



National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche 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.

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

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

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

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

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

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 mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

L'ORDRE DE DIFFICULTE CROISSANTE DES ITEM DANS UN TEST DE
PUISSANCE A TEMPS LIMITE, L'EXAMEN INTERMEDIAIRE OTIS-OTTAWA
D'HABILETE MENTALE: SA CONSTANCE ET SES EFFETS SUR LA
VALEUR DISCRIMINANTE DES ITEM

par André R. Pellerin

Thèse présentée à l'Ecole des études supérieures en vue de
l'obtention du Ph.D. en psychologie clinique

Université d'Ottawa
Canada, 1974

© A.R. Pellerin, Ottawa, Canada, 1978

RECONNAISSANCE

Cette thèse a été préparée sous la direction de Laurent Isabelle, Ph.D., président du Collège Algonquin, Ottawa, et ancien professeur titulaire à la Faculté de Psychologie de l'Université d'Ottawa. Le professeur Isabelle fut secondé par Henry P. Edwards, Ph.D., professeur agrégé et maintenant doyen de la Faculté de Psychologie de l'Université d'Ottawa, et par Charles E. McInnis, Ph.D., professeur adjoint à la même faculté.

Le conseil des Arts du Canada a favorisé cette recherche en attribuant une bourse d'études à son auteur.

L'élaboration du schème expérimental et la cueillette des données ont été réalisées avec la collaboration de Jean-Guy Bonin, M.A. (Ps), qui a complété sa thèse de maîtrise à partir d'une expérience parallèle à celle-ci.

La conduite de l'expérience a été possible grâce à l'autorisation des directeurs des commissions scolaires concernées et à la collaboration de Robert Pelletier, M.A., directeur des services d'Orientation à la Régionale de l'Outaouais, et de William Rodrigue, M.A. (Ps). Les étudiants de première année 1967-1968 et 1969-1970 des Facultés de Psychologie et d'Education ont été les principaux administrateurs des tests.

RECONNAISSANCE

iii

L'analyse des données a été grandement facilitée par le travail de Roland Serrat, L.Sc., et de Robert Thomas, B.Sc., qui en ont fait la programmation pour l'ordinateur.

L'auteur remercie toutes les personnes qui ont été associées de près ou de loin à la réalisation de cette recherche.

CURRICULUM STUDIORUM

André R. Pellerin naquit à Asbestos, Province de Québec, le 1er février 1946. Il obtint son B.A. de l'Université Laval en 1967 et son M.Ps. de l'Université d'Ottawa en 1972.

TABLE DES MATIERES

Chapitres	pages
INTRODUCTION	xii
I.- PRESENTATION DU PROBLEME	1
1. La présentation de l'instrument	1
a. La version originale: <u>Otis Self-Administering Test of Mental Ability</u>	2
b. La traduction: <u>les Examens Otis-Ottawa d'habileté mentale</u>	6
c. <u>L'ordre de difficulté des item des Examens Otis-Ottawa d'habileté mentale.</u>	8
2. Recension des écrits, première partie: constance de l'ordre de difficulté des item	11
a. La mesure de la difficulté des item d'un test à temps limité	12
b. Recension des recherches sur la mesure de la difficulté des item	16
c. Hypothèses de recherche	27
3. Recension des écrits, deuxième partie: effets de l'ordre de difficulté	29
a. Importance de l'ordre de difficulté des item dans les tests de rendement	30
b. Importance de l'ordre de difficulté des item dans les tests d'aptitude	34
c. Recension des recherches sur l'ordre de difficulté des item	37
d. Hypothèses de recherche	56
II.- SCHEME EXPERIMENTAL	59
1. Les échantillons	59
2. Les méthodes statistiques	70
a. Le calcul de l'ordre de difficulté des item	71
b. La mesure de la valeur discriminante des item	73
III.- PRESENTATION ET INTERPRETATION DES RESULTATS . .	94
1. La constance de l'ordre de difficulté croissante des item	94
2. Les effets de l'ordre de difficulté des item sur la valeur discriminante des item	100
3. La discussion des résultats	108
CONCLUSIONS	117
BIBLIOGRAPHIE	120

TABLE DES MATIERES

vi

Appendices

pages

1.	LES STATISTIQUES QUE LES CINQ ECHANTILLONS D'ELEVES DE LA REGIONALE DE L'OUTAOUAIS ONT OBTENUES SUR LES FORMULES A, EXPERIMENTALE I ET EXPERIMENTALE II DE L'EXAMEN INTERMEDIAIRE OTIS-OTTAWA ADMINISTREES EN TRENTE ET EN SOIXANTE MINUTES	127
2.	POLYONES DE FREQUENCES DES SCORES BRUTS QUE LES CINQ ECHANTILLONS D'ELEVES DE LA REGIONALE DE L'OUTAOUAIS ONT OBTENUS SUR LES FORMULES A, EXPERIMENTALE I OU EXPERIMENTALE II DE L'EXAMEN INTERMEDIAIRE OTIS-OTTAWA ADMINISTREES EN TRENTE MINUTES.	130
3.	EXEMPLAIRES DES FORMULES A, EXPERIMENTALE I ET EXPERIMENTALE II DE L'EXAMEN INTERMEDIAIRE OTIS-OTTAWA D'HABILETE MENTALE	136
4.	STATISTIQUES QUE LES ECHANTILLONS III, IV ET V ONT OBTENUS SUR LE <u>CULTURE FAIR, SCALE 2, FORM A</u>	149
5.	ANALYSE DE LA DIFFICULTE DES ITEM: PROPORTIONS CORRIGÉES DE SUCCES SUR LES ITEM, RANGS DES ITEM ET COEFFICIENTS DE CORRELATION RHO ENTRE TOUS LES ORDRES DE DIFFICULTE DES ITEM.	162
6.	ANALYSE DE LA VALEUR DISCRIMINANTE DES ITEM: RESULTATS DE L' <u>ANALYSE PROBIT</u>	176
7.	RESUME DE <u>Influences de l'ordre de difficulté croissante des item sur les qualités psychométriques d'un test d'intelligence du type puissance à temps limité</u> , par Jean-Guy Bonin	189
8.	ABSTRACT OF <u>L'ordre de difficulté croissante des item dans un test de puissance à temps limité, l'Examen Otis-Ottawa d'habileté mentale: sa constance et ses effets sur la valeur discriminante des item</u>	193

LISTE DES TABLEAUX

Tableaux	pages
I.- Moyennes des âges, moyennes et écarts-types des scores bruts de l'échantillon sélectionné et de la population des élèves de septième année de la Régionale de l'Outaouais en 1967-1968	61
II.- Proportions de succès et proportions de succès corrigées pour les réussites dues à la chance, sur les item de la formule A de l' <u>Examen intermédiaire Otis-Ottawa</u> administrée en soixante minutes à 258 élèves de sixième année de la Régionale de l'Outaouais	65
III.- Item de la formule A de l' <u>Examen intermédiaire Otis-Ottawa</u> replacés en ordre de difficulté croissante d'après la performance de 258 élèves de sixième année de la Régionale de l'Outaouais.	66
IV.- Rangs occupés par les item dans les formules A, expérimentale I et expérimentale II de l' <u>Examen intermédiaire Otis-Ottawa</u>	68
V.- <u>Analyse probit</u> de l'item 12 de la formule A à partir des données brutes du troisième échantillon.	79
VI.- Résultats de l' <u>analyse probit</u> de l'item 12 pour les groupes III, IV et V considérés un à un (hétérogénéité) ou tous ensemble (parallélisme)	89
VII.- Tableau de la variance pour les données de l' <u>analyse probit</u> de l'item 12 concernant les groupes III, IV et V	93
VIII.- Coefficients de corrélation <u>rho</u> entre les ordres de difficulté croissante des item, calculés à partir des administrations en trente minutes des formules A, expérimentale I et expérimentale II de l' <u>Examen intermédiaire Otis-Ottawa</u>	96

Tableaux

pages

- IX.- Coefficients de corrélation ρ entre les ordres de difficulté croissante des item, calculés à partir des administrations en soixante minutes des formules A, expérimentale I et expérimentale II de l'Examen intermédiaire Otis-Ottawa. 98
- X.- Statistiques que les cinq échantillons d'élèves ont obtenues sur les formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes 128
- XI.- Statistiques que les cinq échantillons d'élèves ont obtenues sur les formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en soixante minutes 129
- XII.- Statistiques des échantillons III, IV et V sur le Culture Fair, Scale 2, Form A 150
- XIII.- Description statistique de l'équivalence des échantillons III, IV et V sur le Culture Fair, Scale 2, Form A. 151
- XIV.- Proportions de succès, corrigées pour les réussites dues à la chance, obtenues sur les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes aux cinq échantillons. 153
- XV.- Rangs occupés par les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes, lorsqu'ils sont placés en ordre de difficulté croissant selon la proportion corrigée de succès obtenu par les cinq échantillons. 156
- XVI.- Proportions de succès, corrigées pour les réussites dues à la chance, obtenues sur les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en soixante minutes aux quatre derniers échantillons 159

LISTE DES TABLEAUX

ix

Tableaux	pages
XVII.- Rangs occupés par les item des formules A, expérimentale I ou expérimentale II de l' <u>Examen intermédiaire Otis-Ottawa</u> admistrées en soixante minutes, lorsqu'ils sont placés en ordre de difficulté croissante selon la proportion corrigée de succès obtenu par les quatre derniers échantillons	162
XVIII.- Coefficients de corrélation <u>rho</u> entre tous les ordres de difficulté croissante et les ordres de présentation des item des formules A et expérimentale II	165
XIX.- Tests "chi deux" entre les scores probit empiriques et les scores probit théoriques de la fonction ogive normale, pour tous les item dans chacun des groupes. . .	167
XX.- <u>Analyse probit</u> : pentes des droites probit des item dans chacun des groupes	170
XXI.- <u>Analyse probit</u> : pour chacun des item, le nombre d'élèves des groupes I, II et III qui l'ont atteint, la pente commune aux trois groupes et le test F qui détermine si cette pente représente chaque groupe sans erreur significative	173
XXII.- <u>Analyse probit</u> : pour chacun des item, le nombre d'élèves des groupes III, IV et V qui l'ont atteint, la pente commune aux trois groupes et le test F qui détermine si cette pente représente chaque groupe sans erreur significative.	176

LISTE DES FIGURES

Figures	pages
1.- Probabilités de succès sur un item quelconque en fonction des scores bruts totaux, fonction ogive normale	76
2.- Scores probit en fonction des scores bruts totaux, fonction linéaire	78
3.- Proportions de succès obtenues sur l'item 12 par les élèves du troisième échantillon, en fonction de leurs scores bruts totaux. . .	80
4.- Scores probit obtenus sur l'item 12 par les élèves du troisième échantillon, en fonction de leurs scores bruts totaux	82
5.- Probabilités de succès sur un item quelconque en fonction des scores bruts totaux lorsque la moyenne de l'item est de 42 et son écart-type de 15	85
6.- Scores probit sur un item quelconque en fonction des scores bruts totaux lorsque la moyenne de l'item est de 42 et son écart-type de 15	86
7.- <u>Analyse probit</u> de l'item 12: droites probit individuelles des groupes III, IV et V	90
8.- <u>Analyse probit</u> de l'item 12: droites probit parallèles des groupes III, IV et V.	91
9.- Polygones des fréquences théoriques et observées des scores bruts du premier échantillon sur la formule A de l' <u>Examen intermédiaire Otis-Ottawa</u> administrée en trente minutes	131
10.- Polygones des fréquences théoriques et observées des scores bruts du deuxième échantillon sur la formule A de l' <u>Examen intermédiaire Otis-Ottawa</u> administrée en trente minutes	132

LISTE DES FIGURES

xi

Figures	pages
11.- Polygones des fréquences théoriques et observées des scores bruts du troisième échantillon sur la formule A de l' <u>Examen intermédiaire Otis-Ottawa</u> administrée en trente minutes	133
12.- Polygones des fréquences théoriques et observées des scores bruts du quatrième échantillon sur la formule expérimentale I de l' <u>Examen intermédiaire Otis-Ottawa</u> administrée en trente minutes.	134
13.- Polygones des fréquences théoriques et observées des scores bruts du cinquième échantillon sur la formule expérimentale II de l' <u>Examen intermédiaire Otis-Ottawa</u> administrée en trente minutes.	135

INTRODUCTION

Le domaine de la mesure en psychologie évolue rapidement à la suite de recherches les plus diverses. A mesure que les uns s'appliquent à clarifier les qualités de ce qu'ils voudraient mesurer, d'autres s'ingénient à développer des techniques raffinées qui s'adaptent à ces qualités et qui pourront en donner une appréciation de plus en plus juste.

Les techniques de mesure varient en complexité et, en général, elles doivent leur efficacité à leur cohérence interne. Ce raffinement comprend les détails les plus simples comme la bonne disposition des item sur une page jusqu'aux procédures complexes de la standardisation. La présente recherche s'intéresse à l'un de ces aspects qui pourraient influencer la cohérence d'un test de puissance à temps limité; il s'agit de l'ordre de présentation de ses item selon leur difficulté.

L'intérêt pour ce problème a d'abord été suscité par le fait que l'ordre de difficulté des item était lié explicitement à un test de puissance à temps limité en particulier, les Otis Self-Administering Tests of Mental Ability, ou leur version française, les Examens Otis-Ottawa d'habileté mentale. Cette version est reconnue pour sa grande valeur pratique. En effet, bien que les origines de l'Otis remontent à quelque cinquante années et que la conception de ce test ne reflète

pas la sophistication des dernières techniques en psychométrie, il n'en demeure pas moins que l'Otis-Ottawa est encore le test d'intelligence le plus utilisé au Canada français et que son efficacité n'a pas été égalée par un instrument plus moderne.

Dans un premier effort pour situer le problème, une étude préliminaire avait été entreprise dans les cadres du cours Ps 4409 La mesure des aptitudes (1967-1968). Cette analyse de la formule A de l'Examen intermédiaire Otis-Ottawa avait montré que les item n'étaient pas en ordre de difficulté croissante comme l'avait voulu Otis dans la conception du test américain original.

Même si cette analyse a remis en question l'existence de l'ordre de difficulté croissante des item de ce test, elle a surtout débouché sur les deux problèmes théoriques qui sont le centre de la présente recherche. Le premier problème peut être formulé comme suit: à partir d'un test de puissance à temps limité comme l'Otis-Ottawa, est-il possible de calculer et d'établir un ordre de difficulté croissante des item qui soit constant tout en respectant la limite de temps imposée par la standardisation? En d'autres mots, la difficulté relative d'un item, c'est-à-dire le rang de l'item dans l'échelle de difficulté de l'ensemble des item, est-elle influencée par la limite de temps imposée par la standardisation et/ou par la position que l'item occupe dans le test?

Ceci entraîne le deuxième problème. S'il est possible d'établir un ordre de difficulté croissante, quels seront ses effets sur la valeur discriminante des item? Favorisera-t-il la valeur discriminante des item plus que ne le fait l'ordre actuel des item ou plus que ne le ferait un ordre au hasard?

Ces deux problèmes forment le coeur de la présente expérience et ils seront élaborés dans le premier chapitre à travers la présentation de l'instrument et la recension des écrits. Le deuxième chapitre décrira les échantillons, les procédures expérimentales ainsi que les méthodes statistiques utilisées dans l'expérience. Le troisième chapitre présentera les résultats obtenus et l'analyse de ces résultats. Finalement, une conclusion résumera l'expérience et en fera le point.

CHAPITRE PREMIER

PRESENTATION DU PROBLEME

Le problème de l'ordre de difficulté des item peut se retrouver dans une multitude de tests avec une acuité plus ou moins grande. Dans la présente étude, il est mis en relief particulièrement parce qu'il est lié aux caractéristiques et à la conception même de l'instrument utilisé, la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale. Ce problème sera d'abord situé, à travers la présentation de l'instrument, dans les cadres qui l'entourent et qui lui donnent en plus d'une valeur théorique, une importance pratique liée à l'utilisation répandue du test. Les deux aspects théoriques de ce problème seront ensuite étudiés à travers la recension des écrits.

1. La présentation de l'instrument.

Dans la brève histoire de la psychologie moderne, l'évolution a été accélérée. Les théories et les techniques se sont succédées rapidement, les unes plus durables ou plus éphémères que les autres. Ce phénomène se reflète aussi en psychométrie où plusieurs tests sont apparus pour disparaître quelque temps plus tard alors que d'autres parviennent à survivre sans doute pour de multiples raisons. Les Examens Otis-Ottawa d'habileté mentale, bien qu'issus de la première

génération des tests de groupe, demeurent très actuels parce qu'ils présentent des qualités pratiques que les nouvelles générations de tests n'ont pu supplanter. Ce test qui est utilisé depuis plusieurs années sera d'abord présenté ici dans sa version originale, les Otis Self-Administering Tests of Mental Ability, et ensuite dans sa version traduite et adaptée au Canada français, les Examens Otis-Ottawa d'habileté mentale. Finalement, le problème de l'ordre de difficulté des items de ce test sera résumé et discuté.

- a. La version originale: Otis Self-Administering Tests of Mental Ability.

La construction des Otis Self-Administering Tests of Mental Ability a été guidée par les besoins d'une mesure plus précise de l'habileté à apprendre. Otis définit cette habileté mentale selon des qualités mesurables comme suit:

l'habileté mentale est

that quality of the mind which enables a person to apply the knowledge he possesses to the solution of new problems. It is a quality which possesses magnitude in the sense that a child as he grows older shows the ability to apply his knowledge to the solution of more 'difficult' problems. ... Thus, if we draw up a series of questions and problems (let us call them items) ranging from easy to difficult, we find that a child as he grows older can answer or solve more and more of these items. This fact gives us a convenient method of measuring mental ability¹.

¹ Arthur S. Otis, Statistical Method in Educational Measurement, New York, World Book, 1925, p. 136.

PRESENTATION DU PROBLEME

3

Selon cette définition, où le développement de l'habileté mentale est lié au développement chronologique de l'individu, le fait de placer les item en ordre de difficulté croissante est essentiel à l'instrument de mesure. En effet chaque niveau de difficulté représente un stade dans le développement de l'habileté mentale et il serait inutile et illogique de présenter à l'enfant des item d'un stade trop avancé. De plus, il importe que l'élève ne soit pas retardé par des item trop difficiles parce que sa performance est évaluée d'après le nombre d'item qu'il a réussis. Ce score correspond au score moyen d'un groupe d'élèves d'un âge chronologique donné et indique ainsi l'âge mental de l'élève. En divisant cet âge mental par l'âge chronologique de l'élève, Otis obtient un quotient qu'il appelle "degree of brightness"².

Otis construisit ses Self-Administering Tests of Mental Ability en suivant la définition précitée. Il construisit une quantité d'item divers, analogues à ceux que l'élève rencontre dans ses cours. Il inclut des item de connaissance générale, des analogies, des problèmes d'arithmétique, des inférences logiques, etc.. Comme il voulait appliquer son test à différents niveaux scolaires, il

2 A.S. Otis, op. cit., p. 148.

PRESENTATION DU PROBLEME

4

construisit des item de difficulté variée pour que le test discrimine plusieurs niveaux d'habileté, qu'il soit équidiscriminant.

Il administra cette banque d'item à des élèves et il ne conserva que les item qui pouvaient différencier à l'intérieur d'un même niveau académique les élèves brillants, les plus jeunes, des élèves lents, les plus âgés. Dans la présentation finale, il plaça les item en ordre de difficulté croissante et fit suivre un item d'une catégorie par un item d'une autre catégorie et ainsi de suite pour former un test du type spiral-omnibus. De plus en imposant une limite de temps dans l'administration du test, il en fit un test de puissance à temps limité.

La méthode de sélection des item selon leur capacité de discriminer les étudiants jeunes des étudiants plus âgés est l'une des plus grandes faiblesses du test Otis, sinon la plus grande. En construisant son test autour du critère externe très complexe qu'est le succès scolaire, Otis lui donna une validité très vaste et diffuse. Il réunit plusieurs item reliés à différents aspects de la pensée académique et il ne se préoccupa point de leur unité, si ce n'est qu'il retint uniquement ceux qui étaient en accord avec le critère. Les différents types d'item de l'Otis montrent bien une validité diffuse, mais leur plus ou moins grande indépendance les uns des autres peut aussi diminuer la constance du test.

Aujourd'hui, la tendance des constructeurs de tests est de suivre un procédé de construction qui est à l'opposé de celui d'Otis; en faisant une analyse factorielle des item, ils s'assurent d'abord une validité interne suffisante, et par ce fait une constance plus élevée, et ils ne se préoccupent de la validité externe qu'en second lieu. Otis s'est préoccupé uniquement de la validité externe et a laissé la validité interne en suspend.

Pour ajouter à cette ambiguïté interne de l'Otis, les item varient de très faciles à très difficiles, le nombre de choix possibles n'est pas constant d'un item à l'autre et le test doit être administré pendant une période de temps limitée. Toutes ces variables font de l'Otis un test très complexe, difficile à définir et lui enlève ainsi de la valeur sur le plan théorique. Cependant ceci n'enlève rien à son utilité dans une situation scolaire normale et ce point à lui seul était assez important pour motiver Shevenell³ à traduire les Otis Self-Administering Tests of Mental Ability. En plus, l'urgence d'une solution à la mesure de l'habileté mentale se faisait aussi sentir au Canada français.

3 R.-H. Shevenell, Les examens Otis-Ottawa d'habileté mentale, Ottawa, Les Editions de l'Université d'Ottawa, 1948, 40 p.

b. La traduction: Les examens Otis-Ottawa d'habileté mentale.

La plus grande partie des tests psychométriques utilisés au Canada français ont été traduits de tests américains ou canadiens anglais. Les uns ne sont que le résultat d'une simple traduction et doivent emprunter les normes de l'instrument original. D'autres, en plus de cette traduction, offrent des normes établies à partir d'un échantillon de la population francophone à laquelle ils s'adressent. Les examens Otis-Ottawa d'habileté mentale appartiennent à cette dernière catégorie. Ils ont été traduits des Otis Self-Administering Tests of Mental Ability en 1942 et standardisés sur un échantillon de francophones de la région de l'Outaouais. Tout comme Otis l'avait fait dans son milieu, Shevenell procurait à la population scolaire francophone un test de scolapitude qui permettait de mieux planifier l'éducation en fonction des aptitudes intellectuelles des écoliers. Cet instrument fut le premier du genre au Canada français et bien que d'autres tests soient venus s'ajouter, il demeure très en vogue. L'utilisation de la formule A de l'examen intermédiaire s'est accrue constamment depuis sa parution au début des années quarante. De quelque 10,000 exemplaires par année, son tirage passa à 40,000 exemplaires en 1960 et depuis 1965 les Editions de l'Université d'Ottawa en distribuent environ 100,000 exemplaires par année. Cette formule est encore utilisée de

façon systématique comme test de classement à l'entrée des étudiants au niveau secondaire dans certaines régionales du Québec.

Par analogie au critère de validité dont s'était servi Otis, Shevenell utilisa le rang des élèves dans leur classe pour valider son test. Il obtint des coefficients de corrélation variant de .40 à .60, ce qui serait sensiblement du même ordre que ceux qu'avait obtenus Otis⁴. Il calcula la constance en mettant en corrélation des formules équivalentes du test et obtint des coefficients variant de .82 à .89. Ces coefficients transformés selon la formule Brown-Spearman varient de .91 à .94. Il qualifie ces résultats comme suit: "De pareils coefficients témoignent d'un fort degré de précision dans la mesure de l'habileté mentale telle que l'atteignent les tests Otis-Ottawa⁵."

Par ailleurs, Shevenell a souligné ses intentions de retenir le plus possible les qualités du test original:

Dans notre traduction du test, nous avons tenté de suivre son procédé et d'établir une équivalence entre les formules anglaises et françaises. Nos efforts ont été plus ou moins heureux; certains items de notre traduction satisfont pleinement au critère de validité; d'autres, moins. [...]
D'autre part, il est fort difficile, sinon presque impossible, de tenter d'obtenir l'équivalence avec les tests anglais en même temps que la validité des items français⁶.

4 R.-H. Shevenell, op. cit., p. 22.

5 R.-H. Shevenell, op. cit., p. 19.

6 R.-H. Shevenell, op. cit., p. 22.

Selon cette affirmation, il aurait été dans les intentions du traducteur de présenter les item en ordre de difficulté croissante comme l'avait fait Otis: "The items in each form of each examination have been arranged in the order of difficulty⁷." Cependant Shevenell ne mentionne aucune analyse d'item à ce sujet.

c. L'ordre de difficulté des item des Examens Otis-Ottawa.

A la suite de l'intention implicite du traducteur de présenter les item en ordre de difficulté croissante, il devenait intéressant d'analyser les item pour découvrir quelque trente ans après la traduction de l'instrument quel serait leur ordre de difficulté. En effet quel que soit l'ordre de difficulté des item établi au début des années quarante, il aurait pu être bouleversé par les effets de l'évolution sociale, par les changements de méthodes d'enseignement, etc..

Il semblait donc opportun d'analyser l'Otis-Ottawa sous ce point, de le remettre à jour s'il y avait lieu et de déterminer en même temps quels seraient les effets d'un tel changement sur les qualités discriminantes des item. Déjà

⁷ Arthur S. Otis, Otis Self-Administering Tests of Mental Ability, New York, Harcourt, Brace & World, 1928, p. 3.

Chapanis⁸, Crooks et Ferguson⁹, et Hovland et Wonderlic¹⁰ avaient étudié les formules supérieures de l'Otis original et y avaient découvert des erreurs importantes dans l'ordre de difficulté des item.

Une autre recherche conduite par Bonin et Pellerin en 1968, à partir d'un échantillon de cent garçons et de cent filles tirés au hasard de la population des élèves de septième année de la Régionale de l'Outaouais, parvint à la conclusion suivante: "Les résultats obtenus montrent un ordre progressif de difficulté d'allure très générale pour l'ensemble du test, mais nient l'existence de cet ordre progressif d'un item à l'autre¹¹".

⁸ Alphonse Chapanis, A Note on the Validity and Difficulty of Items in Form A of the Otis Self-Administering Tests of Mental Ability, dans The Journal of Experimental Education, vol. 5, n° 3, livraison de mars 1937, p. 246-248.

⁹ William R. Crooks et Leonard W. Ferguson, Item Validities of the Otis Self-Administering Tests of Mental Ability for a College Population, dans The Journal of Experimental Education, vol. 9, n° 3, livraison de mars 1941, p. 229-232.

¹⁰ Carl Iver Hovland et E.F. Wonderlic, A Critical Analysis of the Otis Self-Administering Tests of Mental Ability, Higher Form, dans The Journal of Applied Psychology, vol. 23, n° 3, livraison de juin 1939, p. 367-387.

¹¹ Jean-Guy Bonin et André Pellerin, Ordre et degré de difficulté des item et valeur discriminante des item des Examens Otis-Ottawa d'habileté mentale, examen intermédiaire, formule A, étude (inérite) présentée à la Faculté de Psychologie de l'Université d'Ottawa en mars 1968 et présentée au Congrès de l'ACFAS en novembre 1968, Ottawa, vii-50 p.

Cette étude préliminaire stimula l'intérêt vers une recherche plus approfondie. D'autant plus que si l'Otis est comparé aux autres tests qui ont servi à l'étude de l'ordre de difficulté des item, il apparaît comme un instrument où la variable ordre de difficulté est particulièrement importante. En effet le grand nombre d'item (75), l'étendue de la difficulté des item et la limite de temps imposée dans l'administration du test sont trois facteurs qui sembleraient multiplier les effets de l'ordre de difficulté des item.

Les deux auteurs de cette recherche préliminaire ont donc choisi de poursuivre plus loin en se divisant la tâche. Bonin¹² étudia les influences de l'ordre de difficulté des item sur la moyenne des scores bruts, sur l'écart-type des scores bruts, sur la constance et sur la validité externe. Un résumé de cette recherche est présenté à l'Appendice 7.

L'auteur de la présente recherche choisit d'étudier la constance de l'ordre de difficulté des item, en utilisant l'Examen intermédiaire Otis-Ottawa d'habileté mentale, et les influences de cet ordre sur la valeur discriminante des item. Ces deux problèmes qui soulèvent des difficultés

12 Jean-Guy Bonin, Influences de l'ordre de difficulté croissante des item sur les qualités psychométriques d'un test d'intelligence du type puissance à temps limité, thèse de maîtrise (inédite) présentée à la Faculté de Psychologie de l'Université d'Ottawa, Ontario, 1971, xv-166 p.

théoriques particulières seront discutés dans les deux sections qui suivent.

2. Recension des écrits, première partie:
constance de l'ordre de difficulté des
item.

Le problème de la présentation des item d'un test selon un ordre de difficulté est très spécifique et il y a relativement peu de rapports de recherche qui ont été publiés à ce sujet. De plus, lorsque ce problème est situé dans les cadres d'un test de puissance à temps limité comme l'Otis, une difficulté additionnelle le rend encore plus exceptionnel. En effet, ce genre de test ne permet pas à tous les élèves d'atteindre tous les item et ainsi les item de la fin ne sont pas mesurés d'après exactement le même échantillon que les item du début. Les indices de difficulté risquent donc de varier d'un échantillon à un autre, selon la place que l'item occupe dans le test, et de compromettre ainsi le calcul d'un ordre constant de difficulté croissante.

Dans la recherche des écrits pertinents, on ne trouve qu'un auteur qui a mis en évidence expérimentalement les effets conjoints de l'ordre de difficulté des item et de la période de temps allouée pour passer un test sur la mesure de la difficulté des item. D'autres auteurs ont cependant lancé des affirmations sur la mesure de la difficulté des item dans un test à temps limité sans appuyer leurs énoncés

de recherches expérimentales. Une synthèse de ces énoncés apparaîtra dans la première partie de cette section. Elle sera suivie en deuxième partie par la présentation de rapports de recherche qui ont traité de ce sujet ou plus généralement de la constance des mesures de difficulté. Une dernière partie fera le point sur le problème et proposera des hypothèses qui devront ensuite être éprouvées expérimentalement.

- a. La mesure de la difficulté des item d'un test à temps limité.

La variable principale qui influence la mesure de la difficulté des item des tests à temps limité est la période de temps plus ou moins longue qui empêche un certain nombre de sujets d'atteindre tous les item. Comme le rapportent Freeman¹³ et Guilford¹⁴, les item de la fin sont atteints par un nombre de plus en plus réduit de sujets qui n'est plus représentatif de la population et les mesures de difficulté donnent des résultats apparents qui s'éloignent de plus en plus de la difficulté réelle.

¹³ Frank S. Freeman, Theory and Practice of Psychological Testing, Toronto, Holt, Rinehart and Winston, 1966, p. 116.

¹⁴ J.P. Guilford, Psychometric Methods, Toronto, McGraw-Hill, 1954, p. 436.

Ce problème a été analysé par Horst qui le présente comme ceci:

...if we wish to determine the difficulty of the individual items in a speeded test, we are confronted with the problem that not all persons have attempted all items. We can of course use as an estimate of the difficulty the number of persons who attempted it. This, however, would tend to give a biased estimate of the difficulty of the item. It is probable that, of those persons who did not attempt the item because they were slower and had less ability in the function measured by the item, the proportion who could answer the item correctly is smaller than for those who attempted the item. Therefore the difficulty index for an item which is not attempted by all individuals, based on the number of individuals who do attempt it, would probably indicate that the item is easier than is actually the case for the total group. It is also possible, of course, that those working more rapidly and attempting more items may also be more careless than the slower workers. Practically, we may consider that the item was administered to only those persons who attempted it. Obviously, this is only a subsample of the total group, and a biased sample at that¹⁵.

Pour contourner cette situation, il est possible d'administrer le test sans limite de temps et permettre ainsi à tous les individus d'atteindre tous les items. Mais cette solution crée peut-être des imprécisions encore plus grandes comme le rapporte Henrysson:

¹⁵ Paul Horst, Psychological Measurement and Prediction, Belmont (California), Wadsworth Publishing Company, 1966, p. 397-398.

...if the tryout is conducted with generous time limits, the mental set and rate of work of the examinees may be so different from those likely to exist during administration of the final form the tryout data may not be useful¹⁶.

Ce dilemme est aussi élaboré par Horst:

In the administration of a test for item analysis purposes it is important that most of the group should finish all of the items. This is the only way in which it is possible to calculate the indices for various types of item analysis. The most serious objection to this procedure is that it is ordinarily not typical of the situation when the test is administered for decision purposes. In such cases much less time is usually allowed. There seems, however, to be no entirely satisfactory solution to the problem at this time¹⁷.

Cette impasse semble prendre des proportions encore plus grandes lorsqu'il faut calculer l'ordre de difficulté des item d'un test de puissance à temps limité.

