

Université de Montréal

Influence de la formulation des items d'un questionnaire d'enquête sur les réponses  
fournies par les répondants et sur l'échelle de mesure associée aux catégories de réponse

par

Julie Grondin

Département d'administration et fondements de l'éducation

Faculté des sciences de l'éducation

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès arts (M. A.)

en mesure et évaluation en éducation

Août 2006

© Julie Grondin, 2006



LB

5

U57

2006

V.028

**Direction des bibliothèques**

**AVIS**

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

**NOTICE**

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal  
Faculté des études supérieures

Ce mémoire intitulé :

Influence de la formulation des items d'un questionnaire d'enquête sur les réponses  
fournies par les répondants et sur l'échelle de mesure associée aux catégories de réponse

présenté par

Julie Grondin

a été évalué par un jury composé des personnes suivantes :

Michel D. Laurier

---

président-rapporteur

Jean-Guy Blais

---

directeur de recherche

Nathalie Loye

---

membre du jury

## RÉSUMÉ

L'utilisation des enquêtes par questionnaire pour recueillir des données sur les attitudes et les opinions amène les chercheurs à s'interroger sur les facteurs d'influence pouvant contaminer cet instrument de mesure. Plusieurs facteurs d'influence ont été étudiés. L'objectif de cette recherche consiste à étudier un de ces facteurs : l'influence de la formulation des items. En effet, cette recherche souhaite explorer l'influence d'un changement de mots dans la formulation des items sur les réponses des participants de même que sur l'échelle de mesure associées aux catégories de réponse du questionnaire. Pour ce faire, le questionnaire d'enquête mis au point par le Centre de formation initiale des maîtres de l'Université de Montréal afin de recueillir l'opinion et l'attitude des étudiants et des étudiantes du programme d'enseignement en éducation préscolaire et enseignement primaire a été utilisé. Deux versions de ce questionnaire ont été distribuées. Dans l'une des versions, un changement de mot a été appliqué, mais sans changer le sens général des items. Pour l'analyse des données, le modèle de mesure *Rating Scale* de la famille des modèles de Rasch est mis à contribution.

## MOTS CLÉS

Modèle, Rasch, mesure, opinion, enquête, questionnaire

## **SUMMARY**

The use of questionnaires to gather information on the attitudes and opinions brings some concerns to survey methodologist regarding the influential factors that can cause contamination to this instrument. Many influential factors have been studied. The objective of this research consists in exploring one of these factors: the influence of the formulation of the items. In fact, this research wishes to analyse the influence of a change of words in the formulation of the items on the answers of the participants and also on the measurement scale associated to the categories of answers of the questionnaire. To do so, the questionnaire inquiry put out by the Centre de formation initiale des maîtres of the University of Montreal to gather the opinions and attitudes of students in the kindergarden and primary teaching program has been used. Two versions of this inquiry have been distributed. In one of these versions, a change of words has been applied but without changing the general meaning of the items. For the analysis of the data, the Rating Scale model of the Rasch's family of models was made to contribution.

## **KEYWORDS**

Model, Rasch, measurement, opinion, survey, questionnaire

# TABLE DES MATIÈRES

<b>LISTE DES TABLEAUX</b> .....	<b>IX</b>
<b>LISTE DES FIGURES</b> .....	<b>XIII</b>
<b>LISTE DES SIGLES ET DES ABRÉVIATIONS</b> .....	<b>XVII</b>
<b>REMERCIEMENTS</b> .....	<b>XVIII</b>
<b>INTRODUCTION</b> .....	<b>1</b>
<b>1. PROBLÉMATIQUE</b> .....	<b>3</b>
<b>2. OBJECTIFS ET QUESTIONS DE RECHERCHE</b> .....	<b>12</b>
2.1 OBJECTIF GÉNÉRAL.....	12
2.2 OBJECTIF SPÉCIFIQUE.....	12
2.3 QUESTIONS DE RECHERCHE.....	13
<b>3. CADRE CONCEPTUEL</b> .....	<b>14</b>
3.1 CHOIX TERMINOLOGIQUE.....	15
3.2 FACTEURS D'INFLUENCE SUR LES RÉPONSES RECUEILLIES À L'AIDE D'UNE ENQUÊTE PAR QUESTIONNAIRE AUTO-ADMINISTRÉ.....	17
3.2.1 <i>Processus de réponse</i> .....	17
3.2.2 <i>Caractéristiques du questionnaire</i> .....	21
3.2.3 <i>Ordre de présentation des items</i> .....	23
3.2.4 <i>Échelles de réponses en catégories ordonnées</i> .....	26
3.2.5 <i>Formulation des items</i> .....	29
3.3 MODÈLES DE MESURE.....	32
3.3.1 <i>Modèles de mesure de la famille des modèles de Rasch</i> .....	37
3.3.1.1 <i>Le modèle Rating Scale</i> .....	44
3.3.1.2 <i>Le modèle Partial Credit</i> .....	48

3.3.2 Hypothèses et propriété des modèles de la famille des modèles de Rasch....	49
3.3.2.1 Hypothèse de l'unidimensionnalité.....	49
3.3.2.2 Hypothèse de l'indépendance locale.....	51
3.3.2.3 Invariance (propriété) .....	51
3.3.3 Ajustement entre les données et le modèle .....	53
3.3.3.1 Ajustement pour les items .....	57
3.3.3.2 Ajustement pour les participants .....	60
3.4 SYNTHÈSE DU CADRE CONCEPTUEL .....	61
<b>4. MÉTHODOLOGIE.....</b>	<b>64</b>
4.1 DESCRIPTION DE L'INSTRUMENT DE CUEILLETTE D'INFORMATION.....	64
4.2 POPULATION ÉTUDIÉE ET CUEILLETTE DES DONNÉES .....	70
4.3 CHOIX MÉTHODOLOGIQUE POUR LE TRAITEMENT DES DONNÉES.....	71
<b>5. LIMITES DE LA RECHERCHE .....</b>	<b>73</b>
<b>6. ANALYSE DES DONNÉES.....</b>	<b>75</b>
6.1 DÉMARCHE D'ANALYSE .....	75
6.1.1 Indices statistiques .....	75
6.1.2 Critères d'analyse .....	83
6.1.3 Stratégie d'analyse choisie.....	89
6.2 ANALYSES DES DONNÉES .....	92
6.2.1 Questionnaires 2003.....	93
6.2.1.1 Qualité de l'ajustement pour les questionnaires 2003 .....	93
6.2.1.2 Comparaisons entre les versions A et B du questionnaire 2003 .....	105
6.2.2 Questionnaires 2004.....	115
6.2.2.1 Qualité de l'ajustement pour les questionnaires 2004 .....	115
6.2.2.2 Comparaisons entre les versions A et B du questionnaire 2004 .....	119



6.2.3	Questionnaires 2005.....	122
6.2.3.1	Qualité de l'ajustement pour les questionnaires 2005 .....	122
6.2.3.2	Comparaisons entre les versions A et B du questionnaire 2005 .....	125
6.2.4	Analyses complémentaires.....	128
6.2.4.1	Qualité de l'ajustement de la version A du questionnaire 2004 .....	128
6.2.4.2	Comparaisons entre les versions A et B du questionnaire 2004 .....	130
6.2.4.3	Qualité de l'ajustement de la version A du questionnaire 2005 .....	131
6.2.4.4	Qualité de l'ajustement de la version B du questionnaire 2005 .....	133
6.2.4.5	Comparaisons entre les versions A et B du questionnaire 2005 .....	135
<b>7.</b>	<b>DISCUSSION.....</b>	<b>137</b>
7.1	RETOUR SUR L'ANALYSE DES DONNÉES .....	137
7.2	INTERPRÉTATION DES RÉSULTATS .....	141
7.2.1	Résultats de l'analyse de la qualité de l'ajustement.....	142
7.2.2	Résultats pour les questionnaires 2003 .....	144
7.2.3	Résultats pour les questionnaires 2004 .....	147
7.2.4	Résultats pour les questionnaires 2005 .....	151
7.2.5	Résultats pour les analyses complémentaires .....	153
7.3	RETOUR SUR L'OBJECTIF ET LES QUESTIONS DE RECHERCHE.....	155
7.3.1	Retour sur la première question de recherche .....	156
7.3.2	Retour sur la deuxième question de recherche.....	157
	<b>CONCLUSION .....</b>	<b>159</b>
	<b>RÉFÉRENCES.....</b>	<b>161</b>
	<b>ANNEXES.....</b>	<b>167</b>
	ANNEXE A : VERSION A DU QUESTIONNAIRE 2003 .....	168
	ANNEXE B : FORMULATION DES ITEMS POUR LES QUESTIONNAIRES 2003, 2004 ET 2005 ....	175

---

ANNEXE C : TABLEAUX DE L'ANALYSE DES DONNÉES .....	178
<i>C.1. Tableaux complémentaires de l'analyse des questionnaires 2003 .....</i>	<i>178</i>
<i>C.2. Tableaux de l'analyse des questionnaires 2004 .....</i>	<i>188</i>
<i>C.3. Tableaux de l'analyse des questionnaires 2005 .....</i>	<i>210</i>
ANNEXE D : TABLEAUX DES ANALYSES COMPLÉMENTAIRES .....	231
<i>D.1. Version A du questionnaire 2004 .....</i>	<i>231</i>
<i>D.2. Comparaison entre les versions A et B du questionnaire 2004 .....</i>	<i>237</i>
<i>D.3. Version A du questionnaire 2005 .....</i>	<i>242</i>
<i>D.4. Version B du questionnaire 2005 .....</i>	<i>248</i>
<i>D.5. Comparaison entre les versions A et B du questionnaire 2005 .....</i>	<i>253</i>

# LISTE DES TABLEAUX

TABLEAU 1 : RÉSULTATS POSSIBLES DÉCOULANT DE LA COMPARAISON ENTRE LA FORCE DE DEUX ATHLÈTES.....	39
TABLEAU 2 : RÉPONSES POSSIBLES D'UNE PERSONNE EN FONCTION DE SON AMOUR POUR LES DEVOIRS.....	45
TABLEAU 3 : STATISTIQUES D'AJUSTEMENT FINALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2003 .....	95
TABLEAU 4 : STATISTIQUES D'AJUSTEMENT DES ITEMS POUR LA VERSION A DU QUESTIONNAIRE 2003.....	97
TABLEAU 5 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2003 .....	98
TABLEAU 6 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES ITEMS DE LA VERSION A DU QUESTIONNAIRE 2003 .....	100
TABLEAU 7 : POINTS D'ANCRAGE DE LA VERSION A DU QUESTIONNAIRE 2003.....	103
TABLEAU 8 : SYNTHÈSE DES PRINCIPAUX RÉSULTATS DE L'ANALYSE DE LA QUALITÉ DE L'AJUSTEMENT POUR LES QUESTIONNAIRES 2003, 2004 ET 2005.....	139
TABLEAU 9 : SYNTHÈSE DES PRINCIPAUX RÉSULTATS DE LA COMPARAISON ENTRE LES VERSIONS A ET B DES QUESTIONNAIRES 2003, 2004 ET 2005 .....	140
TABLEAU 10 : FORMULATION DES ITEMS POUR LES VERSIONS A ET B DU QUESTIONNAIRE 2003 .....	175
TABLEAU 11 : FORMULATION DES ITEMS POUR LES VERSIONS A ET B DU QUESTIONNAIRE 2004 .....	176
TABLEAU 12 : FORMULATION DES ITEMS POUR LES VERSIONS A ET B DU QUESTIONNAIRE 2005 .....	177
TABLEAU 13 : STATISTIQUES D'AJUSTEMENT INITIALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2003 .....	178
TABLEAU 14 : STATISTIQUES D'AJUSTEMENT INITIALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2003 .....	180

TABLEAU 15 : STATISTIQUES D'AJUSTEMENT FINALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2003 .....	182
TABLEAU 16 : STATISTIQUES D'AJUSTEMENT DES ITEMS POUR LA VERSION B DU QUESTIONNAIRE 2003 .....	184
TABLEAU 17 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2003 .....	185
TABLEAU 18 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES ITEMS DE LA VERSION B DU QUESTIONNAIRE 2003 .....	185
TABLEAU 19 : POINTS D'ANCRAGE DE LA VERSION B DU QUESTIONNAIRE 2003 .....	187
TABLEAU 20 : STATISTIQUES D'AJUSTEMENT INITIALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	188
TABLEAU 21 : STATISTIQUES D'AJUSTEMENT INITIALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2004 .....	190
TABLEAU 22 : STATISTIQUES D'AJUSTEMENT FINALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	192
TABLEAU 23 : STATISTIQUES D'AJUSTEMENT FINALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2004 .....	194
TABLEAU 24 : STATISTIQUES D'AJUSTEMENT DES ITEMS POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	196
TABLEAU 25 : STATISTIQUES D'AJUSTEMENT DES ITEMS POUR LA VERSION B DU QUESTIONNAIRE 2004 .....	197
TABLEAU 26 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	198
TABLEAU 27 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2004 .....	198
TABLEAU 28 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES ITEMS POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	199
TABLEAU 29 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES ITEMS POUR LA VERSION B DU QUESTIONNAIRE 2004 .....	199
TABLEAU 30 : POINTS D'ANCRAGE DE LA VERSION A DU QUESTIONNAIRE 2004 .....	202
TABLEAU 31 : POINTS D'ANCRAGE DE LA VERSION B DU QUESTIONNAIRE 2004 .....	202

TABLEAU 32 : STATISTIQUES D'AJUSTEMENT INITIALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	210
TABLEAU 33 : STATISTIQUES D'AJUSTEMENT INITIALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	212
TABLEAU 34 : STATISTIQUES D'AJUSTEMENT FINALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	214
TABLEAU 35 : STATISTIQUES D'AJUSTEMENT FINALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	216
TABLEAU 36 : STATISTIQUES D'AJUSTEMENT DES ITEMS POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	217
TABLEAU 37 : STATISTIQUES D'AJUSTEMENT DES ITEMS POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	218
TABLEAU 38 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	219
TABLEAU 39 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	219
TABLEAU 40 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES ITEMS POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	220
TABLEAU 41 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES ITEMS POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	220
TABLEAU 42 : POINTS D'ANCRAGE DE LA VERSION A DU QUESTIONNAIRE 2005 .....	223
TABLEAU 43 : POINTS D'ANCRAGE DE LA VERSION B DU QUESTIONNAIRE 2005 .....	223
TABLEAU 44 : STATISTIQUES D'AJUSTEMENT FINALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	231
TABLEAU 45 : STATISTIQUES D'AJUSTEMENT FINALES DES ITEMS POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	233
TABLEAU 46 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	234
TABLEAU 47 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES ITEMS POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	234
TABLEAU 48 : POINTS D'ANCRAGE DE LA VERSION A DU QUESTIONNAIRE 2004 .....	236

TABLEAU 49 : STATISTIQUES D'AJUSTEMENT FINALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	242
TABLEAU 50 : STATISTIQUES D'AJUSTEMENT FINALES DES ITEMS POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	244
TABLEAU 51 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES PERSONNES POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	245
TABLEAU 52 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES ITEMS POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	245
TABLEAU 53 : POINTS D'ANCRAGE DE LA VERSION A DU QUESTIONNAIRE 2005 .....	247
TABLEAU 54 : STATISTIQUES D'AJUSTEMENT FINALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	248
TABLEAU 55 : STATISTIQUES D'AJUSTEMENT DES ITEMS POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	249
TABLEAU 56 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES PERSONNES POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	250
TABLEAU 57 : STATISTIQUES D'AJUSTEMENT GÉNÉRALES DES ITEMS POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	250
TABLEAU 58 : POINTS D'ANCRAGE DE LA VERSION B DU QUESTIONNAIRE 2005 .....	252

## LISTE DES FIGURES

FIGURE 1 : SCHÉMA SYNTHÈSE DU CADRE CONCEPTUEL .....	63
FIGURE 2 : DISTRIBUTION DES PERSONNES SUR LE CONTINUUM DE L'HABILITÉ ET DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LA VERSION A DU QUESTIONNAIRE 2003 .....	101
FIGURE 3 : COURBES DE PROBABILITÉS DE CHACUNE DES CATÉGORIES DE RÉPONSE DE LA VERSION A DU QUESTIONNAIRE 2003 .....	104
FIGURE 4 : DIAGRAMME DE DISPERSION DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2003 PAR RAPPORT À LEUR NIVEAU DE DIFFICULTÉ .....	106
FIGURE 5 : DIAGRAMME DE COMPARAISON MONTRANT LA DISTRIBUTION DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LES VERSIONS A ET B DU QUESTIONNAIRE 2003.....	108
FIGURE 6 : COURBES DE DISTRIBUTION DES MESURES ESTIMÉES PAR LE MODÈLE POUR LE NIVEAU DE DIFFICULTÉ DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2003 .....	110
FIGURE 7 : DIAGRAMME EN BÂTON REPRÉSENTANT LA VALEUR DES DIFFÉRENCES ENTRE LES MESURES ESTIMÉES PAR LE MODÈLE POUR LES VERSION A ET B DU QUESTIONNAIRE 2003 .....	111
FIGURE 8 : DIAGRAMME DE COMPARAISON MONTRANT LA POSITION DES POINTS D'ANCRAGE ESTIMÉS PAR LE MODÈLE POUR CHACUN DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2003.....	114
FIGURE 9 : DISTRIBUTION DES PERSONNES SUR LE CONTINUUM DE L'HABILITÉ ET DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LA VERSION B DU QUESTIONNAIRE 2003.....	186
FIGURE 10 : COURBES DE PROBABILITÉS DE CHACUNE DES CATÉGORIES DE RÉPONSE DE LA VERSION B DU QUESTIONNAIRE 2003 .....	187
FIGURE 11 : DISTRIBUTION DES PERSONNES SUR LE CONTINUUM DE L'HABILITÉ ET DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	200

FIGURE 12 : DISTRIBUTION DES PERSONNES SUR LE CONTINUUM DE L'HABILITÉ ET DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LA VERSION B DU QUESTIONNAIRE 2004 .....	201
FIGURE 13 : COURBES DE PROBABILITÉS DE CHACUNE DES CATÉGORIES DE RÉPONSE DE LA VERSION A DU QUESTIONNAIRE 2004 .....	203
FIGURE 14 : COURBES DE PROBABILITÉS DE CHACUNE DES CATÉGORIES DE RÉPONSE DE LA VERSION B DU QUESTIONNAIRE 2004 .....	204
FIGURE 15 : DIAGRAMME DE DISPERSION DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2004 PAR RAPPORT AU NIVEAU DE DIFFICULTÉ DES ITEMS.....	205
FIGURE 16 : DIAGRAMME DE COMPARAISON MONTRANT LA DISTRIBUTION DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LES VERSIONS A ET B DU QUESTIONNAIRE 2004.....	206
FIGURE 17 : COURBES DE DISTRIBUTION DES MESURES ESTIMÉES PAR LE MODÈLE POUR LE NIVEAU DE DIFFICULTÉ DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2004 .....	207
FIGURE 18 : DIAGRAMME EN BÂTON REPRÉSENTANT LA VALEUR DES DIFFÉRENCES ENTRE LES MESURES ESTIMÉES PAR LE MODÈLE POUR LES VERSIONS A ET B DU QUESTIONNAIRE 2004 .....	208
FIGURE 19 : DIAGRAMME DE COMPARAISON MONTRANT LA POSITION DES POINTS D'ANCRAGE ESTIMÉS PAR LE MODÈLE POUR CHACUN DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2004 .....	209
FIGURE 20 : DISTRIBUTION DES PERSONNES SUR LE CONTINUUM DE L'HABILITÉ ET DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	221
FIGURE 21 : DISTRIBUTION DES PERSONNES SUR LE CONTINUUM DE L'HABILITÉ ET DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	222
FIGURE 22 : COURBES DE PROBABILITÉS DE CHACUNE DES CATÉGORIES DE RÉPONSE DE LA VERSION A DU QUESTIONNAIRE 2005 .....	224
FIGURE 23 : COURBES DE PROBABILITÉS DE CHACUNE DES CATÉGORIES DE RÉPONSE DE LA VERSION B DU QUESTIONNAIRE 2005 .....	225



FIGURE 24 : DIAGRAMME DE DISPERSION DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2005 PAR RAPPORT AU NIVEAU DE DIFFICULTÉ DES ITEMS.....	226
FIGURE 25 : DIAGRAMME DE COMPARAISON MONTRANT LA DISTRIBUTION DES VERSIONS A ET B DU QUESTIONNAIRE 2005 SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ.....	227
FIGURE 26 : COURBES DE DISTRIBUTION DES MESURES ESTIMÉES PAR LE MODÈLE POUR LE NIVEAU DE DIFFICULTÉ DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2005 .....	228
FIGURE 27 : DIAGRAMME EN BÂTON REPRÉSENTANT LA VALEUR DES DIFFÉRENCES ENTRE LES MESURES ESTIMÉES PAR LE MODÈLE POUR LES VERSIONS A ET B DU QUESTIONNAIRE 2005 .....	229
FIGURE 28 : DIAGRAMME DE COMPARAISON MONTRANT LA POSITION DES POINTS D'ANCRAGE ESTIMÉS PAR LE MODÈLE POUR CHACUN DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2005 .....	230
FIGURE 29 : DISTRIBUTION DES PERSONNES SUR LE CONTINUUM DE L'HABILITÉ ET DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LA VERSION A DU QUESTIONNAIRE 2004 .....	235
FIGURE 30 : COURBES DE PROBABILITÉS DE CHACUNE DES CATÉGORIES DE RÉPONSE DE LA VERSION A DU QUESTIONNAIRE 2004 .....	236
FIGURE 31 : DIAGRAMME DE DISPERSION DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2004 PAR RAPPORT AU NIVEAU DE DIFFICULTÉ DES ITEMS.....	237
FIGURE 32 : DIAGRAMME DE COMPARAISON MONTRANT LA DISTRIBUTION DES VERSIONS A ET B DU QUESTIONNAIRE 2004 SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ.....	238
FIGURE 33 : COURBES DE DISTRIBUTION DES MESURES ESTIMÉES PAR LE MODÈLE POUR LE NIVEAU DE DIFFICULTÉ DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2004 .....	239
FIGURE 34 : DIAGRAMME EN BÂTON REPRÉSENTANT LA VALEUR DES DIFFÉRENCES ENTRE LES MESURES ESTIMÉES PAR LE MODÈLE POUR LES VERSIONS A ET B DU QUESTIONNAIRE 2004 .....	240
FIGURE 35 : DIAGRAMME DE COMPARAISON MONTRANT LA POSITION DES POINTS D'ANCRAGE ESTIMÉS PAR LE MODÈLE POUR CHACUN DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2004 .....	241

FIGURE 36 : DISTRIBUTION DES PERSONNES SUR LE CONTINUUM DE L'HABILITÉ ET DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LA VERSION A DU QUESTIONNAIRE 2005 .....	246
FIGURE 37 : COURBES DE PROBABILITÉS DE CHACUNE DES CATÉGORIES DE RÉPONSE DE LA VERSION A DU QUESTIONNAIRE 2005 .....	247
FIGURE 38 : DISTRIBUTION DES PERSONNES SUR LE CONTINUUM DE L'HABILITÉ ET DES ITEMS SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ POUR LA VERSION B DU QUESTIONNAIRE 2005 .....	251
FIGURE 39 : COURBES DE PROBABILITÉS DE CHACUNE DES CATÉGORIES DE RÉPONSE DE LA VERSION B DU QUESTIONNAIRE 2005 .....	252
FIGURE 40 : DIAGRAMME DE DISPERSION DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2005 PAR RAPPORT AU NIVEAU DE DIFFICULTÉ DES ITEMS.....	253
FIGURE 41 : DIAGRAMME DE COMPARAISON MONTRANT LA DISTRIBUTION DES VERSIONS A ET B DU QUESTIONNAIRE 2005 SUR LE CONTINUUM DU NIVEAU DE DIFFICULTÉ.....	254
FIGURE 42 : COURBES DE DISTRIBUTION DES MESURES ESTIMÉES PAR LE MODÈLE POUR LE NIVEAU DE DIFFICULTÉ DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2005 .....	255
FIGURE 43 : DIAGRAMME EN BÂTON REPRÉSENTANT LA VALEUR DES DIFFÉRENCES ENTRE LES MESURES ESTIMÉES PAR LE MODÈLE POUR LES VERSIONS A ET B DU QUESTIONNAIRE 2005 .....	256
FIGURE 44 : DIAGRAMME DE COMPARAISON MONTRANT LA POSITION DES POINTS D'ANCRAGE ESTIMÉS PAR LE MODÈLE POUR CHACUN DES ITEMS DES VERSIONS A ET B DU QUESTIONNAIRE 2005 .....	257

## LISTE DES SIGLES ET DES ABRÉVIATIONS

Pourcentage d'observations (% d'obs.) .....	103
Courbe caractéristique de l'item (CCI).....	36
Centre de formation initiale des maîtres (CFIM).....	64
Corrélation (Corr.) .....	95
Écart-type (É.-T.).....	96
Fonction caractéristique de l'item (FCI) .....	35
Grand dictionnaire terminologique (GDT).....	15
Kuder-Richardson 20 (KR-20) .....	81
Maximum (MAX.) .....	98
Ministère de l'éducation du Québec (MEQ).....	66
Minimum (MIN.).....	98
Carré moyen (mean square) (mnsq).....	77
Moyenne (MOY.).....	96
Erreur type moyenne (mean standard error) (mse) .....	80
Nombre d'observations (Nb. d'obs.).....	103
Office québécois de la langue française (OQLF).....	15
Partial Credit (PC) .....	10
Rating Scale (RS) .....	10
Théorie classique des tests (TCT) .....	34
Théorie de la réponse à l'item (TRI) .....	35
Indice d'ajustement standardisé (standardized fit index) (zstd).....	78

## REMERCIEMENTS

Je tiens tout particulièrement à remercier mon directeur de recherche Jean-Guy Blais pour toutes les occasions de découvertes et d'enrichissement qu'il m'a offertes. Il a su me guider, me conseiller et m'encourager tout au long de ce rêve maintenant devenu réalité.

Merci aux membres du jury pour les conseils et les commentaires constructifs formulés à l'égard de mon projet de recherche.

Un grand merci à ma famille pour son soutien inconditionnel, ses encouragements et pour son infinie compréhension. Un grand merci également à mon conjoint, qui a bien voulu prendre part à ce rêve, et qui s'est lancé sans hésiter dans cette aventure avec moi.

Enfin, merci à Mathieu, Samuel, Maude et Emmie pour leurs rires joyeux qui ont su m'émerveiller et m'encourager à persévérer.

## INTRODUCTION

L'être humain est de nature curieuse et sa soif de savoir est insatiable. Il peut s'interroger à propos de la couleur d'une pomme, de la montée des taux d'intérêt ou des causes d'un glissement de terrain. Les divers objets, événements ou phénomènes qui l'entourent sont susceptibles d'éveiller sa curiosité. Or, « le processus scientifique s'ouvre habituellement par une interrogation » (Jones, 2000, p. 13). Ainsi, la soif d'apprendre de l'être humain peut enclencher tout un processus de recherche scientifique.

Un processus de recherche scientifique est constitué de diverses activités. De prime abord, la personne s'intéresse à un objet ou à un phénomène et elle s'interroge sur celui-ci. Ensuite, la personne peut être amenée à examiner de façon attentive l'objet à l'origine de son questionnement, à l'explorer. Cette exploration consiste à observer, identifier les faits, les nommer, les comparer et les classer selon des critères (Lamoureux, 2000). L'exploration d'un objet d'étude fait généralement émerger des idées, des intuitions, des possibilités d'explication concernant le comment, le pourquoi ou le quand de cet objet (Jones, 2000). Or, le premier objectif de la science consiste à comprendre la réalité, c'est-à-dire à acquérir des connaissances dans le but d'explorer cette réalité, de la décrire et de l'expliquer (Lamoureux, 2000, p. 9).

Pour atteindre cet objectif, diverses techniques peuvent être mises en œuvre. Les méthodes de recherche utilisées par le chercheur deviennent alors des instruments qui lui permettent de mieux connaître l'objet de son étude, de mieux le décrire ou de mieux l'expliquer. Le choix de la méthode de recherche dépend donc des questions auxquelles le chercheur souhaite répondre (Jones, 2000). Or, « aucune méthode n'est meilleure qu'une autre; chacune est idéale dans les conditions qui en commandent l'utilisation » (Lamoureux,

2000). Ainsi, c'est en fonction d'un contexte de recherche particulier qu'un instrument peut devenir plus approprié qu'un autre.

Par conséquent, la méthode de recherche doit être choisie en fonction de ses caractéristiques. Le chercheur doit évaluer les avantages et les inconvénients de chacune des méthodes et choisir l'instrument qui possède les caractéristiques les plus susceptibles de l'aider à saisir l'objet de son étude et à répondre à ses questions. Or, le choix de l'instrument de recherche doit également pouvoir être mis à l'épreuve des faits (Jones, 2000) : les données qu'il permet de recueillir sur l'objet d'étude sont-elles valables? Les résultats qu'il produit sont-ils valables? Est-ce qu'il serait possible pour un autre chercheur de reproduire ces données (ou ces résultats)? Est-ce que les données recueillies (ou les résultats obtenus) parviennent à rendre suffisamment compte de l'objet étudié pour permettre de mieux le comprendre (objectif de la science)?

C'est à travers un tel questionnement que le chercheur peut prendre un certain recul par rapport au choix de l'instrument et vérifier que la méthode choisie lui permet de fournir une réponse rigoureuse et objective aux questions qu'il se pose. Le choix d'un instrument de recherche est donc une étape déterminante dans le processus de recherche. En effet, la recherche scientifique doit se traduire par une approche systématique et tendre vers l'objectivité (Jones, 2000). Or, plusieurs facteurs peuvent influencer de façon importante les divers instruments de recherche utilisés par un chercheur. Quels sont ces facteurs? Dans quelle mesure peuvent-ils influencer les résultats d'une recherche?

Le but du présent travail est d'explorer ces facteurs et d'acquérir des connaissances sur un des facteurs pouvant influencer un instrument de recherche. C'est donc dans le prolongement d'efforts visant une plus grande objectivité et une meilleure connaissance des instruments de recherche que le présent travail se situe.

# 1. PROBLÉMATIQUE

La nature curieuse de l'être humain est vaste et ses centres d'intérêt ne se limitent pas seulement à quelques objets ou à quelques phénomènes. Il peut s'intéresser à son environnement, à ses enfants, à sa famille, à ses finances, aux affaires politiques, à la biologie ou à la chimie. Tous les objets, événements ou phénomènes qui l'entourent sont susceptibles d'éveiller sa curiosité. Or, parmi les sujets auxquels il s'intéresse, plusieurs traitent de choses qui lui sont familières et qui font partie de sa vie quotidienne (Jones, 2000) : les conflits entre individus, les facteurs de réussite scolaire, les causes de divorce, etc. Ces points d'intérêts sont généralement regroupés sous une même étiquette, c'est-à-dire qu'ils font partie de ce qui s'appelle la recherche sociale.

La recherche sociale a donc pour objectif d'étudier l'être humain, ses comportements, ses aptitudes, ses relations avec les autres, etc. Une des particularités de ce domaine d'étude, est que les humains sont des êtres continuellement en évolution, qui s'adaptent aux situations qui leur sont présentées, qui ont des opinions, des sentiments et qui interprètent la réalité qui les entoure à leur manière. Les objets d'étude des sciences sociales ne possèdent généralement pas de caractéristiques physiques palpables comme la longueur, la couleur, le poids ou la texture.

Les variables qui relèvent du social sont plus souvent des caractéristiques psychologiques qui sont observables de façon indirecte comme l'habileté, l'intelligence ou l'attitude. Ainsi, afin de mieux documenter leur objet d'étude, les chercheurs ont dû mettre au point divers instruments tels que les tests, l'observation ou les simulations. D'autres instruments, tels que les questionnaires d'enquête et les entretiens, ont été mis au point

afin de permettre aux chercheurs de connaître leur objet d'étude, mais d'un autre point de vue, c'est-à-dire du point de vue de la personne interrogée.

Chacun de ces instruments possède des caractéristiques qui lui sont propres et s'utilise dans un contexte particulier. Par conséquent, pour faire son choix, le chercheur doit étudier les caractéristiques de chacun de ces instruments et opter pour celui qui est le plus susceptible de l'aider à saisir son objet d'étude. En effet, aucune méthode n'est parfaite. Cependant, lorsqu'elle est utilisée dans les bonnes conditions, une méthode peut être idéale par rapport à une autre (Lamoureux, 2000). C'est le cas entre autres pour les questionnaires d'enquête. En effet, Blais et Durand (1997, p. 359), de même que Jones (2000, p. 169), soutiennent qu'il s'agit de l'instrument de collecte de données le plus utilisé en recherche sociale.

Les enquêtes par questionnaires sont donc très populaires auprès des chercheurs. Pour certains d'entre eux, cela n'est pas surprenant, car les enquêtes par questionnaires sont très flexibles pour la présentation et la distribution, elles sont polyvalentes et elles permettent aux chercheurs d'obtenir de l'information sur certains sujets privés (politique, perception, opinion, attitude, etc.).

Ainsi, lorsqu'un chercheur utilise l'enquête par questionnaire comme outil de collecte de données, il souhaite généralement se faire une idée de l'opinion des gens sur un certain sujet ou de leur attitude par rapport à une situation donnée. Pour recueillir leurs données, les chercheurs vont fréquemment proposer un choix de réponses aux sujets interrogés. Ce choix correspond normalement à différentes catégories de réponse dont la signification serait saisie à peu près de la même façon par différents groupes de personnes. De plus, les catégories offertes suivent généralement une gradation à laquelle on associe une valeur numérique et une étiquette.



Par exemple, les choix offerts pourraient être du type « 1 = Totalemment en désaccord », « 2 = Plutôt en désaccord », « 3 = Plutôt d'accord » et « 4 = Totalemment d'accord ». Ces catégories de réponse ordonnées sont mieux connues sous le nom d'échelle de Likert. L'utilisation de ce genre de catégories de réponse est avantageuse pour la construction du questionnaire et est généralement facile à comprendre par le répondant. Cependant, « depuis les premiers jours où des chercheurs ont fait appel à la technique de l'enquête pour recueillir des données sur les attitudes et les opinions, on s'est interrogé sur le type de contamination qui pouvait être induit par la procédure » (J.-G. Blais, communication personnelle, 18 octobre, 2004).

En effet, les chercheurs sont conscients que l'utilisation des enquêtes par questionnaire permet d'obtenir des informations riches et variées, mais que plusieurs facteurs peuvent influencer les réponses des sujets interrogés. Par exemple, l'environnement dans lequel les sujets répondent au questionnaire, leur condition physique (fatigue, stress), les pressions sociales, des difficultés de lecture ou de compréhension de la langue utilisée dans le questionnaire, etc. Il existe donc plusieurs facteurs à travers ou autour d'un questionnaire d'enquête qui peuvent avoir une influence sur les réponses des participants.

Parmi tous ces facteurs, plusieurs sont directement liés aux caractéristiques du questionnaire et de ses items. Par exemple, l'influence peut provenir du texte d'introduction présenté aux participants au début du questionnaire (Tourangeau & Rasinski, 1988; Wänke, 2002), de l'ordre de présentation des items (Maheux, 2001; Schuman & Presser, 1996; Tourangeau & Rasinski, 1988), des catégories de réponse de l'échelle proposées (Bradburn & Sudman, 1979; Lam & Kolic, 2006; Schwarz, 1995) ou de la formulation des items (Kahneman & Tversky, 1982; Ng, Pipe, Beath & Holton, 1999).

Les caractéristiques du questionnaire et de ses items constituent donc des facteurs qu'un chercheur doit considérer s'il souhaite utiliser une enquête par questionnaire dans sa recherche. En effet, bien que plusieurs chercheurs aient étudié leurs influences potentielles, les conditions sous-jacentes ne sont toujours pas bien comprises. L'influence demeure complexe à identifier et la direction que prendra cette influence reste difficile à prévoir. Les chercheurs ont donc encore beaucoup de choses à découvrir sur le sujet.

Schuman et Presser (1996) soutiennent cette idée en précisant, qu'à leur avis, l'ordre des items semble être un des secteurs les plus importants de la recherche méthodologique. En effet, les résultats obtenus par différents chercheurs ne sont pas toujours convergents. Il serait donc pertinent de continuer les recherches sur le sujet afin d'approfondir les connaissances sur ce facteur d'influence.

De la même façon, l'influence des catégories de réponse de l'échelle proposée dans le questionnaire doit être étudiée davantage. Tout d'abord, Schwarz (1995), a pu constater qu'une échelle de réponses dont les codes sont des nombres positifs, par exemple de 0 à 10, offre une signification différente aux répondants qu'une échelle dont les codes sont des nombres à la fois négatifs et positifs (de -5 à 5 par exemple) et que cette différence dans l'interprétation que font les sujets de la signification de l'échelle proposée est une source d'influence sur les réponses obtenues.

En effet, la première échelle admet la présence d'un élément nul, mais l'échelle est positive. La seconde échelle admet autant la présence de réponses négatives que de réponses positives. Ainsi, cette dernière échelle permet, par exemple, à des répondants d'indiquer l'ampleur de leur endettement ou de leur richesse alors que la première échelle leur permet seulement de préciser l'ampleur de leur richesse, sinon son absence. Les codes numériques associés aux différentes catégories de l'échelle de réponses peuvent

donc avoir une influence sur l'interprétation que les répondants font de l'échelle et par conséquent, avoir une influence sur leur réponse.

Ensuite, selon Bradburn et Sudman (1979), les catégories offertes dans les échelles de réponses sont généralement ordonnées, mais les intervalles entre chacune des catégories sont inégaux. Par exemple, les résultats de leur recherche ont montré que l'intervalle entre « D'accord » et « Plutôt d'accord » est plus petit que celui entre « Plutôt d'accord » et « Totalement d'accord ». Autrement dit, leur recherche montre que les points d'ancrage de chacune des catégories de l'échelle de réponses ne correspondent pas à la valeur de l'étiquette numérique qui lui est généralement associée. De plus, les résultats de leur recherche ont montré que le nombre de catégories offertes pourrait avoir un effet sur la valeur de l'écart entre chacune des catégories. Cette hypothèse reste cependant à confirmer.

Enfin, l'influence de la formulation des items a pu être observée dans différentes recherches. À titre d'exemple, citons l'expérience rapportée par Rugg (1941). Il s'agit d'une expérience faite aux États-Unis où deux questions similaires, mais dont la formulation était différente, ont été posées au hasard dans la population. Les résultats ont dévoilé que la proportion de la population qui était prête à « ne pas permettre » plutôt que d'« interdire » des discours politiques contre la démocratie était plus grande d'environ 20 %. Cet exemple montre bien que le choix des mots peut avoir un impact important. Ainsi, les chercheurs savent que le choix des mots dans la formulation d'un item peut avoir une influence sur les réponses des sujets. Cependant, il existe très peu d'écrits leur permettant de mieux comprendre la portée de cette influence.

Tous les résultats qui viennent d'être présentés soulèvent plusieurs questions : l'ordre de présentation des items influence-t-il les réponses des participants? Sommes-nous certains que tous les répondants comprennent vraiment la même chose des

catégories de réponse proposées dans une échelle de réponses (étiquettes et codes numériques associés)? Dans quelle mesure le nombre de catégories offertes dans une échelle de réponses influence-t-il les réponses recueillies? Et finalement, dans quelle mesure la formulation des items influence-t-elle les réponses des participants?

Le manque de réponses à ces questions souligne l'importance de poursuivre les recherches sur les facteurs d'influence pouvant contaminer les données recueillies à l'aide d'enquêtes par questionnaire. En effet, il apparaît pertinent d'approfondir les connaissances actuelles sur le sujet afin de permettre aux chercheurs en science sociale de développer un instrument de recherche performant et qui permet d'obtenir des réponses rigoureuses et objectives. Or, parmi tous les facteurs d'influence présentés, l'influence de la formulation des items est celui pour lequel les connaissances actuelles sont les plus limitées. Ainsi, le but du présent travail est d'explorer ce facteur afin d'aider les chercheurs en science sociale à mieux comprendre la portée de son influence.

Pour bien comprendre la portée de l'influence de la formulation des items sur les réponses des participants, il est important d'établir de quelle façon ce facteur doit être exploré et quels sont les éléments à observer. En effet, l'influence de la formulation des items peut se manifester de plusieurs façons. Par exemple, il peut se manifester à travers la formulation des questions (exemple de Rugg, 1941) ou à travers les catégories de réponse de l'échelle proposée (exemples de Bradburn et Sudman, 1979, et de Schwarz, 1995). Ces deux aspects doivent donc être pris en considération pour mieux comprendre dans quelle mesure la formulation des items influence les réponses des participants.

Il est également important de noter que le choix des mots ne se rattache pas seulement à la formulation des items. En effet, les études de Tourangeau et Rasinski (1988), et celle de Wänke (2002), ont montré que le choix des mots peut influencer les participants à travers le texte d'introduction présenté aux participants. De la même façon,

les réponses des participants peuvent également être influencées par d'autres items du questionnaire en raison de leur contiguïté et des mots qui les composent (J.-G. Blais, communication personnelle, 18 octobre, 2004). Ainsi, l'ordre de présentation des items, de même que les caractéristiques du questionnaire sont des facteurs qui peuvent influencer les réponses des sujets de façon similaire à celle que peut produire la formulation des items.

Par conséquent, tous ces facteurs d'influence doivent être étudiés afin d'éclairer et de guider l'exploration que cette recherche se propose de faire. Cependant, puisque le but de ce travail est d'étudier l'influence de la formulation des items sur les réponses des participants, l'analyse qui sera effectuée portera seulement sur l'item et sur les catégories de réponse de l'échelle qui lui est associée.

Or, l'analyse des items et des catégories de réponse de l'échelle proposée dans une enquête par questionnaire peut être effectuée à l'aide de plusieurs modèles de mesure. Parmi ces modèles, les modèles de la « famille des modèles de Rasch » sont particulièrement intéressants. En effet, les modèles de la famille des modèles de Rasch font partie d'une grande famille de modèles de mesure abondamment étudiés ces dernières années. Certains modèles de cette famille de modèles de mesure ont été proposés spécifiquement pour des données obtenues à l'aide d'échelles de réponses de type Likert. Lorsqu'ils sont adéquats, ces modèles permettraient aux chercheurs d'obtenir un résultat invariant, c'est-à-dire un résultat qui est indépendant du groupe de sujets ciblés par l'opération de mesure ou du groupe d'items inclus (Bertrand & Blais, 2004). Par conséquent, si ces modèles sont adéquats pour les données recueillies dans cette étude, il sera possible d'examiner dans quelle mesure la formulation des items aura eu une influence sur les réponses des participants.

De plus, les modèles de la famille des modèles de Rasch permettent aux chercheurs d'obtenir une estimation de la valeur des points d'ancrage de chacune des catégories de réponse. Ainsi, les chercheurs peuvent travailler avec la vraie valeur des intervalles de réponses plutôt que de travailler avec les intervalles découlant des étiquettes numériques associées à chacune des catégories de réponse. L'estimation des points d'ancrage de l'échelle de réponses constitue un avantage intéressant dans cette recherche, car il permettra d'examiner dans quelle mesure la formulation des items aura eu une influence sur l'échelle de mesure associée aux catégories de réponse.

Parmi tous les modèles appartenant à la famille des modèles de Rasch, deux modèles conviendraient mieux à cette étude. En effet, lorsque les échelles de réponses contenues dans les questionnaires utilisés sont composées d'au moins trois catégories de réponse, ce sont habituellement les modèles de mesure *Rating Scale* (RS) et *Partial Credit* (PC) qui sont mis à contribution dans la modélisation des données.

Le modèle *Rating Scale* s'applique lorsque l'intervalle entre chaque catégorie de réponse demeure la même pour tous les items du questionnaire et le modèle *Partial Credit* s'applique lorsque l'intervalle peut varier d'un item à l'autre. Or, les questionnaires d'enquête de cette recherche utilisent une seule échelle de réponses et l'interprétation de cette échelle devrait être la même pour tous les items du questionnaire. (Tous les détails concernant la méthodologie utilisée dans cette recherche seront présentés au chapitre 4.)

Ainsi, c'est le modèle *Rating Scale* (modèle de mesure pour échelles en catégories ordonnées)<sup>1</sup> qui devra être utilisé pour l'analyse de données de cette recherche. Les deux modèles seront toutefois présentés dans le prochain chapitre, car il sera alors plus facile

---

<sup>1</sup> Bien qu'il soit possible d'exprimer le nom du modèle utilisé dans cette recherche en français, le nom « *Rating Scale* » sera conservé afin de ne pas allonger le texte inutilement.

de comprendre le choix du modèle et comment il pourra être mis à contribution pour l'analyse des données.

En résumé, le but du présent travail est d'explorer l'influence de la formulation des items afin de mieux en comprendre la portée. Pour atteindre cet objectif, deux aspects sont à étudier. Tout d'abord, les réponses des participants devront être examinées afin de vérifier dans quelle mesure la formulation des items peut les influencer. Dans un deuxième temps, ce sont les points d'ancrage de l'échelle de mesure associée aux catégories de réponse qui devront être examinées afin de vérifier dans quelle mesure la formulation des items peut les influencer. Enfin, l'examen des réponses des participants et des points d'ancrage de l'échelle de réponses devra être effectué à l'aide du modèle *Rating Scale* qui, lorsqu'il est adéquat, permet d'assurer un résultat invariant. À la lumière de cette conclusion, il devient possible de formuler l'objectif et les questions de cette recherche. Ces éléments sont présentés dans le prochain chapitre.

## **2. OBJECTIFS ET QUESTIONS DE RECHERCHE**

Le but du présent travail est d'explorer l'influence de la formulation des items sur les réponses des participants et sur l'échelle de mesure associée. Le chapitre précédent a permis de présenter plusieurs éléments qui nous permettent maintenant de formuler clairement les objectifs et les questions auxquelles cette recherche souhaite répondre.

### ***2.1 Objectif général***

L'objectif général de cette recherche consiste à étudier jusqu'à quel point des changements de mots apportés aux items d'un questionnaire d'enquête peuvent influencer la modélisation des réponses fournies par les répondants et ainsi modifier l'échelle de mesure associée aux catégories de réponse.

### ***2.2 Objectif spécifique***

Plus précisément, l'objectif de la recherche consiste à étudier jusqu'à quel point des changements de mots apportés aux items d'un questionnaire sur l'opinion, sans changer le sens général de l'item, peuvent être assez importants pour modifier le construit que représenterait l'item et donc modifier le sens des données qu'il permet de recueillir. De plus, l'objectif consiste à mettre à contribution l'utilisation du modèle de mesure *Rating Scale* dans la modélisation des données.



### **2.3 Questions de recherche**

Au regard de cet objectif, il est maintenant possible de formuler nos deux questions de recherche, c'est-à-dire :

- Dans quelle mesure, un changement de mots dans la formulation des items d'un questionnaire sur l'opinion (sans changer le sens général de la question) a-t-il une influence sur les réponses fournies par les répondants?
- Dans quelle mesure, ce changement de mots a-t-il une influence sur les points d'ancrage de l'échelle de mesure associée aux catégories de réponse?

### 3. CADRE CONCEPTUEL

Pour bien circonscrire un objet d'étude, il est essentiel d'établir les termes et les concepts qui le caractérisent. La première partie de ce chapitre précisera donc le choix terminologique qui sera utilisé pour la suite de ce travail.

Ensuite, afin de mieux comprendre comment les caractéristiques d'un questionnaire et de ses items peuvent influencer les réponses des participants, le processus de réponse impliqué dans une enquête par questionnaire sera examiné. Puis, chacune des caractéristiques du questionnaire et des items présentés précédemment sera étudiée de nouveau. En effet, toutes ces caractéristiques constituent des facteurs pouvant influencer les réponses des sujets interrogés lors d'une enquête par questionnaire. Ainsi, leur étude peut apporter des pistes intéressantes permettant d'éclairer et de guider l'exploration que cette recherche se propose de faire.

