



National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services Branch

Direction des acquisitions et
des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file - Votre référence

Our file - Notre référence

NOTICE

AVIS

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

If pages are missing, contact the university which granted the degree.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

COMPARAISON D'UN TEST CONVENTIONNEL ET
D'UNE SIMULATION D'UN TEST ADAPTATIF SUR
ORDINATEUR CHEZ LES ELEVES D'HABILETE
FAIBLE EN MATHEMATIQUE

par Linda Mainville-Lanthier

Thèse présentée à l'Ecole des études supérieures
et de la recherche de l'Université d'Ottawa en vue de
l'obtention du M.A. en Education

Ottawa, Canada, 1993

c Linda Mainville-Lanthier, Ottawa, Canada, 1993



National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Votre référence*

Our file *Notre référence*

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

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

ISBN 0-315-89636-1

Canada



UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

RECONNAISSANCE

Cette thèse a été préparée sous la direction du Docteur Marc Richard de la Faculté d'Education de l'Université d'Ottawa. Cet homme, sur le seuil de la retraite, a su, avec patience, dévouement et amour de son prochain, partager une sagesse acquise au fil des ans. C'est avec beaucoup de reconnaissance que je désire lui exprimer mes remerciements.

Je désire également remercier les professeurs Renée Forgette Giroux et Dany Leveault qui ont pris le temps dans leur horaire très chargé, de m'écouter, de me conseiller et même de m'apporter du matériel dans le but de faciliter ma tâche.

Je tiens aussi à remercier infiniment mes parents, Fernand et Nicole qui ont su, pendant les 20 premières années de mon existence, implanter chez moi, une joie de vivre et un bien-être profond de l'âme avec lesquels je suis maintenant armée contre les intempéries de la vie.

En terminant, je veux exprimer ma reconnaissance à la personne avec qui j'ai choisi de partager ma vie. Sans mon mari, je n'aurais pu faire cette thèse avec sérénité. En s'occupant merveilleusement de nos deux jeunes enfants, Nadia et Érik, il m'a enlevé la culpabilité qu'éprouvent plusieurs mères lorsqu'elles ont une carrière. Comme les mots ne suffisent pas, je lui dis tout simplement: "Merci Luc".

CURRICULUM STUDIORUM

Linda Mainville Lanthier est née à Casselman, Province d'Ontario, le 3 juin 1964. Elle obtient son Baccalauréat es Arts de l'Université de Toronto en 1985 et son Baccalauréat en Education de l'Université d'Ottawa en 1986.

TABLE DES MATIERES

Chapitres

Introduction	vi
I - LE CADRE CONCEPTUEL	1
1. Testing conventionnel	1
A. Testing papier-crayon (TC)	1
B. Testing sur ordinateur (TCO)	3
2. Testing adaptatif	5
A. Testing adaptatif papier-crayon (TA)	5
B. Testing adaptatif sur ordinateur (TAO)	7
3. Synthèse: Problème et hypothèses de recherche	16
II - LA METHODOLOGIE	23
1. Description de l'échantillon	23
2. Instruments	28
A. Test sur les verres de jus de Noelting (1979) ..	28
B. MicroCAT	35
C. Procédure de l'administration des tests	36
D. Plan de l'analyse statistique	37
III - LA PRESENTATION ET L'ANALYSE DES RESULTATS	39
1. La présentation des résultats	39
2. L'interprétation des résultats	46

RESUME ET CONCLUSION	59
----------------------------	----

BIBLIOGRAPHIE	62
---------------------	----

Appendices

A. Résultats bruts des 6 tests	68
--------------------------------------	----

B. Test du T.A.R.A. (sections n2 & n3)	70
--	----

C. Test sur les verres de jus de Noelting (version originale)	72
--	----

LISTE DES TABLEAUX

Tableaux

1 - Répartition du nombre d'élèves par école à travers les six différents tests (90 élèves)	25
2 - Sommaire de l'analyse de variance du T.A.R.A.	27
3 - Items de l'expérience des verres de jus de Noelting (1979)	31
4 - Moyenne et écart-types des six différentes versions du <u>Test sur les verres de jus</u>	41
5 - Sommaire des comparaisons planifiées de chaque hypothèse selon le test F	47

LISTE DES FIGURES

1 - Six différentes versions du Test sur les verres de jus de Noelting (1979) élaborées en fonction de la variable de l'ordre de difficulté des items et la variable de la mobilité de l'élève dans un test	34
2 - Comparaison des moyennes du test 2 au test 6	42
3 - Comparaison des moyennes du test 1 au test 4	42
4 - Comparaison des moyennes du test 2 au test 5	43
5 - Comparaison des moyennes du test 3 au test 6	43
6 - Comparaison des moyennes du test 4 au test 5	44
7 - Comparaison des moyennes du test 4 au test 6	44
8 - Comparaison des moyennes du test 5 au test 6	45

INTRODUCTION

En salle de classe, l'enseignant cherche à découvrir la compétence réelle de l'élève. Aussi, il est en mesure d'exiger un juste rendement. Pour ce faire, il administre des tests. Toutefois, il ne dépend pas totalement de ces derniers pour porter un jugement. Il est en effet tout à fait conscient des facteurs qui peuvent influencer la performance d'un élève.

Il est difficile d'évaluer précisément l'habileté cognitive du l'élève sans se préoccuper de sa personnalité et de la situation du testing. Un test qui provoque beaucoup d'anxiété chez certains élèves peut souvent nuire à leur performance. En outre, le type de test peut aussi influencer le rendement de l'élève. Celui qui a de la difficulté à s'exprimer par écrit sera favorisé par un test à choix multiples et pénalisé par un test composé de questions ouvertes. On doit aussi prendre en considération l'endroit et l'heure: l'administration d'un test le vendredi après-midi dans un endroit différent de la classe habituelle, peut nuire à la concentration de l'élève. De plus, il ne faut pas oublier le côté physique et affectif de l'élève au moment de l'administration du test. A cet égard, l'élève malade ou tracassé par une dispute avec un ami, ne réussira pas aussi bien qu'en temps normal. L'attitude de l'élève envers une matière quelconque peut aussi influencer la performance de celui-ci. Le défi cependant est d'estimer l'impact de ces facteurs sur la performance afin d'identifier le potentiel

cognitif réel de l'étudiant.

La présente étude docimologique a pour but de comparer le testing conventionnel à une simulation du testing adaptatif (TAO) sur ordinateur selon la théorie des réponses aux items (TRI) en fonction des élèves d'habileté faible. Le testing conventionnel se définit par l'administration d'un test de type papier-crayon et dont l'ordre de difficulté des items est aléatoire. Quant à la simulation du TAO selon la TRI envers des élèves d'habileté faible, elle s'effectue ainsi: on administre le test par le biais d'un ordinateur qui ne permet pas à l'élève de réviser ses réponses. De plus, ce test comporte un ordre décroissant de la difficulté de ses items c'est-à-dire du plus difficile au plus facile. Or, le comportement des élèves face à l'ordre de difficulté des items dans un test a fait l'objet de nombreuses études depuis environ 40 ans. On en est arrivé à un certain consensus. L'ordre de difficulté des items se doit d'être du plus facile au plus difficile ou de façon aléatoire. L'ordre inverse, c'est-à-dire du plus difficile au plus facile, augmente la difficulté du test. On a fait cette recommandation pour le testing papier-crayon. En est-il de même pour le testing adaptatif sur ordinateur selon la théorie des réponses aux items envers les élèves d'habileté faible?

Le premier chapitre se divise en trois parties. Tout d'abord, la nature du testing conventionnel avec et sans l'ordinateur est exposée. Par la suite, les recherches empiriques relatives à l'ordre de difficulté des items dans un testing conventionnel sont présentées. La dernière division comporte une synthèse incluant le problème et les hypothèses de recherche. Le deuxième chapitre présente une description de l'échantillon, des instruments et du plan de l'analyse statistique. La description des résultats et l'interprétation de ceux-ci s'effectuent au troisième chapitre. Enfin, le quatrième chapitre expose le résumé et la conclusion de cette thèse.

CHAPITRE 1

LE CADRE CONCEPTUEL

Le fondement théorique de cette recherche comprend les trois aspects suivants: le testing conventionnel, le testing adaptatif et le testing adaptatif sur ordinateur selon la théorie des réponses aux items.

1. Testing conventionnel (TC)

A. Testing papier-crayon (TC)

Le testing conventionnel (TC) consiste à administrer un test au moyen d'un crayon et du papier. Dans un test conventionnel de type papier-crayon, l'élève peut sauter des items, revenir en arrière et même changer des réponses. Il est donc en mesure d'adapter sa propre stratégie de sélection d'items.

Beaucoup de recherches ont porté sur l'ordre de difficulté des items dans un test de type papier-crayon. Hambleton (1968), MacNicol (1956), Mollendopf (1950), Monk & Stallings (1970) et Sax & Carr (1962) ont démontré que l'ordre de difficulté des items allant du plus difficile au plus facile augmente la difficulté du test. Lorsque les items difficiles sont présentés dès le début du test, les élèves d'habileté moyenne et faible en particulier

deviennent plus stressés et ont tendance à se décourager. C'est pourquoi, Ahmann & Glock (1971), Ross & Stanley (1954), et Thorndike & Hagen (1961) recommandent de concevoir des tests commençant par des items faciles. Toutefois, bien que plusieurs recherches empiriques (Buger, Munz, Smouse & Angelino, 1969; Brenner, 1964; Kestenbaum & Weiner, 1970; Munz & Smouse, 1968; Sweeny, Smouse, Rupper & Munz, 1970) n'aient pas encore réussi à démontrer l'influence de l'ordre de difficulté des items d'un test sur le score obtenu par les élèves, elles ont par ailleurs mis en lumière une interaction entre l'ordre de difficulté des items et l'anxiété. D'autres chercheurs soulignent que la limite de temps imposée intensifie l'effet de l'augmentation de la difficulté d'un test (Nelson, Towle & Merrill, 1955). A cet égard, Sax et Cromack (1966) soutiennent que l'ordre de difficulté des items influence la difficulté du test seulement lorsqu'on impose une limite de temps pour répondre aux items. Brenner (1964) a en effet démontré qu'il existe une relation significative entre la performance des élèves et l'ordre de difficulté des items en présence d'une limite de temps. Lors du test dont les items sont placés de façon aléatoire, il obtient 45,7 de moyenne. En présence d'items dont l'ordre de difficulté est croissant, il note une moyenne de 42,6 et lors d'un arrangement décroissant de la difficulté d'item, il obtient 38,2. Depuis environ 40 ans, des chercheurs débattent cette question. Ils ne sont pas parvenus à un accord sur les détails concernant la

limite de temps imposée mais semblent être arrivés à un consensus sur l'ordre de difficulté des items d'un test: celui-ci ne doit pas être du plus difficile au plus facile. L'ordinateur occupe maintenant une place importante en testing comme dans bien d'autres domaines. Mais quelle est au juste l'influence de cet outil dans le domaine du testing?

B. Testing sur ordinateur (TCO)

Il est possible de conceptualiser, d'administrer et de corriger un test conventionnel au moyen d'un ordinateur. L'avantage à utiliser l'ordinateur se situe surtout au niveau de l'efficacité du temps. En effet, le TCO requiert beaucoup moins de temps que le TC et commet moins d'erreur.

Le testing sur ordinateur (TCO) par contre, ne comporte pas toutes les caractéristiques retrouvées normalement dans un test papier-crayon de type conventionnel. Sans un progiciel puissant, les items sont présentés dans un ordre prédéterminé. Dans ces conditions, l'élève ne peut pas sauter ou réviser certains items. Il ne peut pas non plus changer ses réponses. L'élève ne possède alors aucune mobilité au niveau de l'ordre dans lequel il peut répondre aux items dans le test. Le TCO pourrait comporter

les mêmes caractéristiques retrouvées dans un testing conventionnel mais ces progiciels sophistiqués se font rares présentement.

On n'a pas encore fait de recherches empiriques sur l'ordre de difficulté des items dans un TCO. On a cependant tenté de comparer les deux modes de testing. Bunderson & al. (1989) affirment que le TC est similaire au TCO. Par ailleurs, Maseo et Harvey (1988) conclurent après l'étude d'une douzaine de recherches, qu'on ne peut assumer une équivalence entre les deux types de testing. Dans la plupart des cas, la moyenne des élèves qui ont répondu à un TC est plus élevée que celle des élèves qui ont effectué un TCO. Cette moyenne n'est généralement pas significative. De plus, on n'a pas encore observé l'impact de l'absence de mobilité à l'intérieur d'un test sur la performance de l'élève. En effet, lors d'un TCO, l'élève doit répondre aux questions selon l'ordre qu'on les lui présente. Il ne peut alors ni réviser ni changer ses réponses. "Whether item review has an effect on test scores is, however an issue that awaits empirical study and is of substantial importance to CB (computer based) test developers." (Wise & Plake, 1989). On continue cependant à explorer d'autres champs dans le domaine du testing sur ordinateur comme celui du testing adaptatif.

2. Testing adaptatif

A. Testing adaptatif papier-crayon (TA)

Le but du testing adaptatif (TA) est de présenter des items de difficulté appropriée à chaque élève (Weiss et Betz, 1973; Weiss et Vale, 1987). Mais pour bien comprendre la nature et la portée du TA, il importe de retracer brièvement ses origines, et de dégager les principaux postulats de base de ce type de testing.