D'une part, si l'analyse de la difficulté des item se fait en respectant la limite de temps imposée par la standardisation, la mesure de la difficulté d'un item varie selon qu'il apparaît au début, au milieu ou à la fin du test. Dans de telles circonstances, il semble difficile de parvenir à un ordre de difficulté croissante des item qui soit stable. Si les item sont placés en ordre d'après les résultats d'une première analyse, une deuxième analyse risque de montrer que

¹⁶ Sten Henrysson, Gathering, Analyzing, and Using Data on Test Items, dans Educational Measurement, Robert L. Thorndike, éditeur, Washington, American Council on Education, 1971, p. 133.

¹⁷ P. Horst, op. cit., p. 393.

tel item, qui était perçu comme facile parce qu'il était placé à la fin du test, est maintenant perçu comme plus difficile parce qu'il a été devancé dans le test et qu'il a été atteint par des sujets moins doués. Le raisonnement correspondant semble valable pour l'item qui a été reporté plus loin dans le test parce qu'il avait été perçu comme difficile. Si ce phénomène s'avérait réel, il pourrait y avoir une suite infinie d'analyses sans parvenir à un ordre de difficulté croissante des item qui soit stable.

D'autre part si l'analyse de la difficulté des item se fait après avoir permis à tous les sujets d'atteindre tous les item, la mesure de la difficulté sera sans doute plus précise et plus stable. Mais les item qui seront placés en ordre de difficulté croissante selon cette analyse ne maintiendront peut-être pas le même ordre lorsque la limite de temps sera réintégrée à l'administration du test. En effet rien n'indique que la limite de temps a un effet égal sur la difficulté de tous les item indépendamment de leur position dans le test.

L'impasse semble totale: en respectant la limite de temps, on parvient à des résultats imprécis et inconstants alors qu'en supprimant la limite de temps, on obtient des résultats peut-être plus précis et plus stables, mais dont l'application demeure douteuse.

Après cette clarification du problème, l'étape suivante dans la recherche d'une solution est d'analyser les recherches expérimentales susceptibles d'apporter des éléments de réponse.

- b. Recension des recherches sur la mesure de la difficulté des item et sur la constance de cette mesure.

Il ne semble pas y avoir de chercheur qui ait étudié directement la relation entre la constance de l'ordre de difficulté des item et la limite de temps imposée dans l'administration d'un test. Les quelques recherches publiées dans ce domaine portent plutôt sur la constance des indices de difficulté des item. Ces recherches, même si elles semblent relativement périphériques au sujet, situent les cadres qui seront des points de repère dans l'élaboration des hypothèses de recherche de ce travail et dans l'interprétation des résultats.

Le premier auteur qui sera considéré, Gibbons¹⁸, administra un test d'algèbre de quarante-cinq item à 426 étudiants. Après avoir analysé la difficulté des item, il plaça ceux-ci en ordre de difficulté croissante et les administra à un autre groupe de 435 étudiants. Il avait choisi

¹⁸ Charles C. Gibbons, The Predictive Value of the Most Valid Items of an Examination, dans Journal of Educational Psychology, vol. 31, n° 8, livraison de novembre 1940, p. 616-621.

comme indice de difficulté le nombre de personnes qui réussissaient un item. Le coefficient de corrélation entre les indices de difficulté obtenus du premier groupe et les indices de difficulté du deuxième groupe était de 0.97 . Ce résultat montre qu'à partir d'ordres de présentation différents, cet indice de difficulté est stable, mais l'auteur ne précise pas si la plupart des élèves ont terminé l'examen malgré une limite de temps de soixante minutes.

Le nombre de réussites sur un item n'est plus utilisé comme indice de difficulté parce qu'il rend les comparaisons difficiles d'un groupe à l'autre. Il en est de même pour le nombre d'erreurs sur un item qui est utilisé comme indice de difficulté dans la recherche suivante. Dans son étude, Carter¹⁹ analysa la difficulté des item d'un examen construit pour évaluer les étudiants d'un cours de psychologie. Cet examen de quatre-vingts item fut administré à deux groupes d'environ cent élèves. Le premier groupe fut divisé au hasard en quatre sous-groupes égaux. Les coefficients de constance des indices de difficulté varient entre les sous-groupes de 0.92 à 0.94 .

A l'intérieur de chacun des groupes, il sélectionna les 25 élèves qui avaient le mieux réussi et les 25 qui

¹⁹ Harold D. Carter, How Reliable Are the Common Measure of Difficulty and Validity of Objective Test Items, dans The Journal of Psychology, vol. 13, première demie, livraison de janvier 1942, p. 31-39.

avaient le moins bien réussi. Le coefficient de constance entre les meilleurs des deux groupes est de 0.90 alors que celui calculé entre les plus faibles est de 0.91 . A l'intérieur des groupes, les coefficients de corrélation entre les meilleurs élèves et les plus faibles sont de 0.81 et de 0.84 . Ces derniers coefficients indiquent la constance de la difficulté relative des item ou de l'ordre de difficulté des item. En effet, bien que les moyennes des plus faibles (13.6 et 13.5 erreurs) et des meilleurs (8.7 et 8.8 erreurs) soient très différentes, un item qui est perçu comme plus facile que les autres par les meilleurs étudiants le sera probablement aussi par les étudiants faibles. Ces derniers résultats sont les plus originaux de cette étude.

Il serait possible de contester le choix d'un examen dont les questions n'avaient pas été analysées précédemment. Il aurait été possible d'enlever les questions réussies par tous ou échouées par tous qui viennent augmenter les coefficients de constance mais sans posséder de valeur discriminante. D'autre part, le nombre d'élèves dans les sous-groupes est de beaucoup inférieur à ce qui est exigé habituellement pour le calcul de la difficulté des item. En résumé, cette étude apporte des indications sur la constance de la difficulté relative des item, particulièrement entre les élèves les plus forts et les élèves les plus faibles.

Puisque l'auteur ne fait aucune mention de la période de temps allouée pour passer l'examen, il est impossible de préciser si ces résultats s'appliquent aux tests de puissance ou aux tests de puissance à temps limité.

Une recherche plus poussée conduite par Mollenkopf²⁰ mit en évidence les effets de la place des item dans un test et les effets de la durée de l'administration sur des indices de difficulté. Il administra deux formules d'un test d'aptitude verbale et deux formules d'un test d'aptitude mathématique à des étudiants du niveau secondaire. Les item des deux tests étaient à choix multiples. Le test d'aptitude verbale était composé de 85 antonymes alors que le test de mathématique était composé de 36 questions.

Chacun des tests était divisé en trois. Dans le premier et le dernier tiers, les item étaient placés en ordre de difficulté croissante; le tiers central était composé d'item de difficulté moyenne. Dans une première formule, le premier tiers comprenait les item les plus faciles et le dernier tiers, les plus difficiles. Dans la deuxième formule, le tiers le plus facile fut placé à la fin et le tiers le plus difficile fut placé au début.

²⁰ William G. Mollenkopf, An Experimental Study of the Effects on Item-Analysis Data of Changing Item Placement and Test Time Limit, dans Psychometrika, vol. XV, n° 3, livraison de septembre 1950, p. 291-317.

L'échantillon fut divisé au hasard en quatre groupes d'environ cent sujets. Ces groupes se sont révélés équivalents sur trois mesures verbales et sur le Q-score du American Council on Education Psychological Examination. Les premier et deuxième groupes passèrent respectivement la première et la deuxième formules du test d'aptitudes verbales sans être limités par le temps. Le troisième et le quatrième groupes passèrent respectivement la première et la deuxième formules, mais dans une période de temps limitée qui permit à très peu d'étudiants d'atteindre tous les items. Le temps écoulé, les étudiants purent continuer le test en utilisant des crayons d'une autre couleur. Les conditions expérimentales furent les mêmes pour l'administration du test d'aptitudes mathématiques sauf que l'ordre des groupes a été modifié.

Une première constatation générale à l'analyse des résultats fut de découvrir des coefficients de corrélation élevés (0.81, 0.82, 0.91 et 0.94) entre les scores bruts totaux obtenus dans une période de temps limitée et dans une période de temps illimitée. Mollenkopf résume comme suit les effets du temps et de la position des items sur l'indice de difficulté "p" (pourcentage de réussite):

a. When power data only were compared, the p value of an item was slightly higher when the item appeared early than when it came late in the test, for the verbal forms. No similar position effect was noted in the comparisons involving only power data on the mathematics forms.

b. When speed and power data were compared, before drop-out began, the power p's usually exceeded the speed p's for the verbal forms. With high drop-out, the speed p's exceeded the power p's for the verbal tests. No corresponding relationships were found in the comparisons involving speed and power data for the mathematics forms.

c. When speed data for one form were compared with speed data for a second form in which the item arrangement was different, for the verbal forms the p of an item was distinctly higher when the item came late than when it came early. No corresponding trends were noted for comparisons involving speed data on two mathematics forms²¹.

Ces résultats viennent compléter ceux de Gibbons.

En effet, si les indices de difficulté demeurent stables même à partir d'ordres de présentation différents, il semble que cette stabilité est compromise quand il y a une restriction de temps qui est imposée dans l'administration du test. Il faut remarquer cependant qu'il s'agit ici de la difficulté absolue des item et non de la difficulté relative. En d'autres mots, l'indice de difficulté d'un item augmente ou diminue selon que l'item est placé au début ou à la fin d'un test à temps limité, mais cette recherche ne permet pas de spécifier si cette variation est assez marquée pour que l'item change de rang de façon significative dans l'échelle

21 W.G. Mollenkopf, op. cit., p. 310-311.

de difficulté des item. La stabilité ou l'instabilité de la difficulté relative des item reste à être élucidée.

Mollenkopf utilisa aussi un autre indice de difficulté qui s'avéra moins précis à l'analyse.

Une autre recherche fut conduite par Brenner^{22,23} à Ohio State University. Il étudia les effets de différents ordres de difficulté sur la constance, la difficulté et la validité interne d'un test. Il utilisa comme instruments les examens administrés aux étudiants d'un cours de psychologie. Les item à choix multiples de quatre options étaient atteints en général par 99% des étudiants. L'étude s'échelonna sur deux trimestres. Pendant le premier, quatre paires d'examens de quarante item furent administrées à intervalles réguliers de deux semaines à un premier groupe d'étudiants. Ces examens servirent de base aux expériences menées au second semestre chez un autre groupe d'étudiants.

La première expérience consista d'abord en une nouvelle administration de l'un des examens de la première paire. Les item de ce test étaient distribués au hasard et

22 Marshal H. Brenner, Test Difficulty, Reliability, and Discrimination as Functions of Item Difficulty Order, thèse de doctorat (inérite) présentée à Ohio State University, 1962, vii-92 p.

23 -----, Test Difficulty, Reliability, and Discrimination as Functions of Item Difficulty Order, dans Journal of Applied Psychology, vol. 48, n° 2, livraison d'avril 1964, p. 98-100.

furent administrés à 184 étudiants. Ces item furent ensuite placés successivement en ordre de difficulté croissante et en ordre de difficulté décroissante. Ces deux nouvelles formules furent administrées respectivement à 186 et 185 étudiants.

Des quatre-vingts item de la deuxième paire d'examens, quarante furent choisis pour la deuxième expérience. Parmi les item choisis se trouvaient les dix plus faciles et les dix plus difficiles tandis que les vingt autres furent choisis à différents niveaux de difficulté. Dans une première formule administrée à 275 étudiants, les dix item les plus faciles furent placés au début en ordre de difficulté croissante et les autres suivirent au hasard. Dans la deuxième formule administrée à 267 étudiants, les dix item les plus difficiles furent placés au début en ordre de difficulté décroissante et les autres suivirent au hasard.

Lors de la troisième expérience, l'un des examens de la troisième paire fut utilisé. Ses item furent placés en ordre de difficulté croissante dans une première formule et furent administrés à 267 étudiants. La deuxième formule était en ordre de difficulté décroissante et fut administrée à 267 autres étudiants.

La technique de sélection utilisée dans la deuxième expérience fut reprise dans la quatrième expérience. Les item furent présentés en ordre de difficulté croissante à

192 étudiants et une deuxième formule, en ordre de difficulté décroissante, fut administrée à 191 étudiants.

Dans chacune des formules utilisées, la difficulté des item fut calculée par la proportion de succès par rapport au nombre d'individus qui avaient essayé l'item. Dans chacune des expériences, Brenner calcula le ou les coefficients de constance entre les indices de difficulté des diverses formules utilisées. Il calcula aussi les coefficients de constance entre les indices de difficulté des item des expériences du deuxième semestre et les indices de difficulté que ces item avaient obtenus lors du semestre précédent. Tous ces coefficients de constance varient de 0.950 à 0.986, ce qui est très élevé. Brenner interprète ces résultats comme suit:

This extremely small range of very high values indicates convincingly that the difficulty of an item is a very reliable statistic regardless of the ordering of the items by difficulty or the specific content of the other items in the test (as long as all the items in the test are concerned with the same general topic)²⁴.

Il faut rappeler que ces résultats ont été obtenus à partir de tests de puissance où 99% des étudiants atteignaient tous les item.

24 M.H. Brenner, Test Difficulty, Reliability, and Discrimination As Functions of Item Difficulty Order, thèse de doctorat, op. cit., p. 61.

Cette recherche qui vient appuyer les conclusions des études précédentes semble bien conduite. Les échantillons sont assez considérables et l'analyse se ramifie dans quatre expériences successives pour donner plus de poids aux conclusions. En plus de cette étude de constance, cette recherche est l'une des mieux structurées et des plus complètes sur le sujet de l'ordre de difficulté des item d'un test.

Enfin la dernière étude qui sera rapportée dans cette section a été menée par Hilgard et Jandron²⁵ qui s'intéressaient aux influences du milieu sur la stabilité des mesures de difficulté et de valeur discriminante d'un test. Conscients que des variables propres au milieu pouvaient influencer considérablement les résultats d'un examen, ils ont voulu étudier l'ampleur des différences qu'ils pourraient retrouver entre deux universités. Un cours d'introduction à la psychologie fut choisi pour les fins de la recherche et le même volume de base était utilisé dans les deux universités.

Un premier examen composé de 230 questions à quatre choix fut administré à 221 étudiants de l'Université de Stanford, tandis qu'un deuxième examen de 230 questions à

²⁵ Ernest R. Hilgard et Earl L. Jandron, Stability of Item Statistics in Two Institutions, dans Journal of Educational Psychology, vol. 51, n° 4, livraison de août 1960, p. 195-198.

quatre choix fut présenté à 125 étudiants du San Jose State College; 115 questions étaient communes aux deux tests.

Le degré de difficulté d'un item fut calculé par le pourcentage de réussite. Les étudiants de Stanford, qui étaient d'un niveau académique plus élevé, ont obtenu une moyenne (63.5) significativement plus élevée que celle (47.6) des étudiants de San Jose. Malgré cette différence, la difficulté relative des item communs était semblable aux deux endroits et fut représentée par un coefficient de corrélation de 0.80 . Ce coefficient de corrélation semble faible lorsqu'il est comparé à tous ceux qui ont été présentés précédemment. Ceci pourrait s'expliquer par le fait que les groupes ne sont pas équivalents, qu'ils sont de milieux différents, qu'ils sont de niveaux académiques différents, ou qu'ils n'ont pas suivi la même présentation de cours en dépit du volume de base commun.

Cette recherche est assez banale et peu rigoureuse au point de vue méthodologique. Elle serait sans doute facile à reproduire puisqu'elle exige peu de contrôle expérimental; cependant ce manque de contrôle expose les résultats à trop d'instabilité et lui enlève de la valeur. Le coefficient de constance de la difficulté des item peut au plus être interprété vaguement comme un minimum de constance à laquelle on est en droit de s'attendre entre des groupes équivalents.

En résumé, ces recherches, qui pour la plupart ont été faites à partir de tests de rendement, montrent que les mesures de difficulté des item peuvent être très stables quel que soit l'ordre de présentation des item, à la condition que tous les étudiants aient la chance d'atteindre tous les item. Un seul auteur, Mollenkopf²⁶, met en doute la constance des mesures de difficulté des item des tests à temps limité. Cette restriction touche l'essence du problème de la présente recherche et devrait être contournée selon lui en laissant au moins la moitié du groupe d'étudiants atteindre tous les item. Le problème de la mesure de la difficulté des item d'un test de puissance à temps limité qui a été présenté initialement demeure donc le même.

Dans le but éventuel d'analyser ce problème expérimentalement, des hypothèses de recherches seront présentées.

c. Hypothèses de recherche.

Le problème initial de la mesure de l'ordre de difficulté des item dans un test de puissance à temps limité n'a pas été modifié par les recherches qui viennent d'être analysées. D'une part, il semble exister une différence entre la mesure de la difficulté d'un item administré dans un test de puissance où il est atteint par tous les élèves et la

26 W.G. Mollenkopf, op. cit., p. 312.

mesure de la difficulté de cet item administré dans un test limité par le temps où seulement une fraction des élèves l'atteignent. D'autre part, rien n'indique que cette différence soit telle que l'item qui occupait un rang donné dans l'échelle de difficulté d'un test occupe un autre rang dans l'échelle de difficulté de l'autre test. Il est possible que ces différences entre les indices de difficulté soient toujours dans le même sens, par exemple que les item soient identifiés comme de plus en plus faciles lorsqu'ils sont atteints par de moins en moins d'élèves. Il est possible aussi que ces différences ne soient pas assez marquées pour bouleverser l'ordre de difficulté des item d'un test par rapport à l'autre. Il serait donc possible que l'ordre de difficulté des item d'un test de puissance à temps limité puisse atteindre un niveau de constance satisfaisant tout comme dans un test de puissance. Cette possibilité permet de formuler une première hypothèse de travail:

Il n'y a pas de différence significative entre les ordres de difficulté croissante des item calculés à partir d'ordres différents de présentation des item d'un test de puissance à temps limité comme la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale administrée en trente minutes.

Une deuxième hypothèse peut se formuler comme suit:

Il n'y a pas de différence significative entre les ordres de difficulté croissante des item d'un test de puissance à temps limité, comme la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale, administrée dans une période de temps limité et les ordres de difficulté croissante des item d'un tel test alors qu'il est complété par la plupart des élèves.

Ces deux hypothèses qui portent sur la constance de l'ordre de difficulté des item seront étudiées en tenant compte des recherches qui ont été analysées précédemment. L'élaboration du schème expérimental qui permettra d'éprouver ces hypothèses sera présentée au deuxième chapitre.

3. Recension des écrits, deuxième partie:
effets de l'ordre de difficulté des item.

Si les caractéristiques d'un test de puissance à temps limité, tel l'Examen intermédiaire Otis-Ottawa, semblent nuire à la mesure de la difficulté des item et diminuer les possibilités d'établir un ordre constant de difficulté croissante, paradoxalement ces mêmes caractéristiques semblent donner de l'importance à un ordre de difficulté croissante. En effet, si l'on veut évaluer des élèves dont l'habileté mentale varie considérablement, même en utilisant un test de puissance, il semble naturel de leur présenter des item en ordre de difficulté croissante pour éviter que les moins doués se fatiguent inutilement à tenter de répondre à des item trop difficiles pour eux. Cette présentation des item en ordre de difficulté croissante semble logique à plus forte raison quand le test de puissance est limité par le temps. Un tel test dont les item ne seraient pas en ordre de difficulté croissante risquerait d'être injuste pour l'élève moins doué. Celui-ci non seulement pourrait se fatiguer sur des item trop difficiles

pour lui, mais pourrait être empêché par ces item d'atteindre des item plus faciles et plus révélateurs de son niveau de fonctionnement qui apparaissent plus loin.

Mais ceci n'est qu'une déduction qui pourrait être acceptée par les uns et mise en doute par d'autres comme on pourra le constater dans les pages qui suivent. Cette déduction pourrait aussi être élaborée; l'ordre de difficulté des item n'est que l'une des variables qui composent l'administration d'un test et sa valeur est liée aux autres variables du test et aux caractéristiques de l'élève.

Dans le but de situer le cadre du problème de l'ordre de difficulté des item, des énoncés que des auteurs ont élaborés au sujet des tests de rendement et des tests d'aptitude seront maintenant présentés. Ces énoncés seront suivis de descriptions de recherches qui ont relevé les effets de différents ordres de difficulté.

a. Importance de l'ordre de difficulté des item dans les tests de rendement.

Le sujet de l'ordre de difficulté des item dans un test semble se prêter facilement à des déductions et à des énoncés de principes. Certains auteurs en traitent comme par hasard et il serait impossible de relever tous ceux qui ont émis une opinion à ce sujet.

Par exemple, Ruch²⁷ affirma que la présentation des items en ordre de difficulté croissante augmentait à la fois la validité et la constance d'un test. Il n'élabora pas davantage cette affirmation et n'apporta aucune évidence expérimentale pour la supporter.

Plusieurs années plus tard, Furst²⁸ traita aussi ce sujet dans un livre sur les principes de construction de tests. Il fit alors ressortir quelques avantages et désavantages que cette pratique apportait à l'étudiant et au professeur:

Many standardized tests have their items arranged in an increasing order of difficulty - from very easy to moderately difficult to very difficult. This is the common sequence when items are grouped according to difficulty. The advantages are that -

1. The easy items which come early give the student a chance to warm up and to build up confidence rather than to experience discouragement at the outset.

2. It paces the tasks so that the student can work steadily to the limit of his ability rather than get stalled early on hard items. This makes for increased test reliability since the functioning sample of items for each student tends to reach a maximum in the time available.

3. It gives a greater spread of scores than would a jumbled order or an inverse order of difficulty.

27 Giles M. Ruch, The Objective or New-Type Examination, New York, Scott, Foresman and Co., 1929, p. 32.

28 Edward J. Furst, Constructing Evaluation Instruments, New York, David McKay, 1958, xv-334 p.

The disadvantages of this arrangement are that -
1. It may not provide as comfortable, meaningful, or useful a grouping as one by subject matter, type of objective, or form. Thus, it may be fatiguing for the student to have to shift continually from one topic to a different one rather than to concentrate on related material.

2. The very difficult items come at the end of the test when the student reaches the point of greatest fatigue. Actually it might be better for students to encounter these hard items after an early warm-up period; or to work alternately on easy and hard items so as to have a chance to relax between periods of intense concentration.

3. It requires knowing the true order of difficulty before tryout - and this information is usually not available²⁹.

Ces propos sont plausibles et révèlent des caractéristiques qui peuvent être observées dans une situation de testing. On peut remarquer que certains élèves deviennent très anxieux si les premiers item sont trop difficiles et ne peuvent atteindre autant d'item que leur potentiel devrait leur permettre. D'autres préfèrent parcourir tout le test à la recherche d'item d'une même catégorie, revenir au début et recommencer la même stratégie pour une deuxième catégorie et ainsi de suite. D'autres finalement sont si épuisés ou frustrés qu'ils démissionnent avant la fin du test.

Un autre auteur, Stanley³⁰, opte pour la présentation des item en ordre de difficulté croissante alors que

29 E.J. Furst, op. cit., p. 278-279.

30 Julian C. Stanley, Measurement in Today's Schools, New York, Prentice-Hall, 1965, p. 187.

Gerberich, Greene, Jorgensen³¹; et Helmstadter³² préfèrent un ordre de présentation par catégories de contenu.

Helmstadter énonce cependant des règles pour minimiser les effets d'un ordre de difficulté laissé au hasard:

Any unreliability which normally would result from examinees spending too much time on a difficult item while not attempting an easy item can be prevented by setting adequate time limits and by suggesting in the directions that the examinee should not spend too much time on any one item, but rather skip a difficult item and go back to it later if time permits³³.

L'originalité de cette opinion est de souligner une relation entre l'ordre de difficulté des item et la période de temps allouée pour passer le test. D'après ces écrits, l'ordre de difficulté n'aurait d'importance que lorsque la période de temps est trop brève. Ceci sera élaboré, un peu plus loin à partir des expériences qui ont été conduites à ce sujet par d'autres auteurs.

Finalement, Henrysson³⁴ affirme que les item devraient habituellement être présentés selon leur niveau de difficulté, les plus faciles apparaissant les premiers.

31 Raymond L. Gerberich, H.A. Greene and A.N. Jorgensen, Measurement and Evaluation in the Modern School, New York, David McKay, 1962, p. 253.

32 G.C. Helmstadter, Principles of Psychological Measurement, New York, Appleton-Century-Crofts, 1964, p. 173.

33 Idem, ibid., p. 173.

34 S. Henrysson, op. cit., p. 144.

Il ajoute cependant que l'ordre de présentation des item ne peut pas reposer uniquement sur la mesure de la difficulté puisque celle-ci est contaminée par les erreurs d'échantillonnage et par les erreurs de mesure, surtout lorsque l'échantillon est petit et que les item sont peu nombreux. Une mesure précise de l'ordre de difficulté des item est donc rendue impossible dans la plupart des situations scolaires et le problème de l'ordre de difficulté perd ainsi de son attrait, même s'il peut demeurer important.

En résumé, des auteurs recommandent l'application de l'ordre de difficulté croissante dans les tests de rendement, d'autres s'y opposent. Quoi qu'il en soit, ce procédé semble réservé aux tests standardisés sur un grand nombre d'individus.

L'étape suivante de cette recension d'écrits traitera des tests d'aptitudes. Dans l'ensemble, ces tests remplissent mieux les exigences d'une standardisation.

b. Importance de l'ordre de difficulté croissante des item dans les tests d'aptitude.

Le sujet de l'ordre de difficulté croissante est apparu assez tôt dans l'histoire des tests d'aptitudes. Déjà en 1928, Hull encourageait cette présentation:

When organized into a test unit, the various items should be arranged accurately in the order of increasing difficulty. [5.7]

The items retained in the series are then arranged in such a way that each succeeding item is passed by a smaller and smaller number of persons. It is true that the order of increasing difficulty for one person will not be exactly the same as that for another, but the above arrangement reduces the difficulty from this source to a minimum. With the items arranged in this way, the person of limited capacity finds at once those items upon which he can demonstrate his powers without wasting his time and depleting his courage by struggling with items which offer no possibility of success³⁵.

Il soutient son énoncé par un raisonnement qui ressemble fort à ceux qui ont été appliqués aux tests de rendement. Il révèle en plus un aspect technique intéressant: même si une mesure précise de la difficulté des item est calculée et que ceux-ci sont placés ensuite en ordre de difficulté croissante, ceci n'implique pas que cet ordre convienne le mieux à chacun des individus, mais plutôt qu'il convient le mieux à un plus grand nombre d'individus.

Un nouvel avantage de l'ordre de difficulté croissante est introduit par Horst qui voit dans ce procédé un moyen de diminuer les erreurs associées aux limites de temps et de rapprocher ainsi les tests de puissance à temps limité des tests de puissance.

35 Clark L. Hull, Aptitude Testing, New York, World Book Co., 1928, p. 320-321.

...it would probably still be best, in the eventual test format, to arrange the items in order of difficulty, as determined by the item difficulty analysis. This arrangement of items in the finished booklet would then result in the more ideal type of instrument represented by the power test which indicates primarily the degree of item difficulty persons are capable of answering, rather than simply the total number of items they can answer correctly within a given time limit³⁶.

Implicitement, il associe donc le problème de l'ordre de difficulté des item à la limite de temps imposée dans l'administration du test. Il rejoint ainsi Helmstadster qui avait fait la même relation chez les tests de rendement.

En général, dans le cadre du problème de l'ordre de difficulté, il est possible d'appliquer aux tests d'aptitudes les principes qui avaient été élaborés pour les tests de rendement et vice versa. Selon les énoncés qui viennent d'être rapportés, la tendance principale serait de favoriser la construction de tests de rendement et d'aptitudes avec une présentation d'item en ordre de difficulté croissante. C'est aussi ce qui est rapporté par Nunnally:

In many aptitude tests and achievement tests, there is a wide range of item difficulty, and the practice frequently is to order the items in terms of difficulty. Then the items at the beginning of the test are very easy, and the items near the end of the test are very difficult³⁷.

36 P. Horst, op. cit., p. 399-400.

37 Jum C. Nunnally, Psychometric Theory, New York, McGraw-Hill, 1967, p. 571.

Il semble donc préférable, de l'avis de plusieurs auteurs contemporains d'ordonner les item selon leur difficulté. Les rapports de recherche qui suivent révéleront si ce principe demeure au niveau expérimental. Les premières recherches rapportées traiteront de tests de rendement alors que les dernières ont été faites à partir de tests d'intelligence.

c. Recension des recherches sur les effets de l'ordre de difficulté des item.

Le premier rapport de recherche sur ce sujet a été publié en 1933 par Capron³⁸. Il utilisa trois tests: un test d'arithmétique, un test sur les processus fondamentaux en arithmétique et un test d'épellation. Il établit trois ordres de difficulté pour chacun de ces tests, de l'item le plus facile au plus difficile, du plus difficile au plus facile et un ordre au hasard. L'échantillon comprenait quatre cent cinquante-trois élèves de cinquième et huitième année. Des formules différentes étaient employées selon l'année scolaire des sujets. Le temps d'administration des tests fut prolongé de dix minutes pour permettre à la plupart des étudiants de terminer tous les item. Chacun des étudiants

³⁸ Virginia L. Capron, The Relative Effect of Three Orders of Arrangement of Items Upon Pupils' Scores in Certain Arithmetic and Spelling Tests, dans Journal of Educational Psychology, vol. XXIV, n° 9, livraison de décembre 1933, p. 687-695.

passa tous les tests dans leurs trois ordres de présentation. Chacun fit donc face au même item trois fois. L'effet de pratique fut contrôlé par une rotation de l'ordre de présentation des tests pour les différents groupes. En plus d'étudier les effets de l'ordre de difficulté des item sur le groupe total, il les étudia aussi dans des sous-groupes divisés selon l'âge chronologique, l'âge mental, le quotient intellectuel et selon le sexe. Il résume comme suit les résultats qu'il a obtenus:

1. Order of arrangement of items has in general very little effect upon pupils' scores in power (scaled) tests in problems, spelling, and arithmetic fundamentals.

2. On the basis of percentage of error very few items reveal important differences in amount of error. Differences which do occur are random, that is, no one order is responsible for the greatest number of errors, for the maximum percentages of error are distributed about equally among the three orders of arrangement.

3. On the basis of differences in changes in scores, problems, spelling, and arithmetic fundamentals are consistent in showing only slight variations.

4. Although mean differences are small, there is considerable scattering above and below the mean, indicating that individual pupils may be considerably influenced by the factor of order of items, but this influence is very different in the case of different pupils³⁹.

Capron conclut en disant qu'il n'y aurait pas lieu pour les professeurs de placer les item de leurs examens en ordre de difficulté lorsque la majorité des élèves peuvent

39 V.L. Capron, op. cit., p. 694-695.

les compléter dans la période de temps permise. Il faut remarquer qu'il parle alors de tests de puissance. En effet, même s'il a choisi des tests à temps limité, il les a transformés en tests de puissance pour les fins de son expérience.

Capron rejeta de son expérience les scores obtenus par tout étudiant qui n'avait réussi à compléter aucune des trois présentations du test. Cette exclusion est peut-être significative; il aurait été intéressant de comparer ces scores aux scores retenus dans l'expérience pour constater si les étudiants qui n'ont pas réussi à compléter un test ne sont pas aussi ceux qui ont été le plus affectés par les différents ordres de présentation des item.

La plus grande faiblesse de cette expérience a été d'avoir présenté les mêmes tests trois fois aux étudiants. Il n'est pas impensable que l'étudiant qui fait face au même problème une deuxième et une troisième fois mettra moins de temps à le comprendre, à le résoudre, augmentant ainsi ses chances de succès. Ce phénomène d'apprentissage n'aura-t-il pas tendance à contaminer les effets des différents ordres de difficulté? Les formules ont été présentées selon une séquence différente pour chacun des groupes, il est vrai, et cette rotation empêche qu'une formule soit favorisée par la position qu'elle occupe dans l'administration de tous les tests. Mais ce contrôle n'enlève pas l'effet général de

l'apprentissage, il ne fait qu'assurer que toutes les formules soient contaminées également.

Le problème de l'ordre de difficulté des item fut aussi étudié par Mensh⁴⁰ à l'intérieur d'un projet plus vaste concernant les effets de l'abrégement d'un test sur les réponses données aux item. Cette étude sur l'ordre de difficulté fut conduite à partir de douze item "de base" du McGill Verbal Situation Series.

Les cinquante item qui composaient le test original furent réduits successivement à vingt-cinq, à vingt, et finalement à douze item dits de base. Chacune de ces formules abrégées fut ensuite administrée à un tiers d'un groupe de 534 candidats au moment de leur admission à l'Université Northwestern. Les item de base qui étaient en ordre de difficulté croissante de 1 à 12 furent ensuite présentés selon la séquence 4, 5, 6, 1, 2, 7, 9, 8, 11, 3, 12 et 10 pour former trois nouvelles formules qui furent administrées respectivement à trois sous-groupes d'un autre échantillon de 270 candidats.