Enfin, une introduction sur les modèles de mesure sera présentée afin de mieux comprendre quelles sont les propriétés qui les rendent utiles et quelles conditions ont orienté le choix du modèle pour l'analyse des données de cette recherche. Or, le modèle choisi appartient aux modèles de la famille des modèles de Rasch. Ainsi, la famille des modèles de Rasch sera présentée, de même que les modèles *Rating Scale* et *Partial Credit*. Il sera alors plus facile de comprendre le choix du modèle et comment il pourra être mis à contribution pour l'analyse des données. Finalement, ce sont les statistiques d'ajustement de ces modèles qui seront exposées, car les propriétés de ces modèles (comme l'invariance) ne sont obtenues que lorsqu'il y a adéquation avec les données.

### **3.1 Choix terminologique**

L'objectif de la présente recherche consiste à étudier l'influence de la formulation des items d'une enquête par questionnaire sur les réponses fournies par les répondants et sur l'échelle de mesure associée. Ainsi, il est important de définir en quoi consiste une enquête, un questionnaire et un item, et d'établir pourquoi ces termes ont été choisis par rapport aux autres choix possibles.

Tout d'abord, De Landsheere (1979) définit une enquête comme une démarche de recherche qui tend délibérément à recueillir des données. L'enquête est une méthode de recherche descriptive qui vise à obtenir une information de nature privée en interrogeant des personnes (Lamoureux, 2000). Selon le Grand dictionnaire terminologique (GDT) de l'Office québécois de la langue française (OQLF) de 2005, l'enquête est une étude analytique de tous les renseignements qui vise la connaissance objective d'un phénomène. Bref, c'est la collecte d'informations qui est au cœur de l'enquête.

Pour sa part, le sondage est une démarche qui consiste « à interroger un certain nombre d'individus au sein d'une population déterminée afin d'étudier une ou des caractéristiques de cette population en utilisant des méthodes statistiques adaptées » (GDT, 2005). Par conséquent, c'est l'interrogation d'échantillons représentatifs d'une population à propos d'un sujet donné qui est au cœur du sondage (De Landsheere, 1979). L'objectif d'un sondage est donc de pouvoir faire des inférences et de généraliser les résultats à la population entière.

À la lumière de ces définitions, c'est l'expression « enquête » qui sera retenue pour caractériser la méthode de cueillette de données utilisée dans cette recherche. En

effet, le centre d'intérêt de cette recherche est la collecte de données et non pas les inférences et la généralisation des résultats.

Ensuite, selon Legendre (1988), un questionnaire se définit comme « tout instrument servant à la collecte de données et sur lequel une personne inscrit ses réponses à un ensemble de questions ». Or, une enquête par questionnaire peut prendre plusieurs formes. Le questionnaire peut se dérouler sous forme d'entrevue (en face-à-face ou par téléphone) ou être auto-administré (par la poste ou distribué de main-à-main). S'il est auto-administré, le sujet inscrit alors lui-même ses réponses. De plus, le questionnaire peut avoir différents buts : recueillir des données sur la fréquence d'apparition d'un comportement, sur l'attitude ou sur l'opinion des gens. Enfin, la collecte des données effectuée à l'aide d'une enquête par questionnaire peut avoir lieu une seule fois ou bien à plusieurs reprises pour observer une évolution par rapport à certains éléments (A. Blais & Durand, 1997).

Par conséquent, les termes « enquête par questionnaire auto-administré » seront utilisés pour déterminer l'instrument de cueillette d'informations sur lequel porte cette recherche.

Les enquêtes par questionnaire auto-administré ont donc pour objectif d'obtenir des réponses de la part des répondants (A. Blais & Durand, 1997). Or, pour obtenir des réponses de la part des répondants, il serait naturel de considérer qu'un questionnaire est composé de questions auxquelles les répondants fournissent des réponses. Cependant, lorsqu'on s'intéresse à ce type d'instrument de cueillette de données, plusieurs termes sont utilisés : questions, énoncés et items.

Ainsi, une question est une interrogation adressée à quelqu'un pour obtenir un renseignement, une explication ou pour vérifier des connaissances (Druide, 2005). Pour sa

part, l'énoncé se définit comme l'ensemble des données d'un problème, d'une proposition ou d'une relation mathématique (Larousse, 1988). Enfin, l'item se définit comme l'élément d'un test ou d'un questionnaire d'enquête (Druide, 2005). Selon Legendre (1988), l'item se définit comme un élément d'un instrument de mesure. Conséquemment, c'est donc l'expression « item » qui sera utilisée dans cette recherche pour identifier les éléments composants une enquête par questionnaire auto-administré.

### ***3.2 Facteurs d'influence sur les réponses recueillies à l'aide d'une enquête par questionnaire auto-administré***

Le chapitre précédent a permis de montrer que les réponses recueillies à l'aide d'une enquête par questionnaire pouvaient être influencées par les caractéristiques du questionnaire et de ses items. Or, pour mieux comprendre de quelle façon ces caractéristiques peuvent influencer les réponses des participants, il faut d'abord se pencher sur le processus de réponse impliqué dans une enquête par questionnaire auto-administré.

#### **3.2.1 Processus de réponse**

Une enquête par questionnaire auto-administré implique tout un processus de réponse de la part des sujets interrogés. En effet, les participants doivent suivre une série d'opérations avant de répondre aux items du questionnaire qui leur est proposé. Selon Schwarz (1995), la première étape de ce processus est l'interprétation de l'item et la

compréhension de ce qu'il signifie. Ensuite, le répondant doit trouver la réponse demandée, soit en estimant le nombre de fois où un certain comportement se manifeste, soit en se faisant une opinion par rapport à une situation précise ou bien en évaluant son attitude par rapport à un événement particulier. Finalement, il doit ajuster sa réponse selon les catégories de réponse que le questionnaire lui propose.

La première étape du processus de réponse consiste donc, pour le répondant, à comprendre l'item auquel il tente de répondre. Or, pour mieux saisir ce que cette étape implique, il faut tenter de déterminer ce que « comprendre » exige du répondant. En effet, est-ce que « comprendre » signifie que le répondant doit être capable de reconnaître les caractères ou les mots de la langue utilisée dans le questionnaire ou bien est-ce que « comprendre » exige plus que cela du répondant?

Selon Clark et Schober (1992), la compréhension n'est pas du tout une simple question de mots. Au contraire, ils soutiennent que la compréhension est essentiellement liée à l'intention du chercheur qui a rédigé l'item. L'exemple qui suit illustre ce propos. Supposons qu'un des items d'un questionnaire demande aux répondants d'énumérer la liste des gens avec qui ils communiquent habituellement par courriel. Une personne répondant à ce questionnaire n'aura probablement aucune difficulté à comprendre les mots de cette demande. Cependant, afin de fournir une réponse appropriée, la personne cherchera à savoir si elle doit fournir la liste des amis avec qui elle communique habituellement, la liste de ces collègues de travail, la liste des membres de sa famille ou la liste de toutes ces personnes regroupées.

Dans cet exemple, la personne interrogée ne cherche pas seulement à fournir une réponse à l'item qui lui est proposé, mais vise à coopérer avec le chercheur afin de lui fournir les éléments d'informations dont il a besoin pour atteindre les objectifs de sa recherche. En effet, Grice (1975) soutient qu'une fois engagés dans une conversation (ou

dans un processus d'échange d'informations), les deux interlocuteurs considèrent comme acquis que chacun doit contribuer à atteindre le but de cet échange. C'est le principe de coopération qu'il a formulé après avoir étudié le déroulement d'une conversation entre deux individus.

Ainsi, selon Grice, en acceptant de répondre à l'enquête, les répondants doivent ensuite coopérer avec le chercheur afin de l'aider à atteindre ses objectifs. Or, pour contribuer de façon appropriée à l'enquête, les participants doivent avoir une idée de la perspective dans laquelle l'item a été formulé, c'est-à-dire quelle était l'intention du chercheur en rédigeant son questionnaire. Dans l'exemple décrit ci-dessus, la personne interrogée n'arrive pas à déterminer la perspective qu'elle doit donner à l'item et donc ne sait pas comment contribuer.

La contribution de chacun des interlocuteurs est habituellement déterminée à travers le processus d'échange d'informations dans lequel ils se sont engagés. En effet, Clark et Schober (1992) soutiennent que les interlocuteurs en apprennent davantage sur la perspective qu'ils doivent donner à la conversation à travers chacune de leur contribution (principe de l'accumulation). Or, dans une enquête par questionnaire auto-administré, le chercheur n'est généralement pas présent au moment où la personne interrogée complète le questionnaire. Par conséquent, plusieurs études ont montré que lorsqu'un répondant n'arrive pas à comprendre un des items du questionnaire, c'est à travers les caractéristiques du questionnaire et de ses items qu'il tentera de déterminer la perspective que le chercheur souhaite donner au processus d'échange d'informations et qu'il en déduira le sens de ce qui lui est demandé (Clark & Schober, 1992; Schuman & Presser, 1996; Schwarz, 1995; Wänke, 2002). Ainsi, c'est à partir de ce qu'elle perçoit des caractéristiques du questionnaire et de ses items qu'une personne coopère au processus d'échange d'informations déclenché par une enquête.

Une fois que la personne interrogée a pu interpréter l'item auquel elle tente de répondre, elle doit ensuite trouver la réponse demandée et la communiquer (Schwarz, 1995). Pour ce faire, Clark et Schober (1992) suggèrent que les participants cherchent généralement à estimer la réponse demandée plutôt que de fournir une réponse précise. En effet, Clark et Schober ont remarqué que les participants considèrent que le chercheur ne s'attend sûrement pas à une réponse précise puisqu'ils doivent fournir une réponse immédiate et dans le court temps alloué à la complétion du questionnaire. Ainsi, selon Clark et Schober, les participants considèrent que le chercheur s'intéresse seulement aux événements dont ils peuvent se souvenir rapidement et qu'ils doivent ensuite fournir la meilleure estimation possible sur le nombre d'occurrences de cet événement.

Enfin, une fois que la personne a pu estimer la réponse demandée, elle doit se référer aux catégories de réponse de l'échelle proposée afin de communiquer sa réponse au chercheur. Divers facteurs peuvent amener la personne à ajuster sa réponse en fonction de l'échelle qui lui est offerte. Ces facteurs seront étudiés dans la section portant sur les échelles de réponses en catégories ordonnées.

Ainsi, le processus de réponse à une enquête par questionnaire auto-administré permet de mieux comprendre de quelle façon les caractéristiques du questionnaire et de ses items peuvent influencer les réponses des personnes interrogées. Or, Schwarz (1995) définit les caractéristiques du questionnaire et de ses items comme étant composées de l'environnement immédiat d'un item, de la mise en forme du questionnaire, ainsi que des catégories de réponse proposées aux répondants. Selon Schuman et Presser (1996), les répondants interagissent avec les items du questionnaire, c'est-à-dire qu'ils sont influencés par les mots qui les composent de même que par leur ordre de présentation. Cependant, ils précisent que l'influence la plus grande sur les réponses des participants proviendrait de la formulation des items et de la formulation des choix de réponses qui leur sont présentés.



Par conséquent, les caractéristiques d'un questionnaire d'enquête auto-administré, l'ordre de présentation des items, les choix de réponses présentés, de même que la formulation des items constituent des facteurs d'influence importants. La section qui suit explorera donc de quelle façon les caractéristiques d'un questionnaire peuvent influencer les réponses des sujets interrogés lors d'une enquête par questionnaire.

### **3.2.2 Caractéristiques du questionnaire**

La présente section porte principalement sur les caractéristiques d'un questionnaire telles que la mise en forme et la présentation. Les autres caractéristiques (ordre des items, catégories de réponse, formulation des items) seront étudiées dans les prochaines sections de ce chapitre.

Parmi toutes les études recensées dans le cadre de ce travail, seulement quelques unes ont signalé l'influence des caractéristiques d'un questionnaire comme la mise en forme ou la présentation. En effet, la plupart des chercheurs ont plutôt associé l'influence constatée dans leur recherche au choix des mots, à l'ordre des items ou aux catégories de réponse de l'échelle proposée. Par conséquent, seulement quelques exemples ont été retenus pour la présente section. Les exemples choisis permettront de montrer que, bien que certains chercheurs signalent l'influence potentielle des caractéristiques d'un questionnaire, les résultats de ces études n'apportent pas de nouvelles pistes pouvant éclairer l'exploration visée par ce travail.

Ainsi, selon Tourangeau et Rasinski (1988), les facteurs influençant les réponses des sujets interrogés lors d'une enquête par questionnaire peuvent provenir de l'environnement immédiat d'un item, mais ils peuvent également provenir d'autres

caractéristiques du questionnaire comme le texte d'introduction présenté aux participants au début du questionnaire. Cependant, ils estiment également que tous les éléments du questionnaire précèdent l'item auquel un sujet interrogé tente de répondre peuvent avoir une influence sur sa réponse.

Autrement dit, Tourangeau et Rasinski (1988), considèrent que tous les éléments précédant l'item auquel un sujet interrogé tente de répondre forment un cadre interprétatif qui lui permet de déduire la perspective qu'il doit donner à sa réponse. De plus, Tourangeau et Rasinski suggèrent que ce cadre interprétatif pourrait influencer le point de vue des participants sur le contenu des prochaines questions.

Par conséquent, l'étude de Tourangeau et Rasinski abonde dans le même sens que celle de Clark et Schober (1992) et montre que les caractéristiques du questionnaire et de ses items permettent aux sujets interrogés de déterminer l'intention du chercheur et d'en déduire ce qui leur est demandé. Cette étude vient donc soutenir la description du processus de réponse à une enquête par questionnaire auto-administré présenté à la section précédente, mais n'apporte aucun élément nouveau permettant d'expliquer de quelle façon les caractéristiques d'un questionnaire peuvent influencer les réponses des participants.

De la même façon, Wänke (2002) a pu constater que les participants interrogés dans son étude répondaient de façon différente selon le sujet de recherche annoncé dans le texte d'introduction du questionnaire. Parallèlement, les participants interrogés fournissaient des réponses différentes selon la population cible annoncée. Or, encore une fois, cette étude montre que les sujets interrogés dans une enquête cherchent à coopérer avec le chercheur (principe de coopération de Grice (1975) présenté dans la section précédente) et que le texte d'introduction leur permet de déduire l'intention du chercheur et d'orienter leur réponse.

Ainsi, il faudra plutôt étudier l'ordre de présentation des items, les catégories de réponse de l'échelle proposée et la formulation des items afin de voir si ces différents facteurs peuvent fournir d'autres indices permettant de mieux comprendre de quelle façon les caractéristiques du questionnaire et de ses items influencent les réponses des participants.

### **3.2.3 Ordre de présentation des items**

Parmi tous les facteurs étudiés précédemment, l'ordre de présentation des items est un des facteurs d'influence que les chercheurs mentionnent le plus souvent. Le premier chapitre de ce travail a d'ailleurs permis de souligner que, selon Schuman et Presser (1996), l'ordre des items semble être un des secteurs les plus importants de la recherche méthodologique.

Pour mieux comprendre de quelle façon l'ordre des items peut influencer les réponses des participants, Schuman et Presser (1996) ont pu diviser les divers effets que ce facteur pourrait produire selon différentes catégories. Parmi les catégories qu'ils ont pu identifier, les deux chercheurs présentent entre autres l'effet « cohérence partie-partie » qui surviendrait lorsqu'un répondant, qui fait le rapprochement entre deux items de même degré d'importance, s'assurerait que la réponse fournie à chacun de ces items est cohérente. À l'inverse, l'effet « contraste partie-partie » surviendrait lorsqu'un répondant, qui fait le rapprochement entre deux items de même degré d'importance, fournirait des réponses contraires à ces deux items.

Schuman et Presser présentent aussi l'effet « cohérence partie-tout » qui surviendrait lorsque la réponse fournie par un répondant à un item précis, induirait la réponse qu'il doit fournir à un item plus général traitant du même sujet. À l'inverse, l'effet « contraste partie-tout » surviendrait lorsqu'un répondant fournirait des réponses contraires à deux items traitant d'un même sujet, mais où l'un des items est précis, et l'autre plus général.

Enfin, Schuman et Presser ont pu identifier d'autres effets que l'ordre de présentation pourrait produire sur les réponses des participants comme la fatigue de l'individu, la prédominance de certains items par rapport à d'autres (effet « *Saliency* »), etc. Les auteurs précisent toutefois que toutes les catégories présentées ne sont que des effets possibles pouvant être produits par l'ordre de présentation des items et que l'intensité de ces effets pourrait être variable.

Tourangeau et Rasinski (1988) ont également remarqué que les items qui précèdent l'item auquel un sujet interrogé tente de répondre pouvait parfois influencer le sujet et l'amener à répondre de façon cohérente aux items subséquents. Clark et Schober (1992) supportent cette idée. Toutefois, Tourangeau et Rasinski ont aussi constaté que les items qui précèdent pouvaient parfois amener le sujet à répondre de façon incohérente aux items subséquents.

Il apparaît donc que les résultats obtenus par différents chercheurs sur l'influence de l'ordre des items ne sont pas convergents (Maheux, 2001; Schuman & Presser, 1996). Cependant, Maheux (2001) et Blais (communication personnelle, 18 octobre, 2004) pensent que l'effet « contraste partie-tout » serait à étudier davantage. En effet, ils ont remarqué que la question de l'ordre des items pourrait être importante dans certains cas particuliers où la position relative d'un item sur l'échelle des niveaux de difficulté pourrait être influencée par sa contiguïté avec d'autres items en raison des mots qui le composent,

mais surtout en raison de ses liens avec l'item qui le précède. Or, selon Clark et Schober (1992), les participants considèrent que deux items qui se suivent sont nécessairement liés au même sujet, à moins qu'une précision leur indique que ce n'est pas le cas.

Enfin, selon Tourangeau et Rasinski (1988), les items précédents celui auquel un sujet interrogé tente de répondre aideraient le répondant à déterminer quelles informations il serait pertinent de fournir et lesquelles seraient redondantes.

Par conséquent, l'étude de l'influence de l'ordre des items permet de confirmer encore une fois que les caractéristiques du questionnaire et de ses items fournissent un cadre de référence aux sujets interrogés dans une enquête par questionnaire. Or, les sections précédentes avaient montré que ce cadre de référence permettait aux sujets de déduire les intentions du chercheur et les amenait à coopérer avec le chercheur afin de l'aider à atteindre ces objectifs de recherche.

Les études qui viennent d'être présentées apportent de nouvelles pistes d'informations qui pourront peut-être guider l'exploration visée dans ce travail. En effet, ces dernières suggèrent que le cadre de référence fourni par les items qui précèdent l'item auquel un sujet tente de répondre, permettrait aux sujets interrogés de déterminer les informations qu'il serait pertinent de fournir aux items subséquents et celles qui seraient redondantes. De plus, ce cadre de référence pourrait amener certains sujets interrogés à rester cohérent et d'autres, à répondre de façon incohérente.

Enfin, l'étude de l'influence de l'ordre des items suggère qu'il faut rester attentif à l'une des catégories d'effet possible, l'effet « contraste partie-tout », car il s'agit d'un facteur d'influence potentiel qui pourrait être présent dans l'analyse des données de cette recherche.

### 3.2.4 Échelles de réponses en catégories ordonnées

De façon générale, les chercheurs qui utilisent les enquêtes par questionnaire, vont proposer un choix de réponses aux sujets interrogés et c'est à travers ce choix que les participants doivent communiquer leur réponse. Le choix de réponses proposé est habituellement une échelle de type Likert pour laquelle chaque catégorie de réponse est formée d'une étiquette et d'un code numérique (chap. 1, p. 4).

Parmi les facteurs d'influence liée aux caractéristiques du questionnaire et de ses items, les échelles de réponses en catégories ordonnées sont parmi les sources d'influence les plus étudiées. En effet, les chercheurs ont constaté que cette source d'influence pouvait produire plusieurs effets.

Tout d'abord, les catégories de réponse de l'échelle proposée déterminent le domaine de réponses que les sujets interrogés doivent utiliser pour répondre au questionnaire (Clark & Schober, 1992). Or, le domaine proposé peut très bien ne pas correspondre à la réponse que les sujets auraient fourni spontanément. Par exemple, le fruit préféré d'un sujet pourrait très bien être l'ananas alors que l'échelle propose une série de fruits des champs comme les fraises, les framboises et les bleuets.

Par ailleurs, le domaine proposé pourrait très bien ne pas correspondre à la réponse que les sujets auraient fournie spontanément parce que les catégories de réponse sont en rupture sémantique par rapport à l'item proposé (Lam & Kolic, 2006). Par exemple, l'item proposé pourrait être « Je participe en classe » et les catégories de réponse offertes constituées de termes tels que « Faible », « Moyen », « Satisfaisant », etc.

Ainsi, contrairement à une question ouverte où le répondant est libre de répondre comme il le souhaite (Jones, 2000), un item qui propose une échelle de réponses force les participants à ajuster leur réponse en fonction des catégories offertes (Clark & Schober, 1992). Dans le premier exemple présenté ici, le sujet devra ajuster sa réponse à la perspective du chercheur et choisir le fruit des champs qu'il préfère. Dans le deuxième exemple, Lam et Kolic (2006) suggèrent que les répondants seraient amenés à ignorer les étiquettes des catégories proposée. Or, dans les deux cas, l'échelle de réponses contraint les répondants à s'ajuster aux catégories de réponse offertes.

Par ailleurs, Schwarz (1995) a pu montrer que les répondants considèrent comme acquis que les catégories de réponse de l'échelle offerte sont construites de façon à refléter les connaissances du chercheur sur le sujet abordé dans le questionnaire et constituent une norme par rapport à ce qui leur est demandé. En effet, pour les répondants, la catégorie intermédiaire de l'échelle de réponses représenterait la moyenne alors que la première et la dernière catégorie seraient associées à des comportements extrêmes. Ainsi, les répondants compareraient leur réponse à ce qu'ils estiment être une représentation typique des réponses obtenues dans la population et ajusteraient leur réponse en conséquence.

L'étude de Schaeffer (1991) apporte toutefois un point de vue différent. En effet, sa recherche a montré que les répondants utiliseraient plutôt des normes et des références liées principalement à un réseau social de personnes semblables afin d'évaluer la fréquence de leurs comportements. Ainsi, des caractéristiques telles que l'âge, l'éducation ou la race joueraient un rôle dans l'interprétation que les répondants font de l'échelle de réponses qui leur est fournie.

Par conséquent, l'échelle de réponses offerte fournit un cadre de référence aux répondants, soit en déterminant le domaine des réponses possibles, soit en établissant la

distribution des réponses attendues. Or, Schwarz (1995) a établi que les codes numériques associés à chacune des catégories de l'échelle jouaient également un rôle dans l'interprétation que font les répondants de l'échelle qui leur est proposée.

En effet, Schwarz (1995) a demandé à différents sujets à quel point leur vie était réussie et il leur a offert une échelle de réponses allant de « pas réussie du tout » à « extrêmement bien réussie ». Les résultats de son étude ont montré que lorsque l'échelle de réponses était codée de 0 à 10, les répondants interprétaient la première catégorie (0 = pas réussie du tout) comme traduisant l'absence de succès notable. Cependant, lorsque les réponses étaient codées de -5 à 5, les sujets interprétaient la première catégorie (-5 = pas réussie du tout) comme exprimant la présence d'échecs majeurs. Ainsi, les étiquettes numériques des catégories de réponse de l'échelle offrirait pondérerait la valeur que les répondants accordent à chacune des catégories.

Bradburn et Sudman (1979) ont remarqué que les adverbes utilisés pour identifier chacune des catégories, produisaient un effet similaire. En effet, les résultats de leur étude ont démontré, par exemple, que l'adverbe « pas trop » appliqué au marqueur « souvent » avait un effet réducteur sur l'intervalle qui sépare ces deux catégories. Ainsi, lorsque les répondants sont enthousiastes par rapport à une situation donnée, l'intervalle entre « pas trop souvent » et « souvent » est réduit à environ 0,51 par rapport à l'intervalle de 1 que les codes numériques apposés sur chacune des catégories suggèrent. En revanche, l'intervalle est réduit à environ 0,3 lorsque la situation amène un sentiment d'ennui chez les sujets. À l'inverse, l'adverbe « très » avait un effet multiplicateur sur l'intervalle qui sépare les catégories « souvent » et « très souvent ». La grandeur de l'intervalle augmenterait donc à environ 1,37 lorsque les participants sont enthousiastes par rapport à une situation et à environ 1,27 pour une situation qui leur procure un sentiment d'ennui.



Par conséquent, les échelles de réponses offertes dans une enquête par questionnaire fournissent différents cadres de référence aux répondants. En effet, les échelles fournies peuvent déterminer le domaine des réponses possibles ou établir la distribution des réponses attendues. Les étiquettes numériques associées à chacune des catégories peuvent modifier l'interprétation que font les répondants de l'échelle de réponses en pondérant la valeur qu'ils accordent à chacune des catégories. Enfin, les adverbes utilisés pour identifier chacune des catégories auraient également une influence sur l'interprétation de l'échelle en déplaçant les points d'ancrage de chacune des catégories. L'analyse des échelles de réponses en catégories ordonnées offre donc des pistes intéressantes qui pourront certainement être utiles pour la suite de ce travail.

### **3.2.5 Formulation des items**

Les études présentées jusqu'ici laissent entrevoir que le choix des mots utilisés dans une enquête par questionnaire pourrait avoir une influence sur les réponses des participants et ce, de diverses façons. Par exemple, la section 3.2.3 suggérait que les réponses des participants pouvaient être influencées par d'autres items du questionnaire en raison de leur contiguïté sur l'échelle des niveaux de difficulté et des mots qui les composent. Ensuite, la section 3.2.4 a montré que le choix des mots peut également influencer les réponses des participants à travers les catégories de réponse de l'échelle proposée (exemple de Sudman et Bradburn, 1979). L'exemple de Rugg (1941), présenté dans le premier chapitre, confirme que le choix des mots utilisés dans la formulation des items influence les réponses des sujets interrogés. Enfin, la section 3.2.2 a montré que le contenu du texte d'introduction avait une influence sur la façon de répondre des participants. Ainsi, il est possible de se demander si le choix des mots utilisés pour présenter

le contenu ou le sujet pourrait induire une influence similaire à celle qui a été obtenue en présentant des contenus ou des sujets différents.

En effet, selon Clark et Schober (1992), les répondants considèrent que le chercheur choisit les mots utilisés dans son questionnaire de façon intentionnelle et que le sens des mots choisis est conforme à celui habituellement utilisé dans leur culture. De plus, Clark et Schober soutiennent que lorsque des termes chargés de sens comme devoir, pouvoir, anti-, pro-, etc., étaient utilisés, ces termes établissaient une perspective particulière qui devait être utilisée pour répondre à l'item. Par conséquent, le choix des mots utilisé dans un questionnaire est important, car il permet aux répondants de déterminer la perspective que le chercheur souhaite donner au processus d'échange d'information engendré par l'enquête. Ce résultat est cohérent avec celui trouvé lors de l'analyse du processus de réponse de la section 3.2.1.

Or, l'influence de la formulation des items n'est pas seulement liée au sens des mots utilisés. En effet, certains chercheurs ont constaté que la formulation choisie pouvait influencer les réponses des sujets interrogés en rendant certaines données plus ou moins saillantes. C'est ce que, entre autres, Kahneman et Tversky (1982) ont remarqué.

En effet, l'étude de Kahneman et Tversky (1982) a montré que, lorsque les données relatives aux lois de la probabilité étaient relativement saillantes, ces données étaient utilisées par des adultes pour résoudre le problème qui leur était proposé (inférer le domaine de spécialisation d'un étudiant). Cependant, si les données relatives aux lois de la probabilité étaient rendues moins saillantes par l'ajout d'informations stéréotypées sur le domaine de spécialisation le moins populaire, les répondants semblaient alors ignorer la loi des probabilités en faveur du domaine de spécialisation le plus rare (réponse incorrecte). Un effet similaire, quoique plus faible, était obtenu lorsque la formulation utilisée ajoutait

des informations neutres (non stéréotypées), mais non pertinentes, sur l'un ou l'autre des domaines de spécialisation.

Ng, Pipe, Beath et Holton (1999) ont obtenu des résultats similaires. En effet, Ng et al. se sont intéressés à l'effet possible du choix des mots sur les inférences statistiques des enfants. Ils ont pu constater que la formulation des énoncés proposés aux enfants avait une influence plus grande sur leur réponse que les données fournies. Pour leur étude, ils ont demandé aux enfants de réaliser une série de raisonnements inductifs reliés à deux concepts : les probabilités et la loi des grands nombres. Deux aspects de la loi des grands nombres ont été étudiés. Le premier aspect est que la moyenne d'un grand échantillon a moins de chance que la moyenne d'un petit échantillon de s'éloigner de la moyenne de la population. Cette relation entre la taille d'un échantillon et l'écart d'un échantillon par rapport à la moyenne est souvent ignorée par les adultes lorsqu'on leur présente une tâche liée à cette relation (étude de Kahneman et Tversky, 1982). Le deuxième aspect est qu'une prédiction est plus fiable lorsqu'elle est inférée à partir d'un grand échantillon plutôt qu'à partir d'un petit échantillon.

Afin de faire ressortir si la formulation des énoncés avait un impact sur les raisonnements des enfants relativement à ces deux concepts (les probabilités et la loi des grands nombres), les chercheurs ont préparé quatre paires de questions (deux portant sur les taux de base, une sur l'écart d'un échantillon par rapport à la moyenne et une sur la fiabilité d'un échantillon). Pour chacune de ces paires, les énoncés étaient formulés différemment de façon à rendre les données plus ou moins saillantes. Les résultats de cette recherche ont montré que le choix des mots avait un impact important sur les réponses. En effet, les résultats des enfants étaient meilleurs lorsque les données étaient saillantes que lorsque des informations additionnelles étaient ajoutées pour rendre les données moins saillantes.

Par conséquent, ces études confirment que la formulation des items peut avoir une grande influence sur les réponses des participants. Or, l'influence peut provenir de deux sources différentes. Elle peut provenir du sens des mots utilisés pour formuler l'item ou de la formulation elle-même, en rendant certaines informations plus ou moins saillantes.

Enfin, la méthode utilisée par Ng et al. (1999) pour vérifier si la formulation des énoncés avait une influence sur les réponses des enfants est comparable à celle que cette recherche se propose d'utiliser. Leur étude permet donc de soutenir la méthode d'analyse choisie dans ce travail. Afin de mieux comprendre cette méthode, les prochaines sections présenteront le modèle de mesure qui sera mis à contribution dans cette recherche.

### ***3.3 Modèles de mesure***

Les sections précédentes ont permis d'étudier chacune des caractéristiques d'un questionnaire et de ses items et cette étude a permis d'apporter des pistes intéressantes pouvant éclairer l'exploration que cette recherche se propose de faire. Cependant, pour atteindre les objectifs de cette recherche, il est également important de comprendre le choix du modèle de mesure qui sera utilisé pour l'analyse des données.

Avant de présenter le modèle utilisé, il est intéressant de déterminer en quoi consiste un modèle de mesure. La présente section introduira donc ce qu'est un modèle de mesure, quelles sont les propriétés qui les rendent utiles et quelles sont les conditions qui ont orienté le choix du modèle pour l'analyse des données de cette recherche.

Par définition, « un modèle de mesure consiste en un plan formé d'une série de règles à suivre afin de construire un instrument de mesure possédant des propriétés

comme la précision et la validité » (Bertrand & Blais, 2004, p. 30). Ainsi, selon les auteurs, un modèle de mesure peut se définir comme une démarche d'analyse rigoureuse qui consiste à modéliser les données recueillies (à l'aide d'une enquête par questionnaire par exemple), puis à comparer les résultats produits par le modèle à des critères de référence permettant de s'assurer que les conditions de base du modèle sont vérifiées. Cette démarche d'analyse permet alors au chercheur d'obtenir de son outil de cueillette de données des résultats auxquels il peut se fier, et par suite, un instrument de mesure précis et valide (le questionnaire).

Un instrument de mesure précis est donc un instrument dont les données obtenues sont réellement représentatives de l'opinion des participants, de leur attitude, de leur comportement, etc. (Lamoureux, 2000). Si, de surcroît, l'instrument est stable, c'est-à-dire que les données qu'il permet d'amasser sont identiques pour les mêmes participants interrogés à des moments différents, cela indique au chercheur la fidélité de son outil. Cette qualité est souvent considérée par les chercheurs car elle indique jusqu'à quel point l'instrument mesure toujours le même phénomène d'une fois à l'autre (Lamoureux, 2000).

Un instrument de mesure valide est un instrument qui mesure avec justesse ce dont il est censé rendre compte. Ainsi, un instrument valide « doit refléter avant tout la caractéristique qu'il est censé mesurer » (GDT, 2005). Lamoureux (2000) précise que la validité d'un instrument concerne la nature même de l'objet étudié. En effet, la validité exprime jusqu'à quel point l'instrument produit des données qui traduisent la réalité.

Diverses stratégies d'analyse peuvent être mises en œuvre pour la modélisation des données. Par exemple, pour plusieurs chercheurs, un des objectifs de l'utilisation des enquêtes par questionnaire est d'obtenir un résultat total. Ainsi, très souvent, les chercheurs vont faire la somme des codes numériques associés à chacune des catégories de réponse de l'échelle offerte dans le questionnaire. Cela leur permet de résumer, pour

chacun des items ou pour chacun des individus, les réponses des sujets interrogés. Une telle façon d'analyser les données est une des stratégies auxquelles les chercheurs ont fréquemment recours lorsqu'ils utilisent la modélisation de la théorie classique des tests (TCT).

Or, le score global obtenu en faisant la somme des codes numériques associés à chacune des réponses du questionnaire ne permet pas de comparer les résultats facilement. En effet, selon Hambleton, Swaminathan et Rogers (1991), un item peut sembler facile ou difficile selon le niveau d'habileté des sujets interrogés. De même, le niveau d'habileté des participants peut sembler faible ou élevé selon que les items du questionnaire qui leur a été distribué étaient faciles ou difficiles.

Le niveau de difficulté des items modélisé à l'aide de la TCT dépend donc du groupes de sujets qui ont servi à la calibration du test (J.-G. Blais & Ajar, 1992). De même, l'estimation de l'habileté des participants dépend des caractéristiques des items qui ont été utilisées dans le questionnaire. Le score global obtenu à l'aide de la TCT limite donc les comparaisons entre différentes versions d'un même questionnaire, car les résultats obtenus peuvent varier d'un groupe de participants à l'autre. La comparaison se limite à des situations où les conditions entourant le processus de mesure sont les mêmes et où les groupes de candidats possèdent des caractéristiques similaires (J.-G. Blais, 1987).

Une autre critique généralement associée au score global obtenu à l'aide de la TCT est que les échelles utilisées dans les enquêtes par questionnaire ne possèdent habituellement pas la propriété d'addition. Les catégories de réponse des échelles sont généralement ordonnées, mais les intervalles entre chacune des catégories sont inégaux. En effet, l'étude de Bradburn et Sudman (1979) présentée à la section 3.2.4, a montré que les points d'ancrage des catégories de réponse de l'échelle proposée, ne correspondent pas nécessairement aux étiquettes numériques associées par le chercheur à chacune des

catégories. Le score global obtenu en faisant la somme des codes numériques associés à chacune des réponses du questionnaire ne permet donc pas de résumer de façon appropriée les réponses des sujets interrogés.

Les critiques qui viennent d'être présentées permettent de mieux comprendre pourquoi la modélisation de la TCT ne constituerait pas la stratégie la plus utile pour modéliser les données de cette recherche. En effet, le score global obtenu en faisant la somme des codes numériques associés à chacune des réponses du questionnaire ne permet pas de prédire avec succès le résultat d'un participant confronté à un des items d'un questionnaire (J.-G. Blais, 1987). Or, pour être approprié, les scores modélisés devraient permettre au chercheur de faire des prédictions (Thissen & Wainer, 2001).

Selon Wright et Mok (2004), pour qu'un modèle de mesure soit utile, il doit permettre de faire des inférences, d'estimer la précision de ces inférences et fournir des moyens de détecter et d'évaluer les divergences entre les données recueillies et les mesures modélisées. De plus, les paragraphes précédents ont montré que le modèle utilisé devrait permettre d'obtenir des résultats invariants, c'est-à-dire indépendants du contexte à partir duquel les données ont été recueillies.

Une des stratégies d'analyse que les chercheurs peuvent mettre en œuvre et qui répond aux conditions ci-dessus est la modélisation de la théorie de la réponse à l'item (TRI). Selon Hambleton, Swaminathan et Rogers (1991), les modèles de la TRI reposent sur deux postulats : le premier stipule que la performance d'un individu à un item d'un questionnaire peut être prédite sur la base de l'habileté de cette personne. Deuxièmement, la relation entre la performance d'un individu à un item d'un questionnaire et l'habileté de la personne ayant produit cette performance peut être définie à l'aide d'une fonction monotone croissante. Cette fonction est appelée fonction caractéristique de l'item (FCI).

Actuellement, la relation mathématique qui décrit cette fonction est, le plus souvent, une forme de régression logistique.

La FCI se traduit donc par une fonction reliant l'habileté de l'individu à la probabilité de réussir un item. L'objectif de cette fonction est, d'une part, de produire une estimation, de la façon la plus précise possible, de l'habileté des individus à partir de leur réponse aux items; d'autre part, la FCI vise l'évaluation des qualités psychométriques des items (Bertrand & Blais, 2004). La représentation graphique de cette fonction s'appelle la courbe caractéristique de l'item (CCI). Dans la version la plus simple de cette fonction, soit le modèle dichotomique (ce modèle est présenté à la section suivante), la représentation graphique a une forme de « S » où le point d'inflexion correspond à une probabilité de 50 % que le répondant réussisse l'item.

La modélisation faite avec la TRI permet donc d'obtenir une perspective différente de celle offerte par la TCT sur les données recueillies à l'aide d'enquêtes par questionnaire. De plus, le processus stochastique sur lequel se basent les modèles de la TRI aide les chercheurs à transformer les données brutes recueillies à partir des catégories de réponse de l'échelle offerte dans le questionnaire en une échelle dont l'intervalle entre chaque catégorie de réponse est égal (modèle pour des catégories de réponses ordonnées) (Bond & Fox, 2001). L'échelle de réponses ainsi construite possède alors des propriétés plus métriques s'approchant de l'additivité.

Les paragraphes précédents montrent que la modélisation de la TRI constituerait une stratégie utile pour modéliser les données de cette recherche. Parmi les modèles de la TRI, les modèles de la famille des modèles de Rasch sont particulièrement intéressants pour l'analyse des données de cette recherche. En effet, certains modèles de cette famille de modèles de mesure ont été proposés spécifiquement pour des données obtenues à l'aide d'échelles de réponses de type Likert. De plus, lorsqu'ils sont adéquats,



ces modèles permettraient aux chercheurs d'obtenir un résultat invariant. Afin de mieux comprendre pourquoi ces modèles constituent un choix utile pour ce travail, la prochaine section étudiera ces modèles de façon plus détaillée.

Il existe deux courants de pensée au sujet de la famille des modèles de Rasch. En effet, certains chercheurs considèrent que la famille des modèles de Rasch découle de modèles généraux de la TRI tels que les modèles à deux ou à trois paramètres. Pour d'autres, les modèles de la famille des modèles de Rasch ont été développés suivant un paradigme différent de celui de la TRI. Cependant, puisque cette recherche s'intéresse aux modèles et non aux courants de pensée desquels ils ont émergés, les modèles sont exposés ici suivant la présentation des modèles que l'on retrouve le plus souvent dans la littérature.

### **3.3.1 Modèles de mesure de la famille des modèles de Rasch**

La section précédente a permis de faire ressortir que les modèles de la famille des modèles de Rasch sont les modèles qui répondraient le mieux à l'analyse des données de cette recherche. Afin de mieux comprendre pourquoi ces modèles constituent le meilleur choix et comment ils pourront être mis à contribution, la présente section propose une introduction sur ces modèles de mesure. La présentation utilisée pour cette introduction est fortement inspirée du texte de Wright et Mok (2004).

Afin de présenter les modèles de la famille des modèles de Rasch, Wright et Mok (2004) utilisent d'abord un exemple dont les résultats possibles sont dichotomiques. La méthode est ensuite facile à généraliser pour des situations où la gradation de l'échelle est plus précise. L'exemple utilisé est donc le suivant : imaginons des athlètes qui sautent

différentes haies. La force des athlètes peut varier allant de fort à faible. La hauteur des haies varie également afin d'offrir différents niveaux de compétition.

La performance des différents athlètes qui essaient de sauter les diverses haies peut être représentée par une matrice de résultats. Ainsi, pour chacun des athlètes ( $n = 1, \dots, N$ ) qui saute la haie ( $i = 1, \dots, L$ ), le résultat sera soit un succès ( $x_{ni} = 1$ ) ou un échec ( $x_{ni} = 0$ ). Ensuite, il est possible de calculer un score total des sauts réussis ( $R_n = \sum_i x_{ni}$ ). Cependant, puisque les athlètes sont de forces différentes, un tel score brut ne nous permet pas de les comparer. En effet, ils n'ont pas une chance égale de réussir les diverses haies qui leur sont présentées puisque leur chance dépend de leur force.

De la même façon, nous pourrions créer une matrice pour chacune des haies où, pour chaque athlète qui réussit le saut, nous notons un succès, etc. Et, comme nous l'avons fait, nous pourrions également calculer un score total pour chaque haie du nombre d'athlètes ayant réussi le saut. Mais encore une fois, comme les haies n'auront pas été défiées par des athlètes de même calibre, la comparaison du niveau de difficulté relative entre deux haies est impossible.

Par ailleurs, bien que des matrices ainsi créées comportent toutes les informations qui nous intéressent, elles ne nous aideraient pas à faire des prédictions sur les prochains événements. Or, ce que nous voulons vraiment savoir est si l'athlète  $n$  réussira son prochain saut à la haie  $i$ . Afin de connaître la force d'un athlète, il faut le confronter à une haie; de même, pour connaître la hauteur d'une haie, il faut la confronter à un athlète. Ce faisant, nous transformons nos observations en attentes, c'est-à-dire en probabilité de réussite.

Ensuite, la base sur laquelle reposent les modèles de Rasch peut s'établir à l'aide d'une paire de comparaisons similaire à celle utilisée par Wright et Linacre (1987). Ainsi,

considérons que nous voulons comparer la force de deux athlètes ou la hauteur de deux haies. Afin d'établir lequel de ces deux athlètes est le plus fort, nous les faisons sauter la même haie. Le tableau suivant présente les 2 x 2 résultats possibles.

**Tableau 1 : Résultats possibles découlant de la comparaison entre la force de deux athlètes**

	<b>Athlète 2 réussit le saut</b>	<b>Athlète 2 échoue le saut</b>
<b>Athlète 1 réussit le saut</b>	N <sub>11</sub> = le nombre de fois où les deux athlètes réussissent	N <sub>10</sub> = le nombre de fois où l'athlète 1 réussit le saut alors que le deuxième échoue
<b>Athlète 1 échoue le saut</b>	N <sub>01</sub> = le nombre de fois que l'athlète 1 échoue alors que le deuxième réussit	N <sub>00</sub> = le nombre de fois où les deux athlètes échouent

Ainsi, lorsque nous voulons comparer la force des deux athlètes, c'est le ratio  $N_{10}/N_{01}$ , plutôt que la différence entre ces résultats, qui produit la comparaison la plus stable de la force de l'un par rapport à l'autre. Or, ce ratio peut être écrit en termes de probabilités. Pour éviter toute confusion dans l'écriture de l'équation obtenue, le premier athlète sera nommé Mike et le deuxième, Nick. Ainsi, nous obtenons l'équation suivante :

$$N_{10}/N_{01} = [P_{mi} (1 - P_{ni})] / [P_{ni} (1 - P_{mi})] \quad (1)$$

où  $P_{mi}$  correspond à la probabilité que Mike réussisse son saut à la haie  $i$ ,  $P_{ni}$  correspond à la probabilité que Nick réussisse son saut à la haie  $i$ ,  $(1 - P_{ni})$  correspond à la probabilité que Nick échoue son saut à la haie  $i$  et  $(1 - P_{mi})$  correspond à la probabilité que Mike échoue son saut à la haie  $i$ .

Cependant, puisque l'objectif que nous souhaitons atteindre est d'arriver à établir quel athlète entre Mike et Nick est le plus fort, il ne faut pas que la comparaison entre les deux athlètes dépende de la haie sur laquelle ils compétitionnent. Mathématiquement, cela consiste à écrire l'équation 1 comme suit :

$$[P_{mi} (1 - P_{ni})] / [P_{ni} (1 - P_{mi})] = [P_{mj} (1 - P_{nj})] / [P_{nj} (1 - P_{mj})] \quad (2)$$

À partir de cette équation, il est possible d'isoler le ratio entre la probabilité que Nick, par exemple, réussisse son saut à la haie  $i$  et la probabilité qu'il l'échoue d'un seul côté de l'équation :

$$[P_{ni} / (1 - P_{ni})] = [P_{nj} / (1 - P_{nj})] \times [(1 - P_{mj}) / P_{mj}] \times [P_{mi} / (1 - P_{mi})] \quad (3)$$

Les équations 2 et 3 permettent donc de faire la comparaison entre Mike et Nick pour toutes haies  $i$  et  $j$  utilisées. Or, afin d'établir la force de chacun des athlètes, il ne faut pas que celle-ci dépende des athlètes choisis pour la comparaison. En effet, pour établir qui de Nick et de Mike est le plus fort, il faut que la force de Nick, par exemple, soit la même si on le compare avec Mike ou avec tout autre athlète. Par conséquent, il faut que n'importe quelle haie ou n'importe quel athlète puisse être choisi afin de définir le cadre de référence qui permettra d'établir la force d'un athlète.

Il est donc possible d'écrire l'équation 3 en comparant Nick à un athlète 0 (plutôt que Mike), avec lequel il compétitionne sur une haie 0 (plutôt que  $j$ ). De plus, il est possible de fixer que l'athlète 0 et la haie 0 ont un même calibre, c'est-à-dire que la probabilité  $P_{00}$  que l'athlète 0 réussisse un saut à la haie 0 est de 0,5. L'équation 3 s'écrit alors sous la forme 4 présentée à la page suivante.

$$[ P_{ni} / ( 1- P_{ni} ) ] = [ P_{n0} / ( 1- P_{n0} ) ] \times [ ( 1- P_{00} ) / P_{00} ] \times [ P_{0i} / ( 1- P_{0i} ) ] \quad (4)$$

$$= [ P_{n0} / ( 1- P_{n0} ) ] \times [ P_{0i} / ( 1- P_{0i} ) ] \quad (\text{car } P_{00} = 0,5) \quad (5)$$

$$= f(n) \times g(i) \quad (6)$$

où  $f(n)$  est une fonction qui ne dépend que de la force de l'athlète et  $g(i)$  une fonction qui ne dépend que du niveau de difficulté de la haie.

Ensuite, il est possible d'écrire le produit  $f(n) \times g(i)$  comme le ratio entre un paramètre  $b_n$  qui ne dépend que de l'athlète et d'un paramètre  $d_i$  qui ne dépend que de la haie, c'est-à-dire :

$$[ P_{ni} / (1- P_{ni}) ] = b_n / d_i \quad (7)$$

Nous obtenons alors une formule qui exprime mathématiquement la relation entre la force de l'athlète et le degré de difficulté de la haie. Le modèle ainsi proposé possède des paramètres pour l'athlète et pour la haie qui sont complètement séparés. Cela nous permet donc d'estimer la force d'un athlète indépendamment de la difficulté posée par la haie proposée et vice versa.

Or, une fois cette relation établie, ce qui nous intéresse vraiment est d'établir à quel point Mike est plus fort que Nick (par exemple). Pour y arriver, il faut pouvoir établir la différence entre la force de Mike et celle de Nick. Cette différence peut être obtenue en appliquant le logarithme aux deux côtés de l'équation 7 :

$$\text{Ln} [ P_{ni} / (1- P_{ni}) ] = \text{Ln} (b_n / d_i) \quad (8)$$

$$= \text{Ln} (b_n) - \text{Ln} (d_i) \quad (9)$$

$$= B_n - D_i \quad (10)$$

où  $B_n = \ln(b_n)$  et  $D_i = \ln(d_i)$ . Enfin, il est possible de réécrire l'équation 10 sous la forme suivante :

$$P_{ni} = \exp(B_n - D_i) / [1 + \exp(B_n - D_i)] \quad (11)$$

dans lesquelles  $B_n$  représente la force de l'athlète. Ce paramètre ne dépend que des attributs de cette personne et de la référence métrique d'origine. Il en va de même pour le paramètre  $D_i$ , qui représente la hauteur de la haie, et qui ne dépend que des caractéristiques de la haie et de la référence métrique d'origine.