Le Stanford-Binet est à l'origine du testing adaptatif. On utilisait alors, l'âge chronologique comme échelle d'habileté. Les scores des sujets et la difficulté des items du test étaient placés sur une même échelle afin d'indiquer l'âge mental de l'élève. Cette méthode comporte cependant l'inconvénient de corriger la première partie du test pour décider de la prochaine section à administrer. Ce procédé exige ainsi un testing individualisé. Entre 1950 et 1970, on a tenté d'adapter un test en fonction de l'examiné sur une plus grande échelle en mettant au point le test à niveau double (Angoff & Huddleston, 1958), le test pyramidal (Kratwohl & Huyser, 1956) et le test flexi-niveau (Lord, 1971). Le testing adaptatif a émergé de la combinaison du Stanford Binet et de la méthodologie de la théorie des réponses aux items (TRI),

développée vers la fin des années 60 et au début des années 70 (Reckase, 1989).

Weiss (1985) indique que les postulats de base du testing adaptatif sont les suivants: 1) la connaissance a priori de la difficulté et de la discrimination de tous les items parce que le TA cherche à administrer les items qui discriminent bien au niveau estimé de l'habileté de l'élève; 2) la sélection du prochain item selon la réponse de l'élève à l'item précédent; ainsi une bonne réponse engendre un item plus difficile et une mauvaise réponse est suivie de l'administration d'un item plus facile; 3) l'administration d'une combinaison différente d'items pour chaque examiné selon leur niveau d'habileté. Par conséquent, l'élève reçoit les items de son test dans un ordre déterminé par la stratégie de sélection utilisée par la théorie classique des scores (TCS), par la théorie des réponses aux items (TRI), ou encore par une théorie fondée sur autre chose (ex.: une approche cognitive). Habituellement, on utilise le TA à l'aide d'un ordinateur car autrement, il faudrait un examinateur pour chaque élève.

On n'a pas encore effectué de recherches empiriques sur l'ordre de difficulté des items dans un test adaptatif sans ordinateur. On ne connaît pas non plus le rôle de l'absence de mobilité de l'élève dans le choix de l'ordre des items à répondre

dans le test. Comme le TA se fait généralement à l'aide d'un ordinateur, il est préférable d'effectuer des recherches sur ces facteurs dans un testing adaptatif sur ordinateur.

B. Testing adaptatif sur ordinateur (TAO)

L'expression "testing adaptatif sur ordinateur" (TAO) fait référence à plusieurs appellations anglaises telles que "tailored, response-contingent", "programmes individualized", "branched and sequential" (Weiss, 1985). En réalité, le TAO est une forme sophistiquée de testing où l'ordinateur choisit les items à administrer à l'élève en se basant sur la réponse de celui-ci, à l'item précédent (Wise et Plake 1989). Habituellement, le TAO selon la théorie des réponses aux items (TRI) utilise un nombre variable d'items pour chaque élève. Le degré de précision qu'on désire atteindre lorsqu'on mesure le niveau d'habileté de l'élève détermine le nombre d'items à administrer.

La procédure du testing adaptatif comprend quatre éléments méthodologiques: une banque d'items déjà paramétrisés, un processus de sélection d'items durant le testing, un processus d'administration des items et un processus de comptabilisation des résultats (Auger, 1989 p. 54). On doit aussi identifier un moyen qui détermine la fin du test (Reckase 1989).

La banque d'items contient des items déjà paramétrisés par un, deux ou trois paramètres. Le modèle Rasch tient compte seulement du premier paramètre qui est celui de l'indice de difficulté de l'item. La discrimination de l'item constitue le deuxième paramètre et l'indice de probabilité au troisième. Plus il y a de paramètres, plus le nombre de sujets pour valider les items est élevé. Un modèle à un paramètre requiert 250 sujets alors qu'un modèle à trois paramètres nécessite 1000 sujets.

Le processus de sélection d'items durant le testing est conçu de manière à choisir parmi les items qui contiennent le plus haut degré d'information par rapport au niveau d'habileté estimé d'un individu en particulier (Brown & Weiss, 1977; Mautelli & Weiss, 1981; Weiss & Kingsbury, 1984). On administre en premier lieu un item de difficulté moyenne avec un haut degré de discrimination. Advenant la réussite de cet item, la stratégie de sélection offre un item plus difficile et elle offre un item plus facile lors d'un échec. Ce processus continue ainsi tout au long du test.

Le processus d'administration des items se fait par l'entremise de différents progiciels. Présentement, on reconnaît quatre progiciels de bonne qualité: "Academic Instructional Measurement System" (AIMS), "MicroCAT", "Northwest Evaluation Association Test

Director" (Roid, 1989, p. 19) ainsi que le Système Informatisé de Traitement d'Examens (SITE).

Le processus de comptabilisation des résultats peut s'effectuer avec la théorie classique des scores mais on utilise habituellement la théorie des réponses aux items (TRI) pour le testing adaptatif à l'ordinateur (TAO).

La fin d'un TAO dont le nombre d'items varie, peut être déterminé de deux façons: par le manque d'items discriminatifs au niveau de difficulté de l'habileté estimée ou par une erreur de mesure acceptable. En effet, la stratégie de testing fondée sur une fonction de maximum d'information comme la TRI possède la capacité d'estimer l'erreur de mesure associée à chaque niveau de difficulté de l'habileté mesurée.

Plusieurs auteurs (Bejar, Weiss et Gialluca 1977; Kingsbury et Weiss, 1980b; Martin, McBride et Weiss 1983; McBride et Martin, 1983; McKinley et Reckase, 1981a; Reckase, 1981a; Thomson et Weiss, 1980; Vale et Weiss 1975;) ont souligné les avantages du testing adaptatif sur ordinateur. Parmi ces derniers, il importe de mentionner en premier lieu sa capacité d'obtenir une plus grande fidélité et validité que le testing conventionnel, tout en diminuant le nombre d'items. McBride et Martin (1983) ont démontré

qu'un test adaptatif de 15 items avait un coefficient de fidélité équivalent à une forme parallèle d'un test conventionnel avec 25 items pour leur test d'aptitude "Armed Service Vocational Aptitude Battery (ASVAB)". On assiste ici à une diminution de 40% des items. Des résultats similaires ont été obtenus par McKinley & Reckase (1980), Olsen et al. (1986), et Weiss (1992). En second lieu, le TAO discrimine mieux aux extrémités de la distribution normale que le testing conventionnel. Selon la stratégie fondée sur une fonction de maximum d'information (TRI), il est possible de mesurer les élèves indépendamment de leur niveau d'habileté pour un trait en particulier, à un même degré d'erreur de mesure choisi. Kingsbury et Weiss (1981) démontrent que le testing adaptatif est supérieur au testing conventionnel par rapport aux élèves qui se situent aux extrémités de l'habileté mesurée. Ceci est dû à la possibilité de mesurer tous les élèves avec la même erreur de mesure. Ils confirment l'équivalence des deux stratégies pour les élèves d'habileté moyenne. Par conséquent, le TAO est plus précis que le testing conventionnel par rapport aux élèves d'habileté supérieure et inférieure. Enfin, le fait de présenter des items discriminant bien au niveau d'habileté du sujet a pour conséquence de minimiser en quelque sorte l'effet de hasard. L'élève devine moins lorsqu'il est en présence d'items adaptés à son niveau d'habileté (Auger, 1989). De plus, chaque élève perçoit la difficulté du test comme étant appropriée, indépendamment de son

degré élevé ou faible d'habileté par rapport au trait mesuré (Weiss, 1985).

Ainsi par rapport au TC, le TAO avec un nombre variable d'items, diminue le nombre d'items tout en obtenant une plus grande validité et fidélité. De plus, le TAO discrimine mieux aux extrémités de la distribution normale, diminue l'erreur de mesure et le temps de testing, et enfin minimise l'effet de hasard.

Cependant, dans un TAO, le degré de mobilité de l'élève dans le test est nul. Cet aspect du TAO soumet tous les élèves à l'ordre de difficulté des items retrouvés dans le test; ce qui n'est pas le cas avec le testing conventionnel de type papier-crayon. En effet, le test style papier-crayon permet à l'élève de répondre aux items selon sa propre stratégie de sélection. Il peut ainsi sauter des questions, revenir en arrière et changer ses réponses. Il en résulte que plusieurs élèves ne sont pas soumis à l'ordre de difficulté des items présentés dans le test. Par ailleurs, le test adaptatif sur ordinateur sélectionne le prochain item en se basant sur la réponse de l'élève à l'item précédent. Ainsi, l'élève doit répondre à la question présentée par l'ordinateur, dans l'ordre déterminé par l'ordinateur. Le TAO soumet donc tous les élèves, sans exception, à sa stratégie de sélection qui provient de la théorie des réponses aux items (TRI).

La théorie des réponses aux items (TRI) est fondée sur le niveau d'habileté de l'élève. Ainsi, dépendamment du niveau d'habileté de l'élève, une différente stratégie de sélection est utilisée. Globalement, cette stratégie consiste à présenter un item de difficulté moyenne au début du test. Cet item de difficulté moyenne joue le rôle d'item d'ancrage parce que tout en diminuant l'erreur de mesure et le nombre d'items, il permet de bien discriminer dès le début. Advenant la réussite de ce premier item, l'ordinateur sélectionne un item plus difficile. Toutefois, lors d'un échec, un item plus facile est présenté. Ce procédé de sélection continue ainsi jusqu'au moment où une erreur de mesure acceptable est atteinte.

Par exemple, un élève d'habileté élevée reçoit des items allant d'un niveau de difficulté du plus facile au plus difficile. Tout d'abord, l'item d'ancrage de difficulté moyenne est en réalité un item facile pour cet élève. L'examiné répond ensuite à des items de plus en plus difficiles jusqu'à ce qu'il en échoue un. Alors, l'ordinateur sélectionne un item un peu plus facile que l'item échoué. Il est fort probable qu'à cette étape, on soit près du niveau d'habileté réel de l'élève. Après plusieurs échecs, le test est terminé car l'élève a atteint son plafond.

L'élève d'habileté faible est dans la situation opposée. Il

reçoit des items d'un niveau de difficulté allant du plus difficile au plus facile. L'item d'ancrage de difficulté moyenne est maintenant perçu comme un item difficile et cet élève connaît un échec dès le début. De plus, même si la difficulté des items suivants diminue peu à peu, la difficulté de ces items peut être encore trop élevée. Ce n'est qu'au moment où l'ordinateur atteint son véritable niveau d'habileté que cet élève commence à répondre à des items de difficulté moyenne pour lui. Quant à l'élève d'habileté moyenne, il reçoit dès le début, des items adaptés à son niveau d'habileté.

b.3 Recension des écrits au sujet de l'ordre de difficulté et de la mobilité de l'élève dans un TAO selon la TRI

La stratégie de sélection de la TRI semble pénaliser les élèves d'habileté faible en leur administrant un test dont le niveau de difficulté va du plus difficile au plus facile. Plusieurs auteurs ont déjà souligné (Hambleton, 1968; MacNicol, 1956; Mollendopf, 1950; Monk & Stallings, 1970; Sax & Carr, 1962) que cet ordre de difficulté dans un test papier-crayon a pour effet de frustrer l'élève, d'augmenter la difficulté du test et de nuire à sa performance. Retrouve-t-on ce même effet avec un test adaptatif sur ordinateur? Le test adaptatif sur ordinateur permet-il de corriger ces inconvénients?

La question suivante est au centre de la présente recherche: Est-ce que l'ordre de difficulté des items dans le testing adaptatif sur ordinateur selon la théorie des réponses aux items pénalise les élèves d'habileté faible? En effet, seuls les élèves d'habileté faible reçoivent un test dont le niveau de difficulté des items est décroissant. Plusieurs chercheurs mettent en doute, non pas le calcul de l'estimation de l'habileté, mais l'effet de l'ordre de difficulté des items sur la performance des élèves d'habileté faible. Wainer et Kiely (1987) croient que la frustration créée par la stratégie de sélection de la TRI chez les élèves d'habileté faible peut nuire à la performance de celui-ci. "This approach (TRI) to item difficulty ordering raises concerns about possible negative effects for low ability examinee, such as frustration and anxiety, which can hamper performance." (Wainer & Kiely, 1987, p. 189). Quant à Wise et Plake (1989), ils recommandent tout simplement de faire des recherches à ce sujet. "Whether the typical selection strategy has a negative effect on the test scores of low-ability examinee should be investigated." (Wise & Plake, 1989, p. 8)

Une sous-question se pose maintenant au sujet du rôle de l'absence de mobilité de l'élève sur l'effet de l'ordre de difficulté des items. Wise et Plake (1989) mentionnent qu'on ne connaît pas encore le rôle de ce facteur particulier au TAO. Par

exemple, dans un test papier-crayon, les élèves ont la possibilité d'éviter l'ordre de difficulté des items. Ils peuvent sauter des questions, revenir en arrière ou corriger certaines réponses à leur guise. Ainsi, l'élève est en mesure d'adopter sa propre stratégie de sélection. Dans un test adaptatif, l'ordinateur sélectionne le prochain item en se basant sur la réponse de l'élève à l'item précédent. Il en résulte qu'on ne peut permettre aux élèves de sauter des questions, de réviser ou de changer certaines réponses. L'absence de mobilité de l'élève dans le test soumet ainsi tous les élèves à la stratégie de sélection de la théorie utilisée. En ce qui a trait à la présente étude, tous les élèves d'habileté faible subissent un test comportant des items d'ordre de difficulté décroissant. La soumission de tous les élèves à la stratégie de sélection de l'ordinateur peut alors contribuer à l'augmentation de l'effet de l'ordre de difficulté décroissant des items puisqu'aucun élève a la possibilité de déjouer l'ordre de difficulté des items présentés dans le test.