Une analyse de la variance révéla que les item en ordre de difficulté croissante avaient été significativement mieux réussis que les autres. Cette tendance se maintient

⁴⁰ Ivan N. Mensh, The Effects of Test Abbreviation Upon Responses to the Individual Items, dans The Journal of Clinical Psychology, vol. v, n° 1, livraison de janvier 1949, p. 22-36.

lorsque les individus des deux groupes sont ensuite comparés selon leur âge, leur niveau d'éducation et selon leur score sur un test équivalent qui fut utilisé comme critère externe. Mensh conclut que l'ordre de difficulté croissante des item dans un test semble produire une meilleure performance qu'un ordre irrégulier.

Cette recherche semble bien faite même si l'auteur n'a pas rapporté clairement comment il a regroupé (matched) les individus de ses échantillons. Elle est la première qui révèle que l'ordre de difficulté a un effet significatif sur les scores bruts totaux d'un test. Cette relation entre la performance à un test et l'ordre de difficulté des item fut ensuite étudiée par Lund⁴¹ dans sa thèse de doctorat. Il utilisa The Henmon-Nelson Test of Mental Ability, Form B. Ce test spiral-omnibus de soixante-douze item à cinq choix est très semblable à l'Otis. Il est habituellement administré dans une période de temps précise, mais Lund enleva la restriction de temps pour les fins de son expérience.

Il administra ce test à deux groupes de 90 étudiants de première année collégiale choisis dans un cours d'orientation. Ces étudiants avaient déjà passé des tests au niveau secondaire et à leur entrée au niveau collégial.

⁴¹ Kenneth Withelm Lund, Test Performance as Related to Order of Item Difficulty, Anxiety and Intelligence, thèse de doctorat (inédiée) présentée à l'Université-Northwestern, Evanston, Illinois, 1953, iii-66 p.

Les individus des deux groupes ont été pairés selon leur score à The American Council on Education Psychological Examination.

Le premier groupe d'étudiants passa le test original alors que pour le deuxième groupe le quatrième quart du test, soit les item cinquante-cinq à soixante-douze, fut placé au début. Pour les fins de la comparaison, le test fut ensuite divisé en quatre parties. Les scores des étudiants furent cumulés dans chacun des quarts et Lund compara les deux groupes par une analyse de variance et par des tests "t". Il parvint à la conclusion que le nouvel ordre de difficulté avait diminué significativement les scores quels que soient le niveau d'intelligence des étudiants et leur niveau d'anxiété tel que mesuré par le Taylor Anxiety Scale. Toutefois, il ne trouva une différence significative qu'entre le troisième quart du test original et le quatrième quart correspondant de la deuxième formule, à l'avantage du test original.

Lund vient ainsi appuyer les résultats de Mensh qui avait démontré que l'ordre de difficulté croissante des item dans un test favorisait une meilleure performance qu'un ordre irrégulier. Cependant son rapport révèle qu'il a omis de calculer la difficulté des item au début de son expérience. Cette lacune n'infirme pas totalement ses résultats, mais il fut réduit à considérer non pas l'ordre des item les

uns par rapport aux autres, mais plutôt des quartiers de test les uns par rapport aux autres. La précision de la variable indépendante est donc vague, incertaine et rien ne rejette la possibilité que des item du quatrième quart aient été plus faciles que ceux des autres quarts.

L'expérience de French et de Greer⁴² étendit le champ d'exploration du problème de l'ordre de présentation des item: en plus d'étudier les effets de l'ordre de difficulté des item d'un test d'intelligence sur la mesure du quotient intellectuel, ils analysèrent aussi ses effets sur le comportement psychologique et physiologique des élèves.

Ils administrèrent un test individuel d'intelligence, The Pictorial Test of Intelligence (PTI), à 152 élèves de première année dont l'âge variait de six à huit ans et dont le quotient intellectuel mesuré par un autre instrument, variait de cent à cent vingt-cinq. Ce test fut administré sans limite de temps. Des mesures d'anxiété, un galvanomètre et The PTI Behavioral Rating Scale, qui est une mesure qualitative du comportement de l'élève, furent aussi utilisés.

L'échantillon fut divisé en quatre groupes de dix-neuf garçons et dix-neuf filles. Chaque groupe passa dans

⁴² Donald Greer et Joseph L. French, Effect of Test-Item Arrangement on Physiological and Psychological Behavior in Primary-School Children, dans The Journal of Educational Measurement, vol. 1, n° 2, livraison de décembre 1964, p. 151-153.

un ordre différent quatre formules du PII. La première formule était la formule originale, la deuxième formule présentait des item en ordre de difficulté croissante, la troisième présentait des item dont l'ordre avait été laissé au hasard et la quatrième présentait une séquence de groupes d'item composés de deux item faciles et d'un item difficile.

Les résultats d'une analyse de la variance ne montrèrent aucune différence significative entre les formules quant aux quotients intellectuels, aux mesures physiologiques ou aux mesures de comportement. Ces conclusions rejoignent donc celles de Capron et diffèrent des résultats de Mensh et de Lund qui avaient trouvé des effets significatifs de l'ordre de difficulté des item sur le score brut moyen des tests utilisés.

Plusieurs facteurs peuvent avoir affaibli cette expérience. Comme il a déjà été mentionné au sujet de l'expérience de Capron, l'apprentissage, qui se développe lorsqu'un élève passe un item plusieurs fois, est peut-être un facteur qui fait varier la performance de l'élève plus que ne le font les différents ordres de présentation des item, et pourrait donc masquer les résultats. Et que dire des petits groupes de 38 élèves pour ce genre d'expérience et du fait de limiter l'étendue des quotients intellectuels? Ne diminue-t-il pas les effets des différents ordres de difficulté? De plus les mesures physiologiques et les mesures de comportement

étaient-elles suffisamment précises pour déceler les variations possibles engendrées par les différents ordres de difficulté?

Dans les deux expériences suivantes, conduites par Munz et Smouse^{43,44}, l'ordre de difficulté des item d'un examen est étudié en relation avec le rendement des étudiants et l'anxiété qu'ils éprouvent lors d'un examen.

Munz et Smouse ont bâti deux expériences autour d'un même instrument: un examen final d'un cours d'introduction à la psychologie. Dans une étude pilote, ils ont d'abord analysé 197 item et ils en ont choisi cent dont la difficulté définie en fonction des réussites variait de 17.7% à 96.6% . Ils construisirent trois formules dont les item étaient respectivement en ordre de difficulté croissante, décroissante et au hasard. Ces formules furent ensuite administrées sans limite de temps.

43 David C. Munz et Albert D. Smouse, Interaction Effects of Item-Difficulty Sequence and Achievement-Anxiety Reaction on Academic Performance, dans Journal of Educational Psychology, vol. 59, n° 5, livraison de février 1968, p. 370-374.

44 Albert D. Smouse et David C. Munz, The Effects of Anxiety and Item Difficulty Sequence on Achievement Testing Scores, dans The Journal of Psychology, vol. 68, première demie, livraison de janvier 1968, p. 181-184.

Dans la première expérience⁴⁵, ils administrèrent ces formules à 113 étudiants divisés, au hasard en deux groupes. Les étudiants du premier groupe avaient été informés que si leur note à cet examen diminuait par rapport à leurs notes précédentes, ils devraient passer un examen oral supplémentaire parce que l'on soupçonnait qu'il y avait eu du plagiat dans les examens antérieurs. Ce traitement d'anxiété ne fut pas appliqué au deuxième groupe. Chacun des groupes fut divisé au hasard en trois sous-groupes auxquels furent administrés respectivement chacune des trois formules. Chaque sous-groupe comprenait donc environ dix-neuf étudiants.

Une analyse de la variance des scores bruts ne révéla aucune différence significative quant à l'ordre de difficulté, quant à l'anxiété ou quant à l'interaction de ces deux variables. Les auteurs se demandèrent si ces résultats étaient dus à l'inefficacité du traitement pour créer l'anxiété ou à un niveau d'anxiété si élevé chez tous ceux qui ont passé l'examen que l'anxiété additionnelle provoquée par le traitement aurait été négligeable.

⁴⁵ A.D. Smouse et D.C. Munz, The Effects of Anxiety and Item Difficulty Sequence on Achievement Testing Scores, op. cit., p. 181-184.

Dans la deuxième expérience⁴⁶, Munz et Smouse étudièrent les effets des trois ordres de difficulté liés à l'anxiété naturelle des individus, telle que mesurée par The Achievement Anxiety Test (AAT). Ce test distingue deux niveaux d'anxiété, "facilitating or debilitating to test performance".

Des 181 étudiants d'un cours d'introduction à la psychologie qui avaient subi le AAT, 120 furent choisis pour l'expérience. Parmi ceux-ci se trouvaient les dix qui avaient obtenu les scores les plus élevés de l'échelle "facilitating" et les dix qui avaient obtenu les scores les plus élevés sur l'échelle "debilitating". Les trois formules de l'examen final furent administrées au hasard aux étudiants.

Une analyse de la variance des scores bruts ne révéla aucun effet significatif des ordres de difficulté alors que l'effet de la variable anxiété était significatif à un niveau .01 ainsi que l'effet d'interaction. D'une façon générale, les étudiants qui éprouvent une anxiété du type "facilitator" réussissent mieux que les autres dans la formule en ordre de difficulté croissante ou dans la formule au hasard. Il n'y a pas de différence significative entre

⁴⁶ D.C. Munz et A.D. Smouse, Interaction Effects of Item-Difficulty Sequence and Achievement-Anxiety Reaction on Academic Performance, op. cit., p. 370-374.

les résultats des étudiants de divers niveaux d'anxiété sur la formule en ordre de difficulté décroissante.

Ces deux expériences ne rapportent donc aucune différence significative entre les ordres de difficulté croissante, décroissante et au hasard quand l'ordre est l'unique facteur considéré, mais elles montrent aussi que l'anxiété éprouvée par les étudiants peut influencer leur efficacité face à tel ordre de difficulté plutôt qu'à tel autre.

Une grande faiblesse qui marque ces expériences est le petit nombre de sujets dans les sous-groupes. Dans la première expérience de Munz et Smouse, un sous-groupe comprend environ dix-neuf étudiants. Dans la deuxième expérience, le nombre exact n'est pas donné explicitement, mais, par exemple, il est à remarquer que les dix étudiants qui éprouvaient une anxiété "facilitating" ont dû être divisés en trois pour former ainsi des sous-groupes de trois ou quatre étudiants chacun.

Dans une étude plus récente, Marso⁴⁷ s'intéressa aux effets de la présentation des items, selon leur difficulté et leur contenu, sur l'anxiété ressentie par les étudiants, sur leur rendement et sur le temps requis pour compléter un examen.

⁴⁷ Ronald N. Marso, Test Item Arrangement, Testing Time, and Performance, dans Journal of Educational Measurement, vol. 7, n° 2, livraison d'été 1970, p. 113-118.

Dans une première expérience, il analysa d'abord deux cents item du Quick World Test et en choisit cent trente-neuf dont la difficulté variait de zéro à cent pour-cent de réussite. Après avoir arrangé ces item en ordre de difficulté croissante, décroissante et au hasard, il administra ces trois formules, au hasard, à 122 étudiants d'un cours d'introduction à la psychologie de sorte que chacun des étudiants passe l'une des formules. Il n'y avait aucune limite de temps imposée, mais la période de temps requise par chacun des étudiants et son score à un test d'anxiété étaient notés.

Après une analyse de la variance des variables ordre, temps et anxiété, il parvint à la conclusion suivante:

...even when a rather lengthy examination (139 items) with a diverse range in item difficulty (0-100%) is administered as a power or achievement test, arranging items according to difficulty has little or no effect upon either required testing time or upon student performance on the examination. In addition the data indicate that students with greater or less measured test anxiety do not appear to perform differently when item difficulty arrangements are varied⁴⁸.

Dans sa deuxième expérience, il varia l'ordre de présentation des item selon leur contenu. Encore là, il n'obtint aucune différence significative quant à la performance des étudiants ou quant au temps requis pour compléter l'examen.

⁴⁸ R.N. Marso, op. cit., p. 116.

L'expérience de Marso vient donc s'ajouter aux recherches qui n'ont trouvé aucune relation significative entre l'ordre de difficulté des item d'un test et la performance des étudiants sur ce test telle qu'exprimée par le score brut total.

Les auteurs suivants ont étudié eux aussi les effets de différents ordres de difficulté sur le score brut total d'un test, mais en plus, ils ont ajouté les effets conjugués de la limite de temps imposée dans l'administration du test. Cromack et Sax⁴⁹ utilisèrent soixante-dix item des formules A et B des Henmon-Nelson Tests of Mental Ability. Ils construisirent quatre formules où l'ordre de difficulté des item était croissant, décroissant, au hasard et composé d'un item facile inséré à tous les six item pour favoriser la motivation des étudiants.

Les quatre formules furent d'abord administrées en trente minutes à quatre groupes de 69 étudiants d'un cours d'introduction en éducation. Dans une expérience parallèle, les quatre formules furent ensuite administrées en quarante-huit minutes à quatre groupes d'environ 32 étudiants d'un cours d'introduction à la psychologie. Les auteurs

⁴⁹ Theodore R. Cromack et Gilbert Sax, The Effects of Various Forms of Item Arrangement of Test Performance, dans Journal of Educational Measurement, vol. 3, n° 4, livraison d'hiver 1966, p. 309-311.

comparèrent les scores bruts totaux à l'intérieur de chacune des expériences par une analyse de variance.

Ces analyses parvinrent à montrer des différences significatives lorsque le test était administré en trente minutes, alors qu'aucune différence significative n'apparaissait lorsque le test était administré en quarante-huit minutes. Les auteurs conclurent que l'ordre de difficulté des item était important seulement s'il y avait une sévère limite de temps imposée lors de l'administration. Dans de telles circonstances, l'ordre de difficulté croissante donnerait les scores bruts les plus élevés. Les résultats ne montrèrent aucune différence significative entre l'ordre au hasard et l'ordre où un item facile était inséré à tous les six item. Bien que la performance des étudiants les plus forts sur la formule en ordre de difficulté décroissante n'était pas significativement différente de celle des plus faibles lorsque la période de temps était de trente minutes, elle était significativement plus élevée lorsque la période de temps était de quarante-huit minutes.

Cette étude fort simple qui montre la relation entre l'ordre de difficulté des item d'un test et la période de temps allouée pour passer le test aurait eu encore plus d'impact si on avait démontré que les groupes des deux expériences étaient équivalents. A défaut de ce contrôle, il est davantage possible que les différences observées

entre les formules de trente minutes et les formules de quarante-huit minutes soient dues à des différences d'échantillonnage.

Un nouvel aspect du problème est introduit par Mollenkopf⁵⁰ qui étudie les effets de différents ordres de difficulté et de différentes périodes de temps, mais non plus sur les scores bruts, mais sur la validité des item telle que définie par un coefficient de corrélation bi-sériale entre un item et les scores bruts.

Puisque cette recherche a été présentée dans la section précédente, le lecteur pourra se référer aux pages 19-22 pour se retremper au besoin dans les détails de l'expérience. Rappelons seulement que cette étude avait pour objet d'analyser les effets de la place des item, dans des tests d'aptitudes verbale et mathématique, et les effets de la durée d'administration sur les indices de difficulté et de validité des item.

Pour étudier les effets des variables indépendantes, ordre de difficulté et temps, sur la validité des item, il choisit d'utiliser une mesure de corrélation bi-sériale entre les item réussis ou non et le score total comme suit:

50 W.G. Mollenkopf, op. cit., p. 291-317.

For each group tested only under the power condition there was computed one set of item-test biserial coefficients of correlation. On the other hand, for the groups tested first under the speed condition and then under the power condition, four sets of biserial r 's were computed. For each item, both a speed response and a power response were available. Each of these was correlated with both the speed score (number right under speed) and the power score (number right) on the total test¹.

Mollenkopf résuma ses conclusions comme suit:

- a. When power data only were compared, the r 's of an item was not found to vary noticeably with different placements of the item in the test, for either the verbal or the mathematics forms.
- b. When speed and power data were compared, the speed response-speed score biserials generally exceeded the power response-power score biserials for items when drop-out on the items was high (about half the cases or more).
- c. When speed data for one form were compared with speed data for a second form in which item arrangement was different, the r 's of an item was distinctly higher when the item appeared near the end of the test than when it came at the beginning; [...]
- d. These higher biserials for items appearing late in a speeded test were found to be a reflection of the inappropriateness of the biserial coefficient when the continuous-score distribution for those attempting the item is markedly flatter than the normal curve².

L'originalité de cette partie de l'expérience de Mollenkopf est d'avoir analysé les effets conjoints de l'ordre de difficulté et de la limite de temps sur les qualités psychométriques des item. Plus précisément, il montre que

51 W.G. Mollenkopf, op. cit., p. 306.

52 W.G. Mollenkopf, op. cit., p. 311-312.

le coefficient biserial peut être biaisé par la place que l'item occupe dans un test à temps limité. Cependant il est impossible de préciser à partir de cette expérience si la validité d'un item définie autrement serait aussi influencée par la place que l'item occupe dans un tel test.

L'auteur suivant, Brenner^{53,54}, dont la recherche a été étudiée aux pages 22-25 de la section précédente, a élargi les cadres du problème encore davantage. Il administra des examens d'un cours de psychologie sur lesquels 99% des étudiants ont eu l'occasion d'atteindre et d'essayer tous les item. Il administra ces examens à plusieurs groupes en présentant plusieurs variations de l'ordre de difficulté des item.

Brenner compara les effets des différents ordres de difficulté sur la validité des item en évaluant la différence entre les coefficients rpt-bis moyens de chacune des formules.

Il parvint aux conclusions suivantes: dans trois de ses expériences, l'ordre de difficulté croissante était moins discriminante, mais non pas à un niveau significatif

⁵³ M.H. Brenner, Test Difficulty, Reliability, and Discrimination as Functions of Item Difficulty Order, thèse de doctorat, op. cit., vii-92 p.

⁵⁴ -----, Test Difficulty, Reliability, and Discrimination as Functions of Item Difficulty Order, dans Journal of Applied Psychology, op. cit., p. 98-100.

alors que, dans la quatrième expérience, elle s'avéra significativement plus discriminante. Brenner ne put expliquer ces résultats contraires.

Les résultats cumulés des expériences qui viennent d'être présentées montrent que, en ce qui concerne les tests non limités par le temps, deux auteurs, Lund et Mensh, ont trouvé des relations significatives entre l'ordre de difficulté des item et la performance de l'étudiant sur un test. Six autres expériences n'apportèrent aucun résultat significatif alors qu'une dernière faite par Brenner parvint à des résultats contradictoires qui favorisaient l'ordre de difficulté croissante en une occasion et l'ordre de difficulté décroissante en trois autres occasions.

En ce qui concerne les tests administrés à l'intérieur d'une période de temps limitée, ce qui empêche une partie des étudiants d'atteindre tous les item, l'expérience de Cromack et Sax parvint à des différences significatives de scores bruts entre des formules où seul l'ordre de difficulté était différent. Quant à Mollenkopf, il montra que les coefficients de corrélation bisériale entre les item et les scores bruts étaient influencés par la place que les item occupaient dans le test indépendamment de l'ordre général des item.

A la suite de ces résultats et des énoncés présentés antérieurement, une hypothèse de recherche sera maintenant discutée.

d. Hypothèse de recherche.

Les opinions des auteurs et les résultats de recherches assez disparates et souvent d'apparence incohérente qui viennent d'être passés en revue ne font pas le point sur le problème posé par l'ordre de difficulté des item dans un test. Ils laissent bien supposer qu'il y a encore matière à investigation, mais sans apporter des suggestions ou des faits assez précis pour dicter une ligne de conduite définie pour les recherches à venir. Cependant, ces opinions et ces recherches ne vont pas non plus à l'encontre du projet de recherche proposé au début, soit l'étude des effets de l'ordre de difficulté des item sur la valeur discriminante des item d'un test de puissance à temps limité. Au contraire, un tel projet pourrait tout au moins venir compléter l'expérience de Mollenkopf.

Il semble donc approprié de poursuivre cette recherche au niveau expérimental et de la structurer autour de l'hypothèse suivante:

Dans un test de puissance à temps limité, il n'y a pas de différence significative de la valeur discriminante des item entre des formules qui diffèrent uniquement par l'ordre de présentation de leurs item selon leur difficulté.

Formulée en fonction de l'instrument utilisé, cette hypothèse devient:

Il n'y a pas de différence significative entre la valeur discriminante des item de la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale dans sa présentation originale, la valeur discriminante de ces item dans une formule où ils sont en ordre de difficulté croissante et la valeur discriminante de ces item dans une formule où ils sont disposés au hasard.

En plus de l'intérêt théorique sous-jacent à cette hypothèse, ce problème revêt une importance pratique liée à la valeur de l'instrument utilisé. En effet la nature du problème permet de respecter la standardisation de la formule A de l'Examen intermédiaire Otis-Ottawa, de respecter ses qualités originales et de faire ainsi une étude de ce test lui-même. Le fait de respecter la standardisation n'enlève rien à l'expérience sur le plan théorique; au contraire l'Otis-Ottawa possède dans sa nature même des qualités susceptibles de mettre en évidence l'importance de l'ordre de difficulté des item. Ces qualités sont le nombre et la variété des item, l'étendue de la difficulté des item et la limite de temps imposée dans l'administration du test. Ces caractéristiques ne se retrouvent pas ensemble et à des degrés aussi prononcés dans les instruments utilisés jusqu'à maintenant pour analyser le problème de l'ordre de difficulté des item. Ce test semble donc bien choisi pour démontrer, s'il y a lieu, les effets de l'ordre de difficulté des item sur la valeur discriminante des item d'un test de puissance à temps limité.

La dernière hypothèse énoncée sera éprouvée à la suite des deux hypothèses qui ont déjà été présentées, soient:

Il n'y a pas de différence significative entre les ordres de difficulté croissante des item calculés à partir d'ordres différents de présentation des item d'un test de puissance à temps limité comme la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale administrée en trente minutes.

Il n'y a pas de différence significative entre les ordres de difficulté croissante des item d'un test de puissance à temps limité, comme la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale, administré dans une période de temps limité et les ordres de difficulté croissante des item d'un tel test alors qu'il est complété par la plupart des élèves.

Les circonstances expérimentales qui font suite à ces hypothèses seront maintenant présentées dans le chapitre qui suit.

CHAPITRE II

SCHEME EXPERIMENTAL

A partir de l'étude initiale, l'expérience s'est déroulée sur une période de six années. Des étapes successives ont permis de mieux contrôler les variables expérimentales tout en favorisant l'amélioration des méthodes statistiques. Seules les composantes les plus pertinentes de l'expérience ont été retenues. Elles seront présentées dans la description des échantillons et du déroulement de l'administration des différentes formules du test et dans la description des méthodes statistiques.

1. Les échantillons.

L'expérience fut conduite au niveau de la septième année du cours primaire parce que l'Examen intermédiaire Otis-Ottawa révèle une efficacité maximum à ce niveau, c'est-à-dire qu'il n'est ni trop facile ni trop difficile et qu'il distribue habituellement les scores des élèves de ce niveau selon une courbe normale¹. C'est aussi à ce niveau qu'il est le plus utilisé. De plus, son application systématique à tous les élèves qui devaient faire leur entrée au niveau

¹ R.-H. Shevenell, Les Examens Otis-Ottawa d'habileté mentale, Ottawa, Editions de l'Université d'Ottawa, 1948, p. 25 et 28.

secondaire à la Régionale de l'Outaouais permettait de tirer des échantillons d'une population bien définie.

Le premier échantillon, qui fit partie de l'étude initiale², fut prélevé parmi les 2349 élèves de septième année de l'année académique 1967-68 . Il se composait de cent garçons et de cent filles choisis au hasard individuellement après que la population eu été divisée selon le sexe. Ces élèves avaient déjà passé la formule A de l'Examen intermédiaire Otis-Ottawa selon les procédures habituelles.

Il n'y a pas de différence significative entre l'échantillon et la population au niveau de la moyenne d'âge, de la moyenne des scores bruts et au niveau de la variabilité des scores bruts, comme nous pouvons le constater au Tableau I à la page suivante. D'autres statistiques de cet échantillon sont présentées au Tableau X en Appendice 1 avec les statistiques correspondantes des autres échantillons. Le polygone des fréquences des scores bruts est comparé au polygone des fréquences théoriques de la courbe normale à la figure 9 de l'Appendice 2 . Ce premier échantillon a été utilisé comme critère de constance pour les mesures statistiques.

² J.G. Bonin et A. Pellerin, Ordre et degré de difficulté des item et valeur discriminante des item des Examens Otis-Ottawa d'habileté mentale, examen intermédiaire, formule A, travail présenté à l'Université d'Ottawa en mars 1968, non publié, mais présenté au Congrès de l'ACFAS, 9 novembre 1968, Ottawa, vi-51 p.

Tableau I.

Moyennes des âges, moyennes et écarts-types des scores bruts de l'échantillon sélectionné et de la population des élèves de septième année de la Régionale de l'Outaouais en 1967-68.

	M.A.	M	σ_m	σ	σ_σ
Echantillon	12 ans 11 mois	38.59	.7	9.99	.5
Population	12 ans 10.5 mois	38.30	.7	9.90	.5

Un deuxième échantillon de 258 élèves fut prélevé au hasard par classe, soit 10 classes sur 90, parmi les sixièmes années de la Régionale de l'Outaouais au mois de mai 1969. Comme cette partie de l'expérience avait pour but d'établir un ordre de difficulté croissante des item, il fallait que la plupart des élèves atteignent tous les item. En effet, cette méthode est préconisée par la plupart des auteurs qui ont écrit à ce sujet. De plus, la mesure de la difficulté des item passés dans une période de temps limitée devait être analysée plus tard dans l'expérience et n'offrait pas assez de sûreté à ce moment-là. Pour ce deuxième échantillon, la formule A de l'Examen intermédiaire Otis-Ottawa fut d'abord administrée selon les procédures standardisées. Après trente minutes, l'administration fut arrêtée, on changea les crayons des élèves pour des crayons d'une autre couleur et on leur distribua une formule B de l'Examen intermédiaire Otis-Ottawa. Les instructions suivantes leur furent aussi données:

Dans une demi-heure, vous avez pu répondre à plusieurs questions. Nous voudrions savoir combien de questions de plus vous pourriez répondre si nous vous donnions encore une demi-heure. Essayez donc d'en réussir autant que possible. Prenez garde d'aller si vite que vous fassiez des fautes. Ne vous attardez pas trop sur une question aux dépens des autres. Nous vous avons passé un autre cahier qui continue le premier. N'ouvrez pas ce deuxième cahier tant que vous n'aurez pas fini de répondre aux questions du premier cahier. Les réponses aux questions du premier cahier sont plus importantes que celles du deuxième cahier.

Cette demi-heure supplémentaire permit à 95% des élèves d'atteindre l'item 73, alors que 90% atteignaient l'item 75 . La formule B avait pour but de stimuler les élèves et de leur faire ressentir la pression de la variable vitesse dans la formule A même s'ils avaient l'opportunité d'atteindre tous les item. Cette période de temps et cette formule supplémentaire permirent donc à la plupart des élèves d'atteindre tous les item de la formule A tout en respectant le plus possible l'esprit de la standardisation. Les données premières de la formule A administrée en trente minutes et en soixante minutes sont présentées respectivement aux Tableaux X et XI en Appendice 1 . Le polygone de fréquences des scores obtenus en trente minutes est présenté à la figure 10 à l'Appendice 2 .

La difficulté des item fut calculée selon la proportion des élèves qui avaient réussi l'item par rapport à ceux qui l'avaient essayé. Cette proportion fut ensuite transformée par la formule suivante pour corriger l'effet de la chance³:

$$P_c = \frac{K_p - 1}{K - 1}$$

P_c: proportion de réussite corrigée pour les succès dus à la chance

p: proportion initiale de réussite

K: nombre de choix possibles

³ J.P. Guilford, Psychometric Methods, Toronto, McGraw-Hill Book Company, 1954, p. 420.

Les premiers indices de difficulté ainsi que les indices corrigés des item passés en soixante minutes apparaissent au Tableau II à la page 65 . Les item de la formule A ont ensuite été placés en ordre de difficulté croissante selon les indices corrigés. Il est intéressant de noter que ce nouvel ordre de difficulté, présenté au Tableau III à la page 66, n'enlève rien à la qualité spirale-omnibus de l'Otis-Ottawa, au contraire, il l'améliore.

Par mesure de précaution, l'échantillon fut divisé en deux et un ordre de difficulté croissante des item fut calculé à partir des résultats de chacune des parties. Ces deux ordres de difficulté furent comparés par la corrélation rho et nous avons obtenu un coefficient de corrélation de .969 entre eux, alors que nous obtenions des coefficients de .990 et .992 respectivement entre ces deux ordres de difficulté et l'ordre de difficulté calculé à partir de l'échantillon total.

Une nouvelle formule que nous appellerons formule expérimentale I fut construite selon cet ordre de difficulté. Une autre formule que nous appellerons formule expérimentale II fut construite selon une présentation au hasard des item. Ces deux formules ont ensuite été imprimées, de même que la formule A originale, de sorte qu'il n'y ait aucune différence au niveau de la présentation matérielle des formules. Des exemplaires de ces formules sont présentés à l'Appendice 3

Tableau II.

Proportions de succès et proportions de succès corrigées pour les réussites dues à la chance, sur les item de la formule A de l'Examen intermédiaire Otis-Ottawa administrée en soixante minutes à 258 élèves de sixième année de la Régionale de l'Outaouais.

Item	P	PC	Item	P	PC	Item	P	PC
1	.872	.840	26	.837	.795	51	.609	.476
2	.950	.937	27	.465	.330	52	.500	.375
3	.818	.772	28	.833	.833	53	.630	.577
4	.837	.795	29	.756	.721	54	.595	.492
5	.682	.602	30	.764	.705	55	.430	.335
6	.934	.920	31	.764	.705	56	.219	.219
7	.884	.855	32	.450	.340	57	.746	.620
8	.907	.882	33	.209	.000	58	.512	.390
9	.795	.742	34	.686	.607	59	.320	.093
10	.888	.875	35	.508	.385	60	.531	.531
11	.601	.500	36	.550	.437	61	.277	.095
12	.457	.366	37	.570	.426	62	.836	.780
13	.783	.783	38	.267	.082	63	.431	.287
14	.403	.252	39	.628	.574	64	.350	.187
15	.926	.907	40	.488	.360	65	.242	.052
16	.667	.582	41	.326	.157	66	.227	.032
17	.752	.690	42	.798	.747	67	.333	.110
18	.531	.412	43	.550	.437	68	.120	.120
19	.775	.712	44	.798	.747	69	.319	.147
20	.775	.736	45	.775	.550	70	.593	.490
21	.659	.601	46	.791	.737	71	.065	.065
22	.857	.820	47	.531	.412	72	.150	.150
23	.558	.447	48	.682	.573	73	.492	.406
24	.442	.302	49	.554	.554	74	.383	.230
25	.725	.655	50	.283	.102	75	.103	.000

Tableau III.

Item de la formule A de l'Examen intermédiaire Otis-Ottawa
replacés en ordre de difficulté croissante d'après
la performance de 258 élèves de sixième année de
la Régionale de l'Outaouais.

Item	Ordre	Item	Ordre	Item	Ordre
2	1er	57	26è	12	51è
6	2è	34	27	40	52
15	3	5	28	32	53
8	4	21	29	27	54
10	5	16	30	55	55
7	6	39	31	24	56
1	7	48	32	63	57
28	8	53	33	14	58
22	9	45	34	74	59
4	10	49	35	56	60
26	11	60	36	64	61
13	12	11	37	41	62
62	13	54	38	69	63
3	14	70	39	72	64
42	15	51	40	68	65
44	16	23	41	67	66
9	17	36	42	50	67
46	18	43	43	61	68
20	19	37	44	59	69
19	20	18	45	38	70
29	21	47	46	71	71
30	22	73	47	65	72
31	23	58	48	66	73
17	24	35	49	33	74
25	25	52	50	75	75

et la position des item selon ces trois formules est présentée au Tableau IV à la page 68 . Dans ce tableau et dans ceux qui vont suivre, les item sont d'abord identifiés par la position qu'ils occupent dans la formule A. Ainsi lorsqu'il sera fait mention par exemple de l'item 12, l'item désigné sera celui qui apparaît au douzième rang dans la formule A, au cinquante et unième rang dans la formule expérimentale I et au quinzième rang dans la formule expérimentale II.

Ces trois formules ont été utilisées pour comparer des ordres de difficulté croissante établis à partir d'ordres différents de présentation des item et pour éprouver ainsi les deux premières hypothèses de cette recherche. Elles devaient permettre aussi de comparer les effets de ces trois ordres sur la valeur discriminante des item, ce qui se rapporte à la troisième hypothèse.

Pour la partie finale de l'expérience, trois nouveaux échantillons ont été prélevés de la population urbaine et semi-urbaine de la Régionale de l'Outaouais. Certaines commissions scolaires rurales n'ont pas permis à leurs élèves de passer le test, et par conséquent elles ont été éliminées de l'expérience et la population a été forcément restreinte aux villes et aux villages de la Régionale. Ces trois échantillons ont été tirés au hasard par classe de la

Tableau IV.

