleton & Swaminathan, 1985; Séguin & Auger, 1986a). La TRI a pour but d'assurer l'invariance statistique des items et de l'estimation de l'habileté de l'élève lorsqu'il y a une concordance raisonnable entre le modèle et un ensemble d'items mesurant un trait commun (Hambleton & Swaminathan, 1985).

La TRI comporte les avantages suivants : invariance de l'estimation de l'habileté d'un élève quelque soit l'ensemble d'items administré; invariance des paramètres ou caractéristiques des items; indice de précision de l'estimation selon l'habileté. Ainsi, la TRI remplace par un concept statistique d'estimation de l'habileté et d'erreur standard, l'utilisation des tests parallèles sur lesquels repose la théorie classique des tests (Hambleton & Swaminathan, 1985; Séguin & Auger, 1986a).

Les modèles de la TRI sont basés sur des postulats restrictifs tels l'unidimensionalité du trait mesuré, l'indépendance locale des réponses aux items, la forme mathématique de la courbe caractéristique de chaque item. Par unidimensionalité, on entend que l'habileté ou le trait mesuré par les items se situent sur un même et unique continuum. L'indépendance locale signifie qu'il n'y a pas de lien de dépendance entre les items d'un même test. Les caractéristiques des items sont obtenues par calibration auprès d'un grand échantillon d'individus. Elles s'expriment mathématiquement en une courbe propre à chaque item qui relie l'habileté mesurée par un item donné à la probabilité de succès à ce même

item, d'où l'appellation de courbe des caractéristiques des items (CCI). Cette courbe doit être croissante et peut s'exprimer différemment selon l'un ou l'autre des modèles mathématiques retenus. Ainsi, les modèles dits à trois paramètres définissent des indices de difficulté, de discrimination et de l'effet du hasard. Des postulats spécifiques à chacun des modèles doivent être respectés dans l'application de la TRI (Séguin & Auger, 1986a, 1986b).

Malgré les contraintes qu'elle impose, l'utilisation d'une stratégie d'administration des items basée sur la TRI est encore la stratégie dominante du testing adaptatif. Si les diverses contraintes sont respectées, l'information produite par les tests adaptatifs nous indique le niveau d'habileté d'un élève par rapport à un ensemble d'items de même dimension pour un trait donné. Par contre, plusieurs problèmes concernant la violation des postulats et la difficulté à les respecter sont exposés dans les écrits (Hambleton & Swanimathan, 1985; Séguin & Auger, 1986a, 1986b; Blais, 1987; Bunderson *et al.*, 1989).

La TRI requiert la construction de grandes banques d'items paramétrisés. Qui plus est, la paramétrisation ou la calibration des items doit se faire à partir d'un grand échantillon de répondants. Or, la constitution de grands échantillons dans le domaine scolaire est difficilement réalisable. Il s'ensuit une difficulté à s'assurer de l'invariance des paramètres et des qualités métrologiques des items. Notons, que la composition même de l'échantillon d'individus affecte l'estimation des paramètres et entre autre la détermination du point zéro (Blais 1987; Hambleton & Swanimathan, 1985). De plus, la TRI impose des contraintes quant à la forme des tests adaptatifs. Il faut que chacune des banques d'items mesure un même trait et que l'ensemble des items soit représenté sur un continuum d'habileté, ce qui est rarement le cas pour les tests scolaires. Ce problème est d'autant plus sérieux que la violation du postulat d'unidimensionalité crée un biais au niveau de l'estimation (Folk & Green, 1989). De plus, le continuum d'habileté établi pour chaque item à partir du groupe de référence fait en sorte que l'estimation du niveau d'habileté d'un élève correspond à une mesure à

interprétation dite normative. Par ailleurs, les modèles d'estimation les plus utilisés ne traitent que les réponses dichotomiques, d'où une grande perte d'information concernant les profils de catégories de leurres.

Étant donné que dans un test adaptatif, la sélection des items est fonction du niveau d'habileté individuelle de l'élève, les items présentés au cours d'un test diffèrent d'un élève à l'autre. Il en résulte parfois un problème au niveau de la validité du test individualisé particulièrement si la banque d'items est multidimensionnelle.

De plus, l'étude de Blais (1987) soulève le problème de convergence des estimations engendré par les patrons de réponses aberrants, les problèmes de sensibilité ou de robustesse des modèles et le peu de gains présentés par la TRI dans la pratique.

L'usage de l'ordinateur pour les tests de cette deuxième génération, en plus d'enregistrer l'information, d'exécuter les tâches rapidement (traitement des données et production de rapport), de gérer un grand nombre de processus complexes de décision, permet de gérer l'adaptation du test au cheminement particulier de chaque élève. Sans l'ordinateur, le testing adaptatif serait impossible à cause de la complexité des calculs mathématiques nécessaires au cours de l'administration d'un test.

La troisième génération des tests informatisés à mesure continue se distingue des deux premières générations de tests informatisés (CT et CAT) par la prise d'informations continue, c'est-à-dire une prise d'informations au début, en cours et après l'apprentissage dans le cadre d'un curriculum ou d'un programme de formation donné. Ces tests informatisés à mesure continue s'appuient sur une hiérarchisation ou une taxonomie des tâches du domaine de référence pour décider à chaque moment du cheminement ultérieur de l'élève. Ainsi, le cheminement de l'élève s'ajuste à chaque prise d'informations et

l'évaluation de son habileté se fait en fonction d'échelles dont la complexité dépend du schème conceptuel. Les fondements théoriques des tests informatisés à mesure continue reposent sur la validité de construit et sur la modélisation des fonctions cognitives de l'apprenant. La validité de construit concerne l'adéquation entre la structure de l'instrument de mesure et le modèle théorique des caractéristiques comportementales relatives au trait mesuré telles que ses composantes, les interrelations des diverses caractéristiques mesurées et l'importance relative de chacune de ces interrelations (Legendre, 1988).

Cette génération de tests correspond à un processus dynamique d'évaluation, permettant de dégager un profil de l'élève à dimensions multiples. Pour l'instant, il n'existe que des systèmes d'évaluation à mesure continue dits partiels (Bunderson *et al.*, 1989, p.388). Les difficultés rencontrées lors de l'opérationnalisation des tests à mesure continue, sont dues à la conception même de cette instrumentation, à la nature de la validité de construit mais aussi à la structuration des éléments diagnostiques nécessaires à l'orientation du cheminement de l'élève.

La conception de l'instrumentation répondant aux exigences des tests à mesure continue, nécessite la rédaction de nombreuses tâches de mesure, leur expérimentation et leur intégration dans le système informatisé (Laurier, 1993). De plus, il faut prévoir la régulation des apprentissages malgré les parcours différenciés des élèves. Le modèle psychométrique unidimensionnel n'est pas valable pour cette génération de tests tandis que les modèles multidimensionnels posent des problèmes d'ordre théorique (en ce qui a trait aux interrelations des dimensions) et d'ordre pratique (la taille des échantillons nécessaire pour la calibration des items). Malgré l'évolution rapide des technologies, des problèmes subsistent encore dans l'adaptation de l'équipement nécessaire à l'évaluation de certains comportements propres à certains domaines tels que l'apprentissage d'une langue. Même lorsque l'équipement adéquat existe, son utilisation routinière à l'école et l'intégration de ce type de tests à mesure continue dans le

contexte de la classe restent à faire.

La quatrième génération des tests informatisés dits "intelligents" se distingue des générations précédentes par sa fonction d'interprétation. L'information recueillie est interprétée en fonction des composantes d'un modèle élaboré de la connaissance et par le type d'information colligée. La modélisation de la connaissance d'un tel système est déviniée non seulement à partir d'un ensemble de tâches mais de réseaux de tâches. En plus de modéliser la connaissance, ces systèmes tentent d'établir une modélisation de l'apprenant et des relations entre les tâches et l'apprenant en cours de parcours (Laurier, 1997). Pour gérer ces modélisations, des systèmes d'inférences complexes composés d'ensemble de représentations et de stratégies cognitives (règles) sont utilisées. L'interprétation d'une réponse sert non seulement à identifier le niveau d'habileté de l'élève en comparant ses performances à celles attendues (provenant d'experts) mais aussi à modéliser la connaissance de l'apprenant vers un diagnostic cognitif dans le but de dresser un profil complet de l'élève identifiant ses difficultés et déterminant son cheminement ultérieur.

Ici encore, l'ordinateur serait utilisé pour ses capacités d'accumulation de données, de traitement rapide de masses d'informations complexes. Les technologies informatiques restent à être perfectionnées et développées dans le but de permettre une interaction ordinateur-élève la plus complète possible et une plus grande variété de traitement d'information tel l'analyse d'items à réponses construites (Bunderson *et al.*, 1989). Depuis, dans le milieu du développement multimédia, l'interactivité n'est plus seulement considérée comme une action de "feedback" ou de rétroaction provenant de l'ordinateur vers l'élève mais désormais comme un échange, une transaction multisensorielle (Giardina, 1997) entre l'individu et l'ordinateur. Cette nouvelle approche de l'interactivité intégrant le multimédia veut rendre les systèmes futurs plus flexibles et plus adaptatifs à l'élève.

D'autres problèmes d'ordre théorique relevant de la structuration du contenu et du maintien de la validité des mesures, étant donné la complexité des modèles de la connaissance, demeurent sans solution.

1.3.3 Des considérations liées au problème de recherche

Dans le milieu scolaire québécois, les tests informatisés les plus répandus sont de type conventionnel (tests non-adaptatifs à choix de réponse multiples). Leur usage est maintenant bien implanté. Les tests adaptatifs commencent à peine à faire leur apparition dans ce milieu. Pour ce qui est des tests à mesure continue et intelligents, ces deux générations de tests informatisés suscitent des travaux de recherche de pointe mais ne font présentement l'objet d'aucune diffusion au niveau de l'école et, a fortiori, de la classe.

Les problèmes qui retardent la diffusion consécutive des tests adaptatifs en milieu scolaire touchent à la difficulté de produire des instruments adéquats qui respectent les exigences strictes des postulats des modèles de la TRI, ainsi que les contraintes pratiques inhérentes aux tests adaptatifs. En ce qui a trait aux problèmes restreignant le développement et l'implantation de tests à mesure continue, les plus importants touchent au manque de connaissance des processus d'apprentissage, à la difficulté de définir l'ensemble d'un curriculum, aux problèmes théoriques liés aux modèles d'adaptabilité et d'estimation multidimensionnel ainsi qu'à la calibration des items.

De plus, en contexte d'évaluation formative, il est très difficile, sinon impossible, de constituer de très grands échantillons nécessaires à la calibration des items dans les modèles de la TRI. Les problèmes dérivant du respect de l'unidimensionalité des items constituant une banque, l'invariance local des items, la robustesse des modèles disponibles selon le type de données traitées, sont autant d'obstacles dont il faut tenir compte lors de la conception de tests adaptatifs utilisant les modèles de la TRI.

En ce qui a trait au développement de tests à mesure continue, les obstacles majeurs sont reliés à la difficulté de définir un curriculum très élaboré, à établir une taxonomie opérationnelle particulièrement destinée à l'évaluation diagnostique et compatible avec les objectifs des programmes et enfin, à la modélisation des apprentissages des apprenants (le design de la connaissance).

L'utilisation d'une nouvelle technologie amène toujours des contraintes intrinsèques ou organisationnelles (Berthelot, 1987) par rapport au système scolaire et à la planification de l'évaluation dans le temps.

Quant aux contraintes matérielles propres au milieu scolaire québécois, il convient de souligner l'accès très limité aux ordinateurs qui n'est possible que de façon intermittente, le plus souvent dans un local unique réservé à cet effet et en nombre inférieur au nombre d'élèves formant une classe.

Le système de tests informatisés qui fait l'objet de ce mémoire, constitue une tentative de développement de tests adaptatifs basés non pas uniquement sur la TRI mais aussi sur une taxonomisation tridimensionnelle du diagnostic des apprentissages en mathématiques. Le système développé s'inspire de certains aspects caractérisant les tests adaptatifs à mesure continue. Il convient de souligner que ce système porte sur des objectifs significatifs du début du programme d'algèbre et de géométrie de mathématiques au secondaire. Cette partie bien que très signifiante, puisqu'elle sert à constituer les préalables de l'apprentissage de l'algèbre et de la géométrie au secondaire, n'en demeure pas moins une portion réduite du curriculum de toute l'année : cet enseignement requiert environ huit semaines. De plus, une ébauche de design de la connaissance via les catégories de leurres et la navigation par les catégories de leurres seront intégrées au système de tests. De ce point de vue, ce système présente donc des caractéristiques relevant aussi bien de la deuxième génération de tests que de la troisième.

La problématique de la régulation de la performance de l'apprentissage comporte deux aspects indissociables et complémentaires. D'une part, il est important de colliger une information relative au produit des processus cognitifs: maîtrise des objectifs et des contenus, développement des habiletés générales. D'autre part, l'efficacité des aspects correctifs de la régulation repose sur une information relative à la nature même des processus cognitifs et à leur développement. Cependant, il est important de souligner dans ce mémoire que l'information relative à l'évaluation du processus même s'inscrit dans le cadre des contraintes fortement structurées par l'organisation de l'enseignement collectif.

En guise de conclusion de ce chapitre, rappelons qu'il convient de s'assurer que les outils développés répondent aux besoins du milieu scolaire. Il faudra donc en tout temps tenir compte de l'écologie de la classe et notamment du contexte des pratiques d'évaluation relatives aux contenus et aux objectifs des programmes "réels", c'est-à-dire effectivement enseignés en classe.

Chapitre second:
PROBLÈME DE RECHERCHE

Ce chapitre présente le système de tests adaptatifs existant, les problèmes rencontrés, suivis des objectifs de l'étude.

Le système de tests adaptatifs existant a été réalisé, grâce à une subvention du CRSH (1988-92), par le groupe de recherche dirigé par Clément Dassa. Le système de tests adaptatifs a été conçu en collaboration avec des enseignants de mathématiques du secondaire provenant des secteurs public et privé et recrutés sur une base volontaire. Cette forme de coopération entre les chercheurs et les enseignants lors du développement du système de tests adaptatifs a permis d'intégrer "un grand nombre de facteurs touchant le curriculum, le système scolaire, l'organisation de l'enseignement ainsi que les caractéristiques propres des enseignants et des étudiants" (Dassa, Vasquez & Ajar, 1993).

Le rapport final soumis au CRSH, spécifie les objectifs, la méthodologie, les étapes de planification, d'élaboration, la mise à l'essai et l'évaluation du système de tests adaptatifs existants (Dassa & Ajar, 1995). Les trois principaux objectifs de ce projet sont :

- 1) "formuler un modèle d'intégration du diagnostic pédagogique à l'évaluation formative en situation scolaire et particulièrement en mathématiques, au premier cycle du secondaire; produire et valider un matériel d'intervention approprié",
- 2) "valider le modèle par formulation et vérification d'hypothèses",
- 3) "évaluer les effets temporels concernant le maintien des apprentissages; reformuler le modèle dans sa version finale".

C'est donc à partir d'observations des pratiques sur une période de deux années (de 1988 à 1990), d'une planification et d'observations des outils pédagogiques tels que les manuels scolaires utilisés en classe, de la construction de près de 2000 items, de la validation de contenu de ces items

par les différents collaborateurs (membres du groupe de recherche, enseignants et spécialistes de la didactique), d'une expérimentation durant les années scolaires 1990-91 et 1991-92, que le système de tests adaptatifs existant a été réalisé.

Dans ce qui suit, nous présentons l'essentiel des résultats tel que rapportés dans le rapport final:

- 1) "l'intégration du diagnostic pédagogique à l'enseignement en classe de façon instrumentée, structurée et encadrée présente un effet sur le rendement scolaire moyen supérieur à celui de l'utilisation instrumentée mais libre";
- 2) "l'utilisation par l'enseignant de ses propres outils d'évaluation présente un effet moyen sur le rendement scolaire supérieur à celui de l'utilisation d'outils élaborés par d'autres ou non intégrés à sa pratique";
- 3) "les différences individuelles relatives au rendement scolaire sont réduites par le recours à un système de testing informatisé conçu de façon à individualiser le diagnostic pédagogique";
- 4) "l'effet moyen des résultats aux tests informatisés n'est pas supérieur à celui des tests traditionnels mais plusieurs indices, non confirmés statistiquement, tendent à lui conférer des avantages non négligeables" tel un fort taux de satisfaction et de motivation des usagers, enseignants et élèves dans l'utilisation d'un système visant un diagnostic individuel.

La collaboration des enseignants impliqués dans le projet a consisté entre autres, à définir les banques d'items et à les valider, à la mise à l'essai et à l'évaluation du système de tests adaptatifs durant les deux années de l'expérimentation. Avec eux, l'équipe de recherche a construit, à partir des manuels scolaires, d'exams maison et de leur pratique quotidienne, des items

répondant aux cinq premiers objectifs d'algèbre et aux trois premiers objectifs de géométrie de secondaire II. Les objectifs initiaux d'algèbre et de géométrie ont été retenus afin de s'assurer d'évaluer les habiletés initiales des élèves et, si nécessaire, d'intervenir le plus tôt possible en conformité avec les préceptes de la pédagogie de maîtrise.

Les banques d'items ont été constituées à partir des 2000 items développés et portant sur les objectifs retenus. Elles ont été structurées selon une taxonomie diagnostique ad hoc dite "modèle du cube diagnostique" (Auger & Dassa , 1991) fondée sur la définition opérationnelle du diagnostic pédagogique (Dassa, 1988, 1990). De ces 2000 items créés, 591 items ont été retenus après validation pour constituer les 8 banques d'items informatisées utilisées par les 16 tests adaptatifs. Chaque banque est organisée par habileté cognitive ou par objectif. Les objectifs d'algèbre et de géométrie utilisés par les tests adaptatifs correspondent aux objectifs 3.1.3, 3.1.6, 3.1.8, 3.2.3, 3.2.6, 5.5.2, 5.5.6 et 5.5.8 du programme officiel de mathématiques secondaire II (annexe A : Liste des objectifs).

La structure du "cube diagnostique" comporte trois dimensions: le contexte diagnostique, le contenu notionnel et l'habileté cognitive. L'axe du contexte diagnostique exprime toutes les composantes des facettes docimologiques pouvant être utiles au diagnostic pédagogique. L'axe du contenu notionnel relève du contenu de la matière. Le contenu exprime toutes les composantes des notions relatives à l'objectif. L'axe de l'habileté cognitive repose sur l'habileté exprimée par un objectif spécifique provenant du programme officiel. Les composantes d'une même dimension sont ordonnées du plus facile au plus difficile. De même, les items d'une même cellule sont hiérarchisés selon un continuum de difficulté croissante. Le tableau I présente un exemple de cube diagnostique pour l'objectif de géométrie 5.5.2: calculer la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.

Tableau I

Cube diagnostique pour l'objectif de géométrie 5.5.2

(seule la 1ère habileté cognitive est représentée: CALCULER l'aire latérale ou totale d'un prisme droit ou d'un cylindre)

Contextes diagnostiques		Contenu notionnel		CALCUL DE LA CIRCONFÉRENCE (1)	CALCUL DE L'AIRE (2)	
P R O B . A B S T R A I T S	Mesures identifiées sur une figure	1.nombres entiers	254 255 256 291 506 513 701 704 707		257 706 258 259 507 705 290 514 702 703 708 901	
		2.nombres décimaux				
	Mesures identifiées dans un texte	1."aire" présent nombres entiers	206 208 216 217 220 231 234 235 247 248 249 271 285 709 884		(233-237) 225 243 245 296 239 221 299 280 222 (238-201-205) 204 242 203 240 888 244 279 294	
		2.express. équiv. nombres entiers				
		1."aire" présent nombres décimaux	207 209 210 218 232 233 246 250 277 733		241 224 236 721 732 (202 720 722) 723	
		2.express. équiv. nombres.décimaux				
	P R O B . R É L L S	Mesures identifiées sur une figure	1.nombres entiers	260 502 508 710 712 713 714 715 716 718 711		719 (501 726) 511 724
			2.express. équiv.			
		Mesures identifiées dans un texte	1.nombres décimaux	262 503 512 717 725 727 729		263 728 731 730
			2.express. équiv.			
1."aire" présent nombres entiers			270 275 283 509 910		292 276 750 908 736 737 293 295	
2.express. équiv. nombres entiers			269 274 281 282 289 734 902 903 904		297 510 735 288	
1."aire" présent nombres décimaux	251 738 739 740		741 742 743 252			
2.express. équiv. nombres.décimaux	253					

Chacun des items appartient à une et une seule cellule du cube diagnostique. De ce fait, chaque item relève d'un contexte notionnel, d'un contexte diagnostique et d'une habileté cognitive. A l'intérieur d'une même cellule, les items ont été classés en premier lieu (1989-1990) selon le jugement des enseignants. Ce classement a été validé par des experts en 1990-1991. De plus, les composantes du contexte notionnel et du contexte diagnostique étant ordonnées, le niveau de difficulté des contextes notionnels et diagnostiques augmentent de gauche à droite et de haut en bas du cube diagnostique. Idéalement, une telle banque aurait dû faire l'objet d'une détermination empirique des indices docimologiques usuels (indices de difficulté, de discrimination, de l'effet du hasard) par l'entremise d'une validation auprès d'échantillons d'élèves adéquats. La taille de l'échantillon requis pour une telle paramétrisation aurait été de l'ordre de 2000 élèves. Dans le contexte actuel, il est quasiment impossible d'accéder à autant d'élèves lorsqu'il s'agit de valider un aussi grand nombre d'items (environ 2000). En effet, aucun dispositif de prise d'informations ne pourrait assurer l'administration d'autant de tests de façon uniforme et au moment propice après que l'enseignement de la matière pertinente ait été réalisé. Qui plus est, dans le cadre d'une recherche effectuée dans le contexte de l'évaluation faite en classe, où les items proviennent d'un domaine de référence reflétant les pratiques réelles, l'utilisation obligée d'élèves dont les enseignants n'auraient pas participé à la définition du domaine aurait introduit des biais en ce qui a trait à la validité de contenu et de construit. En bref, l'approche de validation retenue reflète la nécessité de tenir compte des pratiques d'évaluation.

Nous remarquons qu'il existe des cellules vides à l'intérieur de la majorité des banques. Nous rappelons ici que l'élaboration des items a été effectué en collaboration avec les enseignants à partir de leur pratique réelle en classe. Ces cellules vides reflètent en quelque sorte l'écart entre le curriculum officiel défini par le MEQ et celui qu'utilise réellement les enseignants.

La batterie de tests adaptatifs comprend 8 tests de modalité "laboratoire" et 8 tests de modalité "récupération". Les 8 tests de modalité "laboratoire" mis ensemble, font appel à 130 items et les 8 tests de modalité "récupération" font appel à 312 items. Le tableau II présente la répartition des items disponibles dans les banques selon les objectifs et des items utilisés par les tests informatisés selon les modalités d'administration (laboratoire ou récupération).

Tableau II

**Répartition des items selon les objectifs et
la modalité d'administration des tests informatisés**

Objectif	Nombre d'items banque informatisée	Nombre d'items test informatisé modalité "laboratoire"	Nombre d'items test informatisé modalité "récupération"
Algèbre 3.1.3	74	12	35
Algèbre 3.1.6	64	18	34
Algèbre 3.1.8	53	15	29
Algèbre 3.2.3	76	18	36
Algèbre 3.2.6	53	16	13
Géométrie 5.5.2	140	16	86
Géométrie 5.5.6	73	17	46
Géométrie 5.5.8	58	18	33
	591	130	312

La modalité d'administration des tests dite "laboratoire" consiste en l'administration collective d'un test adaptatif individualisé. L'adaptabilité du test ou le cheminement particulier de l'élève est déterminé par la stratégie de sélection des items qui sera exposée ultérieurement. Les tests adaptatifs de modalité "laboratoire" produisent une fiche de diagnostic individuel exhibée automatiquement à l'écran à la fin du test. Ces fiches individuelles accessibles

par l'élève et l'enseignant peuvent aussi, si l'enseignant le désire, être imprimées tout de suite. De plus, l'enseignant a accès à un programme de compilation des résultats individuels en des fiches de diagnostic collectif. Ces informations diagnostiques individuelles et collectives accessibles immédiatement après l'administration d'un test permettent à l'enseignant d'établir sans délai des stratégies d'intervention pédagogique auprès des élèves. Les tableaux III et IV présentent des exemples d'une fiche individuelle et des fiches collectives résultant de l'administration du test en mode "laboratoire" correspondant à l'objectif de géométrie 5.5.2.

Tableau III

**Fiche de diagnostic individuel résultant de l'administration du test
correspondant à l'objectif de géométrie 5.5.2**

Test interpretation for élève #1

14 May 92

Projet de recherche subventionné par le CRSH - Canada.
Responsable: Clément Dassa (Université de Montréal).

-- Parcours de l'élève pour le test de l'objectif 5.5.2 --

No item	Cont. notion	Donnée	Regr. cont.	Bon. rép.	Éleve choix	Le résultat	Numéro leurre
Heomb 513	circ.	rayon	A	1	1	réussite	
Heomb 514	aire	diam.	A	3	1	échec	8
Heomb 507	aire	rayon	A	4	1	échec	12
Heomb 225	aire	rayon	B	2	2	réussite	
Heomb 206	circ.	diam.	B	1	2	échec	7
Heomb 508	circ.	diam.	D	4	4	réussite	
Heomb 511	aire	rayon	D	2	4	échec	2
Heomb 276	aire	rayon	E	1	4	échec	12
Heomb 509	circ.	diam.	E	3	1	échec	7

L'élève a réussi 3 questions sur 9; sa moyenne: 33.333%

 -- Description des résultats --

Contenus notionnels :

1. Calcul de la circonférence:
 2 item(s) réussi(s) sur 4 questions posées.
2. Calcul de l'aire:
 1 item(s) réussi(s) sur 5 questions posées.

Contextes diagnostiques :

- Problèmes abstraits
 - A: La mesure du rayon ou du diamètre est identifiée sur la figure.
 Résultat: 1 item(s) réussi(s) sur 3 questions.
 - B: La mesure du rayon ou du diamètre est identifiée dans le texte.
 Résultat: 1 item(s) réussi(s) sur 2 questions.
- Problèmes réels
 - D: La mesure du rayon ou du diamètre est identifiée dans le texte et indiquée sur la figure.
 Résultat: 1 item(s) réussi(s) sur 2 questions.
 - E: La mesure du rayon ou du diamètre est identifiée dans le texte.
 Résultat: 0 item(s) réussi(s) sur 2 questions.

 -- Interprétation en fonction des leurres proposés. --

L'élève a choisi 1 fois sur 2 le leurre 2.
 Desc: élève au carré le diamètre au lieu du rayon dans le calcul de l'aire.
 Form: aire = [(diam) élevé au carré] x pi.

L'élève a choisi 2 fois sur 3 le leurre 7.
 Desc: double le diamètre inutilement dans le calcul de la circonférence.
 Form: circ = (diam x 2) x pi.

L'élève a choisi 1 fois sur 7 le leurre 8.
 Desc: place la virgule décimale à la mauvaise position.

L'élève a choisi 2 fois sur 3 le leurre 12.
 Desc: multiplie le rayon par pi au lieu de calculer l'aire.
 Form: aire = ray x pi.

----- Fin du test pour l'objectif 5.5.2 -----

Tableau IV
Fiches de diagnostic collectif
résultant de l'administration du test
correspondant à l'objectif de géométrie 5.5.2

DIAGNOSTIC COLLECTIF

Projet de recherche CRSH - Clément Dassa - U. de M.

Professeur: *PROF #2*
Groupe : 1
Nb élèves : 18

Objectif : Calculer la circonférence ou l'aire d'un cercle

A) RESULTATS INDIVIDUELS

MAITRISE	NOM ET PRÉNOM	NOTION 1	NOTION 2	CONT. A	CONT. B	CONT. D	CONT. E	TOTAL
	<i>ÉLÈVE #1</i>	4/04	4/04	2/02	2/02	2/02	2/02	08/08
**	<i>ÉLÈVE #2</i>	4/06 *	1/04 **	1/03 **	2/02	1/03 **	1/02 **	05/10
	<i>ÉLÈVE #3</i>	4/04	4/04	2/02	2/02	2/02	2/02	08/08
	<i>ELEVE #4</i>	4/06 *	4/04	2/03 *	2/03 *	2/02	2/02	08/10
*	<i>ELEVE #5</i>	3/05 *	4/05	2/02	1/03 **	2/02	2/03 *	07/10
**	<i>ELEVE #6</i>	4/04	1/06 **	1/03 **	1/02 **	1/03 **	2/02	05/10
*	<i>ELEVE #7</i>	3/06 **	4/04	2/02	1/03 **	2/02	2/03 *	07/10
	<i>ELEVE #8</i>	4/04	4/04	2/02	2/02	2/02	2/02	08/08
	<i>ELEVE #9</i>	4/04	4/04	2/02	2/02	2/02	2/02	08/08
**	<i>ÉLÈVE #10</i>	4/05	1/05 **	1/03 **	2/02	1/03 **	1/02 **	05/10
	<i>ELEVE #11</i>	4/05	4/04	2/03 *	2/02	2/02	2/02	08/09
**	<i>ELEVE #12</i>	4/05	1/05 **	1/03 **	2/02	1/03 **	1/02 **	05/10
*	<i>ELEVE #13</i>	4/04	3/06 **	2/03 *	2/02	2/02	1/03 **	07/10
*	<i>ELEVE #14</i>	4/04	3/06 **	1/03 **	2/02	2/03 *	2/02	07/10
*	<i>ELEVE #15</i>	4/04	3/06 **	2/02	2/03 *	1/03 **	2/02	07/10
**	<i>ELEVE #16</i>	3/05 *	1/05 **	1/03 **	2/02	1/03 **	0/02 **	04/10
	<i>ELEVE #17</i>	4/04	4/04	2/02	2/02	2/02	2/02	08/08
**	<i>ELEVE #18</i>	4/06 *	2/05 **	1/03 **	2/02	2/03 *	1/03 **	06/11

Les résultats se lisent comme suit:

Dans chaque colonne : nb réussites / nb items soumis

* signifie difficulté mineure

** signifie difficulté majeure

DIAGNOSTIC COLLECTIF

Projet de recherche CRSH - Clément Dassa - U. de M.

Professeur:

Groupe : 1

Nb élèves : 18

Objectif : Calculer la circonférence ou l'aire d'un cercle

REGROUPEMENTS	DIAGNOSTIC STRUCTUREL						
	MOYENNES			FREQUENCES (Nb)			NOMBRE
	ITEMS REUSSIS	ITEMS PASSES	RESULTATS INDIVIDUELS	DE 0% A 59%	DE 60% A 79%	DE 80% A 100%	
Notion 1. Calcul de la circonférence.	3.8	4.7	84%	6% (1)	28% (5)	67% (12)	18
Notion 2. Calcul de l'aire.	2.9	4.7	65%	50% (9)	0% (0)	50% (9)	18
Probl. abstraits: mesure identifiée sur une figure.	1.6	2.6	69%	39% (7)	17% (3)	44% (8)	18
Probl. abstraits: mesure identifiée dans le texte, nombres entiers.	1.8	2.2	86%	17% (3)	11% (2)	72% (13)	18
Problèmes réels: mesure identifiée sur une figure.	1.7	2.4	74%	33% (6)	11% (2)	56% (10)	18
Problèmes réels: mesure identifiée dans le texte, nombres entiers.	1.6	2.2	75%	33% (6)	11% (2)	56% (10)	18
Calculer la circonférence ou l'aire d'un cercle.	6.7	9.4	73%	33% (6)	28% (5)	39% (7)	18

Se lit comme suit: Moyennes : items = nombre moyen d'items passés ou réussis par la classe

: résultats individuels = moyenne des résultats de chaque élève

Fréquences: Nombre et pourcentage d'élèves dans chacune des catégories

Total : Nombre d'élèves ayant passé au moins un item pour le regroupement indiqué

TEST: H552

DIAGNOSTIC COLLECTIF

Projet de recherche CRSH - Clément Dassa - U. de M.