Cette étude tente aussi d'apporter plus d'éléments de comparaison entre le TAO selon la TRI et le testing conventionnel (TC). Le testing selon la TRI se révèle supérieur au testing conventionnel par rapport aux extrémités de la distribution normale et équivalent pour les élèves d'habileté moyenne (Kingsbury &

Weiss, 1981). Cette supériorité est attribuable à la possibilité de mesurer tous les élèves avec un même degré d'erreur. Que dire cependant de la précision du TAO par rapport au TC si on démontre que sa stratégie de sélection pénalise les élèves d'habileté faible.

En somme, la recension des écrits nous a permis de mettre en lumière la nécessité d'entreprendre cette étude. Tout d'abord, des questions sont soulevées dans les études de Kiely, Plake, Wainer et Wise au sujet de l'effet de l'ordre de difficulté des items chez les élèves d'habileté faible, lors d'un TAO. De plus, il n'existe aucune étude empirique sur l'impact de la stratégie de sélection de la TRI sur la performance des élèves d'habileté faible. Il faut ajouter aussi qu'on ne connaît pas encore le rôle de certaines caractéristiques du TAO dont celui de la mobilité de l'élève dans un test. Aussi, il nous apparaît important de se demander: quel effet peut avoir cette variable sur l'ordre de difficulté des items? Et pour terminer, cette étude permet de remettre en question la supériorité du TAO selon la TRI par rapport au testing conventionnel, en fonction des élèves d'habileté faible.

3. Synthèse: Problème et hypothèses de recherche

La présente étude docimologique se préoccupe du comportement

des élèves d'habileté faible face à un test adaptatif sur ordinateur, administré selon la TRI. L'élève d'habileté faible reçoit, dans ce genre de testing, des items dont l'indice de difficulté va du plus difficile au plus facile. Or, depuis 40 ans environ, on recommande de faire des tests avec un ordre de difficulté du plus facile au plus difficile ou aléatoire parce que l'ordre de difficulté du plus difficile au plus facile, a pour effet de frustrer l'élève, d'augmenter la difficulté du test et ainsi, de nuire à la performance (Hambleton, 1968; MacNicol, 1956; Mollendopf, 1950; Monk & Stallings, 1970; Sax & Carr, 1962). Sax & Cromack (1966) précisent aussi que ce n'est qu'en présence d'un test à temps limité que l'élève est pénalisé par un ordre de difficulté décroissant. Toutes ces recherches furent effectuées avec des tests conventionnels de type papier-crayon. Que se passe-t-il alors avec l'utilisation de tests adaptifs sur ordinateur?

Le TAO possède par contre certaines caractéristiques différentes du testing conventionnel. Tout d'abord, tous les élèves doivent répondre aux items dans l'ordre déterminé par l'ordinateur sans avoir la possibilité de vérifier ou de changer leurs réponses. On ne connaît pas encore le rôle de cette absence de mobilité. (Wise & Plake, 1989).

Un autre aspect à explorer est celui de la possibilité

d'établir un parallèle entre le rôle de la variable du temps limité sur la variable de l'ordre de difficulté des items lors d'un TC de type papier-crayon et le rôle de la variable de l'absence de mobilité de l'élève sur l'ordre de difficulté des items lors d'un TAO. Sax et Cromack (1966) mentionnent que l'effet de l'ordre de difficulté des items du plus difficile au plus facile existe seulement en présence d'un test conventionnel papier-crayon à temps limité. La participation de l'élève à ce genre de testing le pousse peut-être à répondre aux items dans l'ordre présenté et à ne pas vérifier ses réponses. Ainsi, l'élève se retrouve dans des conditions semblables à l'élève d'habileté faible lors d'un TAO selon la TRI. Il répond à un test dont l'ordre de difficulté des items est du plus difficile au plus facile. Il ne prend pas le temps de répondre aux items selon sa propre stratégie de sélection. Il ne bénéficie donc d'aucune mobilité dans son test.

En résumé, l'élève d'habileté faible reçoit habituellement dans notre système d'éducation, un test de type papier-crayon dont l'ordre de difficulté des items est aléatoire. Qu'advient-il cependant lorsque ces élèves d'habileté faible reçoivent le même test mais dont les items sont présentés dans un ordre de difficulté décroissant et administrés par un ordinateur qui ne permet pas d'adopter une stratégie de sélection différente de celle présentée

et qui ne permet pas non plus de réviser ses réponses. Aussi, il nous apparaît important de poser la question de recherche suivante: Est-ce que les scores des élèves d'habileté faible lors d'un testing adaptatif sur ordinateur (TAO) selon la théorie des réponses aux items (TRI) seront plus bas que les scores des élèves d'habileté faible lors d'un test conventionnel de type papier-crayon?

Hypothèses de recherche

Hypothèse générale

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté décroissant et administrés sur un ordinateur qui ne permet pas la révision, seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont administrés dans un ordre de difficulté aléatoire et au moyen d'un test de type papier-crayon.

Hypothèse 1

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté croissant et administrés sur un **ordinateur** qui ne permet pas la révision seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont administrés dans un ordre de difficulté croissant et au moyen d'un test de type **papier-crayon**.

Hypothèse 2

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté aléatoire et administrés sur un **ordinateur** qui ne permet pas la révision seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont administrés dans un ordre de difficulté aléatoire et au moyen d'un test de type **papier-crayon**.

Hypothèse 3

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté

décroissant et administrés sur un **ordinateur** qui ne permet pas la révision seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont administrés dans un ordre de difficulté décroissant et au moyen d'un test de type **papier-crayon**.

Hypothèse 4

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté aléatoire et administrés sur un **ordinateur** qui ne permet pas la révision seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont présentés dans un ordre de difficulté croissant et administrés sur un **ordinateur** qui ne permet pas la révision.

Hypothèse 5

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté décroissant et administrés sur un **ordinateur** qui ne permet pas la révision seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items seront présentés dans un ordre de difficulté croissant et administrés sur un **ordinateur** qui

ne permet pas la révision.

Hypothèse 6

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté décroissant et administrés sur un **ordinateur** qui ne permet pas la révision seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items seront présentés dans un ordre de difficulté aléatoire et administrés sur un **ordinateur** qui ne permet pas la révision.

CHAPITRE II

LA METHODOLOGIE

La recension des écrits présentée au chapitre précédent soulève le problème de l'ordre de difficulté des items d'un test adaptatif sur ordinateur (TAO) administré aux élèves d'habileté faible. En effet, un test dont l'ordre de difficulté va du plus difficile au plus facile nuit à la performance de ceux-ci. De plus, l'absence de mobilité dans le choix des items à répondre dans le TAO soumet davantage les élèves à l'ordre de difficulté des items que le test conventionnel (TC) de type papier-crayon. Ainsi l'effet de l'ordre de difficulté décroissant des items retrouvés lors d'un TC peut être présent sinon intensifié par l'absence de mobilité lors d'un TAO. Ce chapitre décrit l'échantillon et les instruments et présente le plan de l'analyse statistique.

1. Description de l'échantillon

Dans la présente recherche, tous les élèves d'habileté faible en mathématique, inscrits en 10e année dans le CECSLFPR (Conseil des écoles catholiques séparées de langue française de Prescott-Russell) pendant l'année 1991-1992, participent au projet d'étude. La moyenne en mathématique au niveau général obtenue en 9e année par les élèves permet l'identification de cette population. On déclare avec l'avènement du décroisement qu'il existe un grand

écart entre les élèves provenant de classes avancées et les élèves provenant de classes générales ou fondamentales. Ainsi, on estime qu'une moyenne inférieure ou égale à 70% provenant d'un niveau général représente la population des élèves d'habileté faible.

Parmi les 689 élèves inscrits en 10e année, 112 ont obtenu une moyenne égale ou inférieure à 70% en mathématique générale. Lors de l'administration des tests, 97 élèves se sont présentés. Afin de calculer les comparaisons a priori selon le test F avec le même nombre d'élèves dans chacun des groupes, sept copies furent éliminées. Voici le critère qui a permis cette sélection. Les tests dont la stratégie de l'élève consistait à se débarrasser de la tâche en encerclant toujours la même réponse par exemple, ne furent pas considérés. En éliminant ces tests, les copies de 90 élèves furent conservées. Le tableau 1 démontre la répartition des élèves à travers les six différents tests qui seront expliqués dans la présentation des instruments.

En Ontario, on ne procède pas à l'administration d'examens uniformes et on n'applique pas la méthode de la modération des

scores. Ainsi, le critère de sélection de 70% peut ne pas signifier le même degré de performance dans chacune des écoles utilisées dans cette étude. Pour remédier à ce problème, on a effectué une analyse de variance. Le test F-max de Hartley a tout d'abord été effectué afin de vérifier l'homogénéité de la variance. La valeur du F-max frontière était de 6,9 au seuil de $P \leq 0,01$. Le F-max obtenu fut 2,35 et l'hypothèse nulle fut alors acceptée. Les variances étant égales, on a ensuite tenté de vérifier s'il n'existait pas de différence significative entre les six groupes à l'aide de l'administration d'un test standardisé. Seules les sections n2 et n3 du T.A.R.A. (test analytique de rendement arithmétique) furent utilisées. Ces sections semblaient suffire puisqu'elles mesurent le rendement des élèves sur les fractions et le test de Noelting (1979) retenu dans cette étude, ne comporte que des proportions c'est-à-dire des fractions visuelles. De plus, l'administration du test en entier prenait trop de temps. Selon la méthode des moitiés, le T.A.R.A. a un indice de fidélité de 0,78 qui devient 0,87 avec la correction de Spearman-Brown. L'analyse de variance de ce test révèle qu'il n'existe aucune différence significative entre les groupes. Le tableau 2 présente le sommaire de cette analyse.

Répartition du nombre d'élèves par école à travers les six différents tests (90 élèves)

	Casselman	Embrun	Hawkes- bury	Planta- genet	Rockland	Total
Test 1 Papier- crayon Ordre croissant	7			7	1	15
Test 2 Papier- crayon Ordre aléatoire	9			4	2	15
Test 3 Papier- crayon Ordre décroissant	8			6	1	15
Test 4 Ordinateur Ordre croissant		3	12			15
Test 5 Ordinateur Ordre aléatoire		3	12			15
Test 6 Ordinateur Ordre décroissant		4	11			15
Total	24	10	35	17	4	90

Tableau 2

Sommaire de l'analyse de variance du T.A.R.A.

Source de variation	SC	dl	CM	F obs.
Inter-groupe	1,71	5	0,34	1
Intra-groupe	213,59	84	2,54	
Total	215,30	89		

2. Instruments

La présente recherche vise à démontrer l'influence de la combinaison de l'ordre de difficulté des items et de la mobilité de l'élève lorsqu'un test est administré à l'aide d'un ordinateur sur le score des élèves d'habileté faible. Cette étude nécessite donc un test qui permet de contrôler l'ordre de difficulté des items et un logiciel capable d'administrer ce genre de test. A cet égard, le test hiérarchique sur les proportions de Noelting (1979) comporte un ordre de difficulté des items appropriés à cette étude et le logiciel MicroCAT permet l'administration de ce test sur ordinateur tout en empêchant l'élève de réviser ses réponses. Au préalable cependant, le test standard T.A.R.A. (test analytique de rendement arithmétique) fut utilisé dans le but de vérifier s'il existait des différences significatives entre les groupes.

A. Test sur les verres de jus

A.1 But

Le test de Noelting sur les verres de jus est un test hiérarchique en ce qui a trait au développement du rapport et du raisonnement des proportions. On a choisi ce test car il a des items hiérarchiques essentiels à l'étude. De plus, le fait

d'utiliser un test de mathématique permet de ne pas laisser de place à la subjectivité lors de la correction.

Deux postulats fondamentaux sont décrits par Noelting (1979) et exposés ici afin de comprendre ce qu'est un test hiérarchique:

- 1) Le développement se fait de façon hiérarchique c'est-à-dire que chaque nouvelle structure intègre la précédente comme étant une sous-structure. Cette structure est un stade.
- 2) Le changement d'une structure à une autre se fait par la différenciation du schème existant. Un schème est une structure "behaviorale" qui résout un problème en particulier. Le schème qui ne peut plus comprendre certains aspects d'un problème doit subir un changement afin de comprendre tout le problème. Ce changement permet au schème de saisir le nouvel aspect du problème et de le relier avec l'aspect déjà compris. Ce genre de processus a déjà été décrit théoriquement par Piaget (1975-77) sous le terme d'équilibration ascendante.

En résumé, dans le Test sur les verres de jus de Noelting (1979), l'élève doit maîtriser les items d'un stade avant de passer à l'item du prochain stade. Ainsi, il est possible de déterminer l'ordre de difficulté des items de ce test. De plus, ce test permet d'identifier la difficulté de chaque item par rapport à l'âge de l'élève. En effet, Noelting (1979) a établi un âge

d'accessibilité à chacun des 8 stades cognitifs du développement du rapport et du raisonnement des proportions. Cet âge d'accessibilité est défini par la réussite d'un item par 50% des élèves du même âge.

A.2 Difficulté des items

Dans un test adaptatif sur ordinateur, l'élève d'habileté faible reçoit des items de difficulté moyenne et faible seulement. Afin de simuler cet aspect, le test de Noelting (1979) ne contient que des items de difficulté moyenne et faible en ce qui a trait aux élèves de 15 ans. Comme l'échantillon se compose d'élèves d'habileté faible en mathématique, ces derniers auraient échoué les items de difficulté élevée. Ainsi, il ne serait pas possible de bien discriminer parmi les élèves d'habileté faible. Afin de déterminer la difficulté des items, on s'est servi du tableau de Noelting (tableau 3) représentant l'âge d'accessibilité aux différents stades. Les items de ce tableau représentant l'expérience avec les verres de jus ont été placés dans un ordre selon la fréquence de succès furent ensuite placés dans des catégories afin de former des stades. La fréquence de succès a été établie à partir de 321 sujets entre 6 et 16 ans.