Rangs occupés par les item dans les formules A, expérimentale I et expérimentale II de l'Examen intermédiaire Otis-Ottawa.

Item	Rangs des item dans les formules			Item	Rangs des item dans les formules		
	A	Ex. I	Ex. II		A	Ex. I	Ex. II
1	1ère	7ème	2ème	39	39ème	31ème	7ème
2	2	1	55	40	40	52	44
3	3	14	34	41	41	62	72
4	4	10	71	42	42	15	31
5	5	28	22	43	43	43	45
6	6	2	57	44	44	16	3
7	7	6	46	45	45	34	56
8	8	4	25	46	46	18	14
9	9	17	16	47	47	46	74
10	10	5	6	48	48	32	21
11	11	37	12	49	49	35	54
12	12	51	15	50	50	67	67
13	13	12	17	51	51	40	5
14	14	58	11	52	52	50	33
15	15	3	68	53	53	33	38
16	16	30	58	54	54	38	24
17	17	24	75	55	55	55	69
18	18	45	47	56	56	60	49
19	19	20	28	57	57	26	40
20	20	19	73	58	58	48	61
21	21	29	60	59	59	69	50
22	22	9	53	60	60	36	18
23	23	41	9	61	61	68	30
24	24	56	13	62	62	13	65
25	25	25	51	63	63	57	59
26	26	11	35	64	64	61	29
27	27	54	66	65	65	72	19
28	28	8	36	66	66	73	42
29	29	21	32	67	67	66	62
30	30	22	64	68	68	65	26
31	31	23	48	69	69	63	70
32	32	53	41	70	70	39	23
33	33	74	8	71	71	71	39
34	34	27	10	72	72	64	20
35	35	49	63	73	73	47	1
36	36	42	43	74	74	59	27
37	37	44	52	75	75	75	4
38	38	70	37				

population de septième année de l'année académique 1969-70, et se composaient respectivement de 316, 313 et 301 élèves.

Même si l'échantillonnage au hasard devait assurer l'équivalence de ces trois échantillons, le test Culture Fair, Scale 2, Form A fut administré pour rendre cette équivalence explicite. Ce test fut choisi comme critère d'équivalence parce qu'il est un test d'intelligence générale et qu'il se rapproche ainsi de l'Otis, et parce qu'il est un test non verbal qui risque moins d'introduire des variables confondantes dans l'expérience. Il permet donc d'être utilisé comme critère d'équivalence sans interférer dans l'expérience. Les résultats que les trois échantillons ont obtenus au Culture Fair sont présentés au Tableau XII en Appendice 4 et une description statistique de leur équivalence est présentée au Tableau XIII en Appendice 4. Les moyennes (26.0, 25.4 et 26.2), les écarts-types (5.09, 5.81 et 5.28) et les distributions de scores bruts ne sont pas significativement différents d'un groupe à un autre.

Ces troisième, quatrième et cinquième échantillons passèrent ensuite respectivement les formules A, expérimentale I et expérimentale II. Comme deux formules différentes ne sont jamais administrées à un même groupe, il n'y a pas d'apprentissage d'une formule à une autre, comme nous l'avons retrouvé dans certaines expériences rapportées au chapitre précédent. Les statistiques de ces trois derniers

échantillons sont présentées au Tableaux X et XI à l'Appendice 1 alors que les polygones de fréquences des scores bruts apparaissent aux figures 11, 12 et 13 à l'Appendice 2 .

Quelque cinq semaines après cette administration, environ cent élèves de chacun des groupes répétèrent l'expérience. Bien qu'aucune analyse d'item ne fut complétée à partir de cette nouvelle administration, des coefficients de constance furent calculés entre les scores bruts obtenus lors des deux administrations. Le coefficient de constance de la formule A est de .809 lorsqu'elle est administrée en trente minutes et de .862 lorsqu'elle est administrée en soixante minutes. La formule expérimentale I (en ordre de difficulté croissante) révèle des coefficients de .894 en trente minutes et de .896 en soixante minutes alors que la formule expérimentale II (en ordre au hasard) révèle des coefficients de .851 en trente minutes et de .884 en soixante minutes.

2. Les méthodes statistiques.

Comme l'expérience consiste à faire l'analyse de la constance de l'ordre de difficulté d'une part et les effets de différents ordres de difficulté sur la valeur discriminante des item d'autre part, les méthodes analytiques seront regroupées sous les deux titres correspondants, soient le

calcul de l'ordre de difficulté des item et ensuite le calcul de la valeur discriminante des item.

a. Le calcul de l'ordre de difficulté des item.

La difficulté d'un item a été calculée selon la proportion de ceux qui ont réussi l'item par rapport à ceux qui l'ont essayé, moins la proportion de succès due à la chance. Cette proportion corrigée de succès sert d'indice de difficulté des item. Nous retrouvons donc une série de 75 indices de difficulté pour le premier échantillon, qui a passé la formule A en trente minutes. Il y a aussi deux séries de 75 indices de difficulté pour chacun des quatre autres échantillons puisqu'alors les item ont été évalués à partir de deux administrations: une administration en trente minutes et une administration en soixante minutes des formules A, expérimentale I ou expérimentale II. Chacune des séries d'indices fut ensuite placée en ordre de difficulté croissante. Au total, il y a donc neuf ordres de difficulté croissante des item auxquels viennent s'ajouter l'ordre de la formule A originale et l'ordre au hasard. Ces onze ordres de difficulté des item furent comparés entre eux par des coefficients de corrélation ρ pour lesquels l'erreur-type d'estimation est de $\frac{1}{R-1}$, soit $\frac{1}{75-1}$ ou .116.

Pour éprouver la constance des ordres de difficulté, chacun des échantillons a été divisé en deux et l'ordre obtenu à partir d'une moitié de l'échantillon a été comparé aux ordres obtenus à partir de l'autre moitié et à partir de l'ensemble de l'échantillon. De cette analyse de constance, seulement les coefficients de corrélation rho seront présentés.

Les coefficients de corrélation rho calculés dans cette expérience pourraient être plus élevés et plus stables que ceux qui sont calculés habituellement entre les rangs que des individus obtiennent lors de deux examens. En comparant par une corrélation rho les rangs attribués à des items par deux groupes différents, on diminue ou élimine deux sources de variabilité qui contaminent généralement un coefficient rho. D'une part, en substituant dans les calculs habituels une série d'items invariables à des individus qui de nature sont variables, une première source de variabilité est éliminée. D'autre part, en substituant des groupes, dont la moyenne tend à se stabiliser en fonction du nombre d'individus, à des examens qui ne sont pas parfaitement constants, l'effet d'une deuxième source de variabilité peut être diminué. Ainsi on peut s'attendre à des coefficients rho plus élevés, plus précis et plus stables que ceux que l'on retrouve habituellement lorsque l'on compare les rangs d'individus.

- b. La mesure de la valeur discriminante des item.

La valeur discriminante d'un item est définie en fonction du score brut total des élèves qui ont atteint l'item: un item est discriminant s'il est réussi par ceux qui ont un score brut élevé et failli par ceux qui obtiennent un score brut peu élevé.

À partir de cette définition, l'analyse probit a été utilisée pour mesurer cette valeur discriminante. Cette méthode a été préférée aux corrélations bi-sérielles et de point bi-sériel, plus couramment utilisées, parce que celles-ci sont biaisées par la position qu'un item occupe dans un test à temps limité. En effet, Conrad⁴ et Mollenkopf⁵ montrent que le coefficient de corrélation bi-sériale tend à grandir lorsqu'il est utilisé pour mesurer la valeur discriminante d'un item placé à la fin d'un tel test. De plus, ce défaut était apparent chez les coefficients bi-sériaux et de point bi-sériel calculés précédemment à partir des données de la présente

⁴ H.S. Conrad, Characteristics and Uses of Item-Analysis Data, dans Psychological Monographs, vol. 62, no 8, livraison de 1948, p. 17.

⁵ W.G. Mollenkopf, An Experimental Study of the Effects on Item-Analysis Data of Changing Item Placement and Test Time Limit, dans Psychometrika, vol. XV, no 3, livraison de septembre 1950, p. 312.

recherche⁶. Ainsi, puisque ces corrélations sont contaminées par la variable indépendante, l'ordre de difficulté des item, on ne peut les utiliser pour mesurer la variable dépendante, la valeur discriminante des item, sans introduire une erreur constante.

L'analyse probit fut donc choisie parce que sa mesure de la valeur discriminante s'est avérée plus indépendante de la position de l'item dans le test. Cette méthode utilisée surtout en toxicologie fut introduite en psychométrie par Finney⁷ qui en demeure le principal propagandiste⁸. Lorsqu'elle est appliquée à la psychométrie, l'analyse probit compare les résultats à un item selon les postulats suivants: (1) l'habileté mentale sous-jacente à la performance sur le test est continue et se distribue normalement et (2) la probabilité de répondre correctement à un item est une fonction ogive normale de l'habileté de l'élève. Ce dernier postulat est illustré à

6 André R. Pellerin, L'ordre de difficulté croissante des item de la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale: sa constance et ses effets sur la valeur discriminante des item, rapport intermédiaire (inédit) présenté à la Faculté de Psychologie de l'Université d'Ottawa, 1973, xv-179 p.

7 D.J. Finney, The Application of Probit Analysis to the Results of Mental Tests, dans Psychometrika, vol. 9, no 1, livraison de mars 1944, p. 31-39.

8 _____, Probit Analysis, London, Cambridge University Press, 1971, xv-333 p.

la Figure 1 qui représente la fonction ogive normale pour un item donné. Le score brut de l'élève est porté en abscisse et la probabilité de succès est portée en ordonnée. Nous pouvons remarquer que plus le score brut augmente, plus la probabilité de succès augmente. Nous reviendrons à cette figure plus tard.

Cette probabilité de succès d'un individu sur un item ne peut être connue directement à partir de la réponse puisque celle-ci doit être interprétée d'une manière dichotomique, c'est-à-dire elle est bonne ou mauvaise. L'analyse probit contourne cette difficulté en se basant sur le postulat suivant: tous les élèves qui ont un même score brut possèdent un même degré d'habileté mentale et par conséquent ils sont favorisés par la même probabilité de succès sur un item donné. Selon les lois du hasard, lorsque les réponses de ces individus sur un item sont compilées, on devrait retrouver pour le groupe total une proportion de succès égale à la probabilité de succès de chacun des individus. Par exemple, en compilant les réponses de cent individus qui auraient tous 30 chances sur 100 de réussir l'item, on devrait retrouver trente individus qui auraient effectivement réussi l'item et soixante-dix qui l'auraient failli. Ce postulat nous permet aussi de faire le raisonnement inverse. Si trente élèves sur cent qui possèdent un même score brut réussissent

SCHEME EXPERIMENTAL

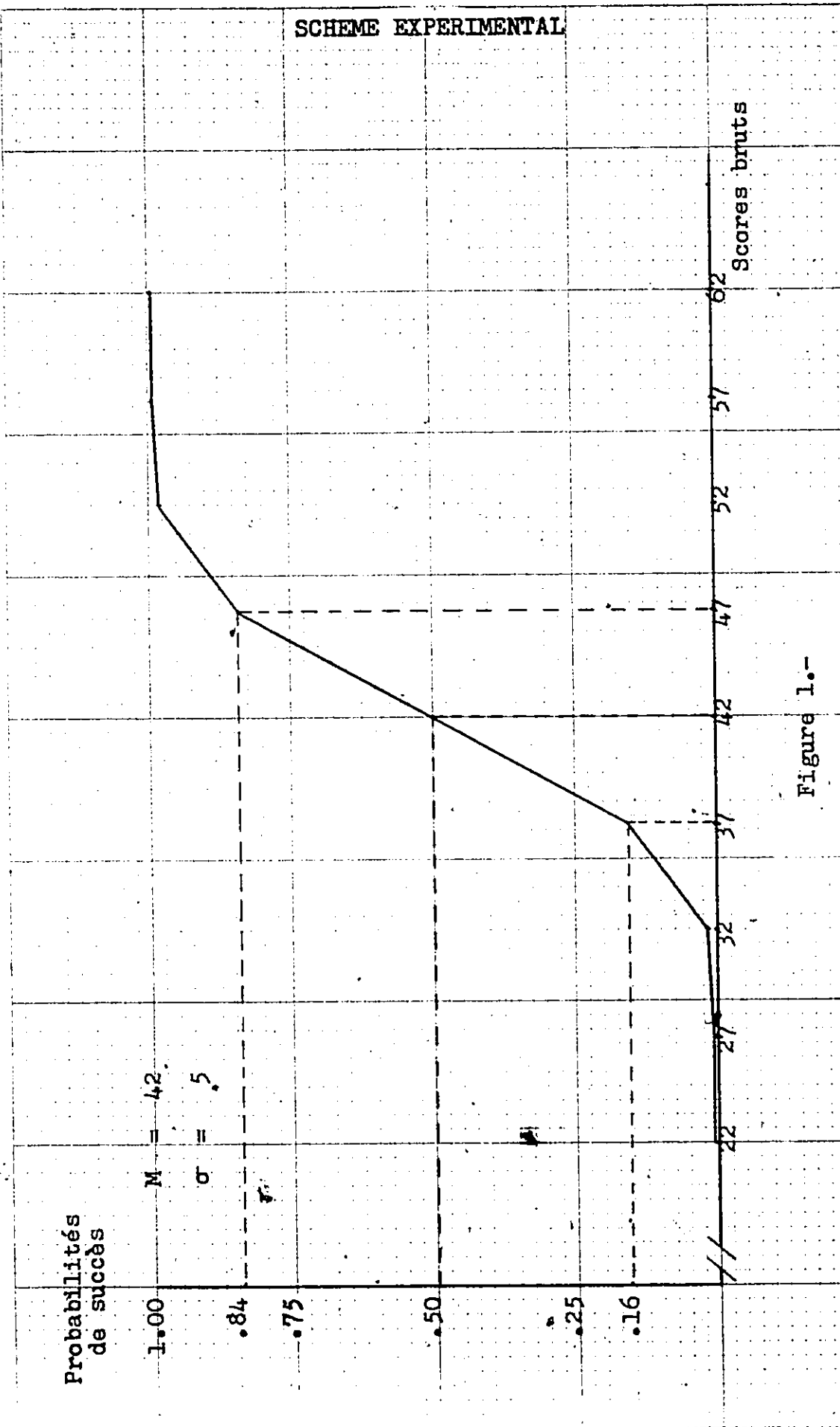


Figure 1.-

Probabilités de succès sur un item quelconque en fonction des scores bruts totaux, fonction ogive normale.

un item, nous pouvons affirmer que chacun de ces cent élèves possédait une probabilité de succès égale à .30 .

Pour un item donné, l'analyse probit calcule la proportion de succès pour chacun des groupes de niveau successif d'habileté et peut reproduire ainsi une fonction comparable à la fonction ogive normale énoncée dans le deuxième postulat et reproduite à la Figure 1.

Pour faciliter l'analyse de l'item, cette fonction ogive est transformée en une droite en traduisant les pourcentages de l'ordonnée en scores-types auxquels est ajoutée une constante de 5 pour éliminer les scores-types négatifs, ex.: .30 devient -.53 en score-type et 4.47 en score probit. La fonction se présente maintenant sous la forme d'une ligne de régression tel qu'on peut le constater à la Figure 2 à la page suivante.

A partir de cette droite, l'analyse probit révèle trois qualités de l'item: son niveau de difficulté, sa valeur discriminante et son degré de conformité au 2^{eme} postulat, la fonction ogive normale. Cette méthode est illustrée au Tableau V à la page 79 où apparaissent les données brutes, la progression des calculs et les résultats de l'analyse de l'item 12 de la formule A. La Figure 3 reproduit les mêmes données sous forme de proportions de succès en fonction des scores bruts. Cette figure devrait reproduire une fonction ogive normale, mais cette dernière est difficile à distinguer à cause des fluctuations des données.

SCHEME EXPERIMENTAL

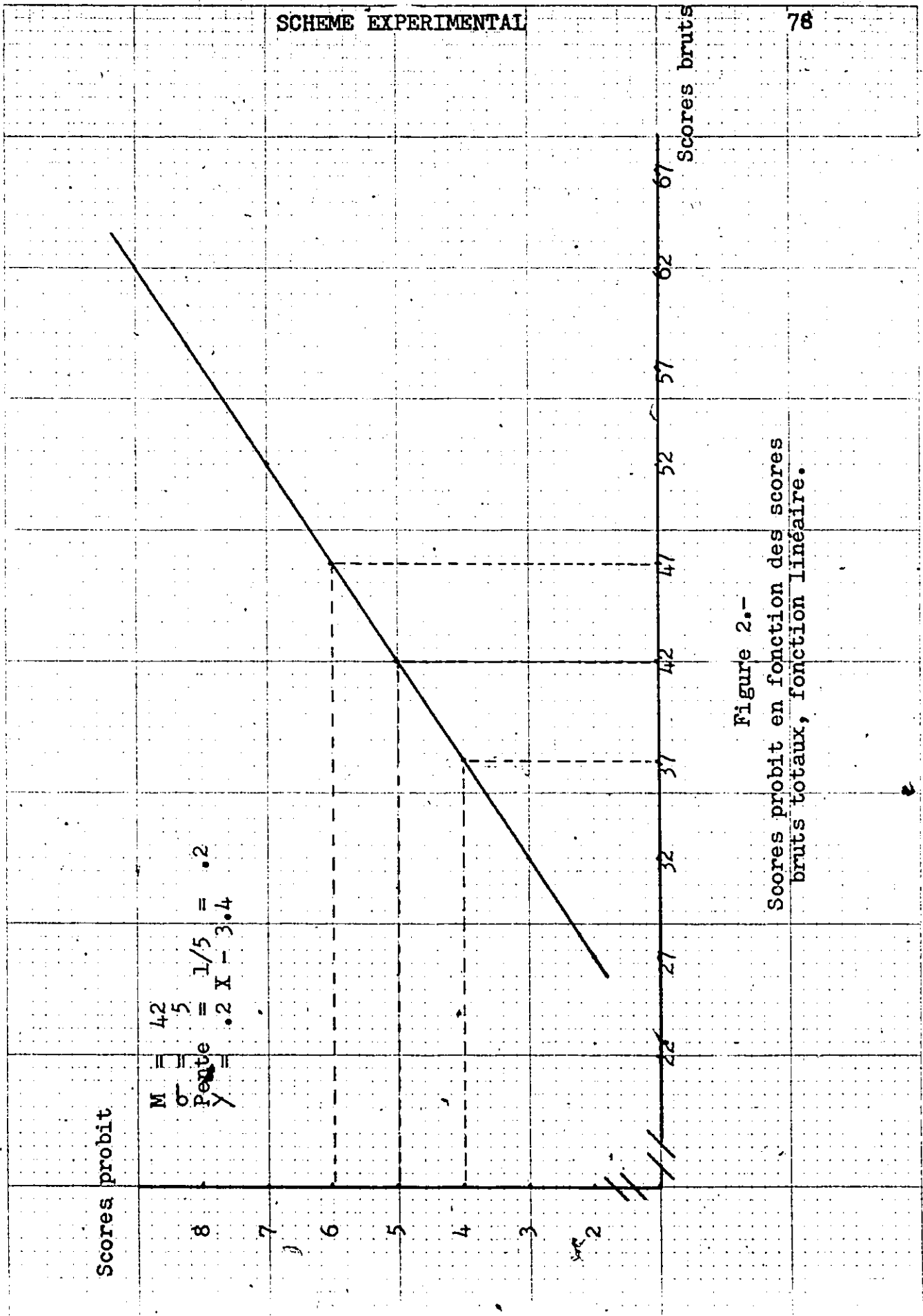


Figure 2.-
 Scores probit en fonction des scores bruts totaux, fonction linéaire.

Tableau V.

Analyse probit de l'item 12 de la formule A à partir des données brutes du troisième échantillon

Points milieux des classes de scores bruts	20.94	27	32	37	42	47	52	59.83
Nombre d'élèves qui ont réussi l'item	7	7	13	18	18	23	18	21
Nombre d'élèves qui ont atteint l'item	34	54	47	54	43	34	27	23
Proportions de réussite	.21	.13	.28	.33	.42	.68	.67	.91
Score probit	4.19	3.87	4.42	4.56	4.80	5.47	5.44	6.34
<p>Equation probit: $Y = 2.524 + 0.058X$ $X^2 = 7.85$ $dl = 6$ $X^2_{.99}(6) = 16.81$ Degré de difficulté: 42.66 Ecart-type: 17.22 Valeur discriminante: $1/ = 0.058$</p>								

Test "chi deux" entre les scores probit empiriques et les scores probit théoriques de la fonction ogive normale.

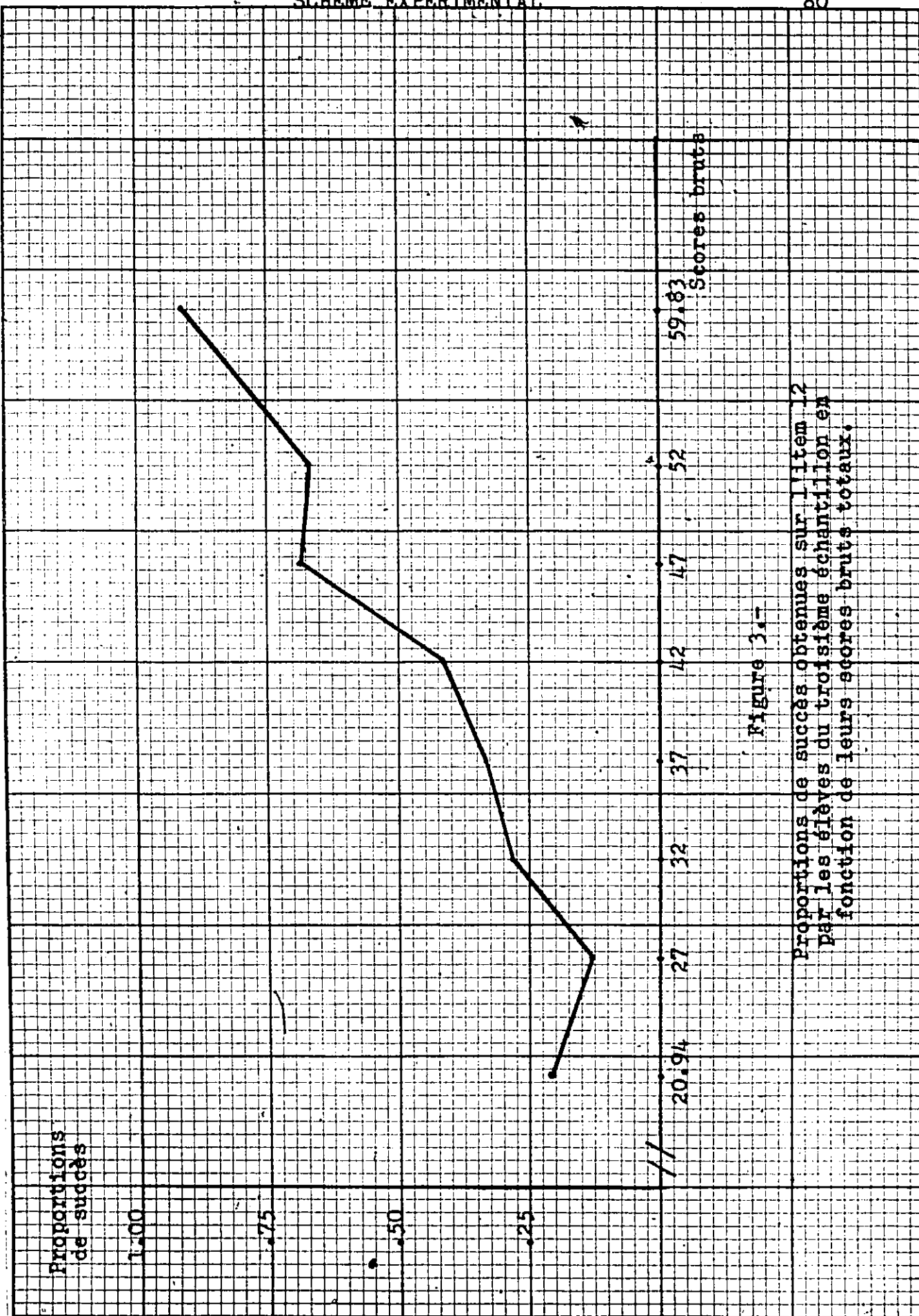


Figure 3.

Proportions de succès obtenues sur l'item 12 par les élèves du troisième échantillon en fonction de leurs scores bruts totaux.

Lorsque les proportions de succès sont transformées en score probit à la Figure 4, la fonction qui devrait être représentée par une droite est plus facile à voir malgré les fluctuations. A partir de cette distribution de scores probit en fonction des scores bruts totaux, l'analyse probit trace la droite la mieux adaptée à tous ces points. Le test "chi deux" permet ensuite d'évaluer si les écarts à la droite peuvent être attribués à des fluctuations d'échantillonnage ou s'ils doivent être attribués à un facteur quelconque. Donc, si le test "chi deux" est significatif, l'item doit être rejeté parce qu'un facteur extérieur l'empêche de reproduire la fonction ogive exigée par l'un des postulats.

On utilise habituellement l'analyse probit dans l'étude d'item à une seule réponse. Dans cette recherche, elle est employée dans l'analyse d'item à choix multiples. L'élément chance inhérent à ces item en rend l'analyse plus difficile et moins précise. Si cet élément chance avait une influence exagérée sur la fonction ogive, il serait possible de la détecter visuellement à partir des proportions de succès et ensuite elle pourrait faire augmenter le test "chi deux" au point de le rendre significatif. Advenant ce dernier cas, l'analyse probit devrait être rejetée comme méthode d'analyse de cet item.

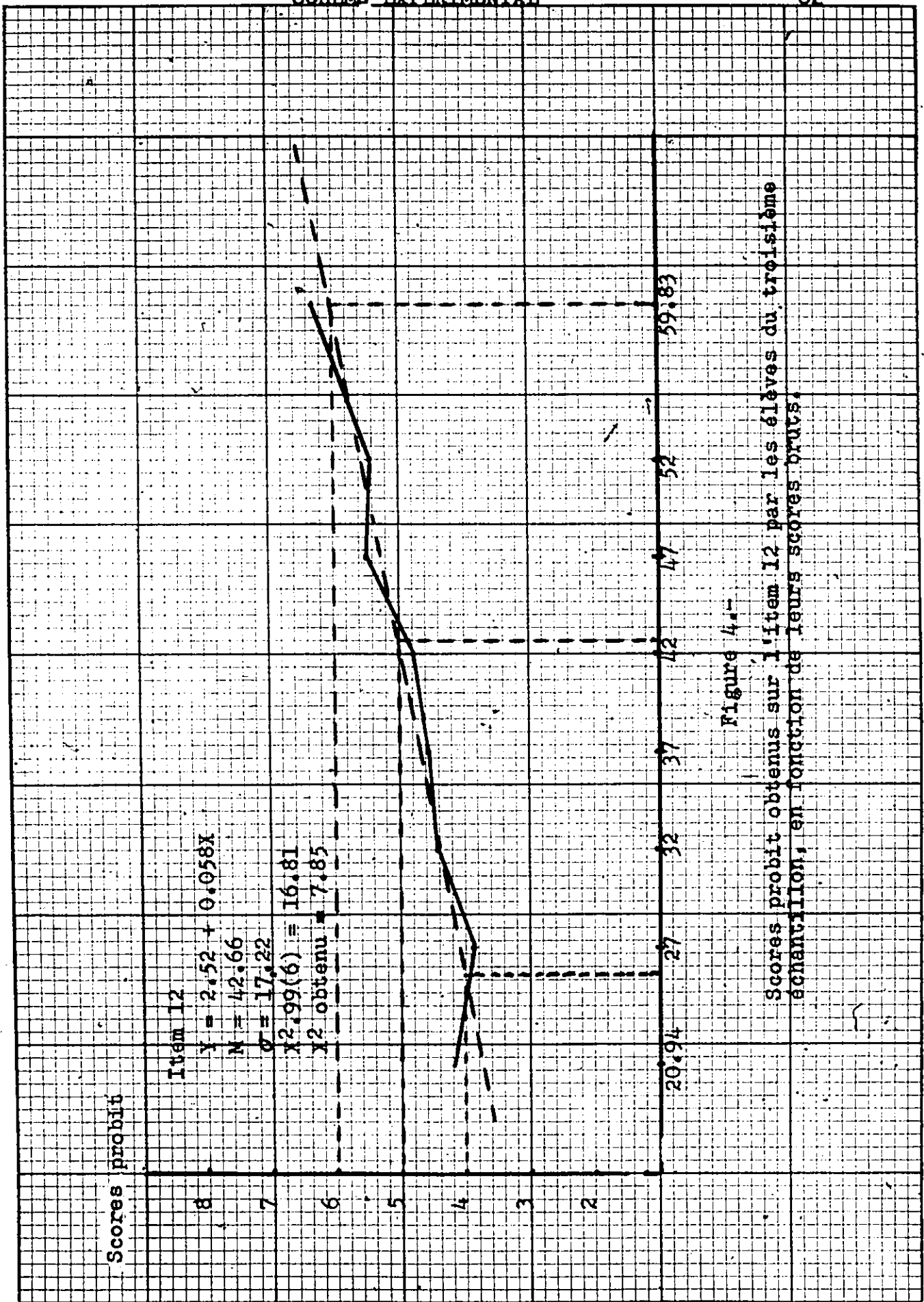


Figure 4.-

Scores probit obtenus sur l'item 12 par les élèves du troisième échantillon, en fonction de leurs scores bruts.

La pente de la droite peut être positive, nulle ou négative selon la distribution des données. Si elle est nulle ou négative, l'item doit être rejeté parce qu'il ne satisfait pas au postulat de l'ogive normale. Si la pente est positive, c'est-à-dire si l'item est de plus en plus réussi à mesure que les scores bruts totaux augmentent, la droite traversera à un endroit quelconque l'ordonnée 5 en score probit. Le score brut qui se situe sous ce point est interprété comme la moyenne de l'item et comme un indice de la difficulté de l'item. Dans l'exemple, cet indice est de 41.55 . Ceux qui auraient obtenu ce score brut total auraient eu 50% de chance de réussir l'item.

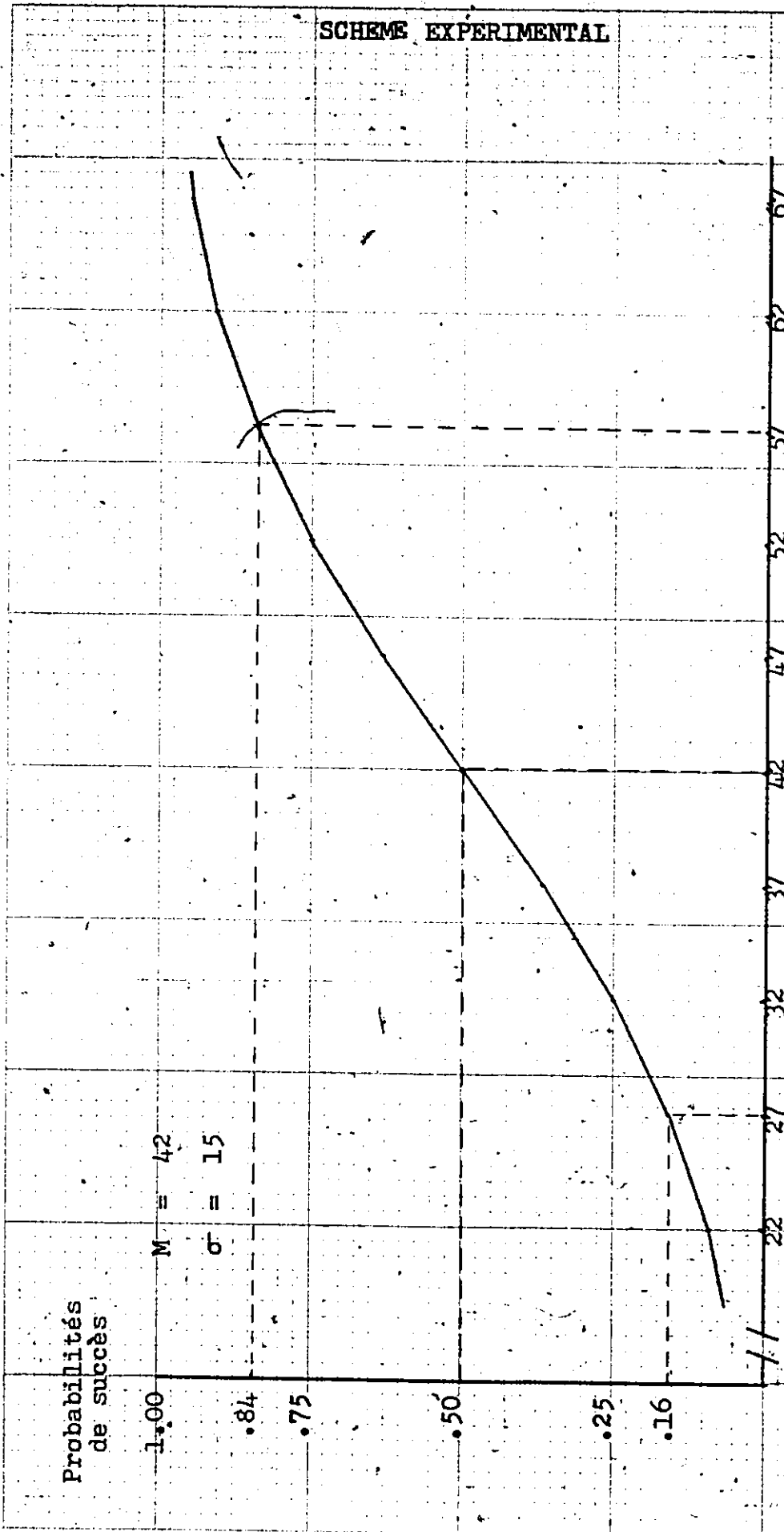
Cette méthode permet d'évaluer les items les plus faciles et les plus difficiles. En effet, puisque la courbe étudiée est une droite, il est possible de la prolonger dans un sens ou dans l'autre jusqu'au point de rencontre hypothétique qui peut se trouver à l'extérieur des limites des scores bruts totaux possibles pour le test Otis: ex. -50 ou 175 . Ces indices de difficulté pourront ensuite être placés en ordre de difficulté croissante.