Ces deux équations forment le modèle dichotomique de la famille des modèles de Rasch et les paramètres  $B_n$  et  $D_i$  sont plus souvent appelés l'habileté de la personne et l'indice de difficulté de l'item respectivement. Ainsi,  $P_{ni}$  est la probabilité que la personne  $n$  avec une habileté  $B_n$  réussisse à l'item  $i$  dont le degré de difficulté est  $D_i$ .

Un autre modèle appartenant à la famille des modèles de Rasch est le modèle *Binomial Trials*. Ce modèle est utilisé lorsqu'il y a deux réponses possibles, que le répondant possède plusieurs essais pour réussir un item, mais que le nombre d'essais est limité.

Enfin, parmi l'ensemble des modèles appartenant à la famille des modèles de Rasch, on retrouve les modèles suivants :

- a) le modèle de Poisson (*Poisson counts*). Ce modèle ressemble au modèle *Binomial Trials*. Cependant, le nombre d'essais est illimité.
- b) le modèle *Rating Scale*. Ce modèle s'applique lorsqu'il y a plus de deux catégories de réponse (pour les échelles de type Likert par exemple) et lorsque l'intervalle entre chaque catégorie de réponse demeure le même pour tous les items du questionnaire. Le modèle *Rating Scale* transforme chacun

des intervalles en situation dichotomique. Il se sert du point où la probabilité d'opter pour la prochaine catégorie est égale à celle de conserver la précédente de telle sorte que ces intervalles peuvent être interprétés comme le succès ou l'échec de passer à la catégorie suivante.

- c) le modèle *Partial Credit*. Ce modèle fonctionne comme le modèle *Rating Scale*. Cependant, les intervalles entre chaque catégorie peuvent varier d'un item à l'autre.

Les modèles présentés à la page précédente sont parmi les modèles unidimensionnels de la famille des modèles de Rasch les plus fréquemment utilisés. Un modèle unidimensionnel est un modèle dont la structure repose sur le fait qu'un seul facteur, l'habileté d'une personne ( $B_n$ ), permet de déterminer sa probabilité de réussir ( $P_{ni}$ ).

À la lumière de cette présentation, ce sont donc les modèles *Rating Scale* et *Partial Credit* qui s'appliquent le mieux aux enquêtes par questionnaire étudiées dans cette recherche. En effet, les échelles de réponses contenues dans les questionnaires utilisés pour cette recherche sont composées d'au moins trois catégories de réponse. Or, les questionnaires d'enquête de cette recherche utilisent une seule échelle de réponses et l'interprétation de cette échelle doit être la même pour tous les items du questionnaire (voir le chapitre 4 pour plus de détails). Par conséquent, c'est le modèle *Rating Scale* qui sera mis à contribution dans cette recherche. Les deux modèles seront toutefois présentés dans les prochaines sections. Il sera alors plus facile de comprendre le choix du modèle et comment il pourra être mis à contribution pour l'analyse des données.

### 3.3.1.1 Le modèle *Rating Scale*

L'introduction précédente sur les modèles de mesure de la famille des modèles de Rasch reposait sur une situation où les résultats possibles sont dichotomiques. Cependant, les chercheurs qui utilisent les questionnaires d'enquête pour recueillir leurs données vont fréquemment proposer un choix de réponses où les catégories offertes suivent une gradation plus fine. Par exemple, les choix offerts pourraient être « 4 = Totalement d'accord », « 3 = Plutôt d'accord », « 2 = Plutôt en désaccord » et « 1 = Totalement en désaccord ».

La section 3.3 avait permis de montrer que, selon Wright et Mok (2004), le modèle de mesure utilisé pour analyser des données recueillies à l'aide de ce type d'échelle doit permettre de faire des inférences, d'estimer la précision de ses inférences et fournir des moyens de détecter et d'évaluer les divergences entre les données recueillies et les mesures modélisées. Or, Wright et Mok (2004) soutiennent également que, pour qu'une mesure soit utile, il faut qu'elle soit linéaire et qu'elle puisse être reproduite. Par conséquent, pour être utile, l'échelle ordinaire des catégories de réponse fournie par une échelle de type Likert doit être transformée en une échelle à intervalles égaux. Le modèle *Rating Scale* permet cette transformation (Andrich, 1978).

Afin d'introduire le modèle *Rating Scale*, Wright et Mok (2004) utilisent l'exemple suivant : « Que pensez-vous de votre charge de travail cette session-ci? » avec les choix de réponses « Trop grande », « Correcte » et « Insuffisante » (dans cet ordre).

Dans cet exemple, si le choix d'une personne se situe entre « Trop grande » et « Correcte », sa décision indique déjà que la charge de travail n'est pas « Insuffisante ».



De même, si son choix se situe entre « Correcte » et « Insuffisante », sa décision indique que la charge de travail n'est pas « Trop grande ». Or, même si le nombre de catégories de réponse était plus grand, la décision et l'interprétation d'une réponse peuvent toujours se trouver entre deux catégories adjacentes (Wright & Mok, 2004).

Le point où la probabilité qu'une personne choisisse la catégorie suivante plutôt que la précédente est appelé « seuil » (*threshold*). Dans l'exemple ci-dessus, il y a deux seuils (F1 et F2). Si une personne n'aime pas les devoirs, alors son choix sera en deçà du premier seuil F1 et la personne adhérera à la catégorie « Trop grande ». Au contraire, si une personne apprécie les devoirs, sans pour autant souhaiter en recevoir une trop grande quantité, alors son choix dépassera le premier seuil sans franchir le second; son choix s'arrêtera donc sur « Correcte ». Enfin, si une personne adore les devoirs, alors son choix franchira les deux seuils et s'arrêtera sur « Insuffisante ».

À la frontière de chaque seuil, il est possible d'associer un score de « 0 » ou de « 1 » selon qu'une personne ait « échoué » à franchir le seuil ou qu'elle ait « réussi ». Ainsi, selon le continuum de réponses illustré au paragraphe précédent, il y aurait trois résultats possibles que Wright et Mok (2004) présentent selon le tableau suivant :

**Tableau 2 : Réponses possibles d'une personne en fonction de son amour pour les devoirs**

« Trop grande »	« Correcte »	« Insuffisante »
Score pour F1 = 0	Score pour F1 = 1	Score pour F1 = 1
Score pour F2 = 0	Score pour F2 = 0	Score pour F2 = 1
Score total = 0	Score total = 1	Score total = 2

F1
F2  
 ————— Amour des devoirs —————>

La probabilité de réussir (ou d'échouer) un seuil peut être décrite à l'aide du modèle dichotomique de la famille des modèles de Rasch. En effet, il est possible de faire un parallèle entre la probabilité de réussir un seuil et la probabilité qu'un athlète réussisse le saut d'une haie (exemple utilisé précédemment pour introduire les modèles de la famille des modèles de Rasch). Dans les deux cas, les résultats possibles sont dichotomiques (la réussite ou l'échec).

Or, le modèle dichotomique de la famille des modèles de Rasch peut s'écrire de la façon suivante (équation 10) :

$$\ln [ P_{ni} / (1 - P_{ni}) ] = B_n - D_i$$

où  $P_{ni}$  représentait la probabilité de réussir le saut d'une haie,  $(1 - P_{ni})$  correspondait à la probabilité de l'échouer,  $B_n$  représentait la force de l'athlète (son habileté) et  $D_i$  le degré de difficulté proposé par la haie.

En ajustant cette formule à l'exemple ci-dessus, les différents paramètres de l'équation pourraient être transposés de la façon suivante. Tout d'abord,  $P_{ni}$  représenterait la probabilité de réussir un seuil (plutôt qu'une haie). En d'autres mots, cela correspondrait à la probabilité de franchir ce seuil et de choisir la catégorie suivante. Ensuite,  $(1 - P_{ni})$  représentait la probabilité d'échouer une haie. Dans notre exemple, cela signifierait ne pas franchir un seuil. Ce paramètre pourrait donc s'écrire comme la probabilité de choisir la catégorie précédent le seuil, c'est-à-dire  $P_{n(i-1)}$ . De plus,  $B_n$  correspondait à l'habileté de l'athlète. Ce paramètre pourrait être transposé pour représenter l'opinion de la personne. Enfin,  $D_i$  correspondrait au degré de difficulté proposé par la question (plutôt que par la haie).

Cette transposition permet alors de réécrire l'équation 10 de la façon suivante :

$$\text{Ln} [ P_{ni} / P_{n(i-1)} ] = B_n - D_i \quad (12)$$

Or, cette dernière formulation constitue un cas particulier de l'équation générale du modèle *Rating Scale* qui est donnée par :

$$\text{Ln} [ P_{nix} / P_{n(i-1)x} ] = B_n - D_i - F_x \quad (13)$$

dans laquelle  $P_{nix}$  représente la probabilité que la personne  $n$  avec une opinion  $B_n$  adhère à la catégorie  $x$  (où  $x = 0$  à  $m-1$ , pour  $m$  catégories de réponse offertes) d'un item  $i$  dont le degré de difficulté est  $D_i$ . Le paramètre  $F_x$  correspond au seuil entre les catégories  $x-1$  et  $x$  ou, plus précisément, au point où la probabilité d'opter pour l'une ou l'autre des catégories est égale.  $F_x$  peut également être interprété comme le degré de difficulté observé dans la catégorie  $x$  par rapport à celui de la catégorie  $x-1$ , c'est-à-dire comme l'écart entre la catégorie  $x-1$  et la catégorie  $x$ . Enfin,  $P_{n(i-1)x}$  représente la probabilité que la personne  $n$  avec une opinion  $B_n$  adhère à la catégorie  $x-1$ .

Ainsi, dans une enquête par questionnaire auto-administré, l'habileté d'une personne ( $B_n$ ) correspond à son degré d'adhésion par rapport aux items qui lui sont proposés et le degré de difficulté d'un item ( $D_i$ ), à la difficulté qu'éprouvent les sujets à endosser cet item. Un item dont le niveau de difficulté est élevé est donc un item auquel il est difficile d'adhérer et une personne dont le niveau d'habileté est élevé est une personne dont l'opinion est généralement favorable aux items proposés.

Enfin, la présentation des modèles de la famille des modèles de Rasch de la section précédente a permis de préciser que le modèle *Rating Scale* s'applique lorsque plusieurs items d'un questionnaire partagent la même échelle de réponses et que les

catégories de réponse offertes possèdent des seuils identiques. Lorsque les seuils ne sont pas les mêmes pour tous les items, alors c'est le modèle *Partial Credit* qui devrait être utilisé.

Or, le chapitre 2 a permis de préciser que les questionnaires d'enquête de cette recherche utilisent une seule échelle de réponses et que l'interprétation de cette échelle devrait être la même pour tous les items du questionnaire. Par conséquent, c'est le modèle *Rating Scale* qui devra être utilisé pour l'analyse des données. La présentation du modèle *Partial Credit* proposée dans la prochaine section ne constituera donc qu'une brève introduction.

### 3.3.1.2 Le modèle *Partial Credit*

Le modèle *Partial Credit* est similaire au modèle *Rating Scale*. Cependant, dans ce cas-ci, les intervalles entre chaque catégorie peuvent varier d'un item à l'autre. Autrement dit, chacun des items possède ses propres paramètres de seuil  $F_x$ . Par conséquent, l'équation générale du modèle *Partial Credit* est donnée par :

$$\ln [ P_{nix} / P_{ni(x-1)} ] = B_n - D_i - F_{ix} \quad (14)$$

dans laquelle  $P_{nix}$  représente la probabilité que la personne  $n$  avec une opinion  $B_n$  adhère à la catégorie  $x$  (où  $x = 0$  à  $m-1$ , pour  $m$  catégories de réponse offertes) d'un item  $i$  dont le degré de difficulté est  $D_i$ . Le paramètre  $F_{ix}$  correspond au point, pour cet item, où la probabilité d'opter pour l'une ou l'autre des catégories est égale. Enfin,  $P_{ni(i-1)x}$  représente la probabilité que la personne  $n$  avec une opinion  $B_n$  adhère à la catégorie  $x-1$ .

### **3.3.2 Hypothèses et propriété des modèles de la famille des modèles de Rasch**

Les sections précédentes ont permis de montrer que les modèles *Rating Scale* et *Partial Credit* de la famille des modèles de Rasch constituent des modèles de mesure utiles pour l'analyse des données de cette recherche. En effet, la présentation des modèles qui a été faite a permis d'exposer comment les mesures produites à l'aide de ces modèles permettent aux chercheurs de faire des prédictions sur les réponses des participants. Or, pour être utiles, les modèles doivent également permettre aux chercheurs d'obtenir des résultats invariants (sect. 3.3).

L'invariance des résultats est une propriété théorique qui peut être obtenue lorsque les postulats de base des modèles sont vérifiés. L'unidimensionnalité et l'indépendance locale sont les deux hypothèses fondamentales des modèles *Rating Scale* et *Partial Credit* et se présentent comme des conditions à remplir afin de pouvoir les utiliser. Les prochaines sections présenteront donc ces différents concepts afin de mieux comprendre en quoi chacun consiste.

#### **3.3.2.1 Hypothèse de l'unidimensionnalité**

La section 3.3.1 avait présenté un modèle unidimensionnel comme un modèle dont la structure repose sur le fait qu'un seul facteur, l'habileté d'une personne (ou son opinion), permet de déterminer sa probabilité de réussir (ou d'adhérer) à un item d'un

questionnaire. Or, plusieurs études présentées dans ce travail ont montré que de nombreux facteurs (stress, fatigue, caractéristiques du questionnaire et de ses items, etc.) peuvent influencer les réponses des sujets interrogés. Ainsi, l'habileté des participants (ou l'opinion) ne constitue pas le seul facteur déterminant la réponse qu'ils fourniront à un item d'un questionnaire.

L'unidimensionnalité est donc une condition qui, strictement parlant, ne peut pas être remplie (Hambleton, Swaminathan & Rogers, 1991). Cependant, divers éléments permettent de mieux définir en quoi consiste l'unidimensionnalité. Selon Blais (1987), l'unidimensionnalité est présente lorsqu'il est possible de supposer qu'un seul facteur (l'habileté ou l'opinion par exemple) est principalement responsable des réponses observées. Ensuite, Blais indique qu'un questionnaire peut être qualifié d'« unidimensionnel » si chacun des items le composant sert à mesurer la même chose (l'habileté ou l'opinion). Blais précise également qu'un ensemble de données est considéré unidimensionnel s'il est possible d'établir, à l'aide d'une technique éprouvée, qu'un seul facteur dominant (l'habileté ou l'opinion) permet d'expliquer de façon satisfaisante les relations entre les mesures associées aux variables. Enfin, Blais mentionne que, de façon générale, il est possible d'assumer que les participants sont influencés par des éléments contextuels comme le stress ou la fatigue de façon uniforme et aléatoire alors qu'un facteur dominant comme l'habileté (ou l'opinion) influence leurs réponses de façon systématique.

Ainsi, puisque l'idéal de l'unidimensionnalité « pure » est difficile à réaliser, dans la pratique, l'hypothèse de l'unidimensionnalité consiste à démontrer qu'une seule dimension dominante (l'habileté ou l'opinion par exemple) est responsable des réponses des candidats interrogés (J.-G. Blais, 1987; Hambleton, Swaminathan & Rogers, 1991).

### **3.3.2.2 Hypothèse de l'indépendance locale**

L'indépendance locale est la condition selon laquelle les réponses d'un individu sont statistiquement indépendantes, c'est-à-dire que la réponse donnée par cet individu à un des items du questionnaire ne permet pas de prédire la réponse qu'il fournira à d'autres items de ce questionnaire. Concrètement, la démonstration de l'unidimensionnalité qui doit être faite garantit théoriquement le respect de la condition d'indépendance locale. L'inverse n'est cependant pas vrai.

En effet, lorsque l'hypothèse de l'unidimensionnalité est vérifiée, cela signifie qu'une seule dimension est responsable de la réponse d'un participant à chacun des items. Autrement dit, seul le niveau d'habileté du sujet (ou son opinion) permet de déterminer sa probabilité de réussir (ou d'adhérer) à un item donné et la réussite ou l'échec (ou l'adhésion) de cet individu à d'autres items du questionnaire n'ajoute rien de plus à notre connaissance (indépendance locale). Par conséquent, lorsque l'hypothèse de l'unidimensionnalité est vérifiée, alors logiquement, l'hypothèse de l'indépendance locale est également vérifiée.

### **3.3.2.3 Invariance (propriété)**

L'invariance est une propriété selon laquelle les estimations des paramètres d'un modèle sont indépendantes, sous certaines conditions, du groupe de sujets ciblés par l'opération de mesure ou du groupe d'items inclus (Bertrand & Blais, 2004; Hambleton, Swaminathan & Rogers, 1991). Ainsi, lorsqu'il y a présence de la propriété d'invariance,

les mesures produites par le modèle resteraient les mêmes peu importe le groupe de participants ayant servi à la calibration du test ou les caractéristiques des items ayant été utilisées dans le questionnaire.

La section 3.3 a permis de montrer que l'invariance n'était pas une propriété caractéristique de tous les modèles. En effet, l'invariance est une des caractéristiques principales qui permet de distinguer les modèles de la théorie classique des tests et ceux de la TRI (Hambleton, Swaminathan & Rogers, 1991). Cependant, la propriété d'invariance n'est pas limitée aux modèles de la TRI. En effet, Hambleton, Swaminathan et Rogers, précisent que l'invariance est une propriété bien connue dans le cas particulier des modèles de régression linéaire (par exemple).

En effet, supposons qu'une relation linéaire existe entre les variables X et Y. Lorsque l'hypothèse de la relation linéaire peut être confirmée (vérifiée statistiquement), la relation établie sera la même peu importe les valeurs prises par la variable X (J.-G. Blais & Ajar, 1992, p. 11; Hambleton, Swaminathan & Rogers, 1991, p. 19). En d'autres mots, les paramètres de la relation posséderont la propriété d'invariance.

L'invariance est donc une propriété théorique qui peut être obtenue lorsqu'il y a adéquation entre les données et le modèle. Comme pour l'exemple de régression linéaire qui vient d'être présenté, l'invariance tient au fait que l'on puisse démontrer que le modèle est raisonnablement adéquat pour représenter les données (J.-G. Blais & Ajar, 1992).

Ainsi, il est primordial d'étudier la qualité de l'ajustement entre les données et le modèle afin de vérifier si la propriété d'invariance peut être obtenue. De plus, bien que les hypothèses d'unidimensionnalité et d'indépendance locale se présentent généralement comme les conditions de base à remplir avant de pouvoir utiliser les modèles, le niveau d'applicabilité de ces conditions dépend du niveau d'ajustement entre les données et le



modèle (Smith Jr., 2004). Par conséquent, la première étape avant d'utiliser le modèle *Rating Scale* (ou *Partial Credit*) consiste à vérifier la qualité de l'ajustement entre les données et le modèle.

### 3.3.3 Ajustement entre les données et le modèle

Chaque fois qu'un chercheur travaille avec un modèle afin d'analyser les données de son étude, il est possible que ce modèle ne soit pas adéquat. En effet, le modèle choisi par le chercheur peut être perçu comme une hypothèse parmi d'autres utilisée pour « formaliser la relation entre la probabilité d'observer une réponse quelconque à un item et la position du candidat sur le continuum de l'habileté » (Bertrand & Blais, 2004, p. 191).

Ainsi, lorsque le chercheur fait le choix du modèle qu'il souhaite utiliser, il sélectionne certaines propriétés du modèle qui le rendent plus désirable que d'autres par rapport à certaines situations. Or, à partir du moment où il arrête son choix sur un modèle qui, à la lumière de ses caractéristiques et de ses propriétés, lui apparaît comme étant le plus approprié pour sa recherche, il doit faire la démonstration empirique que celui-ci est effectivement approprié pour représenter les données qu'il a recueillies (Bertrand & Blais, 2004).

La démonstration que le chercheur doit faire se présente de façon différente selon que le chercheur travaille avec la famille des modèles de Rasch ou non. Généralement, le chercheur utilise les statistiques d'ajustement afin de déterminer si un modèle s'ajuste bien aux données. Différents modèles sont donc comparés afin de déterminer celui qui traduit le mieux les données recueillies. Ainsi, le modèle est choisi en fonction des données à analyser. L'idée sous-jacente est que les données sont fixes et qu'il faut choisir le modèle

qui s'y ajuste le mieux (Andrich, 1988). Une telle approche s'inscrit dans le courant général des travaux de recherche sur l'ajustement statistique des modèles aux données (*goodness-of-fit*).

Cependant, la démonstration qui sera faite par le chercheur travaillant avec la famille des modèles de Rasch est différente. En effet, la relation qui est perçue entre les données et le modèle est plus symétrique (Andrich, 1988). Ainsi, Rasch (1980) soutient que lorsqu'il y a discordance entre les données et le modèle, les sources de discordance peuvent provenir du modèle comme des données. Andrich appuie cette idée en précisant que lorsqu'il y a discordance entre les données et le modèle, il est possible que les données s'égarer de l'objectif de recherche ou que certains problèmes soient survenus lors de leur collecte. Andrich ajoute que, dans ces cas, il est possible que le simple fait de modifier certains aspects de la cueillette des données, de leur interprétation ou de leur cadre de référence produise des résultats plus pertinents que de changer le modèle de mesure utilisé.

Ainsi, le chercheur qui travaille avec la famille des modèles de Rasch a accès à différents indices statistiques afin de détecter s'il existe des sources de discordance entre les données et le modèle. Ces indices statistiques permettent au chercheur d'identifier les données qui ne rencontrent pas les conditions de base du modèle (qui ne s'ajustent pas bien au modèle), d'examiner les caractéristiques de ces données et de vérifier de quelle façon elles contreviennent à la mesure (Smith Jr., 2004). L'analyse de la qualité de l'ajustement entre les données et le modèle permet donc au chercheur de porter un jugement sur la qualité de la prédiction que le modèle lui permettra de réaliser (Smith Jr., 2004).

Suite à une revue de la littérature, Hambleton et Swaminathan (1985) ont pu déterminer que la qualité de l'ajustement entre les données et le modèle devrait être

vérifiée de trois façons différentes. Tout d'abord, le chercheur doit vérifier que les données répondent aux conditions d'application du modèle. Par exemple, cela pourrait correspondre à vérifier les hypothèses d'unidimensionnalité et d'indépendance locale. Ensuite, le chercheur doit vérifier que les avantages découlant de l'utilisation du modèle choisi sont obtenus. Cela pourrait se traduire, par exemple, par la vérification de la propriété d'invariance dans les valeurs estimées pour les paramètres des items et de l'habileté des participants. Finalement, le chercheur doit étudier les différences entre les paramètres estimés par le modèle et les données observées.

La différence entre la valeur modélisée et la valeur concrètement observée fournit donc une certaine indication de la qualité de l'ajustement (Wright & Mok, 2004). En effet, si la différence est grande, alors l'expérience concrète n'est pas un exemple utile de la valeur modélisée. En revanche, une petite différence implique alors que la valeur modélisée est robuste par rapport à l'expérience concrète ou pour des expériences similaires futures.

Or, bien que la matrice des résidus (différences entre la valeur modélisée et la valeur concrètement observée) puisse donner une certaine indication de la qualité de l'ajustement entre les données et le modèle, la matrice des résidus pose plusieurs problèmes (Smith, 2004). En effet, Smith soutient que la variance de chacun des éléments de cette matrice est différente. Ainsi, l'interprétation de chacune de ces valeurs dépend des valeurs modélisées (valeurs estimées pour le degré d'adhésion des personnes, par exemple, ou pour le degré de difficulté posé par les items pour les endosser). De plus, la somme des différences sur chacune des colonnes ou des rangées de la matrice est nulle. Par conséquent, la matrice ne fournit aucune information utile par rapport aux personnes ou aux items.

Afin de contrôler les différentes variances des éléments de la matrice des résidus, Smith (2004) montre qu'il est possible de diviser chacun des éléments par son écart-type

et de créer une matrice des résidus centrée et réduite. Cette nouvelle matrice permet alors de contrôler la variance. Cependant elle ne permet pas de régler les problèmes de somme sur chacune des rangées (personnes) ou des colonnes (items) (Smith, 2004).

Plusieurs solutions mathématiques permettent de corriger ce problème (Smith, 2004). La solution la plus fréquente consiste toutefois à élever les valeurs de la matrice des résidus centrés et réduits au carré. En effet, Smith soutient que cette solution permet au chercheur de faire la somme des éléments sur chacune des rangées ou des colonnes sans annuler le pouvoir de détecter les réponses inattendues de la matrice.

La nouvelle matrice ainsi créée possède donc des caractéristiques intéressantes (Smith, 2004). En effet, la distribution des résidus centrés, réduits et mis au carré s'apparente à celle d'un khi-carré à un degré de liberté. Il faut toutefois être prudent en faisant ce rapprochement, car ce rapprochement est vrai seulement si les trois conditions suivantes soient remplies. Tout d'abord, la probabilité d'obtenir une bonne réponse doit être connue, c'est-à-dire qu'elle doit avoir été calculée en fonction de l'estimation de l'habileté des personnes (ou de leur opinion) et du niveau de difficulté des items (ou de la difficulté à les endosser). Ensuite, les choix de réponses offerts aux participants devraient être continus et non discrets. Enfin, les données doivent s'ajuster au modèle.

Or, puisque les éléments de la matrice des résidus standardisés et mis au carré s'apparentent à des statistique khi-carré, Smith (2004) présente trois façons d'utiliser les éléments de cette matrice afin d'examiner la qualité de l'adéquation entre les données et le modèle. En premier lieu, il est possible de faire la somme de tous les éléments de cette matrice afin de vérifier la qualité de l'ajustement général entre les données et le modèle. Ensuite, il est possible de faire la somme des éléments sur chacune des colonnes afin de vérifier la qualité de l'ajustement des items (*item fit*). Enfin, il est possible de faire la somme

des éléments sur chacune des rangées afin de vérifier la qualité de l'ajustement des personnes (*person fit*).

Selon Wright et Mok (2004), la question de l'ajustement global est principalement fondée sur la qualité des données et sur la validité du construit, c'est-à-dire que l'ajustement global entre les données et le modèle revient surtout à vérifier l'unidimensionnalité de l'ensemble des scores et leur indépendance locale. Par conséquent, la qualité de l'ajustement entre les données et le modèle se fait généralement en étudiant les deux composantes principales suivantes : les items du questionnaire et les sujets.

### **3.3.3.1 Ajustement pour les items**

La qualité de l'ajustement pour les items s'évalue principalement de deux façons. Tout d'abord, une première analyse peut être faite à l'aide de la représentation graphique du modèle. La méthode consiste à comparer les distributions théoriques des items (distributions modélisées ou courbes CCI) et les distributions observées.

Ensuite, une deuxième analyse peut être faite à l'aide d'indices statistiques. Généralement, les indices statistiques pour les items peuvent être classés selon trois types : *total fit*, *between fit* et *within fit* (Smith, 2004). Smith décrit l'indice statistique *total fit* comme correspondant à la somme des résidus sur chacune des colonnes (items) de la matrice des résidus standardisés mis au carré.

L'indice *between fit* se calcule après qu'une des caractéristiques propres aux participants a été utilisée afin de séparer les participants en sous-groupes pertinents (Smith, 2004). La caractéristique utilisée peut être reliée à l'habileté des individus ou à des

caractéristiques comme le sexe, l'âge ou la langue maternelle. L'indice *between fit* peut être particulièrement utile pour déceler des niveaux de performance différents qui seraient liés à des sous-groupes de participants différents.

L'indice *within fit* se calcule de la même façon que l'indice *total fit* (Smith, 2004). Cependant, plutôt que de calculer la somme des résidus de chacun des items pour tous les participants, la somme est faite sur un sous-groupe de participants (comme pour l'indice *between fit*). L'avantage de l'indice *within fit* est qu'il permet d'identifier des patrons de réponses à l'intérieur de sous-groupes alors que ces patrons pourraient passer inaperçus à travers l'ensemble des participants.

Chacun de ces indices (*total fit*, *between fit* et *within fit*) peut être calculé de deux façons différentes (Smith, 2004). D'une part, ils peuvent être calculés à partir de la matrice des résidus standardisés mis au carré tel qu'indiquée dans les paragraphes précédents. Cette première méthode de calcul correspond à la version non pondérée (*unweighted*) des statistiques d'ajustement pour les items. L'indice statistique correspondant est alors appelé l'indice *outfit* (*out-lier sensitive mean square residual goodness of fit statistic*).

L'expérience a démontré que l'indice statistique *outfit* était très sensible aux données aberrantes (Smith, 2004). Par conséquent, une deuxième version de l'indice a été développée. Dans la deuxième version, chacun des éléments de la matrice standardisée mise au carré est divisé par la fonction d'information avant que les indices *total fit*, *between fit* et *within fit* ne soient calculés. L'effet de cette pondération est ensuite annulé en divisant la somme obtenue sur chacune des colonnes par la somme des coefficients (poids) utilisés. Cette méthode de calcul correspond à la version pondérée (*weighted*) des statistiques d'ajustement pour les items. L'indice est alors appelé l'indice *infit* (*information weighted mean square residual goodness of fit statistic*). Les études relatives à cette

deuxième forme ont révélé qu'elle s'avérait un meilleur choix dans l'étude de l'ajustement des données au modèle (Smith, 2004).

L'indice *outfit* est donc un indice d'ajustement externe qui met l'accent sur les réponses inattendues qui s'éloignent du score de la personne (Wright & Masters, 1982). Par exemple, l'*outfit* permet de déceler des erreurs d'inattention, des réponses réussies de façon inattendue à des items difficiles par des sujets de faible habileté ou des mauvaises réponses fournies de façon inattendues par des sujets possédant un niveau d'habileté élevé à des items plus faciles (Linacre & Wright, 1994). Les problèmes signalés par cet indice sont généralement simples à diagnostiquer et à corriger.

Pour sa part, la valeur *infit* est un indice d'ajustement interne (ou pondéré) qui met l'accent sur les réponses inattendues qui sont près du score de la personne (Wright & Masters, 1982). Il s'agit d'une statistique d'ajustement qui permet de concentrer l'analyse sur la performance d'un item ou d'un individu. L'indice *infit* est influencé par les patrons de réponses (Linacre & Wright, 1994). Les problèmes décelés grâce à cet indice sont généralement difficiles à diagnostiquer et à corriger. Ils présentent donc un risque plus grand pour la mesure.

En résumé, l'ajustement pour les items s'évalue à l'aide de deux indices principaux : les indices « *infit* » et « *outfit* ». Les deux indices permettent de diagnostiquer des problèmes différents. Ils doivent donc être utilisés de façon complémentaire dans une analyse de données.

### **3.3.3.2 Ajustement pour les participants**

L'analyse de la qualité de l'ajustement pour les participants est similaire à celle qui est faite pour les items du questionnaire (Daniel, 1999). En effet, le niveau de difficulté d'un item est évalué sur la base de la performance des participants à cet item, alors que leur niveau d'habileté est connu. Daniel précise que, de la même façon, l'habileté d'un participant est évaluée sur la base de la performance de l'individu sur un ensemble d'items dont le niveau de difficulté est connu.

Ainsi, la qualité de l'ajustement pour les participants s'évalue à partir d'indices statistiques calculés de façon parallèle à celle qui vient d'être présentée pour l'ajustement des items. La différence principale entre la qualité de l'ajustement pour les participants et celui pour les items est une question de quantité. En effet, il y a généralement plus de participants inclus dans une enquête ou un examen, par exemple, que d'items inclus dans le questionnaire utilisé (Daniel, 1999).

Tout comme pour l'ajustement pour les items, l'ajustement pour les participants se divise selon trois types d'indices : *total fit*, *between fit* et *within fit*. L'indice statistique *total fit* correspond donc à la somme des résidus sur chacune des rangées (participants) plutôt qu'à la somme sur chacune des colonnes. L'indice *between fit* se base sur une caractéristique propre aux items afin de séparer les items en sous-groupes plutôt que sur une caractéristique propre aux participants. Cette caractéristique peut donc être la difficulté de l'item, sa position dans le questionnaire, le type d'item auquel il correspond, le niveau cognitif requis pour y répondre, etc. (Smith, 2004). Son calcul, de même que celui de l'indice *between fit*, se fait de la même façon que celle décrite pour l'ajustement pour les items.



L'indice *between fit* pour les participants permet de détecter les différences de performance d'un individu par rapport à différents sous-groupes d'items. Cet indice permet donc de vérifier la variance de l'habileté d'un participant sur ces différents sous-groupes. Selon Rasch (1980), l'habileté d'un individu serait invariante; ainsi, elle devrait être la même, peu importe le sous-groupe d'items choisi.

Enfin, tout comme pour les indices statistiques utilisés pour l'analyse de l'ajustement pour les items, les indices statistiques utilisés pour l'analyse de l'ajustement pour les participants peuvent être calculés selon deux versions : la version originale non pondérée (l'indice *outfit*) ou la version pondérée (l'indice *infit*).

### **3.4 Synthèse du cadre conceptuel**

Afin de bien saisir tout ce qui a été vu jusqu'ici, il est important de revenir sur les concepts principaux qui ont été présentés et de mettre en lumière les relations qui les unissent. (La figure 1 de la page 63 permet de présenter visuellement ce résumé.)

Ainsi, dans le présent chapitre, deux sujets principaux ont été abordés. Tout d'abord, les facteurs pouvant influencer les réponses des participants dans une enquête par questionnaires auto-administrés ont été présentés. Ainsi, les caractéristiques d'un questionnaire et de ses items ont été étudiés. Cela a permis de confirmer que l'influence de la formulation des items devait être étudiée de deux façons : en analysant les réponses fournies par les participants et en examinant les points d'ancrage de l'échelle de mesure associée aux catégories de réponse. Ensuite, il a été montré que le modèle *Rating Scale* de la famille des modèles de Rasch constituait un modèle de mesure utile pour l'analyse

les données de cette recherche. En effet, lorsqu'il est adéquat, ce modèle permet d'obtenir des résultats invariants.

Or, un modèle de mesure peut se définir comme une démarche d'analyse rigoureuse qui consiste à modéliser les données recueillies (à l'aide d'une enquête par questionnaire par exemple), puis à comparer les résultats produits par le modèle à des critères de référence permettant de s'assurer que les conditions de base du modèle sont vérifiées (sect. 3.3). Ainsi, bien que le modèle qui sera utilisé dans cette étude ait été introduit (de même que certaines statistiques d'ajustement), les critères de référence qui seront utilisés dans la démarche d'analyse n'ont pas été présentés. En effet, ces critères n'ajoutaient aucune information qui aurait permis de mieux comprendre comment le modèle *Rating Scale* nous permettra d'atteindre les objectifs de cette recherche. Nous avons donc choisi de présenter ces détails un peu plus loin (voir chap. 6).

**Processus de réponse à une enquête par questionnaire auto-administré :**

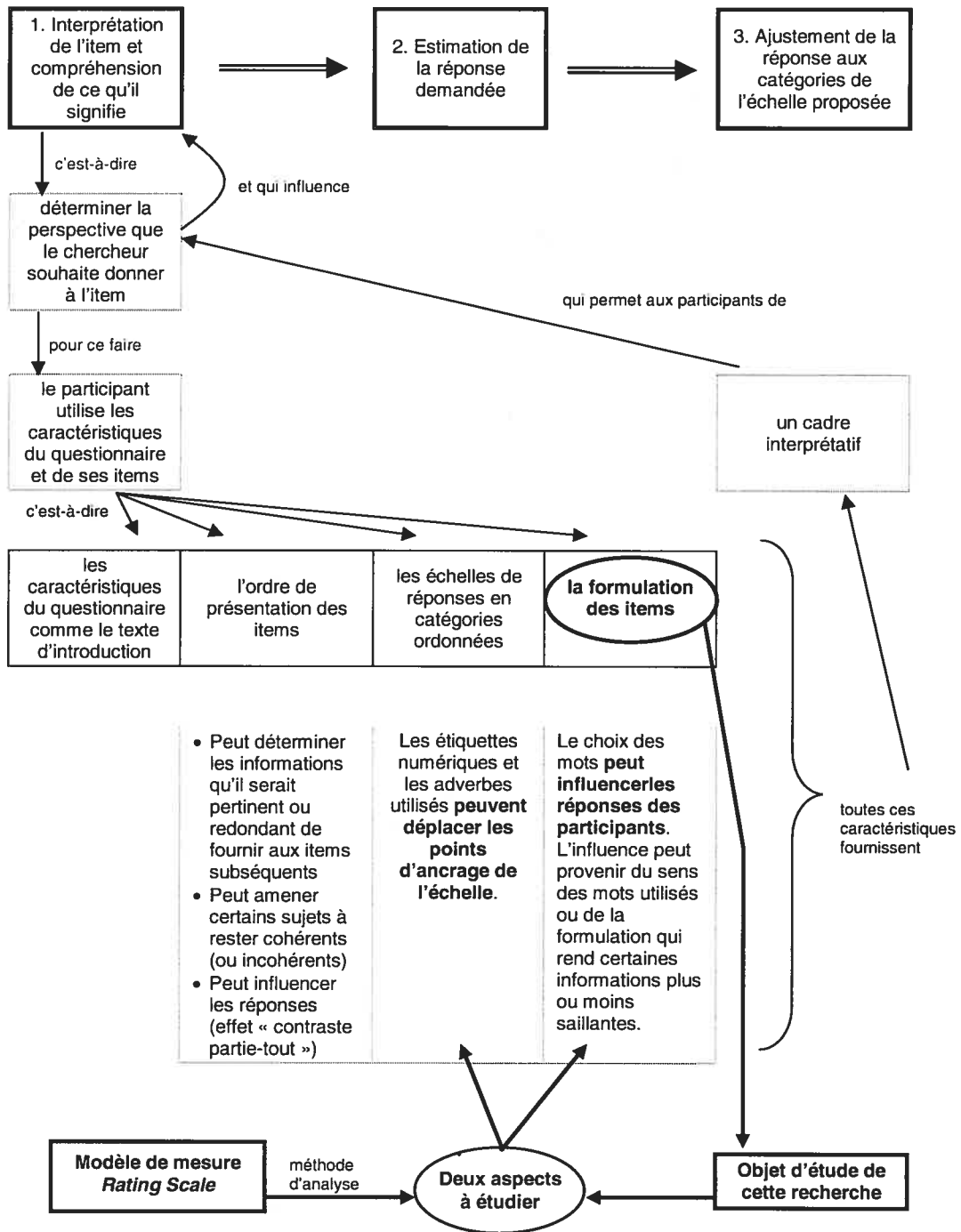


Figure 1 : Schéma synthèse du cadre conceptuel

## **4. MÉTHODOLOGIE**

Le chapitre précédent a permis de bien circonscrire l'objet d'étude de cette recherche en établissant les termes et les concepts qui le caractérisent. Dans ce chapitre, ce sont l'instrument utilisé pour la cueillette des données de cette recherche, l'échantillon, ainsi que le choix méthodologique pour le traitement des données qui seront présentés.

### ***4.1 Description de l'instrument de cueillette d'information***

En 1999, le Centre de Formation Initiale des Maîtres (CFIM) de l'Université de Montréal a mis au point un questionnaire d'enquête auto-administré afin d'évaluer ses programmes de premier cycle universitaire en formation des maîtres (l'annexe A de ce travail présente la version A du questionnaire 2003). L'objectif de ce questionnaire était de recueillir l'opinion et l'attitude des premiers gradués du nouveau programme de 4 ans de formation des maîtres. Le questionnaire a été distribué pour la première fois aux étudiants de troisième et de quatrième année des programmes d'enseignement en éducation préscolaire et enseignement primaire et à ceux en enseignement secondaire au printemps de l'an 2000.

Le questionnaire original était composé de huit sections. La première section portait sur la perception générale des étudiants par rapport à la formation qu'ils ont reçue dans leur programme d'étude. La seconde portait sur la perception des étudiants par rapport à leur préparation à l'enseignement. C'est cette deuxième section qui sera l'objet d'étude de cette recherche. La troisième section s'enquérissait de leur opinion par rapport à

leurs stages. La quatrième partie portait sur leur charge d'étude et leur gestion du temps; la cinquième s'intéressait à la poursuite de leurs études; la sixième les invitait à exprimer leurs commentaires sur leur programme d'études, et les deux dernières sections étaient composées de questions d'ordre plus démographiques. Le questionnaire a été modifié avec les années et la version 2006 ne comporte plus que quatre sections (perception générale de la formation, préparation à l'enseignement, les stages et les renseignements généraux d'ordre plus démographiques).

Seule la section concernant la préparation à l'enseignement a été retenue comme objet d'étude pour le présent travail. Dans cette section, les étudiants doivent répondre à 20 items introduits par la phrase « Je considère que mon programme d'études m'a permis de développer des compétences pour... ». L'échelle de réponses fournie aux répondants est une échelle de type Likert en six points. L'échelle de référence est strictement positive (depuis la version de 2002) et les catégories de réponse sont des valeurs numériques. Seules la première et la dernière catégorie possèdent une étiquette. Ainsi, les catégories de réponse vont de « Tout à fait en désaccord = 1 » à « Tout à fait d'accord = 6 » (les catégories intermédiaires étant seulement identifiées par les chiffres de 2 à 5).

Parmi les 20 items proposés dans la version originale du questionnaire, deux portent sur les relations avec les parents d'élèves (items 2 et 20). Un autre (item 14) porte sur les relations avec les membres de l'équipe-école. Dix items (1, 3, 5, 6, 7, 8, 9, 13, 15 et 19) sont liés au développement du jugement professionnel (évaluer, identifier, diriger vers, etc.). Un des items (item 4) est associé à la formation de base (maîtriser les contenus à enseigner) et quatre portent sur l'acte pédagogique lui-même (items 10, 11, 16 et 17). Enfin, il y en a deux qui portent sur des questions d'éthique (items 12 et 18). Ainsi, tous les items sauf trois abordent un aspect qui est soit en lien avec les élèves soit avec la classe. Par ailleurs, le grand nombre d'items directement liés au développement du jugement professionnel est notable. Or, il s'agit de compétences importantes préconisées par le

Ministère de l'Éducation du Québec (MEQ) pour le programme de formation des maîtres. Il n'est donc pas surprenant que le questionnaire contienne autant d'items sur le sujet.

En observant les items, nous pouvons constater qu'ils sont formés de deux parties. La première partie est constituée d'un verbe (identifier, répondre, construire, etc.), et la deuxième partie vient préciser la première en mettant l'accent sur une compétence en enseignement (discipline, correction, méthodes de travail, évaluations, etc.). Ainsi, une fois décomposé en deux, chacun des items peut être manipulé pour former des items qui diffèrent plus ou moins de l'item de départ. Les modifications qui ont été apportées aux différentes versions de questionnaires ne changent que légèrement le sens de l'item (sauf dans la version 2005 où l'un des items (20) a été complètement modifié). Les mots ont donc été manipulés soit dans le but de préciser l'objet de l'item, soit pour le rendre plus général. L'objectif de ces manipulations étant de provoquer un déplacement notable des positions des items sur le continuum du niveau de difficulté avec un minimum de changement dans la formulation des items.

Le questionnaire est distribué à chaque printemps depuis 2000. Dans la version de l'année 2000, l'ordre des questions de cette section avait été déterminé par les concepteurs du questionnaire et l'échelle de type Likert qui était présentée, était une échelle en quatre points : tout à fait d'accord, plutôt d'accord, plutôt en désaccord, tout à fait en désaccord, codés respectivement 1, 2, 3, 4. Cette version sera identifiée comme la version A du questionnaire.

Au printemps de l'année 2001, deux versions du questionnaire ont été distribuées. La première version était identique à la version de l'année 2000 (version A). Pour la deuxième version (2001), l'ordre de présentation des items de la deuxième section a été modifié. En effet, les items où les réponses étaient les moins favorables et les items qui étaient les plus favorables suite à la collecte de l'année 2000 ont été respectivement

placés au début et à la fin du questionnaire. Cette version modifiée sera identifiée comme la version B de la collecte 2001.

À partir de 2001, l'attribution des versions A et B du questionnaire a été faite de la façon suivante : les paquets de questionnaires distribués aux étudiants comportaient les versions A et B en alternance. Cela permettait d'assurer une distribution aléatoire des questionnaires auprès des étudiants.

Au printemps de l'année 2002, deux versions du questionnaire ont de nouveau été distribuées. La première version était identique aux versions 2000 et 2001A du questionnaire. Seule l'échelle de réponses a été modifiée. En effet, l'échelle de réponses originale allait de « Tout à fait d'accord = 1 » à « Tout à fait en désaccord = 4 ». Pour l'année 2002, l'échelle a été inversée et le nombre de catégories a été augmenté. Ainsi, l'échelle allait de « Tout à fait en désaccord = 1 » à « Tout à fait en accord = 10 ». Cette première version du questionnaire, incluant les changements apportés à l'échelle, correspond donc à la version A de la collecte 2002.

Dans la deuxième version du questionnaire de l'année 2002 (version B), la formulation de quatre items a été modifiée. L'ordre de sept items également. Deux items ont vu leur formulation et leur ordre modifiés. Deux items ont seulement été reformulés alors que pour les cinq autres, c'est seulement leur ordre de présentation qui a été modifié. Les deux items où les réponses étaient les plus favorables en 2001 ont été placés au début de la section.

Il est à noter que les travaux effectués sur les questionnaires de 2000 à 2002 étaient exploratoires et visaient essentiellement à témoigner du réalisme de la recherche. Une fois la possibilité d'utiliser ces questionnaires pour la recherche établie, les questionnaires ont été modifiés de façon à pouvoir étudier l'influence de la formulation des

items sur les réponses des participants et sur l'échelle de mesure associée. Par conséquent, à partir de la version 2003 du questionnaire, seule la formulation des items allait être modifiée d'une année à l'autre (ces changements sont présentés dans l'annexe B de ce travail).

Ainsi, au printemps de l'année 2003, deux versions du questionnaire ont été distribuées. L'échelle de réponses offerte aux répondants a été modifiée dans les deux versions. Le nombre de catégories fournies a été réduit et est passé de dix à six. L'ordre de présentation des items des deux versions est identique et correspond à l'ordre qui avait été utilisé dans les versions 2000, 2001A et 2002A. Dans la version A du questionnaire, la formulation de trois items (items 2, 14 et 20) est différente et correspond à la formulation qui avait été utilisée dans la version 2002B. Pour la version B, c'est la formulation de quatre items qui a été modifiée (items 2, 3, 15 et 20). Les modifications apportées à ces quatre items viennent préciser l'objet de l'item. Par exemple, l'item 2 de la version A se lit « Répondre aux questions des parents lors de la présentation du bulletin » alors que le même item dans la version B est « Aider les parents à comprendre le bulletin de leur enfant ». Ainsi, la formulation est différente, mais le sens de l'item est légèrement modifié et la signification plus précise.

En 2004, l'étude se fait de nouveau dans les mêmes conditions, c'est-à-dire à partir de deux versions du questionnaire. La version 2004A est identique à la version 2003A. Pour la version B, la formulation des items a été modifiée pour 11 des 20 items de la section à l'étude. L'ordre des items demeure inchangé. Les modifications ont été apportées à la formulation des items en fonction de deux critères : l'ajustement statistique des items au modèle et offrir une formulation plus précise aux répondants. Ainsi, la formulation de cinq items a été changée parce que l'étude de l'ajustement statistique de ces items dans les versions précédentes du questionnaire révélait un problème potentiel. Par ailleurs, des changements de mots ont été appliqués à cinq autres items de façon à en



rendre l'objet plus précis. Un des items a cependant été reformulé de façon à ce que le sujet soit plus général (item 15).

En 2005, deux nouvelles versions du questionnaire ont été préparées. Dans la version A du questionnaire, la majorité des items sont identiques à ceux utilisés dans la version 2004A. Toutefois, trois des items proviennent de la version 2004B (items 1, 2, et 17). L'item 3 provient également de la version 2004B. Cependant, un ajout de mots a été apporté afin de le rendre plus précis. L'item 20 a, pour sa part, été complètement modifié<sup>2</sup>. Dans les versions 2004 du questionnaire, l'item 20 portait sur la relation avec les parents (« Discuter avec les parents des difficultés de leur enfant » et « Expliquer aux parents quelles sont les difficultés de leur enfant »). Dans les versions 2005, l'item 20 porte maintenant sur l'acte pédagogique (« Guider les élèves dans leurs apprentissages » et « Discuter avec les élèves de leurs difficultés »). La version B du questionnaire de l'année 2005 est similaire à la version 2005A. Cependant, huit items ont été reformulés. Pour sept de ces items (4, 9, 11, 14, 15, 18 et 20), les changements de mots appliqués rendent l'item plus général. En revanche, pour l'autre item (item 5), les modifications rendent l'objet de l'item plus précis.