Professeur:

Groupe : 2

Nb élèves : 27

Objectif : Calculer la circonférence ou l'aire d'un cercle

DIAGNOSTIC CAUSAL

CATEGORIE DE LEURRE	FREQUENCES DE CHOIX	FREQUENCES D'OCCURRENCE	TAUX DE CHOIX	FREQUENCES 2 CHOIX OU + ELEVES % (Nb)
2. Élève le diamètre au carré dans le calcul de l'aire.	4	65	6%	0% (0)
3. Calcule la circonférence au lieu de l'aire.	23	126	18%	22% (6)
7. Double le diamètre inutilement dans le calcul de la circonférence.	6	70	9%	7% (2)
8. Place la virgule décimale à la mauvaise position.	9	210	4%	7% (2)
9. Calcule l'aire au lieu de la circonférence.	11	96	11%	11% (3)
10. Oublie de doubler le rayon dans le calcul de la circonférence.	11	65	17%	7% (2)
11. Divise le rayon par deux puis l'élève au carré - calcul de l'aire.	1	26	4%	0% (0)
12. Multiplie le rayon par pi au lieu de calculer l'aire.	13	75	17%	19% (5)
15. Ne tient pas compte d'une donnée pertinente dans le calcul de l'air	1	21	5%	0% (0)
19. Double le diam., l'élève au carré, puis multiplie par pi - circonf.	0	14	0%	0% (0)
21. Élève le diamètre au carré, le multiplie par pi - circonférence.	1	15	7%	0% (0)

La fiche individuelle de diagnostic comporte des informations descriptives liées au diagnostic structurel et des informations liées au diagnostic causal. L'information descriptive comprend entre autres, le nombre d'items réussis, la répartition des items réussis en fonction des contenus notionnels et des contextes diagnostiques. La seconde partie de la fiche de diagnostic individuel comprend des informations se rapportant au diagnostic causal soit une explication didactique de l'erreur commise; la fréquence des catégories de leurs choisis par l'élève complète cette fiche.

Les fiches de diagnostic collectif résultent d'une compilation des résultats individuels des élèves. Elles sont au nombre de trois : la compilation des résultats individuels, du diagnostic structurel et du diagnostic causal. La fiche collective des résultats individuels indique selon les notions et les contextes, le nombre d'items réussis par chacun des élèves. De plus, le système de compilation des fiches de diagnostic individuel indique, par des astérisques, les élèves qui ont moins bien réussi. La fiche collective du diagnostic structurel comporte des statistiques descriptives, par exemple, le nombre moyen d'items réussis par la classe pour chaque contexte notionnel, pour l'ensemble de l'objectif. La fiche collective du diagnostic causal donne des statistiques descriptives globales concernant la fréquence des catégories de leurs choisis par les élèves.

La modalité d'administration dite "récupération" s'adresse aux élèves que l'enseignant a identifié comme présentant des difficultés à partir des résultats au test de modalité "laboratoire". La séance de récupération permet alors de recommencer l'évaluation diagnostique portant sur le même objectif, à partir d'un ou de plusieurs contextes diagnostiques ou contenus notionnels déterminés par l'enseignant. Ces tests en mode "récupération" permettent une prise d'informations plus approfondie et s'effectuent en présence de l'enseignant, ce qui renforce l'interaction élève-enseignant. Ce système est

ouvert en ce qui a trait au choix du contenu ou du contexte à revoir, mais il reste que la navigation à l'intérieur d'un axe est déterminé par la programmation de la stratégie de sélection et d'administration des items du système. A la fin de la séance d'un test en mode "récupération", une fiche de diagnostic individuel est produite. Elle comporte de l'information liée aux diagnostics structurel et causal, semblable à celle produite en mode "laboratoire".

Nous retenons que la navigation dans la banque fait partie de la stratégie d'administration des items. La stratégie de sélection et d'administration des items qui a été adoptée est dite "stradaptative", la navigation à l'intérieur d'une banque d'items fixes se fait en fonction d'une facette pour un contexte ou pour un contenu donné. Rappelons que les items ont été classé dans un "cube diagnostique" selon 3 dimensions et que chacune des dimensions fait l'objet d'une stratification mutuellement exclusive. A l'intérieur d'une même strate les items sont ordonnés selon la difficulté. La sélection des items s'effectue comme suit : le premier item présenté à tous les élèves correspond à l'item du centre de la première cellule. Les items subséquents seront sélectionnés selon la réussite ou l'échec à l'item précédent. Ainsi, on présentera un item plus facile si l'élève a échoué; sinon on présentera un item classé plus difficile. On répétera l'opération jusqu'à ce qu'un minimum de réussite ou d'échec et/ou un maximum d'items présentés pour cette facette soit atteint. Par la suite, les items administrés appartiendront à la facette (contexte ou contenu) suivante. Le nombre total d'items administrés à un test donné est limité afin de répondre aux contraintes exposées précédemment (limite de temps, du nombre d'appareils disponibles, disponibilité des appareils). Afin de respecter la contrainte de temps, les tests adaptatifs existants comportaient un maximum de 15 items. Selon les habiletés cognitives (objectifs) testés, le nombre minimum d'items par test fluctuait entre 9 et 15.

Ainsi, une stratégie d'administration d'un test est déterminée en partie par les contraintes organisationnelles. D'autres contraintes, comme le peu d'items contenu dans certaines facettes (contenu ou contexte) ou le manque d'items dans certaines cellules des banques, restreignent la navigation à l'intérieur d'une banque.

La stratégie "stradaptative" a été programmée dans un environnement MicroCat™ version 4. Ce logiciel permet, en principe, la programmation d'une stratégie d'administration des items. Cependant, selon Auger (1990), le logiciel MicroCat™ version 4 ne permet pas de programmer l'algorithme de sélection des items selon la stratégie "stradaptative", pour des raisons techniques liées au très grand nombre d'embranchements possibles et aux limites de mémoire du logiciel. Pour contourner ce problème de programmation de la stratégie de sélection des items, une programmation ad hoc a été effectuée dans laquelle certains parcours ont été planifiés et programmés. La figure 2 et le tableau V illustrent la navigation dans la banque d'items pour le test de modalité "laboratoire" correspondant à l'objectif de géométrie 5.5.2.

Nous observons à la figure 2, les 16 items utilisés pour le test et les parcours planifiés (programmés) entre les items. Le tableau V illustre la navigation programmée pour le test donné à l'intérieur de la banque comprenant tous les items de cet objectif. Nous constatons qu'un parcours parfait illustré par un "trait plein" signifie que l'élève répond correctement à la série d'items #513, #514, #238, #248, #512, #263, #510 et #274. Le second parcours illustré représente l'élève qui échoue tous les items qui lui sont présentés: #513, #506, #507, #225, #206, #508, #511, #276 et #509. Il existe plusieurs autres combinaisons possibles de parcours qui ne sont pas illustrés dans ce tableau et correspondent à l'alternance d'un élève à réussir et à échouer une ou plusieurs items.

Tableau V
Navigation ad hoc selon la stratégie "stradaptable" pour le test de modalité "laboratoire" obj.géom. 5.5.2
(1ère habileté cognitive seul.: CALCULER l'aire latérale ou totale d'un prisme droit, d'un cylindre)

Contextes diagnostiques		Contenu notionnel	CALCUL DE LA CIRCONFÉRENCE (1) DÉBUT	CALCUL DE L'AIRE (2)
P R O B . R É E L S	Mesures identifiées sur une figure	1.nombres entiers	254 255 256 291 506 513 701 704 707	257 706 258 259 507 705 290 514 702 703 708 901
		2.nombres décimaux		
	Mesures identifiées dans un texte	1."aire" présent nombres entiers	206 208 216 217 220 231 234 235 247 248 249 271 285 709 884	(233-237) 225 243 245 296 239 221 299 280 222 238 201-205) 204 242 203 240 888 244 279 294
		2.express. équiv. nombres entiers		
		1."aire" présent nombres décimaux	207 209 210 218 232 233 246 250 277 733	241 224 236 721 732 (202 720 722) 723
		2.express. équiv. nombres.décimaux		
	Mesures identifiées sur une figure	1.nombres entiers	260 502 508 710 712 713 714 715 716 718 711	719 (501 726) 511 724
		2.express. équiv.		
		1.nombres décimaux	262 503 512 717 725 727 729	263 728 731 730
		2.express. équiv.		
Mesures identifiées dans un texte	1."aire" présent nombres entiers	270 275 283 509 910	292 276 750 908 736 737 293 295	
	2.express. équiv. nombres entiers	269 274 281 282 289 734 902 903 904	297 510 735 288	
	1."aire" présent nombres décimaux	251 738 739 740	741 742 743 252	
	2.express. équiv. nombres.décimaux	253		
			FIN	

Légende : réussite → et échec -->

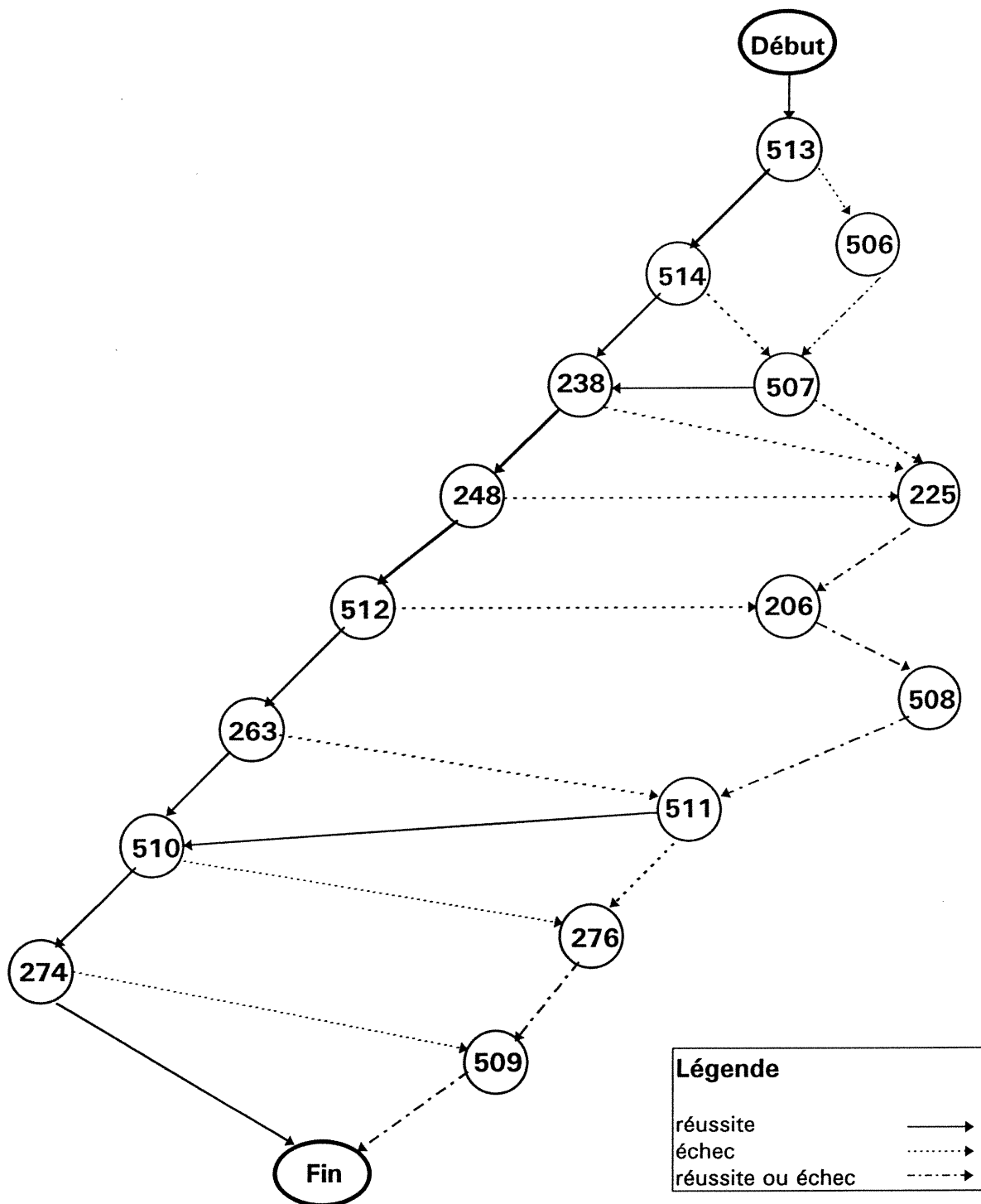


figure 2 - Navigation ad hoc selon la stratégie "stradaptative" pour le test de modalité "laboratoire" correspondant à l'objectif de géométrie 5.5.2

La programmation de la stratégie "stradaptative" actuelle est fermée car tous les embranchements pertinents ont été prévus entre les items d'un test donné (de 12 à 18 items pour les tests de modalité laboratoire et de 13 à 86 items pour les tests de modalité récupération). La programmation des tests dépend donc du contenu de chacune des banques. Tout changement à une banque entraîne des changements à la programmation du test. La programmation de la sélection des items n'est donc pas assez flexible compte tenu des changements pouvant être apportés aux banques d'items et ne peut être transposable à de nouveaux tests sans apporter des modifications majeures à la programmation existante. Pourtant, il est important qu'un système de tests de ce type soit flexible et facilement modifiable (évolutif) pour être viable (Depover, C., 1987).

Le système de tests adaptatifs existant ne comporte que des items à choix multiples : 4 choix de réponses sans l'utilisation de l'énoncé "aucune de ces réponses". Chaque choix de réponses erronées ou catégorie de leurres est lié à une hypothèse de non compréhension équivalente à une conception erronée. Ainsi, à chaque catégorie de leurres est associée un type de difficulté plausible. Les catégories de leurres ont été établies par les chercheurs et des didacticiens experts de la matière, de concert avec les enseignants, à partir d'une analyse didactique approfondie des erreurs habituellement commises par les élèves.

Rappelons que le but principal du système est de renseigner l'enseignant sur l'état des apprentissages de l'élève, par le biais de fiches diagnostiques afin qu'il puisse procéder rapidement à des interventions correctives efficaces. D'où l'importance de constituer un diagnostic précis. Tel que mentionné par Dassa et René De Cotret (1993), afin d'identifier les conceptions erronées des élèves, les seules informations disponibles sont liées aux catégories leurres sans référence aucune aux processus utilisés par l'élève. D'où l'importance

d'associer les catégories de leurres à des conceptions permettant d'induire les processus de résolution de problèmes probablement utilisées par l'élève. Il s'ensuit que l'élaboration et l'agencement des catégories de leurres constituent une phase essentielle à l'obtention d'un diagnostic de qualité.

Comme mentionné (Dassa & René De Cotret, 1993), une catégorie de leurres doit être liée à une erreur et non à une faute. Une faute relève plutôt d'une inattention tandis que l'erreur dépend généralement d'une incompréhension de la matière. De plus, la description des catégories de leurres reportée dans les fiches de diagnostic causal, doit permettre à l'enseignant de reconnaître l'hypothèse de la conception sous-jacente à l'erreur, afin de lui donner une piste de remédiation qui soit efficace. Avant de conclure à une conception erronée particulière, il convient de s'assurer que l'erreur est commise de façon cohérente sinon systématique. Il apparaît donc nécessaire de tenir compte du profil de sélection des catégories de leurres afin de dégager un profil cohérent des difficultés d'apprentissage de l'élève.

Le système de tests adaptatifs existant communique la nature et la fréquence des catégories de leurres choisies par l'élève par l'entremise des fiches de diagnostic causal. Cependant, ces catégories de leurres ne contribuent pas à orienter directement la stratégie de sélection des items et d'administration des tests initiaux. Ils ne confirment pas plus la persistance de l'élève à choisir certaines catégories de leurres. Uniquement leurs fréquences sont compilées et communiquées à la fin de l'unique phase du test. Une utilisation plus systématique des catégories de leurres pour des fins de navigation dans le cube diagnostique et de raffinement du diagnostic causal est donc souhaitable. Une deuxième phase au test visant spécifiquement à améliorer le diagnostic causal sera donc présentée au chapitre troisième.

Les tableaux VI, VII et VIII suivants présentent la répartition des catégories de leurres selon l'objectif visé, la liste des catégories de leurres utilisées pour le test correspondant à l'objectif de géométrie 5.5.2 (les listes des autres objectifs sont placées en annexe A) et leur répartition pour la banque totale d'items, dans les tests des modalités "laboratoire" et "récupération".

Tableau VI
Répartition des catégories de leurres
selon l'objectif visé

Objectif	Nombre de catégories de leurres différentes
algèbre 3.1.3	15
algèbre 3.1.6	17
algèbre 3.1.8	37
algèbre 3.2.3	24
algèbre 3.2.6	21
géométrie 5.5.2	22
géométrie 5.5.6	16
géométrie 5.5.8	17
	169

Pour l'ensemble des banques d'items, 169 catégories de leurres différentes composent 1773 choix de réponses erronées possibles, soit 3 choix erronés par item pour les 591 items de l'ensemble des banques d'items. Chaque catégorie de leurres apparaît un nombre variable de fois dans une banque d'items.

Tableau VII
Liste des catégories de leurres pour le test correspondant
à l'objectif de géométrie 5.5.2

02	:	Élève au carré le diamètre au lieu du rayon dans le calcul de l'aire.
03	:	Calcule la circonférence au lieu de l'aire.
05	:	Double le diamètre et le multiplie par pi au lieu de calculer l'aire.
06	:	Oublie d'élever au carré le rayon dans le calcul de l'aire.
07	:	Double le diamètre inutilement dans le calcul de la circonférence.
08	:	Place la virgule décimale à la mauvaise position.
09	:	Calcule l'aire au lieu de la circonférence.
10	:	Oublie de doubler le rayon dans le calcul de la circonférence.
11	:	Divise le rayon par deux avant de l'élever au carré dans le calcul de l'aire.
12	:	Multiplie le rayon par pi au lieu de calculer l'aire.
13	:	Multiplie au lieu de diviser.
15	:	Ne tient pas compte d'une donnée pertinente dans le calcul de l'aire.
16	:	Oublie de convertir les unités.
19	:	Double le diamètre, l'élève au carré, multiplie le résultat par pi au lieu de calculer la circonférence.
21	:	Élève le diamètre au carré, le multiplie par pi au lieu de calculer la circonférence.
22	:	Élève au carré le produit de pi par le rayon au lieu de calculer l'aire.
24	:	Divise le rayon par deux et multiplie par pi au lieu de calculer la circonférence.
25	:	Divise le diamètre par deux, deux fois de suite et multiplie le quotient par pi au lieu de calculer l'aire.
26	:	Divise le diamètre par deux et multiplie le quotient par pi au lieu de calculer l'aire..
27	:	Double deux fois de suite le rayon et multiplie le produit par pi au lieu de calculer la circonférence.
28	:	Double le rayon, élève le produit au carré, multiplie par pi au lieu de calculer l'aire.
29	:	Calcule l'aire au lieu de la circonférence et place la virgule décimale à la mauvaise position.

Tableau VIII

Répartition des catégories de leurres pour l'objectif de géométrie 5.5.2" dans la banque d'items et les tests de modalités "laboratoire" et "récupération".

no.des catégories de leurres	fréquence des catégories de leurres		
	banques d'items	tests modalité "laboratoire"	tests modalité "récupération"
02	23	4	13
03	64	8	41
05	9	0	6
06	2	0	0
07	31	5	21
08	110	13	67
09	51	5	26
10	37	3	22
11	13	2	7
12	31	5	25
13	1	0	0
15	10	1	5
16	5	0	1
19	3	1	2
21	10	1	9
22	9	0	3
24	2	0	2
25	1	0	1
26	1	0	1
27	2	0	2
28	4	0	4
29	1	0	0
Total	420	48	258

Finalement, mentionnons brièvement que l'étude de Lanoue (1993) qui portait sur la mise à l'essai du système existant, a permis de faire ressortir les avantages des tests adaptatifs par rapport aux tests conventionnels du point de vue des pratiques pédagogiques en classe. À ce chapitre, soulignons l'essentiel : la rapidité de production des résultats d'un test, la participation et l'engagement des élèves dans la vérification de leurs apprentissages, la disponibilité du profil global des difficultés individuelles et collectives des élèves. L'information produite a permis à l'enseignant participant de mieux planifier ses interventions de remédiation collectives et individuelles auprès des élèves en difficulté. Rappelons que parmi les conclusions de l'étude générale (Dassa & Ajar, 1995), il est établi que "l'intégration du diagnostic pédagogique à l'enseignement en classe de façon instrumentée, structurée et encadrée [est supérieure] à l'utilisation instrumentée mais libre".

Le problème de recherche de ce mémoire se situe dans la continuité du développement du système diagnostique informatisé existant. Il porte sur l'élaboration d'une variante de stratégie de testing informatisée et interactive, centrée sur le diagnostic pédagogique.

Nous proposons, en premier lieu, de programmer l'algorithme "stradaptatif" de sélection des items, afin de rendre les banques d'items indépendantes de la programmation.

Les objectifs visés consistent:

- a) à produire un algorithme qui soit indépendant du contenu des banques d'items et compatible avec les contraintes relatives aux banques d'items et à l'administration des tests existants;
- b) à tester et à vérifier son fonctionnement par une étude de simulation.

Dans un deuxième temps, nous proposons d'ajouter au test une seconde stratégie de sélection des items qui utilise la catégorie de leurres la plus fréquente préalablement identifiée par la stratégie "stradaptative" lors d'une première phase du test. Nous utiliserons la catégorie de leurres comme élément d'information pour la sélection des items à présenter à l'élève. Cette stratégie d'administration de tests axée sur la catégorie de leurres devra permettre d'estimer le niveau de difficulté cognitive de l'élève par rapport à cette catégorie de leurres précédemment identifiée. Cette information s'ajoutera à celle déjà produite par le système de tests adaptatifs actuel en ce qui concerne le diagnostic causal et indiquera à l'enseignant la propension de chaque élève à commettre certaines erreurs.

Les objectifs visés consistent:

- a) à programmer un algorithme de sélection des items axé sur la catégorie de leurres;
- b) à intégrer cette stratégie au système de tests adaptatifs;
- c) à tester et à vérifier son fonctionnement dans des situations simulées;
- d) à définir une mesure (un score) de difficulté cognitive à partir du modèle de TRI de Rasch (à un paramètre);
- e) à établir des bornes d'interprétation des scores de difficulté cognitive obtenues à partir de l'étude de simulations.

Le développement du système de tests adaptatifs proposé tiendra compte des contraintes de la classe. Parmi ces contraintes, spécifions le nombre d'appareils disponibles habituellement centralisés dans un laboratoire informatique, la disponibilité des équipements et la durée consacrée à une période de cours. De plus, nous respecterons l'environnement MicroCattm utilisé par le système de tests adaptatifs existant.

Chapitre troisième:
MÉTHODOLOGIE

Ce chapitre présente la méthodologie utilisée pour réaliser et valider le système de tests adaptatifs développé. Dorénavant, un test comporte deux phases au cours d'une même séance d'administration. La première phase du test permet d'établir un diagnostic structurel et de formuler une hypothèse du diagnostic causal. La seconde phase du test sert à confirmer ou infirmer, le diagnostic causal proposé à la fin de la première phase.

La première phase d'un test s'apparente au système de tests adaptatifs existant à la différence que la stratégie "stradaptative" de sélection des items ne dépend plus d'une programmation ad hoc mais est automatisée (par programmation) et rendue indépendante du contenu de la banque d'items. La seconde phase utilise l'information obtenue (l'identification de la catégorie de leurres la plus fréquente) à la fin de la première phase du test, pour sélectionner et administrer un sous-ensemble d'items comportant tous ce leurre relevant d'une même catégorie. Cette seconde phase permet d'estimer le niveau de cohérence de l'élève à choisir une catégorie de leurres donnée et à encadrer l'interprétation de cette mesure à l'aide d'un message qualitatif destiné à l'enseignant. La fiche diagnostique individuelle de l'élève comporte ainsi des informations plus élaborées relatives au diagnostic causal.

3.1 La première phase du test

La première phase a pour objectif d'intégrer l'algorithme "stradaptatif" de sélection des items, programmé en langage Microsoft Pascal™, au système de tests existant et vérifier le fonctionnement du système en utilisant une des banques d'items disponibles.

3.1.1 Le choix de la banque d'items

Nous avons validé le système à partir de la banque d'items et du test

existants, en mode "laboratoire", correspondant à l'objectif de géométrie 5.5.2.: calculer la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre. Nous avons choisi cette banque d'items parce qu'elle comporte le plus grand nombre d'items et de catégories de leurres.

L'organisation des items de la banque de l'objectif 5.5.2 est représentée au tableau I "Cube diagnostique pour l'objectif de géométrie 5.5.2" au chapitre précédent. La création et la disposition des items reposent sur une analyse des pratiques pédagogiques et des pratiques d'évaluation des enseignants impliqués dans le projet de création du système de tests initial et sur l'avis d'experts en didactique des mathématiques. Rappelons que le cube diagnostique comporte trois dimensions : la première dimension est représentée par les différentes notions, la seconde par les différents contextes et la troisième par l'habileté cognitive évaluée par cet objectif. Ainsi, chaque item de la banque mesure l'habileté cognitive définie par l'objectif et est classé dans le cube selon la notion évaluée et le contexte de l'énoncé de l'item (Quesney, 1996). La dimension correspondant à l'habileté cognitive évaluée ne reflète pas nécessairement une dimension unifactorielle, cependant nous pouvons dire qu'elle reflète la pratique pédagogique et d'évaluation des enseignants qui ont participé à l'élaboration du système de tests (Dassa, Ajar, 1995). L'ordre des items d'une même cellule dans le cube diagnostique, c'est-à-dire des items référant à une même notion et à un même contexte pour l'habileté cognitive évaluée, a été établi par les enseignants et les experts en didactique. Les items ainsi ordonnés se répartissent sur un continuum de difficulté du plus facile au plus difficile. Il en est de même pour la hiérarchisation des notions et des contextes. En général, les items appartenant à une notion à gauche du cube sont jugés plus faciles que ceux de droite et les items appartenant à un contexte situé en haut du cube sont jugés plus faciles que ceux en dessous.

Dans le contexte de l'administration du test en mode "laboratoire", nous

avons regroupé certains contextes diagnostiques afin de respecter les contraintes d'ordre pédagogique (de temps, de disponibilité des laboratoires, d'organisation de la classe) et technique liées à certaines limites physiques du système de tests développé. Nous avons considéré les deux notions définies pour l'objectif 5.5.2 et regroupé les 8 contextes (A,C,E,G,I,K,L,M,N) en 5 contextes diagnostics (A,C,E,G-J,K-L-M-N). Par conséquent, la première phase du test peut comprendre entre 15 et 50 items selon le profil de l'élève. Le tableau IX illustre le regroupement des dimensions majeures pour l'objectif de géométrie 5.5.2.

Tableau IX

Regroupement des dimensions majeures de l'objectif de géométrie 5.5.2.

Habileté cognitive évaluée : Calculer la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.

Contenu notionnel	Notion 1: Calcul de la circonférence	Notion 2: Calcul de l'aire
Contextes diagnostiques		
Contexte A: Problèmes abstraits/Mesures identifiées sur une figure.	254 255 256 291 506 513 701 704 707	257 706 258 259 507 705 290 514 702 703 708 901
Contexte C: Problèmes abstraits/Mesures identifiées dans un texte/entiers.	206 208 216 217 219 220 231 234 235 247 248 249 271 285 709 884	(223-237) 225 243 245 296 239 221 299 280 222 (238-201-205) 204 242 203 240 888 244 279 294
Contexte E: Problèmes abstraits/Mesures identifiées dans un texte/décimaux.	207 209 210 218 232 233 246 250 277 733	241 224 236 721 732 (202 720 722) 723
Contexte G-I Problèmes réels/Mesures identifiées sur une figure.	260 502 508 710 712 713 714 715 716 718 711 262 503 512 717 725 727 729	719 (501 726) 511 724 263 728 731 730
Contexte K-L-M-N Problèmes réels/Mesures identifiées dans un texte.	270 275 283 509 910 274 281 282 289 734 902 903 904 251 738 739 740 253	292 276 750 908 736 737 293 295 297 510 735 288 741 742 743 252

Rappelons que dans le contexte de l'administration du test en mode récupération, ces paramètres (techniques et pédagogiques) sont autres et ne s'appliquent plus. Dans ce cas, nous parlons plutôt d'un élève qui chemine à travers une banque d'items avec l'assistance de l'enseignant de façon individuelle ou en petit groupe.

3.1.2 L'algorithme "stradaptatif"

L'ordinogramme, présenté à la figure 3, illustre les principales étapes de la phase I du test: l'identification de l'élève, la sélection des contexte, des notion et des items, la présentation des items, la correction des réponses, la collection des informations, les conditions d'arrêt et les critères d'arrêt.