Tableau 3

Items de l'expérience des verres de jus de Noelting (1979)

Stade	Nom	Item	Fréquence de succès	Age d'accessibilité
0	Symbolique	0	-	(2;0)
IA	Intuitif inférieur	2	319	(3;6)
		6	319	
		4	319	
IB	Intuitif de transition	1	311	(6;4)
		3	307	
		5	305	
IC	Intuitif supérieur	8	295	(7;0)
		13	291	
		10	297	
IIA	Opérationnel concret inférieur	9	251	(8;1)
		11	244	
		7	231	
IIB	Opérationnel concret	12	186	(10;5)
		15	156	
IIIAI	Opérationnel formel inférieur	16	141	(12;2)
		17	131	
		14	107	
		18	88	
IIIA2	Opérationnel formel inférieur	20	87	(12;2)
		22	71	
		19	65	
		21	59	
IIIB	Opérationnel formel supérieur	23	51	(15;1)
		24	-	
		25	-	

Les 15 derniers items du test de Noelting (1979) représentent des items de difficulté moyenne et faible pour l'échantillon de cette présente recherche dont l'âge varie entre 15 et 16 ans. Plus précisément, les items provenant des stades IIA, IIB et IIIA1 sont considérés comme des items de difficulté faible, alors que les items provenant des stades IIIA2 et IIIB sont perçus comme des items de difficulté moyenne.

A.3 La hiérarchisation du Test sur les verres de jus de Noelting (1979)

L'analyse "Guttman-type scalogram" a permis de vérifier la hiérarchisation des items du test. Un coefficient de CR=0,942 et MMR=0,78 fut obtenu. A ce sujet, on considère un coefficient de CR=90 satisfaisant avec un MMR=0,80. De plus, $PPR = \frac{CR-MMR}{1-MMR} = 0,735$

et un coefficient plus grand que 0,70 est considéré comme bon. On peut donc utiliser, dans cette étude, les items du Test sur les verres de jus de Noelting (1979) dans une perspective hiérarchique.

A.4 Arrangement des items du Test sur les verres de jus de Noelting (1979) en six versions d'un même test

Afin de comparer la stratégie du TAO selon la TRI utilisée avec des élèves d'habileté faible au testing conventionnel de type papier-crayon, il importe de manipuler deux variables: l'ordre de difficulté des items et la mobilité de l'élève dans un test. La mobilité fait référence à la possibilité de l'élève à répondre aux items qu'il a choisis et de réviser ses réponses. Lors d'un testing conventionnel de type papier-crayon, la mobilité de l'élève est possible. Par contre, elle n'est plus possible lorsque le test est administré à l'aide d'un ordinateur. La variable de l'ordre croissant, aléatoire ou décroissant de la difficulté des items est présente dans les deux types de test, papier-crayon et ordinateur. La figure 1 démontre visuellement la combinaison des variables des six tests. L'hypothèse générale vise à comparer le test 2 au test 6.

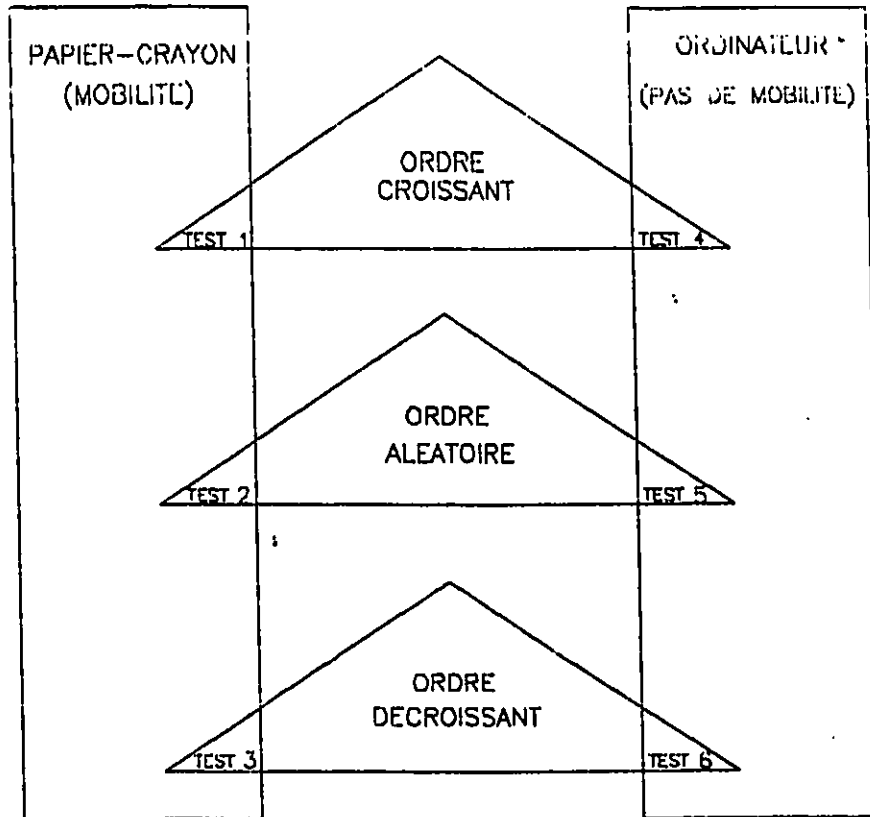


Figure 1 Six différentes versions du Test sur les verres de jus de Noelting (1979) élaborées en fonction de la variable de l'ordre de difficulté des items et la variable de la mobilité de l'élève dans un test.

B. MicroCAT

Ce progiciel permet de concevoir, d'administrer et de corriger un test conventionnel ou adaptatif sur ordinateur selon la théorie classique des scores (TCS) ou la théorie des réponses aux items (TRI). Il fut choisi en tenant compte des raisons suivantes. Chaque item du Test sur les verres de jus de Noelting (1979) comporte un texte et un graphique. Deux facteurs ont empêché la calibration de chaque item du test. Tout d'abord, cela aurait pris trop de temps puisqu'il faut 1000 sujets afin de calibrer un item à trois paramètres. De plus, l'objet de cette étude n'est pas la calibration d'items mais le rôle de la stratégie de sélection de l'ordre de difficulté des items de la TRI sur la performance de l'élève d'habileté faible. Ce progiciel permet alors d'imiter la stratégie de sélection des items de la TRI envers les élèves d'habileté faible en donnant la possibilité au chercheur de contrôler à l'avance l'ordre de sélection des items. De plus, MicroCAT administre des tests adaptatifs sur ordinateur selon la TRI mais aussi des tests conventionnels et les corrige selon la théorie classique des scores. Dans le présent projet de recherche, ce progiciel permet la conception et l'administration des tests 4, 5 et 6.

C. Procédures de l'administration des tests

Tous les élèves inscrits en 10e année générale ayant obtenu 70% et moins de moyenne en mathématique lors de la 9e année générale sont appelés à répondre au test. Dans les écoles secondaire d'Embrun et de Hawkesbury, les tests N.4, N.5 et N.6 furent distribués au hasard en raison de la disponibilité du matériel informatique. En effet, ces écoles possèdent un réseau d'ordinateurs IBM avec écran VGA, nécessaire à l'administration de ces tests. Dans les écoles secondaires de Casselman, Plantagenet et Rockland, les tests N.1, N.2 et N.3 furent administrés au hasard car ils sont de type papier-crayon et ne demandent aucune attention particulière. Ainsi trois groupes furent formés à partir des écoles d'Embrun et de Hawkesbury et les trois autres groupes proviennent des écoles de Casselman, Plantagenet et Rockland. A l'aide des deux premiers exemples des Test sur les verres de jus de Noelting (1979), l'administrateur du test explique le test aux élèves. Les élèves n'ont pas eu de limite de temps pour répondre aux tests. La majorité des élèves ont pris 10 à 15 minutes pour répondre au test. On encourage aussi les élèves à bien réussir. A la suite du test sur les verres de jus, l'élève répond au test standard T.A.R.A. En s'intéressant à la réaction de l'élève par rapport à l'ordre de difficulté des items administrés par un ordinateur, il importe de ne pas administrer le test

standardisé T.A.R.A. avant celui des verres de jus, sinon l'effet de l'ordre de difficulté des items chez l'élève pourrait être biaisé.

D. Plan de l'analyse statistique

Le plan de l'analyse statistique de la présente recherche a été effectué à l'aide de comparaisons planifiées (a priori). Les critères qui, selon Keppel et Zedeck (1989) déterminent l'utilisation de cette méthode sont les suivants: 1) les comparaisons devraient être intégrées dans le cadre expérimental 2) les comparaisons devraient représenter le but principal de l'expérience 3) les comparaisons devraient constituer des tests directs et pertinents en ce qui a trait aux hypothèses de recherche. Cette méthode a donc permis de s'attarder seulement aux comparaisons déjà identifiées au début de l'étude en plus d'offrir un test rigoureux. Le test F, à un degré de liberté unique, a été utilisé afin de comparer les groupes dans l'étude présente.

L'ordre de difficulté des items est la variable indépendante. La mobilité de l'élève à répondre aux questions du test est en quelque sorte une variable modératrice qui agit sur l'effet de l'ordre de difficulté des items. Cette dernière est présente lors

de l'administration des items avec un ordinateur. Les scores des élèves d'habileté faible obtenus lors du Test sur les verres de jus de Noelting (1979) constituent la variable dépendante. Ce test est administré de 6 différentes façons grâce à la manipulation de la variable indépendante (l'ordre de difficulté des items) et la variable modératrice (mobilité de l'élève dans le test).

L'hypothèse générale de recherche prévoit un lien entre la combinaison de l'ordre décroissant de la difficulté des items et l'absence de la mobilité de l'élève dans un test sur le score du test de l'élève. Cette combinaison se produit lors de l'administration des réponses aux items (TRI) en présence d'élèves d'habileté faible. Les 6 autres hypothèses visent à expliquer le résultat obtenu par l'hypothèse de recherche générale. A l'aide de comparaisons planifiées selon le test F, il sera possible de vérifier cette hypothèse. La description et l'interprétation des résultats font l'objet du prochain chapitre.

CHAPITRE III

LA PRESENTATION ET L'INTERPRETATION DES RESULTATS

Les données relatives à l'hypothèse générale et aux 6 autres hypothèses de recherche ont été analysées en utilisant la technique des comparaisons a priori selon le test F à degré de liberté unique. Le but de ce chapitre est de présenter ces résultats et de les interpréter.

1. La présentation des résultats

L'hypothèse générale de cette recherche est la suivante. Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté décroissant et administrés sur un ordinateur qui ne permet pas la révision des réponses (test 6), seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont administrés dans un ordre de difficulté aléatoire et au moyen d'un test de type papier-crayon (test 2). En réalité, on tente de simuler la stratégie de sélection des items d'un test adaptatif sur ordinateur (TAO) selon la théorie des réponses aux items (TRI) utilisée envers les élèves d'habileté faible et de le comparer à un test conventionnel.

Après s'être assuré de l'équivalence des groupes, les moyennes

et les écart-types des six différentes versions du Test sur les verres de jus de Noelting (1979) furent calculés. Les prochains tableaux et figures présentent une analyse descriptive des données relatives à chacune des hypothèses.

TABLEAU 4

Moyenne et écart-types des six différentes versions du Test sur les verres de jus.