Enfin, deux nouvelles versions du questionnaire ont été produites au printemps de l'année 2006. Dans la version A du questionnaire, les items utilisés sont identiques à ceux de la version 2005A sauf pour l'item 3 qui provient de la version 2004B et l'item 15 qui provient de la version 2003B. Dans la version B du questionnaire, les items utilisés sont identiques à ceux de la version A sauf pour les items 5, 14 et 18 dont la formulation correspond à celle de la version 2005B.

---

<sup>2</sup> Le questionnaire n'est pas développé en vase clos. Ainsi, certaines modifications ont été apportées suite à des discussions avec les responsables du CFIM.

Des changements de mots ont également été apportés à la formulation des étiquettes des catégories de réponse de l'échelle proposée. En effet, la section 3.2.4 a permis de présenter des résultats démontrant que les adverbess utilisés pour identifier chacune des catégories de réponse de l'échelle offerte pouvaient provoquer un déplacement des points d'ancrage de l'échelle de réponses. Les questionnaires de l'année 2006 ont donc été modifiés afin d'étudier cette influence. Ainsi, l'échelle de réponses de la version A du questionnaire est constituée de catégories allant de « 1 = Tout à fait en désaccord » à « 6 = Tout à fait en accord » (les catégories intermédiaires étant seulement identifiées par les chiffres de 2 à 5). Dans la version B, l'échelle de réponses est plutôt constituée de catégories allant de « 1 = En désaccord » à « 6 = En accord » (les catégories intermédiaires étant seulement identifiées par les chiffres de 2 à 5).

Ainsi, aucun changement de mots nouveau n'a été appliqué aux items de cette version du questionnaire. En effet, l'objectif principal de ces questionnaires est d'étudier l'influence du choix des mots dans la formulation des étiquettes des catégories de réponse de l'échelle proposée. Les questionnaires de l'année 2006 ne seront donc pas utilisés dans cette recherche.

## ***4.2 Population étudiée et cueillette des données***

Les questionnaires qui seront utilisés dans le cadre de cette recherche sont ceux des années 2003, 2004 et 2005, car ce sont les questionnaires qui ont été modifiés de façon à pouvoir étudier l'influence de la formulation des items sur les réponses des participants et sur l'échelle de mesure associée.

La population utilisée pour cette recherche est celle formée par les étudiants et les étudiantes de troisième et de quatrième année du programme d'enseignement en éducation préscolaire et enseignement primaire. En effet, le nombre d'étudiants et d'étudiantes ayant répondu aux différentes versions du questionnaire pour ce programme était généralement plus élevé que pour celui du programme en enseignement secondaire. Il est à noter que lors du recueil des données de 2004 et de 2005, le CFIM a décidé de ne plus interroger les étudiants de troisième année. Par conséquent, pour ces deux années, la population visée est uniquement composée des étudiants de quatrième année.

Enfin, la section précédente a permis de préciser que les paquets de questionnaires distribués aux étudiants comportent les versions A et B en alternance. Les deux versions du questionnaire ont donc été distribuées aléatoirement parmi la population ciblée. Par conséquent, à moins d'erreurs systématiques, la distribution utilisée a permis d'obtenir deux groupes de répondants à peu près équivalents et relativement homogènes.

### ***4.3 Choix méthodologique pour le traitement des données***

La modélisation mathématique est généralement liée de près à des processus de calcul complexes. Ces calculs nous permettent d'estimer les paramètres, de tracer la courbe caractéristique des items, de vérifier les hypothèses de bases des modèles, etc. Ainsi, l'utilisation de logiciels pour le traitement des données s'avère nécessaire.

Plusieurs logiciels permettant d'utiliser les modèles de la TRI existent sur le marché. Citons par exemple les logiciels BILOG, CONQUEST, RUMM2020 ou WINSTEP. De ces modèles, certains sont plus particulièrement dédiés aux modèles de la famille de Rasch. C'est le cas du logiciel WINSTEP qui intègre, entre autres, les modèles *Rating*

*Scale* et *Partial Credit*. L'objectif de cette recherche étant de mettre à contribution le modèle *Rating Scale*, le logiciel WINSTEP est le programmes retenu pour la modélisation des données.

## 5. LIMITES DE LA RECHERCHE

Cette recherche possède plusieurs limites. Tout d'abord, puisque l'instrument de cueillette de données utilisé dans cette recherche est le questionnaire d'enquête du CFIM visant l'évaluation de ses programmes de premier cycle universitaire en formation des maîtres, nous n'avons que très peu de pouvoir sur le questionnaire. Par conséquent, certains changements apportés aux différentes versions du questionnaire l'ont été suite à des discussions avec les responsables du CFIM et non pas de façon à pouvoir étudier l'influence de la formulation des items.

De la même façon, nous n'avons que très peu de contrôle sur la distribution des questionnaires. Ainsi, nous n'avons pas pu nous assurer que le nombre de questionnaires retournés était maximum. La taille des échantillons est donc restreinte pour certaines des années utilisées dans ce travail.

De plus, les réponses obtenues par l'entremise de ces questionnaires pourraient être influencées par différentes situations liées au programme (problèmes avec certains professeurs, problèmes de stages, grève étudiante, etc.) et sur lesquelles nous n'avons aucune emprise.

En contrepartie, l'utilisation de ces questionnaires nous a permis d'étudier un instrument concret, dynamique et qui devait s'ajuster aux contraintes et aux besoins de la réalité. Les données recueillies à l'aide de ces questionnaires étaient donc tangibles, c'est-à-dire conformes à une situation réelle d'enquête par questionnaire.

Par ailleurs, une autre des limites rencontrées dans cette recherche est l'utilisation de logiciels pour la modélisation des données. En effet, les résultats de notre analyse dépendent des procédures utilisées par le logiciel et des options offertes. La précision des calculs effectués de même que la rigueur de la programmation sont également à considérer.

Enfin, il s'agit d'une recherche exploratoire et les résultats obtenus devront être vérifiés par d'autres recherches.

## **6. ANALYSE DES DONNÉES**

Les chapitres précédents ont permis de montrer que le modèle *Rating Scale* de la famille des modèles de Rasch constituait un modèle de mesure utile pour l'analyse des données de cette recherche. En effet, lorsqu'il est adéquat, ce modèle permet d'obtenir des résultats invariants. Or, bien que le modèle ait été introduit, les critères de référence qui seront utilisés n'ont pas été présentés. Par conséquent, la présente section exposera tous les éléments nécessaires afin d'assurer que la démarche d'analyse utilisée dans cette recherche soit rigoureuse. Ensuite, les résultats de l'analyse des données seront présentés.

### **6.1 Démarche d'analyse**

Afin d'établir la démarche d'analyse qui sera utilisée dans cette recherche, les différents indices statistiques auxquels les chercheurs ont accès lorsqu'ils travaillent avec les modèles de la famille des modèles de Rasch seront présentés. Ensuite, les critères d'analyse généralement utilisés pour juger ces indices seront étudiés. Enfin, la stratégie d'analyse retenue pour cette recherche sera exposée.

#### **6.1.1 Indices statistiques**

La section 3.3.2 a permis de montrer que le modèle *Rating Scale* de la famille des modèles de Rasch se fonde sur deux hypothèses et une propriété de base. Ainsi,

l'unidimensionnalité et l'indépendance locale (hypothèses fondamentales de ces modèles), de même que l'invariance (propriété), se présentent généralement comme des conditions de base à remplir afin de pouvoir utiliser ces modèles. Cependant, puisque le modèle choisi peut très bien ne pas être le bon, le niveau d'applicabilité de ces conditions dépend du niveau d'adéquation entre les données et le modèle (sect. 3.3.3).

Par conséquent, la première étape d'une démarche d'analyse utilisant les modèles de la famille des modèles de Rasch consiste à vérifier la qualité de l'ajustement entre les données et le modèle. L'analyse de la qualité de l'ajustement entre les données et le modèle est une étape importante de la démarche d'analyse d'une recherche, car c'est elle qui permet au chercheur de poser un diagnostic sur la qualité de la prédiction que le modèle lui permettra de réaliser. Le chercheur peut alors identifier les caractéristiques des données qui ne s'ajustent pas bien au modèle (données mal adaptées) et examiner de quelle façon elles peuvent contrevenir à la mesure.

Or, le chercheur qui travaille avec la famille des modèles de Rasch a accès à différents indices statistiques afin d'éclairer l'analyse de la qualité de l'ajustement entre ses données et le modèle choisi. Les statistiques d'ajustement « *infit* » et « *outfit* », de même que celles pour les items et les personnes sont les indices les plus fréquemment utilisés. Cependant, plusieurs autres indices statistiques ont été développés afin d'aider le chercheur à mieux évaluer la qualité de l'ajustement entre les données et le modèle. Ces indices incluent entre autres l'indice de séparation (*separation index*), l'indice de « fidélité » (*reliability*) et le calibrage des points d'ancrage de l'échelle de réponses.

Les statistiques d'ajustement pour les items et pour les personnes sont donc les indices que les chercheurs utilisent le plus souvent. Chacune de ces statistiques (items et personnes), peut être calculée selon deux méthodes différentes (section 3.3.3.1). Ainsi, la qualité de l'ajustement pour les items (ou pour les personnes) peut être étudiée à partir des



indices *infit* et *outfit*. Puisque ces indices sont calculés à partir de la matrice des résidus centrés, réduits et mis au carré, ils peuvent être interprétés comme des statistiques ayant une distribution khi-carré (sect. 3.3.3.1, p. 56). Cependant, Smith (2004) indique que, puisque pour deux degrés de liberté différents, la valeur de la région critique associée sera également différente, il devient impossible de déterminer une valeur unique comme point de référence afin de juger de la qualité de l'ajustement entre les données et le modèle.

Par conséquent, Smith (2004) précise qu'afin de faciliter l'interprétation de ces statistiques d'ajustement, plusieurs transformations ont été développées. La première transformation qui a été appliquée à ces indices (*infit* et *outfit*) consiste à diviser ces indices par leur nombre de degrés de liberté. Les statistiques khi-carré sont donc transformées en carré moyen (« *mean square* » ou *mnsq*). La valeur attendue d'un carré moyen est de 1, son étendue est de 0 à l'infini. Cette statistique n'est pas donc pas symétrique par rapport à la valeur attendue.

Par conséquent, si la référence choisie pour juger de la qualité de l'ajustement est symétrique, le taux d'erreur de Type 1<sup>3</sup> sera différent pour chacune des queues de la distribution. De plus, Smith et Suh (2003) ont démontré que certains problèmes d'ajustement importants n'étaient pas détectés par les statistiques des carrés moyens. En effet, les deux chercheurs ont montré que les statistiques des carrés moyens ne décelaient pas du tout le manque d'invariance dans les estimations des paramètres. Or, la propriété d'invariance est une des propriétés les plus importantes pour les modèles de la famille des modèles de Rasch lorsque les données s'ajustent bien au modèle.

Ainsi, Smith (2004) indique que la transformation la plus courante consiste à transformer le carré moyen en racine cubique. Cette transformation permet de convertir le

---

<sup>3</sup> Dans un test statistique, une erreur de Type 1 consiste à rejeter l'hypothèse nulle du test (et donc à conserver l'hypothèse alternative) alors que l'hypothèse nulle est vraie.

carré moyen en une statistique qui s'apparente à la statistique t de Student<sup>4</sup>. Cette nouvelle statistique est communément appelée l'indice d'ajustement standardisé (« *standardized fit index* » ou *zstd* par *Winsteps*). Grâce à cette statistique, Smith soutient qu'il est possible de développer une valeur de référence possédant un taux d'erreur de Type 1 similaire pour chacune des queues de la distribution, de même que pour plusieurs conditions différentes.

Ainsi, l'indice d'ajustement standardisé posséderait un taux d'erreur de Type 1 assez stable même lorsque la taille de l'échantillon varie. Toutefois, comme pour la plupart des indices statistiques, son interprétation peut varier selon la taille de l'échantillon. Cependant, l'étude de Smith, Schumacker et Bush (1998) montre que la distribution de l'indice d'ajustement standardisé possède des propriétés qui le rendent plus stable que le carré moyen par rapport à des tailles d'échantillon différentes. Par conséquent, l'indice d'ajustement standardisé constitue un meilleur choix pour vérifier la qualité de l'ajustement entre les données et le modèle.

Or, une fois que l'indice statistique à utiliser a été déterminé, le chercheur a le choix de débiter son analyse par l'étude de la qualité de l'ajustement des items ou par l'étude de la qualité de l'ajustement des personnes. Afin de déterminer quelle stratégie les chercheurs devraient favoriser, Tejada, Gómez, García et Meléndez (2002) ont étudié deux stratégies d'analyse de la qualité de l'ajustement dans lesquelles l'ordre d'analyse de ces indices est structuré de façon différente.

Dans la première stratégie d'analyse, la qualité de l'ajustement des items est étudiée en premier. Les items pour lesquels la qualité de l'ajustement est mauvaise sont alors éliminés. Cette démarche d'analyse de la qualité de l'ajustement des items est ensuite reprise jusqu'à ce que tous les items conservés possèdent une bonne qualité

---

<sup>4</sup> Les détails de cette transformation sont donnés par Smith (1991).

d'ajustement. La deuxième étape de cette stratégie consiste alors à étudier la qualité de l'ajustement des personnes en suivant une démarche d'analyse similaire à celle utilisée pour l'ajustement des items. Finalement, c'est la qualité de l'ajustement entre les données et le modèle en général qui est étudiée. Cette stratégie est appelée « *TPI* » (*Total-Persons-Items*).

Dans la deuxième stratégie d'analyse, « *TIP* » (*Total-Items-Persons*), la démarche d'analyse est similaire à la première. Cependant, l'ordre d'analyse est différent. En effet, dans ce cas, c'est la qualité de l'ajustement des personnes qui est étudiée en premier, suivi de la qualité de l'ajustement des items et finalement, du modèle en général.

Dans leur étude, Tejada et al. (2002) ont constaté que la stratégie d'analyse *TPI* permettait de maximiser le nombre de sujets conservés, c'est-à-dire qui possèdent une bonne qualité d'ajustement avec le modèle. À l'opposé, ils ont trouvé que la stratégie *TIP* permettait de maximiser le nombre d'items conservés parce qu'ils possèdent une bonne qualité d'ajustement avec le modèle. Ainsi, selon que le chercheur souhaite maximiser le nombre d'items ou le nombre de personnes dans son étude, il adoptera l'une ou l'autre stratégie d'analyse de la qualité d'ajustement entre les données et le modèle.

Une fois cette première étape terminée, l'analyse des données peut se poursuivre et les différents indices statistiques développés viendront fournir au chercheur de nouveaux éléments à étudier afin de compléter son diagnostic sur la qualité de la prédiction que le modèle lui permettra de réaliser. En effet, puisque les données s'ajustent maintenant bien au modèle, le chercheur peut obtenir la valeur modélisée (estimée par le modèle) de l'habileté des personnes (ou du degré d'adhésion) et du niveau de difficulté des items (difficulté à endosser l'item). Le modèle utilisé fournit également une estimation de l'erreur type sur chacun de ces estimés. Cette estimation de l'erreur type est utile pour le chercheur, car elle permet d'évaluer la précision de chacune des mesures fournies

(Smith Jr., 2004). De plus, l'erreur type peut être utilisée afin de décrire une étendue (ou un intervalle de confiance) à l'intérieur de laquelle la « vraie » habileté des personnes (opinion) ou le « vrai » niveau de difficulté (difficulté à endosser) se trouve (Smith Jr., 2004).

Par ailleurs, Smith Jr. précise que si le chercheur souhaite évaluer la « fidélité » de l'estimation de l'habileté (opinion) d'un ensemble de personnes (par exemple), il est possible de mettre l'erreur type de chacun des estimés de cet ensemble au carré, de les additionner et de diviser cette somme par le nombre d'éléments de l'ensemble. Cet indice fournit une erreur type moyenne (« *mean square measurement error* » ou mse) plus « fidèle » que lorsqu'un chercheur choisit d'utiliser l'erreur type d'une personne moyenne.

L'indice de « fidélité » (*reliability*) de la mesure d'un groupe est alors calculé de la façon suivante. Tout d'abord, le modèle calcule l'erreur type moyenne (mse). Ensuite, cette valeur est soustraite de la variance réelle (observée) du groupe de personnes. L'indice de « fidélité » consiste donc en une variance ajustée à l'erreur de mesure et représente la « vraie » variance des mesures du groupe de personnes. Par conséquent, l'indice de « fidélité » permet au chercheur de déterminer jusqu'à quel point les estimations produites par le modèle sont influencées ou non par des erreurs de mesure.

À partir de cet indice statistique, il est alors possible de calculer l'indice de séparation. En effet, ce dernier est calculé comme le rapport entre la variance ajustée (« vraie ») et la variance observée. Ainsi, l'indice de séparation représente la proportion de la variance qui n'est pas due à l'erreur de mesure. Cet indice statistique permet au chercheur d'évaluer la répartition des personnes (séparation) sur le continuum de l'habileté (degré d'adhésion). De la même façon, cet indice permet d'évaluer la répartition des items sur le continuum du niveau de difficulté (difficulté à endosser).

L'indice de séparation est donc souvent qualifié d'homologue chez les modèles de la famille des modèles de Rasch à l'indice de cohérence interne KR-20 plus connu. Andrich (1982), montre que l'indice de séparation a effectivement la même structure de calcul que l'indice KR-20. De plus, Andrich montre que la valeur obtenue par l'un ou l'autre de ces indices est assez similaires, mais que l'indice KR-20 s'exprime en fonction de la variance des personnes et de celle des items alors que l'indice de séparation s'exprime entièrement en fonction des paramètres associés aux personnes.

Selon Andrich (1982), l'idée générale à la base de l'indice de cohérence interne est que les items d'un test devraient tous mesurer la même chose. Or, pour qu'il y ait cohérence interne, Andrich précise qu'il n'est pas nécessaire que tous les items mesurent exactement la même chose; cependant, toutes les dimensions mesurées par ces items doivent être représentées dans les mêmes proportions pour chacun de ces items. L'idée de base de la cohérence interne est très similaire à celle présentée à la section 3.3.2.1 à propos de l'unidimensionnalité que l'on retrouve chez certains modèles de la famille des modèles de Rasch.

Par conséquent, l'indice de séparation permet au chercheur de confirmer que l'hypothèse d'unidimensionnalité est vérifiée. En effet, si l'indice de séparation est élevé, c'est-à-dire si la répartition des items par rapport à leur niveau de difficulté est grande, alors les réponses des participants seraient influencées par une seule et même dimension dominante, c'est-à-dire la dimension mesurée par le test. L'hypothèse de l'unidimensionnalité est alors confirmée et, par suite, celle de l'indépendance locale également (sect. 3.3.2.2).

La séparation des items (ou l'hypothèse de l'unidimensionnalité) peut aussi être vérifiée graphiquement. En effet, si, en aucun point sur le continuum du niveau de difficulté des items, on ne retrouve deux items très rapprochés, cela confirme que les réponses des

participants seraient influencées par la dimension mesurée par le test seulement et non pas par d'autres facteurs inhérents ou externes au test (Andrich, 1982).

Or, l'analyse graphique ne permet pas seulement de vérifier la répartition des items sur le continuum du niveau de difficulté (difficulté à endosser) ou des personnes sur le continuum de l'habileté (opinion), il permet également d'étudier visuellement les catégories de réponse offertes par le test. En effet, l'analyse graphique des catégories de réponse permet de voir, en un seul coup d'œil, la structure des catégories de même que si toutes les catégories offertes ont été utilisées (Park, 2004).

Park (2004) soutient que l'intérêt principal de l'analyse graphique des catégories de réponse consiste à vérifier si la courbe de probabilités de chacune des catégories de réponse possède un sommet distinct et si les courbes de probabilités de l'ensemble des catégories apparaissent comme une suite de collines également distantes. Si tel est le cas, alors pour les participants appartenant à une certaine portion du continuum de l'habileté (degré d'adhésion), une certaine catégorie de réponse parmi celles offertes aura plus de chance que les autres d'être choisie. En revanche, si la courbe de probabilités d'une des catégories de réponse ne possédait pas un sommet distinct qui s'élève au-dessus des courbes de probabilités des catégories adjacentes, cela indiquerait que cette catégorie n'a jamais plus de chance qu'une autre d'être choisie et ce, sur tout le continuum de l'habileté (degré d'adhésion) des personnes (Park, 2004).

Afin d'étudier les catégories de réponse offertes, le chercheur peut également étudier le calibrage des points d'ancrage de l'échelle de réponses. En effet, les modèles de la famille de modèles de Rasch tels que le modèle *Rating Scale* ou le modèle *Partial Credit* fournissent une estimation du seuil entre les catégories de réponse (sect. 3.3.1.1 et 3.3.1.2). Ce paramètre,  $F_x$ , correspond au point où la probabilité d'opter pour l'une ou l'autre des catégories est égale. Ce paramètre indique le degré de difficulté rencontré afin

d'observer une réponse dans la catégorie  $x$  plutôt que dans la catégorie  $x-1$ . Il n'indique pas le degré de difficulté nécessaire pour faire le choix d'une catégorie par rapport à l'autre (Linacre, 2006).

Selon cette définition, la valeur des points d'ancrage de l'échelle de réponses estimé par le modèle devrait augmenter en même temps que la valeur de chacune des catégories. Si tel n'est pas le cas, cela indique que la catégorie de réponse où l'inversion se produit est peu observée, c'est-à-dire qu'elle occupe un intervalle très mince sur le continuum de l'habileté (degré d'adhésion) (Linacre, 2006). De plus, une telle inversion peut également révéler la présence de problèmes dans la définition des catégories de l'échelle de réponses (Linacre, 2006).

### **6.1.2 Critères d'analyse**

La section précédente a permis de présenter les différents indices statistiques à partir desquels les chercheurs qui travaillent avec les modèles de la famille des modèles de Rasch peuvent analyser la qualité de l'ajustement entre les données et le modèle. Cependant, les indices présentés ne suffisent pas pour permettre au chercheur de poser un diagnostic. En effet, afin de déterminer si, par exemple, les données s'ajustent bien ou non au modèle, le chercheur doit travailler avec des critères (indicateurs). Ces critères permettent d'établir une norme ou un point de référence à partir desquels poser un diagnostic.

Afin de guider les chercheurs dans le choix d'une norme de référence, Linacre (2004) a étudié les divers indices statistiques de même que les différents critères généralement utilisés par les chercheurs. Cette étude l'a amené à présenter une série

d'éléments permettant au chercheur d'examiner ses données et d'améliorer l'efficacité de son diagnostic. Les éléments qu'il a présentés sont repris ici afin de guider l'analyse des données de cette recherche.

Avant même de débiter l'analyse des données, Linacre suggère de s'assurer que tous les items servent à mesurer une même dimension. Les items d'un questionnaire peuvent être formulés positivement ou négativement, utiliser la même échelle de réponses ou des échelles différentes, etc. Il est donc essentiel de faire un retour sur les items du questionnaire afin de les regrouper par blocs d'items ou d'inverser le sens (positif ou négatif) de certains items afin de s'assurer que les items qui seront analysés mesurent une même dimension et sont en consensus sur la façon de la mesurer.

Ensuite, Linacre propose de vérifier que chacune des catégories de réponse possède au moins dix observations. En effet, Linacre explique que le calibrage de l'échelle de réponses est estimé à partir du ratio logarithmique des catégories adjacentes. Ainsi, lorsqu'une des fréquences est basse, le calibrage est imprécis et peut être instable.

En troisième lieu, il est important de vérifier que les observations sont distribuées de façon uniforme dans les différentes catégories de réponse. Linacre précise que lorsque les observations sont distribuées de façon irrégulière, cela peut dénoter une utilisation aberrante de l'échelle de réponses offerte. À l'opposé, une distribution uniforme permet un calibrage optimal de l'échelle de réponses.

Une fois la distribution des observations vérifiée, Linacre suggère d'analyser la mesure moyenne estimée par le modèle afin d'observer les réponses de chacune des catégories de l'échelle. De façon générale, Linacre soutient qu'une réponse appartenant aux catégories les plus « grandes » de l'échelle devrait provenir de mesures plus grandes. Par conséquent, la mesure moyenne estimée pour chacune des catégories devrait croître



de façon monotone avec l'augmentation de la valeur de chacune des catégories de l'échelle de réponses.

Linacre propose ensuite de vérifier les carrés moyens « *outfit* ». La section précédente a permis de préciser que la valeur attendue d'un carré moyen est de 1, ce qui correspond à un niveau raisonnable de hasard attendu dans un modèle stochastique comme le sont les modèles de la famille des modèles de Rasch. Lorsque le hasard n'est pas suffisamment présent dans les données, ces dernières sont alors trop prévisibles. Dans ce cas, Linacre explique que l'étendue des mesures produites est élargie par le modèle et elles semblent alors plus différentes les unes des autres. À l'inverse, lorsque le hasard est trop présent dans les données, l'étendue des mesures produites par le modèle diminue et celles-ci apparaissent alors comme étant plus similaires. Linacre soutient que c'est cette dernière situation qui a l'effet immédiat le plus important sur la mesure.

La section précédente a présenté la statistique des carrés moyens « *outfit* » comme n'étant pas symétrique par rapport à la moyenne. Or, les références utilisées pour juger de la qualité de l'ajustement par rapport à cette statistique sont généralement symétriques. Par exemple, certains chercheurs suggèrent que des données possédant des carrés moyens « *outfit* » entre 0,8 et 1,2 s'ajustent bien au modèle alors que des données à l'extérieur de cet intervalle ne s'ajustent pas bien. Pour d'autres, l'intervalle est de 0,6 à 1,4.

Pour Linacre, lorsque les données possèdent des carrés moyens « *outfit* » excédant 2, c'est-à-dire lorsque les données possèdent un surplus de hasard, cela suggère que certaines données contreviennent à la mesure. Enfin, certains chercheurs préfèrent travailler avec l'indice standardisé afin d'analyser la qualité de l'ajustement entre leurs données et le modèle plutôt que de travailler avec les carrés moyens. En effet, la section 6.1.1 a permis de préciser que l'indice d'ajustement standardisé possède des

propriétés qui le rendent plus stable que le carré moyen. De plus, l'indice d'ajustement standardisé permet d'utiliser un critère symétrique pour chacune des queues de la distribution. Ainsi, l'indice d'ajustement standardisé devrait appartenir à l'intervalle de -2 à 2 afin d'assurer une bonne qualité d'ajustement entre les données et le modèle (Bond & Fox, 2001; Bradley & Sampson, 2005).

L'analyse de la qualité de l'ajustement entre les données et le modèle permet donc de détecter les données susceptibles de contrevenir à la mesure. Lorsque les données se trouvent à l'extérieur de l'intervalle déterminé, il suffit bien souvent de retirer ces observations pour que le reste des données s'ajuste bien au modèle, mais il faut prendre le temps d'analyser les observations problématiques (Linacre, 2004).

Or, afin de détecter les données susceptibles de contrevenir à la mesure, il est également possible d'étudier le degré de corrélation entre la mesure estimée par le modèle pour un individu (ou un item) et la réponse observée pour cet individu (ou cet item). En effet, si la mesure estimée par le modèle est cohérente avec la donnée observée, la corrélation sera positive. À l'inverse, lorsque la corrélation est négative ou près de zéro, c'est qu'il y a une rupture entre la mesure estimée par le modèle et la donnée observée (Smith, 1996). Une telle rupture peut provenir d'une inversion de l'échelle de réponses, par exemple, ou d'une mauvaise compréhension de l'item correspondant. Il est donc important de vérifier que la corrélation entre la mesure et l'observation est positive (Bradley & Sampson, 2005). Non seulement cette vérification permet de détecter des observations problématiques, mais cela permet également d'appuyer l'hypothèse d'unidimensionnalité.

Par ailleurs, Linacre propose de vérifier les points d'ancrage de l'échelle de réponses. En effet, comme cela a été présenté à la section précédente, Linacre soutient que la valeur des points d'ancrage de l'échelle de réponses estimé par le modèle devrait augmenter de façon monotone en fonction de la valeur croissante de chacune des

catégories. De façon plus précise, Linacre suggère que les points d'ancrage de l'échelle de réponses devraient augmenter d'au moins 1 logit<sup>5</sup> pour une échelle comprenant cinq catégories de réponse différentes. En effet, lorsque la différence entre deux points d'ancrage est moindre que 1 logit, il est probable que la définition de ces deux catégories ne soit pas assez claire pour que les participants puissent bien les distinguer. Il vaudrait peut-être mieux les combiner.

En revanche, Linacre a également recommandé que la différence entre deux points d'ancrage ne soit pas être supérieure à 5 logits. En effet, lorsque la distance entre deux catégories est trop grande, s'installe alors un genre de « point mort » entre ces catégories et la mesure de la performance perd de sa précision. Dans ce cas, il est donc préférable d'insérer une catégorie de réponse supplémentaire entre ces deux catégories pour combler ce vide.

Les modèles unidimensionnels de la famille des modèles de Rasch, tels que le modèle *Rating Scale (RS)* ou le modèle *Partial Credit (PC)*, permettent d'estimer l'habileté des participants (degré d'adhésion) et le degré de difficulté des items (difficulté à endosser l'item) à partir de données recueillies à l'aide d'une échelle de réponses en catégories ordonnées. Or, il arrive bien souvent qu'un chercheur doive prendre une décision et agir en fonction des résultats trouvés. Dans une perspective clinique, ces décisions sont souvent basées sur une seule observation. Linacre soutient donc que, de façon générale, il est important que cette seule observation produise une mesure modélisée équivalente. Inversement, il est tout aussi important que la mesure modélisée permette d'inférer le comportement d'une personne, c'est-à-dire qu'elle permette de prédire quelle catégorie de réponse devrait être observée pour cette personne en réponse à un item en particulier.

---

<sup>5</sup> « Logit » est l'expression courte utilisée pour représenter une unité du ratio logarithmique des probabilités utilisé dans les modèles Rasch (*log-odds probability unit*).

Pour les modèles unidimensionnels de la famille des modèles de Rasch tels que le *RS* et le *PC*, c'est la CCI (courbe caractéristique de l'item – sect. 3.3) qui dépeint cette relation entre la mesure et la catégorie de réponse attendue. Afin de vérifier le degré de cohérence empirique entre les mesures modélisées et les observations appartenant réellement à une catégorie donnée, Linacre suggère de vérifier les deux statistiques suivantes : la cohérence mesure-catégorie et la cohérence catégorie-mesure.

Dans le premier cas, la statistique montre la proportion, en pourcentage, entre le nombre de réponses attendues dans une catégorie (mesure) par rapport au nombre de réponses réellement observées dans cette catégorie. Dans le deuxième cas, la statistique montre quel pourcentage des observations trouvées dans une catégorie, ont effectivement été produites par des mesures modélisées appartenant à cette catégorie. Afin de faciliter l'interprétation de ces statistiques, Linacre donne deux exemples. Tout d'abord, il indique qu'un résultat de 63 % pour la statistique mesure-catégorie pour la catégorie 0 (par exemple) permet de conclure que parmi toutes les mesures que le modèle estime appartenir à cette catégorie, 63 % y ont effectivement été retrouvées. Une inférence mesure-catégorie sera donc généralement bonne. Par ailleurs, un résultat de 42 % pour la statistique catégorie-mesure pour cette même catégorie 0 (par exemple), signifie que 42 % des observations trouvée dans cette catégorie ont été produites par des mesures appartenant à cette catégorie. L'inférence catégorie-mesure est donc plus fragile, quoique tout de même acceptable.

Enfin, pour mieux compléter les éléments présentés par Linacre, il est important de déterminer selon quels critères l'indice de séparation et l'indice de fidélité sont considérés comme étant acceptables ou problématiques. Or, avant de déterminer ces critères, il faut noter que l'indice de séparation permet d'évaluer le nombre de niveaux de performance

statistiquement différents pouvant être identifiés par le questionnaire. Ce nombre est donné par la formule suivante<sup>6</sup> :

$$H_i = (4 \times G_i + 1) / 3 \quad (15)$$

où  $G_i$  correspond à l'indice de séparation et  $H_i$  au nombre de niveaux de performance statistiquement différents pouvant être identifiés par le questionnaire.

Ainsi, lorsqu'un chercheur souhaite que le questionnaire permette d'identifier au moins deux niveaux de performance différents, l'indice de séparation doit être plus grand que 1,5. L'indice de « fidélité » correspondant est alors de 0,7, ce qui signifie qu'environ 30 % de la variance est due à des erreurs de mesure (Fisher Jr., 1992). Lorsque l'indice de séparation est plus grand que 2, l'indice de « fidélité » correspondant est alors d'au moins 0,8, ce qui signifie que moins de 20 % de la variance est due à des erreurs de mesure. Selon Jackson et Popovich (2006), un indice de séparation plus grand que 2 et un indice de fidélité plus grand que 0,8 constituent une norme acceptable.

### 6.1.3 Stratégie d'analyse choisie

La première vérification suggérée par Linacre (2004) consiste à s'assurer que tous les items étudiés servent à mesurer une même dimension et qu'ils présentent une échelle de réponses orientée dans le même sens. Or, tous les items des questionnaires utilisés dans cette recherche ont été formulés de façon positive et l'échelle de réponses offerte pour chacun d'eux est orienté dans le même sens. Par conséquent, toutes les données

---

<sup>6</sup> Les détails de cette formule sont donnés par Wright et Masters (1982) et Fisher Jr. (1992).

recueillies pour cette étude répondent à ce critère. Cette vérification ne sera donc pas reprise dans l'analyse des données.

Linacre propose ensuite de vérifier que chacune des catégories de réponse possède au moins 10 observations. De plus, Linacre suggère de vérifier que les observations recueillies sont distribuées de façon uniforme dans les différentes catégories de réponse. Les données recueillies dans cette étude ne répondent pas à ces deux critères. En effet, le nombre de sujets ayant répondu aux différentes versions des questionnaires est limité. Il ne permet donc pas d'assurer un minimum d'observations, ni d'obtenir une distribution uniforme dans chacune des catégories offertes. Par conséquent, ces deux critères ne seront pas retenus pour l'analyse des données.

La première étape de la démarche d'analyse qui sera utilisée dans cette recherche consistera donc à vérifier la qualité de l'ajustement entre les données et le modèle. En effet, la section 6.1.1 a montré qu'il s'agit d'une étape importante de la démarche d'analyse, car elle permet de vérifier le niveau d'applicabilité des conditions de base du modèle. De plus, elle permet au chercheur d'identifier les données qui pourraient contrevenir à la mesure.

Dans cette étude, une seule section des questionnaires distribués a été retenue. Cette section, portant sur la préparation des étudiants à l'enseignement, est constituée de seulement 20 items. Puisque l'objectif de cette recherche consiste à étudier jusqu'à quel point des changements de mots apportés aux items d'un questionnaire peuvent influencer les réponses des participants, de même que l'échelle de mesure associée aux catégories de réponse de l'échelle offerte, il est important de conserver le maximum d'items possible. Par conséquent, la qualité de l'ajustement des personnes sera étudiée en premier, suivie de la qualité de l'ajustement des items (sect. 6.1.1).

Or, étant donné la taille relativement petite des échantillons recueillis, il est également important d'essayer de conserver le maximum de participants possible. Ainsi, la qualité de l'ajustement sera analysée à l'aide de l'indice d'ajustement standardisé. En effet, puisque cet indice se montre plus stable que le carré moyen par rapport à des tailles d'échantillon différentes, il constitue un meilleur choix pour cette recherche.

Smith (1991) a constaté que l'intervalle de -2 à +2 pour l'indice d'ajustement standardisé n'était pas toujours utile pour analyser la qualité de l'ajustement d'un test de 20 items distribué sur un large échantillon. Smith en a conclu que sous de telles conditions, un intervalle de -3 à +2 pouvait s'avérer plus utile. Selon Lawton, Bhakta, Chamberlain et Tennant (2004), l'indice d'ajustement standardisé devrait appartenir à l'intervalle de -2,5 à 2,5 afin d'assurer une bonne qualité d'ajustement entre les données et le modèle. Ainsi, afin de travailler avec un intervalle utile, tout en diminuant le nombre d'items ou de sujets devant être considérés comme ne s'ajustant pas bien au modèle, l'intervalle de -2,5 à +2,5 sera utilisé pour l'analyse de la qualité de l'ajustement de cette recherche plutôt que l'intervalle de -2 à +2 utilisé par plusieurs chercheurs. Enfin, la corrélation entre la mesure et l'observation sera vérifiée afin de s'assurer qu'elle est positive pour tous les sujets et pour tous les items.

Ensuite, afin de déterminer jusqu'à quel point la valeur estimée par le modèle pour l'habileté des personnes et pour le niveau de difficulté des items sont influencées par des erreurs de mesure, l'indice de « fidélité » sera étudiée. Dans cette étude, un indice de « fidélité » d'au moins 0,8 sera considéré acceptable. En effet, cela permet d'assurer que moins de 20 % de la variance est due à des erreurs de mesure. De plus, l'indice de séparation sera également vérifié afin de s'assurer qu'il est d'au moins 2. Cela permet de confirmer que le questionnaire étudié permet d'identifier au moins deux niveaux de performance différents.

Linacre propose également de vérifier les points d'ancrage de l'échelle de réponses. Tout d'abord, la valeur des points d'ancrage de l'échelle de réponses estimée par le modèle doit augmenter de façon monotone en fonction de la valeur croissante de chacune des catégories. Ensuite, cette augmentation doit être d'au moins 1 logit, mais de moins de 5 logits. Ces critères d'analyse seront donc utilisés dans cette recherche afin d'étudier si les changements de mots apportés aux items des questionnaires peuvent influencer les réponses des participants, de même que l'échelle de mesure associée aux catégories de réponse de l'échelle offerte.

Enfin, Linacre soutient qu'il est important de vérifier la relation entre la mesure et la catégorie de réponse attendue. Or, puisque aucune inférence ne sera faite à partir des données recueillies dans cette étude, ni la cohérence mesure-catégorie ni la cohérence catégorie-mesure ne seront analysées.

## **6.2 Analyses des données**

Maintenant que la démarche d'analyse est bien définie, il est possible de poursuivre l'exploration proposée dans cette recherche en présentant les résultats obtenus lors de l'analyse des données. Ces résultats seront présentés selon trois sections, c'est-à-dire une pour chacune des versions du questionnaire (années 2003, 2004 et 2005).

De plus, pour chacune de ces sections, l'analyse sera divisée selon deux étapes : la première étape permettra d'étudier la qualité de l'ajustement entre les données et le modèle. La seconde établira une comparaison entre les versions A et B des questionnaires afin d'étudier l'influence de la formulation des items sur les réponses des participants de même que sur l'échelle de mesure associée. Il est à noter que les résultats présentés dans



cette analyse seront repris au début du prochain chapitre afin de faciliter l'interprétation des résultats qui y sera proposée.

Puisque l'objectif de cette recherche est d'explorer l'influence de la formulation des items sur les réponses des participants à partir des 20 items provenant du questionnaire utilisé par le CFIM pour l'évaluation de ses programmes, tous les items seront conservés pour effectuer ces analyses. Toutefois, afin de compléter l'exploration de cette recherche, une dernière section proposera quelques résultats d'analyses complémentaires effectuées une fois que les items mal ajustés auront été retirés de l'échantillon.

## **6.2.1 Questionnaires 2003**

La cueillette de données de 2003 a permis de récolter les réponses de 124 étudiants et étudiantes provenant du programme d'enseignement en éducation préscolaire et enseignement primaire. Parmi ces 124 étudiants, 55 ont répondu à la version A du questionnaire et 69 ont répondu à la version B.

### ***6.2.1.1 Qualité de l'ajustement pour les questionnaires 2003***

La première étape de la démarche d'analyse choisie consiste à vérifier la qualité de l'ajustement des personnes. Ainsi, suite à une première analyse, 4 personnes ayant répondu à la version A du questionnaire ne s'ajustaient pas bien au modèle (sujets 778, 799, 829 et 877), de même que 11 personnes ayant répondu à la version B (sujets 880,

882, 883, 956, 958, 1086, 1092, 1141, 1161, 1163 et 1168)<sup>7</sup>. De plus, la corrélation était négative pour deux des sujets ayant répondu à la version B du questionnaire (sujets 971 et 1163). Une à une, les personnes possédant les pires indices d'ajustement standardisés (« *infit* » ou « *outfit* ») ou une corrélation négative ont été retirés de l'échantillon. La qualité de l'ajustement était ensuite étudiée de nouveau.

Dans la version A du questionnaire, les 4 personnes pour lesquelles l'ajustement était mauvais ont dû être retirées. Dans la version B du questionnaire, 7 personnes parmi les 11 ne s'ajustant pas bien au modèle ont dû être retirées (sujets 883, 971, 1086, 1092, 1141, 1157 et 1163). À la suite de quoi, la qualité de l'ajustement des personnes répondait au critère choisi. La corrélation entre la mesure estimée par le modèle pour chacune des personnes et les observations recueillies était positive pour toutes les personnes retenues.

Ainsi, 51 sujets ont été conservés dans l'échantillon pour la version A du questionnaire et 62 pour la version B. Le tableau 3 de la page suivante présente les statistiques d'ajustement des personnes résultant de l'analyse pour la version A du questionnaire. (Afin de ne pas alourdir le texte, les tableaux de la version B du questionnaire sont présentés dans l'annexe C.1 de ce travail.)

---

<sup>7</sup> Afin de ne pas alourdir inutilement la présentation des résultats, les tableaux de cette première analyse sont présentés en annexe.

**Tableau 3 : Statistiques d'ajustement finales des personnes pour la version A du questionnaire 2003**

Sujet	Mesure estimée <sup>8</sup>	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
610	-0,83	0,26	0,40	-2,5	0,41	-2,4	0,64
611	-0,56	0,26	1,15	0,6	1,17	0,7	0,18
612	-1,86	0,28	1,03	0,2	1,05	0,3	0,43
613	-0,35	0,26	1,17	0,6	1,15	0,6	0,77
614	1,21	0,28	1,35	1,1	1,25	0,9	0,49
616	0,83	0,27	0,45	-2,2	0,43	-2,3	0,78
617	1,06	0,28	0,74	-0,8	0,74	-0,8	0,80
618	-0,35	0,26	0,66	-1,2	0,66	-1,2	0,57
619	-0,15	0,26	1,23	0,8	1,19	0,7	0,65
621	0,76	0,27	0,59	-1,5	0,59	-1,5	0,70
622	1,80	0,30	1,14	0,5	1,23	0,8	0,21
623	1,29	0,28	0,81	-0,5	0,88	-0,3	0,64
624	0,19	0,26	1,18	0,7	1,13	0,5	0,59
647	1,13	0,28	0,57	-1,5	0,56	-1,6	0,50
648	0,26	0,26	0,76	-0,8	0,74	-0,8	0,63
649	0,91	0,27	1,09	0,4	1,10	0,4	0,61
650	0,83	0,27	1,67	1,9	1,64	1,9	0,87
651	0,19	0,26	0,68	-1,1	0,66	-1,2	0,71
652	1,45	0,29	0,68	-1,0	0,72	-0,9	0,82
774	0,47	0,27	1,00	0,1	0,99	0,1	0,60
775	0,26	0,26	0,49	-2,0	0,48	-2,0	0,74
776	0,83	0,27	0,58	-1,5	0,59	-1,5	0,72
777	-0,83	0,26	0,70	-1,0	0,76	-0,8	0,54
779	-0,15	0,26	1,86	2,4	1,87	2,4	0,63
780	0,91	0,27	0,74	-0,8	0,79	-0,6	0,75
781	1,54	0,29	0,70	-1,0	0,70	-1,0	0,84
782	0,54	0,27	1,73	2,1	1,68	2,0	0,69
783	-0,76	0,26	1,78	2,2	1,79	2,2	0,64
784	1,13	0,28	0,66	-1,1	0,69	-1,0	0,54
790	-0,64	0,27	1,21	0,7	1,24	0,8	0,22
791	-0,42	0,26	1,58	1,7	1,52	1,6	0,66
792	-0,90	0,26	1,29	1,0	1,35	1,2	0,21
793	0,61	0,27	0,96	0,0	0,90	-0,2	0,54
794	0,91	0,27	0,81	-0,6	0,80	-0,6	0,45

<sup>8</sup> (opinion)

Suite du tableau 3

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
795	-1,63	0,28	1,01	0,1	0,99	0,1	0,67
796	-0,15	0,26	1,71	2,1	1,68	2,0	0,79
797	1,45	0,29	0,74	-0,8	0,74	-0,8	0,85
798	0,19	0,26	0,71	-1,0	0,69	-1,0	0,67
800	1,45	0,29	0,84	-0,4	0,82	-0,5	0,47
801	-0,08	0,26	0,55	-1,7	0,55	-1,7	0,69
802	0,98	0,28	1,37	1,2	1,39	1,2	0,59
803	-1,11	0,27	0,54	-1,7	0,54	-1,7	0,71
804	1,54	0,29	0,87	-0,3	0,79	-0,6	0,17
827	1,21	0,28	1,23	0,8	1,32	1,0	0,70
828	2,08	0,31	1,20	0,7	1,16	0,6	0,79
830	1,62	0,29	1,09	0,4	1,03	0,2	0,68
832	0,69	0,27	0,48	-2,0	0,46	-2,1	0,86
873	-0,42	0,26	0,64	-1,3	0,67	-1,1	0,34
874	0,06	0,26	1,70	2,0	1,68	2,0	0,74
775	-0,35	0,26	1,09	0,4	1,11	0,4	0,66
876	-0,01	0,26	1,35	1,2	1,32	1,1	0,60
<b>MOY.</b>	<b>0,37</b>	<b>0,27</b>	<b>0,99</b>	<b>-0,1</b>	<b>0,99</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>0,91</b>	<b>0,01</b>	<b>0,39</b>	<b>1,3</b>	<b>0,39</b>	<b>1,3</b>	

Une fois l'analyse de la qualité de l'ajustement des personnes terminée, l'analyse de la qualité de l'ajustement des items a été effectuée. Le tableau 4 présente les statistiques d'ajustement des items obtenues pour la version A du questionnaire. Tous les items des deux versions du questionnaire s'ajustaient bien au modèle (voir également tabl. 16, p. 184). De plus, la corrélation entre la mesure estimée par le modèle pour chacun des items et les observations recueillies était positive pour tous les items.

**Tableau 4 : Statistiques d'ajustement des items pour la version A du questionnaire 2003**

Item	Mesure estimée <sup>9</sup>	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1	0,44	0,16	0,83	-0,9	0,84	-0,8	0,57
2	2,00	0,17	0,81	-1,0	0,81	-1,0	0,58
3	1,19	0,16	1,49	2,3	1,48	2,3	0,57
4	-0,38	0,17	1,13	0,7	1,08	0,5	0,57
5	-0,52	0,17	1,25	1,3	1,22	1,1	0,56
6	-0,35	0,17	1,13	0,7	1,14	0,8	0,68
7	0,44	0,16	0,78	-1,2	0,78	-1,2	0,68
8	-2,01	0,20	0,92	-0,3	0,84	-0,8	0,68
9	-0,76	0,17	0,84	-0,8	0,83	-0,8	0,63
10	-1,34	0,19	0,86	-0,7	0,87	-0,6	0,73
11	-0,67	0,17	1,01	0,1	1,06	0,4	0,57
12	-0,95	0,18	0,88	-0,6	0,87	-0,6	0,67
13	-0,32	0,17	0,59	-2,4	0,59	-2,4	0,69
14	-0,27	0,17	1,27	1,3	1,26	1,3	0,56
15	0,31	0,17	1,19	1,0	1,19	1,0	0,61
16	0,42	0,16	0,75	-1,4	0,74	-1,5	0,67
17	-0,15	0,17	0,88	-0,6	0,88	-0,6	0,67
18	0,42	0,16	0,98	-0,1	0,98	0,0	0,56
19	1,05	0,16	1,05	0,3	1,09	0,5	0,48
20	1,46	0,17	1,21	1,1	1,23	1,2	0,50
<b>MOY.</b>	<b>0,00</b>	<b>0,17</b>	<b>0,99</b>	<b>-0,1</b>	<b>0,99</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>0,95</b>	<b>0,01</b>	<b>0,21</b>	<b>1,1</b>	<b>0,21</b>	<b>1,1</b>	

Ensuite, le tableau 5 présente les statistiques d'ajustement générales des personnes pour la version A du questionnaire. Le tableau montre que l'ajustement général des personnes est à l'intérieur de l'intervalle choisi et que la valeur moyenne estimée par le modèle pour le degré d'adhésion des personnes (opinion) est de 0,37 avec une erreur type de 0,27.