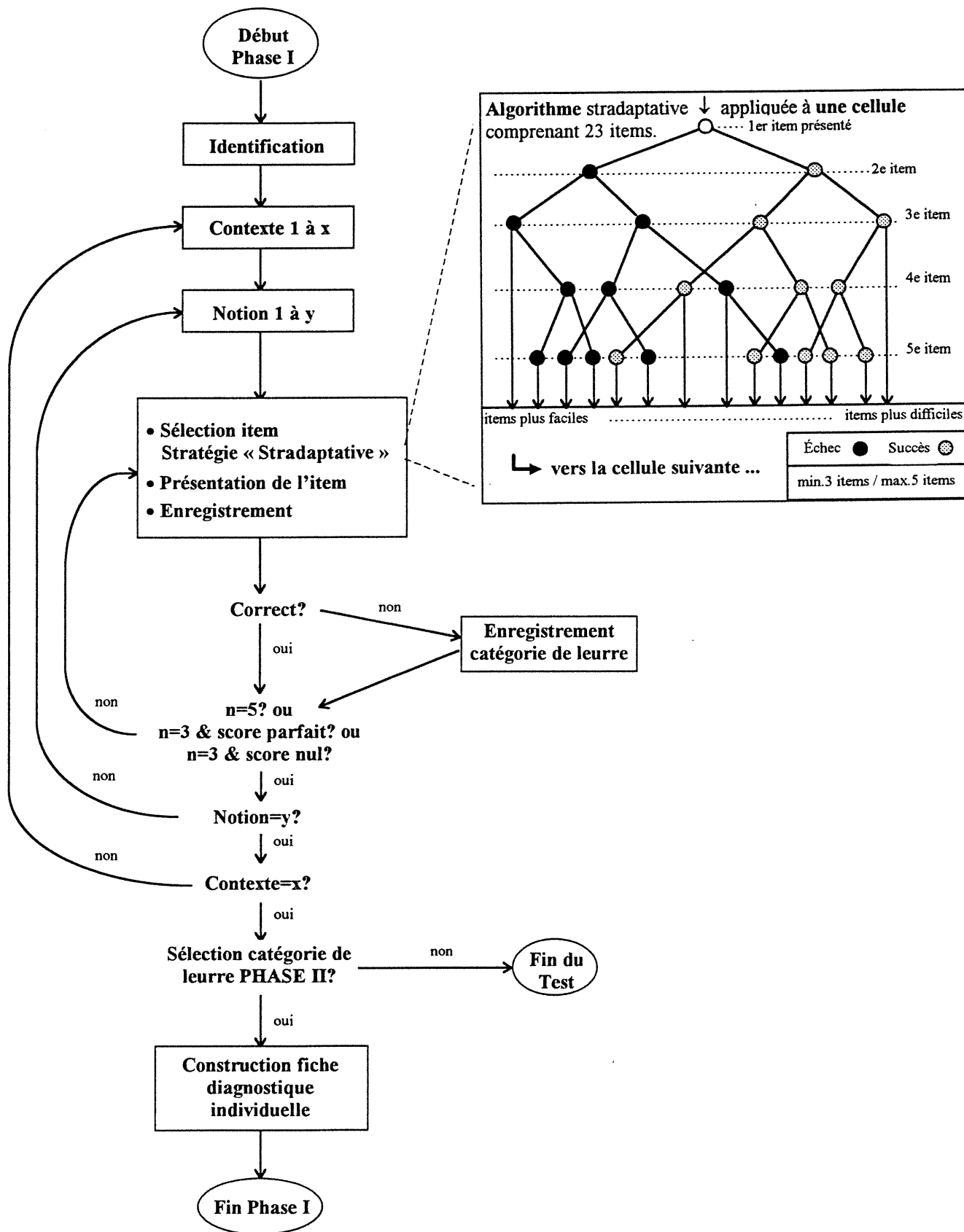


figure 3 - Ordinogramme illustrant la PHASE I du test.

La stratégie de sélection des items

Dans le contexte où les banques d'items ont été développées, la première prise d'information s'effectue par rapport au point milieu d'une cellule définie par le premier contexte, la première notion de ce contexte et l'habileté cognitive évaluée et non à partir d'un niveau général d'habileté de l'élève.

L'algorithme de sélection des items "stradapatif" consiste à présenter un premier item correspondant à l'item milieu d'une cellule ou d'une première facette (tel que structurée par le cube diagnostique d'un objectif donné). Les items subséquents seront sélectionnés selon la réussite ou l'échec de l'élève à l'item précédent. Ainsi, un item plus facile (parmi les items encore disponibles d'une même cellule) sera présenté à l'élève qui a échoué et un item classé plus difficile sera présenté à l'élève qui a réussi l'item précédent. L'algorithme permet de sélectionner l'item situé au milieu d'une cellule ou parmi un ensemble d'items bien délimité (figures 2 et 4 présentées en annexe C) (Lenman, 1994) et ce pour chaque notion et contexte retenus pour le test.

Les cas particuliers

Nous avons prévu, pour des considérations pédagogiques et pour réduire la durée du test, le cas où l'élève échoue systématiquement les trois premiers items relatifs à la première notion d'un contexte donné. Dans ce cas, l'algorithme de sélection des items varie et fait en sorte que les items suivants présentés à l'élève n'appartiennent pas à la notion suivante du contexte couramment testé mais plutôt au contexte suivant. Il nous semblait important de conserver, de la version précédente du système de tests (avec la programmation ad hoc), le parcours particulier où l'élève, ayant des difficultés à résoudre les items d'une notion jugée par les experts et enseignants plus

facile, n'a pas à se voir confronter à des items dont la notion a été jugée plus difficile. Par conséquent, le nombre d'items présenté à l'élève en difficulté sera moindre que pour les autres.

Un problème d'un autre ordre peut survenir lors de l'exécution de la première phase du test sans cependant entraîner l'arrêt. Lorsque le nombre d'items appartenant à une même cellule est réduit au minimum (5 items) et que le patron de réponses (en terme de réussite et d'échec) de l'élève est peu cohérent, l'algorithme peut dégénérer en un phénomène de "*cross over*" documenté par Lenman (1994) (figures 4 et 5 en annexe C), c'est-à-dire que la sélection des items s'effectue parmi les items restants dans la direction d'items plus difficiles ou plus faciles sans considération de la réussite ou de l'échec. Pour limiter ce phénomène, nous avons limité à un seul débordement (ou changement de direction) au-delà de la direction initiale

La programmation de l'algorithme

L'algorithme de sélection des items "stradaptatif" a été programmé à l'aide du langage Pascal (Microsofttm) version 4, compatible avec l'environnement MicroCattm version 4 (ASC, 1989), à partir des notes extraites des documents techniques de référence (Lenman, 1994) et des programmes déjà existants (figure 3 en annexe C). Cet algorithme a été inclus dans un ensemble de programmes externes qui gère une banque d'items différente de celle produite par l'environnement MicroCattm. Cette banque d'items, utilisée par les programmes externes de la phase I, comprend l'ajout d'information relative à la position des items par rapport aux dimensions du cube diagnostique, c'est-à-dire à un contexte, à une notion et au rang de l'item par rapport aux autres items d'une même cellule.

La programmation de la phase I du test nécessite de quitter

temporairement le programme principal géré par le système de gestion de tests MicroCattm, d'exécuter l'algorithme de sélection d'un item par un appel à des programmes externes (la stratégie stradaptative programmée en langage Pascaltm et le programme de gestion de la banque pour le diagnostic) et de revenir au système de gestion MicroCattm pour l'administration de l'item (affichage de l'item, capture de la réponse de l'élève), pour la diffusion des résultats par l'entremise de la fiche diagnostique individuelle à la fin de la première phase et pour l'administration de la seconde phase du test.

Plusieurs modifications concernant les contraintes physiques du système programmé ont été apportées aux programmes initiaux élaborés par Lenman (1994) : le nombre d'items minimum et maximum d'une même cellule, le nombre de notions et de contextes minimum et maximum, le nombre de catégories de leurres pouvant être inclus dans un même test. Nous avons ajouté la procédure de transfert des informations (nom des items, choix de l'élève, réussite ou échec, catégories de leurres et fréquences des catégories de leurres...) nécessaires à la production de la fiche diagnostique individuelle vers l'environnement MicroCattm et modifié la procédure de sélection de la catégorie de leurres la plus fréquente en vue de l'exécution de la seconde phase du test.

La procédure de "sélection de la catégorie de leurres la plus fréquente" intégrée à l'ensemble de programmes externes (langage Pascaltm), a comme fonction de déterminer la catégorie de leurres la plus souvent choisie (leurre spécifique) par l'élève au cours de la phase I du test parmi les catégories de leurres choisies au moins trois fois. S'il y a égalité quant à la fréquence des catégories de leurres, la procédure programmée déterminera aléatoirement la catégorie spécifique. Dans le cas où aucune catégorie de leurres n'a été choisi au moins trois fois par l'élève, la fonction ne retient pas de catégorie de leurres spécifique.

La condition d'arrêt

La première phase du test prend fin lorsque le système a présenté à l'élève des items pour chacune des notions et des contextes préalablement définis dans le test.

3.1.3 Les contraintes et limites

Le système de tests programmé est contraint à respecter certains paramètres : le nombre minimum de notions est de 1 et le nombre maximum est de 10, le nombre minimum de contextes est de 1 et le nombre maximum est de 7, le nombre minimum d'items est de 5 par cellules (une cellule est définie par une même notion, un même contexte, une même habileté cognitive), le nombre de catégories de leurres différents ne peut excéder 50. Lorsque ces paramètres ne sont pas respectés, le système peut ne pas produire toutes les informations dans la fiche de diagnostic individuel, ne pas administrer tous les items des contextes et notions désirés ou refuser d'administrer le test proprement dit. Ce problème survient lorsque l'on ne respecte pas ces paramètres et entraîne un débordement de la mémoire vive utilisée vers des zones physiques du disque dur réservées aux tâches de gestion de l'ordinateur. Il convient de rappeler que l'ensemble du système élaboré fonctionne dans un environnement DOStm et sous le contrôle du système de gestion de tests MicroCattm, qui ne permet pas toujours de gérer adéquatement l'allocation de mémoire additionnelle lors de l'exécution des programmes externes programmés en langage Pascaltm.

3.1.4 La vérification de la phase I du test

Les catégories d'élèves

Pour valider le fonctionnement de l'algorithme "stradaptatif" nous avons effectué une étude de simulation de séances d'administration de la phase I du test pour différents profils d'élèves. Nous avons choisi 6 catégories d'élèves: 1) des élèves ayant atteint une maîtrise parfaite (score égal à 100%), 2) des élèves ayant atteint une maîtrise élevée (score d'environ 90%), 3) des élèves ayant atteint le niveau de maîtrise tel que souvent défini dans le contexte de la pédagogie de maîtrise (score d'environ 80%), 4) des élèves pour lesquels il y a absence de maîtrise (scores d'environ 50%, d'environ 25% et 0%). Ainsi, les élèves fictifs ayant un score égal à 100% ont choisi systématiquement toutes les bonnes réponses des items présentés, tandis que les élèves ayant obtenu un score d'environ 90% auraient choisi 9 fois sur 10 la bonne réponse, les élèves ayant obtenu un score d'environ 80% auraient choisi 4 fois sur 5 la bonne réponse, les élèves ayant obtenu un score d'environ 50% auraient choisi 1 fois sur 2 la bonne réponse, les élèves ayant obtenu un score de 25% auraient choisi 1 fois sur 4 la bonne réponse et les élèves ayant obtenu une note de 0% n'auraient jamais choisi la bonne réponse aux items présentés. La sélection 1 fois sur 4 de la bonne réponse à partir d'items à quatre choix de réponse, est équivalente à un choix de réponse aléatoire. Nous avons considéré 20 élèves pour chaque catégorie de profils différents d'élèves, ce qui constitue un ensemble de 120 élèves.

Le choix de réponse

Pour la réalisation de l'étude de simulation de séances d'administration de la phase I du test, nous avons convenu d'attribuer des choix de réponse des élèves à partir de tables de nombres générés aléatoirement par une fonction

prédéfinie du logiciel SPSS/pc™.

La banque d'items

Contrairement aux tests adaptatifs initiaux (sans l'algorithme de sélection des items mais plutôt avec la programmation ad hoc), nous utilisons l'ensemble des items de la banque (140 items) défini pour l'objectif 5.5.2.

La présentation des résultats

Pour valider la phase I du système de tests, nous comparerons le parcours simulé des élèves (le processus de sélection des items pour chacun des profils d'élèves) pour différentes catégories de maîtrise au parcours attendu par l'algorithme "stradaptatif". A partir de ces résultats, nous produirons des statistiques descriptives tel que le nombre moyen d'items administrés à chaque catégorie.

3.1.5 La communication des résultats de la phase I du test

Toute l'information résultant de l'administration de la première phase du test, en vue de produire la fiche de diagnostic individuel de l'élève à la fin de la séance du test est conservée dans un fichier. En premier lieu, la fiche diagnostique individuelle comporte les mêmes informations produites par le système initial : le nombre d'items réussis, la répartition des items réussis en fonction des contenus notionnels et des contextes diagnostiques présentés à l'élève. La seconde partie de la fiche diagnostique fait état, s'il y a lieu, du numéro et de la fréquence de la catégorie de leurres spécifique retenue, pour la seconde phase du test.

3.2 La seconde phase du test

La seconde phase a pour premier objectif de développer et d'intégrer au système de tests une stratégie adaptative de sélection des items selon une catégorie de leurres donnée (préalablement déterminée à la fin de la phase I). Contrairement au sens habituel donné au θ qui représente l'habileté, selon la stratégie proposée le paramètre θ permet d'estimer le niveau de cohérence de l'élève à choisir une catégorie de leurres donnée. Cette stratégie est basée sur le modèle Rasch à un paramètre (TRI à un paramètre ou indice de difficulté). Un second objectif vise à développer et à communiquer une interprétation qualitative de la valeur du θ calculé.

3.2.1 La constitution des banques d'items spécifiques à une catégorie de leurres donnée

Pour la seconde phase du test, nous avons constitué de nouvelles banques d'items spécifiques aux différentes catégories de leurres (les numéros des catégories sont énumérés au tableau VIII du chapitre 2) associés à la banque d'items correspondant à l'objectif de géométrie 5.5.2. Chaque banque spécifique à une catégorie de leurres donnée a pour caractéristique de regrouper les items de la banque qui comportent un choix de réponse qui relève de cette catégorie.

Pour des considérations liées aux conditions d'arrêt de la phase II (nombre minimal et maximal d'items administrés) du test adaptatif énoncés à la section 3.3.3, nous ne retenons que les banques d'items comportant plus de 20 items spécifiques aux catégories de leurres présentées au tableau X suivant.

Tableau X
Distribution des items parmi les banques
spécifiques aux catégories de leurres retenues

no.de la catégorie de leurres	nombre d'items
02	23
03	63
07	31
08	109
09	51
10	37
12	31

Nous avons construit ces banques en copiant des sous-ensembles d'items provenant de la banque générale de l'objectif 5.5.2. à l'aide d'un programme utilitaire fourni avec le logiciel MicroCattm. De plus nous avons modifié le choix de la bonne réponse par le choix (A,B,C ou D) correspondant à la catégorie de leurres spécifique pour tous les items de ces nouvelles banques.

3.2.2 La calibration des items

Les stratégies adaptatives de sélection des items prédéfinies par MicroCattm, demandent que l'on associe à chaque item des paramètres.

La calibration des items selon une catégorie de leurre donnée n'a pu s'effectuer qu'auprès d'échantillons non-sélectionnés. Le nombre de répondants nécessaires (≈ 250 répondants) à la calibration des items avec 1 paramètre

(modèle le plus simple) n'a pas été atteint lors des mises à l'essai du système de tests initial (1991-1992). Afin de tester l'algorithme, nous avons donc généré l'indice de difficulté des items, en regard de la catégorie de leurres spécifique préalablement choisi, à partir d'une fonction prédéfinie de distribution normale disponible avec le logiciel Spss/pctm.

Nous avons choisi de générer un ensemble de valeurs (égale à 125% du nombre d'items contenu dans la banque) selon une distribution normale pour chacune des banques spécifiques aux catégories de leurres retenus. Nous avons retranché 12.5% des valeurs générées de chaque côté de la distribution afin de ne pas conserver un trop grand nombre d'indices de difficulté représentatifs d'items trop difficiles et trop faciles. La liste des indices de difficulté générés ainsi qu'une représentation graphique de la distribution, avant le retranchement des valeurs extrêmes, pour chacune des banques d'items spécifiques à une catégorie de leurres sont placées à l'annexe D.

Après avoir généré et conservé les valeurs des indices de difficulté, nous avons associé à chaque item un indice de difficulté, selon la position de l'item à l'intérieur du cube diagnostique (tableau I : Cube diagnostique pour l'objectif de géométrie 5.5.2 au chapitre second). Ainsi, l'item situé le plus à gauche et en haut du cube, s'est vu assigné l'indice de difficulté le plus faible et l'item situé le plus en bas et à droite du cube s'est vu assigné l'indice de difficulté le plus élevé. La liste des items avec leur indice de difficulté, pour chacune des banques spécifiques à une catégorie de leurre est présentée à l'annexe E.

La stratégie de sélection des items basée sur le modèle de Rasch à un paramètre (l'indice de difficulté, dénoté "b") utilisée par le logiciel MicroCattm requiert que l'on fixe l'indice de discrimination (dénoté "a") des items à 1 et l'indice attribuable à l'effet du hasard (dénoté "c") à 0 (ASC, 1989, p.119) pour neutraliser leur effet dans le calcul de l'estimation de l'habileté (calibration par

défaut). Nous avons ajouté aux caractéristiques de chaque item, les valeurs de ces paramètres aux items à partir du programme de gestion des banques d'items fourni avec le logiciel MicroCattm.

3.2.3 La stratégie adaptative de sélection des items

L'ordinogramme (figure 4) illustre les principales étapes de la phase II du test: le choix de la banque d'items correspondant à la catégorie de leurres désignée, la méthode de sélection des items selon certaines conditions, les conditions d'arrêt du test et la production de l'information recueillie en diagnostic causal.

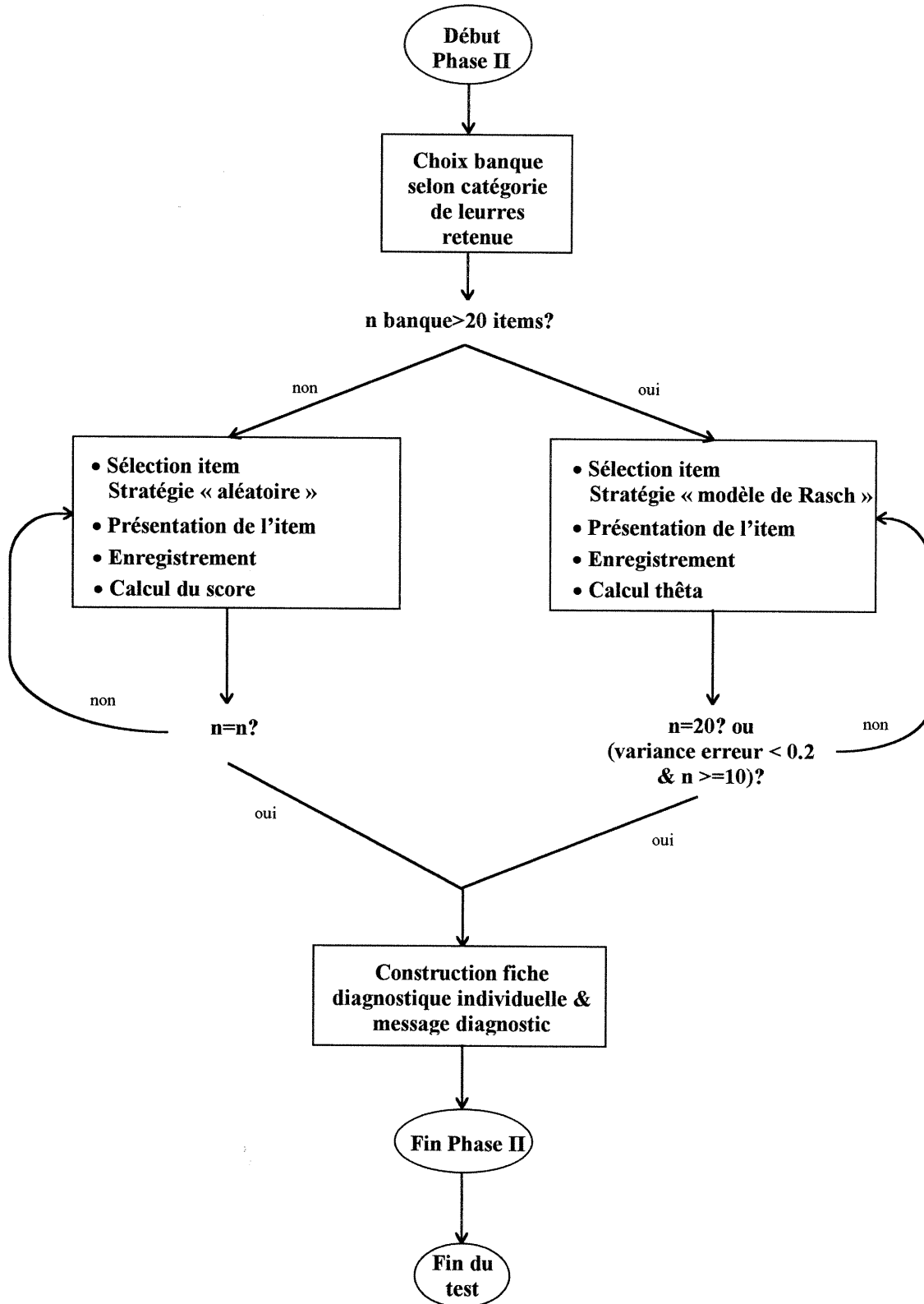


figure 4 - Ordinogramme illustrant la PHASE II du test.

Les conditions préalables

Le système de tests adaptatifs exécute la phase II du test lorsqu'il a détecté des difficultés chez l'élève lors de la première phase du test, c'est-à-dire lorsque l'élève a choisi au moins 3 fois la même catégorie de leurres. Nous rappelons que la catégorie de leurres la plus fréquente est identifiée dans la première phase.

Lorsque l'élève ne présente pas de difficultés à la phase I du test selon les critères préalablement définis, le système n'exécute pas la seconde phase du test et ajoute à la fiche diagnostique individuelle de l'élève le message "aucune hypothèse causale liée à la catégorie de leurres ne peut être émise".

Le système vérifie une autre condition avant d'exécuter la seconde phase du test : la banque d'items spécifiques à la catégorie de leurres doit contenir plus de 20 items (une des conditions d'arrêt du test de la phase II). Lorsque cette condition n'est pas satisfaite, le système administre aléatoirement les items comportant la catégorie de leurres donnée jusqu'à épuisement des items, cumule le score obtenu aux items supplémentaires au score brut de l'élève pour l'ensemble du test et ajoute cette information à la fiche diagnostique individuelle de l'élève.

Lorsque les conditions sont satisfaites, le système procède à l'administration d'items supplémentaires provenant de la banque spécifique à la catégorie de leurres déterminée comme la plus fréquente à la fin de la phase I du test selon une stratégie de sélection des items décrite plus loin. Rappelons que tous les items présentés dans la seconde phase du test comportent la catégorie de leurres spécifique parmi les choix de réponse proposés.

La programmation

La stratégie de sélection des items utilisée dans la seconde phase du test est une stratégie adaptative basée sur le modèle de Rasch à un paramètre déjà prédéfini dans le système de gestion de tests MicroCattm. La phase II du test ne nécessite pas de constituer un ensemble de programmes (langage Pascaltm) externes à MicroCattm pour exécuter la stratégie de sélection des items mais plutôt l'ajout de commandes interprétables par le logiciel MicroCattm au programme principal du système de tests.

La stratégie de sélection des items

Pour rencontrer l'objectif de la seconde phase du test, la programmation (langage MCATLtm) fait en sorte que le système de gestion de l'administration du test MicroCattm présente à l'élève des items qui contiennent toujours la catégorie de leurres donnée (la plus fréquente à la fin de la phase I du test), parmi les choix de réponse proposés. Nous avons défini les caractéristiques des items de la banque de façon à ce que la position de la catégorie de leurres donnée corresponde à la bonne réponse pour le système et qu'à partir de l'indice de difficulté associé à l'item, le système puisse estimer une valeur θ (modèle de Rasch) après chaque réponse de l'élève. Dans le contexte de cette stratégie de sélection des items et de l'utilisation de banque d'items ayant une catégorie de leurres spécifique, nous convenons que la valeur estimée de θ ne correspond pas au niveau d'habileté de l'élève à choisir la bonne réponse mais plutôt à un niveau de difficulté de l'élève relatif à la catégorie du leurre donné.

Nous rappelons que le modèle de Rasch pose que le choix de réponse à un item est un score dichotomique (réussite ou échec) et postule que la

probabilité d'observer une bonne réponse à un item est fonction du niveau d'habileté de l'élève et du niveau de difficulté de l'item (ASC, 1989). Ainsi, après chaque réponse de l'élève, la sélection de l'item suivant est déterminée par la fonction d'information la plus élevée par rapport au niveau d'habileté courant de l'élève.

Les conditions d'arrêt

La phase II du test se termine lorsqu'une des deux conditions suivantes est satisfaite : une variance erreur plus petite que 0.20 et au moins 10 items administrés ou un maximum de 20 items. A la fin de la phase II du test, le système cumule au fichier des résultats de la phase I le message diagnostique relatif au niveau de cohérence de l'élève à choisir la catégorie de leurres spécifique identifiée après la phase II du test. De plus, le système fait apparaître à l'écran les résultats de l'ensemble du test (fiche diagnostique individuelle) de l'élève.

Toute la phase II du test s'exécute sous la gestion du programme MicroCattm : l'administration des items, le calcul du θ , la communication des résultats et du message diagnostique informant l'enseignant du niveau de cohérence de l'élève à choisir une catégorie de leurres donnée.

3.2.4 L'étude de simulations de la phase II du test

L'étude de simulation vise à définir des bornes d'interprétation des valeurs θ obtenues à la fin de la phase II du test. Les résultats de l'étude de simulation permettront de formuler pour chaque élève des messages diagnostiques qualitatifs exprimant le degré de difficulté présenté par une catégorie de leurres donnée. Ces messages, destinés aux enseignants, sont formulés simplement. Ils synthétisent une information probabiliste relative à une

mesure du niveau de cohérence de l'élève à choisir une catégorie de leurres donnée.

Les catégories d'élèves

L'étude de simulation a été réalisée à partir de différentes catégories d'élèves (100 élèves par catégorie). Nous postulons que les élèves fictifs présentent tous une difficulté préalablement identifiée par le système lors de la phase I du test, difficulté qui relève d'une catégorie de leurres la plus fréquemment choisie par l'élève.

Nous avons défini cinq catégories d'élèves : 1) des élèves ayant choisi la catégorie de leurres donnée à tous les items, 2) des élèves ayant choisi 3 fois sur 4 la catégorie de leurres donnée, 3) des élèves ayant choisi 1 fois sur 2 la catégorie de leurres donnée, 4) des élèves ayant choisi aléatoirement une réponse aux items et 5) des élèves n'ayant jamais choisi la catégorie de leurres donnée. Nous considérons que la sélection aléatoire du choix de réponse à un item est équivalente au choix 1 fois sur 4 la catégorie de leurres donnée.

Les banques d'items

Pour les fins de la simulation, nous avons retenu les banques d'items spécifiques aux catégories de leurres 8 et 2 parmi l'ensemble des catégories de leurres correspondant à l'objectif de géométrie 5.5.2 présenté au tableau XI.

Tableau XI

**Distribution des items parmi les banques
spécifiques aux catégories de leurres retenues**

no.de la catégorie de leurres	Description de la catégorie de leurre (cause probable de l'erreur)	nombre d'items
*02	Élève au carré le diamètre au lieu du rayon dans le calcul de l'aire.	23
03	Calcule la circonférence au lieu de l'aire.	63
07	Double le diamètre inutilement dans le calcul de la circonférence.	31
*08	Place la virgule décimale à la mauvaise position.	109
09	Calcule l'aire au lieu de la circonférence.	51
10	Oublie de doubler le rayon dans le calcul de la circonférence.	37
12	Multiplie le rayon par pi au lieu de calculer l'aire.	31

Ces banques spécifiques contiennent le plus grand nombre (catégorie de leurres #8, 109 items) et le plus petit nombre d'items (catégorie de leurres #2, 23 items) parmi celles comportant suffisamment de leurres pour être admissibles à la phase II du test.

Le choix de réponse

Pour les fins de l'étude de simulation, nous avons attribué le choix de réponse des élèves à partir de tables de nombres générés aléatoirement par une fonction prédéfinie du logiciel SPSS/pc™.

La présentation des résultats

Dans un premier temps, nous avons considéré séparément les résultats

obtenus à partir des simulations utilisant la banque d'items spécifique aux catégories de leurres #2 et #8. En second lieu, nous comparerons et commenterons l'ensemble des résultats obtenus.

Pour déterminer les niveaux de cohérence du choix de l'erreur, nous avons calculé des intervalles de confiance pour le θ moyen estimé pour chaque catégorie d'élèves. À partir de ces intervalles, nous avons déterminé des points de césures séparant les niveaux de cohérence.

Finalement, nous avons défini des messages diagnostiques synthétisant l'interprétation des valeurs θ moyennes obtenues.

3.2.5 La communication des résultats de la phase II du test

L'interprétation qualitative de la valeur du θ finale estimée (exprimant le niveau de difficulté cognitive de l'élève par rapport à une catégorie de leurres donnée) a été ajoutée à la fiche diagnostique individuelle de l'élève. À la fin de la phase II du test, le système affiche à l'écran la totalité de la fiche diagnostique individuelle de l'élève. Cette fiche est aussi conservée dans un fichier. Le système de tests diagnostiques cumule toutes les fiches diagnostiques individuelles produites à partir d'un même ordinateur dans un même fichier (ne s'applique pas lorsque le système est utilisé en réseau).

3.3 Le système de tests diagnostiques intégrant les phases I et II

3.3.1 Le fonctionnement général du système de tests diagnostiques

Le système de tests diagnostiques développé procède en deux phases de tests au cours d'une même séance. Lorsque les résultats de la première phase du test remplissent les conditions préalables à la seconde phase, le système procède à l'exécution de la seconde phase du test. A la fin du test, l'élève voit s'afficher à l'écran ses résultats complets (phases I et II) et le système cumule dans un fichier toutes les informations recueillies lors des séances de tests en vue d'une impression ou d'une consultation ultérieure des fiches diagnostiques.

3.3.2 L'information diagnostique produite

Selon le profil de l'élève, la fiche diagnostique individuelle pourra contenir: 1) les informations de la phase I du test seulement, 2) les informations de la phase I incluant les résultats obtenus aux items supplémentaires administrés contenant une catégorie de leurres spécifique mais en nombre insuffisant pour que le système procède à la phase II du test ou 3) les informations de la phase I du test et le message diagnostique relatif au niveau de cohérence de l'élève à choisir une catégorie de leurres donnée déterminée à partir de l'estimation du θ final à l'aide du modèle Rasch basée de la TRI.

3.3.3 Les contraintes du système

À cause de contraintes physiques, le système ne s'applique qu'à un maximum de 10 notions et de 7 contextes diagnostiques pour la phase I du test. La phase II du système de tests actuel est limité à la confirmation du niveau de cohérence de l'élève relevant d'une même catégorie de leurre présente dans au moins 20 items. L'étude de simulations a été réalisée seulement à partir des items de l'objectif 5.5.2 de géométrie et que pour les catégories de leurres #2 et #8 pour ce qui est de la phase II du test.

Chapitre quatrième:

LES RÉSULTATS

Ce chapitre présente les résultats de l'étude de vérification et des études de simulations du système de tests adaptatifs développés. Les limites du système réalisé font l'objet de commentaires à la fin du chapitre.

4.1 La première phase du test

Le système de tests diagnostiques développé comprend un algorithme "stradaptatif" de sélection des items intégré à la première phase du test. Celle-ci produit une fiche diagnostique individuelle de l'élève contenant un diagnostic structurel et une hypothèse du diagnostic causal. Le système détermine à la fin de la première phase la catégorie de leurres la plus souvent choisie par l'élève qui est alors considérée comme le point de départ de la seconde phase du test. Pour vérifier le fonctionnement de l'algorithme "stradaptatif" de sélection des items utilisé dans la première phase du test, nous avons procédé à une étude de vérification des parcours d'élèves.

4.1.1 Les objectifs

L'étude de vérification telle que définie au chapitre précédent, avait pour but de contrôler le fonctionnement de l'algorithme "stradaptatif" de sélection des items intégré au système de tests diagnostiques, en utilisant la banque d'items correspondant à l'objectif de géométrie 5.5.2, pour différentes catégories de maîtrise données .