	Test 1 Papier-crayon croissant	Test 2 Papier-crayon aléatoire	Test 3 Papier-crayon décroissant	Test 4 Ordina- teur crois- sant	Test 5 Ordina- teur aléa- toire	Test 6 Ordinateur décroissant
\bar{x} :	45,07	48,00	40,87	40,40	34,67	31,13
s:	2,92	2,83	1,96	1,71	2,43	1,84

FIGURE 2

Comparaison des moyennes du test 2 au test 6 (hypothèse générale)

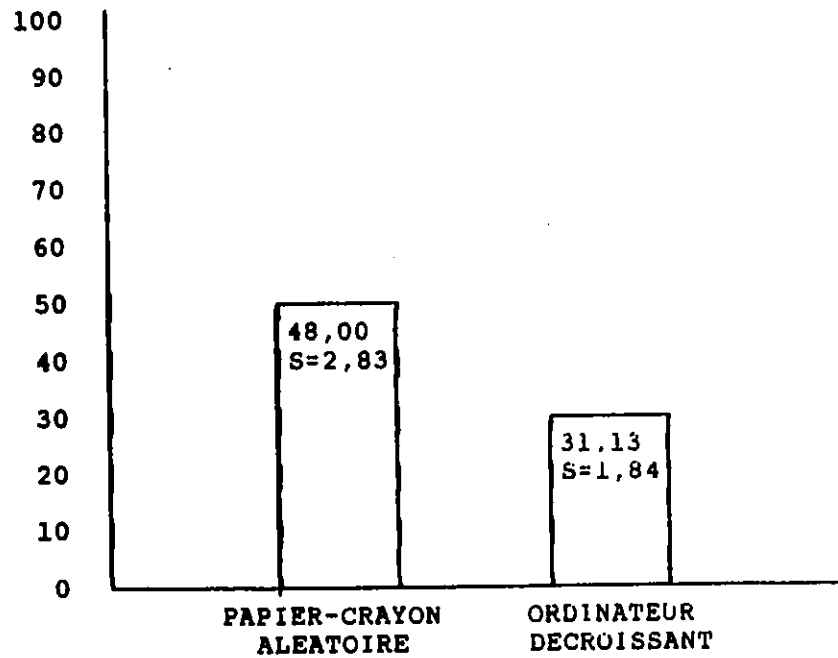


FIGURE 3

Comparaison des moyennes du test 1 au test 4. (hypothèse 1)

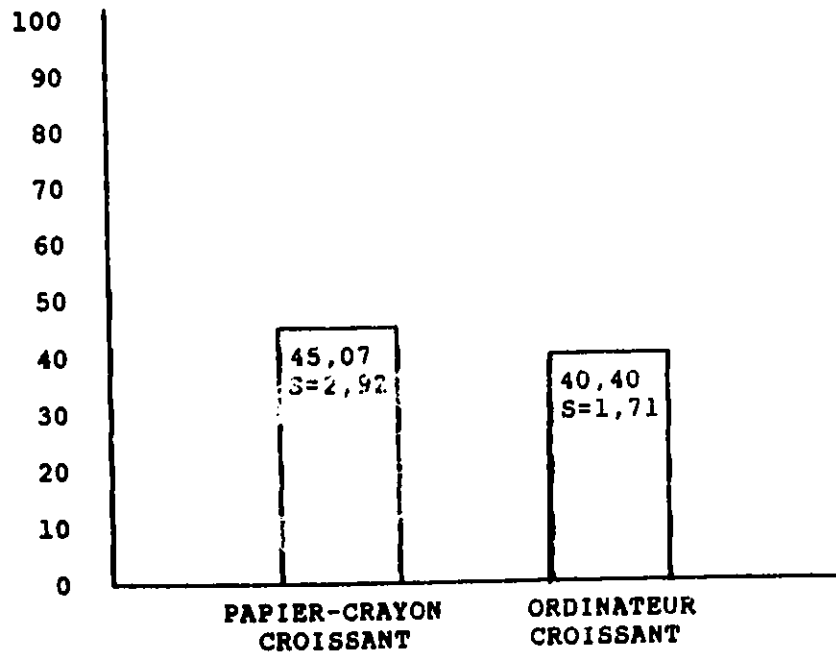


Figure 4 Comparaison des moyennes du test 2 au test 5.
(hypothèse 2)

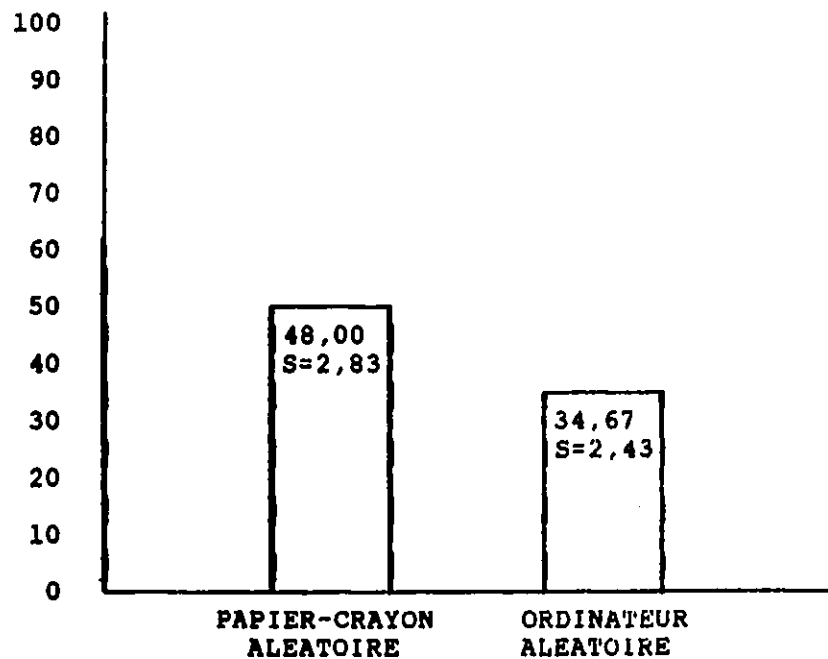


Figure 5 Comparaison des moyennes du test 3 au test 6 (hypothèse 3)

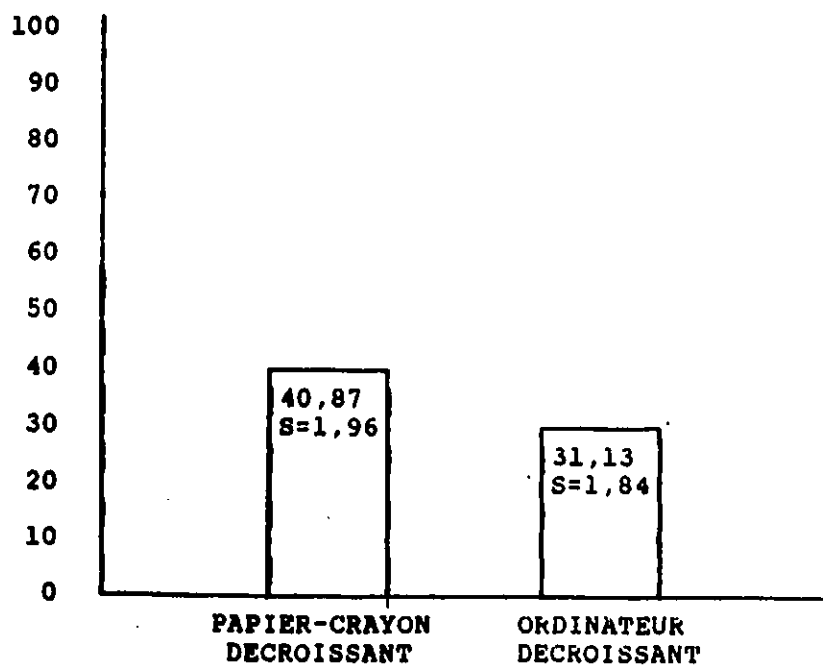


Figure 6 Comparaison des moyennes du test 4 au test 5 (hypothèse 4)

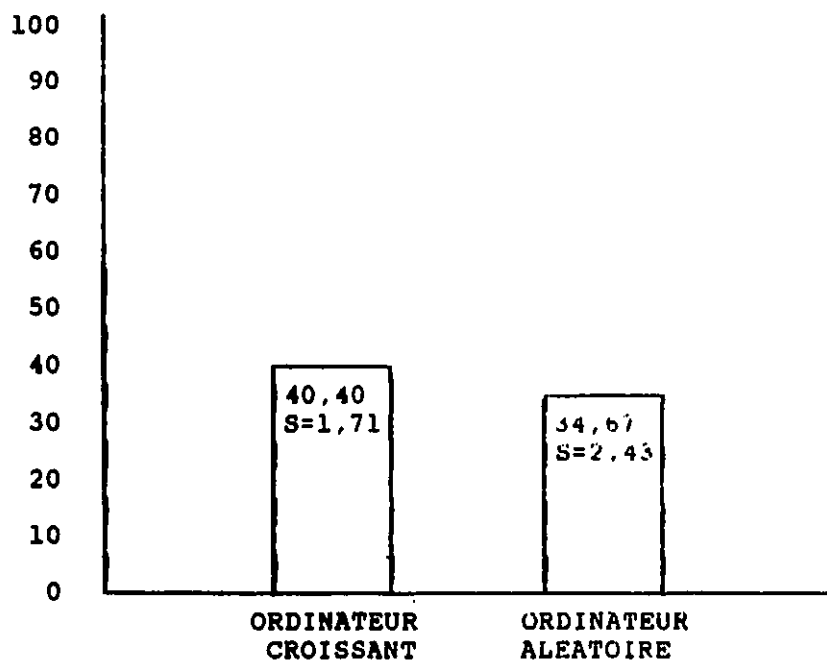


Figure 7 Comparaison des moyennes du test 4 au test 6 (hypothèse 5)

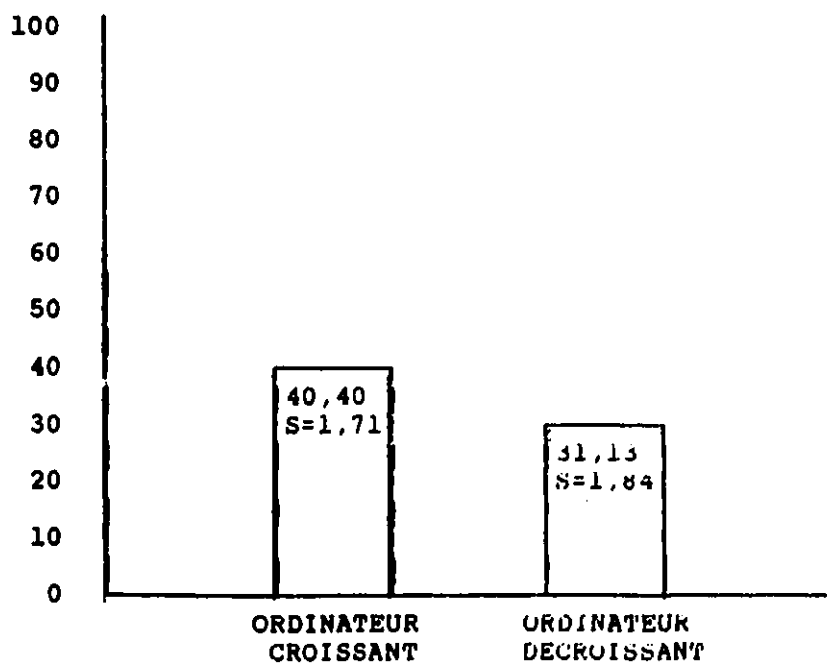
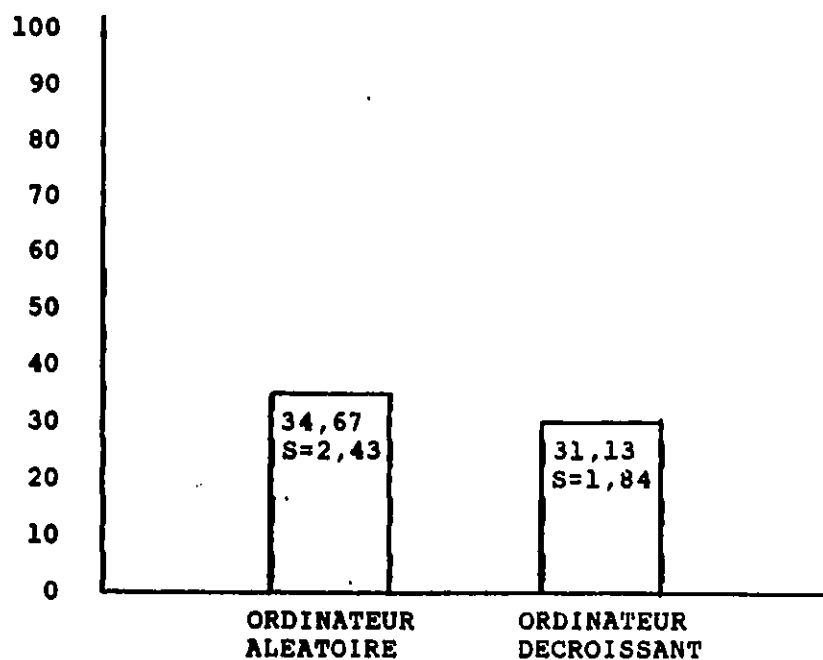


Figure 8 Comparaison des moyennes du test 5 au test 6 (hypothèse 6)



2. Interprétation des résultats

Cette section tente d'expliquer les liens qui semblent exister entre les variables de l'ordre de difficulté des items et de la mobilité de l'élève dans un test sur la performance des élèves d'habileté faible. Ainsi, on pourra argumenter que le testing adaptatif sur ordinateur (TAO) selon la théorie des réponses aux items (TRI) peut pénaliser les élèves d'habileté faible. Pour ce faire, les résultats obtenus lors des comparaisons sont exposés au tableau 5. Ensuite, chaque hypothèse est interprétée. Enfin, une interprétation générale termine cette section.

Hypothèse générale

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté décroissant et administrés sur un **ordinateur** qui ne permet pas la révision (Test 6) seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont administrés dans un ordre de difficulté aléatoire et au moyen d'un test de type **papier-crayon** (Test 2).

Sommaire des comparaisons planifiées
de chaque hypothèse selon le test F

Comparaisons	SS	df	MS	F
<u>Hypothèse générale</u> papier-crayon, aléatoire vs ordinateur décroissant (test 2 vs test 6)	48.01	1	48.01	8.84**
<u>Papier-crayon vs Ordinateur</u> Hypothèse 1 papier-crayon, croissant vs ordinateur, croissant (test 1 vs test 4)	3.36	1	3.36	0.62
Hypothèse 2 papier-crayon, aléatoire vs ordinateur aléatoire (test 2 vs test 5)	30.00	1	30.00	5.52*
Hypothèse 3 papier-crayon, décroissant vs ordinateur, décroissant (test 3 vs test 6)	15.99	1	15.99	2.94*
<u>Ordinateur vs Ordinateur</u>				
Hypothèse 4 Ordinateur, croissant vs ordinateur, aléatoire (test 4 vs test 5)	5.55	1	5.55	1.02
Hypothèse 5 Ordinateur croissant vs ordinateur décroissant (test 4 vs test 6)	14.49	1	14.49	2.67*
Hypothèse 6 ordinateur aléatoire vs ordinateur décroissant (test 5 vs test 6)	2.11	1	2.11	0.39
S/A		84	5.53	

* $p \leq .05$ ** $p \leq .01$

Le test F permet d'établir qu'il semble exister une différence significative à $P \leq 0,01$ entre la simulation d'un test adaptatif sur ordinateur (TAO) selon la théorie des réponses aux items (TRI) et un test conventionnel de type papier-crayon dont l'ordre de difficulté des items est aléatoire lorsqu'on est en présence d'élèves d'habileté faible. Ces élèves obtiennent des scores plus bas lors de la simulation du TAO selon la TRI que sur un test conventionnel de type papier-crayon. Les six prochaines hypothèses tenteront d'examiner le rôle des variables "ordre de difficulté des items" et "mobilité de l'élève" sur la performance des élèves d'habileté faible afin d'expliquer cette différence significative.