Dans cette relation linéaire entre les scores probit et les scores bruts, l'écart-type de l'item correspond à l'écart-type des scores probit. Cet écart-type de l'item peut être mesuré graphiquement puisqu'il est égal à la distance entre les projections en abscisse des deux points

qui sont à l'intersection de la droite et des niveaux 4 et 5 de l'ordonnée ou des niveaux 5 et 6 . L'écart-type de l'item est donc exprimé en unité de score brut et les élèves dont le score brut se situe à moins un écart-type de la moyenne ont 15.87% de chance de réussir l'item alors que ceux dont le score brut se situe à plus un écart-type ont 84.13% de chance de réussir l'item. Puisque l'écart-type des scores probit est égal à une unité, la pente de la droite s'exprime par le rapport $1/\sigma$; dans l'exemple donné, l'écart-type de l'item est de 15.91 et la pente de la droite de $1/15.91$ ou .0628 . Si la pente est très faible, l'écart-type pour dépasser les limites des scores bruts possibles sur le test. Par exemple, il est possible de trouver un écart-type de 100 et une pente de .01 .

La pente de la droite ou l'écart-type de l'item deviennent des indices de la valeur discriminante de l'item. En effet, plus la pente est douce, autrement dit plus l'écart-type est grand, plus la capacité de discrimination diminue. Pour illustrer ceci il est possible de comparer les figures 1 et 2 aux figures 5 et 6 correspondantes. Les item comparés produisent une fonction ogive normale et ont une moyenne constante de 42; ils ont donc le même niveau de difficulté. Cependant la pente passe de .20 à .067 ou l'écart-type passe de 5 unités de scores bruts à 15 unités. De la figure 2 à la figure 6, la pente s'adoucit, se :

SCHEME EXPERIMENTAL



Scores bruts

Figure 5.-

Probabilités de succès sur un item quelconque en fonction des scores bruts totaux lorsque la moyenne de l'item est de 42 et son écart-type de 15.

SCHEME EXPERIMENTAL

Scores probit

$$M = 42$$

$$\sigma_y = 1.5 \quad \text{Pente} = 1/15 = .066$$

$$y = .066 X + 2.2$$

8

7

6

5

4

3

2

1

0

67

62

57

52

47

42

37

32

27

22

17

12

7

2

Scores bruts

Figure 6.-

Scores probit sur un item quelconque en fonction des scores bruts totaux lorsque la moyenne de l'item est de 42 et son écart-type de 1.5.

rapproche de l'horizontale. Comparons maintenant la figure 1 à la figure 5. Selon la figure 1, les élèves qui auraient un score brut de 32 n'auraient que 2.28% de chance de réussir cet item alors que ceux qui auraient un score brut de 52 auraient 97.72% de chance de le réussir. Ce niveau de discrimination, que l'item de la figure 1 parvient à atteindre entre les groupes de scores bruts 32 et 52, ne pourrait être atteint par l'item de la figure 5 qu'entre les groupes de scores bruts 12 et 72.

Dans cette recherche, un item peut avoir cinq indices de valeur discriminante, soit un par groupe. Le but de cette partie de l'étude est de déterminer s'il y a une différence significative entre ces indices. La signification de ces différences s'évalue par une analyse de la variance comme le décrit Finney⁹. On peut calculer si les données représentées par deux indices différents peuvent être représentées par deux indices identiques sans qu'il y ait un accroissement significatif de la somme des carrés des écarts existant entre les données des deux groupes et leur nouvelle droite respective.

Pour illustrer cette méthode, reprenons comme exemple les résultats de l'item 12 analysés dans le groupe III et comparons-les aux résultats de ce même item analysés

⁹ Idem, ibid., p. 100-109.

dans le groupe IV et V . Les données de cette comparaison apparaissent au Tableau VI à la page suivante et les droites probit sont illustrées à la figure 7 à la page 90 . Les pentes 0.058, 0.073 et 0.042 sont-elles significativement différentes? Poursuivons dans le sens de l'hypothèse nulle en supposant que ces pentes ne sont pas significativement différentes les unes des autres, qu'elles représentent une même "vraie" pente et que les différences observées sont attribuables à des erreurs d'échantillonnage. A partir de cette hypothèse, les calculs de l'analyse probit sont refaits avec une nouvelle restriction: les pentes des trois droites doivent être identiques. On calcule ainsi trois droites parallèles qui sont le mieux adaptées respectivement à chacun des groupes, tel qu'illustré à la figure 8 à la page 91, et on obtient trois nouvelles équations de pentes identiques. A cause de la restriction, ces nouvelles droites ne devraient pas être aussi bien adaptées aux données que l'étaient les droites initiales. On évalue donc par un test F si l'accroissement de la variance dû au parallélisme des droites est significatif. Si cet accroissement est significatif, ceci indique que les pentes initiales ne peuvent être représentées par une même pente parce qu'il existe des différences significatives entre elles. Cette comparaison entre les droites probit s'est faite par

Tableau VI.

Résultats de l'analyse probit de l'item 12
pour les groupes III, IV et V considérés
un à un (hétérogénéité) ou tous ensemble (parallélisme)

	Hétérogénéité	Parallélisme
Groupe III		
Equation probit:	$Y = 2.52 + 0.058X$	$Y = 2.64 + 0.055$
Pente:	0.058	0.055
Somme des carrés:	7.85	
Degrés de liberté:	6	
Groupe IV		
Equation Probit:	$Y = 2.12 + 0.073X$	$Y = 2.83 + 0.055$
Pente:	0.073	0.055
Somme des carrés:	10.83	
Degrés de liberté:	6	
Groupe V		
Equation probit:	$Y = 3.26 + 0.042X$	$Y = 2.85 + 0.055$
Pente:	0.042	0.055
Somme des carrés:	13.18	
Degrés de liberté:	7	
Somme des carrés:	31.67 (hétérogénéité)	39.35 (totale)

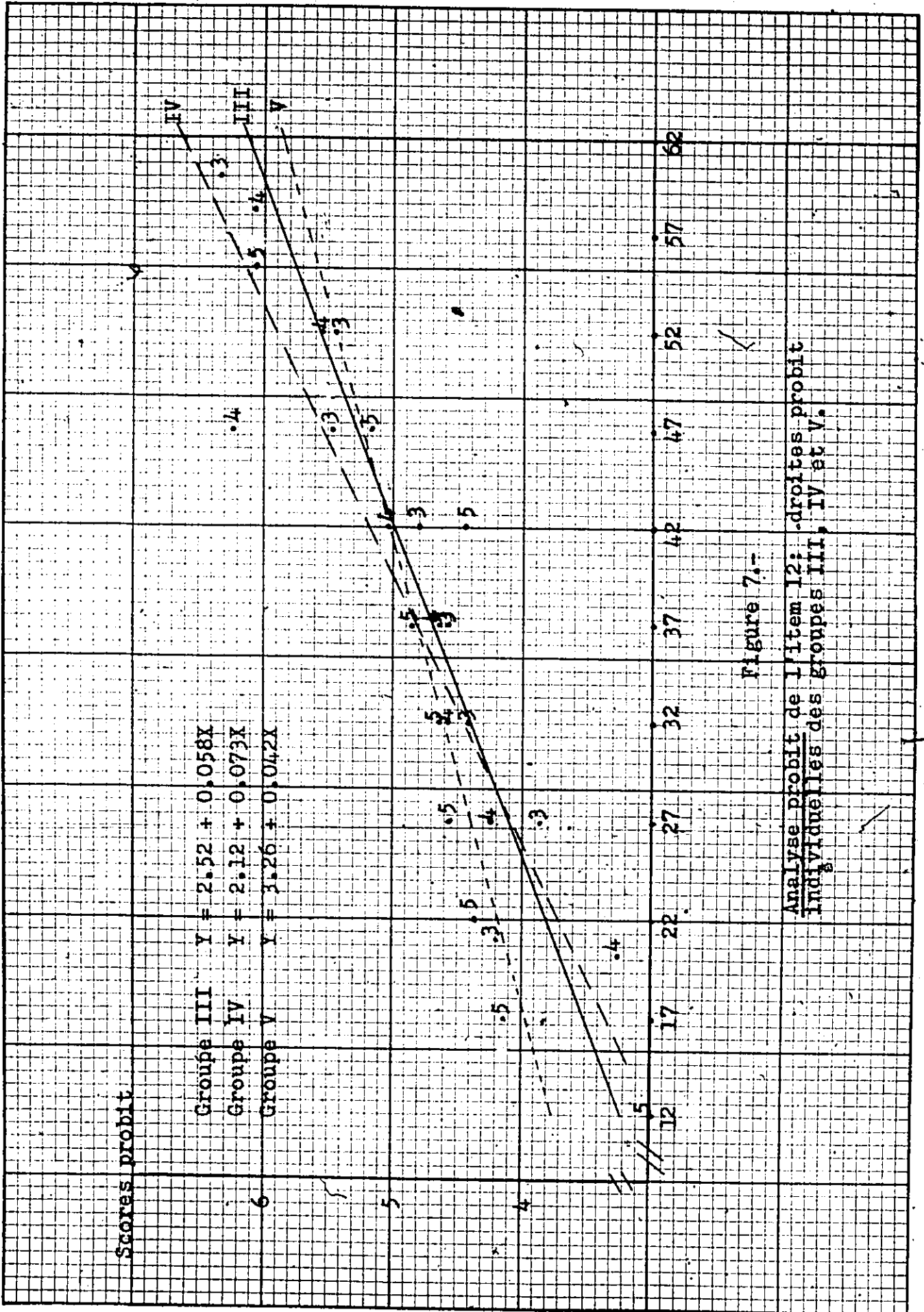


Figure 7.-

Analyse probit de l'item 12: droites probit individuelles des groupes III, IV et V.

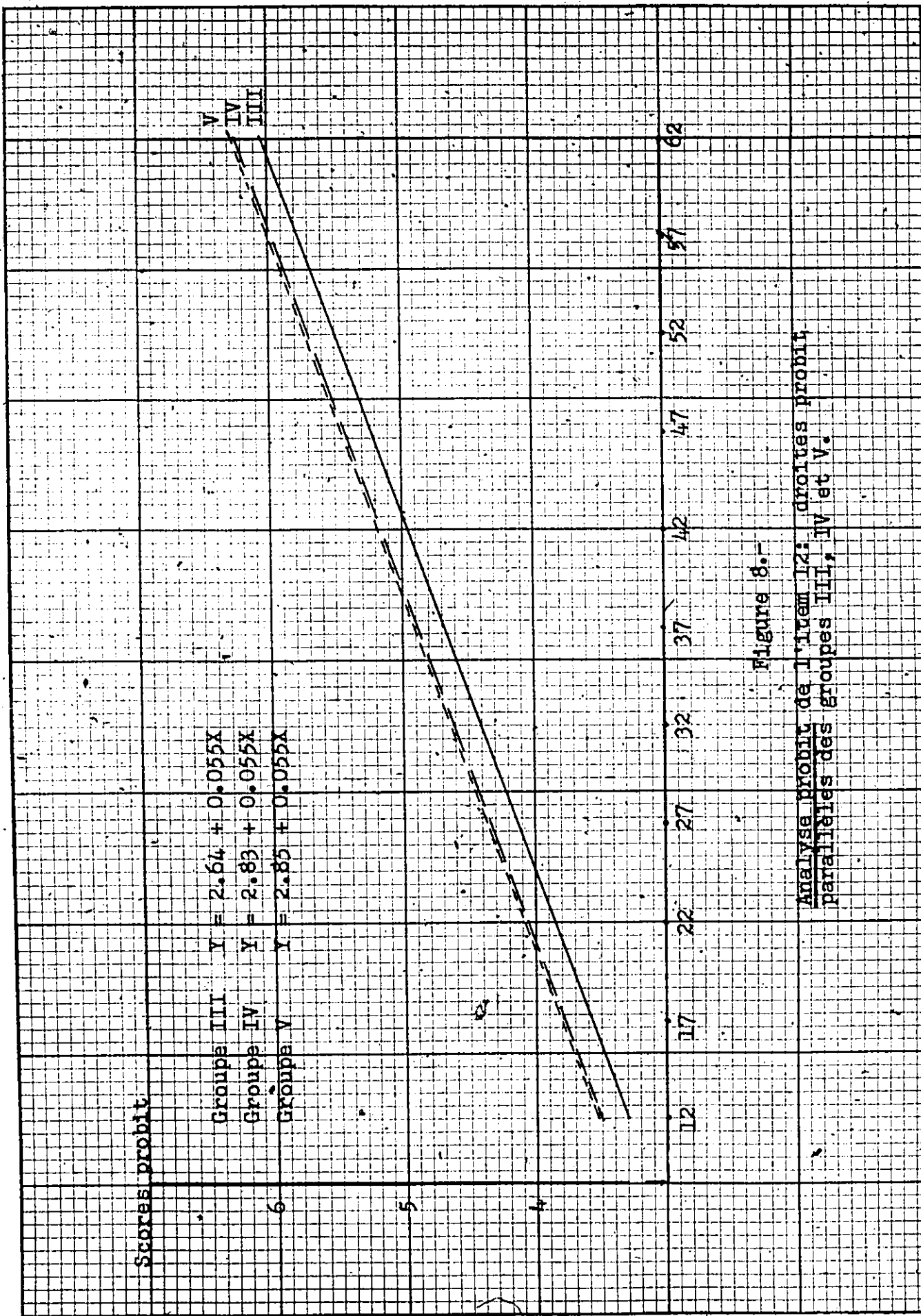


Figure 8.-

ANALYSE PROBIT DE L'ITEM 12: droites probit parallèles des groupes III, IV et V.

calculatrice électronique en élaborant le programme Subroutine PROBT de IBM¹⁰.

Dans l'exemple, l'analyse probit conduit à un rapport F trop petit pour révéler des différences significatives; ces résultats paraissent au Tableau VII à la page suivante. Cet item semble donc avoir conservé une valeur discriminante stable bien qu'il ait changé de position (12ème, 51ème et 15ème) d'une formule à une autre et que les item qui l'entourent aient aussi changé de position.

Au chapitre suivant, le test "chi deux" (entre les scores probit empiriques et les scores probit théoriques de la fonction ogive normale) et la pente initiale de chacun des item dans chacun des groupes seront présentés. Ensuite, pour chacun des item, la pente commune aux trois premiers groupes, qui ont passé la formule A, sera présentée et évaluée par un test F pour établir la constance de ces indices de valeur discriminante. Finalement, pour chacun des item, la pente commune aux trois derniers groupes, qui ont passé des formules différentes, sera présentée et évaluée par un test F pour mettre en relief l'effet de l'ordre de difficulté sur ces indices de valeur discriminante.

¹⁰ Subroutine PROBT, IBM Application Program, System/360 Scientific Subroutine Package, Version III, Programmer's Manual, International Business Machines, New York, p. 44.

Tableau VII.

Tableau de la variance pour les données de
l'analyse probit de l'item 12
concernant les groupes III, IV et V.

La source de variation	La somme des carrés	Les degrés de liberté	L'estimation de la variance
Variance inter-groupes (parallélisme)	7.48	2	3.74
Variance intra-groupes (hétérogénéité)	31.87	19	1.68
Variance totale	39.35	21	

$$F = \frac{3.74}{1.68} = 2.23$$

$$F_{.99}(2,19) = 5.93$$

CHAPITRE III

PRESENTATION ET INTERPRETATION DES RESULTATS

Les résultats seront rapportés selon le plan de présentation suivi jusqu'à maintenant; c'est-à-dire que les résultats reliés à la constance de l'ordre de difficulté croissante des ~~item~~ seront d'abord présentés; ils seront suivis des conclusions reliées aux effets de l'ordre de difficulté sur la valeur discriminante des item et d'une discussion de l'ensemble des résultats.

1. La constance de l'ordre de difficulté croissante des item.

Pour faciliter la compréhension de l'analyse des résultats, rappelons que l'hypothèse principale qui a guidé l'étude de la constance de l'ordre de difficulté des item a été émise comme suit:

Il n'y a pas de différence significative entre les ordres de difficulté croissante des item calculés à partir d'ordres différents de présentation des item d'un test de puissance à temps limité comme la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale administrée en trente minutes.

Pour éprouver cette hypothèse, des indices de difficulté ont été calculés pour chacun des item à partir de la performance de cinq échantillons d'élèves de la Régionale de l'Outaouais. Ces proportions de succès sur un item sont présentées au Tableau XIV à l'Appendice 5. Nous pouvons

remarquer des fluctuations des indices de difficulté pour chacun des item. Ces fluctuations sont plus faciles à voir quand les indices de difficulté sont placés en ordre de difficulté croissante et transformés en rangs tels que présentés au Tableau XV à l'Appendice 5 .

La variabilité entre les rangs calculés individuellement pour certains item est, minime alors que la variabilité entre les rangs calculés pour d'autres item semble beaucoup plus grande. Ceci pourrait s'expliquer sans doute par le fait que la distance en difficulté entre les item pris deux à deux est variable. Ainsi un item qui est beaucoup plus facile ou beaucoup plus difficile que les autres, autrement dit qui est isolé des autres sur l'échelle de difficulté, aura tendance à obtenir un rang stable alors que l'item qui est à peu près de difficulté égale à ses voisins sur l'échelle de difficulté obtiendra un rang qui aura tendance à varier davantage parce qu'il sera plus sensible aux imprécisions de mesure et aux erreurs d'échantillonnage.

Les rangs calculés ont ensuite été comparés d'une formule à l'autre par la corrélation rho. Les coefficients qui ont résultés de ces comparaisons apparaissent au Tableau VIII à la page suivante. Les coefficients qui varient de .91 à .97 ne sont pas significativement différents les uns des autres. En plus de ces coefficients qui démontrent une constance inter-groupes très élevée, d'autres coefficients

Tableau VIII.

Coefficients de corrélation ρ entre les ordres de difficulté croissante des item, calculés à partir des administrations en trente minutes des formules A, expérimentale I et expérimentale II de l'Examen intermédiaire Otis-Ottawa.

Ordres	(1)	(2)	(3)	(4)	(5)
I A 30 (1)					
II A 30 (2)	.94				
III A 30 (3)	.97	.95			
IV ex.I 30 (4)	.91	.92	.94		
V ex.II 30 (5)	.92	.93	.94	.96	

I, II, III, IV et V: échantillons
 A, ex.I et ex.II: formules
 30: temps d'administration

révèlent une constance intra-groupes aussi très élevée. En effet chacun des groupes a été divisé en deux parties et des ordres de difficulté croissante ont été calculés à partir des proportions corrigées de succès de chacune de ces parties. Les coefficients de corrélation rho calculés à l'intérieur des groupes varient de .93 à .98.

Ces coefficients ne permettent donc pas de rejeter la première hypothèse et témoignent ainsi que le calcul de l'ordre de difficulté des item peut être constant entre des groupes équivalents, à partir d'ordres différents de présentation des item, même si tous les élèves ne peuvent atteindre tous les item.

A partir de tels résultats, on ne peut s'attendre à ce que le niveau de constance s'améliore significativement quand le test est administré en soixante minutes. Tout de même, les proportions corrigées de succès sur chaque item administré en soixante minutes ont été calculées, ainsi que les rangs de chacun des item, et sont présentés respectivement aux Tableaux XVI et XVII à l'Appendice 5. Les coefficients rho obtenus entre ces différents ordres de difficulté croissante varient de .92 à .97 comme nous pouvons le constater au Tableau IX à la page suivante. Ces coefficients sont du même ordre que les coefficients précédents calculés à partir d'une administration en trente minutes.

Tableau IX.

Coefficients de corrélation ρ entre les ordres de difficulté croissante des item, calculés à partir des administrations en soixante minutes des formules A, expérimentale I et expérimentale II de l'Examen intermédiaire Otis-Ottawa.

Ordres	(6)	(7)	(8)	(9)
II A 60 (6)				
III A 60 (7)	.97			
IV ex.I 60 (8)	.92	.94		
V ex.II 60 (9)	.94	.96	.97	

II, III, IV et V: échantillons
 A, ex.I et ex.II: formules
 60: administration en 60 minutes

Lorsque les proportions corrigées de succès obtenues après trente minutes, Tableau XIV, sont comparées aux proportions corrigées de succès obtenues en soixante minutes, Tableau XVI, nous remarquons très peu de variations à l'intérieur de chacun des groupes. Ce peu de variation se traduit par des coefficients ρ de .99 présentés au Tableau XVIII à l'Appendice 5 et ne permet pas de rejeter la deuxième hypothèse:

Il n'y a pas de différence significative entre les ordres de difficulté croissante des item d'un test de puissance à temps limité, comme la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale, administré dans une période de temps limité et les ordres de difficulté croissante des item d'un tel test alors qu'il est complété par la plupart des élèves.

Comme il a été souligné dans le chapitre précédent, le fait de comparer des rangs d'item en fonction de groupes, plutôt que des rangs d'individus en fonction d'examens, enlève des sources de variabilité et favorise des coefficients ρ élevés. Mais même en tenant compte de cette particularité, les coefficients obtenus demeurent très élevés.

Par ailleurs, tel qu'attendu, nous pouvons constater au Tableau XVIII à l'Appendice 5 que les coefficients ρ entre l'ordre au hasard et les autres ordres ne sont pas significativement différents de zéro. Nous pouvons remarquer aussi dans le même tableau que les coefficients ρ entre les ordres de difficulté croissante et l'ordre de présentation des item de la formule A originale varient de .61 à .70 .

Il existe donc un écart très marqué entre l'ordre de présentation des item de la formule A et les ordres de difficulté croissante qui ont été calculés. Compte tenu de l'erreur-type du coefficient ρ , cette différence est significative.

A travers les résultats qui suivent, nous essaierons de voir les conséquences de ce désordre.

2. Les effets de l'ordre de difficulté sur la valeur discriminante des item.

L'hypothèse de recherche qui a structuré l'étude des effets de l'ordre de difficulté des item sur la valeur discriminante a été formulée comme suit:

Dans un test de puissance à temps limité, il n'y a pas de différence significative de la valeur discriminante des item entre des formules qui diffèrent uniquement par l'ordre de présentation de leurs item selon leur difficulté.

Formulée en fonction de l'instrument utilisé, cette hypothèse était présentée comme suit:

Il n'y a pas de différence significative entre la valeur discriminante des item de la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale dans sa présentation originale, la valeur discriminante de ces item dans une formule où ils sont en ordre de difficulté croissante et la valeur discriminante de ces item dans une formule où ils sont disposés au hasard.

La valeur discriminante des item a été définie par la pente de la droite probit. Cette droite représente la fonction ogive normale entre les probabilités de succès sur un item et les scores bruts du test. La valeur discriminante a pu être calculée pour tous les item dans chacun des

groupes, excepté pour l'item 75 du premier groupe qui n'avait pas été atteint par un nombre suffisant d'élèves. Ces indices de valeur discriminantes n'ont cependant pas tous la même précision et la même validité. En effet, comme nous l'avons vu au chapitre précédent, l'analyse probit exige d'abord que la fonction entre les succès obtenus sur un item et les scores bruts du test soit une fonction ogive normale. Ainsi, plus la fonction entre les proportions de succès obtenues sur un item et les scores bruts s'éloigne de la fonction ogive normale, plus les résultats de l'analyse probit risquent d'être imprécis et de devenir invalides quand la différence entre les résultats obtenus et la fonction ogive normale est significative.

Cette différence a été évaluée pour chacun des item par des tests "chi deux" qui sont présentés au Tableau XIX à l'Appendice 6. Un premier fait à remarquer dans ce tableau est qu'aucun item n'obtient des valeurs chi deux significatives dans tous les groupes. Les valeurs significatives à un niveau de probabilité de .01 se retrouvent ici et là sans raison apparente. Chacun des item a donc le potentiel pour respecter le postulat de l'ogive normale, mais cela ne suffit pas et si, dans un groupe en particulier, le postulat n'est pas respecté, l'analyse probit ne doit pas être complétée. Ainsi, dans les tableaux qui suivent, nous retrouverons des espaces blancs chez les item dont la valeur chi deux est significative à un niveau de probabilité de .01.

L'étape suivante dans la présentation des résultats permet de considérer la pente de la droite probit de la plupart des item dans chacun des groupes. Ces pentes qui représentent la valeur discriminante des item apparaissent au Tableau XX à l'Appendice 6. Ces indices varient de $-.008$ à $.133$ et montrent que certains item sont plus constants que d'autres. De ce point de vue, les item de la fin semblent moins stables probablement à cause du plus petit nombre d'élèves qui atteignent l'item. Lorsque l'item est atteint par quelque cent élèves, les indices ne semblent pas influencés par la position de l'item dans le test.

Si on utilise les trois premiers groupes comme points de référence, et que l'on y compare le quatrième groupe, on peut remarquer que 25 item administrés au quatrième groupe ont une pente plus élevée que les pentes correspondantes des trois premiers groupes et 16 item ont une pente moins élevée. Par contre, la relation inverse est notée entre les trois premiers groupes et le cinquième groupe; en effet, cette fois-ci, 13 item du cinquième groupe ont une pente plus élevée alors que 31 item ont une pente moins élevée. Lorsque le quatrième groupe est comparé au cinquième, 48 item possèdent une pente supérieure dans le quatrième groupe et 23 item ont une pente inférieure. Ces observations laissent poindre la possibilité qu'un ordre de difficulté croissante des item puisse favoriser la valeur discriminante plus que ne le font un ordre au hasard

ou l'ordre original de la formule A de l'Examen intermédiaire Otis-Ottawa. L'étape suivante permettra de déterminer si cette tendance est significative ou non.

Rappelons brièvement que, pour évaluer la signification de la différence entre les pentes des droites probit, la pente la mieux adaptée aux données de tous les groupes comparés est d'abord établie. Ensuite on évalue par un test F si l'accroissement de la variance provoqué par cette restriction d'une seule pente ou de droites parallèles est significatif par rapport à la variance initiale à l'intérieur des groupes. Si le test F est significatif, ceci permet d'affirmer que les pentes initiales sont significativement différentes et ne peuvent être représentées par une même pente.

En suivant cette méthode de calcul, les trois premiers groupes ont d'abord été comparés. Ces groupes qui avaient tous passé la formule A devaient permettre d'évaluer la constance des pentes calculées par l'analyse probit. Les résultats de cette comparaison sont présentés au Tableau XXI à l'Appendice 6 où apparaissent le nombre d'élèves qui ont atteint l'item dans chacun des groupes, la pente la mieux adaptée aux données ainsi que le test F qui permet d'évaluer si la différence entre les groupes est significative.

Les trois derniers groupes, qui ont passé respectivement les formules A, Expérimentale I et Expérimentale II, ont aussi été comparés pour déterminer les effets de l'ordre

de difficulté sur la valeur discriminante des item. Les résultats de cette comparaison sont présentés au Tableau XXII à l'Appendice 6.

Dans ces deux tableaux, la pente la mieux adaptée aux données des groupes comparés tend à être biaisée légèrement quand les groupes ne sont pas égaux. Elle tend alors à se rapprocher de la pente du groupe le plus nombreux, mais cette tendance est très légère et la pente la mieux adaptée se situe toujours très près de la moyenne des pentes des groupes comparés.

Les tests F sont en général peu élevés. L'un d'eux est même négatif (item 72, groupes III, IV et V). Un tel résultat est illogique et est dû à certaines anomalies.

L'analyse probit parvient à un test F en calculant d'abord la somme des carrés des écarts des scores probit à la droite probit pour chacun des groupes (hétérogénéité) pour ensuite soustraire cette somme de la somme des carrés des écarts de ces mêmes scores probit à la droite probit rendue parallèle aux droites des autres groupes comparés (Sc totale). Pour l'item 72, la somme totale des carrés est inférieure à la somme des carrés due à l'hétérogénéité; ainsi la somme des carrés due au parallélisme des droites apparaît négative et les droites parallèles semblent mieux adaptées aux données que les droites initiales. Ces résultats sont peut-être possibles parce que deux des pentes initiales sont identiques (0.040) et que l'autre est très proche de celles-ci

(0.037). Puisque l'analyse probit calcule la pente d'une droite par approximations successives qui doivent se continuer jusqu'à ce qu'un critère de précision préétabli soit atteint, il est pensable dans un cas comme celui-ci (où les approximations se font simultanément pour les trois droites et où elles doivent se poursuivre jusqu'à ce que le critère de précision soit atteint pour chacune des droites) que le niveau de précision finale dépasse très faiblement le niveau de précision atteint dans le calcul des droites individuelles. De plus le fait que la droite de l'item 72 pour le troisième groupe soit tracée à partir de trois sous-groupes seulement et le fait que les données de cet item pour le cinquième groupe s'éloignent sensiblement de la fonction ogive normale, sans toutefois que cette différence soit suffisante pour rejeter l'item, cela a pu diminuer la précision des calculs.

Ceci permet de souligner les prérequis de l'analyse probit qui se dégagent de l'étude des résultats. D'abord, cette méthode semble exiger un nombre minimum d'environ cent élèves dans chacun des groupes de sorte qu'un nombre minimum de 5 ou 6 sous-groupes puissent servir de base aux calculs de l'analyse. Plus le nombre est inférieur à cent élèves et plus le nombre de sous-groupes est petit, plus la précision des résultats diminue. Il semble aussi très important que les données satisfassent étroitement le postulat de l'ogive normale car plus ils s'en éloignent, plus la précision diminue.

Ces prérequis qui sont importants dans l'analyse individuelle d'un item dans un groupe, deviennent encore plus importants quand des groupes sont comparés entre eux.

La comparaison des trois premiers groupes entre eux pour chacun des item n'a montré aucune différence significative à un niveau de probabilité de .01 . Ce résultat peut illustrer la constance de l'analyse probit quand il s'agit de comparer la pente des item d'une même formule administrée à des groupes semblables.

La comparaison des trois derniers groupes entre eux montre des différences significatives à un niveau de probabilité de .01 pour les item 9, 48 et 62. Lorsque l'on compare les groupes deux à deux, ces différences significatives apparaissent entre les groupes IV et V pour l'item 9 $\sqrt{F(1,14) = 15.59}$; $F_{.99}(1,14) = 8.867$, entre les groupes IV et V pour l'item 48 $\sqrt{F(1,14) = 19.19}$; $F_{.99}(1,14) = 8.867$ et entre les groupes III et IV pour l'item 62 $\sqrt{F(1,12) = 16.51}$; $F_{.99}(1,12) = 9.337$. A cause de la variation des pentes de l'item 62 dans les trois premiers groupes, l'évaluation de cette dernière différence significative est difficile. De plus, ces trois différences significatives perdent du poids quand elle sont considérées par rapport à la somme des différences possibles pour l'ensemble des item, soient plus de deux cents. Ainsi, que trois différences sur plus de deux cents soient significatives à un niveau de

probabilité de .01, c'est à peu près ce à quoi on devrait s'attendre par pur hasard.

Certaines comparaisons ne sont pas possibles en prenant trois groupes à la fois puisqu'il arrive que l'un des trois item soit rejeté parce qu'il ne respecte pas le postulat de l'ogive normale. Dans tous les cas où cette situation se présente, lorsque les deux item qui restent sont comparés, on ne trouve aucune différence significative entre leur pente.

Ces résultats ne permettent pas d'affirmer qu'il y a une différence significative entre la valeur discriminante des item des formules A originale, Expérimentale I, en ordre de difficulté croissante, ou Expérimentale II dont les item sont distribués au hasard. L'hypothèse nulle ne peut donc pas être rejetée à partir de ces résultats.