4.1.2 Les résultats de l'étude de vérification

Pour répondre à cet objectif, nous avons comparé les parcours des différents profils d'élèves enregistrés dans les fiches diagnostiques produites à la fin de la phase I, à la stratégie "stradaptative" de sélection des items prévue. Nous considérons le parcours d'un élève comme étant la séquence

d'items auxquels il a répondu en relation avec la réussite ou l'échec à chaque item.

Le tableau, placé à l'annexe F, présente la compilation de l'étude de vérification de l'algorithme de sélection des items pour différentes catégories de maîtrise. Il comporte les informations suivantes : le nombre d'items et le score individuel de chaque élève regroupés sous une catégorie de maîtrise, le nombre moyen d'items présentés aux élèves par catégorie de maîtrise et le score moyen obtenu par les élèves d'une même catégorie.

En résumé, pour l'ensemble des parcours simulés d'élèves, les élèves se classant au niveau "maîtrise" ont répondu à un plus grand nombre d'items soit en moyenne 37 items. Notons que le système vérifiait trois conditions pour déterminer la sélection de l'item suivant. La première condition consistait à vérifier si les trois premiers items (consécutifs) d'une même cellule (contexte et notion donnés) ont été réussis. Ainsi, un élève ayant réussi l'ensemble des items, a répondu à un nombre fixe d'items soit 30 items (3 items par cellules). La seconde condition vérifiait si les trois premiers items d'une même cellule ont été échoués. Ainsi les élèves qui échouaient la totalité des items présentés, ont répondu au minimum d'items soit 3 items pour la première notion seulement de chacun des contextes (15 items au total). La troisième condition assurait de ne pas dépasser cinq items par cellule. Ainsi les élèves ne réussissaient pas ou n'échouaient pas systématiquement les 3 premiers items d'une même cellule devaient répondre à 5 items par cellules si le nombre d'items le permettait. Le nombre d'items maximum pouvant être présenté à un élève est de 50 items soit 5 items par cellules.

Les différentes catégories de maîtrise

Les différentes catégories de maîtrise s'inspirent du contexte de la pédagogie de maîtrise. Ainsi, la catégorie "Maîtrise" correspond à un score d'environ 80%. Nous avons défini deux catégories supérieures à la "Maîtrise", soit la catégorie "Maîtrise élevée" et la catégorie "Maîtrise parfaite" qui correspond à un score de 100%. Il nous paraissait intéressant de subdiviser la catégorie inférieure à la "Maîtrise", soit la catégorie "Absence de maîtrise" en trois sous-catégories à partir du critère du score attendu : environ 50%, environ 25% et 0%. La catégorie "Absence de maîtrise avec un score d'environ 50%" représente un score de passage minimum où l'élève donne environ 1 fois sur 2 la bonne réponse. La catégorie "Absence de maîtrise avec un score d'environ 25%" peut représenter une sélection aléatoire d'un choix de réponses à partir d'items à 4 choix de réponses ou choisir 1 fois sur 4 la bonne réponse. La catégorie "Absence de maîtrise avec un score nul" représente l'élève qui n'a réussi aucun item.

Le nombre de parcours simulés

Vingt parcours d'élèves ont été simulé pour chaque catégorie de maîtrise à l'exception des catégories "Maîtrise parfaite" et "Absence de maîtrise avec un score égal à 0". Pour ces deux dernières catégories, il n'existe qu'un parcours d'élève possible à l'intérieur d'une même banque d'items.

L'assignation d'un choix de réponse

Pour assigner un choix de réponse à chaque item présenté à un élève lors de la simulation, nous avons établi différentes stratégies de sélection du choix de réponse qui utilisent toutes une table de nombres aléatoires différents selon le score attendu. Ainsi, par exemple, pour obtenir la fiche diagnostique

individuelle d'un élève appartenant à la catégorie "Maîtrise" avec un score d'environ 80%, la stratégie adoptée est d'assigner une bonne réponse 4 fois sur 5. Pour y parvenir, nous avons utilisé une table de nombres aléatoires compris entre 1 et 5 et assigné une bonne réponse à l'item lorsque le nombre de la table correspondait à 1, 2, 3 ou 4 ou une mauvaise réponse à l'item lorsque le nombre de la table correspondait au nombre 5.

Exemples de parcours

Nous examinons en détails les parcours simulés du premier élève de chaque sous-catégorie de maîtrise lors de la phase I du système de tests diagnostiques développé. Les fiches diagnostiques individuelles de ces élèves sont présentées à l'annexe G.

Le système de tests auquel a été intégré l'algorithme "stradaptatif" de sélection des items, a présenté, à l'élève appartenant à la catégorie "Maîtrise parfaite" (score = 100%), trois items par cellules pour toutes les notions et tous les contextes. Le nombre d'items présenté est de 30 items soit un nombre égal au nombre d'items attendu : le minimum de 3 items par cellule pour les 2 notions de chacun des 5 contextes. Le parcours de l'élève fictif présenté au tableau XII correspond en tout point au parcours prévu par l'algorithme "stradaptatif".

Les parcours des élèves no.1 appartenant aux catégories "Maîtrise élevée", "Maîtrise", "Absence de maîtrise" avec un score d'environ 50% et avec un score d'environ 25% sont présentés aux tableaux XIII, XIV, XV et XVI.

Le parcours de l'élève no.1 appartenant à la catégorie "Maîtrise élevée" illustre que l'élève a répondu correctement aux items correspondant aux 9

premières cellules et a échoué un item appartenant à la notion 2 du contexte K-L-M-N pour un total de 32 items. Les items présentés provenaient de tous les contextes et notions planifiés à être évalués.

Le parcours de l'élève no.1 appartenant à la catégorie "Maîtrise" a répondu à 40 items au total. Le nombre d'items présenté à cet élève est supérieur à l'ensemble des autres parcours présentés car rappelons que lorsque les trois premiers items d'une cellule donnée ne sont pas réussis ou échoués systématiquement, le système développé continue à présenter des items jusqu'à concurrence de 5 items. Les items présentés provenaient de tous les contextes et notions planifiés à être évalués.

Les parcours des élèves no.1 appartenant aux catégories "Absence de maîtrise" avec un score d'environ 50% et 25% ont répondu respectivement à 34 items et 31 items au total. Le nombre d'items présenté par cellule variait entre 3 et 5 items. Les items présentés ne provenaient pas de tous les contextes et notions planifiés à être évalués car rappelons que lorsque les 3 premiers items de la première notion d'un contexte évalué sont échoués systématiquement, la sélection des items suivant ne s'effectue pas pour la notion suivante classée plus difficile mais plutôt pour la première notion du contexte suivant.

Le parcours de l'élève no.1 présenté au tableau XVII appartenant à la catégorie "Absence de maîtrise" avec un score nul (0%) correspond en tous points au cheminement prévu par l'algorithme. Le nombre d'items administrés n'est que de 15 items. Tel que prévu, lorsque l'élève échoue tous les items de la première notion pour un contexte donné, le système ne lui présente pas des items appartenant à la notion suivante jugée par les experts plus difficile. Il choisit plutôt les prochains items parmi ceux du contexte suivant et ce pour la notion jugée la plus facile.

En résumé, les élèves de niveau "Maîtrise" se voient présentés le plus grand nombre d'items et seuls les élèves ayant le plus de difficultés ne sont pas évalués à partir d'items correspondant à la notion classée plus difficile d'un contexte donné. Le nombre d'items présentés varie entre 15 et un maximum de 50 items selon le niveau de réussite de l'élève aux items. L'information obtenue présentée à la suite de la phase I du test indique le nombre d'item réussis selon les notions, les contextes évalués et comptabilise les catégories de leurs choisies.

L'étude de vérification des parcours d'élèves appartenant aux différentes catégories de maîtrise a permis de vérifier le fonctionnement attendu de l'algorithme "stradaptatif" de sélection des items programmé et intégré au système de tests diagnostiques. Nous constatons que tous les parcours des élèves appartenant aux différentes catégories de maîtrise ont été vérifiés et que le système se comporte dans tous les cas dans les limites attendues de l'algorithme "stradaptatif".

Tableau XII
Parcours de l'élève no.1 de la catégorie "Maîtrise parfaite"

Habilité cognitive évaluée : Calculer la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.

Contenu notionnel	Notion 1: Calcul de la circonférence	Notion 2: Calcul de l'aire
Contextes diagnostiques		
Contexte A: Problèmes abstraits/Mesures identifiées sur une figure.	254 255 256 291 506 513 701 704 707 DÉBUT →	257 706 258 259 507 705 290 514 702 703 708
Contexte C: Problèmes abstraits/Mesures identifiées dans un texte/entiers.	206 208 216 217 219 220 231 234 235 247 248 249 271 285 709 884	(223-237) 225 243 245 286 239 221 299 280 222 (238-201-205) 204 242 203 240 888 244 279 294
Contexte E: Problèmes abstraits/Mesures identifiées dans un texte/décimaux.	207 209 210 218 232 233 246 250 277 733	241 224 236 721 732 (202 720 722) 723
Contexte G-I Problèmes réels/Mesures identifiées sur une figure.	260 502 508 710 712 713 714 715 716 718 711 262 503 512 717 725 727 729	719 (501 726) 511 724 263 728 731 730
Contexte K-L-M-N Problèmes réels/Mesures identifiées dans un texte.	270 275 283 509 910 274 281 282 289 734 902 903 904 251 738 739 740 253	292 276 750 908 736 737 293 295 297 510 735 288 741 742 743 252

FIN

Tableau XIII

Parcours de l'élève no.1 de la catégorie "Maîtrise élevée"

Habilité cognitive évaluée : Calculer la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.

Contenu notionnel	Notion 1: Calcul de la circonférence DÉBUT	Notion 2: Calcul de l'aire
Contextes diagnostiques		
Contexte A: Problèmes abstraits/Mesures identifiées sur une figure.	254 255 256 291 506 513 701 704 707	257 706 258 259 507 705 290 514 702 703 708
Contexte C: Problèmes abstraits/Mesures identifiées dans un texte/entiers.	206 208 216 217 219 220 231 234 235 247 248 249 271 285 709 884	(223-237) 225 243 245 295 239 221 (299) 280 222 (238-201-205) 204 242 203 240 888 244 279 (294)
Contexte E: Problèmes abstraits/Mesures identifiées dans un texte/décimaux.	207 209 210 218 232 233 246 250 277 733	241 224 236 721 732 202 720 722 723
Contexte G-I Problèmes réels/Mesures identifiées sur une figure.	260 502 508 710 712 713 714 715 716 718 711 262 503 512 717 725 727 729	718 (501 726) 511 724 263 728 731 730
Contexte K-L-M-N Problèmes réels/Mesures identifiées dans un texte.	270 275 283 509 910 274 281 282 289 734 902 903 904 251 738 739 740	292 276 750 908 736 737 293 295 297 510 35 288 741 742 743 252

FIN

Tableau XIV

Parcours de l'élève no.1 de la catégorie "Maîtrise"

Habilité cognitive évaluée : Calculer la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.

Contenu notionnel	Notion 1: Calcul de la circonférence	Notion 2: Calcul de l'aire
Contextes diagnostiques	DÉBUT	
Contexte A: Problèmes abstraits/Mesures identifiées sur une figure.	254 255 256 291 506 513 701 704 707 	257 706 258 259 507 705 290 514 702 703 708
Contexte C: Problèmes abstraits/Mesures identifiées dans un texte/entiers.	206 208 216 217 219 220 231 234 235 247 248 249 271 285 709 884 	(223-237) 225 243 244 296 239 221 289 280 222 (238-201-205) 204 242 203 240 888 244 279 294
Contexte E: Problèmes abstraits/Mesures identifiées dans un texte/décimaux.	207 209 210 218 232 233 246 250 277 733 	241 224 236 721 732 202 720 722 723
Contexte G-I Problèmes réels/Mesures identifiées sur une figure.	260 502 508 710 712 713 714 715 718 718 711 262 503 512 717 725 727 729 	719 501 726 511 724 263 728 731 730
Contexte K-L-M-N Problèmes réels/Mesures identifiées dans un texte.	270 275 283 509 910 274 281 282 289 734 902 903 904 251 738 739 740 253 	292 276 750 908 736 737 293 295 297 510 735 288 241 742 743 252

FIN

Tableau XV

Parcours de l'élève no.1 de la catégorie "Absence de maîtrise" avec un score d'environ 50%

Habileté cognitive évaluée : Calculer la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.

Contenu notionnel	Notion 1: Calcul de la circonférence	Notion 2: Calcul de l'aire
Contextes diagnostiques		
Contexte A: Problèmes abstraits/Mesures identifiées sur une figure.	<p>254 255 256 291 506 513 701 704 707</p>	257 706 258 259 507 705 290 514 702 703 708 901
Contexte C: Problèmes abstraits/Mesures identifiées dans un texte/entiers..	<p>206 208 216 217 219 220 231 234 235 247 248 249 271 285 709 884</p>	(223-237) 225 243 245 296 239 221 299 280 222 (238-201-205) 204 242 203 240 888 244 279 294
Contexte E: Problèmes abstraits/Mesures identifiées dans un texte/décimaux.	<p>207 209 210 218 232 233 246 250 277 733</p>	241 224 236 721 732 (202 720 722) 723
Contexte G-I Problèmes réels/Mesures identifiées sur une figure.	<p>260 502 508 710 712 713 714 715 716 718 711 262 503 512 717 725 727 729</p>	719 (501 726) 511 724 263 728 731 730
Contexte K-L-M-N Problèmes réels/Mesures identifiées dans un texte.	<p>270 275 283 509 510 274 281 282 289 734 902 903 904 251 738 739 740 253</p>	292 276 750 908 736 737 293 295 297 510 735 288 741 742 743 252

FIN

Tableau XVI
Parcours de l'élève no.1 de la catégorie "Absence de maîtrise"
 avec un score d'environ 25%

Habileté cognitive évaluée : Calculer la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.

Contenu notionnel	Notion 1: Calcul de la circonférence	Notion 2: Calcul de l'aire
Contextes diagnostiques		
Contexte A: Problèmes abstraits/Mesures identifiées sur une figure.	254 255 256 257 508 513 701 704 707 	257 706 258 259 507 705 290 514 702 703 708 901
Contexte C: Problèmes abstraits/Mesures identifiées dans un texte/entiers.	206 208 216 217 219 220 231 234 235 247 248 249 271 285 709 884 	(223-237) 225 243 245 296 339 221 299 280 222 (238-201-205) 204 242 203 240 888 244 279 294
Contexte E: Problèmes abstraits/Mesures identifiées dans un texte/décimaux.	207 209 210 218 232 233 246 250 277 733 	241 224 236 721 732 (202 720 722) 723
Contexte G-I Problèmes réels/Mesures identifiées sur une figure.	260 502 508 710 712 713 714 715 716 718 711 262 503 512 717 725 727 729 	719 (501 726) 511 724 263 728 731 730
Contexte K-L-M-N Problèmes réels/Mesures identifiées dans un texte.	270 275 283 289 310 274 281 282 289 734 902 903 904 251 738 739 740 253 	292 276 250 908 736 737 293 295 297 510 735 788 741 742 743 252

FIN

Tableau XVII

Parcours de l'élève no.1 de la catégorie "Absence de maîtrise"

avec un score de 0%

Habilité cognitive évaluée : Calculer la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.

Contenu notionnel	Notion 1: Calcul de la circonférence	Notion 2: Calcul de l'aire
Contextes diagnostiques		
Contexte A: Problèmes abstraits/Mesures identifiées sur une figure.		257 706 258 259 507 705 290 514 702 703 708 901
Contexte C: Problèmes abstraits/Mesures identifiées dans un texte/entiers.		(223-237) 225 243 245 296 239 221 299 280 222 (238-201-205) 204 242 203 240 888 244 279 294
Contexte E: Problèmes abstraits/Mesures identifiées dans un texte/décimaux.		241 224 236 721 732 (202 720 722) 723
Contexte G-I Problèmes réels/Mesures identifiées sur une figure.		719 (501 726) 511 724 263 728 731 730
Contexte K-L-M-N Problèmes réels/Mesures identifiées dans un texte.		292 276 750 908 736 737 293 295 297 510 735 288 741 742 743 252

FIN

Le nombre moyen d'items

Le système de tests diagnostiques est un système de tests adaptatifs qui prend en considération la réponse de l'élève pour sélectionner le prochain item (un item plus facile ou un item plus difficile) non utilisé. Nous ne pouvons pas connaître à priori le nombre d'items qui sera présenté à l'élève mais l'étude de vérification nous a permis d'obtenir un nombre moyen d'items pour chacune des différentes catégories de maîtrise (voir tableau XII à l'annexe F). Ainsi les catégories intermédiaires (maîtrise élevée, maîtrise et absence de maîtrise avec un score de 50%) présentent en moyenne plus d'items à l'élève que pour les élèves catégories extrêmes (maîtrise parfaite et absence de maîtrise avec des scores d'environ 25% et 0%). Le nombre d'items présenté à l'élève est tributaire de l'alternance entre la réussite ou l'échec à un item et des critères même inclus dans l'algorithme de sélection soit un minimum de 3 items et un maximum de 5 items par cellule. Le nombre moyen d'items administrés nous indique qu'il faudrait prévoir une durée d'administration de la phase I du test plus grande que celle de 30 minutes initialement prévue. Rappelons que la durée est fonction du nombre de notions et de contextes du cube diagnostique qui sous-tend la banque utilisée. En pratique, il suffirait de réduire le nombre de contextes évalués au cours d'une même séance d'administration pour respecter les contraintes horaires de l'école.

4.1.3 Une interprétation des résultats

Pour des considérations organisationnels et pédagogiques (nb appareils disponibles, disponibilité des ordinateurs situés dans un laboratoire informatique à utilisation restreinte, durée des périodes de cours), l'administration de la première phase du système de tests telle que réalisée (stratégie "stradaptative", totalité de la banque d'items de l'objectif) devrait se dérouler durant une période de temps plus grande que les 30 minutes prévues pour la version initiale du

test.

Les résultats de vérification des parcours d'élèves appartenant à différentes catégories de profils s'avèrent correspondre au cheminement prévu par l'algorithme "stradaptatif". On peut donc conclure que la programmation contrôle la sélection des items de la façon prévue.

Rappelons que les travaux initiaux concernant la programmation de la stratégie "stradaptative" de sélection des items effectués par Lenman (1994), ont utilisé un nombre réduit d'items (16 items) provenant de la banque évaluant l'objectif de géométrie 5.5.8. Le fait que l'étude de vérification s'est faite à partir d'une autre banque (relevant de l'objectif de géométrie 5.5.2) constitue un élément de preuve de l'indépendance de l'algorithme par rapport à la banque d'items utilisée.

4.2 La seconde phase du test

La seconde phase du test sert à confirmer le diagnostic causal avancé à la fin de la première phase. Dans un premier temps, le système développé administre des items qui comportent tous une même catégorie de leurres (identifiée à la phase I du test) et calcule à partir d'un modèle de Rasch, une valeur de θ après chaque réponse de l'élève à un item. La valeur finale du θ correspond au niveau de cohérence de l'élève à choisir une catégorie de leurres donnée. L'étude de simulations nous a permis de calculer des bornes d'interprétation des valeurs θ obtenues. Nous pouvons ainsi ajouter à la phase II du test des messages qualitatifs résumant le niveau de cohérence de l'élève à choisir une catégorie de leurres destinée à guider l'enseignant dans ses interventions pédagogiques.

4.2.1 Les objectifs

L'étude des simulations décrite au chapitre précédent avait pour but de déterminer des normes permettant d'interpréter les valeurs θ obtenues à la fin de la phase II du test et de calculer leur degré de précision.

4.2.2 Les résultats de l'étude de simulations

Les simulations ont été réalisées à partir des banques d'items spécifiques aux catégories de leurres 2 et 8. Nous avons postulé que tous les élèves présentaient une difficulté cognitive relative aux catégories de leurres 2 ou 8 identifiées à la fin de la phase I du test.

Les différentes catégories de profils d'élèves

Nous avons défini trois catégories de profils d'élèves : 1) des élèves qui répondent aléatoirement ou choisissent 1 fois sur quatre la catégorie de leurres donnée, 2) des élèves qui choisissent 1 fois sur deux la catégorie de leurres donnée et 3) ceux qui choisissent 3 fois sur 4 la catégorie de leurres donnée. Nous avons vérifié aussi les valeurs estimées de θ pour un élève qui choisirait toujours la catégorie de leurres donnée et un élève qui ne la choisirait jamais au cours de la phase II du test.

Le nombre de profils d'élèves

Afin d'établir les caractéristiques de base des distributions des θ ainsi que les intervalles de confiance pour ces θ , nous avons procédé à la simulation de 100 profils d'élèves pour chacune des trois catégories et pour chacune des deux banques d'items spécifiques (catégories de leurres 2 et 8).

L'assignation d'un choix de réponse

Nous avons établi différentes stratégies de sélection du choix de réponse correspondant aux différents profils des élèves. Comme pour la première phase du test, nous nous sommes servis de tables de nombres entiers aléatoires différentes selon la stratégie. Ainsi, lorsque nous voulions simuler le comportement d'un élève qui répondrait au hasard, nous avons utilisé une table de nombres entre 1 et 4 et assigné le choix de réponse correspondant à la catégorie de leurres donnée lorsque le nombre était 1 et un autre choix de réponse pour les nombres 2,3 et 4. Pour simuler un élève qui répond dans un proportion de trois fois sur quatre la catégorie de leurres donnée, nous avons utilisé la même table de nombres mais nous avons fait correspondre les nombres 1, 2 et 3 au choix de réponse contenant la catégorie de leurres

donnée. Pour simuler un élève qui répond dans une proportion d'une fois sur deux la catégorie de leurres donnée, nous avons utilisé une table de nombres entiers aléatoires contenant 0 ou 1 et assigné le nombre 1 au choix de réponse comportant la catégorie de leurres donnée.

Les résultats des simulations

L'ensemble des résultats individuels regroupés selon les catégories d'élèves pour chaque banque d'items utilisée est placé à l'annexe G. Nous présentons au tableau XVII, la valeur moyenne des θ ainsi que le nombre moyen d'items pour chacune des catégories de profils d'élèves et pour chaque banque d'items étudiées.

Tableau XVIII
Compilation des valeurs moyennes de l'étude de simulations PHASE II pour les banques d'items spécifiques aux catégories de leurres 2 et 8.

no. catégorie de leurres	nb total items	Stratégie de sélection de la catégorie de leurres								
		Aléatoire			1 / 2			3 / 4		
		nb moy items	θ moy	s moy	nb moy items	θ moy.	s moy.	nb moy. items	θ moy.	s moy.
2	23	11.9	-0.99	0.180	10.2	-0.11	0.166	12.8	1.19	0.175
8	140	13.3	-1.61	0.178	13.1	-0.07	0.159	10.2	1.42	0.018

Les valeurs moyennes des θ représentent les scores types résultant des diverses contraintes imposées par le test (le nombre d'items dans la banque spécifique à la catégories de leurres, le nombre maximal d'items administrés, le seuil minimal de la variance erreur à atteindre pour terminer la phase II du test). Le score θ calculé par Microcattm utilise la méthode de maximum de la vraisemblance (ASC, 1989). Le score θ obtenu par un élève est un score

observé qui constitue une estimation de son score "vrai". Afin d'estimer la précision de cette estimation, un intervalle de confiance a été calculé pour toutes les catégories d'élèves et pour chaque banque d'items spécifique aux catégories de leurres 2 et 8.

Calcul des intervalles de confiance

Le calcul de l'intervalle de confiance pour le vrai score θ autour du score observé représenté par la moyenne des θ pour chacune des catégories de profils d'élèves et pour chacune des banques d'items (catégories de leurres 2 et 8) a été obtenu à partir de la formule générale (Feldt & Brennan, 1991) donnée par :

$$\theta \pm z \text{ MSE}$$

où θ est le θ moyen d'une catégorie d'élèves
 z est la valeur correspondant au niveau de probabilité désiré
 MSE est une estimation de l'erreur standard de mesure calculée à l'aide de la simulation

Dans un contexte d'évaluation formative et de gestion pédagogique, il nous apparaît acceptable de choisir un niveau de probabilité égal à 80%, cependant pour des considérations liées à l'établissement de normes, nous avons jugé bon de choisir un niveau de probabilité à 90% et en postulant la normalité.

À partir des simulations, nous obtenons une estimation de la variance erreur plutôt que l'erreur standard. La formule générale devient alors pour un niveau de probabilité fixé à 90% :

$$\theta \pm 1.64\sqrt{s}$$

où θ est le θ moyen d'une catégorie d'élèves
 z est la valeur correspondant au niveau de probabilité à 90% (courbe normale)
 s est une estimation de la moyenne des variances erreur des θ calculée à l'aide de la simulation

Le tableau XIX comprend les valeurs utilisées dans les calculs et les intervalles de confiance pour les valeurs des θ moyens résultants d'un profil type d'élève, pour chacune des catégories de leurres utilisées (2 et 8) lors des simulations.

Tableau XIX
Intervalle de confiance pour le vrai score θ moyen
pour chaque catégorie de profils d'élèves à 90%

no. catégorie de leurres	Stratégie de sélection de la catégorie du leurre		
	Aléatoire	1 / 2	3 / 4
	Intervalle de confiance	Intervalle de confiance	Intervalle de confiance
02	-0.99 ± 0,70 [-1.69,-0.29]	-0.11 ± 0,67 [-0.78,0.56]	1.19 ± 0,69 [0.50,1.88]
08	-1.61 ± 0,69 [-2.30,-0.92]	-0.07 ± 0,65 [-0.72,0.58]	1.42 ± 0,67 [0.75,2.09]

À l'aide de cette information, nous avons pu établir des points de césures permettant de catégoriser le degré de cohérence d'une difficulté relative à une catégorie de leurres donnée.

Les figures 5 et 6 illustrent les intervalles de confiance pour chacune des catégories d'élèves et pour chacune des banques (catégories de leurres 2 et 8).

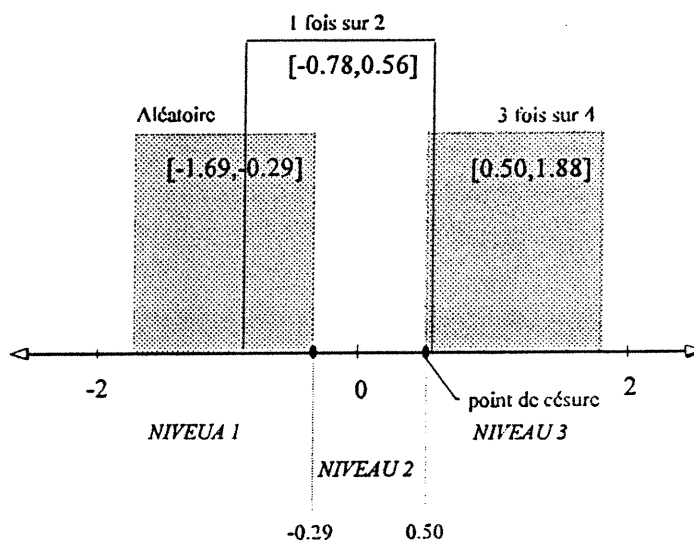


figure 5 - Intervalle de confiance à 90% pour le vrai score θ moyen pour les différentes catégories d'élèves avec la banque d'items spécifique à la catégorie de leurres 2

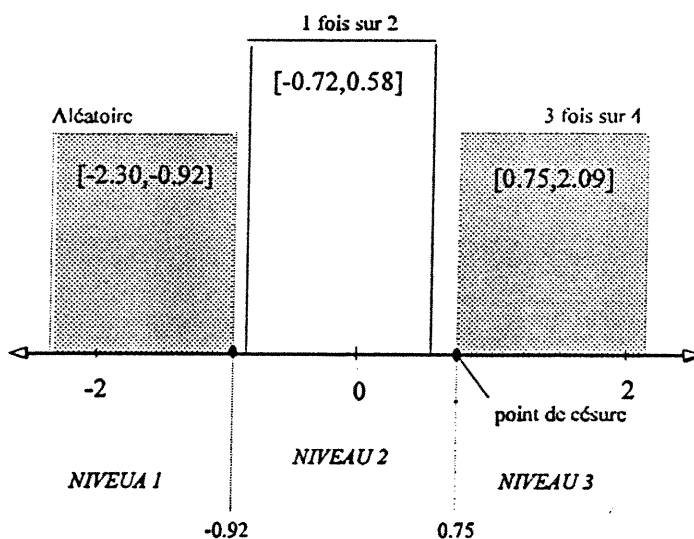


figure 6 - Intervalle de confiance à 90% pour le vrai score θ moyen pour les différentes catégories d'élèves avec la banque d'items spécifique à la catégorie de leurres 8

Si nous décidons de ne pas laisser passer un élève ayant une véritable difficulté cohérente sans l'identifier tout en évitant d'identifier des élèves qui répondent au hasard comme présentant une certaine cohérence dans leur difficulté, nous arrivons aux points de césures présentés au tableau XX et illustrés aux figures 7 et 8.

Tableau XX
Points de césures définis pour chaque banque d'items spécifique à la catégorie de leurres 2 et 8.

Descriptions des niveaux de cohérence	Banque d'items catégorie de leurres 2	Banque d'items catégorie de leurre 8
Niveau 1	$\theta < -0.29$	$\theta < -0.92$
Niveau 2	$-0.29 \leq \theta < 0.50$	$-0.92 \leq \theta < 0.75$
Niveau 3	$\theta \geq 0.50$	$\theta \geq 0.75$

Ces points de césures délimitant les niveaux de cohérence de l'élève à choisir une même catégorie de leurres donnée, servent à définir les critères de sélection du message diagnostique qui est communiqué à l'enseignant par l'entremise de la fiche diagnostique individuelle.

Les messages diagnostiques

Les messages diagnostiques associés aux résultats des simulations pour les trois niveaux de cohérence seraient :

Niveau 1 : "La difficulté n'est pas cohérente, il est inutile de prendre des mesures pédagogiques spécifiques à cette difficulté."

Niveau 2 : "La difficulté présente un certain degré de cohérence qui pourrait nécessiter des mesures correctives (facultatives)."

Niveau 3 : "La difficulté est cohérente, l'hypothèse de diagnostic causal est confirmée pour cet élève et des actions correctives devraient être prises."

4.2.3 Une interprétation des résultats

Nous observons certaines différences entre les intervalles de confiance calculés pour les catégories de profils simulés des deux banques d'items.

Nous constatons un chevauchement entre les intervalles de confiance autour des θ moyens provenant des simulations utilisant la plus petite banque d'items (catégorie de leurres 2). Ces différences semblent provenir des estimations des valeurs θ effectuées à partir des indices de difficulté attribuées aux items de chaque banque. Nous rappelons que les valeurs des indices de difficulté ont été générées selon une distribution normale en fonction du nombre d'items contenu dans la banque. La banque d'items spécifiques à la catégorie de leurres 8 contenant 109 items s'est vu assigner une plus grande variété de valeurs d'indice de difficulté que la banque relevant de la catégorie de leurres 2 (23 items). Les distributions des indices de difficulté générées pour

une plus grande banque d'items (placées à l'annexe D) correspondent mieux à une distribution normale que dans le cas d'une plus petite banque.