Hypothèse 1

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté croissant et administrés sur un **ordinateur** qui ne permet pas la révision (Test 4) seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont administrés dans un ordre de difficulté croissant et au moyen d'un test de type **papier-crayon** (Test 1).

Il ne semble pas y avoir de différence significative entre un test de type papier-crayon et un test sur ordinateur qui ne permet pas la révision lorsque l'ordre de difficulté des items est croissant. La variable mobilité ne semble donc pas jouer un rôle sur la performance des élèves d'habileté faible en présence d'un ordre croissant de la difficulté des items. On suppose que ceci s'explique par le fait qu'un élève répond habituellement aux items dans un ordre croissant de difficulté lorsqu'il a la chance d'être mobile dans le test. Ainsi, en présence d'un ordre croissant de difficulté, la variable mobilité n'influence pas les résultats car l'élève aurait répondu dans cet ordre naturellement.

Hypothèse 2

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté aléatoire et administrés sur un **ordinateur** qui ne permet pas la révision (Test 5) seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont administrés dans un ordre de difficulté aléatoire et au moyen d'un test de type **papier-crayon** (Test 2).

Le résultat de cette hypothèse démontre par contre qu'en présence d'un ordre aléatoire de difficulté des items, il existe une différence significative entre les deux modes de testing. Lors d'un testing sur ordinateur, l'élève ne peut adopter un ordre croissant de difficulté des items à répondre car il n'a pas de mobilité. Le testing conventionnel cependant, permet à l'élève d'être mobile à l'intérieur du test et par ce fait, le laisse choisir sa propre stratégie de sélection des items à répondre. Le test conventionnel de type papier-crayon avec un ordre aléatoire de la difficulté des items semble donc supérieur au test sur ordinateur dont l'ordre de difficulté des items est aléatoire en raison de la mobilité.

Hypothèse 3

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté décroissant et administrés sur un **ordinateur** qui ne permet pas la révision (Test 6) seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont administrés dans un ordre de difficulté décroissant et au moyen d'un test de type **papier-crayon** (Test 3).

Les résultats de cette hypothèse sont significatifs pour les mêmes raisons exposées précédemment (hypothèse 2). Comme le test sur ordinateur ne donne pas accès à la mobilité, l'élève ne peut déjouer la sélection d'items présentés et répondre selon sa propre stratégie de sélection des items. Ainsi, il ne peut répondre aux items dans un ordre croissant de la difficulté de ceux-ci. Encore une fois, le testing de type papier-crayon semble supérieur au testing sur ordinateur lorsqu'on est en présence d'un ordre décroissant de la difficulté des items.

Hypothèse 4

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté aléatoire et administrés sur un **ordinateur** qui ne permet pas la révision (Test 4), seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items sont présentés dans un ordre croissant de difficulté et administrés sur un **ordinateur** qui ne permet pas la révision (Test 5).

L'hypothèse 4 porte sur l'étude de la combinaison de la

variable "absence de mobilité" aux variables "ordre croissant de difficulté des items" et "ordre aléatoire de difficulté des items" sur la performance des élèves d'habileté faible. Le résultat de cette hypothèse n'est pas significatif. Cependant, la moyenne du test sur ordinateur avec un ordre croissant de la difficulté des items est supérieure à celle obtenue par le test administré sur ordinateur et dont les items sont présentés dans un ordre de difficulté aléatoire. Cette tendance suit l'idée de la sélection naturelle d'un élève à choisir un ordre croissant de la difficulté des items. Lors de tests papier-crayon, il ne semble pas non plus y avoir de différence significative entre un ordre de difficulté aléatoire et croissant des items. C'est plutôt en présence d'un ordre de difficulté décroissant des items que plusieurs chercheurs (Hambleton, 1968; MacNicol, 1956; Mollendopf, 1950; Monk & Stallings, 1970; et Sax & Carr, 1962) démontrent l'effet négatif de la variable "ordre de difficulté des items" sur la performance d'un élève.

Hypothèse 5

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté décroissant et administrés sur un ordinateur qui ne permet pas la révision (Test 6) seront plus bas que les scores des élèves

d'habileté faible obtenus lors du même test mais dont les items seront présentés dans un ordre de difficulté croissant et administrés sur un **ordinateur** qui ne permet pas la révision (Test 4). On retrouve une différence significative entre un ordre de difficulté croissant des items et un ordre de difficulté décroissant lorsqu'on est en présence d'un test sur ordinateur. Ainsi, en l'absence de mobilité, les élèves réussissent mieux un test avec un ordre de difficulté croissant qu'un test dont l'ordre de difficulté des items est décroissant. Ce résultat tend aussi vers l'idée de la sélection naturelle de l'élève qui est de répondre aux items dans un ordre croissant.

Hypothèse 6

Les scores des élèves d'habileté faible obtenus lors d'un test dont les items sont présentés dans un ordre de difficulté décroissant et administrés sur un **ordinateur** qui ne permet pas la révision (Test 6) seront plus bas que les scores des élèves d'habileté faible obtenus lors du même test mais dont les items seront présentés dans un ordre de difficulté aléatoire et administrés sur un **ordinateur** qui ne permet pas la révision (Test 5).

Il n'y a pas de différence significative entre les deux ordres de difficulté, aléatoire et décroissant lors de l'administration d'un test sur ordinateur. Dans aucun des cas, l'élève ne peut adopter une stratégie de sélection des items à répondre, du plus facile au plus difficile car il y a absence de mobilité.

Cependant, la moyenne du test avec des items dans un ordre aléatoire est supérieure à celle dont l'ordre est décroissant. Plusieurs chercheurs (Hambleton, 1968; MacNicol, 1956; Mollendopf, 1950; Monk & Stallings, 1970; Sax & Carr, 1962) ont démontré de façon significative la supériorité du test avec un ordre aléatoire de la difficulté des items par rapport au test dont l'ordre de difficulté est décroissant. Ceci se produisit par contre lors de tests de type papier-crayon qui offre une certaine mobilité.

Ainsi, en ce qui concerne les élèves d'habileté faible, il semble que le testing conventionnel de type papier-crayon soit préférable au testing adaptatif sur ordinateur (TAO) selon la théorie des réponses aux items (TRI). La combinaison des variables "ordre de difficulté des items" et "mobilité de l'élève dans un test" sont à la base de ces deux différents modes de testing.

Lorsque des tests papier-crayon sont administrés, la moyenne du test dont l'ordre de difficulté des items est décroissant est inférieure aux deux autres moyennes provenant de tests dont l'ordre est aléatoire ou croissant (Hambleton, 1968; MacNicol, 1956; Mollendopf, 1950; Monk & Stalligns, 1970; et Sax & Carr, 1962). Dans la présente étude, la moyenne des tests papier-crayon est significativement supérieure à tous les tests administrés sur ordinateur quelque soit l'ordre de difficulté des items. Cependant, le test dont la moyenne est la plus basse de tous les tests de cette recherche, combine les variables "absence de mobilité" et "ordre de difficulté décroissant des items". Il est important de souligner la présence de la combinaison de ces deux variables dans un TAO selon la TRI envers les élèves d'habileté faible.

Il importe aussi de mentionner que plusieurs chercheurs (Nelson, Towle & Merrill, 1955; Sax & Cromack, 1966) ont démontré que la variable "ordre de difficulté décroissant des items" n'a d'effet qu'en présence d'un test avec la variable "temps limite". Il se peut que dans ces études, les élèves se comportent comme les élèves qui reçoivent un test sur ordinateur qui ne permet pas la révision. Ceux-ci ne répondent pas aux items selon leur stratégie

de sélection naturelle, c'est-à-dire en ordre de difficulté croissant ou ne révisent pas leurs réponses par manque de temps. La présente étude pourrait alors suggérer que la combinaison de la variable "absence de mobilité" et "ordre de difficulté décroissant des items" qu'on retrouve lors d'un TAO selon la TRI ressemble à la combinaison de la variable "temps limite" et "ordre de difficulté décroissant des items" qu'on retrouve lors d'un testing conventionnel de type papier-crayon. Ainsi, la variable "absence de mobilité dans un test" et la variable de "temps limité" pourraient produire le même effet sur le comportement de l'élève. Il se peut alors que ce ne soit qu'en présence de la combinaison de l'une de ces variables à la variable "ordre de difficulté décroissant des items" que la performance de l'élève soit pénalisée.

Ainsi la variable "mobilité de l'élève ne pourrait-elle pas être considérée comme la variable indépendante? La variable "ordre de difficulté de items" deviendrait alors la variable modératrice agissant sur la variable indépendante en présence d'un test administré sur un ordinateur qui ne permet pas la révision.

En résumé, les élèves d'habileté faible semblent être pénalisés par le TAO selon la TRI par rapport au testing conventionnel. Ceci semble être attribué par contre plutôt à la variable

"absence de mobilité de l'élève dans un test" qu'à la variable "ordre de difficulté décroissant des items". Il est important de noter cependant que ce phénomène ne semble se produire qu'en présence de la combinaison de ces deux variables.

Toutefois, cette étude a été réalisée dans certaines conditions spécifiques. Les caractéristiques de l'échantillon et la simulation du testing adaptatif sur ordinateur (TAO) selon la théorie des réponses aux items (TRI) comportent certaines limites. Tout d'abord, on a pris tous les élèves ayant obtenu une moyenne égale ou inférieure à 70% en mathématique général en 9e année l'an passé dans le Conseil des écoles séparées de langue française de Prescott-Russell. En Ontario, cependant, il n'existe pas d'examen uniforme comme au Québec. A l'aide du T.A.R.A., on n'a obtenu aucune différence significative des moyennes parmi les 6 groupes qui ont effectué ce test standard. Le T.A.R.A. par contre ne mesure qu'une petite partie des connaissances dans le domaine de la proportion en mathématique. Il faut donc prendre cela en considération avant de généraliser les résultats. Il est important de noter également qu'une comparaison a été établie entre le testing conventionnel de type papier-crayon et une simulation du testing adaptatif sur ordinateur selon la TRI chez les élèves d'habileté faible. Comment se comporteront les élèves d'habileté faible lors d'un TAO selon la TRI reste à être examiné. La théorie

des réponses aux items (TRI) calcule le résultat d'un test différemment de la théorie classique des scores. On se demande alors jusqu'à quel point les calculs mathématiques derrière cette théorie peuvent éviter ou contribuer à l'effet retrouvé lors de la combinaison des variables "absence de mobilité de l'élève dans un test" et "ordre de difficulté des items" chez les élèves d'habileté faible.

CHAPITRE IV

RESUME ET CONCLUSION

Cette étude visait à comparer le rendement des élèves d'habileté faible en mathématique sur un test conventionnel (TC) et une simulation d'un testing adaptatif sur ordinateur (TAO) selon la théorie des réponses aux items (TRI) chez les élèves d'habileté faible en mathématique. Le testing conventionnel est représenté par un test de type papier-crayon offrant des items dans un ordre aléatoire de difficulté. Quant à la simulation du TAO selon la TRI, elle s'effectue grâce à l'administration sur ordinateur d'items dont l'ordre de difficulté va du plus difficile au plus facile. On s'assure aussi que l'élève ne peut en aucun temps réviser ses réponses. L'échantillon se compose de tous les élèves inscrits en 10^e année dans le C.E.C.S.L.F.P.R. et ayant obtenu l'an passé une moyenne égale ou inférieure à 70% en mathématique au niveau général. Comme il n'existe pas de test standard en Ontario, la valeur de 70% est relative dans chacune des écoles. Le T.A.R.A. a permis d'établir cependant qu'il n'y avait aucune différence significative entre les groupes. Ainsi, à l'aide de comparaisons a priori selon le test F, on a comparé le testing conventionnel à une simulation du TAO selon la TRI envers les élèves d'habileté faible en mathématique. On a alors obtenu une différence significative au seuil de $P \leq 0,01$ entre les moyennes. Les élèves d'habileté faible en mathématique réussissent mieux lors de l'administration conventionnelle d'un test. Ce genre de test est

représenté par la combinaison des variables "ordre de difficulté aléatoire des items" et "mobilité de l'élève à l'intérieur d'un test". Cette mobilité s'explique par la possibilité de l'élève de répondre dans l'ordre de son choix aux items du test. Le TAO selon la TRI par contre n'offre pas cette possibilité aux élèves d'habileté faible en raison de sa stratégie de sélection. De plus, il présente des items dans un ordre de difficulté décroissant lors de l'administration du test chez les élèves d'habileté faible. Les 6 autres hypothèses ont permis de supposer que la variable "absence de mobilité" dans un test joue un plus grand rôle que la variable "ordre de difficulté décroissant des items". Il est à noter cependant que c'est en combinant les deux variables qu'on obtient la moyenne la plus basse de toutes les combinaisons effectuées dans cette étude. De plus, la combinaison de ces deux variables s'effectue en présence du TAO selon la TRI chez les élèves d'habileté faible. On recommande alors de refaire l'étude mais avec un vrai TAO selon la TRI, c'est-à-dire avec des items calibrés à un, deux ou trois paramètres et une compilation des résultats selon la théorie des réponses aux items (TRI). On ne sait pas jusqu'à quel point les calculs mathématiques de la TRI sont en mesure d'éviter ou d'accentuer la combinaison de ces deux variables "absence de mobilité" et "ordre de difficulté décroissant des items" en présence d'élèves d'habileté faible. On suggère aussi d'utiliser lors d'une prochaine étude, un échantillon plus vaste.