D'autre part, le fait que la formule Expérimentale I a plus d'item de valeur discriminante plus élevée que les deux autres formules et que la formule Expérimentale II a plus d'item de valeur discriminante moins élevée ne semble pas fortuit, mais il ne suffit pas à rejeter l'hypothèse non plus.

L'ensemble des résultats de l'analyse de la valeur discriminante pourrait être résumé comme ceci: il y a des indications qui suggèrent qu'un ordre de difficulté croissante des item favorise une valeur discriminante plus élevée qu'un ordre au hasard ou que l'ordre de la formule A,

mais ces indications ne résistent pas à l'analyse statistique qui exige des différences significatives.

3. La discussion des résultats.

Les résultats qui viennent d'être énumérés peuvent être considérés sous deux angles: un angle pratique et un angle théorique.

Sur le plan pratique, les résultats obtenus peuvent être généralisés à la population totale d'où furent prélevés les échantillons. Cette population se limite à la Régionale de l'Outaouais et plus précisément aux élèves de septième année 1967-1968, aux élèves de sixième année 1968-1969 et aux élèves de septième année des régions urbaines et semi-urbaines de l'année académique 1969-1970.

Pour cette population, nous pouvons affirmer que les item de la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale n'étaient pas en ordre de difficulté croissante. A partir des données de l'expérience, il a été possible de replacer les item et de parvenir à neuf ordres de difficulté croissante d'un niveau de constance élevé entre eux. L'un de ces ordres a constitué la formule expérimentale I. Quel ordre de difficulté croissante conviendrait le mieux à cette population? Il est impossible de le dire puisque l'ordre de difficulté croissante est un phénomène qui varie en fonction de ceux qui passent le test. Dans la population étudiée, la variabilité est faible et l'un ou

L'autre des neuf ordres de difficulté croissante pourrait être choisi dans une réimpression éventuelle du test puisque ces ordres ne sont pas significativement différents les uns des autres, mais qu'ils diffèrent tous significativement de l'ordre des item de la formule A. Un nouvel ordre de difficulté croissante remettrait le test en accord avec la conception de son auteur Otis qui affirma l'avoir construit en plaçant les item en ordres de difficulté croissante.

Les résultats de l'analyse des item ne permettent pas d'affirmer si ce nouvel ordre favoriserait significativement la valeur discriminante des item, mais le test pourrait sans doute être amélioré si les item les moins discriminants, comme les item 14, 27, 41, 54, 66 et 75, étaient modifiés ou rejetés.

Sur le plan théorique, les implications dépassent les cadres de l'instrument et les cadres de la population qui a été choisie pour l'expérience.

Le premier problème qui a été posé à ce niveau se présente comme suit: à partir d'un test de puissance à temps limité, est-il possible de calculer et d'établir un ordre de difficulté croissante des item qui soit constant tout en respectant la limite de temps imposé par la standardisation? Tel que rapporté dans la recension des écrits, aucun auteur ne semble avoir traité ce problème spécifiquement. Des auteurs se sont plutôt intéressés au problème plus général de l'analyse de la difficulté des item dans un

test de puissance à temps limité. Ces auteurs affirment qu'une analyse de la difficulté des item est biaisée lorsque les item de la fin ne sont pas atteints par tous les élèves et ils suggèrent d'éliminer la limite de temps pour faire une analyse d'item plus précise et plus constante. Cependant, certains de ces auteurs reconnaissent que les résultats d'une telle analyse ne sont peut-être plus valides lorsque le test est de nouveau administré avec une limite de temps.

Ces opinions aboutissent à une impasse qui n'a pas été résolue au niveau expérimental. En effet, Mollenkopf (p. 19-22), le seul auteur qui a étudié expérimentalement la difficulté des item dans des tests avec et sans limite de temps, parvint à des résultats divisés. En utilisant un test d'aptitude verbale, il montra que la difficulté des item de la fin était sous-évaluée quand il y avait une limite de temps qui empêchait des élèves d'atteindre tous les item, mais il ne retrouva pas cette tendance lorsqu'il utilisa un test de mathématique.

La recension des écrits laissait donc présager que le calcul de l'ordre de difficulté croissante des item dans un test à temps limité pourrait être imprécis, qu'il pourrait varier selon la position des item dans le test et que cet ordre de difficulté croissante serait moins précis que s'il avait été calculé dans un test sans limite de temps.

A la suite de cet état de la question, les résultats obtenus sont très inattendus. Ils montrent en effet un niveau

de constance très élevé du calcul de l'ordre de difficulté croissante des item de trois formules administrées avec une limite de temps. Le niveau de constance est si élevé qu'il n'a pu être amélioré même quand les formules ont été administrées de façon à permettre à la plupart des élèves d'atteindre tous les item.

Bien que l'indice de difficulté d'un item varie en plus ou en moins quand le test est administré sans limite de temps plutôt qu'avec une limite de temps, cette variation est si petite que l'item garde à peu près la même place dans l'ordre de difficulté des item. Ce niveau de constance montre donc que l'ordre de difficulté des item peut être calculé à l'intérieur de la limite de temps imposée par la standardisation et qu'il est inutile dans ce cas-ci de supprimer cette limite de temps.

Ces résultats sont peut-être liés à la proportion d'élèves que la limite de temps empêche d'atteindre tous les item. En d'autres mots, plus la limite de temps est sévère, moins les élèves auront la chance d'atteindre tous les item et plus grande pourra être la différence entre les résultats avec limite de temps et les résultats sans limite de temps.

Dans la présente expérience, une période de trente minutes empêcha certains élèves de se rendre plus loin que les item 30-35; environ cinquante pourcent des élèves atteignirent l'item 60 alors que moins de vingt-cinq pourcent des élèves atteignirent tous les 75 item. Une période de soixante

minutes permit à plus de quatre-vingt dix pourcent des élèves d'atteindre tous les item. De ce point de vue, la différence entre les administrations en trente et en soixante minutes semble grande. Il faut cependant remarquer que même si moins de vingt-cinq pourcent des élèves ont atteint tous les item dans chacune des formules administrées en trente minutes, ce pourcentage représente environ soixante-quinze élèves dans les trois derniers groupes et semble suffisant pour assurer une certaine constance dans la mesure de la difficulté des item.

Ces résultats viennent appuyer une remise en question des opinions selon lesquelles il faudrait supprimer la limite de temps d'un test de puissance à temps limité pour analyser efficacement la difficulté de ses item. Ces résultats montrent que ces opinions générales pourraient être nuancées puisqu'il a été possible d'analyser l'ordre de difficulté des item de la formule A de l'Examen intermédiaire Otis-Ottawa d'une façon précise et constante tout en respectant la limite de temps imposée par la standardisation. Il semblerait donc souhaitable que des recherches semblables à celle-ci soient entreprises dans le but de confirmer ces résultats et d'apporter d'autres précisions à ce sujet.

D'autre part, le problème concernant les effets de l'ordre de difficulté des item sur leur valeur discriminante a été posé comme suit: dans un test de puissance à temps limité, l'ordre de difficulté croissante des item favorise-t-il

la valeur discriminante des item plus que ne le fait un ordre au hasard ou plus que ne le fait l'ordre actuel de l'instrument utilisé?

Chez les auteurs qui parlent de l'ordre de présentation des item, la tendance principale est de favoriser la construction de tests de rendement et d'aptitudes avec une présentation d'item en ordre de difficulté croissante. Certains préfèrent cependant une présentation d'item par catégories de contenu. Ces affirmations ne sont pas fondées sur des expériences, mais elles sont plutôt des opinions tirées de déductions logiques selon lesquelles, par exemple, une présentation des item en ordre de difficulté croissante serait plus naturelle, elle favoriserait davantage l'élève, elle augmenterait la validité et la constance d'un test, etc.

Il y eu des recherches au sujet de la présentation des item selon un ordre de difficulté, mais elles furent peu nombreuses et, lorsqu'on limite le problème aux tests de puissance à temps limité, il n'en reste que deux où sont étudiés les effets de l'ordre de difficulté croissante des item. L'une de ces recherches, conduite par Cromack et Sax (p. 50-52), met en évidence des différences significatives entre les scores bruts de formules dont l'ordre de difficulté des item est différent. L'auteur de l'autre recherche, Mollenkopf (p. 52-54), montre que les coefficients de corrélation bi-sériale entre les item et les scores bruts sont influencés par la place que les item occupent dans le test

indépendamment de l'ordre général des item. Il ne peut ainsi trouver d'effets significatifs de l'ordre de difficulté des item sur la valeur discriminante.

Après cet état de la question, le problème initial concernant les effets de l'ordre de difficulté des item sur la valeur discriminante demeurait entier et il surgissait même un nouveau problème, celui de trouver une mesure de la valeur discriminante des item qui ne soit pas biaisée par la position des item dans un test de puissance à temps limité. A partir de ces deux problèmes, il est maintenant possible de reprendre les résultats de l'expérience pour les discuter.

Au niveau de la méthode utilisée, il faut rappeler que les corrélations bi-sérielles et de point bi-sériel ont été considérées, mais qu'elles ont été rejetées principalement parce que leurs coefficients variaient directement en fonction de la position de l'item dans le test de sorte qu'il était possible de deviner si un item avait un coefficient plus élevé qu'un autre simplement en regardant si cet item était placé plus loin dans le test. L'analyse probit, qui a été retenue dans cette expérience comme méthode de mesure de la valeur discriminante, ne présente pas les mêmes lacunes. En effet, comme on peut le constater au Tableau XXI à l'Appendice 6, il est impossible de deviner systématiquement si un item a un indice de valeur discriminante supérieur à celui d'un autre item seulement en considérant s'il est placé plus loin dans le test.

A cause de cette plus grande indépendance face à la position des item, l'analyse probit semble plus appropriée que les corrélations bi-sérielales et de point bi-sériel pour mesurer la valeur discriminante des item dans un test de puissance à temps limité. Cette méthode offre donc de nouvelles possibilités de recherche et elle serait ainsi un apport original au problème qui a été laissé en plan par Mollenkopf faute d'une méthode de calcul adéquate pour continuer. A ce titre, l'analyse probit semble mériter une attention particulière de la part des chercheurs qui voudraient l'approfondir comme méthode d'analyse d'item ou l'utiliser pour continuer l'étude entreprise dans cette expérience. Tout au moins, il serait souhaitable que d'autres recherches viennent appuyer ou nuancer, d'une part, cette indépendance de l'analyse probit par rapport à la position des item dans un test de puissance à temps limité et, d'autre part, les résultats qu'elle a fournis sur les effets de l'ordre de difficulté des item sur la valeur discriminante des item.

Si l'on porte maintenant l'attention sur ces derniers résultats, on peut voir que ceux-ci n'apportent pas des conclusions définitives au problème posé. Ces résultats présentent d'abord des indices qui peuvent suggérer qu'un ordre de difficulté croissante des item favorise davantage la valeur discriminante des item que ne le font un ordre au hasard ou l'ordre original de l'instrument utilisé. La formule en ordre de difficulté croissante possède plus d'item de valeur

discriminante maximale par rapport aux deux autres formules; la formule au hasard possède plus d'item de valeur discriminante minimale et l'ordre original de la formule A se situe entre ces deux autres formules. Ces différences apparentes ne sont cependant pas confirmées par les calculs subséquents de sorte que l'hypothèse nulle ne peut pas être rejetée.

Ces résultats invitent à de nouvelles recherches, d'autant plus qu'ils semblent être les premiers résultats concernant les effets de l'ordre de difficulté des item sur la valeur discriminante des item. Idéalement, ces recherches pourraient mieux faire ressortir, si possible, les effets de l'ordre de difficulté des item sur la valeur discriminante en choisissant un instrument qui retiendrait les qualités de l'Otis à ce sujet, soient le grand nombre d'item, l'étendue de leur difficulté et la limite de temps imposée, et dont les item auraient une bonne validité, ne représenteraient qu'une catégorie de contenu et ne seraient pas à choix multiple. Au niveau expérimental, un ordre de difficulté décroissante des item pourrait aussi être ajouté.

Dans l'ensemble, les résultats de cette recherche, sur la constance de l'ordre de difficulté des item et sur les effets de cet ordre sur la valeur discriminante des item dans un test de puissance à temps limité, pourraient susciter de nouvelles expériences tant au niveau des méthodes de mesure de la difficulté ou de la valeur discriminante qu'au niveau de la théorie qui structure ce domaine.

CONCLUSIONS

Le but de cette recherche était d'étudier la constance d'un ordre de difficulté croissante des item et les effets de cet ordre sur la valeur discriminante des item dans un test de puissance à temps limité.

Ces deux problèmes révélaient des difficultés particulières parce qu'aucun auteur ne semblait avoir étudié spécifiquement la constance de l'ordre de difficulté des item dans un test de puissance à temps limité, et que l'unique recherche sur les effets de l'ordre de difficulté sur la valeur discriminante des item d'un tel test n'avait fourni aucun résultat satisfaisant faute d'une méthode de mesure adéquate pour analyser ce problème.

Pour assurer plus de validité à cette recherche, un ordre de difficulté croissante des item a d'abord été calculé selon la méthode traditionnelle qui consiste à supprimer la limite de temps pour permettre à la plupart des élèves d'atteindre tous les item. Une nouvelle formule fut réorganisée selon cet ordre et une autre selon un ordre au hasard.

La formule A originale fut administrée à trois échantillons d'élèves de sixième ou de septième année de la Régionale de l'Outaouais. Le premier échantillon passa le test selon les procédures de la standardisation. Les deuxième et troisième échantillons passèrent la formule A selon la standardisation d'abord et poursuivirent ensuite

pendant trente minutes additionnelles, ce qui permit à la plupart des élèves d'atteindre tous les item. Les nouvelles formules en ordres de difficulté croissante et en ordre au hasard furent administrées respectivement à un quatrième et à un cinquième échantillons en trente et soixante minutes.

Un ordre de difficulté croissante des item fut calculé à partir de chacune des administrations en trente et en soixante minutes alors que la valeur discriminante des item fut calculée à partir des administrations en trente minutes.

Les résultats révèlent que les ordres de difficulté croissante des item qui furent calculés à partir des administrations en trente minutes sont très constants, qu'ils ne sont pas significativement différents des ordres de difficulté croissante calculés à partir des administrations en soixante minutes, mais qu'ils diffèrent de l'ordre de présentation des item de la formule A originale.

D'autre part, les résultats présentent des indices qui suggèrent qu'un ordre de difficulté croissante favoriserait une bonne valeur discriminante plus que ne le ferait l'ordre au hasard ou l'ordre original de l'instrument utilisé. Cependant, ces indices s'estompent devant une analyse statistique plus poussée qui ne révèle aucune différence significative quant à la valeur discriminante des item entre les formules.

Sur le plan pratique, cette expérience pourrait être considérée comme une étape vers une réorganisation de la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale. Cette réorganisation exclurait certains item invalides et placerait les item retenus en ordre de difficulté croissante tel que souhaité par Otis.

Sur le plan théorique, cette expérience démontre qu'une analyse de la difficulté des item peut produire des résultats valides et constants même si une partie de l'échantillon n'atteint pas tous les item. De plus, elle introduit l'analyse probit comme une méthode plus appropriée que les corrélations bi-sérielles et de point bi-sériel pour étudier les effets de l'ordre de difficulté des item sur la valeur discriminante des item dans un test de puissance à temps limité. Cette méthode n'a révélé aucune différence significative de la valeur discriminante des item entre les formules utilisées.

BIBLIOGRAPHIE

Bonin, Jean-Guy et André Pellerin, Ordre et degré de difficulté des item et valeur discriminante des item des Examens Otis-Ottawa d'habileté mentale, examen intermédiaire, formule A, étude (inédiée) présentée à la Faculté de Psychologie de l'Université d'Ottawa en mars 1968, et présentée au Congrès de l'ACEAS en novembre 1968 à Ottawa, vii-50 p.
Etude préliminaire qui a marqué le début de cette recherche.

Bonin, Jean-Guy, Influences de l'ordre de difficulté croissante des item sur les qualités psychométriques d'un test d'intelligence du type puissance à temps limité, thèse de maîtrise (inédiée) présentée à la Faculté de Psychologie de l'Université d'Ottawa, Ontario, 1971, xv-166 p.

Cette thèse a été faite parallèlement à la présente recherche à partir des mêmes données. L'auteur y traite un sujet complémentaire au problème analysé ici; il étudie les effets de l'ordre de difficulté des item sur la validité externe du test en utilisant le Culture Fair, Scale 2 et le succès scolaire comme critère.

Brenner, Marshal Hallock, Test Difficulty, Reliability, and Discrimination as Functions of Item Difficulty Order, thèse de doctorat (inédiée) présentée à Ohio State University, 1962, vii-92 p.

Cette étude est l'une des mieux structurées et des plus complètes sur le sujet de l'ordre de difficulté des item d'un test de puissance.

-----, Test Difficulty, Reliability, and Discrimination as Functions of Item Difficulty Order, dans The Journal of Applied Psychology, vol. 48, n° 2, livraison d'avril 1964, p. 998-100.

Cet article résume la thèse de doctorat présentée ci-haut.

Capron, Virginia L., The Relative Effect of Three Orders of Arrangement of Items upon Pupils' Scores in Certain Arithmetic and Spelling Tests, dans The Journal of Educational Psychology, vol. 24, n° 9, livraison de décembre 1933, p. 687-695.

L'auteur étudie les effets de trois ordres de difficulté sur la moyenne, la variabilité des scores bruts et sur le nombre d'erreurs commises par les étudiants. Elle administre les mêmes item trois fois aux étudiants; l'apprentissage peut donc masquer considérablement les effets des différents ordres de difficulté des item. Elle n'obtient aucun résultat significatif.

Carter, Harold D., How Reliable Are the Common Measure of Difficulty and Validity of Objective Test Items, dans The Journal of Psychology, vol. 13, première demie, livraison de janvier 1942, p. 31-39.

Cette recherche peu rigoureuse au point de vue méthodologique fait ressortir la constance d'indices de difficulté entre des étudiants faibles et des étudiants plus forts.

Chapanis, Alphonse, A Note on the Validity and Difficulty of Items in Form A of the Otis Self-Administering Tests of Mental Ability, dans The Journal of Experimental Education, vol. 5, n° 3, livraison de mars 1937, p. 246-248.

Dans ce rapport de recherche, l'auteur ne révèle pas s'il a administré la forme A de l'examen supérieur avec ou sans limite de temps. Il affirme que certains item ne respectent pas l'ordre de difficulté croissante. Les coefficients r_{bis} calculés entre les item et les scores bruts sont très élevés.

Conrad, H.S., Characteristics and Uses of Item-Analysis Data, dans Psychological Monographs, vol. 62, n° 8, 1948, vi-49 p.

L'auteur traite de plusieurs aspects de l'analyse des item, mais étudie plus méticuleusement la corrélation bi-sériale comme mesure de validité.

Cromack, Theodore R. et Gilbert Sax, The Effects of Various Forms of Item Arrangement on Test Performance, dans Journal of Educational Measurement, vol. 3, n° 4, livraison d'hiver 1966, p. 309-311.

Etude fort simple qui montre une relation entre l'ordre de difficulté des item et la période de temps allouée pour passer le test. L'auteur néglige cependant de s'assurer de l'équivalence de ses groupes.

Crooks, William R. et Leonard W. Ferguson, Item Validities of the Otis Self-Administering Tests of Mental Ability for a College Population, dans The Journal of Experimental Education, vol. 9, n° 3, livraison de mars 1941, p. 229-232.

L'ordre de difficulté croissante des item qui a été calculé dans cette étude diffère de ceux qui ont été proposés par Otis et par Chapanis. Les coefficients r_{bis} sont plus petits que ceux qu'a obtenus Chapanis et ils varient selon la place que l'item occupe dans le test.

Finney, D.J., Probit Analysis, London, Cambridge University Press, 1971, xv-333 p.

Nouvelle édition du livre de base sur l'analyse probit. On y retrouve les élaborations théoriques qui sont à l'origine de cette méthode, une description détaillée des différentes étapes de cette analyse statistique, des applications particulières tant sur le plan expérimental que sur le plan technique, et les derniers développements apportés par l'utilisation de l'ordinateur.

-----, The Application of Probit Analysis to the Results of Mental Tests, dans Psychometrika, vol. 9, n° 1, livraison de mars 1944, p. 31-39.

L'auteur adapte l'analyse probit, méthode d'abord utilisée en toxicologie, à l'étude des tests d'habileté mentale.

Freeman, Frank S., Theory and Practice of Psychological Testing, Toronto, Holt, Rinehart and Winston, 1966, xix-697 p.

Dans ce volume de base en psychométrie, l'auteur souligne en quelques mots l'imprécision de la mesure de la difficulté d'un item qui n'est pas atteint par tous les membres de l'échantillon.

Furst, Edward J., Constructing Evaluation Instruments, New York, David McKay, 1958, xv-334 p.

L'auteur énonce des avantages et des désavantages possibles de l'ordre de difficulté croissante des item, sans appuyer ses affirmations d'une recherche empirique.

Gerberich, Raymond L., Harry A. Greene et Albert N. Jorgensen, Measurement and Evaluation in the Modern School, New York, David McKay, 1962, xviii-622 p.

Ces auteurs préfèrent arranger les item d'un test selon leur contenu plutôt que selon leur difficulté.

Gibbons, Charles C., The Predictive Value of the Most Valid Items of an Examination, dans The Journal of Educational Psychology, vol. 31, n° 8, livraison de novembre 1940, p. 616-621.

Ce rapport de recherche imprécis montre que le nombre de personnes qui réussissent un item est constant entre deux formules dont l'ordre de difficulté des item est différent.

Greer, Donald et Joseph L. French, Effect of Test-Item Arrangement on Physiological and Psychological Behavior in Primary-School Children, dans The Journal of Educational Measurement, vol. 1, n° 2, livraison de décembre 1964, p. 151-153.

Etude faite à partir de petits groupes où les effets de l'apprentissage ne sont pas convenablement contrôlés et où la validité des critères externes n'est pas assurée.

Guilford, J.P., Psychometric Methods, Toronto, McGraw-Hill, 1954, x-597 p.

Livre de base qui couvre le domaine de la psychométrie.

Helmstadter, G.C., Principles of Psychological Measurement, New York, Appleton-Century-Crofts, 1964, xxii-248 p.

L'auteur opte pour la présentation des item selon leur contenu plutôt que selon leur difficulté.

Henrysson, Sten, Gathering, Analysing, and Using Data on Test Items, dans Educational Measurement, Robert L. Thorndike, éditeur, Washington, American Council on Education, 1971, p.

Dans ce chapitre d'un livre de base très complet, l'auteur opte pour l'ordre de difficulté croissante des item dans un test sans toutefois appuyer son option d'une recherche empirique.

Hilgard, Ernest R., et Earl L. Jandron, Stability of Item Statistics in Two Institutions, dans The Journal of Educational Psychology, vol. 51, n° 4, livraison d'août 1960, p. 195-198.

Cette recherche banale et peu rigoureuse met en corrélation les indices de difficulté des item d'un test qui fut administré à deux groupes hétérogènes.

Horst, Paul, Psychological Measurement and Prediction, Belmont (California), Wadsworth, 1966, xii-455 p.

L'auteur fait des remarques intéressantes sur la mesure de la difficulté des item d'un test à temps limité, mais sans les appuyer d'une recherche empirique.

Hovland, Carl Iver et E.F. Wonderlic, A Critical Analysis of the Otis Self-Administering Tests of Mental Ability, Higher Form, dans The Journal of Applied Psychology, vol. 23, n° 3, livraison de juin 1939, p. 367-387.

L'une des études les mieux conduites sur l'ordre de difficulté des item de l'Otis, Higher Form. Elle démontre que les item ne sont pas en ordre de difficulté croissante.

Hull, Clark L., Aptitude Testing, New York, World Book, 1928, xiv-535 p.

L'auteur se prononce en faveur de l'ordre de difficulté croissante des item sans rapporter une recherche expérimentale qui appuierait son affirmation.

Lund, Kenneth Withelm, Test Performance as Related to Order of Item Difficulty, Anxiety and Intelligence, thèse de doctorat (inédiée) présentée à l'Université Northwestern, Evanston, Illinois, 1953, iii-66 p.

Cette thèse de doctorat montre une relation entre l'ordre de difficulté des item et le score brut total d'un test de puissance. L'auteur a cependant négligé de mesurer la difficulté des item au début de son expérience.

Marso, Ronald N., Test Item Arrangement, Testing Time, and Performance, dans The Journal of Educational Measurement, vol. 7, n° 2, livraison d'été 1970, p. 113-118.

L'auteur ne trouve aucune relation significative entre l'ordre de difficulté des item, la performance des étudiants et le temps requis pour compléter un examen.

Mensh, Ivan N., The Effects of Test Abbreviation upon Responses to the Individual Items, dans The Journal of Clinical Psychology, vol. 5, n° 1, livraison de janvier 1949, p. 22-36.

Cette recherche semble bien faite et est la première à révéler un effet significatif de l'ordre de difficulté des item sur les scores bruts totaux d'un test.

Mollenkopf, William G., An Experimental Study of the Effects on Item-Analysis Data of Changing Item Placement and Test Time Limit, dans Psychometrika, vol. 15, n° 3, livraison de septembre 1950, p. 291-317.

Cette étude bien conduite est l'une des plus importantes qui traitent de l'ordre de difficulté des item. Elle montre les effets conjoints de l'ordre de difficulté des item et de la limite de temps sur les mesures de difficulté, et sur les coefficients rbis.

Munz, David C. et Albert D. Smouse, Interaction Effects of Item-Difficulty Sequence and Achievement-Anxiety Reaction on Academic Performance, dans The Journal of Educational Psychology, vol. 59, n° 5, livraison de février 1968, p. 370-374.

A partir de très petits échantillons, les auteurs ne trouvent aucune relation significative entre l'ordre de difficulté des item d'un examen et la performance des étudiants.

Nunnally, Jum C., Psychometric Theory, New York, McGraw-Hill, 1967, xiv-640 p.

Dans ce livre bien structuré, l'auteur rappelle en quelques mots que les item des tests d'aptitude et de rendement sont souvent placés en ordre de difficulté croissante.

Otis, Arthur S., Statistical Method in Educational Measurement, New York, World Book, 1925, xi-337 p.

L'auteur consacre un chapitre au développement de l'habileté mentale et un autre à la mesure de "brightness". Ces deux notions sont à la base de son test.

-----, Otis Self-Administering Tests of Mental Ability, New York, Harcourt, Brace & World, 1928, 12 p.
Ce manuel présente les Otis Self-Administering Tests of Mental Ability.

Pellerin, André R., L'ordre de difficulté croissante des item de la formule A de l'Examen intermédiaire Otis-Ottawa d'habileté mentale: sa constance et ses effets sur la valeur discriminante des item, rapport intérimaire (inédit) présenté à la Faculté de Psychologie de l'Université d'Ottawa, 1973, sv-179 p.

Ruch, Giles M., The Objectives or New-Type Examination, New York, Scott, Foresman and co., 1929, x-478 p.

Livre périmé sur la construction des tests. L'auteur y affirme, sans apporter d'évidence expérimentale, que l'ordre de difficulté croissante des item augmente à la fois la validité et la constance d'un test.

Shevenell, R.-H., Les examens Otis-Ottawa d'habileté mentale, Ottawa, Editions de l'Université d'Ottawa, 1948, viii-40 p.

Shevenell présente en détails la version française de Otis Self-Administering Tests of Mental Ability et met en relief ses caractéristiques de test standardisé.

Smouse, Albert D. et David C. Munz, The Effects of Anxiety and Item Difficulty Sequence on Achievement Testing Scores, dans The Journal of Psychology, vol. 68, première partie, livraison de janvier 1968, p. 181-184.

Ce rapport de recherche ne révèle aucun effet significatif de l'ordre de difficulté des item ou de l'anxiété des étudiants sur les résultats d'un examen. Les auteurs remettent en question le traitement qui avait pour but de provoquer l'anxiété.

Stanley, Julian C., Measurement in Today's Schools,
New York, Prentice-Hall, 1965, xviii-414 p.

Dans ce livre de principes généraux sur la construction des tests, l'auteur affirme que les item devraient être placés en ordre de difficulté croissante.

APPENDICE 1

STATISTIQUES QUE LES CINQ ECHANTILLONS D'ELEVES DE
LA REGIONALE DE L'OUTAOUAIS ONT OBTENUES SUR LES
FORMULES A, EXPERIMENTALE I, OU EXPERIMENTALE II
DE L'EXAMEN INTERMEDIAIRE OTIS-OTTAWA
ADMINISTREES EN TRENTE ET EN
SOIXANTE MINUTES.

Tableau X.

Statistiques que les cinq échantillons d'élèves ont obtenues sur les formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes.

Echantillon Formule	I A	II A	III A	IV Ex. I	V Ex. II
N	200	258	316	313	301
Année acadé.	7ème	6ème	7ème	7ème	7ème
$\chi^2_{.99} () =$	(9)=21.67	(6)=16.81	(8)=20.09	(8)=20.09	(8)=20.09
χ^2 échantillon ^a	(9)= 8.43	(6)=12.05	(8)=22.55	(8)= 6.21	(8)=24.70
Sk	.5	-.05	.34	.08	.39
OSk	.99	.86	.138	.138	.141
RCSk	.505	.058	2.46	.58	2.76
Ku	.272	.30	-.43	-.40	-.53
OKu	.019	.017	.27	.27	.28
RCKu	.489	2.22	1.56	1.44	1.88
Mo	37	32	37	32	32
Md	38	34.2	36.60	36.20	30.20
M	38.59	35.3	37.52	36.46	31.24
OM	.7	.706	.62	.64	.73
σ	9.99	11.32	11.11	11.40	12.63
$\sigma\sigma$.5	.5	.44	.45	.51

a Tests "chi deux" entre les fréquences empiriques des scores bruts et les fréquences théoriques de la courbe normale correspondante.

Tableau XI.

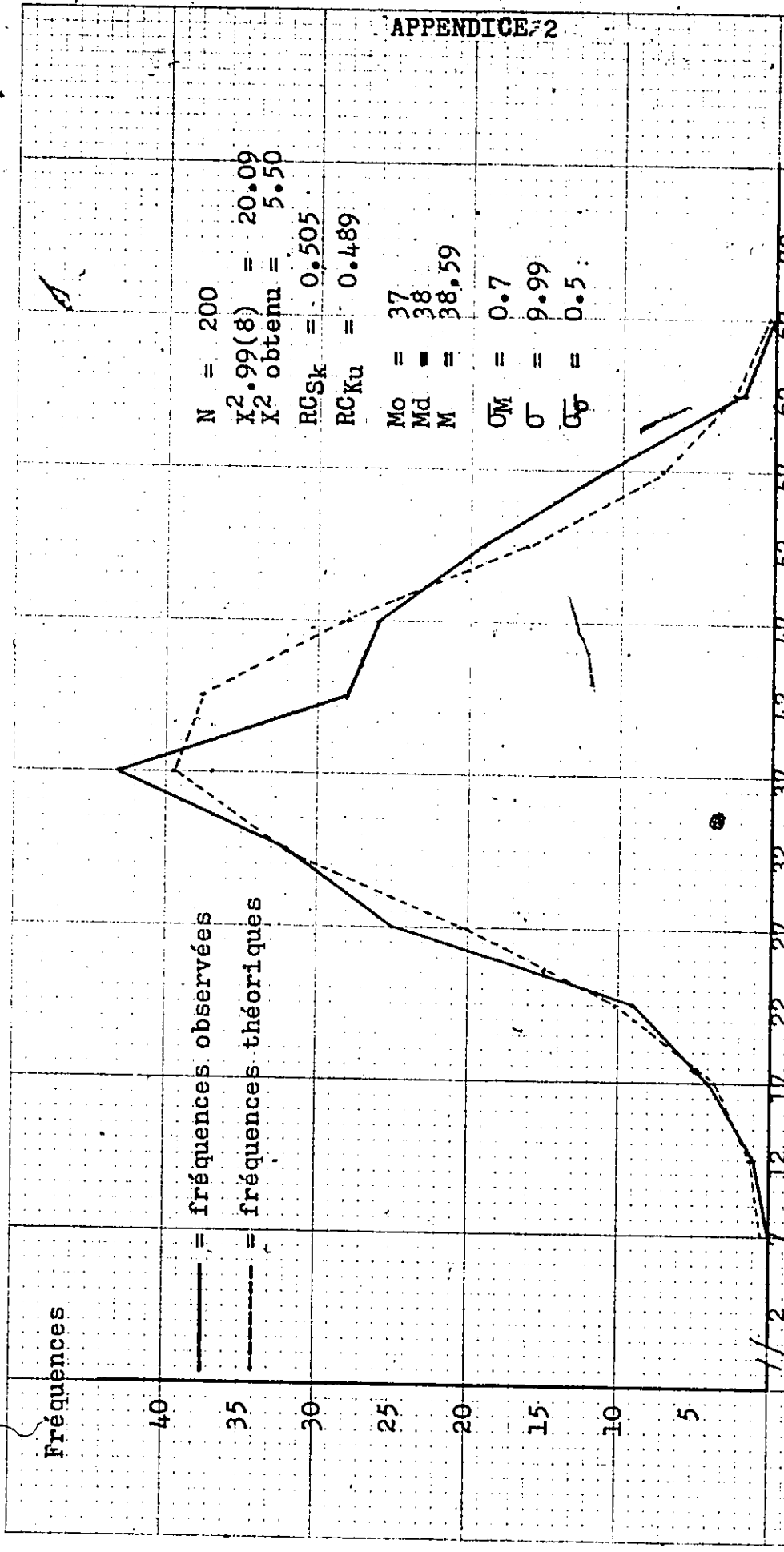
Statistiques que les cinq échantillons d'élèves ont obtenues sur les formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en soixante minutes.