Nous convenons d'une incidence présumée de la grosseur de la banque et de la variété des indices de difficulté sur l'étendue des intervalles de confiance. Les intervalles de confiance obtenus à partir de la plus petite banque (catégorie de leurres 2) sont légèrement plus grands que les intervalles correspondants de la plus grande banque (catégorie de leurre 8).

Les distributions des valeurs θ (figures 7 et 8) à partir desquelles nous avons calculé les valeurs des θ moyens et les intervalles de confiance diffèrent selon les catégories mais particulièrement par rapport à la grandeur de la banque. Les simulations utilisant la plus grande banque d'items (catégorie de leurres 8) produisent moins de valeurs de θ nulles que lors des simulations utilisant la plus petite banque d'items (catégorie de leurres 2). Ces valeurs de θ forment deux distributions aux extrémités. Nous trouvons intéressant d'observer que pour la distribution des valeurs θ pour la catégorie "l'élève choisit 1 fois sur 2 la catégorie de leurres 8", le système discrimine fortement (à droite ou à gauche) les valeurs θ estimées à partir de la plus grande banque d'items. Nous considérons qu'il serait important de prévoir des banques d'items assez grandes qui contiendraient une grande diversité d'indices de difficulté pour permettre à la stratégie adaptative de bien discriminer la zone de niveau intermédiaire de cohérence.

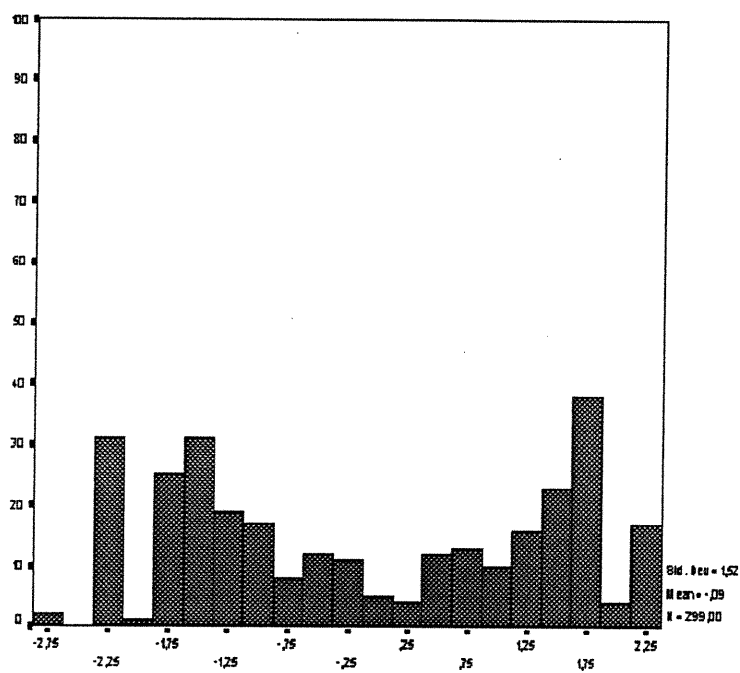


figure 7 - Distribution des valeurs θ estimées à partir de la plus grande banque d'items (catégorie de leurres 8)

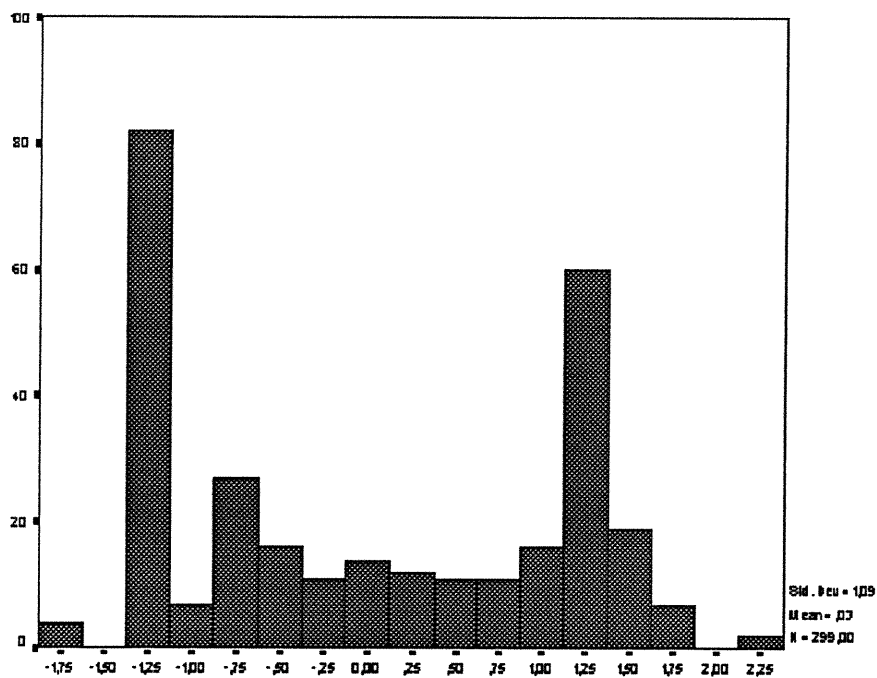


figure 8 - Distribution des valeurs θ estimées à partir de la plus petite banque d'items (catégorie de leurres 2)

Pour tous les cas particuliers où l'élève choisit systématiquement la catégorie de leurres donnée parmi les choix de réponse ou à l'opposé jamais, le système estime la valeur du θ final à 4 et -4 avec une variance erreur extrême entre 6.5 et 7.5. Notons que lorsqu'un élève choisit 19 fois sur 20 la catégorie de leurres spécifique, la valeur estimée du θ chute à environ 2.2 avec une variance erreur d'environ 0.377. Ainsi le système renvoie une interprétation valable pour ces valeurs de θ extrêmes qui correspondent à un niveau 1 ou 3 de cohérence à choisir une catégorie de leurres donnée.

Nous devons prendre note, que les conditions d'arrêt pour la seconde phase du test ont une influence sur la procédure d'estimation des valeurs finales des θ .

Il nous semble important de considérer ces éléments de divergence entre les différentes grandeurs de banques d'items, afin de mieux contrôler la qualité des estimations des valeurs θ et ainsi améliorer la qualité du diagnostic causal.

4.3 Limites du système développé

Le système de tests diagnostiques développé comporte deux phases de tests pour diagnostiquer, dans un contexte d'évaluation formative, les difficultés de l'élève pour l'objectif de géométrie 5.5.2. et confirmer une difficulté diagnostiquée à la première phase relevant des catégories de leurres 2 et 8 seulement. Le système de tests actuel ne fonctionne que pour cet objectif et ne confirme la cohérence des difficultés que vis-à-vis les catégories de leurres 2 et 8. Pour compléter et utiliser ce système avec d'autres objectifs,

il faudrait modifier le programme principal du test dans un environnement MicroCat™. Il faudrait alors modifier l'appel des banques d'items, les messages diagnostiques selon les notions, les contextes et les catégories de leurres ainsi que constituer les banques spécifiques aux autres catégories de leurres.

CONCLUSION

ET

DISCUSSIONS

La problématique de la recherche

Depuis l'adoption au Québec d'une politique générale d'évaluation pédagogique (MEQ, 1981) l'évaluation formative des apprentissages est prônée par les systèmes scolaires. Dans le contexte de la classe, l'enseignant doit assumer le rôle de premier intervenant auprès de l'élève en difficulté; pour être efficace, ses interventions correctives gagneraient à s'intégrer à une démarche diagnostique.

L'étude de Gullickson (1982) souligne l'importance des informations informelles colligées par les enseignants dans le but d'évaluer les apprentissages. Des études québécoises (Scallon, 1988a; Dassa & Dumoulin, 1990; Fontaine, 1990; Lanoue, 1993) montrent que les enseignants manquent de formation concernant les principes et la pratique de l'évaluation formative. Ces lacunes s'étendent à l'élaboration d'outils de mesure notamment en ce qui concerne les instruments de mesure diagnostique adaptés à leurs pratiques évaluatives (Anderson, 1989).

Le manque d'outils d'évaluation formels ou informels est à l'origine de notre intérêt pour les systèmes de tests diagnostiques qui auraient les caractéristiques suivantes : leur contenu serait fondé sur les programmes d'étude, ils respecteraient les contraintes pédagogiques et organisationnelles du système scolaire québécois, ils refléteraient le plus possible les pratiques évaluatives des enseignants ainsi que les recommandations du MEQ qui préconise des interventions correctives rapides.

Le système de test réalisé

Notre recherche s'inscrit dans la continuité du développement d'un système de tests diagnostiques destiné à l'évaluation pédagogique des apprentissages. L'ordinateur a été utilisé comme support à la conception, à la réalisation et à l'administration des tests initialement développés. Ce médium a été utilisé pour ses capacités de traitement rapide et complexe (Boulet *et al.*, 1990; Dassa, Vazquez & Ajar, 1993; Bordeleau, 1994a).

Le système de tests diagnostiques développé dans le cadre de ce mémoire propose une nouvelle stratégie de testing informatisé de nature interactive et centrée sur le diagnostic pédagogique réalisé par l'enseignant dans le contexte de la classe. Il porte sur l'objectif de géométrie 5.5.2. inscrit au programme de mathématiques de la deuxième année du secondaire du MEQ.

La version initiale des tests a été réalisée en collaboration avec les enseignants en ce qui a trait à la création des banques d'items et à l'expérimentation du système auprès des élèves. La nouvelle version du système de tests diagnostiques présenté dans cette étude, comporte deux phases de testing au cours d'une même séance d'administration. La première phase permet d'identifier la nature de la difficulté de l'élève et de formuler une hypothèse concernant sa cause possible. La seconde phase permet de confirmer (ou d'infirmer) cette hypothèse en estimant un score qui permet une explication psychopédagogique de l'erreur commise. Cette information est communiquée à l'enseignant par l'entremise d'une fiche diagnostique individuelle produite à la fin de la séance de test.

Nous avons intégré une nouvelle stratégie "stradaptative" en ajoutant une programmation externe en langage Pascaltm. L'algorithme de sélection des items sur lequel repose cette stratégie a l'avantage d'être indépendante du

contenu des banques d'items. Le nouveau système qui en résulte pourrait être utilisé sans modifications avec d'autres banques relevant d'autres objectifs. La stratégie adaptative de sélection des items utilisée lors de la seconde phase du test permet d'obtenir une mesure de la difficulté de l'élève relative à une catégorie de leurres donnée. Le système de tests comporte un ajout majeur : la possibilité d'obtenir un message diagnostique qui confirme le niveau de difficulté spécifique de l'élève par le biais d'une mesure de cohérence à choisir une catégorie de leurres donnée.

Les études de vérification et de simulations réalisées permettent de conclure qu'il est possible d'intégrer deux stratégies différentes mais complémentaires de sélection des items. La première stratégie s'appuie sur une taxonomie tridimensionnelle du cube diagnostique et établit un diagnostic structurel qui permet d'identifier la nature de l'erreur observée en relation avec ce cube diagnostique; il permet également de formuler une hypothèse sur la cause de l'erreur. La seconde stratégie basée sur la théorie des réponses aux items (TRI), utilisant un modèle de Rasch à un paramètre, permet d'estimer un score qui mesure la cohérence de l'élève à choisir une même catégorie de leurres représentative d'une difficulté identifiée dans la première phase du test. Cette information constitue le diagnostic causal.

Le système de tests développé reflète les pratiques d'évaluation des enseignants qui ont collaboré au projet, une partie du programme d'études actuel en mathématiques de secondaire II et respecte les principales contraintes organisationnelles liées au système éducatif québécois. Notre contribution originale au domaine des tests diagnostiques informatisés, se caractérise par l'utilisation des mauvaises réponses (des catégories de leurres), à l'aide d'un modèle de Rasch, comme élément d'information visant à modéliser le diagnostic causal. Les deux diagnostics, structurel et causal, permettent de formuler un message diagnostique destiné à informer précisément l'enseignant des

difficultés probables d'un élève donc à l'aider dans sa tâche de régulation des apprentissages.

Les résultats des simulations montrent que le système développé est opérationnel et discrimine assez bien le niveau de cohérence de l'erreur. Les valeurs θ obtenues résultant des simulations nous ont permis de calculer des intervalles de confiance pour les trois niveaux de cohérence de l'erreur définis en terme de la valeur θ . En fin de parcours, le système associe le niveau de cohérence de l'erreur de l'élève à un message diagnostique qui est communiqué par la fiche de diagnostique individuelle.

Les limites techniques du système réalisé

Le temps d'administration du nouveau test dépasse du double la durée de la version initiale (environ 25 minutes). Le système de tests développé ne devrait donc pas prendre plus de temps que l'équivalent d'une période de cours complète (50 minutes) dans un laboratoire disposant d'un ordinateur par élève.

Le système de tests réalisé limite l'intégration des éléments du cube diagnostique à un nombre maximal de contextes (7), de notions (10) et des catégories de leurres (50). La mise en mémoire des informations relatives aux items à l'aide d'une structure dynamique plutôt qu'en tableau pourrait pallier ce type de limites. La programmation d'un système de tests sous un environnement qui tiendrait compte de la gestion de la mémoire vive (Windows95tm) plutôt que sous DOStm et MicroCattm serait à considérer.

Le problème majeur du système de tests développé (sous DOStm et MicroCattm) est un problème de duplication dans la structure. Le système utilise un système de gestion de banques d'items différent pour la première et la seconde phase du test. Il serait préférable qu'il n'y ait qu'un seul système de

gestion des banques et qu'une seule banque d'items qui contiendrait à la fois les informations pertinentes au diagnostic relevant du cube diagnostique (notion, contexte, catégorie de leurres, messages diagnostiques) et les informations spécifiques aux catégories de leurres (indice de difficulté de l'item par rapport aux items de même catégorie de leurres). Idéalement, un tel programme unique de gestion du test devrait comprendre la gestion des banques d'items, la création d'un test, l'administration du test et la présentation des résultats.

Les limites pédagogique du système réalisé

Le système de tests diagnostiques développé ne reflète pas de façon universelle les pratiques d'évaluation. Du fait que le système de test prend en compte la gestion des items distribués dans une structure tridimensionnelle du cube diagnostique (habileté, contexte et notion) et des catégories de leurres plausibles associés à chacun des items, il reflète les pratiques pédagogiques de l'utilisateur. Cet outil d'évaluation pédagogique à fonction diagnostique permet d'informer l'enseignant des difficultés d'un élève par l'intermédiaire d'une fiche diagnostique individuelle. Cependant la tâche de régulation des apprentissages ne peut s'effectuer sans l'intervention de l'enseignant. D'où l'importance d'augmenter la convivialité du système de façon à permettre aux enseignants de s'approprier ses différentes composantes: la constitution des banques d'items ainsi que la définition des cubes diagnostiques (Dassa, 1997). À partir des fiches diagnostiques, l'enseignant peut procéder rapidement à interventions pédagogiques individuelles ou collectives de nature corrective.

Recommandations : Un système de tests diagnostiques

Dans l'éventualité de la conception d'un nouveau système de tests diagnostiques, nous recommanderions que de maintenir la stratégie développée, à savoir deux phases dans une même séance : une phase descriptive et une phase confirmative des difficultés. La première donne un portrait général des apprentissages de l'élève alors que la deuxième fournit une information plus précise des difficultés. Cette segmentation permettrait de mieux gérer des périodes d'administration plus courtes en permettant de séparer la phase structurelle de la phase causale.

Pour contrer les problèmes techniques du système actuel et ses limites, nous recommanderions que le système de tests soit développé et utilisé dans un environnement plus ouvert et plus convivial (outils de développement tels VisualBasictm, C++tm, environnement de type Windows95tm). Également, qu'un tel système puisse intégrer plusieurs types de stratégies permettant de confirmer les difficultés des élèves (d'autres modèles), un module de gestion des banques d'items, des messages diagnostiques, différentes banques d'items, la gestion de l'administration des tests et de la communication des résultats. L'intégration de différents médiums (son, image de qualité, vidéo) serait souhaitable dans le cas de systèmes de tests destinés à évaluer d'autres types de tâches que celles qui peuvent être testées par le système et l'environnement utilisés dans notre recherche. Des matières telles que l'histoire, la géographie, les langues les sciences pourraient grandement bénéficier d'une approche multimédia.

Afin de réaliser un système de tests diagnostiques s'inspirant du système développé, il faudrait que les tâches évaluées relèvent d'objectifs connus et que l'on connaisse suffisamment les conceptions erronées afin de définir des catégories de leurs qui les expriment.

Finalement, dans le cadre des programmes d'études actuels, il conviendrait qu'un outil d'évaluation diagnostique soit construit sur la base des pratiques évaluatives des enseignants, dans le respect des objectifs des programmes ainsi que des contraintes pédagogiques et organisationnelles du système éducatif en vigueur.

Discussions

Nous convenons qu'un système informatisé d'évaluation pédagogique doit être ouvert aux pratiques des enseignants, il doit refléter les façons de faire de ceux-ci, l'écologie de la classe. L'implication des enseignants est donc non seulement essentielle au développement d'un tel système mais aussi à l'accueil positif et à l'intégration de cet outil d'évaluation pédagogique à fonction diagnostique dans leur pratique quotidienne.

Dans l'éventualité d'un système diagnostique adaptatif simple, plus complet et ouvert, l'enseignant devrait pouvoir définir lui-même les banques d'items, la structure d'un cube diagnostique et la stratégie du test à être utilisée. Le système devrait être alors convivial, flexible et interactif afin qu'il se prête (sur mesure) aux besoins de l'enseignant.

Un système de diagnostic pédagogique a pour but premier de proposer aux enseignants des pistes d'informations afin qu'ils puissent apporter les correctifs par des actions pédagogiques conséquentes et bien ciblées auprès des élèves en difficultés mais il est aussi approprié pour servir de point d'entrée à des modules correctifs. Dans une perspective plus large d'un système informatisé intégrant évaluation et apprentissage, nous pensons l'intégration de modules initiaux d'apprentissages et finaux de correctifs au système de tests diagnostiques est possiblement réalisable compte tenu des développements technologiques rapides des aspects touchant l'interactivité et le multimédia.

RÉFÉRENCES BIBLIOGRAPHIQUES

BIBLIOGRAPHIE

Allal, L.; Cardinet, J. & Perrenoud, P. (1979). *L'évaluation formative dans un enseignement différencié*. Actes du Colloque à l'Université de Genève, mars 1978: Édition Peter Lang.

Allal, L. (1988). *Vers un élargissement de la pédagogie de maîtrise : processus de régulation interactive, rétroactive et proactive*. In M. Huberman, Assurer la réussite des apprentissages scolaires?, Paris: Delachaux et Niestlé, pp.86-126.

Anderson, J.O. (1989). *Evaluation of Student Achievement : Teacher Practives and Educational Measurement*. The 1986 British Columbia Science Assessment.

Auger, R. (1990). Communication personnelle de R. Auger à C. Dassa, été 1990, Université de Montréal.

Auger, R. (1989). *Étude de praticabilité du testing adaptatif de maîtrise en regard des apprentissages scolaires au Québec*. Thèse de doctorat, Montréal, U.Q.A.M.

Auger, R. & Dassa, C. (1991). *L'intégration du diagnostic interactif à l'évaluation formative informatisée des apprentissages scolaires*. Colloque international de l'Association Francophone Internationale de la Recherche Scientifique en Éducation, Carcassonne, France.

Assessment System Corporation (ASC) (1989). *User's Manual for the MicroCat*. MicroCat Testing System, 3rd edition, St-Paul, Minnesota.

Berthelot, S. (1987). *L'ordinateur compatible avec l'éducation*. Collection Éducation, Montréal : Agence D'Arc.

Binet, A. (1909). *Les idées modernes sur les enfants*. Paris: Flammarion.

Birnbaum, A. (1968). *Some Latent Trait Models and their Uses in Inferring an Examinee's Ability*. In F.M Lord & M.R. Novick, Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley Ed.

Birzée, C. (1982). *La pédagogie du succès*. Paris: Presses Universitaires de France.

Blais, J.G. (1987). *Effets de violation du postulat d'unidimensionalité dans la théorie des réponses aux items*. Thèse de doctorat, Montréal, Université de Montréal, Août.

Bloom, B.S. et al.(1956). *Taxonomie des objectifs pédagogiques, tome 1 : domaine cognitif*. Traduit de l'anglais, Québec : Presses de l'Université du Québec.

Bloom, B.S.; Madaus, G.F. & Hastings, J.T. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. N.Y.: McGraw-Hill.

Bloom, B.S.; Madaus, G.F. & Hastings, J.T. (1981). *Evaluation to Improve Learning*. N.Y.: McGraw-Hill.

Bloom, B.S. (1984). The 2 Sigma Problem, The Search for Methods of Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 41(8), pp.4-17.

Bordeleau, P. (1994a). *Apprendre dans des environnements pédagogiques informatisés*. Montréal: Édition Logiques.

Bordeleau, P. (1994b). *Des outils pour apprendre avec l'ordinateur*. Collectif sous la direction de Pierre Bordeleau, Montréal: Édition Logiques.

Boulet, M-M., Lavoie, L., Labbé, P. & Lemay, F.(1990). Système informatisé pour l'évaluation formative. *Mesure et évaluation en éducation*, 13(1), pp.29-55.

Bunderson, C.V. et al.(1989). *The Four Generations of Computerized Educational Measurement*. In R.L.Linn. N.Y.: American Council of Education / MacMillan.

Burelle, R.; Gadbois, L; Parent, C. & Séguin, S.P. (1991a). Un instrument de mesure des croyances et attitudes des enseignants à l'égard des pratiques d'évaluation formative des apprentissages de leurs élèves. *Mesure et évaluation en éducation*. 14(2), Été, pp.5-24.

Burelle, R.; Gadbois, L.; Parent, C. & Séguin, S.P. (1991b). Mesurer des croyances et des attitudes pour mieux comprendre des comportements. *Mesure et évaluation en éducation*. 14(1), Printemps, pp.61-77.

Conseil Supérieur de l'Éducation (CSE) (1990). *La pédagogie, un défi majeur de l'enseignement supérieur*. Avis au Ministre de l'enseignement supérieur et de la science, Québec : CSE.

Crooks, J. Terence (1988). The Impact of Classroom Evaluation Practices on Students, Classroom Evaluation Practices. *Review of Educational Research*, 58(4), pp.438-481.

Danvoye, P. (1987). *Évaluation de l'implantation de la micro-informatique dans les commissions scolaires*. Québec, MEQ : Direction générale de l'éducation et des ressources didactiques, Direction de la technologie éducative.

Danvoye, P. (1990). *La situation en micro-informatique scolaire dans les écoles publiques en 1989-1990. Document #1 : le parc informatique, ses caractéristiques et son utilisation*. Québec, MEQ : Direction des ressources didactiques.

Dassa, C.; Lavallée, M.; Parent, C. & Séguin, S.P. (1987). Recension : Changing Concepts of Educational Evaluation, de TYLER, R.W. *Mesure et évaluation en éducation*, 10(2), Été, pp.83-88.

Dassa, C. (1988). L'intégration du diagnostic pédagogique aux apprentissages scolaires: de la théorie à la voie informatique. *Mesure et évaluation en éducation*, 11(1), Printemps, pp.7-26.

Dassa, C. & Dumoulin, J. (1990). *Les pratiques de l'évaluation dans le contexte de la classe*. Communication présentée dans le cadre du XIIIe Colloque annuel de l'ADMÉE, Montréal.

Dassa, C. (1990). From a Horizontal to a Vertical Method of Integrating Educational Diagnosis with Classroom Assessment. *The Alberta Journal of Educational Research*, 36(1), pp.35-44.

Dassa, C. & René De Cotret, S. (1993). Validation d'un système informatisé de diagnostic en mathématiques au secondaire : *une approche centrée sur l'analyse didactique*. *Mesure et évaluation en éducation*, 16(3-4), Automne 1993 - Hiver 1994, pp.5-26.

Dassa, C.; Vazquez-Abad, J. & Ajar, D. (1993). Formative Assessment in a Classroom Setting : From Practices to computer Innovations. *The Alberta Journal of Educational Research*, pp.111-125.

Dassa, C. & Ajar, D. (1995). *Élaboration et validation d'un modèle d'intégration du diagnostic pédagogique aux apprentissages scolaires au premier cycle du secondaire, en mathématiques*. Rapport final de productivité présenté au CRSH (No 41-89-1010).

Dassa, C. (1997). *L'évaluation des apprentissages telle qu'elle peut se présenter dans le contexte: Entretien avec C.Dassa*. Édit. Meunier, C. "Points de vue sur le multimédia interactif en éducation". Montréal, édition Chenelière/McGraw-Hill, 1er trimestre 1997, pp.162-174.

De Landsheere, G. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : Presses Universitaires de France, 2e édition.

Delisle, C. (1992). *Répertoire de logiciels pour l'enseignement des sciences*. Québec, MEQ: Direction des ressources didactiques.

Depover, C. (1987). *L'ordinateur média d'enseignement, un cadre conceptuel*. Collection Pédagogies en développement. Bruxelles: De Boeck-Wesmael, 4e édition.

Dumoulin, J. (1991). *Étude des pratiques de l'évaluation des apprentissages réalisées par les enseignants de mathématiques et de sciences au secondaire*. Mémoire de maîtrise, Montréal, Université de Montréal, Avril.

Feldt L.S. & Brennan R.L. (1991). *Reliability, In Educational Measurement*. 3rd Ed., R.L.Linn Ed., New York: MacMillan.

Folk, V.G. & Green, B.F. (1989). Adaptive Estimation when the Unidimensionality Assumption of IRT is Violated. *Applied Psychological Measurement*, 13(4), pp.373-393.

Fontaine, S. (1990). *Élaboration et validation d'une grille d'observation visant l'identification des actes évaluatifs des enseignants dans un contexte d'évaluation formative*. Mémoire de maîtrise, Montréal, Université de Montréal, Septembre.

Giardina, M. (1997). *Revenir à l'interactivité in Meunier, C.: Points de vue sur le multimédia interactif en éducation: Entretien avec M.Gardina*. Édit. Meunier, C. "Points de vue sur le multimédia interactif en éducation". Montréal, édition Chenelière/McGraw-Hill, 1er trimestre 1997, pp.195-216.

Gullickson, A.R. (1982). *The Practice of Testings in Elementary and Secondary Schools*. Paper presented at the 1982 Rural Education Conference, Kansas State University, Manhattan.

Gullickson, A.R. & Ellwain, M.C. (1985). Post hoc Analysis of Teacher-Made Tests : The Goodness-of-fit Between Prescription and Practice. *Educational Measurement: Issues and practice*, 4(1), pp. 15-18.

Hambleton, R.K. & Swaminathan, H. (1985). *Item Response Theory, Principles and Applications*. Kluwer: Nijhoff Publishing.

Kapes, J.T. & Vansickle, T.R. (1992). Comparing Paper-Pencil and Computer, version of the Harrington - O'Shea Career Decision Making System. *Measurement and Evaluation in Counseling & Development*, 25(1), Avril, pp.5-13.

Kulik, C-L.C. & Kulik, J.A. (1987). Review of Recent Research Literature on Computer-Based Instruction. *Contemporary Educational Psychology*, 12(3), pp.222-230.

Lanoué, M. (1993). *Mise à l'essai d'un modèle d'intégration de l'évaluation diagnostique à l'enseignement en classe, en mathématiques de deuxième secondaire*. Mémoire de maîtrise, Montréal, Université de Montréal, Août.

Laurier, M. (1992) *Application des techniques de "tests adaptatifs" en langue seconde*. Actes du colloque du CIPTE, Ste-Foy, Québec: Télé-Université 1992.

Laurier, M. (1993). *L'évaluation en langue seconde : vers des instruments qui s'adaptent à l'apprenant*. Communication présentée au CCCELAO ,Victoria.

Laurier, M. (1996). *L'évaluation en langue seconde : vers des instruments qui s'adaptent à l'apprenant*. Communication présentée au CCCELAO ,Victoria.

Laurier, M. (1997). *L'évaluation des apprentissages telle qu'elle peut se présenter dans le contexte: Entretien avec M.Laurier*. Édité. Meunier, C. "Points de vue sur le multimédia interactif en éducation". Montréal, édition Chenelière/McGraw-Hill, 1er trimestre 1997, pp.175-192.

Legendre, R. (1988). *Dictionnaire actuel de l'éducation*. Boucherville, Québec: Librairie Larousse.

Lenman, S. (1994). Rapport technique (CRSH-No 410-89-1010).

Lord, F.M. (1952). A Theory of Test Scores. *Psychometric Monograph*, 7.

Lord, F.M. & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley Ed.

Lord, F.M. (1980). *Applications of Item Response to Theory to Practical Testing Problems*. Hillsdale, N.J.: Erlbaum Ass.

Miles, M.B. & Huberman, A.M. (1984). *Qualitative Data Analysis : an expanded source book*. 2nd Edition, Thousand Oaks; London; New Delhi : Sage Publications.

Ministère de l'Éducation du Québec (MEQ) (1979). *L'école québécoise, Énoncé de politique et plan d'action*. Québec : MEQ.

Ministère de l'Éducation du Québec (MEQ) (1981). *Politique générale d'évaluation pédagogique*. Québec : MEQ.

Morrisette, D. (1984). *La mesure et l'évaluation en enseignement*. Québec, Presses de l'Université Laval.

Perrenoud, P. (1986). *La part d'évaluation formative dans toute évaluation continue*. Les actes de la rencontre Internationale "Évaluer l'évaluation", Dijon, France.

Perrenoud, P. (1991). Pour une approche pragmatique de l'évaluation formative. *Mesure et évaluation en éducation*, 13(4), pp.49-81.

Perrenoud, P. (1993). Touche pas à mon évaluation! Pour une approche systématique du changement. *Mesure et évaluation en éducation*, 16(1-2), pp.107-132, printemps-été.

Quesney, J.L. (1996). *Élaboration d'un système expert de diagnostic pédagogique dans le contexte de la classe de mathématiques du deuxième secondaire*. Mémoire de maîtrise, Montréal, Université de Montréal, mai.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago : University of Chicago Press.

Samejima, F. (1977). A Use of the Information function in Tailored Testing. *Applied Psychological Measurement*, 1 pp.233-247.

Scallon, G. (1988a). *L'évaluation formative des apprentissages : la réflexion*. Québec: Presses de l'Université Laval.

Scallon, G. (1996). L'évaluation formative et le temps d'enseigner. *Vie Pédagogique*, 99 (mai-juin) 1996, pp. 4-9.

Scriven, M. (1967). The Methodology of evaluation. *AERA, Monograph Series on Curriculum Evaluation*, 1, pp.39-83.

Séguin, P.-S.& Auger, R. (1986a). Introduction à la théorie des réponses aux items TRI. *Mesure et évaluation en éducation*, 9(1), pp.7-44.

Séguin, P.-S.& Auger, R. (1986b). Le modèle de Rash et la paramétrisation d'une banque d'items et d'instruments de mesure. *Mesure et évaluation en éducation*, 9(2-3), pp.59-97.

Stiggins, J.R. (1985). The Ecology of Classroom Assessment. *Journal of Educational Measurement*, 2(4), pp.271-286.