Cette étude permet donc de supposer que le TAO selon la TRI est en mesure de pénaliser les élèves d'habileté faible. De par sa stratégie de sélection, le TAO selon la TRI, offre aux élève d'habileté faible, la combinaison des variables "ordre décroissant de la difficulté des items" et "absence de mobilité de l'élève dans le test". On suppose aussi que cette situation de testing ressemble à celle où les élèves reçoivent des items dont l'ordre de difficulté des items est décroissant et où il s'agit d'un test à temps limité. Dans les études de Nelson, Towle & Merrill (1955) et Sax & Cromack (1966), avec des tests de type papier-crayon, la performance de l'élève était pénalisée aussi par ce genre de testing. De plus, il faut remarquer que le TAO selon la TRI se présente comme étant supérieur au testing conventionnel envers les extrémités de la distribution normale grâce à sa possibilité de mesurer les élèves avec la même erreur de mesure (Kingsbury et Weiss, 1981). Que dire alors, si la sélection et l'administration des items selon cette théorie influence la performance des élèves d'habileté faible?

Il importe maintenant de continuer à explorer tous les aspects, positifs ou négatifs, que peut apporter cette nouvelle approche au domaine de la mesure. Ainsi, il sera possible de bien connaître les avantages et les limites du testing adaptatif sur ordinateur selon la théorie des réponses aux items dans le but d'en

faire un bon usage.

BIBLIOGRAPHIE

62

- Ahmann, J.S., & Glock M.D. (1971) (1975) Measuring and Evaluating Educational Achievement, Boston: Allyn and Bacon, Inc. (in Nelson J.T. & Merrill P.F. Effects of Anxiety Type and Item-Difficulty Sequencing on Mathematics Test Performance. Journal of Educational Measurement, 12, 241-249.
- Angoff, W.H., & Huddleston, E.M. (1958). The multi-level experiment: A study of a two-level test system for the College Board Scholastic Aptitude Test. (Statistical Report No. SR-58-21). Princeton, NJ: Educational Testing Service (Reckase M.D. Adaptive Testing: The Evolution of a Good Idea, Educational Measurement Issues & Practice, 8, 11-15).
- Auger, J. (1989) Etude de praticabilité du testing adaptatif de maîtrise des apprentissages scolaires du Québec. Thèse de doctorat non publiée, Université Laval, Québec.
- Berger, V.F. Munz, D.C. Smoux, A.D., & Angeline H. (1969) The effects of item difficulty sequencing on anxiety reaction type in aptitude test performance. The Journal of Psychology, 71, 253-258. (in Nelson J.T. and Merrill P.F., Effects of Anxiety Type and Item-Difficulty Sequencing on Mathematics Test Performance, Journal of Educational Measurement, 12, 241-249.
- Brenner, M.H. (1975) Test difficulty, reliability, and discrimination as function of item difficulty order. Journal of Applied Psychology, 1964, 48, 98-100. (in Nelson J. T. and Merrill P.F. Effects of Anxiety Type and Item-Difficulty Sequencing on Mathematics Test Performance, Journal of Educational Measurement, 12, 241-249.
- Brown, J.M., & Weiss, D.J. (1977). An adaptive testing strategy for achievement test batteries (Research Rep. No. 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program (in Weiss D.J. Adaptive Testing by Computer, Journal of Consulting and Clinical Psychology, 1985, 53, 774-789.
- Bunderson, C.V., & al. The four generations of computerised educational measurement. In R.L. Linn (Ed.). Educational Measurement (3rd ed., pp. 367-408). New York; MacMillan (in Wise, S.L. & Plake B.S. Research on the Effect of Administering Tests via Computers, Educational measurement Issues & Practice, 1989, 8, 5-10.

- Galassi, J.P., Frierson, H.T. & Sharer, R. (1981) Behavior of High, Moderate, and Low Test Anxious Students During an Actual Test Situation, Journal of Consulting and Clinical Psychology, 49, 51-62.
- Gershon, R.C. (1989) Test Anxiety and Item Order New Parameters for Items Response Theory, Paper Presented at the Annual Meeting of The American Educational Research Association.
- Hambleton, R.K. (1970) Effects of Item Order and Anxiety on Test Performance and Stress, Paper presented to Division D, American Educational Research Association Annual Meeting. Chicago, 1968 (in Mon J.J. & Stallings W.M. Effects of Item Order on Test Scores, The Journal of Educational Research, 63, 463-465).
- Kent, T.H., & Albanese, M.A. (1987) A Comparison of the Relative Efficiency and Validity of Tailored Tests and Conventional Quizzes, Evaluation & Health Professions, 10, 67-79.
- Keppel, G. & Zedeck, S. (1989) Data Analysis for Research Designs, W.H. Freeman and Company, 159).
- Kestenbaum, J.M., & Weiner, B. (1970) Achievement performance related to achievement motivation and test anxiety. Journal of Consulting and Clinical Psychology, 34, 343-344 (in Nelson J.T. and Merrill P.F. Effects of Anxiety Type and Item-Difficulty Sequencing on Mathematics Test Performance, Journal of Educational Measurement, 1975, 12, 241-249).
- Kingsbury, G.G., & Weiss, D.J. (1981). A validity comparison of adaptive and conventional strategies for mastery testing (Research Report No. 81-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Method Program (in Weiss D.J. & Kingsbury G.G. Application of Computerized Adaptive Testing to Educational Problems, Journal of Educational Measurement, 1984, 21, 361-375).
- Krathwohl, D.R., & Huyser, R.J. (1956). The sequential item test (SIT) American Psychologist, 2, 419 (in Reckase M.D. Adaptive Testing: The Evolution of a Gook Idea, Educational Measurement Issues & Practice, 1989, 8, 11-15).
- Lord, F.M. (1971). The self-scoring flexilevel test. Journal of Educational Measurement, 8, 147-151.

- McBride, J.R. & Martin, J.T. (1983) Reliability and validity of adaptive ability tests in a military settings. In D.J. Weiss (Ed.), New horizons in testing: Latent trait test theory and computerized adaptive testing (pp. 223-236). New York: Academic Press).
- McKinley, R.L. & Reckase, M.D. (1980). Computer applications to ability testing. Association for Educational Data System Journal, 13, 193-203. (in Wise, S.L. & Plake B.S. Research in the Effect of Administrating Tests via Computers, Educational Measurement Issues & Practice, 1989, 8, 5-10).
- MacNicol, K. Effects of Varying Order of Item Difficulty in an Unspeeded Verbal Test, (1956) unpublished manuscript, Educational Testing Service, Princeton, New Jersey, (in Monk J.J. & Stallings W.M. Effects of Item Order on Test Scores, The Journal of Educational Research, 1970, 63, 463-465).
- Mandler, G. & Sarason, S.B. (1952) A Study of Anxiety and Learning, Journal of Abnormal and Social Psychology, 47, 166-173.
- Mann, J. (1972) Vicarious Desensitization of Test Anxiety Through Observation of Videotaped Treatment, Journal of Counselling Psychology, 19, 1-7.
- Martin, J.T., McBride, J.S. & Weiss, D.J. (1983). Reliability and validity of adaptive and conventional test in a military recruit population (Research Rep., No. 83-1) Minneapolis: University of Minnesota, Department of Psychology, Psychometrics Methods Program, Computerized (in Weiss D.J. & Kingsbury G.G., Application of Computerized Adaptive Testing to Educational Problems, Journal of Educational Measurement, 1984, 21, 361-375.
- Maurelli, V.A. & Weiss D.J. (1981). Factors influencing the psychometric characteristics of an adaptive testing strategy for test batteries (Research Rep. No. 81-4). Minneapolis: University of Minnesota, Department of Psychology, Computerized Adaptive Testing Laboratory. (in Weiss D.J. Adaptive Testing by Computer, Journal of Consulting and Clinical Psychology, 1985, 53, 774-789.)

- Mazzeo, J. & Harvey, A.L. (1988) The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature (College Board Rep. No. 88-8, EIS RR No. 88-21). Princeton, NJ: Educational Testing Service (in Wise, S.L. & Plake B.S. Research in the Effect of Administrating Tests via Computers, Educational Measurement Issues & Practice, 1989, 8, 5-10).
- Mollenkopf, W.G. (1950). An experimental study of the effects on item analysis data of changing item placement and test-time limit: Psychometrika, 15 291-315.
- Monk, J.J. & Stallings, (1970) W.M. Effects of Item Order on Test Scores, The Journal of Educational Research, 63, 463-465.
- Munz, D.C. & Smouse, A.D. (1968) Interaction effects of item-difficulty sequence and achievement - anxiety reaction on academic performance, Journal of Educational Psychology, 59, 370-374. (in Nelson J.T. and Merrill P.F. Effects of Anxiety Type and Item-Difficulty Sequencing on Mathematics Test Performance, Journal of Educational Measurement, 1975, 12, 241-249.
- Nelson, J.T. & Merrill, P.F. (1975) Effects of Anxiety Type and Item-Difficulty Sequencing on Mathematics Test Performance, Journal of Educational Measurement, 12, 241-249.
- Olsen, J.B. & al. (1986). Comparison and equating of paper-administered, computer-administered, and computerized adaptive tests of achievement. Paper presented at the meeting of the American Educational Paper esearch Association, San Francisco. (in Wise, S.L. & Plake B.S. Research in the Effect of Administrating Tests via Computers, Educational Measurement Issues & Practice, 1989, 8, 5-10).
- Reckase, M.D. (1989) Adaptative Testing: The Evolution of a Good Idea, Educational Measurement Issues & Practice, 8, 11-15.
- Roid, G.H. (199) Item Writing and Item Banking by Microcomputer: An Update Educational Measurement Issues & Practice, 8, 33-38.

- Ross, C.C. & Stanley, J.C. (1954) Measurement in today's schools, Englewood Cliffs: Prentice-Hall Inc. (in Nelson J.T. & Merrill P.F. Effects of Anxiety Type and Item-Difficulty Sequencing on Mathematics Test Performance, Journal of Educational Measurement, 1975, 12, 241-249.)
- Sax, G. & Carr, A. (1962). An investigation of response sets on altered parallel forms. Educational and Psychological Measurement, 22, 371-376. (in Wainer H. & Kiely G.L. Item clusters and Computerized Adaptive Testing: A Case for Testlets, Journal fo Educational Measurement, 1987, 24, 185-201.)
- Sax, G. & Cromack, T.R. (1966) The effects of various forms of item arrangements on test performance, Journal of Educational Measurement, 3, 309-311. (in Nelson, J.T. & Merrill, P.F. Effects of Anxiety Type and Item-Difficulty Sequencing on Mathematics Test Performance, Journal of Educational Measurement, 1975, 92, 241-249.)
- Sneider, H.G. & Ferrante, A.P. (1983) Luck and Learning: Feedback Contingencies and Initial Success in Verbal Discrimination Learning, The Journal of Psychology, 113, 53-59.
- Stone, C.A. (1989) Testing Software Review: MicroCAT Version 3.0, Educational Measurement Issues & Practice, 8, 33-38.
- Sweeney, D.J., Smouse, A.D., Rupiper, O. & Mung, D.C. (1970) A test of the inverted-U. hypothesis relating achievement anxiety and academic test performance. The Journal of Psychology, 74, 267-273.
- Thompson, J.G. & Weiss, D.J. (1980). Criterion-related validity of adoptive testing strategies (Research Rep. No 80-3) Minneapolis: University of Minnesota, Department of Psychology, Computerised Adaptive Testing Laboratory (in Weiss, D.J. Adaptive Testing by Computer, Journal of Consulting and Clinical Psychology, 1985, 53, 774-789).
- Thorndike, R.C. & Hagen, E. Measurement and Evaluation in Psychology and Education, (2nd ed.) New York: John Wiley & Sons, 1961 (in Nelson, J.T. & Merrill P.F. Effects of Anxiety Type and Item-Difficulty Sequencing on Mathematics Test Performance, Journal of Educational Measurement, 1975, 12, 241-249).

- Towle, N.J. & Merrill, P.F. (1975). Effects of anxiety type and item difficulty sequencing on mathematics test performance. Journal of Educational Measurement, 12, 241-249. (in Wainer, H. & Kiely, G.L., Item Clusters and Computerized Adaptive Testing: A Case for Testlets, Journal of Educational Measurement, 1987, 24, 185-201.
- Vale, C.D. & Weiss, D.J. (1975). A study of computer-administered stradaptive ability testing (Research Rep. No. 75-4). Minneapolis. University of Minnesota, Department of Psychology, Psychometric Methods Program, (in Weiss, D.J. Adaptive Testing by Computer, Journal of Consulting and Clinical Psychology, 1985, 53, 774-789).
- Wainer, H. & Kiely, G.L. (1987) Item Clusters and Computerized Adaptive Testing; A Case for Testlets, Journal of Educational Measurement, 24, 185-201.
- Weiss, D.J. (1985) Adaptive Testing by Computer. Journal of Consulting and Clinical Psychology, 53, 774-789.
- Weiss, D.J., (1982) Improving Measurement Quality and Efficiency with Adaptive Testing, Applied Psychological measurement, 6, 473-492.
- Weiss, D.J. & Kingsbury, G.G. (1984) Application fo Computerized Adaptive Testing to Educational Problems, Journal of Educational Measurement, 21, 361-375.
- Wise, S.L. & Plake, B.S. (1981) Research on the Effect of Administrating Tests via Computers, Educational Measurement Issues & Practice, 8, 5-10.