<sup>9</sup> (difficulté à endosser un item)

Le tableau 5 révèle également que l'indice de « fidélité » réel, c'est-à-dire l'indice pour lequel l'erreur type a été ajustée de façon à prendre en considération que les données mal ajustées proviennent d'un mauvais ajustement entre les données et le modèle, est de 0,9. L'indice de « fidélité » réel permet de rendre compte de la « fidélité » minimale qu'il est possible d'obtenir avec les données utilisées (Linacre, 2006). L'indice de séparation correspondant est de 2,93. L'indice de « fidélité » du modèle, c'est-à-dire l'indice calculé en considérant que les données s'ajustent bien au modèle et donc que les données mal ajustées proviennent de la nature stochastique du modèle, est de 0,91. L'indice de « fidélité » du modèle permet de rendre compte de la « fidélité » maximale qu'il est possible d'obtenir avec les données utilisées (Linacre, 2006). L'indice de séparation correspondant est de 3,17.

Ainsi, la version A du questionnaire répond au seuil d'acceptabilité choisi et permet de bien discriminer l'opinion des individus sur au moins quatre niveaux différents<sup>10</sup>. Autrement dit, la version A du questionnaire permet de faire la distinction entre quatre niveaux d'opinion différents par rapport aux six niveaux proposés dans l'échelle de réponses du questionnaire.

**Tableau 5 : Statistiques d'ajustement générales des personnes pour la version A du questionnaire 2003**

	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	0,37	0,27	0,99	-0,1	0,99	-0,1
<b>É.-T.</b>	0,91	0,01	0,39	1,3	0,39	1,3
<b>MAX.</b>	2,08	0,31	1,86	2,4	1,87	2,4
<b>MIN.</b>	-1,86	0,26	0,40	-2,5	0,41	-2,4

<b>Réel</b>	<b>Séparation</b>	2,93	<b>« Fidélité »</b>	0,90
<b>Modèle</b>	<b>Séparation</b>	3,17	<b>« Fidélité »</b>	0,91

<sup>10</sup> Selon la formule  $H_i = (4 \times G_i + 1) / 3$ .

Pour la version B du questionnaire, l'ajustement général des personnes est également à l'intérieur de l'intervalle choisi. La valeur moyenne estimée par le modèle pour le degré d'adhésion des personnes est de 0,4 avec une erreur type de 0,29 (tabl. 17, p. 185). Il est donc légèrement plus difficile pour les individus ayant répondu à la version B du questionnaire d'adhérer aux différents items proposés que pour les individus ayant répondu à la version A du questionnaire. L'indice de « fidélité » réel est de 0,94 et l'indice de séparation correspondant est de 3,85. L'indice de « fidélité » du modèle est de 0,95 et l'indice de séparation correspondant est de 4,15. Ainsi, la version B du questionnaire répond au seuil d'acceptabilité choisi et permet une meilleure discrimination de l'opinion des individus que la version A en discriminant leur opinion sur au moins cinq niveaux différents.

Le tableau 6 de la page suivante présente les statistiques d'ajustement générales des items pour la version A du questionnaire. L'ajustement général des items est à l'intérieur de l'intervalle choisi. La valeur moyenne du niveau de difficulté des items (difficulté à endosser) a été fixée à 0 pour les deux versions du questionnaire. C'est la valeur moyenne généralement utilisée afin de faciliter l'interprétation des résultats. L'indice de « fidélité » réel des items de la version A du questionnaire est de 0,96 et l'indice de séparation correspondant est de 5,25. L'indice de « fidélité » du modèle est de 0,97 et l'indice de séparation correspondant est de 5,46.

La version A du questionnaire répond donc au seuil d'acceptabilité choisi et permet de bien discriminer le niveau de difficulté des items sur au moins sept niveaux différents. Autrement dit, les 20 items de la version A du questionnaire se divisent selon sept niveaux d'endossement. Ce résultat suggère qu'il y a probablement une certaine redondance dans le type d'information que certains items de la version A du questionnaire permet d'obtenir.

**Tableau 6 : Statistiques d'ajustement générales des items de la version A du questionnaire 2003**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	194,0	51,0	0,00	0,17	0,99	-0,1	0,99	-0,1
<b>É.-T.</b>	33,3	0,2	0,95	0,01	0,21	1,1	0,21	1,1
<b>MAX.</b>	258,0	51,0	2,00	0,20	1,49	2,3	1,48	2,3
<b>MIN.</b>	122,0	50,0	-2,01	0,16	0,59	-2,4	0,59	-2,4

<b>Réel</b>	<b>Séparation</b>	5,25	« Fidélité »	0,96
<b>Modèle</b>	<b>Séparation</b>	5,46	« Fidélité »	0,97

La figure 2 de la page suivante permet de se faire une représentation visuelle de cette redondance. En effet, la figure présente la répartition des personnes sur le continuum de l'opinion, de même que la répartition des items sur le continuum du niveau de difficulté (difficulté à endosser), pour la version A du questionnaire. (Dans cette figure, « M » correspond à la moyenne, « S » correspond à une distance d'un écart-type par rapport à la moyenne et « T », à une distance de deux écarts-types.)

Ainsi, la figure 2 montre que, pour la version A du questionnaire, les items 1, 7, 16 et 18 occupent la même position sur le continuum du niveau de difficulté des items. Or, les items 1 (identifier les contenus difficiles à faire apprendre aux élèves), 7 (intervenir individuellement auprès des élèves à risque d'échouer) ou 16 (aider les élèves à développer leurs méthodes de travail) portent sur un construit similaire. Il n'est donc pas surprenant que le niveau de difficulté requis pour endosser ces items soit à peu près équivalent.





La figure 2 de la page précédente permet également de constater que les items les plus faciles à endosser par les étudiants sont les items 8 (planifier le déroulement d'activités d'apprentissage) et 10 (établir les règles de fonctionnement de la classe). À l'opposé, les étudiants trouvent les items 19, 3, 20 et 2 (dans cet ordre) plus difficiles à endosser. Ces items portent sur les relations avec les parents (items 2 et 20), la construction d'une évaluation sommative (item 3) et l'orientation des élèves vers les services d'aide appropriés (item 19). Il est à noter que les items 2, 3 et 20 sont au nombre des quatre items pour lesquels la formulation a été modifiée.

L'ajustement général des items de la version B du questionnaire est également à l'intérieur de l'intervalle choisi (tabl. 18, p. 185). L'indice de « fidélité » réel est de 0,98 et l'indice de séparation correspondant est de 6,33. L'indice de « fidélité » du modèle est également de 0,98 et l'indice de séparation correspondant est de 6,69. Ainsi, les indices de la version B du questionnaire répondent au seuil d'acceptabilité fixé. De plus, la version B du questionnaire permet de discriminer le niveau de difficulté des items (difficulté à endosser) sur au moins huit niveaux différents. La redondance serait donc un peu moins présente dans la version B du questionnaire que dans la version A.

La figure 9 (p. 186) révèle que dans la version B du questionnaire, ce sont les items 8 (planifier le déroulement d'activités d'apprentissage), 10 (établir les règles de fonctionnement de la classe) et 12 (respecter les différences ethniques ou culturelles des élèves) qui sont les plus faciles à endosser par les étudiants. En revanche, ce sont également les items 19, 3, 20 et 2 (dans cet ordre) qui sont les plus difficiles à endosser.

Enfin, le tableau 7 de la page suivante montre que la valeur des points d'ancrage de l'échelle de réponses estimé par le modèle pour la version A du questionnaire augmente de façon monotone croissante en fonction de la valeur de chacune des

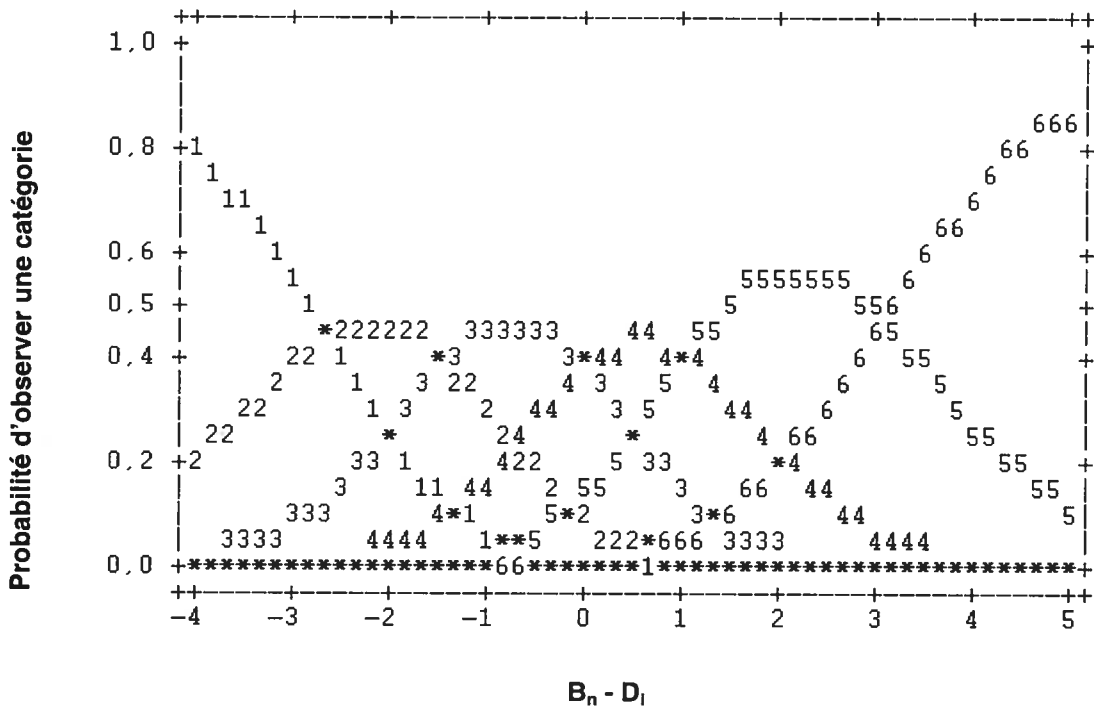
catégories. L'augmentation est d'au moins 1 logit pour chacune des catégories, mais de moins de 5 logits. La même constatation s'applique pour la version B du questionnaire (tabl. 19, p. 187). Les points d'ancrage de l'échelle répondent donc aux critères suggérés par Linacre.

**Tableau 7 : Points d'ancrage de la version A du questionnaire 2003**

Catégorie	Nb. d'obs.	% d'obs.	Mesure moyenne observée	Mesure moyenne attendue	Infit MNSQ	Outfit MNSQ	Points d'ancrage	Erreur type
1	36	4	-1,66	-1,89	1,28	1,23	NONE	
2	116	11	-1,09	-1,07	0,92	0,91	-2,64	0,19
3	251	25	-0,30	-0,26	0,98	0,97	-1,43	0,11
4	288	28	0,54	0,54	1,02	1,01	0,00	0,08
5	261	26	1,29	1,31	1,01	1,01	1,02	0,08
6	67	7	2,20	2,07	0,85	0,89	3,05	0,14

Or, il est également possible d'étudier les points d'ancrage de l'échelle de réponses graphiquement. La figure 3 de la page suivante présente donc la structure des catégories de réponse, c'est-à-dire les courbes de probabilités associées à chacune des catégories de réponse de la version A du questionnaire.

L'analyse des catégories de réponse de la version A du questionnaire (fig. 3) montre que chaque catégorie possède un sommet distinct et que l'ensemble des catégories apparaît comme une série de collines assez bien distancées. Cependant, la probabilité d'observer la quatrième catégorie est plus faible que pour les autres catégories de l'échelle. La probabilité d'observer la cinquième catégorie est un peu plus élevée que la probabilité d'observer la deuxième ou la troisième catégorie. Enfin, la probabilité d'observer la sixième catégorie est un peu plus grande que la probabilité d'observer la première.



**Figure 3 : Courbes de probabilités de chacune des catégories de réponse de la version A du questionnaire 2003**

Pour la version B du questionnaire (fig. 10, p. 187), l'analyse des catégories de réponse montre que chaque catégorie possède également un sommet distinct et que l'ensemble des catégories apparaît comme une série de collines assez bien distancées. La probabilité d'observer la cinquième catégorie est plus grande que pour les deuxième, troisième et quatrième catégories de l'échelle. Enfin, la probabilité d'observer la première catégorie est un peu plus élevée que la probabilité d'observer la sixième.

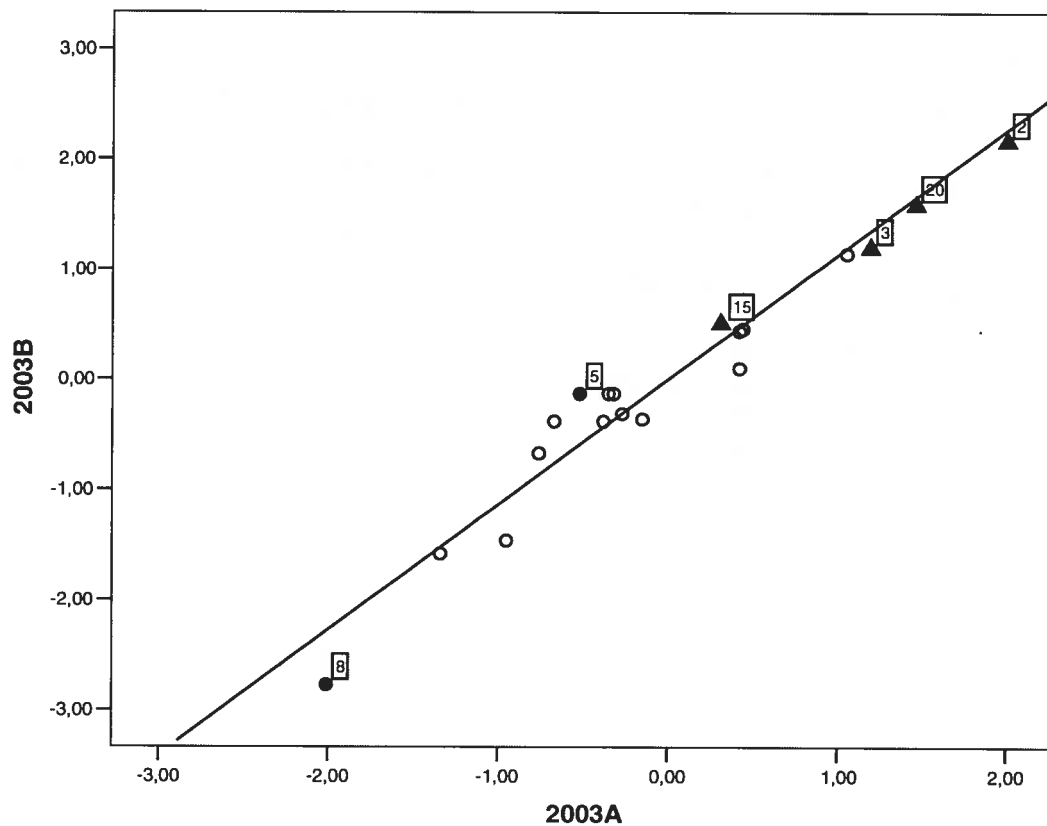
### **6.2.1.2 Comparaisons entre les versions A et B du questionnaire 2003**

L'analyse de la qualité de l'ajustement entre les données et le modèle est une étape importante, car elle permet de démontrer que le modèle choisi est effectivement approprié pour représenter les données recueillies (sect. 3.3.3). Les résultats de la section précédente ont montré que le modèle *Rating Scale* était adéquat pour modéliser les réponses de 51 sujets ayant répondu à la version A du questionnaire 2003 et 62 ayant répondu à la version B. La prochaine étape de la démarche d'analyse consiste donc à étudier l'influence de la formulation des items sur les réponses des participants de même que sur l'échelle de mesure associée.

Pour y arriver, les mesures estimées par le modèle pour le niveau de difficulté des items de chacune des versions du questionnaire devront tout d'abord être comparées. La comparaison porte sur le niveau de difficulté des items plutôt que sur l'habileté des participants (opinion), car l'objectif est d'établir si des différences existent entre les items de chacune des versions du questionnaire et non pas d'étudier si l'habileté (opinion) des participants varie d'un groupe à l'autre. Ensuite, ce sont les points d'ancrage estimés par le modèle qui devront être comparés afin d'examiner si le choix des mots exerce une influence sur l'échelle de mesure associée aux catégories de réponse.

Le diagramme de dispersion présenté à la page suivante (fig. 4) permet de visualiser les variations qui peuvent exister entre les items des versions A et B du questionnaire 2003 par rapport à la mesure estimée par le modèle pour le niveau de difficulté. Dans ce diagramme, les items qui s'éloignent le plus de la droite de dispersion ont été noircis. Les items pour lesquels la formulation a changé d'une version à l'autre ont été marqués par des triangles noirs.

L'étude de ce diagramme (fig. 4 ci-dessous) révèle que les items 5 (corriger la langue écrite des élèves) et 8 (planifier le déroulement d'activités d'apprentissage) sont les items qui s'éloignent le plus de la droite de dispersion. Les items 2, 3, 15 et 20, items pour lesquels la formulation a été modifiée, ne varient pas beaucoup d'une version à l'autre du questionnaire.



**Figure 4 : Diagramme de dispersion des items des versions A et B du questionnaire 2003 par rapport à leur niveau de difficulté**

Afin d'en savoir un peu plus sur les variations qui existent entre les deux versions du questionnaire, le diagramme de la figure 5 (page suivante) présente la distribution des items de chacune des versions du questionnaire sur le continuum du niveau de difficulté. Dans ce diagramme, les flèches continues marquent les variations des items pour lesquels la formulation a changé.

L'analyse de ce diagramme (fig. 5) permet de voir que les items 5 et 8 varient effectivement beaucoup. L'item 5 est beaucoup plus difficile à endosser dans la version B du questionnaire que dans la version A<sup>11</sup>. À l'inverse, l'item 8 est beaucoup plus facile à endosser dans la version B que dans la version A. Ces résultats sont surprenants, car la formulation de ces items est identiques pour les deux versions du questionnaire.

Le diagramme de la figure 5 permet également de confirmer que les items 2, 3, 15 et 20, items pour lesquels la formulation a été modifiée, ne varient pas beaucoup d'une version à l'autre du questionnaire. En effet, les items 2, 15 et 20 sont plus difficiles à endosser dans la version B du questionnaire que dans la version A, mais la différence est petite. La position de l'item 3 sur le continuum est la même pour les deux versions du questionnaire.

Enfin, le diagramme de la figure 5 permet de constater que pour les deux versions du questionnaire, les items 8 et 10 sont les items les plus faciles à endosser par les répondants alors que les items 20 et 2, sont les plus difficiles.

---

<sup>11</sup> Le nombre de sujets ayant répondu aux différentes versions du questionnaire étant limité, il serait préférable d'indiquer que l'item 5 est beaucoup plus difficile à endosser par le groupe de participants ayant répondu à la version B du questionnaire que par celui ayant répondu à la version A. En effet, la population étudiée ne permet pas de généraliser le niveau de difficulté des items. Cependant, afin de ne pas allonger le texte inutilement, la première forme sera celle utilisée pour le reste de ce travail.

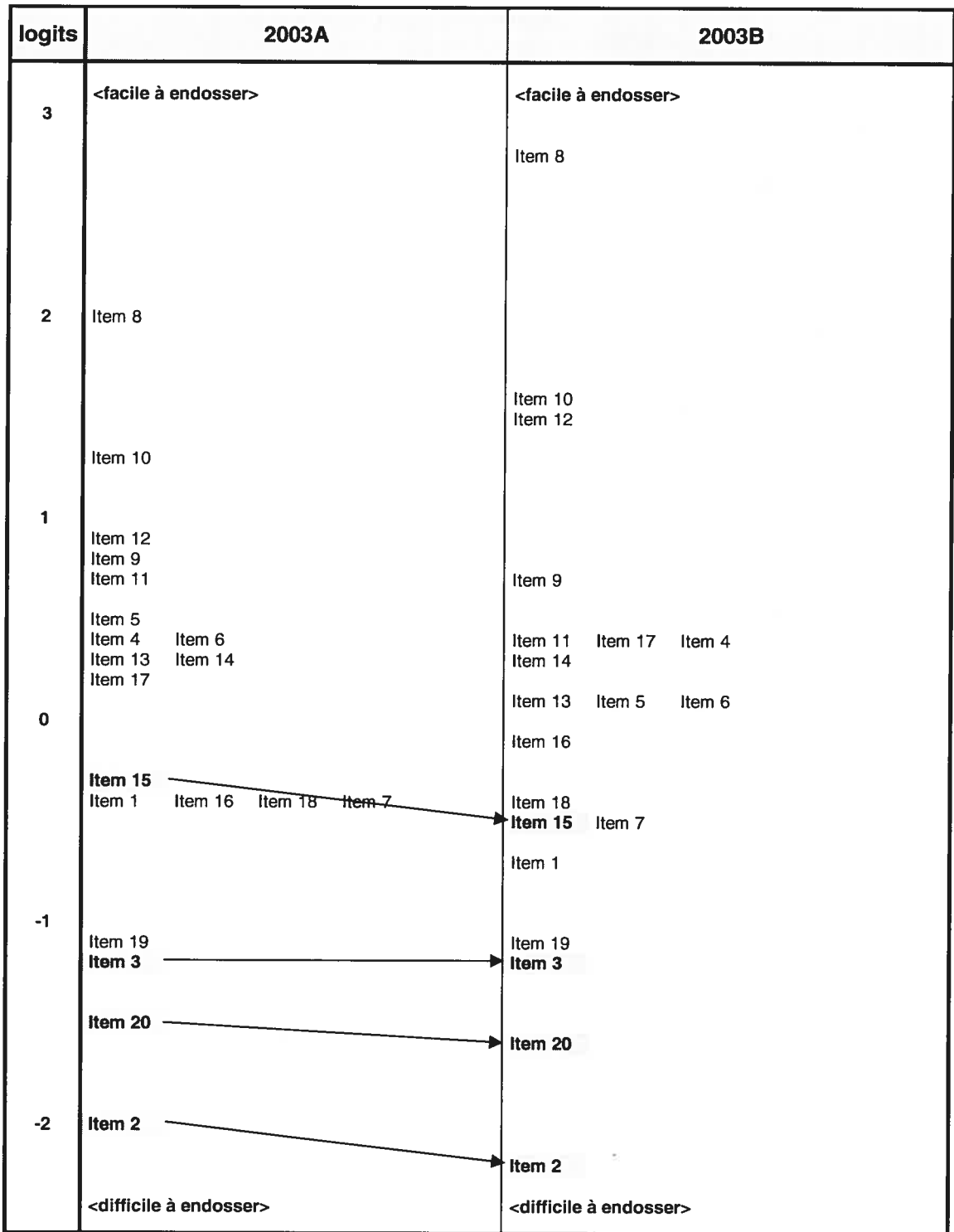
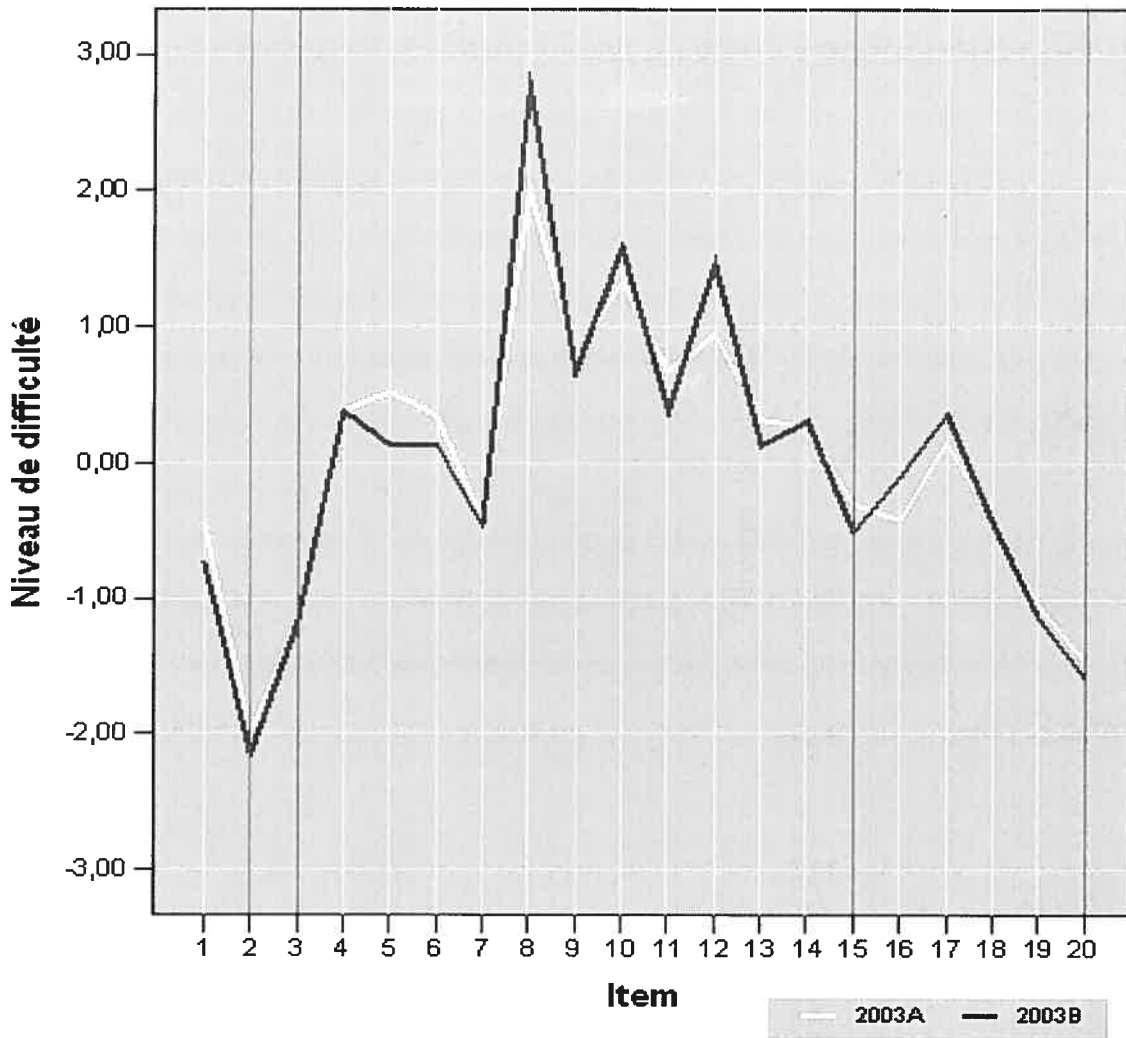


Figure 5 : Diagramme de comparaison montrant la distribution des items sur le continuum du niveau de difficulté pour les versions A et B du questionnaire 2003



Le graphique de la figure 6 (page suivante) présente les courbes de distribution des mesures estimées par le modèle pour le niveau de difficulté des items de chacune des versions du questionnaire. Cette représentation graphique permet de visualiser l'ampleur des variations, pour chacun des items, entre les deux versions du questionnaire. Pour faciliter la lecture de ce graphique, l'échelle utilisée a été inversée de façon à être identique à celle de la figure précédente. Ainsi, l'échelle est en logits et les valeurs négatives de l'échelle correspondent à des items qui sont difficiles à endosser alors que les valeurs positives correspondent à des items plus faciles.

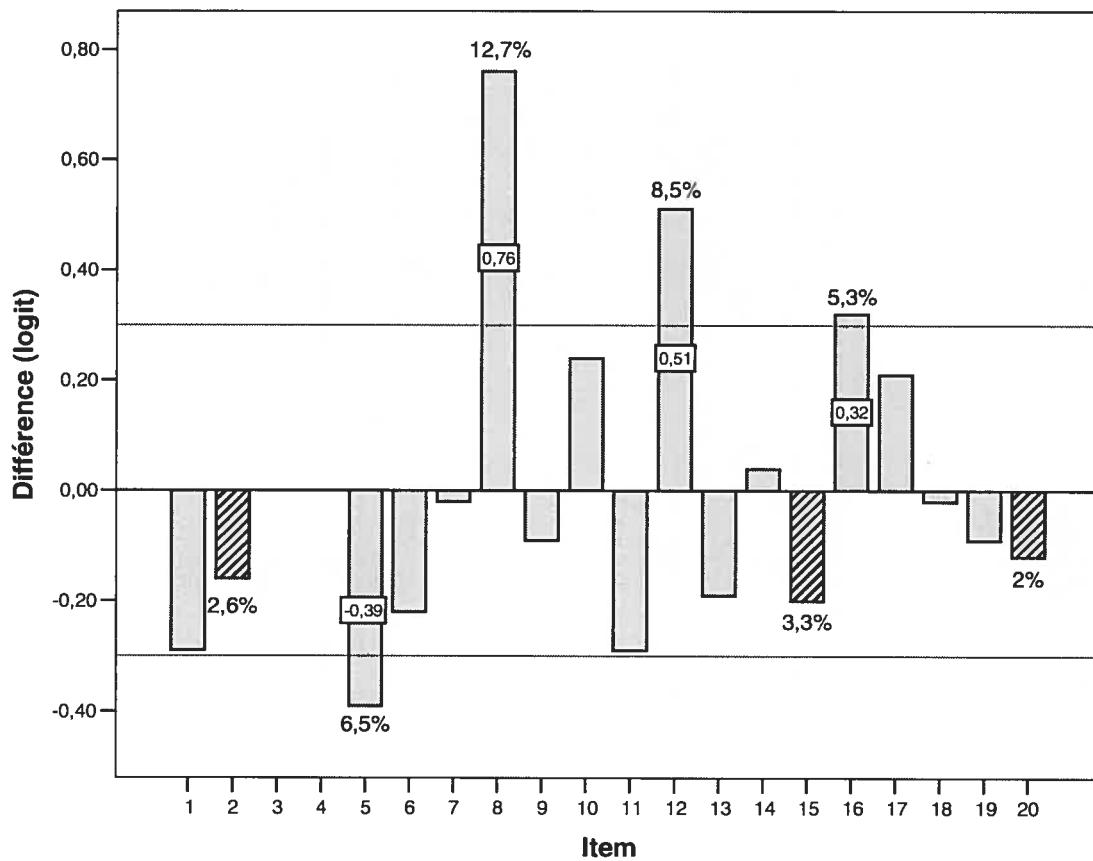
L'analyse de cette figure (fig. 6) montre que l'écart est généralement petit entre les deux versions du questionnaire. Les amplitudes les plus grandes sont observées pour les items 5 et 8 (items identifiés précédemment). Une différence notable existe également pour les items 12 et 16.



**Figure 6 : Courbes de distribution des mesures estimées par le modèle pour le niveau de difficulté des items des versions A et B du questionnaire 2003**

Le diagramme en bâton présenté à la page suivante (fig. 7) permet d'examiner plus en détail l'ampleur des différences qui existent, pour chacun des items, entre les deux versions du questionnaire. Dans ce diagramme, la hauteur des bâtons représente la valeur de la différence entre la mesure estimée par le modèle pour le niveau de difficulté des items de la version B par rapport à ceux de la version A. L'échelle de ce diagramme a également été inversée de façon à être identique aux échelles des précédents graphiques.

Les bâtons au-dessus de l'axe horizontal, correspondent donc aux items qui sont plus faciles à endosser dans la version B que dans la version A. À l'inverse, les bâtons au-dessous de l'axe correspondent aux items qui sont plus difficiles à endosser dans la version B que dans la version A. Les lignes de référence qui ont été ajoutées déterminent un intervalle de plus ou moins 5 % de variation entre les deux versions du questionnaire par rapport à l'échelle en 6 points utilisée dans les graphiques précédents. Les bâtons représentant les items dont la formulation a changé ont été hachurés.



**Figure 7 : Diagramme en bâton représentant la valeur des différences entre les mesures estimées par le modèle pour les version A et B du questionnaire 2003**

Le diagramme de la page précédente (fig. 7) montre que seulement quatre items ont une variation supérieure à 5 % (items 5, 8, 12 et 16). Les items 8, 12 et 16 sont plus faciles à endosser dans la version B que dans la version A. À l'inverse, l'item 5 est plus difficile dans la version B que dans la version A. Il est à noter que la formulation de ces quatre items n'a pas été changée. Il sera donc intéressant de s'interroger sur ces résultats dans le prochain chapitre.

La variation observée pour les items dont la formulation a été modifiée est faible. La variation la plus grande est de 3,3 % pour l'item 15 alors que la variation est nulle pour l'item 3. Il semble donc que le choix des mots dans la formulation des items n'a pas eu beaucoup d'influence sur les réponses des participants. Ces résultats seront discutés au prochain chapitre.

Enfin, le diagramme de la figure 8 (p. 114) permet de comparer la position des points d'ancrage estimés par le modèle, pour chacun des items, entre les deux versions du questionnaire. Les courbes de probabilités de chacune des catégories de réponse (fig. 3, p. 104) permettaient d'étudier les points d'ancrage de l'ensemble des items d'un questionnaire. La figure 8 permet maintenant de comparer la position des points d'ancrage des items dont la formulation a été modifiée par rapport à ceux qui sont demeurés inchangés. L'échelle sur ce diagramme est la même que celle de la figure 3 (p. 104) qui illustre les courbes de probabilités des différentes catégories de réponse de l'ensemble des items de la version A du questionnaire.

L'analyse de ce diagramme (fig. 8) montre que, de façon générale, les points d'ancrage des items de la version B du questionnaire sont plus dilatés que ceux des items de la version A. Le diagramme permet également de voir que les points d'ancrage des items les plus faciles à endosser, comme les items 8, 10 ou 12, se situent davantage vers

la partie gauche de l'échelle alors que les items les plus difficiles à endosser, comme les items 2 et 20, se situent davantage vers la partie droite.

De la même façon, le diagramme permet de visualiser le glissement vers la partie droite de l'échelle des points d'ancrage des items qui sont plus difficiles à endosser dans la version B du questionnaire que dans la version A (item 5 par exemple). Il est alors plus difficile pour les participants ayant répondu à la version B du questionnaire d'accéder aux catégories supérieures de l'échelle de réponses et donc, de fournir une réponse favorable envers ces items. À l'inverse, les points d'ancrage des items qui sont plus faciles à endosser dans la version B du questionnaire se situent davantage vers la partie gauche de l'échelle.

Le diagramme de la figure 8 montre également que, de façon générale, les points d'ancrage des catégories 4 et 5 sont plus rapprochés que les autres. Il est donc possible que la définition de ces catégories ne soit pas assez claire pour que les participants puissent bien les distinguer. Le rapprochement entre les catégories 4 et 5 est un peu plus prononcé pour les items de la version A du questionnaire que pour ceux de la version B. Or, les résultats de la section précédente ont justement permis de constater que la probabilité d'observer la quatrième catégorie était plus petite que pour les autres catégories. Les points d'ancrage de tous les items (versions A et B) répondent cependant aux critères proposés par Linacre, c'est-à-dire que l'intervalle entre chacun d'eux est plus grand que 1 logit, mais de moins de 5 logits.

Enfin, les points d'ancrage des items dont la formulation a été modifiée sont assez similaires à ceux des autres items étudiés. Ainsi, il semble que le choix des mots dans la formulation des items n'ait pas eu beaucoup d'influence sur l'échelle de mesure associée aux catégories de réponse. Ces résultats seront discutés au prochain chapitre.

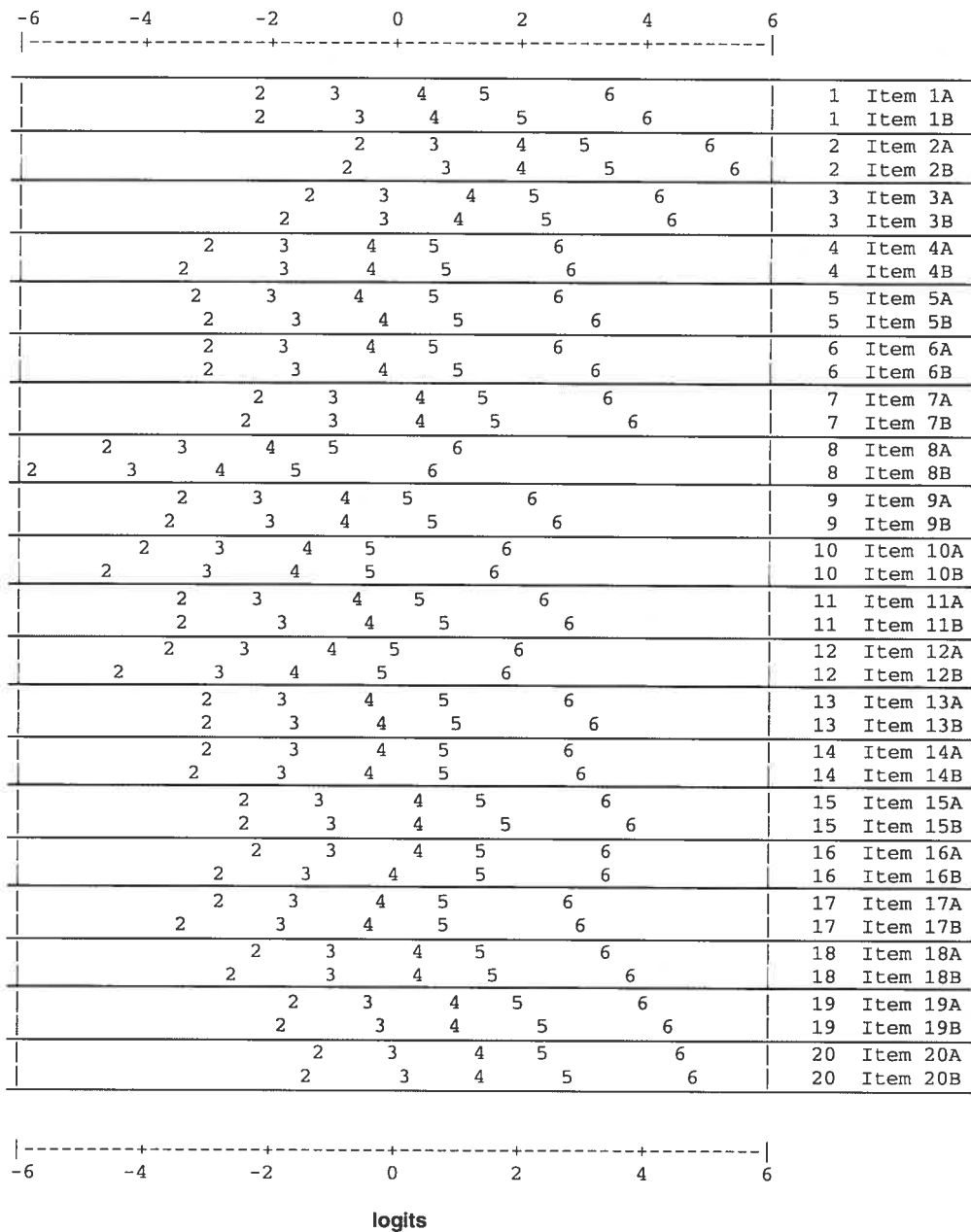


Figure 8 : Diagramme de comparaison montrant la position des points d'ancrage estimés par le modèle pour chacun des items des versions A et B du questionnaire 2003

## 6.2.2 Questionnaires 2004

La cueillette de données de 2004 a permis de récolter les réponses de 125 étudiants et étudiantes provenant du programme d'enseignement en éducation préscolaire et enseignement primaire. Parmi ces 125 étudiants, 63 ont répondu à la version A du questionnaire et 62 ont répondu à la version B.

### 6.2.2.1 Qualité de l'ajustement pour les questionnaires 2004

En premier lieu, la qualité de l'ajustement des personnes a été effectuée. Afin de ne pas alourdir la présentation des résultats, tous les tableaux seront dorénavant présentés en annexe. (Les tableaux de l'analyse des questionnaires 2004 se trouvent donc dans l'annexe C.2 de ce travail.)

Les tableaux 20 et 21 (p. 188 et 190) montrent que suite à une première analyse, 11 sujets ayant répondu à la version A du questionnaire ne s'ajustaient pas bien au modèle (sujets 6, 36, 61, 63, 71, 72, 73, 81, 84, 85 et 151), de même que 9 sujets ayant répondu à la version B (sujets 205, 233, 234, 235, 278, 286, 301, 303 et 344). De plus, la corrélation était négative pour un des participants ayant répondu à la version B du questionnaire (sujet 301).

Une à une, les personnes possédant les pires indices d'ajustement standardisés (« *infit* » ou « *outfit* ») ou une corrélation négative ont été retirées de l'échantillon. La qualité de l'ajustement était ensuite étudiée de nouveau. Ce processus a fait en sorte que

certaines sujets qui ne s'ajustaient pas bien au modèle initialement, se sont bien ajustés au modèle par la suite. De même, le processus a permis de faire ressortir de nouveaux sujets qui ne s'ajustaient plus au modèle. Ainsi, à la fin du processus, 12 participants ont dû être retirés de l'échantillon pour la version A du questionnaire (sujets 6, 9, 32, 36, 72, 73, 84, 85, 128, 151, 152 et 160). Dans la version B du questionnaire, 10 sujets ont dû être retirés de l'échantillon (sujets 207, 233, 234, 235, 263, 278, 286, 301, 303 et 344). À la suite de quoi, la qualité de l'ajustement des personnes répondait au critère choisi et la corrélation entre la mesure estimée par le modèle pour chacune des personnes et les observations recueillies était positive pour toutes les personnes retenues.

Par conséquent, 51 sujets ont été conservés dans l'échantillon pour la version A du questionnaire et 52 pour la version B. Les tableaux 22 et 23 (p. 192 et 194) présentent les statistiques d'ajustement finales résultant de cette analyse.

Une fois l'analyse de la qualité de l'ajustement des personnes terminée, l'analyse de la qualité de l'ajustement des items a été effectuée. Le tableau 24 (p. 196) montre que, dans la version A du questionnaire, deux items ne s'ajustent pas bien au modèle. Les items 12 (respecter les différences ethniques ou culturelles des élèves) et 13 portant sur l'identification des points forts et des points faibles des élèves (item dont la formulation a été modifiée) ne répondent donc pas au critère choisi. Dans la version B du questionnaire, tous les items s'ajustaient bien au modèle (tabl. 25, p. 197). La corrélation entre la mesure estimée par le modèle était positive pour tous les items des deux versions du questionnaire.

Il est important de rappeler qu'aucun item n'a été retiré de l'échantillon. Le nombre de participants ayant répondu à ces questionnaires étant limité, il est possible que les problèmes d'ajustement soient liés à ce faible nombre et à la distribution non uniforme des réponses obtenues à travers les catégories de réponse de l'échelle (voir sect. 6.1.2).



Toutefois, afin de compléter l'exploration de cette recherche, les résultats obtenus une fois que les items mal ajustés ont été retirés de l'échantillon seront présentés un peu plus loin.

Ensuite, le tableau 26 (p. 198) montre que, pour la version A du questionnaire, l'ajustement général des personnes est à l'intérieur de l'intervalle choisi et que la valeur moyenne estimée par le modèle pour le degré d'adhésion (opinion) des personnes est de -0,76 avec une erreur type de 0,28. L'indice de « fidélité » réel est de 0,93 alors que l'indice de « fidélité » du modèle est de 0,94. L'indice de séparation réel est de 3,79 et l'indice de séparation du modèle est de 4,07. Ainsi, la version A du questionnaire répond au seuil d'acceptabilité choisi et permet de discriminer l'opinion des individus sur au moins cinq niveaux différents par rapport aux six niveaux proposés dans l'échelle de réponses du questionnaire.

Le tableau 27 (p. 198) montre que, pour la version B du questionnaire, l'ajustement général des personnes est à l'intérieur de l'intervalle choisi et que la valeur moyenne estimée par le modèle pour le degré d'adhésion était de -0,36 avec une erreur type de 0,26. Il est donc un peu plus difficile pour les individus ayant répondu à la version B du questionnaire d'être en accord avec les différents items proposés que pour les individus ayant répondu à la version A du questionnaire. L'indice de « fidélité » réel est de 0,93 et l'indice de « fidélité » du modèle est de 0,94. Les indices de séparation correspondant sont de 3,59 et de 3,82 respectivement. Ainsi, la version B du questionnaire répond au seuil d'acceptabilité choisi et permet de discriminer l'opinion des individus sur au moins cinq niveaux différents.

Le tableau 28 (p. 199) présente les statistiques d'ajustement générales des items pour la version A du questionnaire. Le tableau 29 présente celles de la version B du questionnaire. L'ajustement général des items des deux versions du questionnaire est à l'intérieur de l'intervalle choisi. La valeur moyenne du niveau de difficulté des items est

toujours fixée à 0 afin de faciliter l'interprétation des résultats. L'indice de « fidélité » réel des items de la version A du questionnaire est de 0,96 par rapport 0,97 pour la version B. L'indice de « fidélité » du modèle est identique pour les deux versions du questionnaire avec 0,97. L'indice de séparation réel des items de la version A est de 5,07 par rapport à 5,43 pour la version B. L'indice de séparation du modèle pour la version A est de 5,31 par rapport à 5,74 pour la version B. Ainsi, les deux versions du questionnaire répondent au seuil d'acceptabilité choisi et permettent de discriminer le niveau de difficulté des items sur au moins sept niveaux différents.

L'analyse graphique de la répartition des personnes sur le continuum de l'opinion ou des items sur le continuum du niveau de difficulté de la version A du questionnaire (fig. 11, p. 200) montre que les items les plus faciles à endosser sont les items 8 (planifier le déroulement d'activités d'apprentissage) et 12 (respecter les différences ethniques ou culturelles des élèves). À l'opposé, les items 20 et 2 (dans cet ordre) sont les plus difficiles à endosser. Ces items portent sur les relations avec les parents. De plus, ces items sont au nombre des 11 items pour lesquels la formulation a été modifiée. Dans la version B du questionnaire (fig. 12, p. 201), ce sont également les items 8 et 12 qui sont les plus faciles à endosser et les items 20 et 2 (dans cet ordre), qui sont les plus difficiles à endosser.

La valeur des points d'ancrage de l'échelle de réponses estimé par le modèle augmente de façon monotone croissante en fonction de la valeur de chacune des catégories pour les deux versions du questionnaire (tabl. 30 et 31, p. 202 et 202). Cependant, dans la version A du questionnaire l'augmentation est seulement de 0,85 logit entre la quatrième et la cinquième catégorie de l'échelle. De même, pour la version B du questionnaire, l'augmentation est seulement 0,96 logit entre la deuxième et la troisième catégorie, et de 0,9 entre la troisième et la quatrième catégorie. Les points d'ancrage des échelles des deux versions du questionnaire ne répondent donc pas aux critères suggérés par Linacre. Il est possible que ces résultats soient causé par le mauvais ajustement des

items 12 et 13. Sinon, selon Linacre, il est probable que la définition de ces catégories ne soit pas assez claire pour que les participants puissent bien les distinguer.

Enfin, l'analyse graphique des catégories de réponse de la version A du questionnaire (fig. 13, p. 203) permet de constater que, la première catégorie n'a que très peu de chance d'être utilisée alors que la sixième n'en a aucune. La cinquième catégorie est celle qui a le plus de chance d'être utilisée, suivi de la deuxième, de la troisième et enfin, de la quatrième catégorie.

Dans la version B du questionnaire (fig. 14, p. 204), la sixième catégorie n'a que très peu de chance d'être utilisée. La première et la cinquième catégorie ont le plus de chance d'être utilisées suivi de la quatrième, de la deuxième et enfin, de la troisième catégorie.

### **6.2.2.2 Comparaisons entre les versions A et B du questionnaire 2004**

Les résultats de la section précédente ont montré que le modèle choisi était adéquat pour modéliser les réponses de 51 sujets retenus pour la version A du questionnaire 2004 et des 52 retenus pour la version B. En revanche, les résultats ont également montré que deux items, les items 12 et 13, ne s'ajustaient pas bien au modèle. De plus, l'analyse des points d'ancrage des catégories de réponse a révélé que l'intervalle entre certaines catégories ne répondaient pas aux critères suggérés par Linacre. Il sera donc important de garder ces observations en mémoire lors de l'interprétation des résultats présentée dans le prochain chapitre.