Stiggins, J.R. (1988). Revitalizing Classroom Assessment : the Highest instructional priority. *Phi Delta Kappan*, 65, pp.363-372.

Tousignant, R. & Morrisette, D. (1990). *Les principes de la mesure et de l'évaluation des apprentissages*. 2e édition revue et augmentée, Boucherville: Gaetan Morin éditeur.

Weiss, D.J.(1974). *Strategies of Adaptive Ability Measurement*. Minneapolis: University of Minnesota, Department of Psychology.

Weiss, D.J. & Kingsbury, G.G. (1984). Application of Computerized Adaptive Testing to Educational Problems. *Journal of Educational Measurement*, 21, pp.361-375.

Wise, S.L. & Plake, B.S. (1990). Computer-Based Testing in Higher Education. *Measurement and Evaluation in Counseling and Development*, 23, April.

Wise, S.L.; Plake, B.S. & al. (1992). A Comparaison of Self Adapted and Computerized Adaptive Tests. *Journal of Educational Measurement*, 29(4), pp..329-339.

Wright, B.D. & Stone, M.H. (1979). *Best Test Design: Rasch measurement*. Chicago: MESA Press.

LES ANNEXES

- A. Liste des objectifs**
- B. Liste des catégories de leurres par objectif**
- C. Figures 2,3,4 et 5 tirées du document Lenman (1994)**
- D. Distributions des indices de difficulté générés**
- E. Liste des items et de leur indice de difficulté**
- F. Compilation de l'étude de vérification de l'algorithme "stradaptatif" de la PHASE I du test.**
- G. Fiches diagnostiques individuelles**
- H. Profils des élèves à la PHASE II.**

A. Liste des objectifs**Module d'algèbre :**

- 3.1.3 CALCULER la somme des termes semblables
- 3.1.6 CALCULER le quotient d'expressions algébriques simples.
- 3.1.8 CALCULER la valeur numérique d'une expression algébrique donnée.
- 3.2.3 CONSTRUIRE une équation ou inéquation simple à une variable à partir d'une donnée textuelle.
- 3.2.6 ISOLER la variable dans une équation ou une inéquation simple.

Module de géométrie :

- 5.5.2 CALCULER la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.
- 5.5.6 Étant donné le périmètre ou l'aire d'une figure géométrique plane et certaines mesures pertinentes, CALCULER la mesure de l'une de ses dimensions ou CALCULER l'aire ou le périmètre respectif.
- 5.5.8 CALCULER l'aire latérale ou l'aire totale d'un prisme droit ou d'un cylindre droit.

Objectif d'algèbre 3.1.3

Description: CALCULER la somme des termes semblables

Contenus notionnels:

1. 1 variable
2. 2 variables
3. 3 variables
4. plus de 3 variables

Contextes diagnostiques:

expressions algébriques sans facteur :

- A. 1 exposant
- B. plusieurs exposants

expressions algébriques avec facteurs :

- C. 1 exposant
- D. plusieurs exposants

Objectif d'algèbre 3.1.6

Description: CALCULER le quotient d'expressions algébriques simples.

Contenus notionnels:

1. opérations sur les monômes
2. opérations sur les polynômes

Contextes diagnostiques:

expressions algébriques (variable au numérateur, au dénominateur) :

- A. 1 variable à simplifier
- B. 2 variables à simplifier
- C. 3 variables à simplifier
- D. 4 variables à simplifier

expressions algébriques :

- E. 1 variable à simplifier
- F. 2 variables à simplifier
- G. 3 variables à simplifier
- H. 4 variables à simplifier

expressions linéaires (facteur manquant):

- I. 1 variable à simplifier
- J. 2 variables à simplifier
- K. 3 variables à simplifier
- L. 4 variables à simplifier

Objectif d'algèbre 3.1.8

Description: CALCULER la valeur numérique d'une expression algébrique donnée.

Contenus notionnels:

1. nombres naturels
2. entiers relatifs
3. variables rationnelles

Contextes diagnostiques:

expressions algébriques :

- A. 1 variable, 1 exposant
- B. 1 variable, plusieurs exposants
- C. 2 variables, 1 exposant
- D. 2 variables, plusieurs exposants
- E. 3 variables, 1 exposant
- F. 3 variables, plusieurs exposants

formules de référence géométrique :

- G. 1 variable, 1 exposant
- H. 1 variable, plusieurs exposants
- I. 2 variables, 1 exposant
- J. 2 variables, plusieurs exposants
- K. 3 variables, 1 exposant
- L. 3 variables, plusieurs exposants

Objectif d'algèbre 3.2.3

Description: CONSTRUIRE une équation ou inéquation simple à une variable à partir d'une donnée textuelle.

Contenus notionnels:

1. équations
2. inéquations

Contextes diagnostiques:

expressions littérales :

- A. 1 donnée à mathématiser (trouver une solution)
- B. 2 données à mathématiser (trouver une solution)
- C. 3 données et plus à mathématiser (trouver une solution)

Objectif d'algèbre 3.2.6

Description: ISOLER la variable dans une équation ou une inéquation simple.

Contenus notionnels:

1. équations
2. inéquations
3. formules

Contextes diagnostiques:

- A. expressions algébriques à 1 variable
- B. expressions algébriques à 2 variables
- C. expressions algébriques à 3 variables

Objectif de géométrie 5.5.2

Description: CALCULER la circonférence ou l'aire d'un cercle à partir du rayon ou du diamètre.

Contenus notionnels:

1. Calcul de la circonférence
2. Calcul de l'aire

Regroupement des contextes diagnostiques:

Problèmes abstraits:

- A. Mesures identifiées sur une figure
- B. Mesures identifiées dans le texte par des nombres entiers
- C. Mesures identifiées dans le texte par des nombres décimaux

Problèmes réels:

- D. Mesures identifiées sur une figure
- E. Mesures identifiées dans le texte par des nombres entiers
- F. Mesures identifiées dans le texte par des nombres décimaux

Objectif de géométrie 5.5.6

Description: Étant donné le périmètre ou l'aire d'une figure géométrique plane et certaines mesures pertinentes, CALCULER la mesure de l'une de ses dimensions ou CALCULER l'aire ou le périmètre respectif.

Contenus notionnels:

items sans figures géométriques :

1. Formules présentes
2. Formules absentes

Contextes diagnostiques:

- A. Mesures sur une figure, nombres entiers
- B. Mesures sur une figure, nombres décimaux
- C. Mesures dans le texte, nombres entiers
- D. Mesures dans le texte, nombres décimaux
- E. Mesures sur une figure, nombres entiers
- F. Mesures sur une figure, nombres décimaux

Objectif de géométrie 5.5.8

Description: CALCULER l'aire latérale ou l'aire totale d'un prisme droit ou d'un cylindre droit.

Contenus notionnels:

prisme droit :

1. Aire latérale
2. Aire totale

cylindre droit:

1. Aire latérale
2. Aire totale

Contextes diagnostiques:

problèmes abstraits:

- A. Mesures identifiées sur une figure, nombres entiers
- B. Mesures identifiées sur une figure, expressions équivalentes, nombres entiers
- C. Mesures identifiées sur une figure, nombre décimaux
- D. Mesures identifiées sur une figure, expressions équivalentes, nombres décimaux
- E. Mesures identifiées dans le texte, nombres entiers
- F. Mesures identifiées dans le texte, expressions équivalentes, nombres entiers
- G. Mesures identifiées dans le texte, nombres décimaux
- H. Mesures identifiées dans le texte, expressions équivalentes, nombre décimaux

problèmes réels:

- I. Mesures identifiées sur une figure, nombres entiers
- J. Mesures identifiées sur une figure, expressions équivalentes, nombres entiers
- K. Mesures identifiées sur une figure, nombre décimaux
- L. Mesures identifiées sur une figure, expressions équivalentes, nombres décimaux
- M. Mesures identifiées dans le texte, nombres entiers
- N. Mesures identifiées dans le texte, expressions équivalentes, nombres entiers
- O. Mesures identifiées dans le texte, nombres décimaux
- P. Mesures identifiées dans le texte, expressions équivalentes, nombre décimaux

B. Liste des catégories de leures par objectif**Liste des catégories de leures pour le test correspondant à l'objectif d'algèbre 3.1.3**

No.Catégorie	Description
02	Associe des termes non semblables lors de l'add. et de la soust.
03	Ignore le coefficient (1 ou -1) lors de la résolution.
04	Additionne les exposants des variables.
05	Ignore que la soustraction de termes identiques est nulle.
06	Erreur de calcul.
07	Oublie la retenue dans l'addition.
08	Oublie d'effectuer la distributivité.
09	Erreur dans les signes.
10	Ignore les parenthèses lors de la résolution.
11	Add. des termes sembl. sans tenir compte des signes et des coeff.
12	Erreur lors de l'addition des fractions.
13	Sa réponse ne comporte pas les termes qui ne peuvent s'additionner.
14	Multiplie au lieu d'additionner.
15	Divise au lieu d'additionner.
16	Erreurs 3 et 4.

**Liste des catégories de leures pour le test correspondant à
l'objectif d'algèbre 3.1.6**

No.Catégorie	Description
02	Interchange diviseur et dividende.
03	Laisse tomber la partie littérale dans la réponse.
04	Multiplie au lieu de diviser.
05	Fait une erreur de signe.
06	Fait une erreur en divisant.
07	Additionne des termes non semblables.
08	Oublie de diviser un ou deux termes.
09	Additionne les exposants au lieu de soustraire.
10	Oublie que l'exposant 1 est sous-entendu.
11	Fait une erreur de calcul.
12	Soustrait les coefficients au lieu de diviser.
13	Erreurs 2 et 9.
14	Erreurs 5 et 10.
15	Considère le 0 absorbant pour la partie numérique seulement.
16	Ne reconnaît pas la propriété d'élément absorbant du 0 dans la mult.
17	Erreur 2 et mauvaise décomposition.
18	Soustrait les termes.

**Liste des catégories de leures pour le test correspondant à
l'objectif d'algèbre 3.1.8**

No.Catégorie	Description
02	Ne tient pas compte des parenthèses ou en ajoute.
03	Ne fait pas la bonne opération (refus de soustraire un grand d'un petit).
04	Remplace une variable par le mauvais nombre.
05	Ne fait pas l'opération, remplace seulement la variable par le nombre.
06	Ne tient pas compte du coefficient d'un terme dans l'expression.
07	Oublie la virgule décimale.
08	Fait une erreur dans la priorité d'opération.
09	Oublie de diviser par 2.
10	Oublie d'élever au carré.
11	Erreur de signe.
12	Erreur lors de l'addition ou de la soustraction d'un négatif.
13	Confond puissance et produit. (ex.: $5^2 = 10$)
14	Croit que l'exposant se rapporte à toute l'expression ou à tout le terme.
15	Oublie que le 0 est élément absorbant dans la multiplication.
16	Simplifie les termes d'un quotient (seuls les facteurs peuvent se simplifier).
17	Erreurs 2 et 5.

- 18 Erreurs 3 et 5.
- 19 Erreurs 3 et 7.
- 20 Erreurs 2 et 10.
- 21 Erreurs 2 et 11.
- 22 Erreurs 7 et 16.
- 23 Erreurs 7 et 11.
- 24 Erreurs 2 et 7.
- 25 Erreurs 8 et 12.
- 26 Erreurs 3 et 13.
- 27 Erreurs 7 et 15.
- 28 Erreurs 13 et 15.
- 29 Erreurs 2 et 14.
- 30 Erreurs 11 et 30.
- 31 Erreurs 11 et 15.
- 32 Méprise de la virgule pour une barre de fraction.
- 33 Concaténation.
- 34 Erreurs 13 et 11.
- 35 Inversion des variables.
- 36 Erreurs 33 et 12.
- 37 Refus de soustraire un grand d'un petit.
- 38 Erreurs 13 et 10.

Liste des catégories de leures pour le test correspondant à l'objectif d'algèbre 3.2.3

No.Catégorie	Description
01	Multiplie au lieu d'additionner ou vice versa.
02	Soustrait au lieu de multiplier ou vice versa.
03	Divise au lieu de multiplier ou vice versa.
04	Inverse les deux termes d'une soustraction.
05	Mêle puissance et produit ($x \text{ exp.}n$ et nx).
06	Ne tient pas compte de la virgule après la variable dans le texte (ou de la priorité d'opération).
07	Fait une erreur de symbolisme ($<$, \leq , $>$, \geq).
08	Applique les lois de la multiplication sur une addition (ex: $x + x = 9x^2$).
09	Met x pour chacune des inconnues.
10	Ne comprend pas la signification du mot consécutif.
11	Traduit incorrectement une donnée du problème.
12	Mêle fraction, produit et/ou puissance (ex: tiers, triple, cube).
13	Oublie de répéter la variable ou une quantité dans l'identification des inconnues.
14	Inverse les deux termes d'une soustraction et mêle puissance et produit ($x \text{ exp.}n$ et nx).
15	Mêle puissance et produit ($x \text{ exp.}n$ et nx) ne tient pas compte de la virgule après la variable dans le texte (ou de la priorité d'opération).
16	Mêle puissance et produit ($x \text{ exp.}n$ et nx) et traduit incorrectement une donnée du problème.

- 17 Multiplie au lieu d'additionner ou vice versa et applique les lois de la multiplication sur une addition (ex: $9x + x = 9x^2$).
- 18 Inverse les deux termes d'une soustraction et fait une erreur de symbolisme ($<$, \leq , $>$, \geq).
- 19 Inverse les deux termes d'une soustraction et traduit incorrectement une donnée du problème.
- 20 Inverse les deux termes d'une soustraction et ne tient pas compte de la virgule après la variable dans le texte (ou de la priorité d'opération).
- 21 Multiplie au lieu d'additionner ou vice versa et fait une erreur de symbolisme ($<$, \leq , $>$, \geq).
- 22 Fait une erreur de symbolisme ($<$, \leq , $>$, \geq) et applique les lois de la multiplication sur une addition (ex: $9x + x = 9x^2$).
- 23 Mêlé puissance et produit ($x \text{ exp.}n$ et nx) et fait une erreur de symbolisme ($<$, \leq , $>$, \geq).
- 24 Mêlé puissance et produit ($x \text{ exp.}n$ et nx) et oublie de répéter la variable ou une quantité dans l'identification des inconnues.

Liste des catégories de leurres pour le test correspondant à l'objectif d'algèbre 3.2.6

No.Catégorie	Description
02	Erreur de calcul.
03	Multiplie au lieu de diviser ou vice versa.
04	Soustrait au lieu de diviser.
05	Interchange le dividende et le diviseur (ou un facteur du dividende et du diviseur).
06	Soustrait au lieu d'additionner.
07	Divise au lieu d'additionner ou vice versa.
08	Multiplie au lieu d'additionner ou vice versa.
09	Additionne au lieu de soustraire.
10	Divise au lieu de soustraire.
11	Multiplie au lieu de soustraire ou vice versa.
12	Erreur ou mauvaise compréhension du symbole (\leq , \geq , ...).
13	Oublie de diviser.
14	Oublie de multiplier.
15	Erreur de signe.
16	Erreur de calcul et additionne au lieu de soustraire.
17	Soustrait au lieu de diviser et soustrait au lieu d'additionner.
18	Soustrait au lieu de diviser et additionne au lieu de soustraire.
19	Erreur avec la virgule décimale.
20	Multiplie ou divise avant d'additionner ou soustraire, et avant d'isoler le terme de la variable.
21	Oublie de mettre au même dénominateur.
22	Additionne au lieu de soustraire et oublie de mettre au même dénominateur.

Liste des catégories de leurres pour le test correspondant à l'objectif de géométrie 5.5.2

No.Catégorie	Description
02	Élève au carré le diamètre au lieu du rayon dans le calcul de l'aire.
03	Calcule la circonférence au lieu de l'aire.
05	Double le diamètre et le multiplie par pi au lieu de calculer l'aire.
06	Oublie d'élever au carré le rayon dans le calcul de l'aire.
07	Double le diamètre inutilement dans le calcul de la circonférence.
08	Place la virgule décimale à la mauvaise position.
09	Calcule l'aire au lieu de la circonférence.
10	Oublie de doubler le rayon dans le calcul de la circonférence.
11	Divise le rayon par deux avant de l'élever au carré dans le calcul de l'aire.
12	Multiplie le rayon par pi au lieu de calculer l'aire.
13	Multiplie au lieu de diviser.
15	Ne tient pas compte d'une donnée pertinente dans le calcul de l'aire.
16	Oublie de convertir les unités.
19	Double le diamètre, l'élève au carré, multiplie le résultat par pi au lieu de calculer la circonférence.
21	Élève le diamètre au carré, le multiplie par pi au lieu de calculer la circonférence.

- 22 Élève au carré le produit de pi par le rayon au lieu de calculer l'aire.
- 24 Divise le rayon par deux et multiplie par pi au lieu de calculer la circonférence.
- 25 Divise le diamètre par deux, deux fois de suite et multiplie le quotient par pi au lieu de calculer l'aire.
- 26 Divise le diamètre par deux et multiplie le quotient par pi au lieu de calculer l'aire..
- 27 Double deux fois de suite le rayon et multiplie le produit par pi au lieu de calculer la circonférence.
- 28 Double le rayon, élève le produit au carré, multiplie par pi au lieu de calculer l'aire.
- 29 Calcule l'aire au lieu de la circonférence et place la virgule décimale à la mauvaise position.

**Liste des catégories de leurres pour le test correspondant à
l'objectif de géométrie 5.5.6**

No.Catégorie	Description
02	Calcule le rayon au lieu du diamètre.
03	Erreur dans la position de la virgule décimale.
04	Multiplie au lieu de diviser.
05	Confond pentagone et hexagone.
06	Confond hexagone et octogone.
09	Soustrait au lieu de diviser.
10	Soustrait le double de la base de l'aire.
11	Oublie de multiplier par 2.
14	Confond la mesure de l'aire avec celle de la base.
16	Oublie de diviser par 2.
17	Utilise la formule de l'aire au lieu de celle du périmètre.
18	Soustrait la base ou la diagonale du double de l'aire.
19	Soustrait la base de l'aire.
20	Soustrait la somme des bases de l'aire.
21	Multiplie l'aire par la somme des deux bases.
22	Soustrait la somme de deux côtés parallèles de l'aire.
23	Soustrait la diagonale de l'aire.
24	Multiplie la base ou la somme des bases par l'aire.

- 25 Additionne la hauteur avec l'aire.
- 26 Utilise la formule du périmètre au lieu de celle de l'aire.
- 28 Divise l'aire par deux.
- 31 Confond hexagone et carré.
- 35 Divise par deux le résultat de la soustraction de l'aire par le double de la diagonale.
- 36 Multiplie la circonférence par pi.
- 37 Multiplie le périmètre par la somme des bases.
- 38 Soustrait la base du double de l'aire et place la virgule décimale à la mauvaise position.
- 39 Multiplie l'aire par la hauteur.

Liste des catégories de leurres pour le test correspondant à l'objectif de géométrie 5.5.8

No.Catégorie	Description
02	Calcule l'aire des bases au lieu de A.L. ou A.T.
03	Calcule A.L. au lieu de A.T. ou vice versa.
04	Oublie une base.
05	Oublie de faire la conversion d'unités.
06	Erreur dans la conversion d'unités.
07	Utilise le diamètre au lieu du rayon.
08	Oublie ou ajoute une ou plusieurs faces.
09	Erreur dans le nombre de côtés du polygone.
10	Calcule le périmètre d'une face au lieu de A.L. ou A.T.
11	Oublie de faire la conversion d'unités et oublie ou ajoute une ou plusieurs faces.
12	Oublie de faire la conversion d'unités et utilise le diamètre au lieu du rayon.
13	Calcule A.L. au lieu de A.T. ou vice versa et oublie de faire la conversion d'unités.
14	Calcule A.L. au lieu de A.T. ou vice versa et utilise le diamètre au lieu du rayon.
15	Erreur dans la conversion d'unités et utilise le diamètre au lieu du rayon.
16	Utilise $A = c \times c$ pour l'aire du losange.
17	Calcule A.L. au lieu de A.T. ou vice versa et utilise $A = c \times c$ pour l'aire du losange.
18	Oublie ou ajoute une ou plusieurs faces et utilise $A = c \times c$ pour l'aire du losange.

Légende : A.T. Aire totale
A.L. Aire latérale

C. Figures 2,3,4 et 5 tirées du document Lenman (1994)

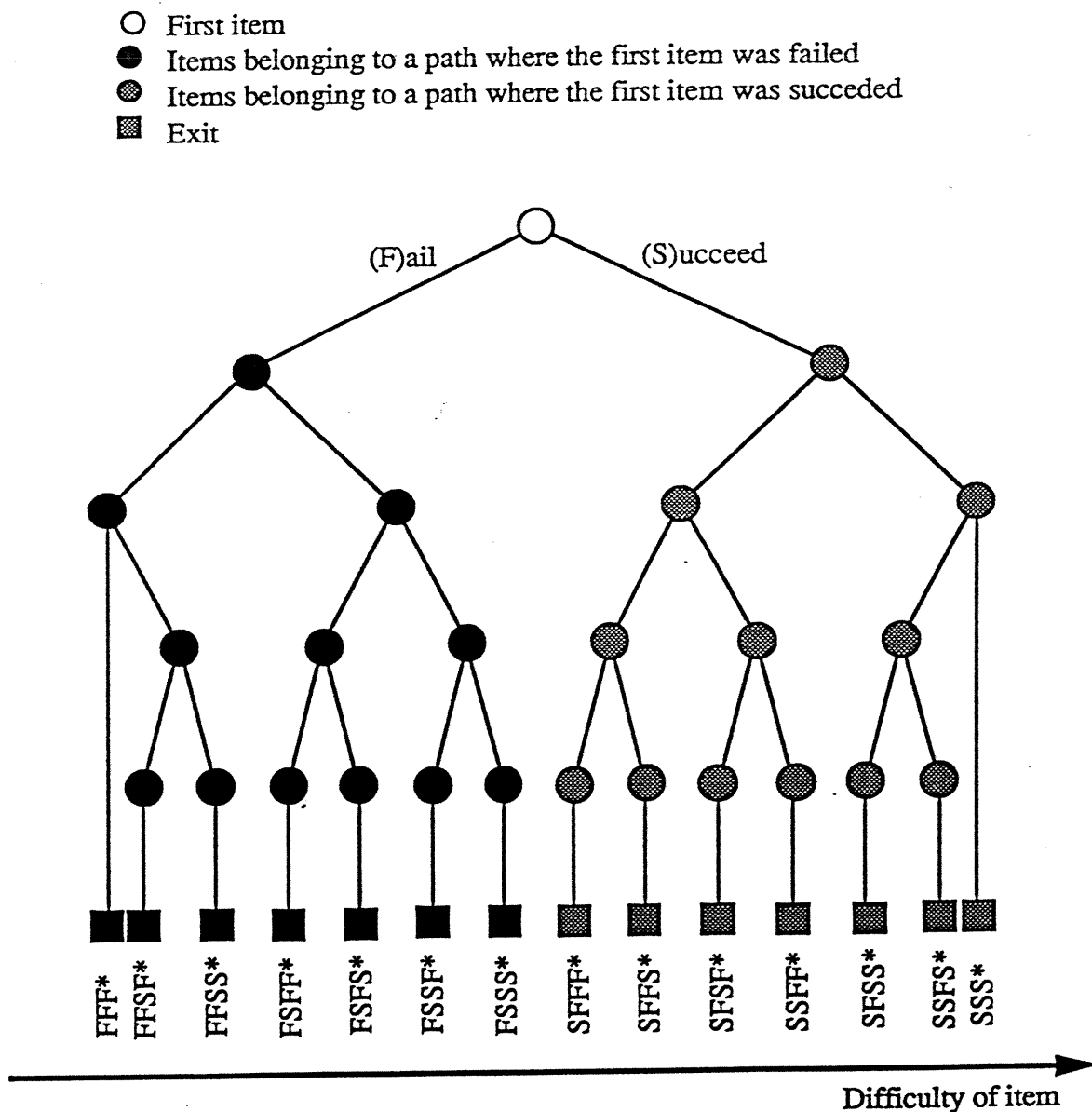


Fig 2. The decision tree for the basic "split - half" testing strategy, as it was implemented in MCATL. A minimum of 3 and a maximum of 5 items are given.

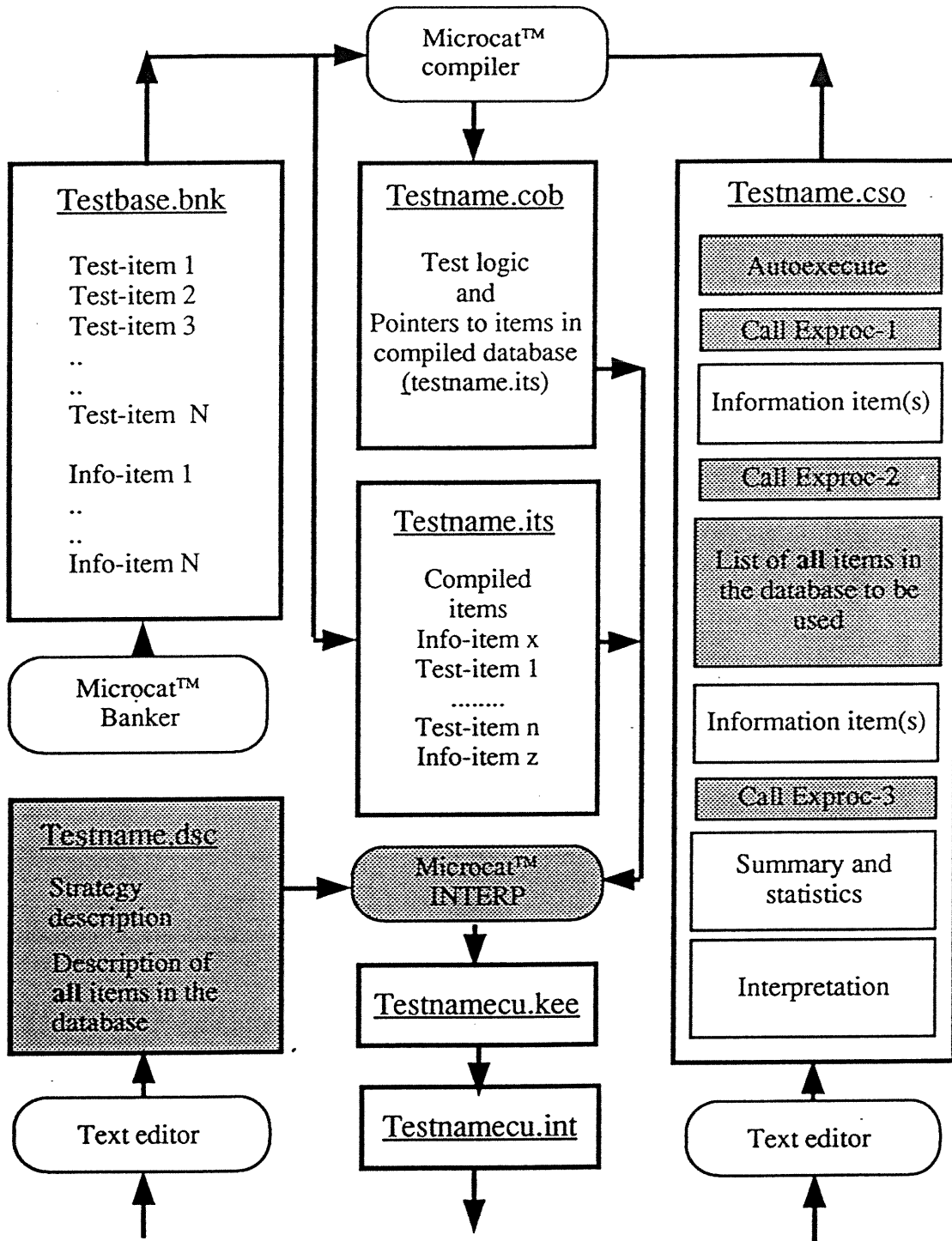


Fig. 3. The modified testing system. Files are indicated by squares and a thick outline, applications by rounded rectangles with thin outline. Sections to which changes have been made are shaded.

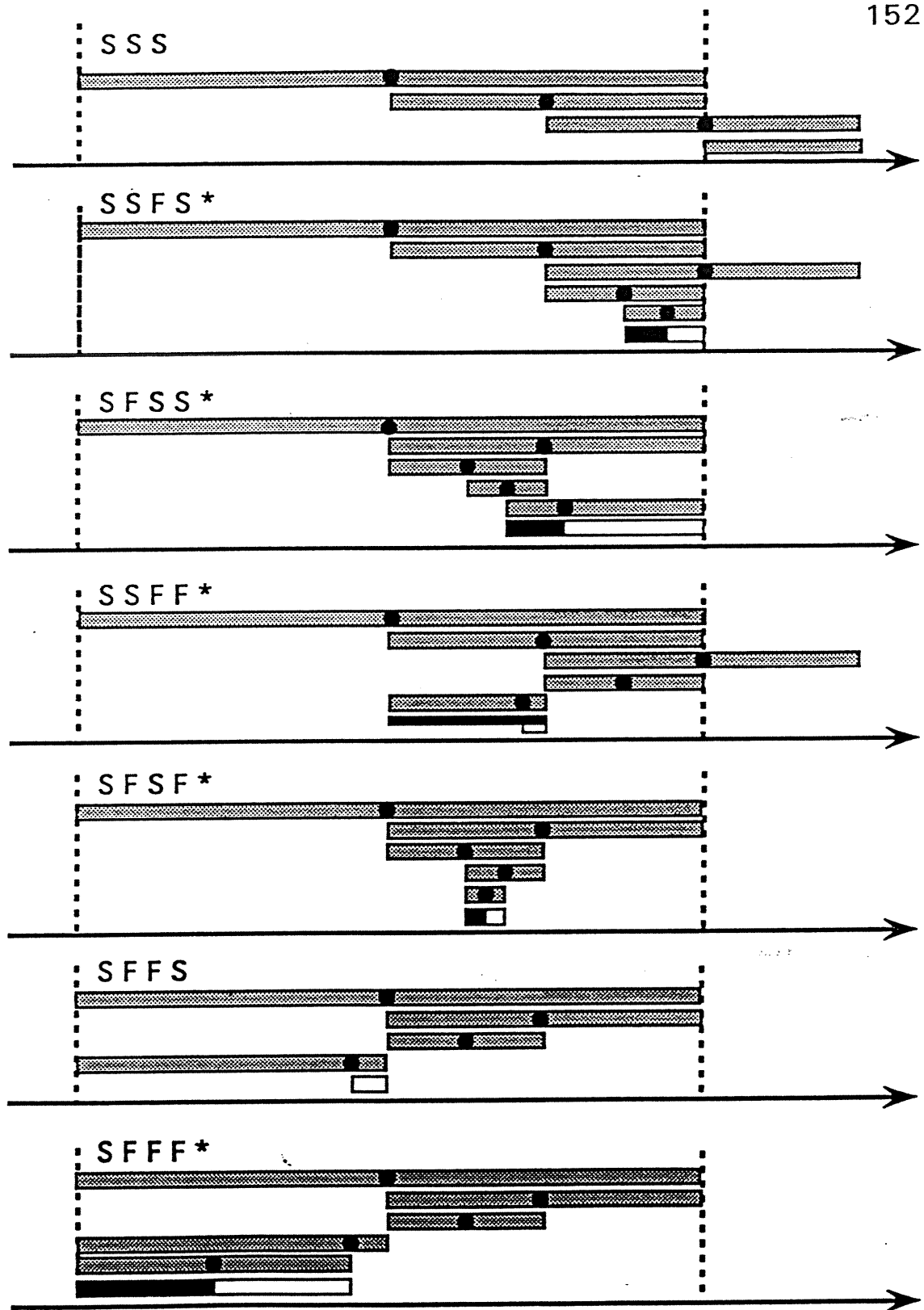


Fig. 5 The partitioning of the "cognitive ability continuum for different paths through the strategy tree. If the last item was failed, the ability is in the black area. If the last item was succeeded the ability is in the white area.