Echantillon Formule	II A	III A	IV Ex. I	V Ex. II
N	258	316	313	301
Année acadé.	6ème	7ème	7ème	7ème
$\chi^2_{.99} () =$	(7)=18.48	(8)=20.09	(8)=20.09	(8)=20.09
χ^2 échantillon ^a	(7)= 8.86	(8)=11.42	(8)=16.11	(8)=14.31
Sk	.38	-.01	-.15	-.20
\overline{OSk}	1.00	.138	.138	.141
RCSk	.38	.07	1.09	1.42
Ku	.29	-.65	-.61	-.71
\overline{OKu}	.017	.276	.277	.282
RCKu	1.58	2.35	2.20	2.52
Mo	42	42	47	47
Md	42.8	42.50	40.80	41.10
M	43.1	42.44	40.51	39.96
\overline{OM}	.724	.626	.687	.695
$\overline{\sigma}$	11.61	11.15	12.23	12.10
$\overline{\sigma\sigma}$.511	.442	.486	.492

^a Tests "chi deux" entre les fréquences empiriques des scores bruts et les fréquences théoriques de la courbe normale correspondante.

APPENDICE 2

POLYGONES DE FREQUENCES DES SCORES BRUTS QUE LES CINQ
ECHANTILLONS D'ELEVES DE LA REGIONALE DE L'OUTAOUAIS
ONT OBTENUS SUR LES FORMULES A, EXPERIMENTALE I OU
EXPERIMENTALE II DE L'EXAMEN INTERMEDIAIRE OTIS-
OTTAWA ADMINISTREES EN TRENTE MINUTES.



$N = 200$
 $\chi^2_{99(8)} = 20.09$
 $\chi^2_{\text{obtenu}} = 5.50$
 $RC_{Sk} = 0.505$
 $RC_{Ku} = 0.489$
 $Mo = 37$
 $Md = 38$
 $M = 38.59$
 $\sigma_M = 0.7$
 $\sigma = 9.99$
 $\sigma_{\sigma} = 0.5$

Figure 9.-

Polygones des fréquences théoriques et observées des scores bruts du premier échantillon sur la formule A de l'Examen intermédiaire Otis-Ottawa administrée en trente minutes.

APPENDICE 2

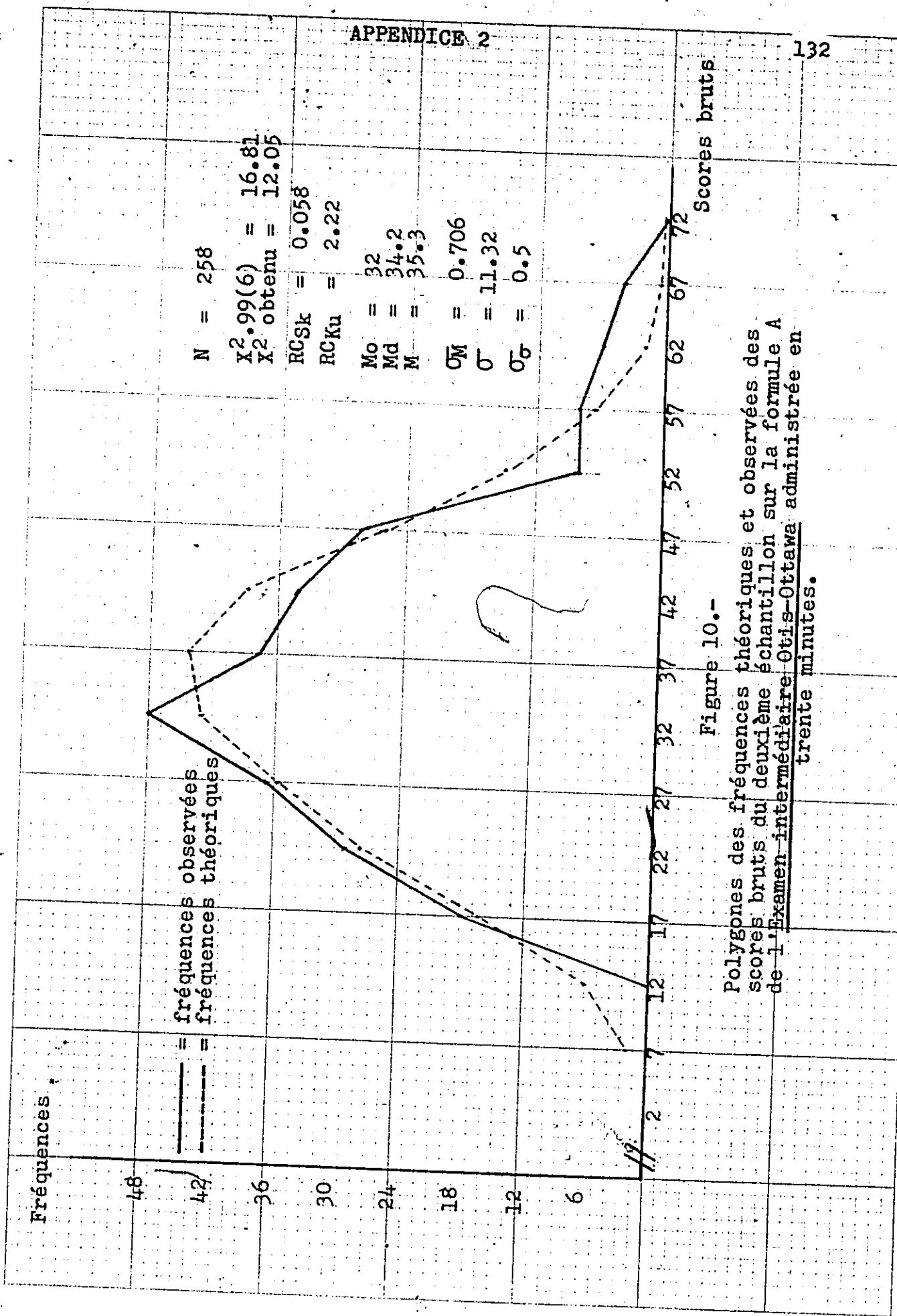
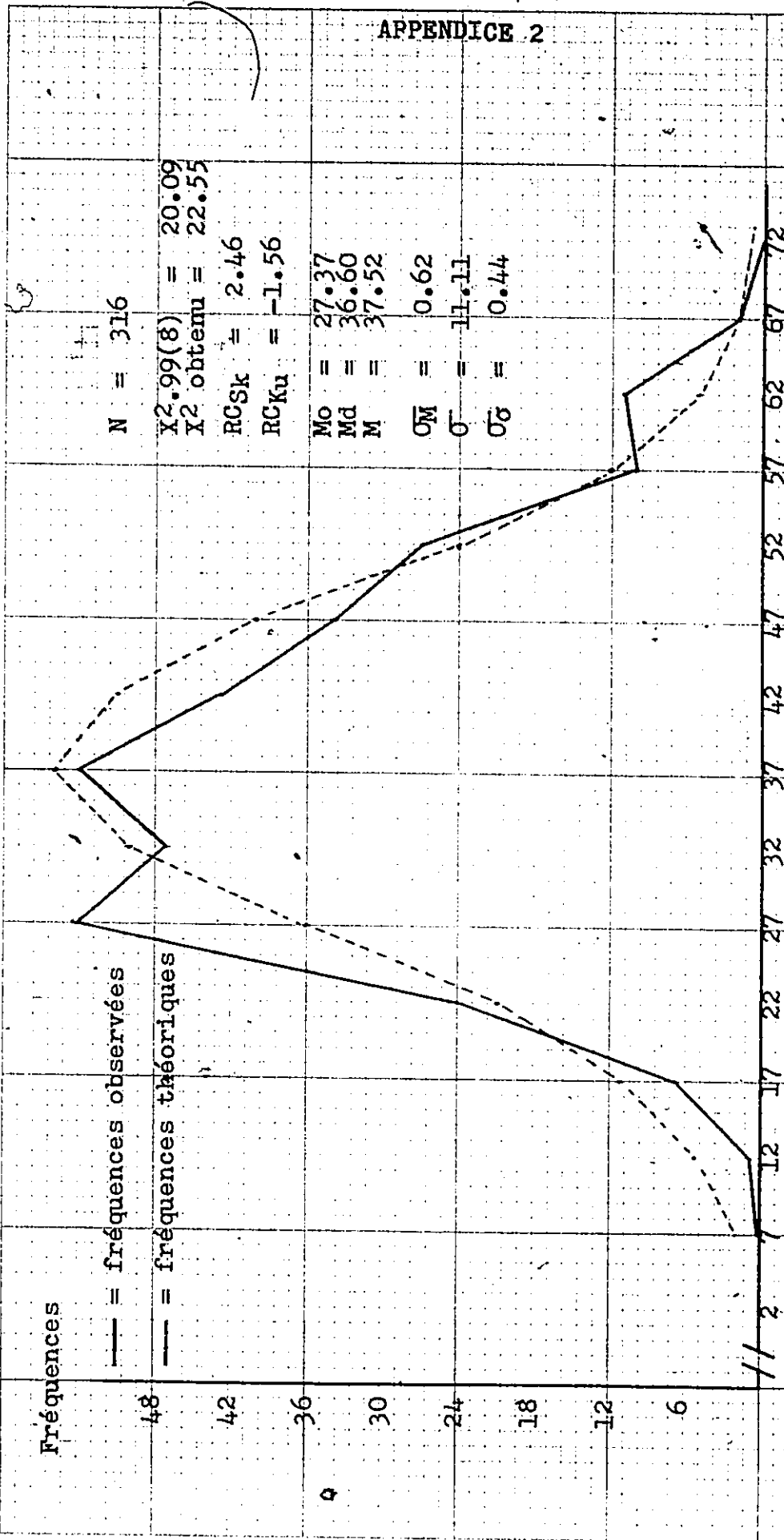


Figure 10.-

Polygones des fréquences théoriques et observées des scores bruts du deuxième échantillon sur la formule A de l'Examen intermédiaire Otis-Ottawa administrée en trente minutes.



N = 316

$\chi^2_{99(8)} = 20.09$

$\chi^2_{\text{obtenu}} = 22.55$

RCsk = 2.46

RCku = -1.56

Mo = 27.37

Md = 36.60

M = 37.52

$\sigma_M = 0.62$

$\sigma = 11.11$

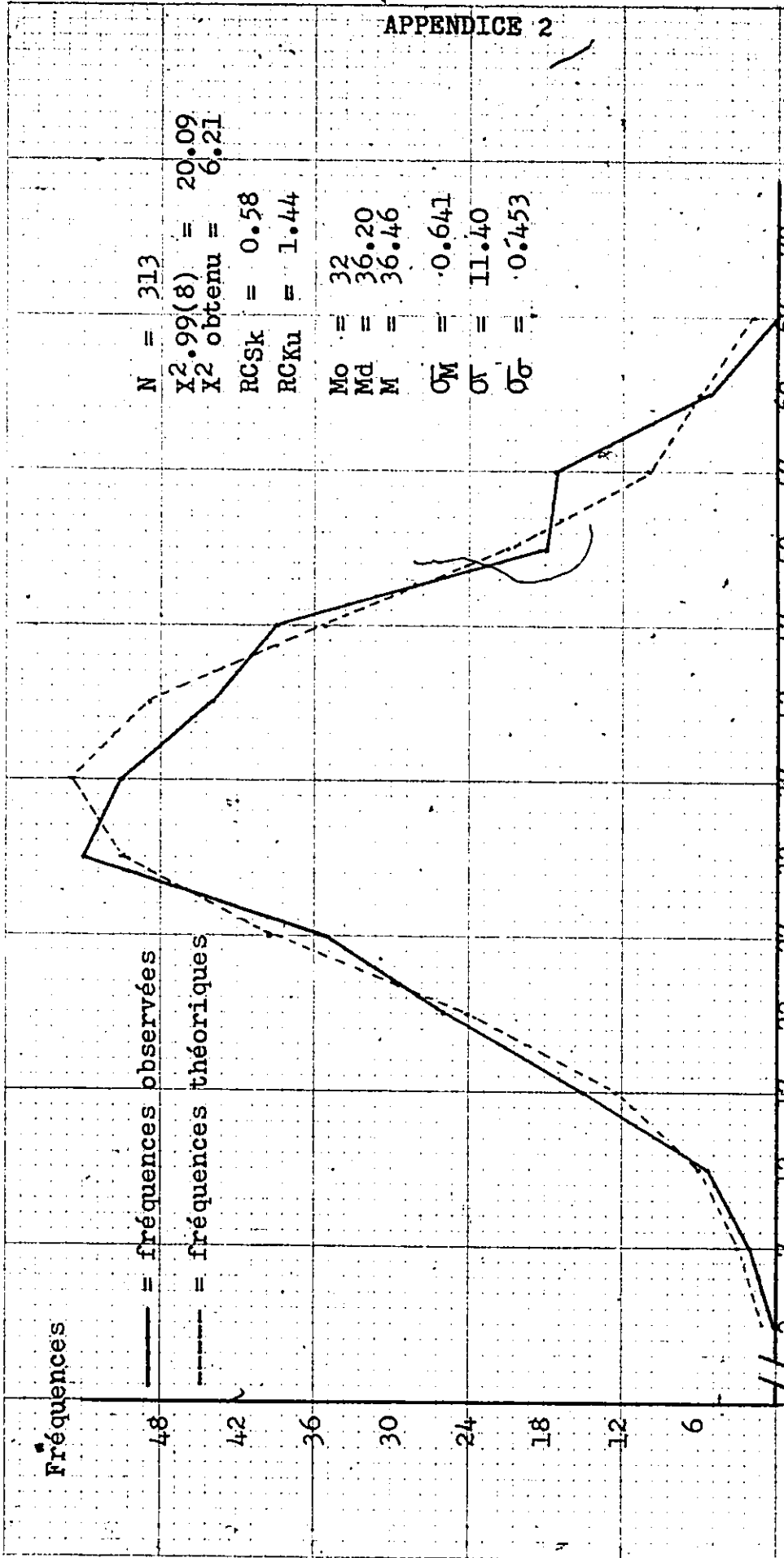
$\sigma_{\sigma} = 0.44$

Scores bruts

Figure 11.-

Polygones des fréquences théoriques et observées des scores bruts du troisième échantillon sur la formule A de l'Examen-intermédiaire Otis-Ottawa administrée en trente minutes.

APPENDICE 2



N = 313

$\chi^2_{99(8)} = 20.09$

$\chi^2_{\text{obtenu}} = 6.21$

RCSk = 0.58

RCKu = 1.44

Mo = 32

Md = 36.20

M = 36.46

$\overline{OM} = 0.641$

$\overline{OV} = 11.40$

$\overline{OG} = 0.453$

Scores bruts

Figure 12.-

Polygones des fréquences théoriques et observées des scores bruts du quatrième échantillon sur la formule expérimentale I de l'Examen intermédiaire Otis-Ottawa administrée en trente minutes.

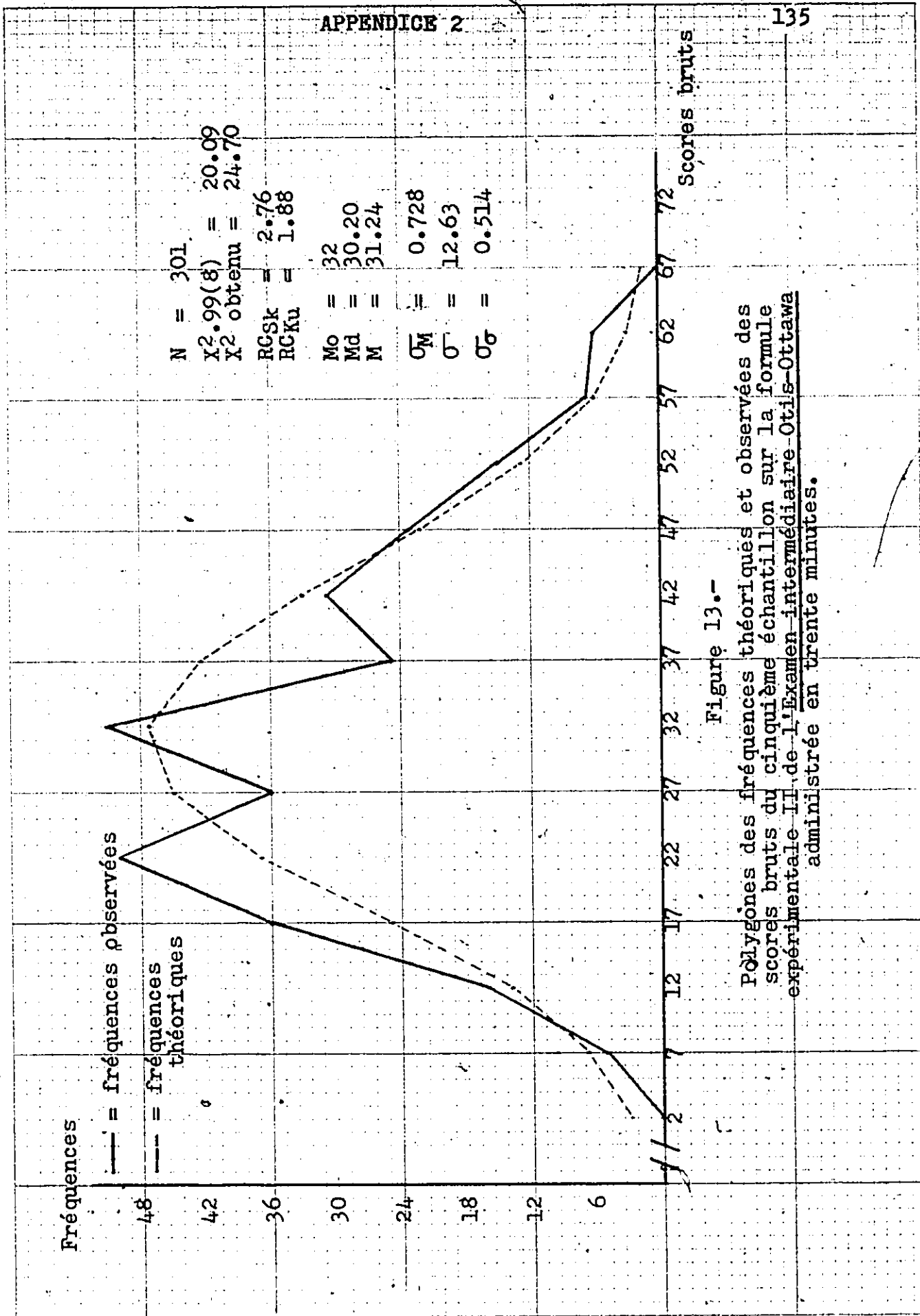


Figure 13.-

Polygones des fréquences théoriques et observées des scores bruts du cinquième échantillon sur la formule expérimentale II de l'Examen intermédiaire-Ottawa administrée en trente minutes.

APPENDICE 3

EXEMPLAIRES DES FORMULES A, EXPERIMENTALE I ET
EXPERIMENTALE II DE L'EXAMEN INTERMEDIAIRE
OTIS-OTTAWA D'HABILETE MENTALE.

DROIT D'AUTEUR RESERVÉ

APPENDICE 3, FEUILLES 136, 137, 138, 139,
140, 141, 142, 143, 144, 145, 146, 147, 148.

N'A PAS ETÉ MICROFILMÉ

PEUT ETRE OBTENU DE:

L'Institut de Psychologie de
l'Université d'Ottawa.

APPENDICE 4

STATISTIQUES QUE LES ECHANTILLONS III, IV ET V ONT
OBTENUES SUR LE CULTURE FAIR, SCALE 2, FORM A.

Tableau XII.- Statistiques des échantillons III, IV et V sur le Culture Fair, Scale 2, Form A.

Echantillons	III	IV	V
N	316	313	301
M âge	12 ans, 8 mois	12 ans, 8 mois	12 ans, 9 mois
$\chi^2_{.99(5)}=15.09^a$	13.25	(7)=18.48	(5) 4.08
Sk	.27	.15	.20
Sk	.138	.139	.141
RCSk	1.96	1.15	1.42
Ku	.27	.27	.03
Ku	.276	.278	.282
RCKu	.98	.97	.106
Mo	27.00	27.00	27.00
Md	26.40	25.70	26.30
M	26.00	25.40	26.20
σ_M	.287	.330	.305
σ	5.09	5.81	5.28
σ_σ	.203	.234	.215

a Tests "chi deux" entre les fréquences empiriques des scores bruts et les fréquences théoriques de la courbe normale correspondante.

Tableau XIII.

Description statistique de l'équivalence des échantillons
III, IV et V sur le Culture Fair, Scale 2, Form A.

Echantillons Statistiques	III/IV	III/V	IV/V
$\chi^2_{.99} (8) = 20.09^a$	11.36	6.24	5.64
DM	.6	.2	.8
σ_{DM}	.43	.41	.44
RC_{DM}	1.39	.50	1.81
$D\sigma$.72	.19	.53
$\sigma_{D\sigma}$.30	.28	.32
$RC_{D\sigma}$	2.40	.68	1.66

a Tests "chi deux" entre les fréquences empiriques
des scores bruts des groupes pris deux à deux.

APPENDICE 5

ANALYSE DE LA DIFFICULTE DES ITEM: PROPORTIONS
CORRIGÉES DE SUCCÈS SUR LES ITEM, RANGS DES
ITEM ET COEFFICIENTS DE CORRELATION ρ ENTRE
TOUS LES ORDRES DE DIFFICULTE DES ITEM.

Tableau XIV..

Proportions de succès, corrigées pour les réussites dues à la chance, obtenues sur les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes aux cinq échantillons.

Echantillons	I	II	III	IV	V
N	200	258	316	313	301
Formules	A	A	A	Ex.I	Ex.II
Item					
1	.8125	.8401	.8151	.8197	.7837
2	.9562	.9370	.9090	.9401	.9077
3	.7939	.7754	.8016	.7524	.7895
4	.8187	.7998	.7619	.8283	.6591
5	.7250	.6050	.6815	.7484	.7459
6	.9437	.9170	.9048	.8518	.8475
7	.8625	.8589	.8101	.8482	.8305
8	.9312	.8837	.8116	.8322	.7918
9	.7125	.7510	.6463	.6086	.5432
10	.8537	.8735	.8357	.8362	.7738
11	.7000	.5049	.5168	.4758	.4631
12	.3350	.3593	.3146	.4094	.2579
13	.8950	.7882	.8344	.7774	.8267
14	.1812	.2530	.2326	.2486	.1514
15	.9312	.9076	.8972	.8842	.8333
16	.6187	.5898	.5860	.5393	.4709
17	.8312	.6936	.6984	.6955	.6875
18	.4875	.4133	.4984	.3638	.3946
19	.8375	.7134	.7937	.8003	.7306
20	.7667	.7367	.7637	.6981	.6111
21	.8483	.6064	.7974	.6575	.6393
22	.7187	.8243	.8031	.8077	.6995
23	.5062	.4439	.4569	.5423	.4833
24	.3750	.3034	.3470	.3235	.3305
25	.6062	.6547	.5104	.4071	.4458

Tableau XIV (suite).

Proportions de succès, corrigées pour les réussites dues à la chance, obtenues sur les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes aux cinq échantillons.

Echantillons	I	II	III	IV	V
N	200	258	316	313	301
Formules	A	A	A	Ex.I	Ex.II
Item					
26	.8500	.7990	.7587	.7324	.7579
27	.2562	.3254	.2755	.0035	.0000
28	.8450	.8313	.8506	.8553	.8393
29	.7657	.7279	.7179	.6896	.6505
30	.6500	.7070	.6497	.6885	.6824
31	.7250	.7085	.6497	.6061	.5852
32	.3427	.3256	.3250	.4991	.4536
33	.0151	.0000	.0000	.1244	.0180
34	.6859	.6158	.6845	.5994	.6125
35	.3147	.3957	.3215	.2720	.4146
36	.5622	.4554	.4221	.4345	.4559
37	.4898	.4319	.3682	.3565	.4735
38	.0051	.0694	.0385	.0000	.1364
39	.6425	.5858	.4531	.4156	.6025
40	.3620	.3341	.3441	.3002	.2274
41	.1184	.1629	.0864	.0294	.0407
42	.7059	.7365	.7340	.7636	.6715
43	.4399	.4333	.3362	.2795	.3620
44	.7596	.7548	.7483	.7324	.7625
45	.6536	.5849	.4879	.2000	.2336
46	.7798	.7755	.6745	.6845	.6686
47	.5468	.4121	.5652	.4577	.6429
48	.6386	.5509	.5669	.4598	.4773
49	.6329	.5427	.4943	.4352	.5434
50	.1935	.1214	.0474	.0888	.0022

Tableau XIV (suite).

Proportions de succès, corrigées pour les réussites dues à la chance, obtenues sur les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes aux cinq échantillons.

Echantillons	I	II	III	IV	V
N	200	258	316	313	301
Formules	A	A	A	Ex.I	Ex.II
Item					
51	.5463	.5041	.4300	.4792	.5795
52	.4708	.3590	.4796	.3254	.4806
53	.6580	.5414	.4973	.2618	.3981
54	.6107	.4776	.4774	.4449	.3896
55	.3865	.3359	.2695	.2458	.2443
56	.2105	.2598	.2723	.2384	.2489
57	.6416	.5943	.5735	.2573	.3587
58	.2961	.4091	.4054	.3369	.3522
59	.3127	.1333	.2316	.1111	.1838
60	.5581	.3750	.5122	.4089	.3697
61	.1297	.0489	.1095	.0000	.1121
62	.8222	.6554	.7037	.4466	.6437
63	.5070	.2812	.3619	.3280	.2289
64	.2647	.1616	.1511	.2321	.1322
65	.2614	.0909	.2480	.1833	.0000
66	.0678	.0278	.0090	.0411	.0144
67	.2533	.0000	.2564	.1192	.1897
68	.0465	.1449	.0612	.1017	.1067
69	.2683	.1458	.1366	.2247	.0407
70	.6528	.3750	.6348	.4643	.5153
71	.2059	.1053	.0575	.0625	.0763
72	.2581	.1158	.1875	.1500	.1573
73	.3778	.5780	.4510	.4166	.4284
74	.3750	.1009	.4318	.2642	.1565
75	.0000	.0000	.0000	.0000	.0000

Tableau XV.

Rangs occupés par les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes, lorsqu'ils sont placés en ordre de difficulté croissante selon la proportion corrigée de succès obtenu par les cinq échantillons.

Echantillons	I	II	III	IV	V
N	200	258	316	313	301
Formules	A	A	A	Ex.I	Ex.II
Item					
1	14	7	8	9	9
2	1	1	1	1	1
3	16	13	10	14	8
4	15	10	14	7	20
5	22	28	22	15	12
6	2	2	2	3	2
7	6	6	9	4	4
8	4	4	4	8	7
9	23	15	25	24	30
10	7	5	6	6	10
11	26	38	33	30	38
12	55	50	55	43	52
13	5	12	7	12	5
14	68	59	62	57	62
15	3	3	3	2	6
16	35	29	29	27	36
17	11	23	20	18	16
18	46	44	36	44	44
19	12	20	13	11	14
20	18	17	15	19	25
21	9	27	11	23	22
22	24	9	12	10	15
23	44	41	42	28	33
24	53	56	51	48	51
25	36	24	34	42	39

Tableau XV (suite).

Rangs occupés par les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes, lorsqu'ils sont placés en ordre de difficulté croissante selon la proportion corrigée de succès obtenu par les cinq échantillons.

Echantillons N° Formules	I 200 A	II 258 A	III 316 A	IV 313 Ex.I	V 301 Ex.II
Item					
26	8	11	16	16	13
27	64	55	57	72	73
28	13	8	5	5	3
29	20	19	19	20	21
30	31	21	26	21	17
31	21	22	27	25	28
32	51	53	56	29	40
33	73	74	74	64	70
34	27	26	23	26	26
35	56	47	54	51	43
36	41	40	47	36	41
37	45	42	49	45	37
38	74	70	72	73	63
39	33	30	43	39	27
40	54	54	52	50	57
41	70	60	68	70	68
42	25	18	18	13	18
43	49	43	53	52	48
44	19	16	17	17	11
45	28	32	39	61	55
46	17	14	24	22	19
47	42	45	30	32	23
48	32	34	32	33	34
49	39	35	38	38	31
50	67	65	71	68	72

Tableau XV (suite).

Rangs occupés par les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en trente minutes, lorsqu'ils sont placés en ordre de difficulté croissante selon la proportion corrigée de succès obtenu par les cinq échantillons.

Echantillons	I	II	III	IV	V
N	200	258	316	313	301
Formules	A	A	A	Ex.I	Ex.II
Item					
51	40	37	45	31	29
52	47	51	40	49	35
53	29	36	37	53	45
54	37	39	41	37	46
55	50	52	58	56	54
56	65	58	59	58	53
57	34	31	31	54	49
58	58	46	48	46	50
59	57	64	63	66	59
60	38	48	35	41	47
61	69	71	67	74	65
62	10	25	21	35	24
63	43	57	50	47	56
64	60	61	65	59	64
65	62	69	60	62	74
66	71	72	73	71	71
67	59	73	61	65	58
68	72	63	69	67	66
69	61	62	66	60	69
70	30	49	28	34	32
71	66	67	70	69	67
72	63	66	64	63	60
73	52	33	44	40	42
74	48	68	46	55	61
75	75	75	75	75	75

Tableau XVI.

Proportions de succès, corrigées pour les réussites dues à la chance, obtenues sur les items des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en soixante minutes aux quatre derniers échantillons.

Echantillons	II	III	IV	V
N	258	316	313	301
Formules	A	A	Ex.I	Ex.II
Item				
1	.8401	.8101	.8203	.7840
2	.9370	.9090	.9401	.8833
3	.7723	.7983	.7524	.7756
4	.7965	.7627	.8283	.7491
5	.6027	.6796	.7484	.7468
6	.9176	.9051	.8482	.8500
7	.8547	.8101	.8482	.8090
8	.8837	.8616	.8322	.7924
9	.7432	.6440	.6086	.5432
10	.8735	.8362	.8366	.7738
11	.5010	.5174	.4768	.4643
12	.3669	.3281	.3643	.2596
13	.7829	.8323	.7732	.8239
14	.2539	.2326	.2581	.1486
15	.9079	.8972	.8842	.8421
16	.5833	.5847	.5367	.5083
17	.6899	.6994	.6965	.6570
18	.4138	.4976	.3411	.3978
19	.7190	.7943	.8003	.7342
20	.7377	.7637	.6981	.5961
21	.6021	.7969	.6571	.6137
22	.8207	.8062	.8083	.6885
23	.4477	.4541	.5198	.4809
24	.3023	.3473	.2686	.3355
25	.6560	.5134	.4050	.4394

Tableau XVI (suite).

Proportions de succès, corrigées pour les réussites dues à la chance, obtenues sur les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en soixante minutes aux quatre derniers échantillons.

Echantillons	II	III	IV	V
N	258	316	313	301
Formules	A	A	Ex.I	Ex.II
Item				
26	.7965	.7587	.7324	.7550
27	.3314	.2761	.0444	.0358
28	.8333	.8544	.8530	.8405
29	.7209	.7179	.6896	.6469
30	.7045	.6519	.6885	.6229
31	.7045	.6479	.6086	.5722
32	.3395	.3241	.4387	.4020
33	.0000	.0000	.0842	.0255
34	.6076	.6835	.5966	.6138
35	.3847	.3275	.2829	.2881
36	.4380	.4146	.4409	.4186
37	.4264	.3797	.3696	.4330
38	.0843	.0506	.0542	.1279
39	.5748	.4430	.4194	.6013
40	.3605	.3434	.3449	.2234
41	.1570	.1060	.0000	.0200
42	.7481	.7302	.7644	.6677
43	.4380	.3294	.2732	.3314
44	.7481	.7412	.7324	.7591
45	.5504	.4713	.2002	.2200
46	.7384	.6696	.6845	.6678
47	.4138	.5581	.4569	.5968
48	.5762	.5541	.4587	.4773
49	.5543	.4682	.4249	.5133
50	.1037	.0605	.0341	.0657

Tableau XVI (suite).

Proportions de succès, corrigées pour les réussites dues à la chance, obtenues sur les item des formules A, expérimentale I ou expérimentale II de l' xamen intermédiaire Otis-Ottawa administrées en soixante minutes aux quatre derniers échantillons.