Le diagramme de dispersion des versions A et B du questionnaire 2004 (fig. 15, p. 205) révèle que la variation est plus grande en 2004 que ce qui avait été observé en 2003. Il est à noter que seulement 4 items avaient été modifiés dans la version 2003 par rapport à 11 dans la version 2004. L'analyse de la figure 15 montre également que les items 5 (item dont la formulation a été modifiée) et 12 sont ceux qui s'éloignent le plus de la droite de dispersion.

Le diagramme de la figure 16 (p. 206) permet de voir que l'item 5 varie effectivement beaucoup d'une version à l'autre du questionnaire. Il s'agit même de la variation la plus importante constatée sur ce diagramme. La variation des items 2, 7, 11, 12, 19 et 20 semble également notable par rapport aux autres items de la distribution. Dans les deux versions du questionnaire, les items 8 et 12 sont les plus faciles à endosser par les répondants alors que les items 20 et 2 sont les plus difficiles. Enfin, parmi les items ayant été modifiés, les items 2, 7, 11, 13, 19 et 20 sont plus faciles à endosser dans la version B du questionnaire que dans la version A. Les items 1 et 15 occupent sensiblement la même position. Seuls les items 3, 4 et 5 sont plus difficiles à endosser dans la version B du questionnaire que dans la version A.

L'analyse des courbes de distribution des mesures estimées par le modèle pour le niveau de difficulté (fig. 17, p. 207) confirme que l'écart le plus important correspond à l'item 5. L'item 5 de la version B est beaucoup plus difficile à endosser que celui de la version A. Cependant, des différences notables peuvent également être constatées pour les items 2, 7, 11, 12, 14, 16, 18, 19 et 20. Parmi ces items, cinq sont des items pour lesquels la formulation a été modifiée (items 2, 7, 11, 19 et 20).

Le diagramme en bâton de la figure 18 (p. 208) permet d'observer les grandes variations décrites jusqu'ici. En effet, la figure 18 indique que la valeur de la mesure estimée par le modèle pour le niveau de difficulté de l'item 5 est 20 % plus élevée dans la

version B que dans la version A. Ainsi, il est beaucoup plus difficile d'endosser l'item 5 dans la version B que dans la version A. De plus, deux items pour lesquels la formulation n'a pas été changée ont une variation supérieure à 5 % (items 12 et 17). Parmi les items pour lesquels la formulation a été modifiée, cinq ont une variation supérieure à 5 % (items 2, 5, 7, 11 et 20), cinq ont une variation inférieure à ce seuil (items 1, 3, 4, 13 et 19). Enfin, la variation de l'item 11 est nulle. À l'exception de l'item 5, la variation est toujours inférieure à 10 %.

Le diagramme de comparaison de la figure 19 (p. 209) révèle finalement que, de façon générale, les points d'ancrage des items de la version B du questionnaire sont un plus comprimés que ceux des items de la version A. De plus, les différences les plus grandes entre les points d'ancrage des versions A et B se trouvent généralement aux extrémités des échelles, c'est-à-dire pour la deuxième et la sixième catégorie plus particulièrement. À l'inverse, les points d'ancrage des catégories intermédiaires (3, 4 et 5) sont plus rapprochés. De façon générale, l'intervalle entre ces catégories est plus petit ou égal à 1 logit, ce qui ne permet pas de répondre au critère proposé par Linacre.

Ainsi, l'analyse du diagramme ne permet pas d'établir facilement dans quelle mesure le choix des mots dans la formulation des items aurait pu avoir une influence sur l'échelle de mesure associée aux catégories de réponse. En effet, les variations entre les points d'ancrage des deux versions du questionnaire sont assez similaires pour tous les items.

### 6.2.3 Questionnaires 2005

La cueillette de données de 2005 a permis de récolter les réponses de 113 étudiants et étudiantes provenant du programme d'enseignement en éducation préscolaire et enseignement primaire. Parmi ces 113 étudiants, 72 ont répondu à la version A du questionnaire et 41 ont répondu à la version B.

#### 6.2.3.1 Qualité de l'ajustement pour les questionnaires 2005

La première étape de la démarche d'analyse choisie consiste à vérifier la qualité de l'ajustement des personnes. Ainsi, les tableaux 32 et 33 (annexe C.3, p. 210 et 212) montrent que suite à une première analyse, 12 sujets ayant répondu à la version A du questionnaire ne s'ajustaient pas bien au modèle (sujets 1, 11, 50, 67, 74, 80, 88, 127, 129, 131, 136 et 138), de même que 8 sujets ayant répondu à la version B (sujets 523, 525, 541, 576, 577, 596, 610 et 621). La corrélation était négative pour un des participants ayant répondu à la version A du questionnaire (sujet 129).

Une à une, les personnes possédant les pires indices d'ajustement standardisés (« *infit* » et « *outfit* ») ou une corrélation négative ont été retirées de l'échantillon. La qualité de l'ajustement était ensuite étudiée de nouveau. Ce processus a fait en sorte que certains sujets qui ne s'ajustaient pas bien au modèle initialement se sont bien ajustés au modèle par la suite. De même, le processus a permis de faire ressortir de nouveaux sujets qui ne s'ajustaient plus au modèle. Ainsi, à la fin du processus, 11 sujets ont dû être retirés de l'échantillon pour la version A du questionnaire (sujets 1, 3, 11, 67, 74, 88, 127, 129, 131, 136 et 138). Dans la version B du questionnaire, 10 sujets ont dû être retirés de

l'échantillon (sujets 525, 541, 558, 572, 574, 577, 596, 607, 610 et 621). À la suite de quoi, la qualité de l'ajustement des personnes répondait au critère choisi et la corrélation entre la mesure estimée par le modèle pour chacune des personnes et les observations recueillies était positive pour toutes les personnes retenues.

Par conséquent, 61 sujets ont été conservés dans l'échantillon pour la version A du questionnaire et 31 pour la version B. Les tableaux 34 et 35 (p. 214 et 216) présentent les statistiques d'ajustement finales résultant de cette analyse.

Une fois l'analyse de la qualité de l'ajustement des personnes terminée, l'analyse de la qualité de l'ajustement des items a donc ensuite été effectuée. Le tableau 36 (p. 217) montre que, dans la version A du questionnaire, trois items ne s'ajustent pas bien au modèle (items 1, 3 et 13). Dans la version B (tabl. 37, p. 218), ce sont les items 6, 8 et 19 qui ne s'ajustent pas bien au modèle. Ainsi, six items ne répondent pas au critère choisi pour l'année 2005. La corrélation était cependant positive pour tous les items. Comme il a été mentionné précédemment, aucun item n'a été retiré de l'échantillon. Il est à noter que le petit nombre de sujet retenus dans l'échantillon (seulement 31 pour la version B du questionnaire) peut avoir une influence sur les résultats. Toutefois, puisqu'il s'agit d'une recherche exploratoire, ces résultats seront conservés en mémoire afin d'éclairer l'interprétation des résultats qui sera proposée au chapitre suivant.

Le tableau 38 (p. 219) montre que, pour la version A du questionnaire, l'ajustement général des personnes est à l'intérieur de l'intervalle choisi et que la valeur moyenne estimée par le modèle pour le degré d'adhésion des personnes est de 0,24 avec une erreur type de 0,28. L'indice de « fidélité » réel est de 0,93 alors que l'indice de « fidélité » du modèle est de 0,94. Les indices de séparation correspondant sont de 3,62 et 3,9 respectivement. Ainsi, la version A du questionnaire répond au seuil d'acceptabilité choisi et permet de discriminer l'opinion des individus sur au moins cinq niveaux différents.

Le tableau 39 (p. 219) montre que, pour la version B du questionnaire, l'ajustement général des personnes est à l'intérieur de l'intervalle choisi et que la valeur moyenne estimée par le modèle pour le degré d'adhésion est de -0,02 avec une erreur type de 0,31. L'indice de « fidélité » réel est de 0,97, de même que l'indice de « fidélité » du modèle. Les indices de séparation correspondant sont respectivement de 5,75 et 6,13. Par conséquent, la version B du questionnaire répond au seuil d'acceptabilité choisi et permet de discriminer l'opinion des individus sur au moins huit niveaux différents. Ce résultat est surprenant puisque l'échelle proposée ne contient que six niveaux d'opinion différents.

Le tableau 40 (p. 220) présente les statistiques d'ajustement générales des items pour la version A du questionnaire. Le tableau 41 présente celles de la version B du questionnaire. L'ajustement général des items des deux versions du questionnaire est à l'intérieur de l'intervalle choisi. La valeur moyenne du niveau de difficulté des items est toujours fixée à 0 pour les deux versions du questionnaire. L'indice de « fidélité » réel des items de la version A du questionnaire est de 0,95, de même que l'indice de « fidélité » du modèle. Les indices de séparation correspondant sont de 4,24 et 4,48 respectivement. Pour la version B du questionnaire, l'indice de « fidélité » réel est de 0,93 alors que l'indice du modèle est de 0,94. Les indices de séparation correspondant sont respectivement de 3,63 et 3,93. Ainsi les deux questionnaires répondent au seuil d'acceptabilité choisi et permettent de discriminer le niveau de difficulté des items sur au moins cinq niveaux différents.

L'analyse graphique de la répartition des items sur le continuum du niveau de difficulté de la version A du questionnaire (fig. 20, p. 221) montre que les items 8 et 12 sont encore une fois les items les plus faciles à endosser. À l'opposé, l'item 2 est encore celui qui est le plus difficile à endosser. Dans la version B du questionnaire (fig. 21, p. 222), les



items les plus faciles à endosser sont les items 8 et 10. À l'opposé, ce sont les items 3 et 19 qui sont les plus difficiles à endosser.

La valeur des points d'ancrage de l'échelle de réponses estimée par le modèle augmente de façon monotone croissante en fonction de la valeur de chacune des catégories pour les deux versions du questionnaire (tabl. 42 et 43, p. 223). De plus, l'augmentation est d'au moins 1 logit, mais de moins de 5 logits pour toutes les catégories de l'échelle. Les points d'ancrage estimés par le modèle répondent donc au critère suggéré par Linacre (malgré les six items mal ajustés).

Enfin, l'analyse graphique des catégories de réponse de la version A du questionnaire (fig. 22, p. 224) permet de constater que la première et la dernière catégorie n'ont jamais plus de chance qu'une autre d'être utilisée. En effet, elles ne possèdent pas un sommet distinct. La deuxième catégorie est celle qui a le plus de chance d'être utilisée, suivi de la cinquième, de la quatrième et de la troisième catégorie.

Dans la version B du questionnaire (fig. 23, p. 225), la situation est similaire. Ainsi, la première et la sixième catégorie n'ont jamais plus de chance que les autres d'être utilisées. Les autres catégories ayant des chances à peu près égales d'être choisies.

### **6.2.3.2 Comparaisons entre les versions A et B du questionnaire 2005**

Les résultats de la section précédente ont montré que le modèle choisi était adéquat pour modéliser les réponses des 61 sujets retenus pour la version A du questionnaire et des 31 sujets retenus pour la version B. En revanche, les résultats ont également montré que trois items, les items 1, 3 et 13 ne s'ajustaient pas bien dans la

version A du questionnaire. De même, trois autres items, les items 6, 8 et 19, ne s'ajustaient pas bien dans la version B du questionnaire.

Le diagramme de dispersion des versions A et B du questionnaire 2005 (fig. 24, p. 226) révèle que, comme pour l'année 2004, la variation est plus grande que celle qui avait été observée en 2003. Il est à noter que quatre items avaient été modifiés en 2003 par rapport à sept en 2005. Le diagramme montre également que les items 5 et 20 sont les items qui varient le plus par rapport à la droite de dispersion. L'item 9 varie également un peu plus que les autres.

Le diagramme de la figure 25 (p. 227) permet de voir que ce sont effectivement les items 5, 9 et 20 qui varient le plus d'une version à l'autre du questionnaire. L'item 9 est beaucoup plus facile à endosser dans la version B du questionnaire que dans la version A. À l'inverse, les items 5 et 20 sont beaucoup plus difficiles à endosser dans la version B que dans la version A. La formulation de ces trois items a été modifiée. Il faut d'ailleurs rappeler que la formulation de l'item 20 a été complètement changée en 2005 et que les deux versions de cet item ne sont pas équivalentes. Les autres items pour lesquels la formulation a été modifiée varient peu entre les deux versions du questionnaire.

L'analyse des courbes de distribution des mesures estimées par le modèle pour le niveau de difficulté (fig. 26, p. 228) révèle que des différences importantes existent entre les deux versions du questionnaire. La figure 26 permet de confirmer que les écarts les plus importants sont observés pour les items 5, 9 et 20. Toutefois, il existe également des différences notables pour presque tous les items, sauf pour les items 4, 12, 14, 15 et 18 qui semblent assez stables d'une version à l'autre du questionnaire. Parmi ces derniers, trois items, les items 4, 14 et 15 sont des items pour lesquels la formulation a été modifiée.

Le diagramme en bâton de la figure 27 (p. 229) permet d'observer les variations décrites jusqu'ici. En effet, la figure 27 indique que la valeur de la mesure estimée par le modèle pour le niveau de difficulté de l'item 5 est 20,5 % plus élevée dans la version B que dans la version A. De même, la valeur estimée pour l'item 20 est 21 % plus élevée dans la version B que dans la version A. Il est important de rappeler encore une fois que l'item 20 est complètement différent dans les deux versions du questionnaire. Six items pour lesquels la formulation des items n'a pas été modifiée ont une variation supérieure à 5 % (items 3, 6, 7, 9, 11 et 19). Cette variation n'excède toutefois pas les 10 %. Parmi les items pour lesquels la formulation a été modifiée, trois items ont une variation supérieure à 5 % (items 5, 9 et 20). Dans les trois cas, la variation est même supérieure à 10 %.

Enfin, le diagramme de comparaison de la figure 28 (p. 230) révèle que, de façon générale, les points d'ancrage des items de la version B du questionnaire sont un peu plus dilatés que ceux des items de la version A. La différence est toutefois assez petite.

De plus, les points d'ancrage des catégories 3 et 4 sont généralement un peu plus rapprochés que les autres. Ce rapprochement est un plus prononcé dans la version A du questionnaire que dans la version B. L'intervalle entre les points d'ancrage répond cependant au critère proposé par Linacre.

Les différences les plus grandes entre les points d'ancrage des versions A et B peuvent être observées pour les items 5, 8, 9 et 20. Parmi ces items, seuls les items 5 et 20 sont des items pour lesquels la formulation a été modifiée. Par conséquent, il est difficile d'établir dans quelle mesure le choix des mots dans la formulation des items aurait pu avoir une influence sur l'échelle de mesure associée aux catégories de réponse.

## **6.2.4 Analyses complémentaires**

Afin de compléter l'exploration de cette recherche, la présente section propose quelques résultats d'analyses complémentaires effectuées une fois que les items mal ajustés ont été retirés de l'échantillon. Les résultats des sections précédentes ont montré que deux items de la version A du questionnaire 2004, trois items de la version A du questionnaire 2005 et trois items de la version B du questionnaire 2005 étaient mal ajustés. Par conséquent, la présente section porte sur les versions 2004 et 2005 des questionnaires. (Les tableaux de cette analyse sont présentés dans l'annexe D de ce travail.)

### ***6.2.4.1 Qualité de l'ajustement de la version A du questionnaire 2004***

L'analyse de la qualité de l'ajustement des items de la version A du questionnaire 2004 a permis de déceler que deux items, les items 12 et 13, ne s'ajustaient pas bien au modèle (sect. 6.2.2.1). L'item qui s'ajustait le moins au modèle (item 13) a d'abord été retiré de l'échantillon. Il s'agit d'un des items pour lequel la formulation avait été modifiée. Ensuite, la qualité de l'ajustement a été étudiée de nouveau. Suite à ce retrait, la qualité de l'ajustement des items répondait au critère choisi.

Par conséquent, dans l'analyse qui suit, l'échantillon retenu est constitué de 51 sujets et de 19 items. Les tableaux 44 et 45 (annexe D.1, p. 231 et 233) présentent les statistiques d'ajustement finales pour les personnes et pour les items résultant de cette analyse.

Le tableau 46 (p. 234) montre que la valeur moyenne estimée par le modèle pour le degré d'adhésion (opinion) des personnes est maintenant de -0,75. La valeur obtenue précédemment en conservant les 20 items était de -0,76. L'indice de séparation réel du modèle est maintenant de 3,61 par rapport à la valeur de 3,79 obtenue précédemment. L'indice de séparation du modèle est maintenant de 3,9 par rapport à 4,07. Les indices de « fidélité » correspondants sont identiques à ceux obtenus précédemment, c'est-à-dire 0,93 et 0,94 respectivement.

Le tableau 47 (p. 234) indique que l'indice de séparation réel des items est de 5,15 par rapport à la valeur de 5,43 obtenue précédemment. L'indice de séparation du modèle est de 5,35 par rapport à 5,74. Les indices de « fidélité » correspondants sont de 0,97. Les valeurs obtenues précédemment étaient de 0,96 et 0,97 respectivement.

Ainsi, les résultats obtenus à l'aide de l'échantillon de 19 items sont à peu près équivalents à ceux obtenus précédemment à partir de l'échantillon de 20 items. La figure 29 (p. 235) permet de constater qu'effectivement, la distribution des personnes sur le continuum de l'habileté et celle des items sur le continuum du niveau de difficulté sont assez similaires à ce qui avait été obtenu précédemment (fig. 11, p. 200).

Le tableau 48 (p. 236) présente la valeur des points d'ancrage de l'échelle de réponses estimé par le modèle pour l'échantillon de 19 items. Les valeurs obtenues sont, encore une fois, assez similaires à celles obtenues avec l'échantillon de 20 items (tab. 30, p. 202). En revanche, l'analyse graphique des catégories de réponse (fig. 30, p. 236) montre que toutes les courbes de probabilités possèdent maintenant un sommet distinct, ce qui n'était pas le cas précédemment.

#### **6.2.4.2 Comparaisons entre les versions A et B du questionnaire 2004**

Afin de compléter cette analyse, les comparaisons entre les versions A et B du questionnaire 2004 ont été reprises. Dans cette comparaison, l'item 13 a été exclu pour les deux versions du questionnaire. L'analyse suivante porte donc sur 19 items plutôt que 20 comme dans les comparaisons initiales effectuées pour le questionnaire 2004. Enfin, il est important de rappeler que l'échantillon retenu est constitué de 51 sujets pour la version A du questionnaire et de 52 pour la version B.

Les analyses de la section précédente ont permis de constater que le retrait de l'item mal ajusté (item 13) permet d'obtenir des résultats assez similaires aux résultats obtenus précédemment avec l'échantillon de 20 items. La seule différence importante étant que toutes les courbes de probabilités des catégories de réponse possèdent maintenant un sommet distinct, ce qui n'était pas le cas précédemment pour la version A du questionnaire.

Le diagramme de la figure 31 (annexe D.2, p. 237) montre qu'effectivement, la dispersion des items est à peu près la même que celle obtenue initialement (fig. 15, p. 205). De la même façon, le diagramme de comparaison de la figure 32 (p. 238) est assez similaire à celui de la figure 16 (p. 206). Les courbes de probabilités de la figure 33 (p. 239) sont également très comparables à celles de la figure 17 (p. 207). Le diagramme en bâton de la figure 34 (p. 240) permet de voir que la valeur des différences entre les versions A et B obtenues pour l'échantillon de 19 items sont de moins de 1 % par rapport aux valeurs obtenues initialement pour l'échantillon de 20 items (fig. 18, p. 208). Ce constat permet de confirmer que les résultats obtenus en analysant les 19 ou les 20 items sont à peu près équivalents.

Enfin, le diagramme de la figure 35 (p. 246) montre que le retrait de l'item 13 n'a eu que très peu d'impact sur les catégories de réponse de l'échelle. Les résultats sont effectivement assez similaires à ceux de la figure 19 (p. 209). Pour la version A du questionnaire, il est toutefois possible de voir que la quatrième et la cinquième catégorie sont un peu plus rapprochées lorsque l'item 13 est retiré.

#### **6.2.4.3 Qualité de l'ajustement de la version A du questionnaire 2005**

L'analyse de la qualité de l'ajustement des items de la version A du questionnaire 2005 a permis de déceler que trois items, les items 1, 3 et 13, ne s'ajustaient pas bien au modèle. Les items possédant les pires indices d'ajustement standardisés ont été retirés de l'échantillon un à un. La qualité de l'ajustement était ensuite étudiée de nouveau. Ce processus a fait en sorte que certains items qui ne s'ajustaient pas bien au modèle initialement, se sont ajustés par la suite. De même, le processus a permis de faire ressortir de nouveaux items qui ne s'ajustaient plus au modèle.

À la fin du processus, deux items ont dû être retirés de l'échantillon (items 3 et 13). À la suite de quoi, la qualité de l'ajustement répondait au critère choisi (autant pour les items que pour les sujets) et la corrélation était positive pour tous les items et tous les sujets. Ainsi, dans l'analyse qui suit, l'échantillon retenu est constitué de 61 sujets et de 18 items. Les tableaux 49 et 50 (annexe D.3, p. 242 et 244) présentent les statistiques d'ajustement finales pour les personnes et pour les items obtenus lorsque les items 3 et 13 sont retirés de l'échantillon.

Le tableau 51 (p. 245) montre que la valeur moyenne estimée par le modèle pour le degré d'adhésion (opinion) des personnes est maintenant de 0,29. La valeur obtenue

précédemment en conservant les 20 items était de 0,24. L'indice de séparation réel du modèle est maintenant de 3,45 par rapport à la valeur de 3,62 obtenue précédemment. L'indice de séparation du modèle est maintenant de 3,76 par rapport à 3,9. Les indices de « fidélité » correspondants sont de 0,92 et 0,93 respectivement. Ces indices étaient initialement de 0,93 et 0,94.

Le tableau 52 (p. 245) indique que l'indice de séparation réel des items est de 4,35 par rapport à la valeur de 4,24 obtenue précédemment. L'indice de séparation du modèle est de 4,58 par rapport à 4,48. Les indices de « fidélité » correspondants sont de 0,95. Ce résultat est identique à ce qui avait été obtenu précédemment pour les 20 items. Ces résultats indiquent que la version A du questionnaire 2005 discrimine donc maintenant l'opinion des participants sur les six niveaux de l'échelle (par rapport aux cinq niveaux obtenus précédemment pour les 20 items).

Ainsi, les résultats obtenus à l'aide de l'échantillon de 18 items sont à peu près équivalents à ceux obtenus précédemment à partir de l'échantillon de 20 items. La figure 36 (p. 246) permet de constater qu'effectivement, la distribution des personnes sur le continuum de l'habileté et celle des items sur le continuum du niveau de difficulté est assez similaire à ce qui avait été obtenu précédemment (fig. 20, p. 221).

Le tableau 53 (p. 247) présente la valeur des points d'ancrage de l'échelle de réponses estimé par le modèle pour l'échantillon de 18 items. Les valeurs obtenues sont, encore une fois, assez similaires à celles obtenues avec l'échantillon de 20 items (tab. 42, p. 223). En revanche, l'analyse graphique des catégories de réponse (fig. 37, p. 247) montre que toutes les courbes de probabilités possèdent maintenant un sommet distinct et sont distancées de façon à peu près égale sur le continuum, ce qui n'était pas le cas précédemment.



#### **6.2.4.4 Qualité de l'ajustement de la version B du questionnaire 2005**

L'analyse de la qualité de l'ajustement des items de la version B du questionnaire 2005 a permis de déceler que trois items, les items 6, 8 et 19, ne s'ajustaient pas bien au modèle. Ainsi, les items possédant les pires indices d'ajustement standardisés ont été retirés de l'échantillon un à un. La qualité de l'ajustement était ensuite étudiée de nouveau. Ce processus a fait en sorte que certains items qui ne s'ajustaient pas bien au modèle initialement, se sont ajustés par la suite. De même, le processus a permis de faire ressortir de nouveaux items qui ne s'ajustaient plus au modèle. À la fin du processus, trois items ont dû être retirés de l'échantillon (items 6, 7 et 8).

En revanche, le retrait de ces items a fait apparaître de nouveaux participants qui ne s'ajustaient plus au modèle. Les sujets 543, 544, 580 et 620 ont donc également dû être retirés de l'échantillon. À la suite de quoi, la qualité de l'ajustement répondait au critère choisi (autant pour les items que pour les sujets) et la corrélation était positive pour tous les items et tous les sujets.

Par conséquent, dans l'analyse qui suit, l'échantillon retenu est constitué de 27 sujets et de 17 items. Les tableaux 54 et 55 (annexe D.4, p. 248 et 249) présentent les statistiques d'ajustement finales pour les personnes et pour les items résultant de cette analyse.

Le tableau 56 (p. 250) montre que la valeur moyenne estimée par le modèle pour le degré d'adhésion (opinion) des personnes est maintenant de -0,21. La valeur obtenue précédemment en conservant les 20 items était de -0,02. L'indice de séparation réel du modèle est maintenant de 6,35 par rapport à la valeur de 5,75 obtenue précédemment.

L'indice de séparation du modèle est maintenant de 6,78 par rapport à 6,13. Les indices de « fidélité » correspondants sont 0,98. La valeur de ces indices était de 0,97 initialement. Ainsi, la version B du questionnaire 2005 discrimine toujours l'opinion des participants sur huit niveaux différents. Le retrait des items mal ajustés n'a donc pas eu d'influence sur ce résultat.

Le tableau 57 (p. 250) indique que l'indice de séparation réel des items est de 3,42 par rapport à la valeur de 3,63 obtenue précédemment. L'indice de séparation du modèle est de 3,6 par rapport à 3,93. Les indices de « fidélité » correspondants sont de 0,92 et 0,93 respectivement. (Les valeurs obtenues précédemment étaient de 0,93 et 0,94.)

Les résultats obtenus à l'aide de l'échantillon de 17 items sont donc légèrement différents de ceux obtenus précédemment à partir de l'échantillon de 20 items. Ces différences sont un peu plus marquées pour les personnes que pour les items. La figure 38 (p. 251) permet de constater qu'effectivement, la distribution des personnes sur le continuum de l'habileté est plus grande qu'elle ne l'était initialement (fig. 21, p. 222). En revanche, la distribution des items sur le continuum du niveau de difficulté est assez similaire à ce qui avait été obtenu précédemment.

Le tableau 58 (p. 252) présente la valeur des points d'ancrage de l'échelle de réponses estimé par le modèle pour l'échantillon de 17 items. Les valeurs obtenues sont, légèrement différentes de celles obtenues avec l'échantillon de 20 items (tab. 43, p. 223). L'analyse graphique des catégories de réponse (fig. 37, p. 252) montre qu'effectivement toutes les courbes de probabilités possèdent maintenant un sommet distinct et distancé à peu près également sur le continuum, ce qui n'était pas le cas précédemment.

#### **6.2.4.5 Comparaisons entre les versions A et B du questionnaire 2005**

Les sections précédentes ont permis d'étudier la qualité de l'ajustement lorsque les items mal ajustés sont retirés pour les versions A et B du questionnaire 2005. Dans la présente section, ce sont les comparaisons entre les versions A et B du questionnaire 2005 qui sont présentées. Pour effectuer ces comparaisons, les items 3, 6, 7, 8 et 13 ont été exclus pour les deux versions du questionnaire.

L'analyse qui suit porte donc sur 15 items plutôt que 20 comme dans les comparaisons initiales effectuées pour le questionnaire 2005. De plus, il est important de rappeler que l'échantillon retenu est constitué de 61 sujets pour la version A du questionnaire et de 27 pour la version B.

Le diagramme de la figure 40 (annexe D.5, p. 253) montre que la dispersion des items est plus grande que celle qui avait été obtenue initialement en analysant les 20 items (fig. 24, p. 226). En effet, le nombre d'items qui s'éloignent de la droite de dispersion est plus élevé.

Le diagramme de la figure 41 (p. 254) permet de voir que la position de certains items sur le continuum varie effectivement beaucoup par rapport à ce qui avait été obtenu initialement (fig. 25, p. 227). Les courbes de distribution de la figure 42 (p. 255) confirment qu'il existe des différences importantes pour les items 1, 5, 9, 10, 11, 12, 19 et 20. (Il faut toutefois rappeler que l'item 20 a été complètement changé en 2005.)

Le diagramme en bâton de la figure 43 (p. 256) précise que la valeur de la différence entre la version B du questionnaire par rapport à la version A dépasse les 10 %

pour les items 1, 4, 11 et 19. Cette valeur atteint les 20 % pour les items 5, 10 et 20. La valeur de cette différence atteint même 27 % pour l'item 9.

Enfin, le diagramme de la figure 44 (p. 257) montre que les différences entre la version B et la version A du questionnaire 2005 lorsque seulement 15 items sont analysés sont également perceptibles à travers les points d'ancrage. Les points d'ancrage de la version A du questionnaire sont assez similaires à ce qui avait été obtenu en analysant les 20 items (fig. 28, p. 230). En revanche, les points d'ancrage de la version B sont beaucoup plus dilatés qu'initialement.

## 7. DISCUSSION

Le chapitre précédent a permis de présenter plusieurs résultats intéressants. Dans le présent chapitre, une interprétation de ces résultats est proposée. Or, pour y arriver, il serait utile de revenir sur les éléments importants qui ont été trouvés. Ainsi, la première section de ce chapitre présentera un résumé des principaux résultats obtenus lors de l'analyse des données. Une fois ce retour effectué, les résultats seront discutés afin d'en faire ressortir la signification. Enfin, un retour sur l'objectif de cette recherche ainsi que sur les questions auxquelles elle souhaite répondre sera effectué.

### ***7.1 Retour sur l'analyse des données***

Avant de présenter les principaux résultats obtenus dans le chapitre précédent, deux remarques s'imposent. Tout d'abord, il est important de rappeler qu'il s'agit ici d'une recherche exploratoire. Ainsi, les résultats de cette recherche permettront de découvrir les influences potentielles liées au choix des mots dans la formulation des items. En revanche, les résultats ne permettront pas d'établir les conséquences d'un choix de mots par rapport à un autre.

De plus, il est important de souligner que la taille des échantillons utilisés dans cette recherche est restreinte. La portée des résultats obtenus dans cette étude est donc limitée. En effet, le nombre de participants retenus pour l'analyse des données oscille entre 31 et 62 selon la version du questionnaire (voir tabl. 8, p. 139). Or, selon Linacre, chacune des catégories de réponse devrait posséder au moins 10 observations afin d'assurer une

certaine précision et une certaine stabilité dans le calibrage des points d'ancrage de l'échelle (sect. 6.1.2).

Ainsi, pour assurer une certaine précision et une certaine stabilité dans les résultats, il aurait fallu un minimum de 60 participants (10 participants x 6 catégories) pour chacune des versions du questionnaire (en considérant que les réponses des participants sont également distribuées à travers les différentes catégories de réponse de l'échelle). Ce minimum de 60 participants a été atteint pour les versions 2003B et 2005A. Toutefois, la distribution des réponses obtenues n'est pas nécessairement uniforme à travers les catégories de réponse de l'échelle. Les résultats obtenus dans cette recherche peuvent donc être instables et le calibrage de l'échelle de réponses imprécis.

Par conséquent, la portée de cette recherche est limitée et les résultats qui en découlent ne devraient pas être généralisés à d'autres situations. D'autres recherches sur le sujet seraient nécessaires pour confirmer les résultats trouvés dans ce travail.

Ces précisions étant faites, il est maintenant approprié de présenter les résultats principaux obtenus lors de l'analyse des données et résumés à la page suivante. Afin de présenter ces résultats, deux tableaux ont été créés. Le premier tableau (tabl. 8, p. 139) fait la synthèse des résultats de l'analyse de la qualité de l'ajustement pour les trois années étudiées (2003, 2004 et 2005). Le second tableau (tabl. 9, p. 140) présente, pour ces trois années, les éléments importants résultants de la comparaison entre les versions A et B des questionnaires. Ces tableaux seront ensuite discutés dans la prochaine section.

**Tableau 8 : Synthèse des principaux résultats de l'analyse de la qualité de l'ajustement pour les questionnaires 2003, 2004 et 2005**

		2003	2004	2005
Nombre d'items étudiés	A	20	20	20
	B	20	20	20
Nombre d'items modifiés	A	4	11	7
	B	4	11	7
Items modifiés	A	2-3-15-20	1-2-3-4-5-7-11-13-15-19-20	4-5-9-11-14-15-20
	B	2-3-15-20	1-2-3-4-5-7-11-13-15-19-20	4-5-9-11-14-15-20
Nombre de réponses recueillies	A	55	63	72
	B	69	62	41
Nombre de personnes retirées pour mauvais ajustement au modèle	A	4	12	11
	B	7	10	10
Nombre de participants retenus	A	51	51	61
	B	62	52	31
Nombre d'items qui ne s'ajustaient pas au modèle	A	0	2	3
	B	0	0	3
Items qui ne s'ajustaient pas au modèle	A	-	12-13	1-3-13
	B	-	-	6-8-19
Valeur moyenne estimée pour le degré d'adhésion des participants (opinion)	A	0,37	-0,76	0,24
	B	0,40	-0,36	-0,02
Nombre de niveaux d'opinion discriminés	A	4	5	5
	B	5	5	8
Valeur moyenne pour le niveau de difficulté des items (fixée)	A	0	0	0
	B	0	0	0
Nombre de niveaux de difficulté discriminés	A	7	7	5
	B	8	7	5
Items les plus faciles à endosser	A	8-10	8-12	8-12
	B	8-10-12	8-12	8-10
Items les plus difficiles à endosser	A	2-3-19-20	2-20	2
	B	2-3-19-20	2-20	3-19
Points d'ancrage qui répondaient au critère	A	2-3-4-5-6	2-3-4-6	2-3-4-5-6
	B	2-3-4-5-6	2-5-6	2-3-4-5-6
Nombre de catégories de réponse qu'il est possible d'observer	A	6	5 <sup>12</sup>	4
	B	6	6	4
Catégories potentiellement problématiques	A	4	1-6	1-6
	B	-	6	1-6

<sup>12</sup> Il est toutefois à noter que la probabilité d'observer la première catégorie est très petite.

**Tableau 9 : Synthèse des principaux résultats de la comparaison entre les versions A et B des questionnaires 2003, 2004 et 2005<sup>13</sup>**

	2003	2004	2005
<b>Items dont la variation entre les deux versions dépasse les 10 %</b>	8	5	5-9-20
<b>Items dont la variation entre les deux versions se situe entre 5 % et 10 %</b>	5-12-16	2-7-11-12-18-20	3-6-7-8-10-11-19
<b>Items dont la variation entre les deux versions est de moins de 5 %</b>	1-2-6-7-9-10-11-13-14-15-17-18-19-20	1-3-4-6-8-9-10-13-14-16-17-19	1-2-4-12-13-14-15-16-17-18
<b>Items dont la variation entre les deux versions est nulle</b>	3	15	-
<b>Items dont les points d'ancrage varient le plus entre les des deux versions</b>	8	(à peu près tous les items)	5-8-9-20
<b>Points d'ancrage dont l'intervalle est d'au moins 1 logit mais de moins de 5</b>	tous	2-6	tous
<b>Points d'ancrage dont l'intervalle est de 1 logit ou moins</b>	-	3-4-5	-
<b>Points d'ancrage généralement plus rapprochés que les autres</b>	4-5	3-4-5	3-4
<b>Points d'ancrage généralement plus éloignés que les autres</b>	6	2-6	-

<sup>13</sup> Dans ce tableau, les items mis en caractères gras sont plus difficiles à endosser dans la version B du questionnaire que dans la version A. (Les autres, sont plus difficiles dans la version A.)



## ***7.2 Interprétation des résultats***

La discussion proposée dans cette section sera divisée en quatre parties. Tout d'abord, les principaux résultats de l'analyse de la qualité de l'ajustement (tabl. 8, p. 139) seront examinés afin de faire ressortir les caractéristiques globales des différents questionnaires. Ensuite, pour chacune des années du questionnaire étudiées dans cette recherche, les résultats de la comparaison entre les versions A et B des questionnaires seront analysés afin d'essayer d'en décoder le sens.

Il est important de noter que les interprétations proposées dans cette section ne sont que des hypothèses. Ainsi, malgré le ton parfois affirmatif utilisé dans la discussion des résultats, les interprétations fournies ne constituent tout de même que des hypothèses formulées dans le but d'expliquer les résultats.

Enfin, avant de conclure ce travail, une dernière section proposera un retour sur les analyses complémentaires effectuées au chapitre précédent. Cependant, puisque l'objectif de cette recherche consiste à explorer l'influence de la formulation des items sur les réponses des participants à partir des 20 items provenant du questionnaire utilisé par le CFIM pour l'évaluation de ses programmes et non pas à étudier les résultats obtenus lorsque les items mal ajustés sont retirés, la discussion proposée sera brève.

### 7.2.1 Résultats de l'analyse de la qualité de l'ajustement

La section 7.1 a permis de préciser qu'un nombre minimum de 60 participants aurait été nécessaire pour assurer des résultats stables et précis. Cependant, les résultats de cette recherche signalent qu'il serait possible d'obtenir une certaine stabilité dans les résultats même si le nombre de participants est un peu moindre. Les données suivantes, tirées du tableau 8 (p. 139), permettent d'expliquer cette hypothèse.

En effet, le tableau 8 montre que l'analyse de la qualité de l'ajustement des items pour les questionnaires 2003 avait révélé que tous les items s'ajustaient bien au modèle, que tous les points d'ancrage des catégories de l'échelle répondaient au critère choisi et qu'il était possible d'observer les six catégories de l'échelle de réponses. Seule la quatrième catégorie était légèrement moins probable que les autres dans la version A du questionnaire. Or, seulement 51 sujets ont été retenus pour la version A du questionnaire par rapport à 62 pour la version B. De plus, seulement 4 items parmi les 20 items du questionnaire du CFIM retenus pour cette étude ont été modifiés.

En revanche, pour les questionnaires 2004, deux items de la version A du questionnaire ne s'ajustaient pas bien au modèle et certains points d'ancrage de l'échelle ne répondaient pas au critère choisi. L'analyse avait révélé qu'il était seulement possible d'observer cinq des six catégories de l'échelle. Dans la version B du questionnaire, certains problèmes avaient également été décelés, en particulier pour les catégories aux deux extrémités de l'échelle. Or, en 2004, seulement 51 sujets ont été retenus pour la version A du questionnaire et 52 pour la version B. De plus, la formulation de 11 items parmi les 20 ont été modifiés.

Enfin, pour les questionnaires 2005, trois items ne s'ajustaient pas bien au modèle pour les deux versions du questionnaire de 2005. Tous les points d'ancrage répondaient au critère choisi, mais l'analyse avait révélé qu'il était seulement possible d'observer quatre des six catégories de l'échelle. En effet, la première et la sixième catégorie n'avaient jamais plus de chance que les autres d'être observées. Or, en 2005, 61 sujets ont été retenus pour la version B du questionnaire par rapport à seulement 31 sujets pour la version A.

Ainsi, les résultats obtenus en 2003 indiquent qu'il a été possible d'obtenir une bonne qualité d'ajustement alors que seulement 51 sujets ont été retenus pour la version A du questionnaire. En comparaison, la qualité de l'ajustement des questionnaires 2004 était moins bonne. Le nombre de sujets retenus en 2004 était pourtant équivalent au nombre de sujets retenus pour l'année 2003. Par contre, le nombre d'items ayant été modifiés en 2004 était beaucoup plus élevés que celui de 2003.

De même, la qualité de l'ajustement était moins bonne en 2005 qu'en 2003. Le nombre de sujets retenus pour la version A du questionnaire était pourtant de 61 sujets, ce qui correspond au nombre minimum suggéré précédemment. En revanche, le nombre de sujets retenus pour la version B du questionnaire était seulement de 31 sujets. Enfin, le nombre d'items ayant été modifiés pour l'année 2005 était plus grand que celui de 2003, mais moindre que celui de 2004.

Les comparaisons des paragraphes précédents soulèvent une question. Serait-il possible d'obtenir une certaine précision dans les résultats avec un échantillon d'au moins 50 sujets si le nombre d'items modifiés est petit? Cette hypothèse serait à vérifier par d'autres recherches.

Le tableau 8 de la page 125 révèle également que la valeur moyenne estimée par le modèle pour le degré d'adhésion des participants (opinion) est plus élevé dans la version B du questionnaire que dans la version A pour les années 2003 et 2004. La situation inverse se produit en 2005. Ainsi, il était un peu plus facile d'avoir une opinion favorable dans la version A du questionnaire pour les deux premières années et dans la version B en 2005. De façon générale, en revanche, il était un peu plus facile d'avoir une opinion favorable avec les questionnaires 2004, 2005 et 2003 (dans cet ordre).

Les résultats de l'analyse montrent par ailleurs que la version A du questionnaire ne discrimine que quatre niveaux d'opinion par rapport aux six niveaux proposés dans l'échelle de réponses en 2003, mais cinq niveaux pour les questionnaires 2004 et 2005. La version B du questionnaire discriminait l'opinion sur cinq niveaux en 2003 et en 2004, mais sur huit niveaux en 2005.

Enfin, le tableau 8 de la page 125 montre que la version A du questionnaire permet de discriminer le niveau de difficulté des items (difficulté à endosser) sur au moins sept niveaux en 2003 et 2004, mais sur seulement cinq niveaux en 2005. La version B du questionnaire permettait de discriminer le niveau de difficulté des items sur au moins huit niveaux en 2003, sur sept niveaux en 2004, mais sur seulement cinq niveaux en 2005. Par conséquent, les informations fournies par les items de la version 2005 des questionnaires sont un peu plus redondantes que pour les années précédentes.

### **7.2.2 Résultats pour les questionnaires 2003**

La comparaison entre les versions A et B des questionnaires 2003 a dévoilé que, de façon générale, les items variaient peu d'une version à l'autre du questionnaire sur le

continuum du niveau de difficulté. En effet, pour 14 des 20 items étudiés (items 1, 2, 6, 7, 10, 11, 13, 14, 15, 17, 18, 19 et 20), la variation observée était de moins de 5 %. La variation de l'item 3 était nulle. Un seul item, l'item 8, variait beaucoup (variation supérieure à 10 %) entre les deux versions du questionnaire. Une variation moindre, mais tout de même notable, a également pu être observée pour les items 5, 12 et 16.

En 2003, la formulation de 4 items a été modifiée (items 2, 3 15 et 20). Parmi ces items, la position de l'item 3 est identique pour les deux versions du questionnaire. Ainsi, le niveau de difficulté exigé pour endosser ces items est le même. La formulation utilisée dans la version A, « Construire des outils pour l'évaluation sommative (contrôles, examens, etc.) », est plus générale que celle utilisée dans la version B, « Construire des examens écrits pour l'évaluation sommative ». Or, il est généralement plus facile pour les participants d'avoir une opinion favorable envers un item dont la formulation est plus générale. Ce n'est cependant pas le cas ici. En analysant un peu plus ces deux versions, il apparaît que les deux formulations réfèrent à un même construit, c'est-à-dire la construction d'examens pour l'évaluation sommative. De plus, les résultats révèlent que le niveau de difficulté requis pour endosser ces items est le même. Le sens de ces deux items est donc équivalent<sup>14</sup>.

Ensuite, les items 2, 25 et 20 sont plus difficiles à endosser dans la version B du questionnaire que dans la version A, mais la différence est petite (variation de moins de 5 %). La formulation utilisée dans la version A pour l'item 2, « Répondre aux questions des parents lors de la présentation du bulletin », est plus générale que celle utilisée dans la version B, « Aider les parents à comprendre le bulletin de leur enfant ». Ainsi, la précision

---

<sup>14</sup> Il est important de rappeler que les interprétations proposées dans cette section ne sont que des hypothèses. Ainsi, malgré le ton parfois affirmatif utilisé, les interprétations fournies ne constituent tout de même que des hypothèses formulées dans le but d'expliquer les résultats.

« Aider les parents » est plus difficile à endosser, car elle demande plus de la part des participants que le fait de répondre aux questions qui leur sont posées.

Pour l'item 15, la formulation utilisée dans la version A, « Construire des outils pour l'évaluation formative (exercices, devoirs, etc.) », est plus précise que celle utilisée dans la version B, « Planifier l'évaluation formative ». Dans ce cas-ci, la précision « Construire des outils » pour l'évaluation formative est plus facile à endosser que le fait de la planifier. En effet, planifier suggère que les participants doivent déterminer à l'avance les évaluations formatives. Or, en situation de classe, cette tâche peut s'avérer assez ardue puisque l'évaluation formative est généralement adaptée aux situations d'apprentissage, aux difficultés rencontrées, etc.

Pour l'item 20, la formulation utilisée dans la version A, « Discuter avec les parents des difficultés de leur enfant », est plus générale que celle utilisée dans la version B, « Préparer une rencontre avec les parents d'un élève ayant des difficultés ». Comme pour l'item 2, la précision « préparer une rencontre » est plus difficile à endosser, car elle demande plus de la part des participants que le fait de discuter. En effet, discuter est un processus d'échange moins formel que celui impliqué par la préparation d'une rencontre qui exige de planifier le contenu de la rencontre, de prendre rendez-vous, etc.

Enfin, les variations constatées pour les items 5, 8, 12 ou 16 sont plutôt surprenantes et difficiles à expliquer. En effet, ni la formulation, ni l'ordre de présentation des items, ni l'échelle de réponses n'ont été modifiés pour ces items entre les deux versions du questionnaire. Or, il est également à noter que les variations constatées pour ces items sont seulement de l'ordre de 10 %. Ainsi, plutôt que de s'interroger sur les causes de ces variations, il apparaît pertinent de se demander à quel point une telle variation peut avoir des conséquences sur les résultats d'enquête et donc avoir un impact pour le chercheur.

### 7.2.3 Résultats pour les questionnaires 2004

La comparaison entre les versions A et B des questionnaires 2004 a dévoilé une plus grande variation que celle observée en 2003. Les résultats de l'analyse ont permis de constater que la variation observée pour 12 des 20 items étudiés (items 1, 3, 4, 6, 8, 9, 10, 13, 14, 16, 17 et 19) était de moins de 5 %. La variation de l'item 15 était nulle. Un seul item, l'item 5, variait beaucoup (variation supérieure à 10 %) entre les deux versions du questionnaire. En revanche, une variation notable (entre 5 % et 10 %) a été observée pour six items (les items 2, 7, 11, 12, 18 et 20).

En 2004, la formulation de 11 items a été modifiée (items 1, 2, 3, 4, 5, 7, 11, 13, 15, 19 et 20). La formulation utilisée pour l'item 1 dans la version A du questionnaire, « Identifier les contenus difficiles à faire apprendre aux élèves », est plus précise que celle utilisée dans la version B, « Identifier les contenus plus difficiles pour les élèves ». Or, la formulation de cet item est légèrement plus difficile à endosser dans la version B que dans la version A. Cependant, la différence est de l'ordre de 1 %. Par conséquent, la différence entre les deux formulations n'était pas suffisamment importante, par rapport au construit visé, pour changer le sens de l'item entre les deux versions du questionnaire et avoir un impact notable sur les réponses des participants.

Pour l'item 2, la formulation utilisée dans la version A, « Répondre aux questions des parents lors de la présentation du bulletin », est plus précise que celle utilisée dans la version B, « Répondre aux questions lors des rencontres avec les parents ». Dans ce cas-ci, la formulation de la version A est plus difficile à endosser que celle de la version B et la différence est notable. La précision sur « la présentation du bulletin » semble imposer un

aspect plus formel à la rencontre et donc plus difficile à endosser que lors des rencontres habituelles avec les parents.

Pour l'item 3, la formulation utilisée dans la version A, « Construire des outils pour l'évaluation sommative (contrôles, examens), etc. », est plus précise que celle utilisée dans la version B, « Élaborer des outils pour l'évaluation sommative ». La version B de cet item est plus difficile à endosser que la version A, mais la différence est de moins de 5 %. Une des explications potentielles de ce résultat est que la forme générale de la version B ne permettrait pas aux participants de déterminer facilement en quoi consiste l'élaboration des outils ni de quels outils il est question. Ainsi, il serait alors plus difficile pour les participants d'adhérer à cet item que lorsque des précisions sont ajoutées à la formulation.

Pour l'item 4, la formulation utilisée dans la version A, « Maîtriser les contenus que j'enseignerai en conformité avec les programmes du ministère de l'Éducation », est plus précise que celle utilisée dans la version B, « Maîtriser les contenus que j'enseignerai ». La version B de cet item est plus difficile à endosser que la version A, mais la différence est de moins de 5 %. La version B est plus difficile à endosser, car elle implique que le participants doit maîtriser tous les contenus qu'il aura à enseigner. La précision « en conformité avec les programmes du ministère de l'Éducation » implique que seuls les contenus du programme du ministère sont à maîtriser, ce qui est plus facile à endosser.