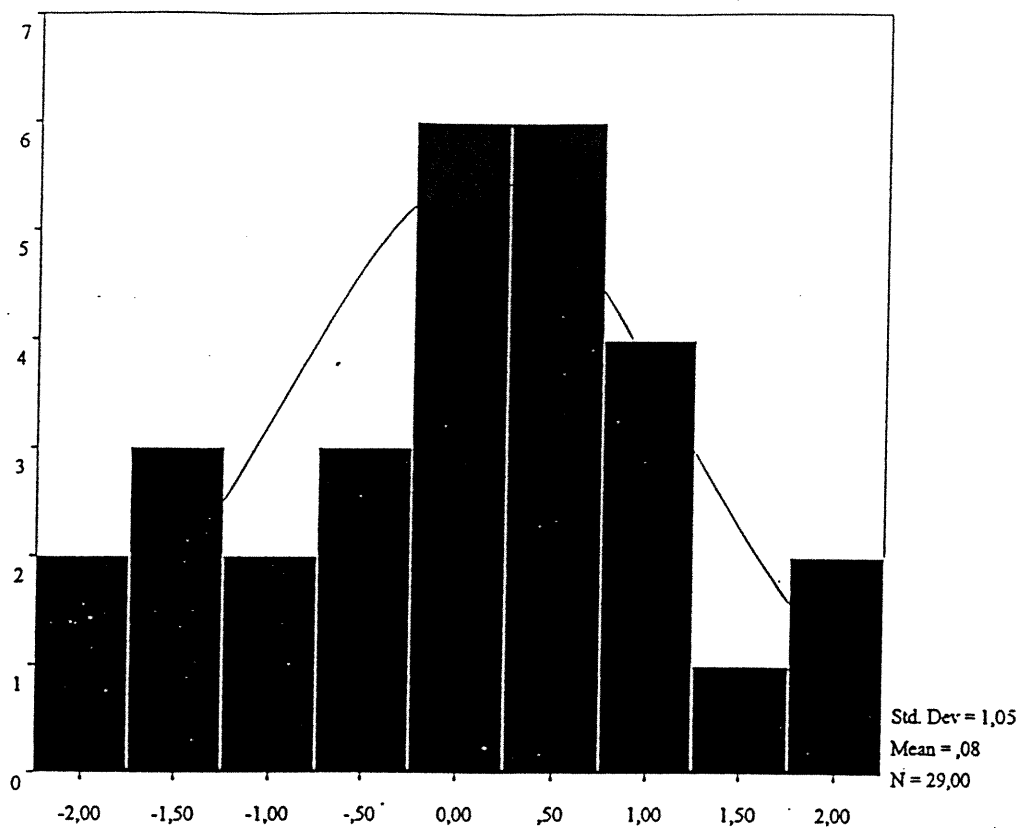
D. Distributions des indices de difficulté générés

Valeurs des indices de difficulté générés pour la banque d'items relative à la catégorie de leurres 2 (23 items)

Variable LDEUX

Mean	,080	Std Dev	1,052
Kurtosis	-,392	S.E. Kurt	,845
Skewness	-,210	S.E. Skew	,434
Minimum	-2,045	Maximum	2,173

Valid observations - 29 Missing observations - 0



LDEUX
(Valeurs des indices de difficultés générées pour la catégorie de leurre #2)

-2,045*
-1,856*
-1,313*
-1,281
-1,256
-1,067
-,780
-,562
-,447
-,311
-,081
-,015
,040
,122
,127
,207
,351
,392
,568
,623
,708
,740
,843
1,034
1,067
1,164
1,416*
1,761*
2,173*

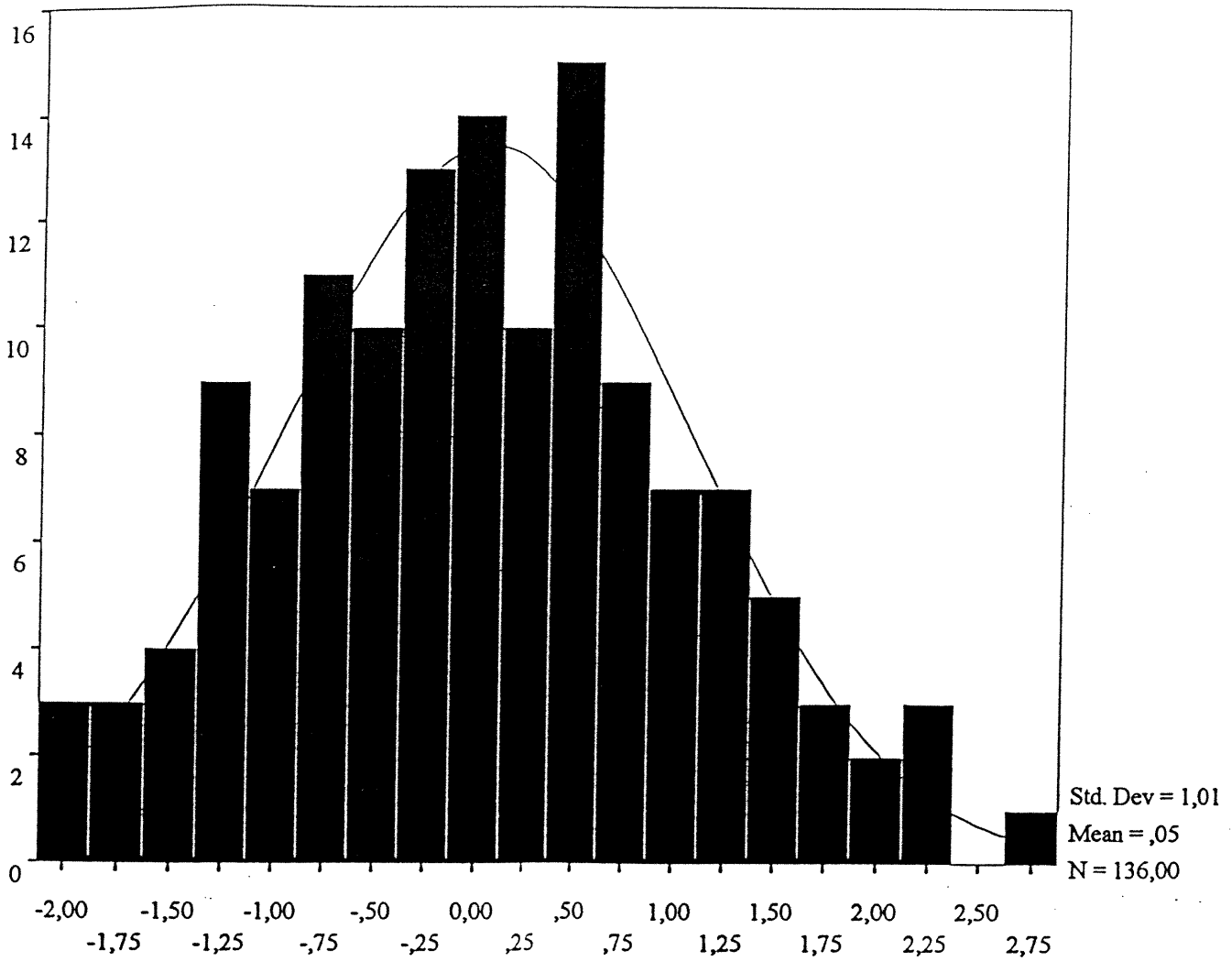
* =valeurs aux extrémités retirées

Valeurs des indices de difficulté générés pour la banque d'items relative à la catégorie de leurres 8 (109 items)

Variable LHUIT

Mean	,054	Std Dev	1,007
Kurtosis	-,273	S.E. Kurt	,413
Skewness	,238	S.E. Skew	,208
Minimum	-1,994	Maximum	2,840

Valid observations - 136 Missing observations - 0



LHUIT

(Valeurs des indices de difficultés générées pour la catégorie de leurre #8)

-1,994*	-,443	,392	1,575*
-1,985*	-,429	,408	1,642*
-1,878*	-,418	,436	1,684*
-1,716*	-,399	,446	1,758*
-1,685*	-,375	,470	1,952*
-1,680*	-,372	,489	2,007*
-1,590*	-,363	,490	2,289*
-1,484*	-,362	,527	2,328*
-1,460*	-,358	,538	2,367*
-1,385*	-,348	,561	2,840*
-1,331*	-,304	,581	
-1,276*	-,295	,585	
-1,273*	-,276	,592	
-1,240*	-,276	,594	
-1,222	-,267	,601	
-1,201	-,213	,658	
-1,189	-,204	,663	
-1,184	-,202	,684	
-1,155	-,109	,705	
-1,080	-,067	,707	
-1,074	-,065	,760	
-1,060	-,054	,794	
-1,000	-,053	,795	
-,991	-,034	,847	
-,902	-,033	,887	
-,894	-,016	,904	
-,830	-,003	,969	
-,823	,002	,988	
-,744	,023	1,027	
-,721	,026	1,077	
-,696	,092	1,113	
-,678	,099	1,139	
-,672	,135	1,140	
-,656	,144	1,170	
-,636	,164	1,221	
-,630	,179	1,237	
-,629	,182	1,333	
-,559	,193	1,349	
-,516	,312	1,389	
-,495	,322	1,427*	
-,471	,357	1,517*	
-,449	,366	1,520*	

* = valeurs aux extrémités retirées.

E. Liste des items et de leur indice de difficulté**BANQUE LDEUX (dont les items comprennent la catégorie de leurre #2 parmi les choix de réponse.)****# items Indice difficulté (b)**

#LDEUX290 = -1,281
#LDEUX514 = -1,256
#LDEUX702 = -1,067
#LDEUX238 = -,780
#LDEUX201 = -,562
#LDEUX205 = -,447
#LDEUX204 = -,311
#LDEUX242 = -,081
#LDEUX203 = -,015
#LDEUX240 = ,040
#LDEUX888 = ,122
#LDEUX244 = ,127
#LDEUX279 = ,207
#LDEUX202 = ,351
#LDEUX720 = ,392
#LDEUX722 = ,568
#LDEUX723 = ,623
#LDEUX511 = ,708
#LDEUX263 = ,740
#LDEUX283 = ,843
#LDEUX908 = 1,034
#LDEUX736 = 1,067
#LDEUX737 = 1,164

BANQUE LHUIT (dont les items comprennent la catégorie de beurre #8 parmi les choix de réponse.)

# items	Indice difficulté (b)			
#LHUIT254 =	-1,22	#LHUIT240 =	-,109 #LHUIT910 =	,795
#LHUIT255 =	-1,20	#LHUIT244 =	-,067 #LHUIT269 =	,847
#LHUIT256 =	-1,18	#LHUIT279 =	-,065 #LHUIT274 =	,887
#LHUIT291 =	-1,18	#LHUIT294 =	-,054 #LHUIT282 =	,904
#LHUIT506 =	-1,15	#LHUIT207 =	-,053 #LHUIT289 =	,969
#LHUIT513 =	-1,08	#LHUIT209 =	-,034 #LHUIT734 =	,988
#LHUIT701 =	-1,07	#LHUIT210 =	-,033 #LHUIT904 =	1,027
#LHUIT707 =	-1,06	#LHUIT218 =	-,016 #LHUIT251 =	1,077
#LHUIT257 =	-1,00	#LHUIT232 =	-,003 #LHUIT738 =	1,113
#LHUIT706 =	-,991	#LHUIT233 =	,002 #LHUIT739 =	1,139
#LHUIT258 =	-,902	#LHUIT246 =	,023 #LHUIT740 =	1,140
#LHUIT259 =	-,894	#LHUIT250 =	,026 #LHUIT737 =	1,170
#LHUIT507 =	-,830	#LHUIT277 =	,092 #LHUIT295 =	1,221
#LHUIT705 =	-,823	#LHUIT733 =	,099 #LHUIT735 =	1,237
#LHUIT290 =	-,744	#LHUIT241 =	,135 #LHUIT741 =	1,333
#LHUIT514 =	-,721	#LHUIT224 =	,144 #LHUIT742 =	1,349
#LHUIT703 =	-,696	#LHUIT236 =	,164 #LHUIT743 =	1,389
#LHUIT708 =	-,678	#LHUIT721 =	,179	
#LHUIT206 =	-,672	#LHUIT732 =	,182	
#LHUIT208 =	-,656	#LHUIT720 =	,193	
#LHUIT216 =	-,636	#LHUIT722 =	,312	
#LHUIT217 =	-,630	#LHUIT723 =	,322	
#LHUIT219 =	-,629	#LHUIT502 =	,357	
#LHUIT220 =	-,559	#LHUIT508 =	,366	
#LHUIT231 =	-,516	#LHUIT710 =	,392	
#LHUIT234 =	-,495	#LHUIT712 =	,408	
#LHUIT235 =	-,471	#LHUIT713 =	,436	
#LHUIT247 =	-,449	#LHUIT714 =	,446	
#LHUIT248 =	-,443	#LHUIT715 =	,470	
#LHUIT271 =	-,429	#LHUIT716 =	,489	
#LHUIT285 =	-,418	#LHUIT718 =	,490	
#LHUIT709 =	-,399	#LHUIT711 =	,527	
#LHUIT884 =	-,375	#LHUIT262 =	,538	
#LHUIT223 =	-,372	#LHUIT503 =	,561	
#LHUIT237 =	-,363	#LHUIT512 =	,581	
#LHUIT225 =	-,362	#LHUIT717 =	,585	
#LHUIT243 =	-,358	#LHUIT725 =	,592	
#LHUIT245 =	-,348	#LHUIT727 =	,594	
#LHUIT296 =	-,304	#LHUIT729 =	,601	
#LHUIT239 =	-,295	#LHUIT719 =	,658	
#LHUIT221 =	-,276	#LHUIT501 =	,663	
#LHUIT299 =	-,276	#LHUIT726 =	,684	
#LHUIT280 =	-,267	#LHUIT731 =	,705	
#LHUIT222 =	-,213	#LHUIT270 =	,707	
#LHUIT238 =	-,204	#LHUIT275 =	,760	
#LHUIT242 =	-,202	#LHUIT509 =	,794	

F. Compilation de l'étude de vérification de l'algorithme "stradaptatif" de la phase I du test

Score attendu	Réussite 1/1		Réussite 9/10		Réussite 4/5		Réussite 1/2		Réussite 1/4		Réussite 0/1	
	Maîtrise parfaite		Maîtrise élevée		Maîtrise		Absence de maîtrise		Absence de maîtrise		Absence de maîtrise	
	nb items	score	nb items	score	nb items	score	nb items	score	nb items	score	nb items	score
1	30	100.0	32	90.6	40	80.0	34	50.0	31	35.5	15	0.0
2	-	-	32	93.8	36	80.6	28	42.9	25	28.0	-	-
3	-	-	36	86.1	39	79.5	34	55.9	25	28.0	-	-
4	-	-	35	85.7	44	81.8	31	48.4	27	33.3	-	-
5	-	-	32	93.8	36	83.3	26	74.3	21	28.6	-	-
6	-	-	32	96.9	37	83.8	34	61.8	23	21.7	-	-
7	-	-	36	88.9	32	78.1	32	37.5	26	30.8	-	-
8	-	-	34	94.1	41	78.0	41	61.0	26	30.8	-	-
9	-	-	32	96.9	40	80.0	33	57.6	29	31.0	-	-
10	-	-	34	94.1	38	81.6	33	57.6	25	32.0	-	-
11	-	-	34	91.2	31	83.9	39	69.2	26	26.9	-	-
12	-	-	36	91.7	38	89.5	31	67.7	24	29.2	-	-
13	-	-	36	86.1	34	82.4	33	69.7	23	17.4	-	-
14	-	-	34	94.1	31	80.6	38	55.3	23	30.4	-	-
15	-	-	34	91.2	33	84.8	33	63.6	27	11.8	-	-
16	-	-	32	90.6	36	77.8	38	72.5	26	30.8	-	-
17	-	-	34	91.2	40	82.5	40	74.3	21	19.0	-	-
18	-	-	32	93.8	34	88.3	36	74.4	21	23.8	-	-
19	-	-	36	86.1	38	89.5	27	44.0	26	30.8	-	-
20	-	-	34	91.2	36	77.8	43	60.5	17	6.0	-	-
VALEURS MOYENNES			33.9	91.4	36.7	82.2	34.2	59.9	24.6	26.3	-	-

G. Fiches diagnostiques individuelles à la PHASE I

Test interpretation for eleve001

29 May 96

Projet de recherche subventionné par le CRSH - Canada.
Responsable: Clément Dassa (Université de Montréal).

----- Résultats de la phase I du test -----

-- Parcours de l'élève pour le test de l'objectif 5.5.2 --

No item	Contenu notionnel	Contexte diagnost.	Bonne rép.	Choix élève	Le résultat	Numéro leurre
HEOMB 506	1	A	3	3	1	
HEOMB 701	1	A	3	3	1	
HEOMB 707	1	A	1	1	1	
HEOMB 507	2	A	4	4	1	
HEOMB 514	2	A	3	3	1	
HEOMB 901	2	A	2	2	1	
HEOMB 219	1	C	3	3	1	
HEOMB 234	1	C	2	2	1	
HEOMB 884	1	C	2	2	1	
HEOMB 296	2	C	2	2	1	
HEOMB 299	2	C	2	2	1	
HEOMB 294	2	C	3	3	1	
HEOMB 233	1	E	4	4	1	
HEOMB 250	1	E	3	3	1	
HEOMB 733	1	E	3	3	1	
HEOMB 721	2	E	4	4	1	
HEOMB 202	2	E	3	3	1	
HEOMB 723	2	E	1	1	1	
HEOMB 716	1	G	2	2	1	
HEOMB 512	1	G	4	4	1	
HEOMB 729	1	G	3	3	1	
HEOMB 263	2	G	3	3	1	
HEOMB 731	2	G	3	3	1	
HEOMB 730	2	G	1	1	1	
HEOMB 289	1	K	4	4	1	
HEOMB 251	1	K	2	2	1	
HEOMB 253	1	K	2	2	1	
HEOMB 297	2	K	4	4	1	
HEOMB 741	2	K	1	1	1	
HEOMB 252	2	K	4	4	1	

L'élève a réussi 30 questions sur 30; sa moyenne: 100%

*** Excellent ! Aucune erreur dans ton test. ***

----- Fin du test pour l'objectif 5.5.2 -----

Test interpretation for eleve001

29 May 96

Projet de recherche subventionné par le CRSH - Canada.
Responsable: Clément Dassa (Université de Montréal).

----- Résultats de la phase I du test -----

-- Parcours de l'élève pour le test de l'objectif 5.5.2 --

No item	Contenu notionnel	Contexte diagnost.	Bonne rép.	Choix élève	Le résultat	Numéro leurre
HEOMB 506	1	A	3	3	1	
HEOMB 701	1	A	3	3	1	
HEOMB 707	1	A	1	1	1	
HEOMB 507	2	A	4	4	1	
HEOMB 514	2	A	3	3	1	
HEOMB 901	2	A	2	2	1	
HEOMB 219	1	C	3	3	1	
HEOMB 234	1	C	2	2	1	
HEOMB 884	1	C	2	2	1	
HEOMB 296	2	C	2	2	1	
HEOMB 299	2	C	2	2	1	
HEOMB 294	2	C	3	3	1	
HEOMB 233	1	E	4	4	1	
HEOMB 250	1	E	3	3	1	
HEOMB 733	1	E	3	3	1	
HEOMB 721	2	E	4	4	1	
HEOMB 202	2	E	3	3	1	
HEOMB 723	2	E	1	1	1	
HEOMB 716	1	G	2	2	1	
HEOMB 512	1	G	4	4	1	
HEOMB 729	1	G	3	3	1	
HEOMB 263	2	G	3	3	1	
HEOMB 731	2	G	3	3	1	
HEOMB 730	2	G	1	1	1	
HEOMB 289	1	K	4	4	1	
HEOMB 251	1	K	2	2	1	
HEOMB 253	1	K	2	2	1	
HEOMB 297	2	K	4	4	1	
HEOMB 741	2	K	1	1	1	
HEOMB 252	2	K	4	1	0	11
HEOMB 742	2	K	3	1	0	12
HEOMB 737	2	K	3	1	0	3

L'élève a réussi 29 questions sur 32; sa moyenne: 90.625%

-- Description des résultats --

Contenus notionnels :

2. Calcul de l'aire:

14 item(s) réussi(s) sur 17 questions posées.

Contextes diagnostiques :

K: La mesure du rayon ou du diamètre est identifiée
dans le texte en nb.entiers ou décimaux.

Résultat: 5 item(s) réussi(s) sur 8 questions.

-- Interprétation en fonction des leurres proposés. --

L'élève a choisi 1 fois sur 16 le leurre 3.

Desc: calcule le double du rayon au lieu de l'élever au carré dans le calcul de l'aire.

Form: aire = (ray x 2) x pi.

L'élève a choisi 1 fois sur 2 le leurre 11.

Desc: divise le rayon par deux avant de l'élever au carré dans le calcul de l'aire.

Form: aire = [(ray divisé par 2) élevé au carré] x pi.

L'élève a choisi 1 fois sur 11 le leurre 12.

Desc: multiplie le rayon par pi au lieu de calculer l'aire.

Form: aire = ray x pi.

----- Fin de la phase I du test -----

----- Fin de la phase II du test -----

----- Fin du test pour l'objectif 5.5.2 -----

Test interpretation for eleve01

30 May 96

Projet de recherche subventionné par le CRSH - Canada.
Responsable: Clément Dassa (Université de Montréal).

----- Résultats de la phase I du test -----

-- Parcours de l'élève pour le test de l'objectif 5.5.2 --

No item	Contenu notionnel	Contexte diagnost.	Bonne rép.	Choix élève	Le résultat	Numéro leurre
HEOMB 506	1	A	3	3	1	
HEOMB 701	1	A	3	3	1	
HEOMB 707	1	A	1	1	1	
HEOMB 507	2	A	4	4	1	
HEOMB 514	2	A	3	2	0	2
HEOMB 705	2	A	2	2	1	
HEOMB 702	2	A	2	2	1	
HEOMB 901	2	A	2	2	1	
HEOMB 219	1	C	3	3	1	
HEOMB 234	1	C	2	1	0	10
HEOMB 220	1	C	2	2	1	
HEOMB 235	1	C	3	3	1	
HEOMB 884	1	C	2	2	1	
HEOMB 296	2	C	2	2	1	
HEOMB 299	2	C	2	2	1	
HEOMB 294	2	C	3	2	0	15
HEOMB 245	2	C	3	3	1	
HEOMB 239	2	C	3	2	0	3
HEOMB 233	1	E	4	4	1	
HEOMB 250	1	E	3	3	1	
HEOMB 733	1	E	3	3	1	
HEOMB 721	2	E	4	4	1	
HEOMB 202	2	E	3	4	0	3
HEOMB 732	2	E	3	1	0	12
HEOMB 224	2	E	3	3	1	
HEOMB 236	2	E	3	3	1	
HEOMB 716	1	G	2	2	1	
HEOMB 512	1	G	4	1	0	10
HEOMB 711	1	G	4	4	1	
HEOMB 503	1	G	3	3	1	
HEOMB 725	1	G	2	3	0	7
HEOMB 263	2	G	3	3	1	
HEOMB 731	2	G	3	3	1	
HEOMB 730	2	G	1	1	1	
HEOMB 289	1	K	4	4	1	
HEOMB 251	1	K	2	2	1	
HEOMB 253	1	K	2	2	1	
HEOMB 297	2	K	4	4	1	
HEOMB 741	2	K	1	1	1	
HEOMB 252	2	K	4	4	1	

L'élève a réussi 32 questions sur 40; sa moyenne: 80%

-- Description des résultats --

Contenus notionnels :

1. Calcul de la circonférence:
16 item(s) réussi(s) sur 19 questions posées.

2. Calcul de l'aire:
16 item(s) réussi(s) sur 21 questions posées.

Contextes diagnostiques :

- Problèmes abstraits

A: La mesure du rayon ou du diamètre est identifiée sur la figure.
Résultat: 7 item(s) réussi(s) sur 8 questions.

C: La mesure du rayon ou du diamètre est identifiée dans le texte et l'aire est demandé en nb.entiers.
Résultat: 7 item(s) réussi(s) sur 10 questions.

E: La mesure du rayon ou du diamètre est identifiée dans le texte et l'aire est demandé en nb.décimaux
Résultat: 6 item(s) réussi(s) sur 8 questions.

- Problèmes réels

G: La mesure du rayon ou du diamètre est identifiée sur la figure en nb.entiers ou décimaux.
Résultat: 6 item(s) réussi(s) sur 8 questions.

-- Interprétation en fonction des leurre proposés. --

L'élève a choisi 1 fois sur 4 le leurre 2.
Desc: élève au carré le diamètre au lieu du rayon dans le calcul de l'aire.
Form: aire = [(diam) élevé au carré] x pi.

L'élève a choisi 2 fois sur 20 le leurre 3.
Desc: calcule le double du rayon au lieu de l'élever au carré dans le calcul de l'aire.
Form: aire = (ray x 2) x pi.

L'élève a choisi 1 fois sur 8 le leurre 7.
Desc: double le diamètre inutilement dans le calcul de la circonférence.
Form: circ = (diam x 2) x pi.

L'élève a choisi 2 fois sur 9 le leurre 10.
Desc: oublie de doubler le rayon dans le calcul de la circonférence.
Form: circ = ray x pi.

L'élève a choisi 1 fois sur 15 le leurre 12.
Desc: multiplie le rayon par pi au lieu de calculer l'aire.
Form: aire = ray x pi.

L'élève a choisi 1 fois sur 3 le leurre 15.
Desc: ne tient pas compte d'une donnée pertinente dans le calcul de l'aire.

----- Fin de la phase I du test -----

----- Fin de la phase II du test -----

----- Fin du test pour l'objectif 5.5.2 -----

Test interpretation for

21 Jun 96

Projet de recherche subventionné par le CRSH - Canada.
Responsable: Clément Dassa (Université de Montréal).

----- Résultats de la phase I du test -----

-- Parcours de l'élève pour le test de l'objectif 5.5.2 --

No item	Contenu notionnel	Contexte diagnost.	Bonne rép.	Choix élève	Le résultat	Numéro leurre
HEOMB 506	1	A	3	1	0	8
HEOMB 256	1	A	2	2	1	
HEOMB 291	1	A	1	4	0	8
HEOMB 255	1	A	4	3	0	7
HEOMB 254	1	A	1	1	1	
HEOMB 219	1	C	3	3	1	
HEOMB 234	1	C	2	2	1	
HEOMB 884	1	C	2	4	0	8
HEOMB 217	1	C	1	2	0	7
HEOMB 206	1	C	1	2	0	7
HEOMB 233	1	E	4	4	1	
HEOMB 250	1	E	3	2	0	10
HEOMB 210	1	E	3	3	1	
HEOMB 232	1	E	3	1	0	8
HEOMB 218	1	E	1	1	1	
HEOMB 721	2	E	4	1	0	3
HEOMB 224	2	E	3	4	0	12
HEOMB 241	2	E	2	3	0	3
HEOMB 716	1	G	2	2	1	
HEOMB 512	1	G	4	1	0	10
HEOMB 711	1	G	4	4	1	
HEOMB 503	1	G	3	3	1	
HEOMB 725	1	G	2	3	0	7
HEOMB 263	2	G	3	3	1	
HEOMB 731	2	G	3	3	1	
HEOMB 730	2	G	1	1	1	
HEOMB 289	1	K	4	4	1	
HEOMB 251	1	K	2	1	0	10
HEOMB 902	1	K	3	3	1	
HEOMB 904	1	K	4	4	1	
HEOMB 739	1	K	4	4	1	
HEOMB 297	2	K	4	1	0	12
HEOMB 736	2	K	2	4	0	3
HEOMB 292	2	K	4	3	0	3

L'élève a réussi 17 questions sur 34; sa moyenne: 50%

-- Description des résultats --

Contenus notionnels :

1. Calcul de la circonférence:
14 item(s) réussi(s) sur 25 questions posées.
2. Calcul de l'aire:
3 item(s) réussi(s) sur 9 questions posées.

Contextes diagnostiques :

- Problèmes abstraits

A: La mesure du rayon ou du diamètre est identifiée sur la figure.

Résultat: 2 item(s) réussi(s) sur 5 questions.

C: La mesure du rayon ou du diamètre est identifiée dans le texte et l'aire est demandé en nb.entiers.

Résultat: 2 item(s) réussi(s) sur 5 questions.

E: La mesure du rayon ou du diamètre est identifiée dans le texte et l'aire est demandé en nb.décimaux.

Résultat: 3 item(s) réussi(s) sur 8 questions.

- Problèmes réels

G: La mesure du rayon ou du diamètre est identifiée sur la figure en nb.entiers ou décimaux.

Résultat: 6 item(s) réussi(s) sur 8 questions.

K: La mesure du rayon ou du diamètre est identifiée dans le texte en nb.entiers ou décimaux.

Résultat: 4 item(s) réussi(s) sur 8 questions.

-- Interprétation en fonction des leurres proposés. --

L'élève a choisi 4 fois sur 9 le leurre 3.

Desc: calcule le double du rayon au lieu de l'élever au carré dans le calcul de l'aire.

Form: aire = (ray x 2) x pi.

L'élève a choisi 4 fois sur 13 le leurre 7.

Desc: double le diamètre inutilement dans le calcul de la circonférence.

Form: circ = (diam x 2) x pi.

L'élève a choisi 4 fois sur 29 le leurre 8.

Desc: place la virgule décimale à la mauvaise position.

L'élève a choisi 3 fois sur 11 le leurre 10.

Desc: oublie de doubler le rayon dans le calcul de la circonférence.

Form: circ = ray x pi.

L'élève a choisi 2 fois sur 8 le leurre 12.

Desc: multiplie le rayon par pi au lieu de calculer l'aire.

Form: aire = ray x pi.

----- Fin de la phase I du test -----

----- Fin de la phase II du test -----

----- Fin du test pour l'objectif 5.5.2 -----

Test interpretation for eleve001

29 May 96

Projet de recherche subventionné par le CRSH - Canada.
Responsable: Clément Dassa (Université de Montréal).