Echantillons	II	III	IV	V
N	258	316	313	301
Formules	A	A	Ex. I	Ex. II
Item				
51	.4780	.4140	.4760	.5748
52	.3750	.4586	.3038	.4560
53	.5775	.4450	.2587	.3583
54	.4942	.5008	.4329	.3871
55	.3346	.2433	.2586	.2342
56	.2188	.2716	.2194	.2425
57	.6191	.5399	.2620	.3122
58	.3896	.3850	.3109	.3308
59	.0938	.1651	.0833	.1362
60	.5313	.4551	.4121	.3654
61	.0967	.0865	.0482	.1030
62	.7813	.6795	.4505	.5420
63	.2892	.3109	.2944	.2182
64	.1880	.1066	.1599	.1279
65	.0526	.1077	.0687	.0000
66	.0339	.0394	.0066	.0199
67	.1111	.1586	.0616	.1412
68	.1205	.0617	.0877	.1063
69	.1482	.1124	.2005	.0733
70	.4909	.5277	.4569	.5017
71	.0645	.0787	.0395	.0897
72	.1504	.1705	.1234	.1528
73	.4071	.3819	.3999	.4263
74	.2300	.2669	.3065	.1570
75	.0000	.0000	.0000	.0000

Tableau XVII.

Rangs occupés par les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en soixante minutes, lorsqu'ils sont placés en ordre de difficulté croissante selon la proportion corrigée de succès obtenue par les quatre derniers échantillons.

Echantillons N Formules Item	II 258 A	III 316 A	IV 313 Ex.I	V 301 Ex.II
1	7	8	9	8
2	1	1	1	1
3	14	11	14	9
4	10	14	8	13
5	28	23	15	14
6	2	2	3	2
7	6	9	4	6
8	4	4	7	7
9	17	27	24	30
10	5	6	6	10
11	37	33	29	36
12	51	51	44	52
13	12	7	12	5
14	58	60	54	59
15	3	3	2	3
16	30	28	27	31
17	24	20	18	19
18	45	36	46	42
19	20	13	11	15
20	19	16	19	24
21	29	12	23	22
22	9	10	10	16
23	41	41	28	35
24	56	49	53	47
25	25	34	41	38

Tableau XVII (suite).

Rangs occupés par les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en soixante minutes, lorsqu'ils sont placés en ordre de difficulté croissante selon la proportion corrigée de succès obtenue par les quatre derniers échantillons.

Echantillons N Formules Item	II 258 A	III 316 A	IV 313 Ex.I	V 301 Ex.II
26	11	15	16	12
27	54	56	70	70
28	8	5	5	4
29	21	19	20	20
30	22	25	21	21
31	23	26	25	28
32	53	54	35	43
33	74	74	64	71
34	27	22	26	23
35	49	53	51	51
36	42	45	36	41
37	44	47	43	39
38	70	72	68	63
39	31	43	38	25
40	52	50	45	55
41	62	67	74	72
42	15	18	13	17
43	43	52	52	48
44	16	17	17	11
45	34	37	59	56
46	18	24	22	18
47	46	29	32	26
48	32	30	31	34
49	35	38	39	32
50	67	70	72	69

Tableau XVII (suite).

Rangs occupés par les item des formules A, expérimentale I ou expérimentale II de l'Examen intermédiaire Otis-Ottawa administrées en soixante minutes, lorsqu'ils sont placés en ordre de difficulté croissante selon la proportion corrigée de succès obtenue par les quatre derniers échantillons.

Echantillons N Formules Item	II 258 A	III 316 A	IV 313 Ex.I	V 301 Ex.II
51	40	44	30	27
52	50	39	49	37
53	33	42	56	46
54	38	35	37	44
55	55	59	55	54
56	60	57	58	53
57	26	31	57	50
58	48	46	47	49
59	69	62	65	62
60	36	40	40	45
61	68	68	69	66
62	13	21	34	29
63	57	55	50	57
64	61	66	61	64
65	72	65	66	74
66	73	73	73	73
67	66	63	67	61
68	65	71	63	65
69	63	64	60	68
70	39	32	33	33
71	71	69	71	67
72	64	61	62	60
73	47	48	42	40
74	59	58	48	58
75	75	75	75	75

Tableau XVIII.

Coefficients de corrélation ρ entre tous les ordres de difficulté croissante et les ordres de présentation des item des formules A et expérimentale II.

Ordres	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
I A 30 (1)										
II A 30 (2)	.94									
III A 30 (3)	.97	.95								
IV Ex.I 30 (4)	.91	.92	.94							
V Ex.II 30 (5)	.92	.93	.94	.96						
II A 60 (6)	.96	.99	.96	.92	.93					
III A 60 (7)	.96	.96	.99	.94	.95	.97				
IV Ex.I 60 (8)	.92	.92	.94	.99	.96	.92	.94			
V Ex.II 60 (9)	.93	.94	.95	.97	.99	.94	.96	.97		
Formule A (10)	.61	.69	.65	.70	.66	.66	.68	.70	.67	
Form. Ex.II (11)	-.03	-.02	-.02	.05	.03	-.02	-.03	.06	.05	-.02

I, II, III, IV et V: échantillons
 A, Ex.I et Ex.II: formules
 30, 60: administration en trente ou soixante minutes
 Formule A: ordre de présentation des item de la formule A
 Form. Ex.II: ordre de présentation des item au hasard

APPENDICE 6

ANALYSE DE LA VALEUR DISCRIMINANTE DES ITEM:
RESULTATS DE L'ANALYSE PROBIT.

Tableau XIX.

Tests "chi deux" entre les scores probit empiriques et les scores probit théoriques de la fonction ogive normale, pour tous les item dans chacun des groupes.

Groupes	I		II		III		IV		V ●	
	A		A		A		Ex. 1		Ex. II	
	dl ^a	X ²	dl	X ²	dl	X ²	dl	X ²	dl	X ²
1	4	0.49	6	3.40	6	5.12	7	6.46	7	9.70
2	4	6.45	6	0.99	6	12.52	7	8.12	5	8.36
3	4	1.45	6	0.70	6	3.42	7	15.42	6	2.59
4	4	1.12	6	8.53	6	8.42	7	7.67	2	0.33
5	4	4.43	6	9.72	6	4.43	7	14.33	7	14.12
6	4	2.17	6	7.07 ^b	6	5.47	7	4.68	5	5.46
7	4	0.87	6	27.45 ^b	6	8.65	7	8.24	6	7.11 ^b
8	4	5.41	6	5.34	6	13.96	7	14.02	7	21.26 ^b
9	4	3.17 ^b	6	9.48	6	5.03	7	2.34	7	6.40
10	4	17.00 ^b	6	7.02	6	2.38	7	6.56	7	6.09
11	4	1.94	6	2.47	6	2.52	7	3.16	7	10.81
12	4	0.71	6	1.81	6	7.85	6	10.83	7	13.18
13	4	5.68	6	3.48	6	8.19	7	4.36	7	16.08
14	4	12.60	6	2.85	6	4.68	6	9.77	7	2.87
15	4	7.30	6	8.94	6	6.57	7	13.58	3	2.51
16	4	6.38	6	10.08	6	2.05	7	4.40	5	13.31
17	4	2.37	6	7.46	6	2.06	7	5.26	1	0.16
18	4	0.81	6	2.12	6	5.37	6	5.86	6	5.30
19	4	3.47	6	3.58	6	16.11	7	3.09	7	13.54
20	4	5.08	6	9.92	6	9.18	7	3.89	2	1.36
21	4	4.94	6	4.73	6	2.55	7	4.88	4	11.83
22	4	4.84	6	2.88	6	3.73	7	14.02	5	4.18
23	4	2.84	6	12.93	6	3.87	7	9.39	7	6.95
24	4	6.28	6	5.76	6	9.72	6	1.66	7	5.30
25	4	5.16	6	8.70	6	8.56	7	7.93	6	3.02

a degrés de liberté

b Tests "chi deux" significatifs à un niveau de probabilité de .01

Tableau XIX (suite).

Tests "chi deux" entre les scores probit empiriques et les scores probit théoriques de la fonction ogive normale, pour tous les item dans chacun des groupes.

Groupes	I		II		III		IV		V	
	A		A		A		Ex. I		Ex. II	
Form.	dl ^a	X ²	dl	X ²	dl	X ²	dl	X ²	dl	X ²
26	4	6.56	6	6.01	6	4.12	7	17.66	6	6.19
27	4	3.24	6	5.20	6	6.95	6	6.86	3	3.02
28	4	8.68	6	3.26	6	16.73	7	14.14	6	2.74
29	4	2.82	6	3.27	6	3.22	7	2.17	7	4.44
30	4	3.24	6	5.20	6	7.34	7	11.79	3	3.65
31	4	3.19	6	14.33	6	5.29	7	1.99	6	11.90
32	4	0.90	6	1.94	6	9.45	6	9.73	6	7.08
33	4	2.19	6	2.90	6	6.60	1	2.08 ^b	7	4.90
34	4	2.64	6	5.67	6	5.40	7	19.43 ^b	7	16.23
35	4	2.45	6	4.54	6	6.16	6	5.63	3	5.64
36	4	1.98	6	5.59	6	3.94	7	3.79	6	7.37
37	4	2.80	5	0.84	6	4.35	7	7.78	6	3.78
38	4	3.65	5	1.23	6	6.30	2	1.31	6	8.89
39	4	8.81	5	10.06	6	2.35	7	4.46	7	7.69
40	4	1.18	5	6.56	6	8.55	6	15.56	6	7.21
41	4	1.74	5	2.10	6	7.90	4	4.17	2	0.91
42	4	4.83	5	5.57	6	13.46	7	3.32 ^b	7	8.43
43	4	5.73	5	1.67	6	9.34	7	19.85 ^b	6	12.20
44	4	8.61	5	4.99	6	12.98	7	1.96	7	6.85
45	4	9.01	5	1.84	6	4.32	7	2.10	5	0.67
46	4	12.08	5	9.37	6	5.48	7	11.21	7	18.58 ^b
47	4	0.40	5	0.60	6	4.26	6	5.87	1	0.01
48	4	1.60	5	0.78	6	4.92	7	6.19	7	8.16
49	4	1.76	5	5.07	6	7.31	7	11.34	5	7.97
50	4	0.27	5	3.78	6	9.40	3	3.45	3	4.82

a degrés de liberté

b Tests "chi deux" significatifs à un niveau de probabilité de .01

Tableau XIX (suite).

Tests "chi deux" entre les scores probit empiriques et les scores probit théoriques de la fonction ogive normale, pour tous les item dans chacun des groupes.

Groupes	I		II		III		IV		V	
	A		A		A		Ex. I		Ex. II	
	dl ^a	X ²	dl	X ²	dl	X ²	dl	X ²	dl	X ²
51	3	0.75	4	4.28	6	3.75	7	4.58	7	4.33
52	3	0.74	4	2.12	6	14.80	6	6.81	6	4.15
53	3	2.53	4	1.77	6	8.84	7	16.18	6	4.29
54	3	1.29	4	2.00	5	9.17	7	4.72	7	5.09
55	3	0.47	4	2.56	5	9.40	6	8.88	3	2.15
56	3	6.58	4	2.19 _b	5	2.22	5	5.46	6	8.55
57	3	3.51	3	25.90 _b	5	11.15	7	6.82	6	3.00
58	2	1.12	3	1.37	5	7.20	6	7.68	3	1.24
59	2	0.06	2	2.69	5	1.37	2	7.70	6	8.03
60	2	0.14	2	0.26	5	2.65	7	6.58	7	3.80
61	1	0.30	2	1.60	5	2.18	3	2.02	7	15.05
62	1	1.71	2	1.44	5	12.77	7	3.42	3	4.66
63	1	1.71	2	2.65	5	5.69	6	1.47	5	3.00
64	1	0.34	2	5.79 _b	5	8.84	4	1.86	7	4.28
65	3	2.36	1	7.91 _b	4	1.96	2	4.98	7	5.20
66	3	1.26	1	0.03	3	3.34	2	1.20	6	4.99
67	1	4.85	1	6.24	3	0.47	3	0.68	3	1.50
68	1	0.21	1	0.01	2	0.77	3	0.91	7	3.17
69	1	0.00	2	0.42	2	1.94	4	3.58	2	1.65
70	3	0.21	2	0.20	1	0.03	7	15.58	7	7.93
71	3	0.40	2	5.37	2	0.76	2	0.11	6	8.86
72	3	3.22	3	5.38	1	1.85	3	4.74	7	11.89
73	2	5.14	4	5.35	1	0.96	6	8.68	7	7.47
74	2	3.63	3	2.23	1	4.16	5	7.70	7	7.31
75	-	-	3	4.29	1	0.04	2	1.28	7	3.49

a degrés de liberté
 b Tests "chi deux" significatifs à un niveau de probabilité de .01

Tableau XX.

Analyse probit: pentes des droites probit
des item dans chacun des groupes

Groupes	I	II	III	IV	V	
Formules	A	A	A	Ex. 1	Ex. II	
Item						
1	.029	.042	.035	(7) ^a	.027 (2) ^b	.030
2	.033	.042	.031	(1)	.025 (55)	.050
3	.066	.070	.058	(14)	.035 (34)	.039
4	.059	.064	.060	(10)	.068 (71)	.047
5	.055	.037	.045	(28)	.040 (22)	.032
6	.016	.043	.029	(2)	.026 (57)	.045
7	.105		.072	(6)	.062 (46)	.057
8	.029	.074	.073	(4)	.028 (25)	
9	.038	.059	.038	(17)	.054 (16)	.024
10		.035	.015	(5)	.011 (6)	.005
11	.039	.028	.033	(37)	.039 (12)	.027
12	.043	.057	.058	(51)	.073 (15)	.042
13	.029	.034	.041	(12)	.057 (17)	.033
14	-.005	.019	.005	(58)	.034 (11)	.004
15	.044	.013	.032	(3)	.064 (68)	.060
16	.031	.023	.017	(30)	.035 (58)	.025
17	.028	.058	.039	(24)	.053 (75)	.091
18	.018	.024	.023	(45)	.046 (47)	.036
19	.047	.029	.031	(20)	.059 (28)	.041
20	.041	.046	.038	(19)	.049 (73)	.030
21	.021	.042	.044	(29)	.054 (60)	.064
22	.021	.017	.013	(9)	.028 (53)	.044
23	.028	.036	.025	(41)	.034 (9)	.009
24	.036	.058	.033	(56)	.035 (13)	.041
25	.040	.045	.041	(25)	.033 (51)	.047

a Position de l'item dans la formule expérimentale I
b Position de l'item dans la formule expérimentale II

Tableau XX (suite).

Analyse probit: pentes des droites probit
des item dans chacun des groupes

Groupes	I	II	III	IV	V		
Formules	A	A	A	Ex. 1	Ex. II		
Item							
26	.050	.034	.055	(11) ^a	.071	(35) ^b	.041
27	.010	.024	.008	(54)	-.003	(66)	.012
28	.059	.058	.048	(8)	.040	(36)	.042
29	.028	.033	.025	(21)	.029	(64)	.093
30	.040	.029	.047	(22)	.055	(64)	.093
31	.021	.028	.033	(23)	.057	(48)	.033
32	.072	.062	.060	(53)	.086	(41)	.061
33	.023	.050	.043	(74)	.034	(8)	.032
34	.027	.048	.055	(27)	.064	(10)	.041
35	.020	.032	.034	(49)	.032	(63)	.036
36	.059	.070	.051	(42)	.071	(43)	.040
37	.030	.039	.037	(44)	.028	(52)	.027
38	.019	.028	.041	(70)	.025	(37)	.023
39	.057	.054	.060	(31)	.052	(7)	.036
40	.019	.023	.032	(52)	.028	(44)	.030
41	-.003	.017	.009	(62)	-.008	(72)	.022
42	.037	.030	.034	(15)	.054	(31)	.037
43	.073	.133	.080	(43)	.082	(45)	.069
44	.031	.035	.048	(16)	.044	(3)	.023
45	.072	.084	.080	(34)	.082	(56)	.079
46	.065	.041	.079	(18)	.051	(14)	.048
47	.033	.058	.055	(46)	.038	(74)	.064
48	.026	.039	.041	(32)	.063	(21)	.020
49	.077	.093	.061	(35)	.062	(54)	.064
50	.061	.072	.040	(67)	.060	(67)	.042

a Position de l'item dans la formule expérimentale I
b Position de l'item dans la formule expérimentale II

Tableau XX (suite).

Analyse probit: pentes des droites probit
des item dans chacun des groupes

Groupes	I	II	III	IV	V		
Formules	A	A	A	Ex. 1	Ex. II		
Item							
51	.043	.051	.031	(40) ^a	.031	(5) ^b	.024
52	.039	.089	.054	(50)	.042	(33)	.037
53	.097	.080	.079	(33)	.085	(38)	.057
54	.017	.000	.021	(38)	.022	(24)	.004
55	.058	.083	.051	(55)	.065	(69)	.053
56	.064	.056	.036	(60)	.037	(49)	.045
57	.105	.108	.093	(26)	.106	(40)	.072
58	.081	.050	.052	(48)	.043	(61)	.052
59	.063	.084	.057	(69)	.043	(50)	.047
60	.089	.103	.100	(36)	.076	(18)	.068
61	.003	.007	.028	(68)	.014	(30)	.011
62	.059	.078	.048	(13)	.086	(65)	.097
63	.067	.062	.051	(57)	.051	(59)	.040
64	.037	.044	.020	(61)	.055	(29)	.038
65	.116	.055	.076	(72)	.035	(19)	.057
66	.039	.028	-.001	(73)	.012	(42)	.010
67	.066	.113	.056	(66)	.060	(62)	.086
68	.072	.131	.048	(65)	.105	(26)	.045
69	.125	.024	.042	(63)	.035	(70)	.030
70	.057	.050	.065	(39)	.062	(23)	.047
71	.063	.038	.056	(71)	.088	(39)	.036
72	.099	.052	.040	(64)	.040	(20)	.037
73	.018	.031	.048	(47)	.054	(1)	.040
74	.010	.012	.060	(59)	.065	(27)	.035
75	----	.024	.001	(75)	.005	(4)	-.002

a Position de l'item dans la formule expérimentale I
b Position de l'item dans la formule expérimentale II

Tableau XXI.

Analyse probit: pour chacun des item, le nombre d'élèves des groupes I, II et III qui l'ont atteint, la pente commune aux trois groupes et le test F qui détermine si cette pente représente chaque groupe sans erreur significative

Item	Groupe I	Groupe II	Groupe III	Pente probit	Degrés liberté	Test F
1	200	258	316	.036	2,16	0.40
2	200	258	316	.035	2,16	0.33
3	200	258	316	.064	2,16	1.70
4	200	258	316	.061	2,16	0.08
5	200	258	316	.044	2,16	0.66
6	200	258	316	.031	2,16	0.64
7	200	258	316			
8	200	258	316	.063	2,16	0.66
9	200	258	316	.044	2,16	1.07
10	200	258	316			
11	200	258	316	.032	2,16	0.70
12	200	258	316	.054	2,16	0.83
13	200	258	316	.036	2,16	0.34
14	200	258	316	.008	2,16	1.63
15	200	258	316	.026	2,16	1.60
16	200	258	316	.022	2,16	0.60
17	200	258	316	.043	2,16	2.54
18	200	258	316	.022	2,16	0.27
19	200	258	316	.033	2,16	0.41
20	200	258	316	.041	2,16	0.35
21	200	258	316	.039	2,16	1.94
22	200	258	316	.017	2,16	0.28
23	200	258	316	.029	2,16	0.12
24	200	258	316	.041	2,16	1.94
25	200	258	316	.042	2,16	0.12

Tableau XXI (suite).

Analyse probit: pour chacun des item, le nombre d'élèves des groupes I, II et III qui l'ont atteint, la pente commune aux trois groupes et le test F qui détermine si cette pente représente chaque groupe sans erreur significative

Item	Groupe I	Groupe II	Groupe III	Pente probit	Degrés liberté	Test F
26	200	258	316	.046	2,16	0.96
27	200	258	316	.014	2,16	1.39
28	200	257	316	.054	2,16	0.09
29	200	257	316	.028	2,16	0.45
30	200	256	315	0.39	2,16	1.65
31	200	255	315	.029	2,16	0.32
32	199	254	315	.063	2,16	0.83
33	199	251	314	.040	2,16	2.65
34	199	250	314	.046	2,16	2.69
35	197	248	310	.030	2,16	0.80
36	197	244	309	.059	2,16	1.14
37	196	239	309	.036	2,15	0.47
38	196	234	308	.032	2,15	1.79
39	195	233	307	.058	2,15	0.40
40	192	230	305	.026	2,15	0.57
41	190	227	304	.009	2,15	1.48
42	187	225	297	.034	2,15	0.07
43	183	219	296	.089	2,15	6.12
44	182	214	293	.040	2,15	0.81
45	179	212	292	.079	2,15	1.19
46	176	206	289	.063	2,15	5.30
47	171	199	285	.050	2,15	4.75
48	166	197	277	.037	2,15	0.92
49	158	187	274	.072	2,15	2.42
50	155	182	272	.053	2,15	2.90

Tableau XXI (suite).

Analyse probit: pour chacun des item, le nombre d'élèves des groupes I, II et III qui l'ont atteint, la pente commune aux trois groupes et le test F qui détermine si cette pente représente chaque groupe sans erreur significative

Item	Groupe I	Groupe II	Groupe III	Pente probit	Degrés liberté	Test F
51	144	171	266	.039	2, 13	1.62
52	137	157	260	.059	2, 13	3.41
53	127	138	250	.083	2, 13	0.94
54	122	132	235	.015	2, 12	1.20
55	116	120	224	.060	2, 12	1.63
56	114	116	216	.046	2, 12	1.11
57	113	113	210			
58	103	107	196	.057	2, 10	1.37
59	97	95	183	.064	2, 9	1.37
60	86	87	170	.098	2, 9	0.37
61	79	85	158	.019	2, 8	2.39
62	75	81	154	.056	2, 8	0.82
63	71	76	144	.057	2, 8	0.27
64	68	71	139	.029	2, 8	0.71
65	66	64	127			
66	59	62	116	.013	2, 7	2.44
67	50	55	109	.070	2, 5	0.70
68	43	52	101	.079	2, 4	6.37
69	41	47	99	.042	2, 5	7.31
70	36	43	91	.059	2, 6	3.76
71	34	42	89	.048	2, 7	0.80
72	31	38	85	.050	2, 7	0.48
73	30	41	85	.039	2, 7	0.44
74	22	33	77	.038	2, 6	1.55
75	—	30	75	—	—	—

Tableau XXII.

Analyse probit: pour chacun des item, le nombre d'élèves des groupes III, IV et V qui l'ont atteint, la pente commune aux trois groupes et le test F qui détermine si cette pente représente chaque groupe sans erreur significative.

Item	Groupe III	Groupe IV	Groupe V	Pente probit	Degrés liberté	Test F
1	316	313	301	.030	2,20	0.03
2	316	313	179	.034	2,18	0.26
3	316	313	289	.043	2,19	1.54
4	316	313	89	.059	2,15	0.62
5	316	313	301	.038	2,20	0.78
6	316	313	169	.032	2,18	1.23
7	316	313	242	.064	2,19	0.29
8	316	313	301			
9	316	313	301	.037	2,20	7.10 ^a
10	316	313	301	.010	2,20	0.86
11	316	310	301	.033	2,20	1.07
12	316	244	301	.055	2,19	2.23
13	316	313	301	.044	2,20	2.37
14	316	179	301	.011	2,19	4.33
15	316	313	110	.051	2,16	0.45
16	316	313	168	.026	2,18	1.45
17	316	313	72	.051	2,14	5.60
18	316	283	232	.034	2,18	2.89
19	316	313	300	.043	2,20	2.98
20	316	313	81	.041	2,15	1.56
21	316	313	158	.053	2,17	0.39
22	316	313	188	.027	2,18	3.08
23	316	297	301	.021	2,20	4.19
24	316	196	301	.037	2,19	0.55
25	316	313	206	.039	2,19	0.95

^a Test F significatif à un niveau de probabilité de .01 .

Tableau XXII (suite).

Analyse probit: pour chacun des item, le nombre d'élèves des groupes III, IV et V qui l'ont atteint, la pente commune aux trois groupes et le test F qui détermine si cette pente représente chaque groupe sans erreur significative.

Item	Groupe III	Groupe IV	Groupe V	Pente probit	Degrés liberté	Test F
26	316	313	287	.055	2,19	2.02
27	316	212	120	.006	2,15	0.60
28	316	313	285	.043	2,19	0.09
29	316	313	294	.026	2,20	0.45
30	315	313	129	.058	2,16	3.20
31	315	313	229	.041	2,19	3.05
32	315	221	270	.066	2,18	1.56
33	314	67	301	.036	2,14	0.63
34	314	313	301			
35	310	261	132	.034	2,15	0.02
36	309	296	262	.053	2,19	5.23
37	309	288	200	.031	2,19	0.85
38	308	82	281	.030	2,14	1.75
39	307	313	301	.049	2,20	4.47
40	305	234	260	.030	2,18	0.04
41	304	136	87	.007	2,12	1.82
42	297	313	296	.041	2,20	2.04
43	296	292	250			
44	293	313	301	.037	2,20	3.55
45	292	310	173	.081	2,18	0.00
46	289	313	301			
47	285	277	70	.048	2,13	2.50
48	277	313	301	.038	2,20	10.31 ^a
49	274	310	182	.062	2,18	0.04
50	272	107	114	.046	2,12	0.50

^a Test F significatif à un niveau de probabilité de .01 .

Tableau XXII (suite).

Analyse probit: pour chacun des item, le nombre d'élèves des groupes III, IV et V qui l'ont atteint, la pente commune aux trois groupes et le test F qui détermine si cette pente représente chaque groupe sans erreur significative.

Item	Groupe III	Groupe IV	Groupe V	Pente probit	Degrés liberté	Test F
51	266	299	301	.029	2,20	0.54
52	260	253	290	.043	2,18	0.21
53	250	312	277	.072	2,19	0.91
54	235	306	301	.014	2,19	2.52
55	224	201	98	.057	2,14	0.12
56	216	153	226	.040	2,16	0.03
57	210	313	272	.089	2,18	1.68
58	196	266	147	.048	2,14	0.47
59	183	84	220	.049	2,13	0.61
60	170	310	301	.077	2,19	2.62
61	158	99	298	.016	2,15	0.87
62	154	313	123	.077	2,15	11.29 ^a
63	144	188	162	.047	2,16	0.66
64	139	142	299	.038	2,16	3.87
65	127	75	301	.056	2,13	4.18
66	116	73	265	.008	2,11	0.36
67	109	114	138	.067	2, 9	4.92
68	101	119	301	.053	2,12	4.63
69	99	131	95	.035	2, 8	0.33
70	91	304	301	.055	2,15	0.93
71	89	81	273	.044	2,10	1.55
72	85	124	301	.038	2,11	-0.02
73	85	271	301	.046	2,14	0.93
74	77	161	301	.046	2,13	2.43
75	75	71	301	-.0001	2,10	0.25

a Test F significatif à un niveau de probabilité de .01 .

APPENDICE 7

EXTRAIT DE

Influences de l'ordre de difficulté croissante des item sur
les qualités psychométriques d'un test d'intelligence du
type puissance à temps limité, par Jean-Guy Bonin.

APPENDICE 7

EXTRAIT DE

Influences de l'ordre de difficulté croissante des item sur les qualités psychométriques d'un test d'intelligence du type puissance à temps limité¹

La rareté des écrits et les conclusions d'une recherche antérieure, faite par l'auteur, sur la séquence des item d'un test d'intelligence à temps limité ont conduit à l'étape actuelle qui consiste en la vérification expérimentale des influences exercées par la présentation des item en ordre de difficulté croissante. Ces influences sont exprimées en termes d'élévation de la moyenné des scores, d'augmentation de la constance du test, d'amélioration de sa validité externe et de dispersion plus étendue des scores.

Pour ce faire, trois ordres différents de présentation des item des Examens Otis-Ottawa d'Habilitété Mentale, Intermédiaire, Formule A ont été utilisés. Dans un premier mode de présentation des item, la Formule A a été conservée dans sa forme originale. Une Formule B en a été dérivée en modifiant la présentation des item en un nouvel ordre de difficulté croissante. La loi du hasard régit la distribution de ces mêmes item dans la Formule C. Le test Culture

¹ Jean-Guy Bonin, Thèse de maîtrise présentée à la Faculté de Psychologie de l'Université d'Ottawa, 1971, xv-166 p.

Fair, Scale 2, Form A a été choisi pour répondre à deux buts bien précis: établir l'équivalence des groupes et servir de critère externe de validité. Les succès scolaires en français et en mathématiques de sixième et septième année ont servi de ~~seul~~ critère externe de validité pour les Otis Formules A, B et C. L'étude a été réalisée avec 926 étudiants de septième année, répartis en trois groupes de onze classes chacun, de la ville de Hull et des environs. Les groupes I, II et III passent respectivement les Formules A, B et C de l'Otis.

Les résultats indiquent une élévation significative de la moyenne des scores bruts des Formules A (groupe I) et B (groupe II). Ceci porte à croire qu'une présentation des item en ordre de difficulté croissante est importante pour un test d'intelligence avec temps limite. Cependant, quand le test en est un de puissance, ce mode de présentation semble perdre de son importance.

Aucune augmentation significative de la constance n'a été observée, sinon une tendance positive en faveur de la Formule B.

L'absence de différence significative entre les Formules A, B et C et le Culture Fair, entre ces mêmes Formules et les succès scolaires de sixième année jette le doute sur l'importance de l'ordre de difficulté croissante des item dans sa capacité d'améliorer la validité externe

d'un test. Par ailleurs, ce même ordre de présentation des item semble être un facteur déterminant dans cette amélioration lorsque les Formules A et B, en trente et soixante minutes sont mises en corrélation avec le français de septième année. Toutefois, ceci n'est que partiellement vrai quand les mathématiques de septième année servent de critère externe.

Loin d'accroître la variabilité des individus, la présentation des item en ordre de difficulté croissante a tendance à restreindre leur dispersion autour de la moyenne.

Dans les conditions de l'étude rapportée ici, la présentation des item en ordre de difficulté croissante n'a pu confirmer clairement son importance que dans l'élévation de la moyenne des scores bruts d'un test d'intelligence à temps limité.

APPENDICE 8

ABSTRACT OF

L'ordre de difficulté croissante des item dans un test de
puissance à temps limité, l'Examen intermédiaire Otis-Ottawa
d'habileté mentale: sa constance et ses effets sur la
valeur discriminante des item.

APPENDICE 8

L'ordre de difficulté croissante des item dans un test de puissance à temps limité, l'Examen intermédiaire Otis-Ottawa d'habileté mentale: sa constance et ses effets sur la valeur discriminante des item.¹

The aim of the present research was to study the order of item difficulty in a timed-power test, the Examen intermédiaire Otis-Ottawa d'habileté mentale, formule A, an adaptation of the Otis Self-Administering Tests of Mental Ability. The order of item difficulty is an essential quality of this valuable test and the problem was to find a reliable order of increasing difficulty within the period of time required by the standardization of the test and to measure the effects of the order of difficulty on the discriminant value of the items. This study was prompted by the absence of agreement among authors on the way to measure the difficulty of an item in a timed-power test, and by the lack of experiments demonstrating the effects of an order of item difficulty on the discriminant value of the items of such a test.

The experiment was conducted in three stages. First, the original test was administered in thirty minutes to 200

¹ André R. Pellerin, thèse présentée à l'École des Etudes Supérieures en vue de l'obtention du Ph.D. en psychologie clinique, Université d'Ottawa, Canada, 1974, xiv-186 p.

seventh grade students according to the procedures of the standardization. The results of this administration showed that the items were not in order of increasing difficulty. Secondly, the original test was given with a sixty minute time limit to a new equivalent group of 258 students; the purpose of this longer time limit was to eliminate the speed factor. From these results, a new order of increasing difficulty was established; a new form was printed according to this order as was a form which presents the items randomly. Thirdly, the original form and the two new forms were administered in thirty and sixty minutes to three groups of 316, 313 and 301 students randomly distributed by classes.

Analysis of the data shows that the orders of increasing difficulty calculated from the five thirty minute administrations are highly reliable, that they are not significantly different from the orders of increasing difficulty calculated from the four sixty minute administrations, but that they are significantly different from the original order of item difficulty presented in Form A.

The study reveals a few signs suggesting that an order of increasing difficulty could enhance the discriminant value of the items more than a random order of item difficulty. However, these signs fade under the scrutiny of the statistical analysis and lose their significance.

From a theoretical point of view, this experiment shows that an item analysis of items in orders of difficulty in a timed-power test could give valid and reliable results even though a number of the subjects do not attempt every item. This conclusion was unexpected because the review of the literature did not point in that direction. This study also introduces the probit analysis as a more adequate method than the biserial and point-biserial correlational methods to measure the item discriminant value in a timed-power test because it seems unbiased by the position an item holds in the test.