Pour l'item 5, la formulation utilisée dans la version A, « Corriger la langue écrite des élèves », est plus précise que celle utilisée dans la version B, « Corriger les productions écrites des élèves ». La version B de cet item est beaucoup plus difficile à endosser que la version A. La différence observée est près de 20 %. Or, en analysant ces deux formulations, il apparaît que le construit visé n'est pas tout à fait le même. Le sens de ces deux items n'est donc pas le même. En effet, « corriger les productions écrites » implique autant de corriger la langue écrite des élèves, mais aussi la production en elle-



même. Ainsi, la tâche exigée par la version B est plus laborieuse et donc plus difficile à endosser par les participants.

Pour l'item 7, la formulation utilisée dans la version A, « Intervenir individuellement auprès des élèves à risque d'échouer », est plus précise que celle utilisée dans la version B, « Planifier une intervention auprès d'un élève en difficulté ». La formulation de la version A est plus difficile à endosser que celle de la version B et la différence est notable. En effet, « planifier une intervention » implique d'établir les stratégies les plus pertinentes, de prévoir un rendez-vous, etc. En revanche, l'intervention auprès de l'élève implique qu'il faut utiliser la meilleure stratégie, ce qui est plus difficile à endosser.

Pour l'item 11, la formulation utilisée dans la version A, « Motiver les élèves à s'engager dans leur apprentissage », est plus précise que celle utilisée dans la version B, « Motiver les élèves dans leur apprentissage ». La formulation de la version A est plus difficile à endosser que celle de la version B et la différence est notable. En effet, « motiver » implique de stimuler, d'encourager ou d'intéresser les élèves à leur apprentissage. En revanche, « motiver les élèves à s'engager » implique que la motivation fournie doit être assez forte pour amener l'élève à jouer un rôle concret dans ses apprentissages et à maintenir ses efforts. Cette version est donc plus difficile à endosser par les participants.

Pour l'item 13, la formulation utilisée dans la version A, « Identifier les points forts et les points faibles des élèves », est plus précise que celle utilisée dans la version B, « Reconnaître les points forts et les points faibles des élèves ». La formulation de la version A est plus difficile à endosser que celle de la version B, mais la différence est petite. Or, « reconnaître » implique de percevoir les forces et les faiblesses des étudiants. En revanche, « identifier » consiste à établir la nature de ces points forts ou de ces points faibles. Cette version est donc plus difficile à endosser par les participants.

Pour l'item 15, la formulation utilisée dans la version A, « Construire des outils pour l'évaluation formative (exercices, devoirs, etc.) », est plus précise que celle utilisée dans la version B, « Élaborer des outils pour l'évaluation formative ». La position de cet item est identique pour les deux versions du questionnaire. Or, l'analyse effectuée pour l'item 3 avait révélé que la forme générale de la version B ne permettait pas aux participants de déterminer facilement en quoi consiste l'élaboration des outils, ni de quels outils il est question et qu'il était alors un plus difficile pour les participants d'adhérer à cet item. L'analyse avait également permis de montrer que la différence était toutefois petite. L'analyse de l'item 15 confirme donc que l'influence de ce choix de mots est minime, voire nulle.

Pour l'item 19, la formulation utilisée dans la version A, « Orienter les élèves vers les services d'aide appropriés », est plus précise que celle utilisée dans la version B, « Suggérer aux élèves des services d'aide appropriés ». La formulation de la version A est plus difficile à endosser que celle de la version B, mais la différence est de moins de 5 %. En effet, « suggérer des services » implique de faire des propositions, de conseiller. En revanche, « orienter » implique d'aider l'élève et de le diriger vers les bons services. Cette version est donc plus difficile à endosser par les participants.

Enfin, pour l'item 20, la formulation utilisée dans la version A, « Discuter avec les parents des difficultés de leur enfant », est plus générale que celle utilisée dans la version B, « Expliquer aux parents quelles sont les difficultés de leur enfant ». La formulation de la version A est plus difficile à endosser que celle de la version B et la différence est notable. En effet, « discuter des difficultés » implique un processus d'échange où des questions auxquelles le participant ne serait pas préparé (à propos des difficultés de l'enfant) pourraient survenir. En revanche, « expliquer les difficultés »

implique que le participant fait connaître les difficultés qu'il a identifiées. Cette version est donc plus facile à endosser par les participants.

## 7.2.4 Résultats pour les questionnaires 2005

La comparaison entre les versions A et B des questionnaires 2005 a dévoilé une plus grande variation que celle observée en 2003 ou en 2004. Les résultats de l'analyse ont permis de constater que la variation observée pour 10 des 20 items étudiés (items 1, 2, 4, 12, 13, 14, 15, 16, 17 et 18) était de moins de 5 %. Trois items, les items 5, 9 et 20, variaient beaucoup (variation supérieure à 10 %) entre les deux versions du questionnaire. De plus, une variation notable (entre 5 % et 10 %) a été observée pour sept items (les items 3, 6, 7, 8, 10, 11 et 19).

En 2005, la formulation de sept items a été modifiée (items 4, 5, 9, 11, 14, 15 et 20). La formulation utilisée pour l'item 4 dans la version A du questionnaire, « Maîtriser les contenus que j'enseignerai en conformité avec les programmes du ministère de l'Éducation », est plus précise que celle utilisée dans la version B, « Maîtriser les contenus que j'enseignerai ». La formulation de la version A est plus difficile à endosser que celle de la version B, mais la différence est de moins de 5 %. Ce constat est intéressant puisqu'il vient contredire les résultats observés pour ce même item en 2004.

Pour l'item 5, la formulation utilisée dans la version A, « Corriger la langue écrite des élèves », est plus précise que celle utilisée dans la version B, « Corriger les productions écrites des élèves ». Encore une fois, la version B de cet item est beaucoup plus difficile à endosser que la version A. La différence observée est de l'ordre de 20 %. Ce résultat vient donc soutenir celui observé en 2004.

Pour l'item 9, la formulation utilisée dans la version A, « Adapter mes activités d'enseignement aux caractéristiques des élèves », est plus précise que celle utilisée dans la version B, « Adapter mes activités d'enseignement ». La formulation de la version A est plus difficile à endosser que celle de la version B et la différence est notable. En effet, « adapter » les activités d'enseignement impliquent une action qu'un enseignant doit continuellement réaliser pour ajuster son enseignement au rythme des élèves, à la difficulté des contenus, à l'horaire du cours, etc. En revanche, adapter les activités aux caractéristiques des élèves implique que les activités doivent être ajustées en fonction des traits propres aux élèves. Or, ces traits peuvent être d'ordre ethnique ou culturel, être liés à l'habileté des élèves, etc. Cette version est donc plus difficile à endosser par les participants.

Pour l'item 11, la formulation utilisée dans la version A, « Motiver les élèves à s'engager dans leur apprentissage », est plus précise que celle utilisée dans la version B, « Maintenir un climat d'apprentissage en classe ». La formulation de la version A est plus difficile à endosser que celle de la version B et la différence est notable. En effet, « maintenir un climat d'apprentissage en classe » est un des objectifs qu'un enseignant s'efforce généralement d'atteindre. En revanche, « motiver les élèves à s'engager » implique que l'enseignant doit fournir un effort supplémentaire afin d'amener l'élève à jouer un rôle concret dans ses apprentissages et à maintenir ses efforts. Cette version est donc plus difficile à endosser par les participants.

Pour l'item 14, la formulation utilisée dans la version A, « Collaborer avec les autres enseignants et enseignantes », est plus précise que celle utilisée dans la version B, « Collaborer avec les membres de l'équipe-école ». La formulation de la version B est plus difficile à endosser que celle de la version A, mais la différence est petite. Collaborer avec les membres de l'équipe-école implique que l'enseignant doit collaborer avec le

psychologue de l'école, l'intervenant social, le directeur, pour n'en nommer que quelques-uns, de même qu'avec les autres enseignants. Cette version est donc un peu plus difficile à endosser par les participants.

Pour l'item 15, la formulation utilisée dans la version A, « Construire des outils pour l'évaluation formative (exercices, devoirs, etc.) », est plus précise que celle utilisée dans la version B, « Élaborer des outils pour l'évaluation formative (exercices, devoirs, etc.) ». La formulation de la version B est plus difficile à endosser que celle de la version A, mais la différence est petite. Or, « construire » un outil d'évaluation consiste à créer l'outil en soit. En revanche, « élaborer » un outil implique qu'il faut planifier, préparer et ensuite construire l'outil. Cette version est donc un peu plus difficile à endosser par les participants.

Enfin, pour l'item 20, la formulation utilisée dans la version A, « Guider les élèves dans leurs apprentissage », est différente de celle utilisée dans la version B, « Discuter avec les élèves de leurs difficultés ». La formulation de la version B est plus difficile à endosser que celle de la version A et la différence est importante (de l'ordre de 20 %). Toutefois, le construit visé par ces deux versions n'est pas le même. Il serait donc délicat d'essayer d'établir une comparaison entre les deux.

### **7.2.5 Résultats pour les analyses complémentaires**

L'analyse de la qualité de l'ajustement des versions A des questionnaires 2004 et 2005, effectuée lorsque les items mal ajustés étaient retirés de l'échantillon, ont donné des résultats assez similaires à ceux obtenus pour l'échantillon de 20 items. La différence principale observée était que les courbes de probabilités des catégories de réponse possédaient un sommet distinct, ce qui n'était pas le cas initialement.

L'analyse de la qualité de l'ajustement de la version B du questionnaire 2005, a également permis de constater que les courbes de probabilités des catégories de réponse obtenues possédaient maintenant un sommet distinct et distancé à peu près uniformément sur le continuum, ce qui n'était pas le cas pour l'échantillon de 20 items.

En revanche, les résultats obtenus pour la version B du questionnaire 2005 étaient légèrement différents de ceux obtenus initialement. Ainsi, les comparaisons entre les versions A et B du questionnaire (effectuées une fois les items mal ajustés retirés de l'échantillon), ont permis de constater des différences dépassant les 10 %, ce qui n'était pas le cas pour l'analyse effectuée sur l'échantillon de 20 items. Il est à noter que, dans les analyses complémentaires de la version B du questionnaire 2005, seulement 27 sujets avaient été retenus.

Les comparaisons effectuées entre les versions A et B du questionnaire 2004 ont donné des résultats assez similaires à ceux obtenus avec l'échantillon de 20 items. En effet, les comparaisons entre les versions A et B du questionnaire (effectuées une fois l'item mal ajusté retiré), ont révélé des différences de moins de 1 % par rapport aux résultats qui avaient été obtenus initialement.

Par conséquent, les résultats des analyses complémentaires suggèrent que les différences importantes observées pour la version B du questionnaire 2005 seraient davantage liées à l'échantillon restreint utilisé pour cette version du questionnaire qu'à la qualité de l'ajustement des items.

### **7.3 Retour sur l'objectif et les questions de recherche**

L'objectif général de cette recherche consistait à étudier jusqu'à quel point des changements de mots apportés aux items d'un questionnaire d'enquête pouvaient influencer la modélisation des réponses fournies par les répondants et ainsi modifier l'échelle de mesure associée aux catégories de réponse.

Une partie de cet objectif a été atteint à travers le cadre conceptuel (chap. 3). En effet, le chapitre 3 a permis de préciser que les changements de mots apportés aux items d'un questionnaire pouvaient influencer les réponses des participants en modifiant le cadre de référence à utiliser pour contribuer de façon appropriée au processus d'échange d'information enclenché par le questionnaire. De plus, le chapitre 3 a permis de préciser que les changements de mots apportés aux items pouvaient modifier l'échelle de mesure en déplaçant les points d'ancrage de ses catégories.

Cependant, une autre partie de cet objectif consistait à répondre aux deux questions de recherche qui ont été formulées :

1. Dans quelle mesure, un changement de mots dans la formulation des items d'un questionnaire sur l'opinion (sans changer le sens général de la question) a-t-il une influence sur les réponses fournies par les répondants?
2. Dans quelle mesure, ce changement de mots a-t-il une influence sur les points d'ancrage de l'échelle de mesure associée aux catégories de réponse?

C'est à ces questions que les prochaines sections tenteront de répondre. Pour ce faire, le modèle de mesure *Rating Scale* a été mis à contribution dans la modélisation des

données recueillies à partir d'un questionnaire sur l'opinion dans lequel la formulation des items a été modifiée. Le sens général des items a été conservé, mais un changement de mots a été appliqué.

### **7.3.1 Retour sur la première question de recherche**

Pour répondre à la première question de recherche, les mesures estimées par le modèle pour le niveau de difficulté des items (difficulté à endosser) ont été examinées. Or, les résultats ont permis de révéler que, de façon générale, ce n'est pas tant la formulation plus ou moins précise utilisée pour désigner un même construit qui influence les réponses des participants, mais que c'est plutôt le sens des termes qui le composent qui influence les réponses recueillies.

Par exemple, « Identifier les points forts et les points faibles des élèves » est une formulation plus précise que « Reconnaître les points forts et les points faibles des élèves ». Dans ce cas-ci, les résultats ont révélé que la première version était plus difficile à endosser que la seconde.

Voici un autre exemple. « Corriger la langue écrite des élèves » est une formulation plus précise que « Corriger les productions écrites des élèves ». Toutefois, dans ce cas-ci, les résultats ont révélé que la seconde version était plus difficile à endosser que la première.

Or, en analysant ces deux exemples, il est possible de constater que le sens de l'item est un peu différent d'une formulation à l'autre. En effet, le sens de « reconnaître » n'est pas le même que celui d'« identifier ». « Reconnaître » implique simplement de



percevoir (les forces et les faiblesses des étudiants en l'occurrence). En revanche, « identifier » consiste à établir la nature de ces points forts ou de ces points faibles. Ainsi, dans ce cas-ci, la formulation plus précise est donc plus difficile à endosser par les participants.

À l'inverse, « corriger les productions écrites » implique autant de corriger la « langue écrite des élèves » que la production en elle-même. Ainsi, dans cet exemple, la formulation plus générale est la plus difficile à endosser par les participants.

Par conséquent, comme la section 3.2.5 l'avait mentionné, le choix des mots utilisé dans la formulation des items d'un questionnaire est important, car il permet aux répondants de déterminer la perspective (le sens) de ce qui leur est demandé et c'est cette perspective qui influence les réponses des participants. Enfin, les résultats de cette recherche ont permis d'observer que les variations produites par les changements de mots étaient généralement de l'ordre de 5 % à 10 % sur le continuum du niveau de difficulté des items.

### **7.3.2 Retour sur la deuxième question de recherche**

Pour répondre à la deuxième question de recherche, les points d'ancrage des catégories de réponse estimés par le modèle ont été examinées. Or, les variations constatées entre les deux versions du questionnaire étaient généralement similaires pour tous les items, c'est-à-dire autant pour les items dont la formulation avait été modifiée que pour les items qui n'avaient pas changé. Par conséquent, les résultats de cette recherche étaient peu concluants et ne nous ont pas permis d'établir dans quelle mesure un

changement de mots dans la formulation des items influence l'échelle de mesure associées aux catégories de réponse.

En revanche, les résultats ont montré que certaines catégories de réponse étaient parfois plus rapprochées que d'autres. Par exemple, les catégories 4 et 5 des questionnaires 2003 étaient plus rapprochées que les autres. Ainsi, les résultats de cette recherche ont permis d'observer que le nombre ou la définition des catégories de réponse offertes n'était peut-être pas assez clair pour que les participants puissent bien les distinguer.

## CONCLUSION

En conclusion, les résultats de cette recherche nous ont permis de découvrir que le choix des mots utilisé dans la formulation des items d'un questionnaire est important, car il permet aux répondants de déterminer la perspective (le sens) de ce qui leur est demandé et que c'est cette perspective qui influence les réponses des participants. De plus, les résultats ont permis de constater que les variations produites par les changements de mots étaient généralement de l'ordre de 5 % à 10 % sur le continuum du niveau de difficulté des items. Ces résultats nous ont donc permis de répondre à notre première question de recherche.

En revanche, les variations constatées entre les points d'ancrage des deux versions du questionnaire ne nous ont pas permis de répondre à notre deuxième question de recherche. En effet, les variations étaient généralement similaires pour tous les items. En conséquence, les résultats n'étaient pas concluants. Il serait donc intéressant de reprendre les analyses de cette recherche en utilisant le modèle *Partial Credit* pour voir si des conclusions différentes pourraient être dégagées. De même, il serait intéressant de reprendre les analyses en diminuant le nombre de catégories de réponse de l'échelle puisque la définition de certaines catégories ne semblait pas suffisamment claire pour que les participants puissent bien les distinguer.

Par ailleurs, l'analyse des résultats de cette recherche nous amène à soulever quelques questions intéressantes. Tout d'abord, l'analyse des résultats a permis de constater qu'il serait possible d'obtenir une certaine précision dans les résultats avec un échantillon d'au moins 50 sujets (plutôt que 60) lorsque le nombre d'items modifiés est

petit. Ce constat fournit une base méthodologique intéressante pour de futures recherches sur le sujet. L'hypothèse reste cependant à être confirmée.

Ensuite, l'analyse des résultats de cette recherche a permis de faire ressortir qu'il serait pertinent de vérifier jusqu'à quel point une variation de 10 % sur le continuum du niveau de difficulté peut avoir des conséquences sur les résultats d'une enquête par questionnaire et, par suite, avoir un impact sur les résultats de recherche du chercheur.

Enfin, il est important de rappeler qu'il s'agissait d'une recherche exploratoire effectuée avec un nombre limité de participants. Par conséquent, les résultats de cette recherche ne devraient pas être généralisés à d'autres situations. De plus, il serait pertinent que les résultats de cette recherche soient confirmés par d'autres études.

## RÉFÉRENCES

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561-573.
- Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR-20 index, and the Guttman scale response pattern. *Education research and perspectives*, 9(1), 95-104.
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage Publications.
- Bertrand, R. & Blais, J.-G. (2004). *Modèles de mesure : l'apport de la théorie des réponses aux items. (Avec la collaboration de Gilles Raïche.)*. Sainte-Foy: Presses de l'Université du Québec.
- Blais, A. & Durand, C. (1997). Le sondage. Dans B. Gauthier (dir.), *Recherche sociale : de la problématique à la collecte des données* (pp. 357-400). Sainte-Foy: Presses de l'Université du Québec.
- Blais, J.-G. (1987). *Effets de la violation du postulat d'unidimensionnalité dans la théorie des réponses aux items*. Thèse de doctorat en éducation non publiée, Université de Montréal, Montréal.
- Blais, J.-G. & Ajar, D. (1992). Théorie des réponses aux items et modélisation. *Mesure et évaluation en éducation*, 14(4), 5-18.
- Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch model : fundamental measurement in the human sciences*. Mahwah, N.J.: Lawrence Erlbaum Associates.

- Bradburn, N. M. & Sudman, S. (1979). *Improving interview method and questionnaire design : response effects to threatenng questions in survey research*. San Francisco: Jossey-Bass Publications.
- Bradley, K. D. & Sampson, S. O. (2005). *Validating an assessment through Rasch techniques: the process of measurement, feedback, reflection and change*. Compte rendu de la réunion annuelle du Mid-Western Educational Research Association (MWERA) tenue à Colombus en Ohio. Page consultée le 17 mai 2006. <http://www.uky.edu/~kdbrad2/ValidatingAnAssessment.pdf>.
- Clark, H. H. & Schober, M. F. (1992). Asking questions and influencing answers. Dans J. M. Tanur (Éd.), *Questions about questions* (pp. 15-48). NY: Russel Sage foundation.
- Daniel, M. H. (1999). Behind the sciences: using new measurement methods on DAS and KAIT. Dans S. E. Embretson & S. L. Hershberger (Éds.), *The new rules of measurement : what every psychologist and educator should know* (pp. 37-64). Mahwah, N.J.: Lawrence Erlbaum Associates.
- De Landsheere, G. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation, avec lexique anglais-français* (1<sup>re</sup> éd.). Paris: Presses universitaires de France.
- Druide. (2005). Dictionnaire du logiciel Antidote prisme (Version 6): Druides informatique.
- Fisher Jr., W. (1992). Reliability statistics. *Rasch measurement transactions*, 6(3), 238.
- GDT. (2005). *Grand dictionnaire terminologique*, sur le site de l'Office québécois de la langue française. Page consultée entre août et décembre 2005. [http://www.granddictionnaire.com/btml/fra/r\\_MotClef/index800\\_1.asp](http://www.granddictionnaire.com/btml/fra/r_MotClef/index800_1.asp)
- Grice, H. P. (1975). Logic and conversation. Dans P. Cole & J. L. Morgan (Éds.), *Syntax and semantics: speech acts* (Vol. 3, pp. 41-58). NY: Academic Press.

- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory : principles and applications*. Hingham, MA: Kluwer-Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Newbury Park, CA: SAGE Publications.
- Jackson, T. R. & Popovich, N. G. (2006). The development, implementation, and evaluation of a self-assessment instrument for use in a pharmacy student competition. *American journal of pharmaceutical education*, 67(2), article 57.
- Jones, R. A. (Ed.). (2000). *Méthodes de recherche en sciences humaines* (Traduction et adaptation de la deuxième édition américaine par Nathalie Burnay et Olivier Servais avec la collaboration de Florence Vandendorpe ed.). Paris: De Boeck Université.
- Kahneman, D. & Tversky, A. (1982). On the psychology of prediction. Dans D. Kahneman, P. Slovic & A. Tversky (Éds.), *Judgment under uncertainty: heuristics and biases* (pp. 48-68). Cambridge: Cambridge University Press.
- Lam, T. C. M. & Kolic, M. (2006). *Effects of semantic incompatibility on rating response*. Compte rendu de la réunion annuelle de l'American Educational Research Association (AERA) tenue à San Francisco en Californie.
- Lamoureux, A. (2000). *Recherche et méthodologie en sciences humaines* (2<sup>e</sup> éd.). Laval, Québec: Éditions Études vivantes.
- Larousse. (1988). *petit Larousse illustré*. Paris: Librairie Larousse.
- Lawton, G., Bhakta, B. B., Chamberlain, M. A. & Tennant, A. (2004). The Behçet's disease activity index. *Rheumatology*, 43(1), 73-78.
- Legendre, R. (1988). *Dictionnaire actuel de l'éducation*. Paris: Larousse.

- Linacre, J. M. (2004). Optimizing rating scale category effectiveness. Dans E. V. Smith Jr. & R. M. Smith (Éds.), *Introduction to Rasch measurement: theory, models and applications* (pp. 258-278). Maple Grove, MN: JAM Press.
- Linacre, J. M. (2006). Aide en ligne du logiciel Winsteps Rasch Measurement (Version 3.60.1).
- Linacre, J. M. & Wright, B. D. (1994). Dichotomous infit and outfit mean-square fit statistics. *Rasch measurement transactions*, 8(2), 350-360.
- Maheux, P. (2001). *Étude exploratoire sur l'utilisation du Rating Scale Model de la famille de Rasch dans le processus de validation d'un instrument de cueillette d'information*. Mémoire de maîtrise en éducation non publié, Université de Montréal, Montréal.
- Ng, S. H., Pipe, M.-E., Beath, B. & Holton, D. (1999). Framing the problem: effects of wording on children's statistical inferences. *Educational psychology*, 19(4), 489-499.
- Park, T. (2004). An investigation of an ESL placement test of writing using many-facet Rasch measurement. *Working papers in TESOL & applied linguistics*, 4(1).
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Rugg, D. (1941). Experiments in wording questions: II. *Public opinion quarterly*, 5(1), 91-92.
- Schaeffer, N. C. (1991). Hardly ever or constantly? Group comparisons using vague quantifiers. *Public opinion quarterly*, 55(3), 395-423.



- Schuman, H. & Presser, S. (1996). *Questions and answers in attitude surveys : experiments on question form, wording, and context*. Thousand Oaks, CA: Sage Publications.
- Schwarz, N. (1995). What respondents learn from questionnaires: the survey interview and the logic of conversation. *International statistical review*, 63(2), 153-177.
- Smith Jr., E. V. (2004). Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective. Dans E. V. Smith Jr. & R. M. Smith (Éds.), *Introduction to Rasch measurement: theory, models and applications* (pp. 93-122). Maple Grove, MN: JAM Press.
- Smith, R. M. (1991). The distributional properties of Rasch item fit statistics. *Educational and psychological measurement*, 51(3), 541-565.
- Smith, R. M. (1996). Polytomous mean-square fit statistics. *Rasch measurement transactions*, 10(3), 516-517.
- Smith, R. M. (2004). Fit analysis in latent trait measurement models. Dans E. V. Smith Jr. & R. M. Smith (Éds.), *Introduction to Rasch measurement: theory, models and applications* (pp. 73-92). Maple Grove, MN: JAM Press.
- Smith, R. M., Schumacker, R. E. & Bush, M. J. (1998). Using item mean squares to evaluate fit to the Rasch model. *Journal of outcome measurement*, 2(1), 66-78.
- Smith, R. M. & Suh, K. K. (2003). Rasch fit statistics as a test of the invariance of item parameter estimates. *Journal of applied measurement*, 4(2), 153-163.
- Tejada, A. J. R., Gómez, A. G., García, J. L. P. & Meléndez, C. P. (2002). Two strategies for fitting real data to Rasch polytomous models. *Journal of applied measurement*, 3(2), 129-145.

- Thissen, D. & Wainer, H. (2001). *Test scoring*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Tourangeau, R. & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological bulletin*, 103(3), 299-314.
- Wänke, M. (2002). Conversational norms and the interpretation of vague quantifiers. *Applied cognitive psychology*, 16, 301-307.
- Wright, B. D. & Linacre, J. M. (1987). Rasch model derived from objectivity. *Rasch measurement transactions*, 1(1), 5-6.
- Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago: Mesa Press.
- Wright, B. D. & Mok, M. M. C. (2004). An overview of the family of Rasch measurement models. Dans E. V. Smith Jr. & R. M. Smith (Éds.), *Introduction to Rasch measurement: theory, models and applications* (pp. 1-24). Maple Grove, MN: JAM Press.



# ANNEXES

ANNEXE A : VERSION A DU QUESTIONNAIRE 2003 .....	168
ANNEXE B : FORMULATION DES ITEMS POUR LES QUESTIONNAIRES 2003, 2004 ET 2005 ....	175
ANNEXE C : TABLEAUX DE L'ANALYSE DES DONNÉES .....	178
<i>C.1. Tableaux complémentaires de l'analyse des questionnaires 2003 .....</i>	<i>178</i>
<i>C.2. Tableaux de l'analyse des questionnaires 2004 .....</i>	<i>188</i>
<i>C.3. Tableaux de l'analyse des questionnaires 2005 .....</i>	<i>210</i>
ANNEXE D : TABLEAUX DES ANALYSES COMPLÉMENTAIRES .....	231
<i>D.1. Version A du questionnaire 2004.....</i>	<i>231</i>
<i>D.2. Comparaison entre les versions A et B du questionnaire 2004.....</i>	<i>237</i>
<i>D.3. Version A du questionnaire 2005.....</i>	<i>242</i>
<i>D.4. Version B du questionnaire 2005.....</i>	<i>248</i>
<i>D.5. Comparaison entre les versions A et B du questionnaire 2005.....</i>	<i>253</i>

## **Annexe A : Version A du questionnaire 2003**

A-2003

### **QUESTIONNAIRE D'ENQUÊTE AUPRÈS DES ÉTUDIANTS ET DES ÉTUDIANTES DE LA FORMATION INITIALE DES MAÎTRES**

**3<sup>e</sup> et 4<sup>e</sup> années**

Nous nous adressons à vous en tant qu'étudiantes ou étudiants du Centre de formation initiale des maîtres (CFIM) de la Faculté des sciences de l'éducation. Nous cherchons à recueillir des informations sur votre programme d'études. Votre participation à cette enquête est primordiale pour les personnes touchées par la qualité de la formation en enseignement.

Nous vous remercions de bien vouloir répondre à ce questionnaire d'évaluation; il vous suffira de dix à quinze minutes.

Il n'y a pas de bonnes ou de mauvaises réponses; l'important est de répondre en fonction de votre expérience personnelle. Nous vous assurons que vos réponses ne seront utilisées qu'aux fins de cette étude et qu'elles seront traitées dans la plus grande confidentialité.

#### **QUELQUES DIRECTIVES AVANT DE RÉPONDRE AU QUESTIONNAIRE**

1) Pour chaque question, répondez :

- en encerclant le chiffre qui correspond à la réponse que vous choisissez,
- ou en écrivant une brève réponse, s'il y a lieu.

2) À moins d'avis contraire, n'indiquez qu'une seule réponse par question.

#### **NOUS VOUS REMERCIONS DE VOTRE COLLABORATION.**

Pour de plus amples informations sur la présente enquête, veuillez communiquer avec :

Monsieur Jean-Pierre Charland, directeur du CFIM et vice-doyen aux études de 1<sup>er</sup> cycle 343-6652

Madame Colette Gervais, directrice adjointe du CFIM 343-7615

## 1. PERCEPTION GÉNÉRALE DE LA FORMATION

Indiquez votre degré d'accord ou de désaccord avec les énoncés suivants selon la consigne suivante: le chiffre **1** indique que vous êtes **tout à fait en désaccord** avec l'énoncé et le chiffre **6** indique que vous êtes **tout à fait en accord avec l'énoncé**.

	Tout à fait en désaccord					Tout à fait en accord
	1	2	3	4	5	6
1. De manière générale, je suis satisfait(e) de <u>l'ensemble de la formation</u> (cours, encadrement, stages).	1	2	3	4	5	6
2. De manière générale, je suis satisfait(e) des <u>cours</u> que j'ai suivis à la Faculté des sciences de l'éducation.	1	2	3	4	5	6
3. De manière générale, je suis satisfait(e) du support pédagogique que j'ai reçu de mes professeurs à la Faculté des sciences de l'éducation.	1	2	3	4	5	6
4. Mon programme d'études m'a amené(e) à faire des liens entre les résultats de la recherche sur l'enseignement et la pratique.	1	2	3	4	5	6
5. Mon programme d'études m'a permis de prendre contact avec les contenus que j'aurai à enseigner.	1	2	3	4	5	6
6. Mon programme d'études m'a permis d'examiner des façons de préparer les rencontres avec les parents.	1	2	3	4	5	6
7. Mon programme d'études m'a permis d'enrichir ma culture générale.	1	2	3	4	5	6
8. Mon programme d'études m'a sensibilisé(e) à l'importance de l'éthique professionnelle.	1	2	3	4	5	6
9. Mon programme d'études m'a sensibilisé(e) à l'importance de la formation continue.	1	2	3	4	5	6
10. Mon programme d'études a répondu à mes attentes de futur enseignant ou enseignante.	1	2	3	4	5	6

## 2. PRÉPARATION À L'ENSEIGNEMENT

Cette section a pour objet de connaître ce que vous pensez de votre degré de préparation pour la réalisation de certaines tâches. Vous devez répondre en tenant compte de la phrase d'introduction suivante :

Je considère que mon programme d'études m'a permis de développer des compétences pour .....

	Tout à fait en <u>désaccord</u>					Tout à fait en <u>accord</u>
	1	2	3	4	5	6
11. Identifier les contenus difficiles à faire apprendre aux élèves	1	2	3	4	5	6
12. Répondre aux questions des parents lors de la présentation du bulletin	1	2	3	4	5	6
13. Construire des outils pour l'évaluation sommative (contrôles, examens, etc.)	1	2	3	4	5	6
14. Maîtriser les contenus que j'enseignerai en conformité avec les programmes du ministère de l'Éducation	1	2	3	4	5	6
15. Corriger la langue écrite des élèves	1	2	3	4	5	6
16. Corriger la langue orale des élèves	1	2	3	4	5	6
17. Intervenir individuellement auprès des élèves à risque d'échouer	1	2	3	4	5	6
18. Planifier le déroulement d'activités d'apprentissage	1	2	3	4	5	6
19. Adapter mes activités d'enseignement aux caractéristiques des élèves	1	2	3	4	5	6
20. Établir les règles de fonctionnement de la classe	1	2	3	4	5	6
21. Motiver les élèves à s'engager dans leur apprentissage	1	2	3	4	5	6
22. Respecter les différences ethniques ou culturelles des élèves	1	2	3	4	5	6
23. Identifier les points forts et les points faibles des élèves	1	2	3	4	5	6
24. Collaborer avec les autres enseignantes et enseignants	1	2	3	4	5	6
25. Construire des outils pour l'évaluation formative (exercices, devoirs, etc.)	1	2	3	4	5	6
26. Aider les élèves à développer leurs méthodes de travail	1	2	3	4	5	6
27. Sanctionner les problèmes de discipline chez les élèves	1	2	3	4	5	6
28. Sensibiliser les élèves aux situations de discrimination qui existent entre eux	1	2	3	4	5	6
29. Orienter les élèves vers les services d'aide appropriés	1	2	3	4	5	6
30. Discuter avec les parents des difficultés de leur enfant	1	2	3	4	5	6

### 3. LES STAGES

Indiquez votre degré d'accord ou de désaccord avec les énoncés suivants :

	Tout à fait en <u>désaccord</u>	1	2	3	4	5	Tout à fait en <u>accord</u>	6
31. Je suis satisfait(e) de la façon dont mes stages ont été organisés (choix de l'école, gestion des dossiers, etc.).								
32. Je suis satisfait(e) de l'accueil que j'ai reçu dans mes milieux de stage.	1	2	3	4	5	6		
33. Je suis satisfait(e) de l'encadrement de mes maîtres associés.	1	2	3	4	5	6		
34. Je suis satisfait(e) de l'encadrement de mes superviseurs de stage à l'université.	1	2	3	4	5	6		
35. Je suis satisfait(e) de mes apprentissages d'enseignant lors de mes stages.	1	2	3	4	5	6		
36. Je suis satisfait(e) de la façon dont mes stages ont été évalués (évaluation globale).	1	2	3	4	5	6		
37. Mes stages m'ont permis de transférer le contenu de mes cours dans la pratique.	1	2	3	4	5	6		
38. Mon programme d'études m'a bien préparé(e) à effectuer mes stages.	1	2	3	4	5	6		
39. Dans l'ensemble, mes stages étaient placés à des moments opportuns de ma formation.	1	2	3	4	5	6		
40. Le nombre d'heures consacrées aux activités de stage était suffisamment élevé.	1	2	3	4	5	6		



#### 4. POURSUITE DES ÉTUDES

Avez-vous pensé à poursuivre vos études au 2<sup>e</sup> cycle en éducation après avoir complété votre baccalauréat?

41. *Encercler une des réponses proposées.*

1. Oui, je continuerai mes études immédiatement après le baccalauréat.
2. Oui, j'aimerais poursuivre immédiatement mais ma situation personnelle ne me le permet pas (emploi, famille, etc.).
3. Oui, mais pas immédiatement. Je préfère prendre quelques années d'expérience (2 à 3) avant de poursuivre mes études.
4. Oui, mais plus tard, dans cinq ou six ans.
5. Non, à l'heure actuelle je ne pense pas du tout à poursuivre mes études au 2<sup>e</sup> cycle en éducation.

42. Quelles sont les raisons qui motivent votre réponse à la question précédente ?

---

---

---

---

---



## 6. RENSEIGNEMENTS GÉNÉRAUX

44. Votre âge \_\_\_\_\_ ans
53. Votre sexe  
 Féminin 1  
 Masculin 2
45. Votre programme d'études
- |  |   |
|--|---|
| Baccalauréat en enseignement secondaire .....                        | 1 |
| Baccalauréat en éducation préscolaire et enseignement primaire ..... | 2 |
| Baccalauréat en orthopédagogie .....                                 | 3 |
| Baccalauréat en enseignement de l'éducation physique et santé .....  | 4 |
| Baccalauréat en enseignement du français langue seconde .....        | 5 |
46. En quelle année êtes-vous? Troisième année 1  
 Quatrième année 2

## 7. VOLONTAIRES RECHERCHÉ(E)S

47. Afin de compléter les informations recueillies à l'aide de ce questionnaire, nous pourrions faire appel à des volontaires pour participer à des entrevues de groupe. Si vous êtes intéressé(e) à participer à des échanges sur votre formation et votre programme d'études, nous vous prions de bien vouloir nous laisser vos coordonnées.

Nom : .....

Numéro de téléphone :  
 (incluant le code régional) .....

Courriel : .....

**MERCI BEAUCOUP DE VOTRE COLLABORATION.**

## **Annexe B : Formulation des items pour les questionnaires 2003, 2004 et 2005**

**Tableau 10 : Formulation des items pour les versions A et B du questionnaire 2003**

<b>2003 A</b>	<b>2003 B</b>
1. Identifier les contenus difficiles à faire apprendre aux élèves	1. Identifier les contenus difficiles à faire apprendre aux élèves
<b>2. Répondre aux questions des parents lors de la présentation du bulletin</b>	<b>2. Aider les parents à comprendre le bulletin de leur enfant</b>
3. Construire des <b>outils</b> pour l'évaluation sommative ( <b>contrôles, examens, etc.</b> )	3. Construire des <b>examens écrits</b> pour l'évaluation sommative
4. Maîtriser les contenus que j'enseignerai en conformité avec les programmes du ministère de l'Éducation	4. Maîtriser les contenus que j'enseignerai en conformité avec les programmes du ministère de l'Éducation
5. Corriger la langue écrite des élèves	5. Corriger la langue écrite des élèves
6. Corriger la langue orale des élèves	6. Corriger la langue orale des élèves
7. Intervenir individuellement auprès des élèves à risque d'échouer	7. Intervenir individuellement auprès des élèves à risque d'échouer
8. Planifier le déroulement d'activités d'apprentissage	8. Planifier le déroulement d'activités d'apprentissage
9. Adapter mes activités d'enseignement aux caractéristiques des élèves	9. Adapter mes activités d'enseignement aux caractéristiques des élèves
10. Établir les règles de fonctionnement de la classe	10. Établir les règles de fonctionnement de la classe
11. Motiver les élèves à s'engager dans leur apprentissage	11. Motiver les élèves à s'engager dans leur apprentissage
12. Respecter les différences ethniques ou culturelles des élèves	12. Respecter les différences ethniques ou culturelles des élèves
13. Identifier les points forts et les points faibles des élèves	13. Identifier les points forts et les points faibles des élèves
14. Collaborer avec les autres enseignantes et enseignants	14. Collaborer avec les autres enseignantes et enseignants
<b>15. Construire des outils pour l'évaluation formative (exercices, devoirs, etc.)</b>	<b>15. Planifier l'évaluation formative</b>
16. Aider les élèves à développer leurs méthodes de travail	16. Aider les élèves à développer leurs méthodes de travail
17. Sanctionner les problèmes de discipline chez les élèves	17. Sanctionner les problèmes de discipline chez les élèves
18. Sensibiliser les élèves aux situations de discrimination qui existent entre eux	18. Sensibiliser les élèves aux situations de discrimination qui existent entre eux
19. Orienter les élèves vers les services d'aide appropriés	19. Orienter les élèves vers les services d'aide appropriés
<b>20. Discuter</b> avec les parents <b>des difficultés de leur enfant</b>	<b>20. Préparer une rencontre</b> avec les parents <b>d'un élève ayant des difficultés</b>

Tableau 11 : Formulation des items pour les versions A et B du questionnaire 2004

2004 A (items identiques à la version 2003A)	2004 B
1. Identifier les contenus difficiles à faire apprendre aux élèves	1. Identifier les contenus plus difficiles pour les élèves
2. Répondre aux questions des parents lors de la présentation du bulletin	2. Répondre aux questions lors des rencontres avec les parents
3. Construire des outils pour l'évaluation sommative (contrôles, examens, etc.)	3. Élaborer des outils pour l'évaluation sommative
4. Maîtriser les contenus que j'enseignerai en conformité avec les programmes du ministère de l'Éducation	4. Maîtriser les contenus que j'enseignerai
5. Corriger la langue écrite des élèves	5. Corriger les productions écrites des élèves
6. Corriger la langue orale des élèves	6. Corriger la langue orale des élèves
7. Intervenir individuellement auprès des élèves à risque d'échouer	7. Planifier une intervention auprès d'un élève en difficulté
8. Planifier le déroulement d'activités d'apprentissage	8. Planifier le déroulement d'activités d'apprentissage
9. Adapter mes activités d'enseignement aux caractéristiques des élèves	9. Adapter mes activités d'enseignement aux caractéristiques des élèves
10. Établir les règles de fonctionnement de la classe	10. Établir les règles de fonctionnement de la classe
11. Motiver les élèves à s'engager dans leur apprentissage	11. Motiver les élèves dans leurs apprentissages
12. Respecter les différences ethniques ou culturelles des élèves	12. Respecter les différences ethniques ou culturelles des élèves
13. Identifier les points forts et les points faibles des élèves	13. Reconnaître les points forts et les points faibles des élèves
14. Collaborer avec les autres enseignantes et enseignants	14. Collaborer avec les autres enseignantes et enseignants
15. Construire des outils pour l'évaluation formative (exercices, devoirs, etc.)	15. Élaborer des outils pour l'évaluation formative
16. Aider les élèves à développer leurs méthodes de travail	16. Aider les élèves à développer leurs méthodes de travail
17. Sanctionner les problèmes de discipline chez les élèves	17. Sanctionner les problèmes de discipline chez les élèves
18. Sensibiliser les élèves aux situations de discrimination qui existent entre eux	18. Sensibiliser les élèves aux situations de discrimination qui existent entre eux
19. Orienter les élèves vers les services d'aide appropriés	19. Suggérer aux élèves des services d'aide appropriés
20. Discuter avec les parents des difficultés de leur enfant	20. Expliquer aux parents quelles sont les difficultés de leur enfant

Tableau 12 : Formulation des items pour les versions A et B du questionnaire 2005

2005 A (items identiques à la version 2003A <sup>15</sup> )	2005 B
1. Identifier les contenus plus difficiles pour les élèves (2004B)	1. Identifier les contenus plus difficiles pour les élèves
2. Répondre aux questions lors des rencontres avec les parents (2004B)	2. Répondre aux questions lors des rencontres avec les parents
3. Élaborer des outils pour l'évaluation sommative (contrôles, examens, etc.) (nouvel item)	3. Élaborer des outils pour l'évaluation sommative (contrôles, examens, etc.)
4. Maîtriser les contenus que j'enseignerai <b>en conformité avec les programmes du ministère de l'Éducation</b>	4. Maîtriser les contenus que j'enseignerai
5. Corriger la <b>langue écrite</b> des élèves	5. Corriger <b>les productions écrites</b> des élèves
6. Corriger la langue orale des élèves	6. Corriger la langue orale des élèves
7. Planifier une intervention auprès d'un élève en difficulté (2004B)	7. Planifier une intervention auprès d'un élève en difficulté
8. Planifier le déroulement d'activités d'apprentissage	8. Planifier le déroulement d'activités d'apprentissage
9. Adapter mes activités d'enseignement <b>aux caractéristiques des élèves</b>	9. Adapter mes activités d'enseignement
10. Établir les règles de fonctionnement de la classe	10. Établir les règles de fonctionnement de la classe
11. <b>Motiver les élèves à s'engager dans leur apprentissage</b>	11. <b>Maintenir un climat d'apprentissage en classe</b>
12. Respecter les différences ethniques ou culturelles des élèves	12. Respecter les différences ethniques ou culturelles des élèves
13. Identifier les points forts et les points faibles des élèves	13. Identifier les points forts et les points faibles des élèves
14. Collaborer avec les <b>autres enseignantes et enseignants</b>	14. Collaborer avec les <b>membres de l'équipe-école</b>
15. <b>Construire</b> des outils pour l'évaluation formative (exercices, devoirs, etc.)	15. <b>Élaborer</b> des outils pour l'évaluation formative (exercices, devoirs, etc.)
16. Aider les élèves à développer leurs méthodes de travail	16. Aider les élèves à développer leurs méthodes de travail
17. Sanctionner les problèmes de discipline chez les élèves	17. Sanctionner les problèmes de discipline chez les élèves
18. Sensibiliser les élèves aux situations de discrimination qui <b>existent</b> entre eux	18. Sensibiliser les élèves aux situations de discrimination qui <b>peuvent exister</b> entre eux
19. Orienter les élèves vers les services d'aide appropriés	19. Orienter les élèves vers les services d'aide appropriés
20. <b>Guider les élèves dans leurs apprentissages</b> (nouvel item)	20. <b>Discuter avec les élèves de leurs difficultés</b>

<sup>15</sup> Sauf si une indication précise une que l'item provient d'une autre version du questionnaire.