----- Résultats de la phase I du test -----

-- Parcours de l'élève pour le test de l'objectif 5.5.2 --

No item	Contenu notionnel	Contexte diagnost.	Bonne rép.	Choix élève	Le résultat	Numéro leurre
HEOMB 506	1	A	3	2	0	8
HEOMB 256	1	A	2	2	1	
HEOMB 291	1	A	1	3	0	9
HEOMB 255	1	A	4	3	0	7
HEOMB 254	1	A	1	4	0	9
HEOMB 219	1	C	3	3	1	
HEOMB 234	1	C	2	3	0	9
HEOMB 220	1	C	2	2	1	
HEOMB 235	1	C	3	1	0	8
HEOMB 296	2	C	2	2	1	
HEOMB 299	2	C	2	1	0	8
HEOMB 239	2	C	3	3	1	
HEOMB 280	2	C	4	3	0	11
HEOMB 233	1	E	4	4	1	
HEOMB 250	1	E	3	3	1	
HEOMB 733	1	E	3	4	0	10
HEOMB 277	1	E	2	3	0	9
HEOMB 218	1	E	1	4	0	8
HEOMB 716	1	G	2	3	0	8
HEOMB 712	1	G	3	2	0	8
HEOMB 260	1	G	2	3	0	10
HEOMB 289	1	K	4	1	0	8
HEOMB 910	1	K	4	4	1	
HEOMB 281	1	K	2	2	1	
HEOMB 251	1	K	2	1	0	10
HEOMB 902	1	K	3	3	1	
HEOMB 297	2	K	4	3	0	11
HEOMB 736	2	K	2	2	1	
HEOMB 293	2	K	3	1	0	12
HEOMB 737	2	K	3	2	0	8
HEOMB 750	2	K	2	3	0	12

L'élève a réussi 11 questions sur 31; sa moyenne: 35.484%

-- Description des résultats --

Contenus notionnels :

1. Calcul de la circonférence:
8 item(s) réussi(s) sur 22 questions posées.
2. Calcul de l'aire:
3 item(s) réussi(s) sur 9 questions posées.

Contextes diagnostiques :

- Problèmes abstraits

A: La mesure du rayon ou du diamètre est identifiée sur la figure.
 Résultat: 1 item(s) réussi(s) sur 5 questions.

C: La mesure du rayon ou du diamètre est identifiée dans le texte et l'aire est demandé en nb.entiers.
 Résultat: 4 item(s) réussi(s) sur 8 questions.

E: La mesure du rayon ou du diamètre est identifiée dans le texte et l'aire est demandé en nb.décimaux
 Résultat: 2 item(s) réussi(s) sur 5 questions.

- Problèmes réels

G: La mesure du rayon ou du diamètre est identifiée sur la figure en nb.entiers ou décimaux.
 Résultat: 0 item(s) réussi(s) sur 3 questions.

K: La mesure du rayon ou du diamètre est identifiée dans le texte en nb.entiers ou décimaux.
 Résultat: 4 item(s) réussi(s) sur 10 questions.

 -- Interprétation en fonction des leures proposés. --

L'élève a choisi 1 fois sur 10 le leurre 7.
 Desc: double le diamètre inutilement dans le calcul de la circonférence.
 Form: $\text{circ} = (\text{diam} \times 2) \times \pi$.

L'élève a choisi 8 fois sur 25 le leurre 8.
 Desc: place la virgule décimale à la mauvaise position.

L'élève a choisi 4 fois sur 17 le leurre 9.
 Desc: calcule l'aire au lieu de la circonférence.

L'élève a choisi 3 fois sur 12 le leurre 10.
 Desc: oublie de doubler le rayon dans le calcul de la circonférence.
 Form: $\text{circ} = \text{ray} \times \pi$.

L'élève a choisi 2 fois sur 2 le leurre 11.
 Desc: divise le rayon par deux avant de l'élever au carré dans le calcul de l'aire.
 Form: $\text{aire} = [(\text{ray} \text{ divisé par } 2) \text{ élevé au carré}] \times \pi$.

L'élève a choisi 2 fois sur 5 le leurre 12.
 Desc: multiplie le rayon par pi au lieu de calculer l'aire.
 Form: $\text{aire} = \text{ray} \times \pi$.

----- Fin de la phase I du test -----

----- Fin de la phase II du test -----

----- Fin du test pour l'objectif 5.5.2 -----

H. Profils des élèves à la PHASE II.

Profil des élèves à la PHASE II
Banque d'items avec le leurre no.2
Sélection aléatoire du leurre

#	Nb.d'items présentés	Consistance à l'erreur	Estimation difficulté cognitive (théta)	Variance Erreur	#	Nb.d'items présentés	Consistance à l'erreur	Estimation difficulté cognitive (théta)	Variance Erreur
1	10	30.0	-1.147	0.188	51	12	25.0	-0.828	0.181
2	12	25.0	-1.208	0.178	52	16	18.8	-1.279	0.167
3	11	27.3	-1.179	0.182	53	20	15.0	-1.314	0.162
4	11	27.3	-1.179	0.182	54	10	40.0	0.215	0.158
5	10	40.0	-0.848	0.167	55	10	30.0	-1.147	0.188
6	11	27.3	-1.179	0.182	56	10	30.0	-0.228	0.191
7	10	30.0	-1.147	0.188	57	12	25.0	-1.208	0.178
8	10	30.0	-1.147	0.188	58	12	25.0	-1.164	0.185
9	10	30.0	-1.147	0.188	59	15	20.0	-1.249	0.171
10	11	27.3	-1.179	0.182	60	15	20.0	-1.249	0.171
11	10	30.0	-1.147	0.188	61	20	10.0	-1.632	0.219
12	10	40.0	0.255	0.157	62	20	10.0	-1.632	0.219
13	10	30.0	-1.147	0.188	63	20	15.0	-1.314	0.162
14	11	36.4	-0.811	0.172	64	11	36.4	-0.811	0.172
15	11	27.3	-1.118	0.193	65	10	40.0	0.248	0.165
16	10	30.0	-1.147	0.188	66	10	40.0	0.456	0.156
17	10	30.0	-1.147	0.188	67	13	23.1	-1.233	0.174
18	10	30.0	-0.623	0.186	68	10	30.0	0.042	0.176
19	10	30.0	-1.147	0.188	69	10	40.0	-0.848	0.167
20	19	15.8	-1.307	0.163	70	10	30.0	-1.147	0.188
21	10	30.0	-1.147	0.188	71	10	30.0	-1.147	0.188
22	12	25.0	-1.208	0.178	72	10	40.0	-0.848	0.167
23	10	40.0	-0.848	0.167	73	18	16.7	-1.299	0.164
24	11	27.3	-1.118	0.193	74	10	40.0	-0.848	0.167
25	11	27.3	-1.179	0.182	75	10	30.0	-1.147	0.188
26	11	27.3	-1.179	0.182	76	13	23.1	-1.198	0.179
27	10	30.0	-1.147	0.188	77	18	16.7	-1.277	0.167

28	11	27.3	-1.179	0.182	78	13	23.1	-1.233	0.174
29	10	30.0	-1.147	0.188	79	12	25.0	-1.208	0.178
30	10	30.0	-1.147	0.188	80	10	30.0	-1.147	0.188
31	10	30.0	-1.147	0.188	81	12	25.0	-1.208	0.178
32	10	30.0	-1.147	0.188	82	10	50.0	0.208	0.152
33	10	30.0	-1.147	0.188	83	13	23.1	-1.233	0.174
34	10	30.0	-1.147	0.188	84	20	10.0	-1.632	0.219
35	11	27.3	-1.179	0.182	85	17	17.7	-1.290	0.165
36	10	30.0	-1.147	0.188	86	10	40.0	-0.848	0.167
37	10	30.0	-1.147	0.188	87	20	10.0	-1.632	0.219
38	11	27.3	-1.179	0.182	88	17	17.7	-1.290	0.165
39	10	30.0	-1.147	0.188	89	10	30.0	-1.147	0.188
40	10	30.0	0.140	0.178	90	10	40.0	-0.848	0.167
41	11	27.3	-1.179	0.182	91	10	40.0	-0.848	0.167
42	10	30.0	-1.147	0.188	92	13	23.1	-1.024	0.181
43	12	25.0	-1.164	0.185	93	14	21.4	-1.009	0.183
44	10	30.0	-1.147	0.188	94	15	20.0	-1.267	0.169
45	11	27.3	-1.179	0.182	95	12	25.0	-0.828	0.181
46	10	40.0	0.456	0.156	96	17	17.7	-1.290	0.165
47	10	30.0	-1.147	0.188	97	10	30.0	-0.623	0.186
48	10	40.0	-0.686	0.168	98	10	60.0	0.973	0.154
49	15	20.0	-1.267	0.169	99	12	25.0	-1.208	0.178
50	11	27.3	-0.824	0.183	100	11	27.3	-1.118	0.193

Profil d'élèves à la PHASE II
Banque d'items avec le leurre no.2
 Sélection attendue du leurre : 1fois sur 2

#	Nb.d'items présentés	Consistance à l'erreur	Estimation de difficulté cognitive (théta)	Variance Erreur	#	Nb.d'items présentés	Consistance à l'erreur	Estimation de difficulté cognitive (théta)	Variance Erreur
1	10	30.0	-1.147	0.188	51	11	27.3	-1.179	0.182
2	10	40.0	-0.848	0.167	52	10	40.0	-0.848	0.167
3	10	50.0	-0.570	0.161	53	10	70.0	1.249	0.174
4	10	60.0	-0.973	0.154	54	10	50.0	0.718	0.149
5	10	70.0	1.249	0.174	55	10	30.0	0.176	0.176
6	10	60.0	-0.295	0.166	56	10	40.0	-0.848	0.167
7	10	40.0	-0.848	0.167	57	10	30.0	-1.147	0.188
8	12	75.0	1.336	0.163	58	13	76.9	1.367	0.159
9	10	40.0	0.362	0.157	59	10	40.0	0.461	0.155
10	10	40.0	0.362	0.157	60	10	40.0	0.461	0.155
11	10	50.0	-0.570	0.161	61	10	50.0	-0.570	0.161
12	10	30.0	-1.147	0.188	62	10	50.0	0.718	0.149
13	10	70.0	1.249	0.174	63	10	50.0	-0.570	0.161
14	10	50.0	0.217	0.171	64	10	40.0	0.461	0.155
15	10	60.0	0.970	0.154	65	10	50.0	-0.304	0.155
16	12	25.0	-1.280	0.178	66	10	30.0	-1.147	0.188
17	10	40.0	0.421	0.156	67	10	50.0	-0.570	0.161
18	10	40.0	-0.031	0.157	68	11	27.3	-1.118	0.193
19	10	50.0	-0.570	0.161	69	13	23.1	-1.233	0.174
20	10	50.0	-0.039	0.156	70	10	50.0	0.718	0.149
21	10	50.0	0.718	0.149	71	10	40.0	0.293	0.155
22	10	30.0	-1.147	0.188	72	10	30.0	-1.147	0.188
23	10	30.0	-1.147	0.188	73	10	50.0	-0.570	0.161
24	10	70.0	-1.249	0.174	74	10	30.0	-1.147	0.188
25	10	50.0	0.718	0.149	75	10	70.0	1.249	0.174
26	10	60.0	-0.147	0.164	76	10	50.0	-0.570	0.161
27	10	40.0	-0.848	0.167	77	14	78.6	1.395	0.156
28	10	40.0	-0.848	0.167	78	10	50.0	-0.570	0.161

29	10	50.0	0.718	0.149	79	10	70.0	1.249	0.174
30	10	70.0	0.000	0.185	80	10	40.0	-0.848	0.167
31	10	60.0	0.973	0.154	81	10	30.0	-1.147	0.188
32	10	60.0	0.973	0.154	82	10	50.0	-0.570	0.161
33	10	60.0	0.973	0.154	83	10	70.0	0.442	0.185
34	10	60.0	0.973	0.154	84	10	40.0	-0.848	0.167
35	10	60.0	-0.295	0.166	85	13	23.1	-1.233	0.174
36	11	72.3	1.299	0.167	86	10	70.0	1.249	0.174
37	10	40.0	-0.848	0.167	87	11	63.6	0.827	0.167
38	12	75.0	1.336	0.167	88	10	40.0	0.042	0.155
39	10	40.0	-0.051	0.156	89	10	30.0	-1.147	0.188
40	10	70.0	0.000	0.185	90	10	40.0	-0.224	0.162
41	10	50.0	-0.570	0.161	91	10	40.0	-0.224	0.162
42	10	40.0	-0.848	0.167	92	10	40.0	-0.848	0.167
43	10	60.0	0.973	0.154	93	10	40.0	-0.327	0.167
44	11	72.3	1.299	0.167	94	10	50.0	-0.570	0.161
45	10	30.0	-1.147	0.188	95	10	70.0	1.249	0.174
46	10	40.0	-0.848	0.167	96	10	40.0	-0.848	0.167
47	10	50.0	-0.057	0.161	97	10	50.0	-0.570	0.161
48	10	50.0	0.718	0.149	98	10	60.0	-0.295	0.166
49	10	30.0	-1.147	0.188	99	10	50.0	-0.570	0.161
50	10	60.0	-0.038	0.159	100	10	50.0	0.718	0.149

Profil d'élèves à la PHASE II
Banque d'items avec le leurre no.2
Sélection attendue du leurre : 3 fois sur 4

#	Nb. d'items présentés	Consistance à l'Erreur	Estimation difficulté cognitive (thêta)	Variance Erreur	#	Nb.d'items présentés	Consistance à l'Erreur	Estimation difficulté cognitive (thêta)	Variance Erreur
1	12	75.0	1.336	0.163	51	13	76.9	1.014	0.179
2	10	70.0	1.249	0.174	52	20	95.0	2.226	0.377
3	11	72.8	1.299	0.167	53	17	82.3	1.449	0.151
4	12	75.0	1.336	0.163	54	17	82.3	1.449	0.151
5	10	70.0	1.249	0.174	55	10	70.0	1.249	0.174
6	11	72.7	1.023	0.185	56	12	75.0	1.336	0.163
7	10	70.0	0.831	0.191	57	13	76.9	1.367	0.159
8	11	72.7	1.299	0.167	58	10	70.0	1.249	0.174
9	12	75.0	1.336	0.163	59	20	85.0	1.472	0.148
10	10	70.0	1.249	0.174	60	18	83.3	1.459	0.150
11	10	70.0	1.249	0.174	61	16	81.2	1.437	0.152
12	10	70.0	0.039	0.189	62	19	64.2	1.467	0.149
13	10	70.0	1.249	0.174	63	16	81.2	1.399	0.155
14	10	70.0	1.249	0.174	64	16	81.2	1.437	0.152
15	12	75.0	1.336	0.163	65	20	90.0	1.766	0.205
16	10	70.0	1.249	0.174	66	10	70.0	1.249	0.174
17	10	70.0	0.000	0.185	67	20	90.0	1.766	0.205
18	13	76.9	1.367	0.159	68	20	90.0	1.763	0.205
19	10	70.0	1.249	0.174	69	10	60.0	0.063	0.157
20	10	70.0	1.249	0.174	70	10	70.0	0.000	0.185
21	12	75.0	1.336	0.163	71	10	60.0	-0.295	0.166
22	11	72.7	1.299	0.167	72	20	95.0	2.224	0.377
23	10	70.0	1.249	0.174	73	18	83.3	1.459	0.150
24	11	72.7	1.299	0.167	74	18	83.3	1.459	0.150
25	11	72.7	1.299	0.167	75	10	70.0	1.249	0.174
26	11	72.7	1.299	0.167	76	20	90.0	1.766	0.205
27	10	70.0	1.249	0.174	77	20	90.0	1.472	0.205
28	11	72.7	1.299	0.167	78	20	85.0	1.472	0.148

29	10	70.0	1.249	0.174	79	10	60.0	0.739	0.155
30	10	70.0	1.249	0.174	80	10	70.0	1.108	0.178
31	10	70.0	1.249	0.174	81	15	80.0	1.418	0.154
32	10	70.0	0.000	0.185	82	20	100.0	4.000	6.531
33	11	72.7	1.299	0.167	83	10	70.0	1.249	0.174
34	11	72.7	1.299	0.167	84	20	90.0	1.766	0.205
35	11	72.7	1.299	0.167	85	11	72.3	1.124	0.186
36	11	72.7	1.124	0.186	86	15	80.0	1.418	0.154
37	10	60.0	0.521	0.159	87	17	82.3	1.449	0.151
38	11	72.7	1.023	0.185	88	20	90.0	1.766	0.205
39	10	60.0	0.521	0.159	89	10	70.0	1.249	0.174
40	11	72.7	1.023	0.185	90	10	70.0	1.231	0.176
41	10	40.0	0.418	0.156	91	13	76.9	1.367	0.159
42	10	60.0	0.585	0.176	92	15	80.0	1.371	0.158
43	10	70.0	1.249	0.174	93	10	70.0	0.246	0.178
44	10	60.0	0.973	0.154	94	15	80.0	1.418	0.154
45	12	75.0	1.336	0.163	95	10	70.0	0.246	0.178
46	11	72.7	1.299	0.167	96	14	78.6	1.395	0.156
47	10	70.0	1.249	0.174	97	20	85.0	1.472	0.148
48	11	72.7	1.023	0.185	98	10	60.0	-0.295	0.166
49	12	75.0	1.336	0.163	99	20	90.0	1.766	0.205
50	11	72.7	1.299	0.167	100	11	72.3	1.299	0.167

Profil d'élèves à la PHASE II
Banque d'items avec le leurre no.8
Sélection aléatoire du leurre

#	Nb.d'items présentés	Consistance à l'erreur	Estimation difficulté cognitive (thêta)	Variance Erreur	#	Nb.d'items présentés	Consistance à l'erreur	Estimation difficulté cognitive (thêta)	Variance Erreur
1	10	30.0	-1.484	0.173	51	16	18.8	-1.875	0.144
2	16	18.8	-1.875	0.144	52	13	23.1	-1.468	0.172
3	16	12.5	-2.161	0.200	53	10	30.0	-1.560	0.170
4	10	30.0	-1.352	0.180	54	15	20.0	-1.839	0.147
5	10	30.0	-1.271	0.177	55	18	11.1	-2.214	0.197
6	16	12.5	-2.161	0.200	56	16	18.8	-1.875	0.144
7	10	30.0	-1.467	0.173	57	14	21.4	-1.799	0.149
8	10	30.0	-1.560	0.170	58	16	12.5	-2.161	0.200
9	10	30.0	-1.560	0.170	59	16	12.5	-2.161	0.200
10	10	30.0	-1.484	0.173	60	12	25.0	-1.700	0.157
11	16	18.8	-1.875	0.144	61	10	50.0	-0.650	0.148
12	16	12.5	-2.161	0.200	62	15	20.0	-1.829	0.147
13	10	30.0	-1.352	0.180	63	16	12.5	-2.161	0.200
14	10	30.0	-1.271	0.177	64	10	30.0	-1.528	0.172
15	16	12.5	-2.161	0.200	65	10	30.0	-1.336	0.175
16	16	12.5	-2.161	0.200	66	18	11.1	-2.214	0.197
17	10	30.0	-1.560	0.170	67	16	12.5	-2.161	0.200
18	10	30.0	-1.560	0.170	68	10	40.0	-1.090	0.152
19	19	10.5	-2.237	0.196	69	16	18.8	-1.866	0.145
20	10	30.0	-1.081	0.178	70	10	40.0	0.797	0.157
21	16	12.5	-2.161	0.200	71	10	30.0	-1.528	0.172
22	16	18.8	-1.875	0.144	72	10	30.0	-0.533	0.185
23	16	12.5	-2.161	0.200	73*	20	5.0	-2.701	0.366
24	13	23.1	-1.752	0.153	74	12	25.0	-1.675	0.158
25	11	27.3	-1.635	0.163	75	16	12.5	-2.161	0.200
26	10	40.0	-0.568	0.160	76	16	12.5	-2.161	0.200
27	19	10.6	-2.237	0.196	77	10	30.0	-1.560	0.170
28	10	30.0	-0.414	0.187	78	10	30.0	-1.560	0.170

29	17	11.8	-2.184	0.198	79	13	23.1	-1.752	0.153
30	10	30.0	-1.560	0.170	80	10	30.0	-1.560	0.170
31	15	20.0	-1.829	0.147	81	20	10.0	-2.258	0.194
32	16	12.5	-2.161	0.200	82	16	12.5	-2.161	0.200
33	10	50.0	0.475	0.154	83	10	50.0	0.613	0.153
34	10	30.0	-1.560	0.170	84	10	30.0	-1.528	0.172
35	20	5.0	-2.701	0.366	85	11	27.3	-1.607	0.164
36	10	30.0	-1.560	0.170	86	16	12.5	-2.161	0.200
37	16	12.5	-2.161	0.200	87	11	27.3	-1.625	0.163
38	15	20.0	-1.829	0.147	88	17	11.8	-2.184	0.198
39	10	30.0	-1.560	0.170	89	17	11.8	-2.189	0.198
40	10	40.0	-0.989	0.157	90	16	12.5	-2.161	0.200
41	15	20.0	-1.829	0.147	91	17	11.8	-2.179	0.199
42	17	11.8	-2.184	0.198	92	10	30.0	-1.329	0.179
43	13	23.1	-1.752	0.153	93	10	40.0	0.541	0.157
44	10	30.0	-1.560	0.170	94	12	25.0	-0.341	0.181
45	11	27.3	-1.653	0.163	95	16	12.5	-2.161	0.200
46	14	21.4	-1.799	0.149	96	16	12.5	-2.161	0.200
47	20	15.0	-1.891	0.143	97	15	20.0	-1.839	0.147
48	12	25.0	-1.700	0.157	98	10	50.0	1.054	0.148
49	17	11.8	-2.189	0.198	99	10	50.0	-0.606	0.150
50	10	30.0	-1.560	0.170	100	10	30.0	-0.839	0.194

Profil d'élèves à la PHASE II
Banque d'items avec le leurre no.8
Sélection attendue du leurre : 1 fois sur 2

#	Nb.d'items présentés	Consistance à l'erreur	Estimation de difficulté cognitive (thêta)	Variance Erreur	#	Nb.d'items présentés	Consistance à l'erreur	Estimation de difficulté cognitive (thêta)	Variance Erreur
1	10	60.0	-0.505	0.154	51	10	50.0	-1.038	0.147
2	10	50.0	1.042	0.148	52	10	50.0	-1.038	0.147
3	10	60.0	1.025	0.159	53	10	60.0	1.319	0.151
4	10	40.0	0.668	0.158	54	10	60.0	0.883	0.155
5	13	76.9	1.774	0.153	55	11	72.7	1.662	0.163
6	10	50.0	-1.000	0.149	56	10	30.0	-1.528	0.172
7	10	60.0	1.319	0.151	57	10	60.0	-0.686	0.157
8	10	40.0	0.797	0.157	58	10	40.0	0.797	0.157
9	10	30.0	-1.182	0.176	59	10	50.0	-0.837	0.148
10	10	70.0	1.590	0.171	60	10	50.0	0.328	0.152
11	10	60.0	1.310	0.152	61	10	50.0	1.045	0.148
12	10	60.0	1.016	0.155	62	10	30.0	-1.560	0.170
13	10	30.0	-0.947	0.184	63	10	30.0	-1.560	0.170
14	10	70.0	1.590	0.171	64	10	40.0	-1.169	0.153
15	10	40.0	-1.238	0.153	65	10	40.0	-1.255	0.153
16	10	40.0	0.696	0.159	66	10	60.0	-0.514	0.158
17	10	30.0	-0.358	0.181	67	10	40.0	-0.416	0.172
18	10	70.0	1.590	0.171	68	11	27.3	-1.469	0.171
19	10	40.0	-1.289	0.151	69	10	70.0	1.590	0.171
20	11	72.7	1.256	0.194	70	10	40.0	-1.289	0.151
21	10	50.0	-1.038	0.147	71	10	60.0	-0.584	0.156
22	10	50.0	-1.038	0.147	72	10	30.0	-1.484	0.173
23	10	40.0	-1.289	0.151	73	10	40.0	0.394	0.157
24	10	50.0	0.725	0.150	74	10	60.0	0.917	0.156
25	10	40.0	-1.256	0.152	75	11	27.3	-1.635	0.163
26	10	50.0	-1.020	0.147	76	11	72.7	1.662	0.163
27	10	40.0	-1.289	0.151	77	10	60.0	-0.554	0.154
28	10	70.0	1.590	0.171	78	13	76.9	1.774	0.153

29	10	70.0	0.476	0.176	79	10	40.0	0.434	0.156
30	10	50.0	-1.038	0.147	80	10	60.0	-0.249	0.154
31	10	70.0	1.590	0.171	81	10	60.0	-0.173	0.156
32	10	60.0	-0.249	0.154	82	10	40.0	-1.289	0.151
33	10	50.0	1.067	0.148	83	10	50.0	-0.945	0.149
34	10	30.0	-1.560	0.170	84	10	60.0	-0.764	0.156
35	10	30.0	-1.560	0.170	85	12	25.0	-0.925	0.178
36	11	72.7	1.256	0.194	86	10	40.0	-0.245	0.163
37	10	40.0	-1.289	0.151	87	11	27.3	-1.635	0.163
38	10	50.0	-0.539	0.151	88	10	40.0	0.553	0.160
39	10	60.0	-0.697	0.157	89	10	40.0	-0.511	0.160
40	10	40.0	-0.789	0.155	90	10	60.0	1.289	0.153
41	10	50.0	-0.048	0.156	91	10	50.0	0.331	0.151
42	10	50.0	-0.739	0.152	92	10	40.0	0.207	0.165
43	11	72.7	1.662	0.163	93	10	40.0	-1.256	0.152
44	10	70.0	1.380	0.180	94	10	60.0	1.279	0.153
45	10	50.0	-1.028	0.147	95	10	30.0	-1.543	0.171
46	10	70.0	1.590	0.171	96	10	60.0	1.319	0.151
47	10	50.0	-1.038	0.147	97	10	50.0	-1.038	0.147
48	10	30.0	0.120	0.178	98	10	50.0	1.058	0.148
49	10	60.0	1.289	0.153	99	11	27.3	-1.635	0.163
50	13	76.9	1.433	0.172	100	10	50.0	-1.038	0.147

Profil d'élèves à la PHASE II
Banque d'items avec le leurre no.8
Sélection attendue du leurre : 3 fois sur 4

#	Nb.d'items présentés	Consistance à l'erreur	Estimation de difficulté cognitive (thêta)	Variance Erreur	#	Nb.d'items présentés	Consistance à l'erreur	Estimation de difficulté cognitive (thêta)	Variance Erreur
1	16	81.3	1.842	0.147	51	10	60.0	1.170	0.154
2	17	88.2	2.209	0.199	52	10	70.0	1.587	0.171
3	15	80.0	1.847	0.147	53	10	70.0	1.590	0.171
4	19	89.5	2.257	0.196	54	17	88.2	2.205	0.199
5	18	88.9	2.235	0.197	55	16	81.3	1.557	0.161
6	12	75.0	1.724	0.158	56	11	72.7	1.640	0.164
7	15	80.0	1.859	0.147	57	17	82.4	1.921	0.143
8	18	83.3	1.837	0.147	58	10	60.0	0.670	0.187
9	17	88.2	2.209	0.199	59	10	70.0	1.590	0.171
10	14	78.6	1.820	0.150	60	17	88.2	2.209	0.199
11	17	88.2	2.209	0.199	61	10	60.0	-0.207	0.154
12	10	70.0	-0.328	0.181	62	10	60.0	-0.322	0.166
13	12	75.0	1.706	0.159	63	17	82.4	1.925	0.143
14	10	70.0	-0.228	0.180	64	15	80.0	1.859	0.147
15	10	70.0	1.590	0.171	65	10	70.0	1.590	0.171
16	10	70.0	0.036	0.177	66	15	80.0	1.859	0.147
17	13	76.9	1.774	0.153	67	10	70.0	1.590	0.171
18	10	70.0	0.407	0.177	68	14	78.6	0.798	0.180
19	18	88.9	2.235	0.197	69	15	80.0	1.859	0.147
20	10	70.0	0.330	0.176	70	10	70.0	0.407	0.177
21	14	78.6	1.820	0.150	71	11	72.7	1.662	0.163
22	13	76.9	1.774	0.153	72	11	72.7	1.640	0.164
23	11	72.7	1.662	0.163	73	10	60.0	-0.453	0.154
24	10	70.0	0.686	0.182	74	14	78.6	1.810	0.150
25	10	70.0	1.590	0.171	75	10	70.0	0.407	0.177
26	13	76.9	1.774	0.153	76	10	70.0	1.142	0.179
27	17	88.2	2.209	0.199	77	14	78.6	1.810	0.150
28	16	81.3	1.451	0.167	78	10	70.0	1.590	0.171

29	10	40.0	-1.289	0.151	79	17	88.2	2.209	0.199
30	10	60.0	1.319	0.151	80	15	80.0	1.859	0.147
31	11	63.6	0.849	0.170	81	15	80.0	1.859	0.147
32	10	70.0	1.590	0.171	82*	20	100.0	4.000	2.902
33	10	50.0	0.626	0.149	83	10	60.0	0.048	0.156
34	17	88.2	2.209	0.199	84	13	76.9	1.774	0.153
35	10	60.0	1.319	0.151	85	12	75.0	0.418	0.182
36	19	89.5	2.257	0.196	86	17	88.2	2.209	0.199
37	15	80.0	1.672	0.159	87	10	70.0	0.028	0.191
38	19	89.5	2.257	0.196	88	15	80.0	0.979	0.178
39	17	82.4	1.925	0.143	89	12	75.0	0.814	0.179
40	12	75.0	0.603	0.182	90	15	80.0	1.859	0.147
41	14	78.6	1.820	0.150	91	11	72.7	1.662	0.163
42	12	75.0	1.706	0.159	92	19	89.5	2.257	0.196
43	12	75.0	1.706	0.159	93	17	88.2	2.209	0.199
44	15	80.0	1.555	0.162	94	10	70.0	1.590	0.171
45	16	81.3	1.704	0.155	95	11	72.7	1.256	0.194
46	18	88.9	2.235	0.197	96	11	72.7	1.256	0.194
47	10	70.0	-0.328	0.181	97	12	75.0	0.814	0.179
48	11	72.7	1.245	0.193	98	12	75.0	1.724	0.158
49	17	82.4	1.925	0.143	99	13	76.9	1.774	0.153
50	10	70.0	1.562	0.172	100	13	76.9	1.774	0.153

Exemples de fiches diagnostiques à la PHASE II.

Résultats au test spécifique au leurre no. 2

Nombre d'items administrés à la Phase II : 20

Estimation de la difficulté cognitive -4
 Erreur final de la variance estimé 7.545

L'élève a choisi 0 fois le leurre no.2 sur 20
 items présentés, soit 0%

Résultats au test spécifique au leurre no. 2

Nombre d'items administrés à la Phase II : 10

Estimation de la difficulté cognitive -1.147
 Erreur final de la variance estimé 0.188

L'élève a choisi 3 fois le leurre no.2 sur 10
 items présentés, soit 30%

Résultats au test spécifique au leurre no. 2

Nombre d'items administrés à la Phase II : 11

Estimation de la difficulté cognitive -0.811
 Erreur final de la variance estimé 0.172

L'élève a choisi 4 fois le leurre no.2 sur 11
 items présentés, soit 36.364%

Résultats au test spécifique au leurre no. 2

Nombre d'items administrés à la Phase II : 20

Estimation de la difficulté cognitive 4
 Erreur final de la variance estimé 6.531

L'élève a choisi 20 fois le leurre no.2 sur 20
 items présentés, soit 100%