APPENDICE A

RESULTATS BRUTS DES 6 TESTS

Résultats bruts des six versions du Test sur les verres de jus
(n = 90)

	Tests					
	1	2	3	4	5	6
	6	11	7	4	8	8
	12	9	6	6	7	5
	4	12	8	5	4	2
	4	6	7	4	8	1
	10	5	5	6	4	5
	8	6	10	10	6	6
	6	6	4	8	9	7
	7	7	6	6	1	6
	4	8	7	8	2	3
	2	3	4	5	5	4
	8	6	4	4	5	5
	7	6	9	5	3	4
	6	12	4	7	5	4
	5	8	4	7	8	4
	12	3	7	6	3	6
Total:	101,00	108,00	92,00	91,00	78,00	70,00
Moyenne:	6,73	7,20	6,13	6,06	5,20	4,67
Variance:	8,53	8,03	3,86	2,92	5,89	3,38

APPENDICE B

TEST DU T.A.R.A. (SECTION n2 & n3)

N.C.M.:

Eude:

Date de naissance:

NOM DU TEST:

Encerle la bonne réponse.

a) $\frac{3}{4} - \frac{1}{4}$

- A. $\frac{3}{4}$
 B. $\frac{3}{2}$
 C. $\frac{5}{8}$
 D. $\frac{1}{8}$
 E. aucune

b) $\frac{1}{3} + \frac{2}{3}$

- A. $\frac{7}{6}$
 B. $\frac{4}{3}$
 C. $\frac{5}{6}$
 D. $\frac{8}{6}$
 E. aucune

c) $\frac{2}{3} + \frac{1}{3}$

- A. $\frac{25}{21}$
 B. $\frac{20}{21}$
 C. $\frac{21}{28}$
 D. $\frac{23}{21}$
 E. aucune

d) $\frac{1}{2} - \frac{2}{3}$

- A. $\frac{15}{35}$
 B. $\frac{17}{35}$
 C. $\frac{11}{35}$
 D. $\frac{10}{35}$
 E. aucune

e) $\frac{1}{3} \div \frac{1}{4}$

- A. $\frac{1}{12}$
 B. $\frac{2}{3}$
 C. $\frac{4}{3}$
 D. $1\frac{1}{3}$
 E. aucune

f) $\frac{2}{3} \div 5$

- A. $\frac{10}{7}$
 B. $\frac{1}{35}$
 C. $\frac{2}{35}$
 D. $\frac{3}{35}$
 E. aucune

g) $2\frac{1}{2} \times 1\frac{1}{2}$

- A. $\frac{38}{9}$
 B. 5
 C. $\frac{37}{9}$
 D. $\frac{35}{9}$
 E. aucune

h) $2\frac{3}{4} - 1\frac{1}{4}$

- A. $\frac{30}{28}$
 B. $\frac{27}{28}$
 C. $\frac{13}{8}$
 D. $\frac{20}{28}$
 E. aucune

i) $\frac{3}{4} + \frac{2}{3} + 1\frac{1}{4}$

- A. $3\frac{1}{4}$
 B. $2\frac{3}{4}$
 C. $1\frac{3}{4}$
 D. $2\frac{1}{4}$
 E. aucune

j) $160\% \div 20\%$

- A. $7\frac{1}{54}$
 B. $7\frac{51}{45}$
 C. $7\frac{45}{51}$
 D. $7\frac{51}{54}$
 E. aucune

k) $(\frac{3}{4})^2$

- A. $\frac{3}{18}$
 B. $\frac{2}{7}$
 C. $\frac{9}{14}$
 D. $\frac{9}{49}$
 E. aucune

l) $2\frac{3}{4} \div 5\frac{3}{4}$

- A. $\frac{3}{4}$
 B. $\frac{3070}{10704}$
 C. $\frac{5077}{10708}$
 D. $\frac{5075}{10702}$
 E. aucune

APPENDICE C

TEST SUR LES VERRES DE JUS DE NOELTING

(VERSION ORIGINALE)

CONCENTRATIONS / ORANGE JUICE

Date: _____

Nom / Name: _____

Anniversaire de naissance /

Age: _____

Date of birth: _____

Ecole / School: _____

Classe / Class: _____

I

A

B

+

=

+

?

II

A

B

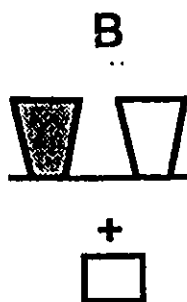
+

=

+

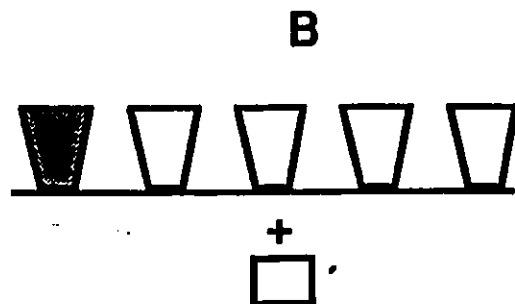
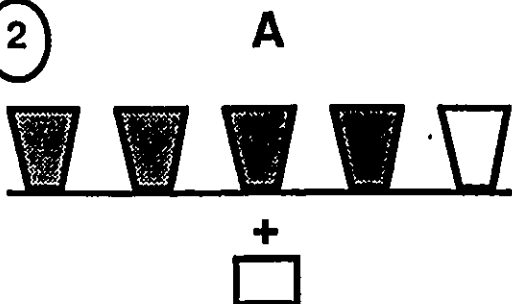
?

1



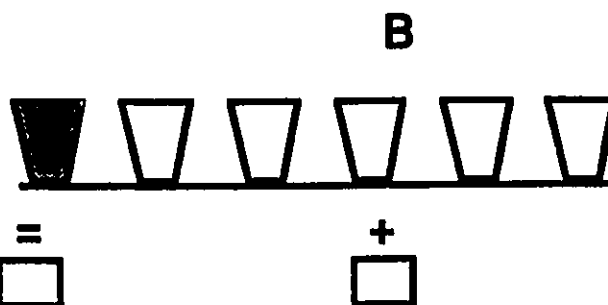
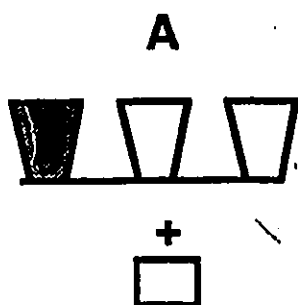
?

2



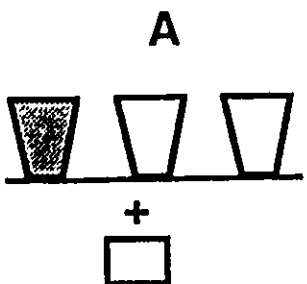
?

3

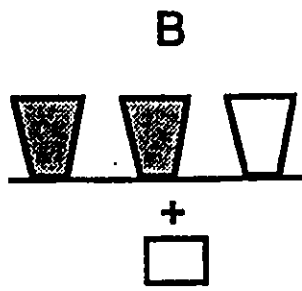


?

4

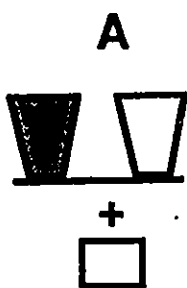


=

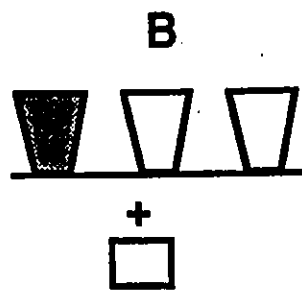


?

5

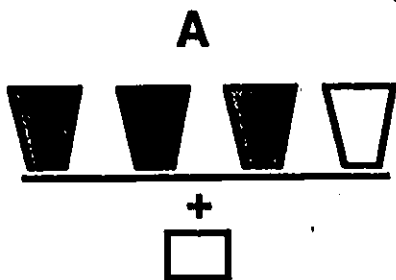


=

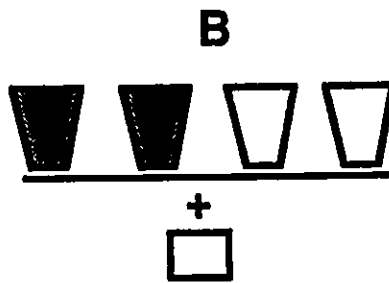


?

6




=




?

7

A



B



+


=

+


?

8

A



B



+


=

+


?

9

A



B



+

=

+

?

10

A B

+

=

+

?

11

A B

+

=

+

?

12

A B

+

=

+

?

13

A B

\square $=$ \square

?

14

A B

\square $=$ \square

?

15

A B

\square $=$ \square

?

16

A B

+

+

=

?

17

A B

+

+

=

?

18

A B

+

+

=

?

19

A



+



B



=



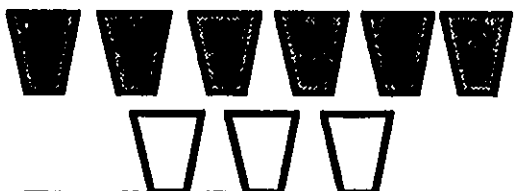
+



?

20

A



+



B



=




+



?

(21)

A B




+ = +

□ □ □

?

(22)

A B

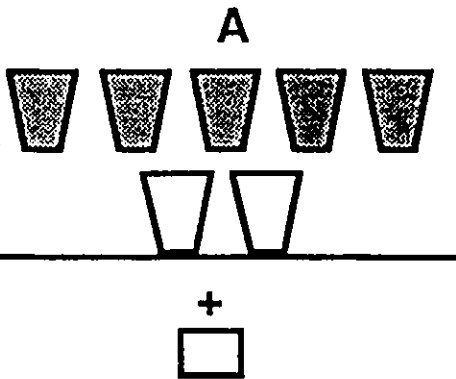


+ = +

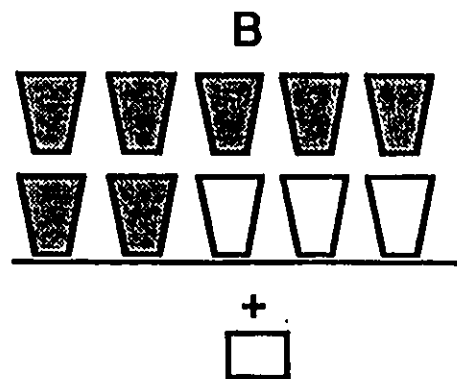
□ □ □

?

23

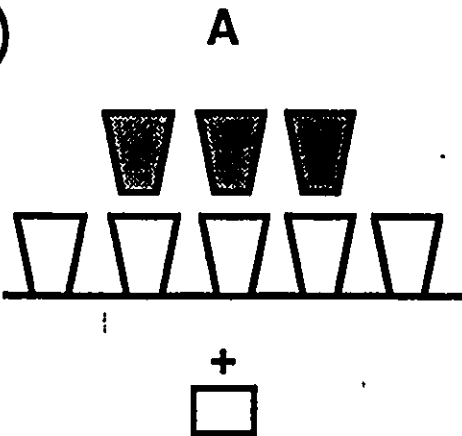


=

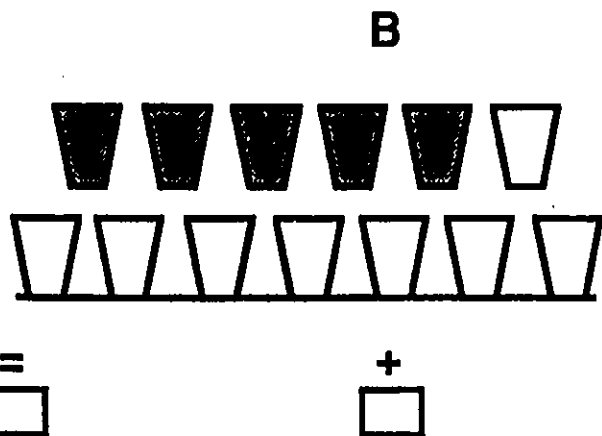


?

24



=

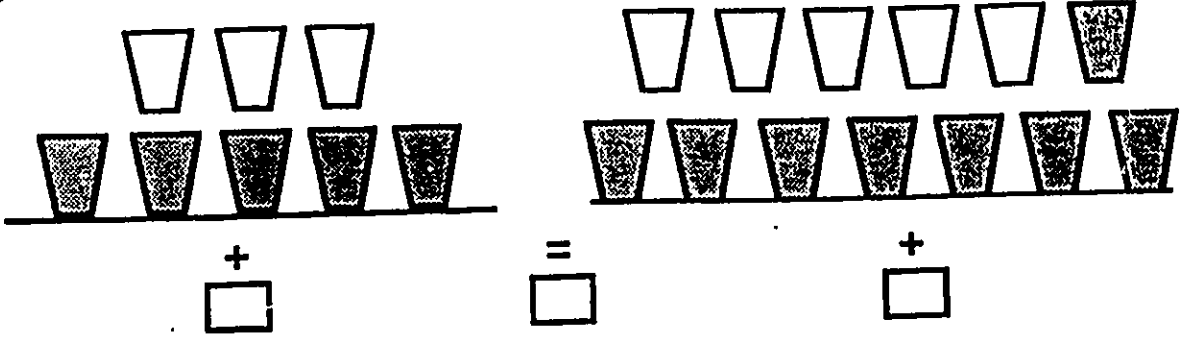


?

25

A

B



?