## **Annexe C : Tableaux de l'analyse des données**

### **C.1. Tableaux complémentaires de l'analyse des questionnaires 2003**

**Tableau 13 : Statistiques d'ajustement initiales des personnes pour la version A du questionnaire 2003**

<b>Sujet</b>	<b>Mesure estimée</b>	<b>Erreur type</b>	<b>Infit MNSQ</b>	<b>Infit ZSTD</b>	<b>Outfit MNSQ</b>	<b>Outfit ZSTD</b>	<b>Corr.</b>
610	-0,82	0,26	0,40	-2,5	0,42	-2,4	0,63
611	-0,55	0,26	1,15	0,6	1,19	0,7	0,17
612	-1,85	0,28	1,00	0,1	1,03	0,2	0,44
613	-0,35	0,26	1,16	0,6	1,15	0,6	0,76
614	1,21	0,28	1,32	1,0	1,22	0,8	0,50
616	0,83	0,27	0,47	-2,1	0,45	-2,2	0,77
617	1,05	0,28	0,76	-0,7	0,75	-0,8	0,80
618	-0,35	0,26	0,66	-1,2	0,67	-1,2	0,57
619	-0,15	0,26	1,22	0,8	1,18	0,7	0,65
621	0,75	0,27	0,56	-1,6	0,56	-1,6	0,71
622	1,79	0,30	1,15	0,6	1,24	0,8	0,19
623	1,28	0,28	0,82	-0,5	0,89	-0,2	0,63
624	0,19	0,26	1,13	0,5	1,08	0,4	0,61
647	1,13	0,28	0,56	-1,6	0,55	-1,6	0,51
648	0,26	0,26	0,77	-0,7	0,76	-0,7	0,62
649	0,90	0,27	1,07	0,3	1,08	0,3	0,62
650	0,83	0,27	1,65	1,9	1,62	1,8	0,87
651	0,19	0,26	0,69	-1,1	0,67	-1,2	0,70
652	1,45	0,29	0,70	-1,0	0,74	-0,8	0,82
774	0,47	0,26	0,99	0,1	0,98	0,0	0,60
775	0,26	0,26	0,49	-2,0	0,47	-2,1	0,75
776	0,83	0,27	0,57	-1,6	0,57	-1,5	0,73
777	-0,82	0,26	0,69	-1,1	0,74	-0,8	0,54
<b>778</b>	<b>0,54</b>	<b>0,27</b>	<b>0,27</b>	<b>-3,4</b>	<b>0,26</b>	<b>-3,4</b>	<b>0,86</b>
779	-0,15	0,26	1,82	2,3	1,84	2,4	0,63
780	0,90	0,27	0,75	-0,8	0,81	-0,6	0,74
781	1,53	0,29	0,71	-0,9	0,71	-0,9	0,84

Suite du tableau 13

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
782	0,54	0,27	1,71	2,0	1,66	1,9	0,69
783	-0,75	0,26	1,77	2,2	1,80	2,3	0,64
784	1,13	0,28	0,66	-1,1	0,70	-1,0	0,53
790	-0,62	0,27	1,20	0,7	1,24	0,8	0,22
791	-0,42	0,26	1,60	1,8	1,53	1,6	0,65
792	-0,89	0,26	1,28	1,0	1,37	1,2	0,20
793	0,61	0,27	0,94	-0,1	0,88	-0,3	0,55
794	0,90	0,27	0,80	-0,6	0,79	-0,6	0,45
795	-1,61	0,28	1,01	0,2	1,00	0,1	0,67
796	-0,15	0,26	1,70	2,0	1,66	1,9	0,79
797	1,45	0,29	0,73	-0,8	0,73	-0,8	0,86
798	0,19	0,26	0,69	-1,1	0,67	-1,1	0,67
<b>799</b>	<b>0,19</b>	<b>0,26</b>	<b>2,99</b>	<b>4,5</b>	<b>2,92</b>	<b>4,4</b>	<b>0,49</b>
800	1,45	0,29	0,82	-0,5	0,80	-0,6	0,48
801	-0,08	0,26	0,53	-1,8	0,54	-1,8	0,69
802	0,97	0,27	1,35	1,1	1,38	1,2	0,59
803	-1,10	0,27	0,54	-1,7	0,54	-1,7	0,71
804	1,53	0,29	0,86	-0,4	0,77	-0,7	0,18
827	1,21	0,28	1,23	0,8	1,32	1,0	0,70
828	2,07	0,31	1,24	0,8	1,21	0,7	0,78
<b>829</b>	<b>-2,98</b>	<b>0,34</b>	<b>0,31</b>	<b>-2,7</b>	<b>0,32</b>	<b>-2,4</b>	<b>0,76</b>
830	1,62	0,29	1,12	0,5	1,06	0,3	0,67
832	0,68	0,27	0,49	-1,9	0,47	-2,1	0,85
873	-0,42	0,26	0,64	-1,3	0,68	-1,1	0,34
874	0,06	0,26	1,69	2,0	1,66	1,9	0,74
775	-0,35	0,26	1,08	0,4	1,09	0,4	0,66
876	-0,01	0,26	1,36	1,2	1,33	1,1	0,59
<b>877</b>	<b>1,28</b>	<b>0,28</b>	<b>0,33</b>	<b>-2,8</b>	<b>0,32</b>	<b>-2,9</b>	<b>0,69</b>
<b>MOY.</b>	<b>0,32</b>	<b>0,27</b>	<b>0,98</b>	<b>-0,2</b>	<b>0,98</b>	<b>-0,2</b>	
<b>É.-T.</b>	<b>0,98</b>	<b>0,02</b>	<b>0,49</b>	<b>1,5</b>	<b>0,48</b>	<b>1,5</b>	



**Tableau 14 : Statistiques d'ajustement initiales des personnes pour la version B du questionnaire 2003**

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
880	1,04	0,27	0,34	-2,8	0,42	-2,3	0,63
881	0,13	0,25	0,51	-1,9	0,49	-2,0	0,68
882	1,49	0,28	0,32	-2,8	0,37	-2,5	0,62
883	1,26	0,27	0,26	-3,3	0,27	-3,2	0,78
884	-0,92	0,25	1,01	0,1	1,01	0,1	0,83
885	-0,55	0,25	0,90	-0,2	0,91	-0,2	0,70
886	-0,92	0,25	0,78	-0,7	0,84	-0,5	0,62
887	-1,46	0,27	0,89	-0,3	0,95	0,0	0,71
888	-0,86	0,25	0,68	-1,1	0,74	-0,9	0,67
889	0,90	0,26	0,53	-1,7	0,50	-1,9	0,69
956	-0,12	0,25	0,36	-2,8	0,42	-2,4	0,66
957	1,41	0,28	1,19	0,7	1,11	0,4	0,59
958	-0,86	0,25	0,33	-3,0	0,35	-2,8	0,80
959	0,57	0,25	0,84	-0,4	0,83	-0,5	0,41
960	-0,36	0,25	0,76	-0,8	0,80	-0,6	0,66
961	1,26	0,27	0,38	-2,5	0,38	-2,5	0,67
962	1,26	0,27	1,44	1,3	1,30	1,0	0,85
971	-2,55	0,32	0,90	-0,2	1,12	0,4	-0,08
973	-0,92	0,25	0,86	-0,4	1,17	0,6	0,57
974	-1,25	0,26	0,96	0,0	0,98	0,0	0,56
975	1,11	0,27	1,35	1,1	1,29	1,0	0,41
976	0,13	0,25	0,62	-1,4	0,72	-0,9	0,35
977	1,26	0,27	0,55	-1,6	0,57	-1,5	0,76
978	1,49	0,28	1,82	2,2	1,72	2,0	0,31
979	1,91	0,30	0,47	-1,9	0,51	-1,7	0,79
980	0,38	0,25	0,92	-0,2	0,92	-0,2	0,77
981	1,11	0,27	0,69	-1,0	0,72	-0,9	0,39
982	1,82	0,30	1,59	1,6	1,66	1,8	0,49
1023	-0,24	0,25	0,67	-1,1	0,65	-1,2	0,77
1086	0,38	0,25	2,21	3,1	2,20	3,1	0,53
1087	0,13	0,25	0,73	-0,9	0,69	-1,1	0,71
1088	-0,55	0,25	0,52	-1,8	0,55	-1,7	0,81
1089	-0,12	0,25	0,99	0,1	0,99	0,1	0,77
1090	-0,24	0,25	0,92	-0,2	0,95	-0,1	0,76
1091	2,50	0,33	0,67	-1,0	0,69	-0,9	0,52
1092	-0,12	0,25	2,27	3,3	2,27	3,3	0,69
1093	0,19	0,25	0,73	-0,9	0,70	-1,0	0,68
1094	2,50	0,33	0,56	-1,4	0,55	-1,5	0,73

Suite du tableau 14

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1134	-0,06	0,25	0,77	-0,7	0,83	-0,5	0,30
1135	1,65	0,29	1,00	0,1	0,98	0,0	0,43
1136	0,00	0,25	1,04	0,2	1,04	0,2	0,78
1137	-1,18	0,26	0,52	-1,8	0,55	-1,7	0,63
1138	-1,05	0,26	0,92	-0,2	0,91	-0,2	0,62
1139	1,04	0,27	1,10	0,4	1,03	0,2	0,65
1140	0,13	0,25	1,16	0,6	1,19	0,7	0,73
<b>1141</b>	<b>-0,30</b>	<b>0,25</b>	<b>1,93</b>	<b>2,6</b>	<b>1,92</b>	<b>2,5</b>	<b>0,79</b>
1142	0,13	0,25	0,92	-0,2	0,92	-0,2	0,76
1143	-0,92	0,25	1,57	1,7	1,56	1,7	0,65
1144	1,04	0,27	0,74	-0,8	0,77	-0,7	0,71
1145	0,06	0,25	1,56	1,7	1,52	1,6	0,66
1146	2,29	0,32	1,07	0,3	1,03	0,2	0,62
1147	0,83	0,26	0,79	-0,6	0,79	-0,6	0,87
1148	-0,73	0,25	0,85	-0,4	0,94	-0,1	0,60
1149	0,76	0,26	1,21	0,7	1,22	0,8	0,80
1155	-0,12	0,25	0,53	-1,8	0,53	-1,8	0,79
1156	0,19	0,25	0,87	-0,3	0,86	-0,4	0,46
1157	0,44	0,25	1,79	2,2	1,70	2,0	0,75
1158	1,33	0,28	0,40	-2,4	0,43	-2,2	0,72
1159	-0,24	0,25	0,57	-1,6	0,62	-1,4	0,58
1160	0,76	0,26	0,47	-2,0	0,47	-2,0	0,68
<b>1161</b>	<b>-0,30</b>	<b>0,25</b>	<b>0,42</b>	<b>-2,4</b>	<b>0,40</b>	<b>-2,6</b>	<b>0,81</b>
1162	-0,42	0,25	0,67	-1,2	0,68	-1,1	0,72
<b>1163</b>	<b>-0,12</b>	<b>0,25</b>	<b>5,68</b>	<b>8,0</b>	<b>6,66</b>	<b>9,0</b>	<b>-0,72</b>
1164	-1,61	0,27	1,03	0,2	1,08	0,3	0,36
1165	0,76	0,26	1,31	1,0	1,25	0,9	0,47
1166	-0,73	0,25	0,68	-1,1	0,63	-1,3	0,66
1167	0,19	0,25	0,90	-0,3	0,90	-0,2	0,80
<b>1168</b>	<b>2,09</b>	<b>0,31</b>	<b>0,30</b>	<b>-2,9</b>	<b>0,49</b>	<b>-1,8</b>	<b>0,62</b>
1169	1,61	0,30	1,18	0,6	1,12	0,4	0,67
<b>MOY.</b>	<b>0,28</b>	<b>0,26</b>	<b>0,96</b>	<b>-0,4</b>	<b>0,99</b>	<b>-0,3</b>	
<b>É.-T.</b>	<b>1,05</b>	<b>0,02</b>	<b>0,72</b>	<b>1,8</b>	<b>0,81</b>	<b>1,8</b>	

**Tableau 15 : Statistiques d'ajustement finales des personnes pour la version B du questionnaire 2003**

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
880	1,24	0,29	0,43	-2,2	0,52	-1,8	0,62
881	0,15	0,27	0,63	-1,3	0,61	-1,4	0,67
882	1,77	0,30	0,38	-2,4	0,47	-2,0	0,62
884	-1,10	0,28	1,21	0,8	1,20	0,7	0,82
885	-0,66	0,27	0,98	0,0	0,97	0,0	0,73
886	-1,10	0,28	0,91	-0,2	0,94	-0,1	0,63
887	-1,74	0,29	1,04	0,2	1,13	0,5	0,71
888	-1,03	0,27	0,89	-0,3	0,93	-0,1	0,63
889	1,08	0,29	0,65	-1,2	0,62	-1,3	0,68
956	-0,14	0,27	0,44	-2,3	0,50	-1,9	0,66
957	1,68	0,30	1,38	1,2	1,30	1,0	0,60
958	-1,03	0,27	0,45	-2,2	0,46	-2,2	0,78
959	0,68	0,28	1,05	0,3	1,04	0,2	0,38
960	-0,44	0,27	0,88	-0,3	0,93	-0,1	0,68
961	1,50	0,30	0,49	-1,9	0,51	-1,8	0,65
962	1,50	0,30	1,72	2,0	1,55	1,6	0,84
973	-1,10	0,28	1,08	0,4	1,38	1,2	0,56
974	-1,49	0,28	1,07	0,3	1,09	0,4	0,59
975	1,33	0,29	1,61	1,7	1,57	1,7	0,41
976	0,15	0,27	0,78	-0,7	0,88	-0,3	0,33
977	1,50	0,30	0,63	-1,2	0,64	-1,2	0,76
978	1,77	0,30	2,00	2,5	1,87	2,3	0,36
979	2,26	0,32	0,55	-1,6	0,58	-1,4	0,78
980	0,45	0,28	1,01	0,1	1,00	0,1	0,80
981	1,33	0,29	0,86	-0,3	0,91	-0,2	0,37
982	2,16	0,32	1,93	2,4	1,99	2,5	0,45
1023	-0,29	0,27	0,91	-0,2	0,89	-0,3	0,73
1087	0,15	0,27	0,81	-0,6	0,78	-0,7	0,73
1088	-0,66	0,27	0,71	-1,0	0,73	-0,9	0,77
1089	-0,14	0,27	1,09	0,4	1,08	0,4	0,80
1090	-0,29	0,27	1,04	0,2	1,08	0,4	0,77
1091	2,93	0,35	0,77	-0,7	0,76	-0,6	0,51
1093	0,23	0,27	0,95	0,0	0,91	-0,2	0,65
1094	2,93	0,35	0,56	-1,5	0,55	-1,5	0,76
1134	-0,07	0,27	0,97	0,0	1,05	0,3	0,27
1135	1,96	0,31	1,18	0,6	1,21	0,7	0,42
1136	0,00	0,27	1,14	0,6	1,15	0,6	0,81
1137	-1,41	0,28	0,58	-1,5	0,61	-1,4	0,66

Suite du tableau 15

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1138	-1,26	0,28	1,02	0,2	1,02	0,2	0,65
1139	1,24	0,29	1,35	1,1	1,28	0,9	0,63
1140	0,15	0,27	1,29	1,0	1,31	1,0	0,76
1142	0,15	0,27	1,11	0,5	1,10	0,4	0,76
1143	-1,10	0,28	1,77	2,2	1,74	2,1	0,68
1144	1,24	0,29	0,80	-0,6	0,83	-0,5	0,74
1145	0,08	0,27	1,89	2,4	1,83	2,3	0,65
1146	2,70	0,34	1,18	0,6	1,12	0,5	0,63
1147	1,00	0,28	0,94	-0,1	0,93	-0,1	0,86
1148	-0,88	0,27	1,00	0,1	1,09	0,4	0,61
1149	0,92	0,28	1,43	1,3	1,42	1,3	0,80
1155	-0,14	0,27	0,68	-1,1	0,67	-1,1	0,77
1156	0,23	0,27	0,99	0,1	0,98	0,0	0,48
1158	1,59	0,30	0,46	-2,0	0,52	-1,7	0,72
1159	-0,29	0,27	0,73	-0,9	0,77	-0,7	0,57
1160	0,92	0,28	0,53	-1,7	0,54	-1,7	0,69
1161	-0,36	0,27	0,53	-1,8	0,51	-1,9	0,80
1162	-0,51	0,27	0,82	-0,5	0,82	-0,5	0,72
1164	-1,90	0,29	1,20	0,7	1,28	0,9	0,37
1165	0,92	0,28	1,50	1,5	1,43	1,3	0,49
1166	-0,88	0,27	0,72	-0,9	0,68	-1,1	0,70
1167	0,23	0,27	1,12	0,5	1,11	0,5	0,78
1168	2,47	0,33	0,36	-2,5	0,65	-1,1	0,61
1169	1,92	0,33	1,40	1,2	1,31	0,9	0,66
<b>MOY.</b>	<b>0,40</b>	<b>0,29</b>	<b>0,98</b>	<b>-0,1</b>	<b>0,99</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>1,23</b>	<b>0,02</b>	<b>0,39</b>	<b>1,3</b>	<b>0,37</b>	<b>1,2</b>	

**Tableau 16 : Statistiques d'ajustement des items pour la version B du questionnaire 2003**

Item	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1	0,73	0,16	1,15	0,9	1,17	1,0	0,55
2	2,16	0,17	1,36	1,9	1,42	2,2	0,57
3	1,19	0,16	1,31	1,7	1,32	1,7	0,67
4	-0,38	0,16	0,65	-2,2	0,63	-2,4	0,75
5	-0,13	0,16	0,98	-0,1	0,98	0,0	0,79
6	-0,13	0,16	0,97	-0,1	0,98	-0,1	0,78
7	0,46	0,16	0,84	-0,9	0,81	-1,1	0,83
8	-2,77	0,21	1,51	2,3	1,42	1,9	0,41
9	-0,67	0,16	0,73	-1,6	0,77	-1,4	0,72
10	-1,58	0,18	0,73	-1,6	0,75	-1,5	0,70
11	-0,38	0,16	0,72	-1,7	0,72	-1,7	0,76
12	-1,46	0,17	0,87	-0,7	0,91	-0,4	0,66
13	-0,13	0,16	0,74	-1,6	0,76	-1,5	0,72
14	-0,31	0,16	1,15	0,9	1,11	0,7	0,75
15	0,51	0,16	1,17	1,0	1,19	1,1	0,69
16	0,10	0,16	0,67	-2,0	0,66	-2,2	0,79
17	-0,36	0,16	0,99	0,0	1,06	0,4	0,68
18	0,44	0,16	0,87	-0,7	0,85	-0,8	0,76
19	1,14	0,16	0,85	-0,9	0,83	-1,0	0,73
20	1,58	0,16	1,41	2,2	1,43	2,3	0,58
<b>MOY.</b>	<b>0,00</b>	<b>0,16</b>	<b>0,98</b>	<b>-0,2</b>	<b>0,99</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>1,10</b>	<b>0,01</b>	<b>0,26</b>	<b>1,4</b>	<b>0,26</b>	<b>1,4</b>	

**Tableau 17 : Statistiques d'ajustement générales des personnes pour la version B du questionnaire 2003**

	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	0,40	0,29	0,98	-0,1	0,99	-0,1
<b>É.-T.</b>	1,23	0,02	0,39	1,3	0,37	1,2
<b>MAX.</b>	2,93	0,35	2,00	2,5	1,99	2,5
<b>MIN.</b>	-1,90	0,27	0,36	-2,5	0,46	-2,2

<b>Réel</b>	<b>Séparation</b>	3,85	« Fidélité »	0,94
<b>Modèle</b>	<b>Séparation</b>	4,15	« Fidélité »	0,95

**Tableau 18 : Statistiques d'ajustement générales des items de la version B du questionnaire 2003**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	233,5	61,9	0,00	0,16	0,98	-0,2	0,99	-0,1
<b>É.-T.</b>	41,5	0,3	1,10	0,01	0,26	1,4	0,26	1,4
<b>MAX.</b>	327,0	62,0	2,16	0,21	1,51	2,3	1,43	2,3
<b>MIN.</b>	149,0	61,0	-2,77	0,16	0,65	-2,2	0,63	-2,4

<b>Réel</b>	<b>Séparation</b>	6,33	« Fidélité »	0,98
<b>Modèle</b>	<b>Séparation</b>	6,69	« Fidélité »	0,98



Tableau 19 : Points d'ancrage de la version B du questionnaire 2003

Catégorie	Nb. d'obs.	% d'obs.	Mesure moyenne observée	Mesure moyenne attendue	Infit MNSQ	Outfit MNSQ	Points d'ancrage	Erreur type
1	50	4	-1,94	-2,16	1,24	1,20	NONE	
2	169	14	-1,40	-1,32	0,82	0,81	-2,95	0,16
3	290	23	-0,48	-0,44	0,96	0,96	-1,42	0,10
4	335	27	0,54	0,53	0,97	1,00	-0,11	0,08
5	292	24	1,68	1,62	0,98	1,01	1,20	0,08
6	102	8	2,80	2,89	1,02	1,03	3,28	0,13

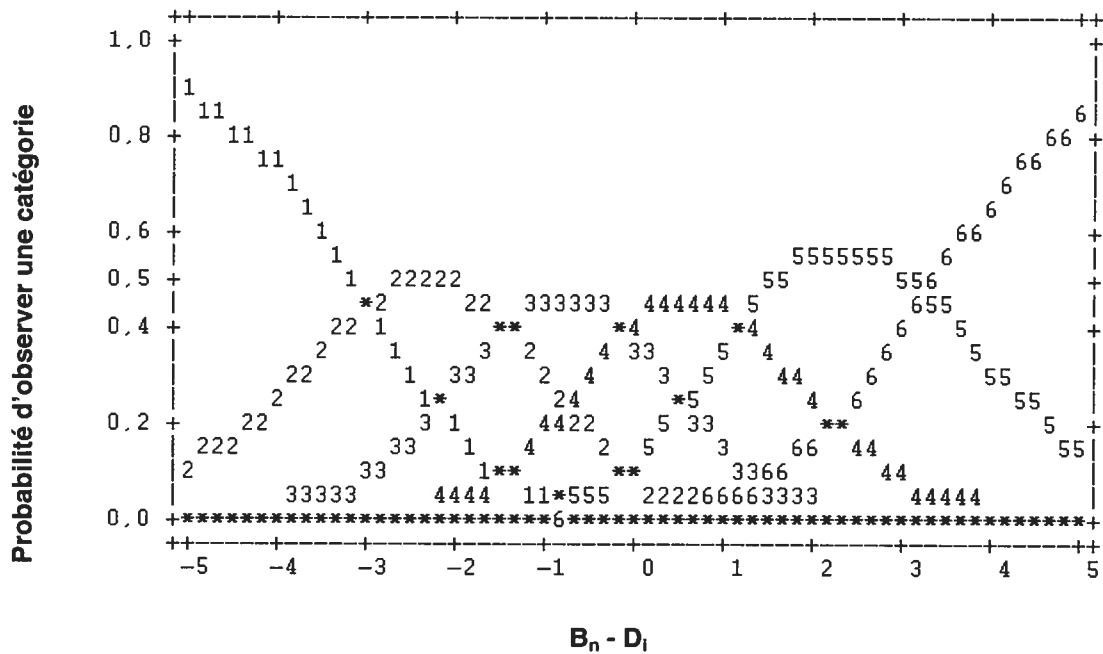


Figure 10 : Courbes de probabilités de chacune des catégories de réponse de la version B du questionnaire 2003



## C.2. Tableaux de l'analyse des questionnaires 2004

Tableau 20 : Statistiques d'ajustement initiales des personnes pour la version A du questionnaire 2004

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1	-0,54	0,23	1,08	0,4	1,05	0,3	0,77
2	-0,75	0,23	1,38	1,2	1,32	1,1	0,60
3	0,93	0,24	0,76	-0,7	0,73	-0,8	0,67
4	-0,75	0,23	0,52	-1,9	0,49	-2,0	0,65
5	-1,04	0,24	0,92	-0,2	0,99	0,1	0,37
<b>6</b>	<b>-0,54</b>	<b>0,23</b>	<b>2,09</b>	<b>2,9</b>	<b>1,98</b>	<b>2,7</b>	<b>0,72</b>
7	0,64	0,23	0,50	-1,9	0,46	-2,1	0,65
8	0,53	0,23	0,88	-0,3	0,83	-0,5	0,46
9	0,43	0,23	1,45	1,4	1,68	2,0	0,03
10	-0,39	0,23	1,43	1,4	1,40	1,3	0,60
11	-0,44	0,23	1,23	0,8	1,21	0,8	0,78
12	-0,86	0,24	0,79	-0,6	0,75	-0,8	0,78
32	0,87	0,24	1,64	1,8	1,48	1,5	0,73
33	-0,70	0,23	1,06	0,3	1,07	0,3	0,82
34	-0,86	0,24	0,76	-0,7	0,77	-0,7	0,52
35	-0,69	0,24	0,51	-1,8	0,53	-1,7	0,80
<b>36</b>	<b>-0,65</b>	<b>0,23</b>	<b>2,33</b>	<b>3,4</b>	<b>2,30</b>	<b>3,3</b>	<b>0,33</b>
<b>61</b>	<b>-1,15</b>	<b>0,25</b>	<b>0,37</b>	<b>-2,6</b>	<b>0,38</b>	<b>-2,5</b>	<b>0,71</b>
62	-1,15	0,25	0,86	-0,4	0,78	-0,7	0,67
<b>63</b>	<b>0,17</b>	<b>0,23</b>	<b>0,35</b>	<b>-2,9</b>	<b>0,43</b>	<b>-2,4</b>	<b>0,57</b>
64	-0,01	0,23	0,92	-0,2	0,94	-0,1	0,81
65	-0,18	0,22	1,15	0,6	1,16	0,6	0,76
66	-0,29	0,23	0,75	-0,8	0,74	-0,9	0,79
67	-1,76	0,28	0,43	-2,1	0,47	-1,8	0,55
68	-1,15	0,25	0,51	-1,8	0,46	-2,1	0,62
69	1,52	0,27	0,72	-0,8	0,67	-1,0	0,71
70	-0,39	0,23	0,64	-1,3	0,76	-0,8	0,48
<b>71</b>	<b>-0,34</b>	<b>0,23</b>	<b>0,37</b>	<b>-2,8</b>	<b>0,44</b>	<b>-2,4</b>	<b>0,58</b>
<b>72</b>	<b>-0,03</b>	<b>0,22</b>	<b>0,28</b>	<b>-3,4</b>	<b>0,27</b>	<b>-3,5</b>	<b>0,87</b>
<b>73</b>	<b>0,02</b>	<b>0,22</b>	<b>2,19</b>	<b>3,2</b>	<b>2,20</b>	<b>3,2</b>	<b>0,49</b>
74	-0,65	0,23	0,73	-0,9	0,77	-0,7	0,77
75	0,75	0,24	0,73	-0,9	0,66	-1,2	0,53
76	-0,18	0,22	0,81	-0,6	0,84	-0,5	0,69
77	-1,09	0,24	0,65	-1,2	0,64	-1,2	0,65
78	-0,81	0,23	1,12	0,5	1,10	0,4	0,74

Suite du tableau 20

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
79	0,75	0,24	0,52	-1,8	0,57	-1,5	0,37
80	-2,49	0,33	1,54	1,4	1,40	1,1	0,51
<b>81</b>	<b>-1,22</b>	<b>0,25</b>	<b>0,30</b>	<b>-3,0</b>	<b>0,32</b>	<b>-2,9</b>	<b>0,64</b>
82	-1,09	0,24	1,16	0,6	1,23	0,8	0,70
83	-0,98	0,24	0,70	-1,0	0,77	-0,7	0,14
<b>84</b>	<b>1,84</b>	<b>0,29</b>	<b>2,63</b>	<b>3,4</b>	<b>2,25</b>	<b>2,8</b>	<b>0,43</b>
<b>85</b>	<b>1,38</b>	<b>0,26</b>	<b>4,30</b>	<b>5,8</b>	<b>3,77</b>	<b>5,2</b>	<b>0,60</b>
124	-1,68	0,27	0,70	-0,9	0,60	-1,3	0,68
125	-0,03	0,22	0,56	-1,7	0,57	-1,6	0,63
126	-0,03	0,22	0,92	-0,2	0,92	-0,2	0,32
127	-1,34	0,25	1,53	1,6	1,38	1,2	0,64
128	0,17	0,23	1,90	2,5	1,86	2,4	0,55
144	-0,86	0,24	0,58	-1,6	0,61	-1,4	0,52
145	0,27	0,23	1,15	0,6	1,16	0,6	0,45
146	-0,23	0,23	1,14	0,6	1,15	0,6	0,54
147	1,52	0,27	0,62	-1,2	0,55	-1,5	0,29
148	-1,04	0,24	0,52	-1,8	0,53	-1,8	0,53
149	0,22	0,23	0,75	-0,8	0,80	-0,6	0,40
150	0,53	0,23	0,49	-2,0	0,47	-2,1	0,67
<b>151</b>	<b>1,17</b>	<b>0,25</b>	<b>0,16</b>	<b>-4,2</b>	<b>0,19</b>	<b>-3,8</b>	<b>0,75</b>
152	-1,15	0,25	1,94	2,5	1,88	2,4	0,68
153	-0,44	0,23	0,86	-0,4	0,83	-0,5	0,66
160	-3,01	0,39	2,08	2,1	1,49	1,1	0,36
161	-1,22	0,25	0,40	-2,4	0,42	-2,3	0,58
162	-1,15	0,25	0,73	-0,9	0,79	-0,6	0,43
163	-3,01	0,39	0,78	-0,4	0,70	-0,6	0,51
164	-2,86	0,37	0,49	-1,4	0,58	-1,0	0,65
165	-0,70	0,23	1,10	0,4	1,39	1,3	0,36
<b>MOY.</b>	<b>-0,45</b>	<b>0,25</b>	<b>1,01</b>	<b>-0,3</b>	<b>0,98</b>	<b>-0,3</b>	
<b>É.-T.</b>	<b>1,02</b>	<b>0,04</b>	<b>0,69</b>	<b>1,9</b>	<b>0,61</b>	<b>1,8</b>	

**Tableau 21 : Statistiques d'ajustement initiales des personnes pour la version B du questionnaire 2004**

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
201	-1,09	0,23	0,48	-2,0	0,51	-1,9	0,80
202	-0,19	0,22	0,70	-1,1	0,70	-1,0	0,81
203	-2,50	0,34	1,00	0,2	1,14	0,5	0,37
204	0,11	0,22	0,93	-0,2	0,94	-,1	0,66
<b>205</b>	<b>0,75</b>	<b>0,24</b>	<b>0,35</b>	<b>-2,7</b>	<b>0,38</b>	<b>-2,5</b>	<b>0,64</b>
206	0,06	0,22	0,81	-0,6	0,76	-0,8	0,67
207	-0,33	0,22	1,71	2,1	1,77	2,3	0,30
208	0,47	0,23	0,88	-0,3	0,88	-0,3	0,68
209	0,64	0,24	0,82	-0,5	0,78	-0,7	0,60
210	-0,99	0,23	0,88	-0,3	0,93	-0,1	0,82
223	-0,68	0,22	0,56	-1,7	0,55	-1,8	0,69
231	-1,78	0,27	1,04	0,2	1,01	0,1	0,61
232	-1,09	0,23	0,64	-1,3	0,64	-1,3	0,71
<b>233</b>	<b>-1,57</b>	<b>0,26</b>	<b>2,10</b>	<b>2,7</b>	<b>1,84</b>	<b>2,2</b>	<b>0,53</b>
<b>234</b>	<b>-0,43</b>	<b>0,22</b>	<b>2,33</b>	<b>3,5</b>	<b>2,52</b>	<b>3,8</b>	<b>0,29</b>
<b>235</b>	<b>-0,58</b>	<b>0,22</b>	<b>2,64</b>	<b>4,1</b>	<b>2,80</b>	<b>4,3</b>	<b>0,30</b>
259	-2,19	0,30	0,76	-0,6	0,72	-0,6	0,63
260	-0,19	0,22	0,60	-1,5	0,58	-1,6	0,64
261	-0,88	0,23	0,61	-1,4	0,57	-1,6	0,66
262	-0,09	0,22	1,02	0,2	1,03	0,2	0,60
263	0,42	0,23	1,81	2,3	1,76	2,2	0,64
276	-0,33	0,22	0,57	-1,7	0,56	-1,7	0,72
277	-0,83	0,23	1,58	1,8	1,60	1,8	0,63
<b>278</b>	<b>0,26</b>	<b>0,23</b>	<b>2,19</b>	<b>3,1</b>	<b>2,04</b>	<b>2,8</b>	<b>0,48</b>
279	0,26	0,23	0,48	-2,1	0,45	-2,2	0,78
280	0,99	0,25	1,05	0,3	1,07	0,3	0,49
281	-0,63	0,22	1,10	0,4	1,15	0,6	0,65
282	0,81	0,24	1,35	1,1	1,27	0,9	0,60
283	-0,63	0,22	0,96	0,0	0,94	-0,1	0,75
284	-1,21	0,24	1,02	0,2	1,17	0,6	0,52
285	0,42	0,23	1,44	1,4	1,48	1,5	0,19
<b>286</b>	<b>0,11</b>	<b>0,22</b>	<b>0,32</b>	<b>-3,1</b>	<b>0,34</b>	<b>-2,9</b>	<b>0,81</b>
287	-0,28	0,22	0,69	-1,1	0,69	-1,1	0,77
293	-0,04	0,22	0,81	-0,6	0,85	-0,4	0,45
294	-0,73	0,22	1,53	1,7	1,47	1,5	0,75
295	-0,73	0,22	0,72	-0,9	0,84	-0,5	0,60
296	0,11	0,22	0,85	-0,4	0,93	-0,1	0,27
297	1,26	0,26	0,57	-1,4	0,66	-1,0	0,51
298	-0,88	0,23	0,50	-2,0	0,56	-1,7	0,67

Suite du tableau 21

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
299	0,01	0,22	1,00	0,1	0,99	0,1	0,56
300	0,75	0,24	1,16	0,6	1,08	0,4	0,52
<b>301</b>	<b>-1,25</b>	<b>0,26</b>	<b>3,03</b>	<b>4,3</b>	<b>4,19</b>	<b>5,7</b>	<b>-0,10</b>
302	-0,68	0,22	0,46	-2,3	0,48	-2,1	0,82
<b>303</b>	<b>-0,14</b>	<b>0,22</b>	<b>0,31</b>	<b>-3,2</b>	<b>0,30</b>	<b>-3,3</b>	<b>0,78</b>
304	0,69	0,24	0,70	-1,0	0,72	-0,9	0,48
305	-0,19	0,22	0,63	-1,4	0,60	-1,5	0,71
306	-0,58	0,22	0,58	-1,6	0,77	-0,7	0,00
307	-3,42	0,47	0,76	-0,3	0,47	-0,8	0,76
308	-1,04	0,23	1,39	1,3	1,58	1,8	0,53
309	-0,38	0,22	0,62	-1,4	0,67	-1,2	0,67
341	-1,04	0,23	0,53	-1,8	0,58	-1,5	0,70
343	-0,28	0,22	1,29	1,0	1,21	0,8	0,65
<b>344</b>	<b>1,06</b>	<b>0,25</b>	<b>0,32</b>	<b>-2,8</b>	<b>0,31</b>	<b>-2,9</b>	<b>0,72</b>
345	1,26	0,26	0,55	-1,5	0,53	-1,6	0,59
346	-0,33	0,22	0,89	-0,3	0,87	-0,4	0,69
347	0,47	0,23	0,95	0,0	0,94	-0,1	0,60
348	0,31	0,23	1,53	1,6	1,47	1,5	0,41
349	0,16	0,22	0,70	-1,0	0,70	-1,0	0,60
350	-0,33	0,22	1,45	1,5	1,45	1,5	0,45
351	0,42	0,23	0,82	-0,5	0,83	-0,5	0,74
352	-0,19	0,22	0,75	-0,9	0,74	-0,9	0,72
353	-0,09	0,22	0,82	-0,5	0,84	-0,5	0,54
<b>MOY.</b>	<b>-0,31</b>	<b>0,24</b>	<b>0,99</b>	<b>-0,2</b>	<b>1,02</b>	<b>-0,2</b>	
<b>É.-T.</b>	<b>0,87</b>	<b>0,04</b>	<b>0,56</b>	<b>1,7</b>	<b>0,65</b>	<b>1,7</b>	

**Tableau 22 : Statistiques d'ajustement finales des personnes pour la version A du questionnaire 2004**

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1	-0,74	0,26	1,41	1,3	1,36	1,2	0,77
2	-1,01	0,26	1,74	2,1	1,66	1,9	0,60
3	1,14	0,28	0,99	0,1	1,00	0,1	0,68
4	-1,01	0,26	0,53	-1,8	0,50	-1,9	0,72
5	-1,37	0,27	1,13	0,5	1,16	0,6	0,38
7	0,76	0,27	0,57	-1,5	0,53	-1,7	0,67
8	0,62	0,26	1,17	0,6	1,12	0,5	0,42
10	-0,55	0,25	1,71	2,1	1,66	2,0	0,64
11	-0,61	0,26	1,59	1,8	1,56	1,7	0,78
12	-1,15	0,27	1,06	0,3	1,01	0,2	0,76
33	-0,94	0,26	1,38	1,2	1,38	1,2	0,81
34	-1,15	0,27	0,89	-0,3	0,84	-0,4	0,56
35	-0,94	0,27	0,75	-0,8	0,75	-0,8	0,76
61	-1,52	0,28	0,45	-2,1	0,48	-2,0	0,70
62	-1,52	0,28	1,06	0,3	0,95	0,0	0,68
63	0,15	0,25	0,46	-2,2	0,56	-1,7	0,48
64	-0,08	0,26	1,17	0,7	1,22	0,8	0,82
65	-0,29	0,25	1,38	1,2	1,43	1,4	0,81
66	-0,42	0,25	1,06	0,3	1,02	0,2	0,76
67	-2,27	0,31	0,56	-1,5	0,60	-1,3	0,51
68	-1,52	0,28	0,60	-1,4	0,54	-1,7	0,64
69	1,95	0,32	1,17	0,6	1,13	0,5	0,64
70	-0,55	0,25	0,81	-0,6	0,99	0,1	0,45
71	-0,48	0,25	0,46	-2,2	0,54	-1,8	0,54
74	-0,88	0,26	0,91	-0,2	0,97	0,0	0,78
75	0,91	0,27	1,00	0,1	0,89	-0,2	0,49
76	-0,29	0,25	1,01	0,1	1,05	0,3	0,71
77	-1,44	0,27	0,81	-0,5	0,80	-0,6	0,65
78	-1,08	0,26	1,35	1,1	1,34	1,1	0,77
79	0,91	0,27	0,55	-1,6	0,62	-1,3	0,40
80	-3,13	0,36	1,75	1,9	1,51	1,3	0,55
81	-1,59	0,28	0,39	-2,4	0,41	-2,4	0,62
82	-1,44	0,27	1,28	0,9	1,30	1,0	0,78
83	-1,29	0,27	0,83	-0,5	0,90	-0,2	0,16
124	-2,17	0,30	0,78	-0,6	0,67	-1,1	0,73
125	-0,10	0,25	0,84	-0,5	0,86	-0,4	0,54
126	-0,10	0,25	1,06	0,3	1,08	0,4	0,35

Suite du tableau 22

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
127	-1,75	0,28	1,94	2,4	1,78	2,1	0,64
144	-1,15	0,27	0,64	-1,2	0,63	-1,3	0,58
145	0,29	0,26	1,34	1,1	1,36	1,2	0,50
146	-0,36	0,25	1,34	1,2	1,39	1,3	0,57
147	1,95	0,32	0,67	-0,9	0,58	-1,3	0,35
148	-1,37	0,27	0,62	-1,3	0,58	-1,5	0,56
149	0,22	0,26	0,77	-0,7	0,87	-0,3	0,48
150	0,62	0,26	0,50	-1,9	0,49	-2,0	0,72
153	-0,61	0,26	1,18	0,7	1,13	0,5	0,61
161	-1,59	0,28	0,57	-1,5	0,59	-1,4	0,51
162	-1,52	0,28	0,84	-0,4	0,89	-0,2	0,48
163	-3,73	0,42	0,90	-0,1	0,81	-0,3	0,50
164	-3,56	0,40	0,55	-1,3	0,60	-1,0	0,67
165	-0,94	0,26	1,39	1,3	1,63	1,8	0,35
<b>MOY.</b>	<b>-0,76</b>	<b>0,28</b>	<b>0,98</b>	<b>-0,1</b>	<b>0,98</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>1,16</b>	<b>0,03</b>	<b>0,39</b>	<b>1,3</b>	<b>0,37</b>	<b>1,2</b>	

**Tableau 23 : Statistiques d'ajustement finales des personnes pour la version B du questionnaire 2004**

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
201	-1,27	0,25	0,48	-2,1	0,51	-1,9	0,83
202	-0,22	0,24	0,80	-0,6	0,80	-0,6	0,82
203	-2,82	0,35	1,02	0,2	1,25	0,7	0,42
204	0,12	0,24	1,12	0,5	1,14	0,6	0,63
205	0,87	0,26	0,43	-2,2	0,45	-2,1	0,65
206	0,06	0,24	0,93	-0,1	0,87	-0,3	0,68
208	0,55	0,25	1,04	0,2	1,02	0,2	0,68
209	0,74	0,26	0,93	-0,1	0,87	-0,3	0,63
210	-1,14	0,25	0,95	-0,1	1,01	0,1	0,84
223	-0,79	0,24	0,66	-1,2	0,65	-1,3	0,69
231	-2,04	0,29	1,19	0,7	1,11	0,4	0,61
232	-1,27	0,25	0,68	-1,1	0,67	-1,1	0,74
259	-2,49	0,32	0,76	-0,6	0,78	-0,4	0,64
260	-0,22	0,24	0,74	-0,8	0,72	-0,9	0,61
261	-1,02	0,24	0,68	-1,1	0,64	-1,3	0,67
262	-0,11	0,24	1,12	0,5	1,14	0,5	0,64
276	-0,39	0,24	0,67	-1,2	0,66	-1,2	0,72
277	-0,96	0,24	1,77	2,2	1,80	2,3	0,64
279	0,30	0,25	0,62	-1,3	0,59	-1,5	0,74
280	1,15	0,27	1,29	0,9	1,34	1,1	0,45
281	-0,73	0,24	1,27	1,0	1,32	1,1	0,66
282	0,94	0,26	1,60	1,7	1,49	1,5	0,59
283	-0,73	0,24	1,04	0,2	1,02	0,2	0,77
284	-1,39	0,25	1,05	0,3	1,20	0,7	0,58
285	0,48	0,25	1,59	1,7	1,57	1,7	0,25
287	-0,33	0,24	0,78	-0,7	0,78	-0,7	0,77
293	-0,05	0,24	1,01	0,2	1,10	0,4	0,40
294	-0,85	0,24	1,69	2,1	1,63	1,9	0,76
295	-0,85	0,24	0,84	-0,5	0,96	0,0	0,60
296	0,12	0,24	1,03	0,2	1,09	0,4	0,27
297	1,46	0,28	0,68	-1,0	0,79	-0,6	0,50
298	-1,02	0,24	0,53	-1,8	0,58	-1,6	0,71
299	0,01	0,24	1,22	0,8	1,22	0,8	0,53
300	0,87	0,26	1,36	1,1	1,27	0,9	0,52
302	-0,79	0,24	0,48	-2,1	0,52	-1,9	0,85
304	0,81	0,26	0,78	-0,6	0,79	-0,6	0,51
305	-0,22	0,24	0,76	-0,8	0,72	-1,0	0,70

Suite du tableau 23

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
306	-0,67	0,24	0,72	-1,0	0,91	-0,2	0,00
307	-3,81	0,48	0,68	-0,5	0,40	-0,9	0,76
308	-1,20	0,25	1,55	1,7	1,73	2,1	0,55
309	-0,45	0,24	0,72	-0,9	0,76	-0,8	0,67
341	-1,20	0,25	0,57	-1,6	0,64	-1,3	0,72
343	-0,33	0,24	1,42	1,4	1,33	1,1	0,68
345	1,46	0,28	0,65	-1,1	0,63	-1,2	0,57
346	-0,39	0,24	1,06	0,3	1,03	0,2	0,68
347	0,55	0,25	1,11	0,4	1,08	0,4	0,60
348	0,36	0,25	1,74	2,1	1,66	1,9	0,44
349	0,18	0,24	0,84	-0,4	0,83	-0,5	0,60
350	-0,39	0,24	1,66	2,0	1,66	2,0	0,47
351	0,48	0,25	0,99	0,1	0,99	0,1	0,72
352	-0,22	0,24	0,86	-0,4	0,84	-0,5	0,73
353	-0,11	0,24	0,99	0,1	1,03	0,2	0,52
<b>MOY.</b>	<b>-0,36</b>	<b>0,26</b>	<b>0,98</b>	<b>-0,1</b>	<b>0,99</b>	<b>0,0</b>	
<b>É.-T.</b>	<b>1,02</b>	<b>0,04</b>	<b>0,35</b>	<b>1,1</b>	<b>0,35</b>	<b>1,1</b>	



**Tableau 24 : Statistiques d'ajustement des items pour la version A du questionnaire 2004**

<b>Item</b>	<b>Mesure estimée</b>	<b>Erreur type</b>	<b>Infit MNSQ</b>	<b>Infit ZSTD</b>	<b>Outfit MNSQ</b>	<b>Outfit ZSTD</b>	<b>Corr.</b>
1	0,38	0,17	0,86	-0,7	0,86	-0,7	0,70
2	1,78	0,20	1,11	0,6	1,09	0,5	0,63
3	0,46	0,17	1,29	1,4	1,32	1,5	0,62
4	-0,43	0,16	1,24	1,3	1,18	1,0	0,65
5	-0,22	0,16	1,10	0,6	1,05	0,3	0,68
6	0,02	0,16	0,80	-1,1	0,78	-1,2	0,75
7	0,32	0,17	0,62	-2,2	0,64	-2,1	0,76
8	-2,42	0,19	1,01	0,1	1,05	0,3	0,65
9	-0,62	0,16	0,72	-1,6	0,75	-1,4	0,80
10	-0,75	0,16	1,03	0,2	1,06	0,4	0,65
11	-0,28	0,16	1,07	0,4	1,03	0,2	0,65
<b>12</b>	<b>-1,46</b>	<b>0,17</b>	<b>1,58</b>	<b>2,6</b>	<b>1,54</b>	<b>2,5</b>	<b>0,55</b>
<b>13</b>	<b>-0,30</b>	<b>0,16</b>	<b>0,57</b>	<b>-2,6</b>	<b>0,55</b>	<b>-2,8</b>	<b>0,79</b>
14	-0,41	0,16	1,25	1,3	1,28	1,4	0,60
15	0,38	0,17	1,10	0,6	1,01	0,1	0,71
16	0,26	0,17	0,61	-2,3	0,59	-2,4	0,76
17	0,43	0,17	1,04	0,3	1,02	0,2	0,66
18	0,28	0,17	0,99	0,0	1,01	0,1	0,67
19	0,97	0,18	0,77	-1,2	0,70	-1,6	0,75
20	1,59	0,19	1,05	0,3	0,99	0,0	0,64
<b>MOY.</b>	<b>0,00</b>	<b>0,17</b>	<b>0,99</b>	<b>-0,1</b>	<b>0,98</b>	<b>-0,2</b>	
<b>É.-T.</b>	<b>0,92</b>	<b>0,01</b>	<b>0,25</b>	<b>1,3</b>	<b>0,25</b>	<b>1,3</b>	

**Tableau 25 : Statistiques d'ajustement des items pour la version B du questionnaire 2004**

Item	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1	0,45	0,15	1,03	0,2	1,06	0,4	0,54
2	1,30	0,16	1,31	1,5	1,48	2,0	0,40
3	0,70	0,15	1,25	1,3	1,17	0,9	0,61
4	-0,25	0,15	0,79	-1,1	0,83	-0,9	0,65
5	0,93	0,15	1,25	1,3	1,19	1,0	0,66
6	0,24	0,15	1,23	1,2	1,20	1,0	0,73
7	-0,05	0,15	0,67	-1,9	0,65	-2,0	0,77
8	-2,35	0,19	1,09	0,5	1,11	0,6	0,64
9	-0,46	0,15	0,77	-1,3	0,73	-1,5	0,70
10	-0,88	0,16	0,58	-2,4	0,59	-2,4	0,82
11	-0,63	0,16	0,86	-0,7	0,84	-0,8	0,71
12	-1,87	0,18	1,33	1,5	1,35	1,7	0,49
13	-0,44	0,15	0,76	-1,3	0,75	-1,4	0,70
14	-0,21	0,15	1,15	0,8	1,13	0,7	0,70
15	0,38	0,15	1,38	1,9	1,32	1,6	0,59
16	0,42	0,15	0,71	-1,7	0,70	-1,7	0,65
17	0,18	0,15	1,01	0,1	1,11	0,6	0,58
18	0,70	0,15	0,71	-1,7	0,70	-1,6	0,72
19	0,70	0,15	0,86	-0,7	0,83	-0,8	0,69
20	1,17	0,16	1,20	1,1	1,11	0,6	0,64
<b>MOY.</b>	<b>0,00</b>	<b>0,16</b>	<b>1,00</b>	<b>-0,1</b>	<b>0,99</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>0,91</b>	<b>0,01</b>	<b>0,25</b>	<b>1,3</b>	<b>0,26</b>	<b>1,3</b>	

**Tableau 26 : Statistiques d'ajustement générales des personnes pour la version A du questionnaire 2004**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	60,5	20,0	-0,76	0,28	0,98	-0,1	0,98	-0,1
<b>É.-T.</b>	15,6	0,2	1,16	0,03	0,39	1,3	0,37	1,2
<b>MAX.</b>	97,0	20,0	1,95	0,42	1,94	2,4	1,78	2,1
<b>MIN.</b>	28,0	19,0	-3,73	0,25	0,39	-2,4	0,41	-2,4

<b>Réel</b>	<b>Séparation</b>	3,79	« Fidélité »	0,93
<b>Modèle</b>	<b>Séparation</b>	4,07	« Fidélité »	0,94

**Tableau 27 : Statistiques d'ajustement générales des personnes pour la version B du questionnaire 2004**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	65,9	20,0	-0,36	0,26	0,98	-0,1	0,99	0,0
<b>É.-T.</b>	15,5	0,0	1,02	0,04	0,35	1,1	0,35	1,1
<b>MAX.</b>	94,0	20,0	1,46	0,48	1,77	2,2	1,80	2,3
<b>MIN.</b>	25,0	20,0	-3,81	0,24	0,43	-2,2	0,40	-2,1

<b>Réel</b>	<b>Séparation</b>	3,59	« Fidélité »	0,93
<b>Modèle</b>	<b>Séparation</b>	3,82	« Fidélité »	0,94

**Tableau 28 : Statistiques d'ajustement générales des items pour la version A du questionnaire 2004**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	154,2	50,9	0,00	0,17	0,99	-0,1	0,98	-0,2
<b>É.-T.</b>	32,0	0,3	0,92	0,01	0,25	1,3	0,25	1,3
<b>MAX.</b>	238,0	51,0	1,78	0,20	1,58	2,6	1,54	2,5
<b>MIN.</b>	97,0	50,0	-2,42	0,16	0,57	-2,6	0,55	-2,8

<b>Réel</b>	<b>Séparation</b>	5,07	« Fidélité »	0,96
<b>Modèle</b>	<b>Séparation</b>	5,31	« Fidélité »	0,97

**Tableau 29 : Statistiques d'ajustement générales des items pour la version B du questionnaire 2004**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	171,4	52	0,00	0,16	1,00	-0,1	0,99	-0,1
<b>É.-T.</b>	36,8	0	0,91	0,01	0,25	1,3	0,26	1,3
<b>MAX.</b>	259,0	52	1,30	0,19	1,38	1,9	1,48	2,0
<b>MIN.</b>	117,0	52	-2,35	0,15	0,58	-2,4	0,59	-2,4

<b>Réel</b>	<b>Séparation</b>	5,43	« Fidélité »	0,97
<b>Modèle</b>	<b>Séparation</b>	5,74	« Fidélité »	0,97

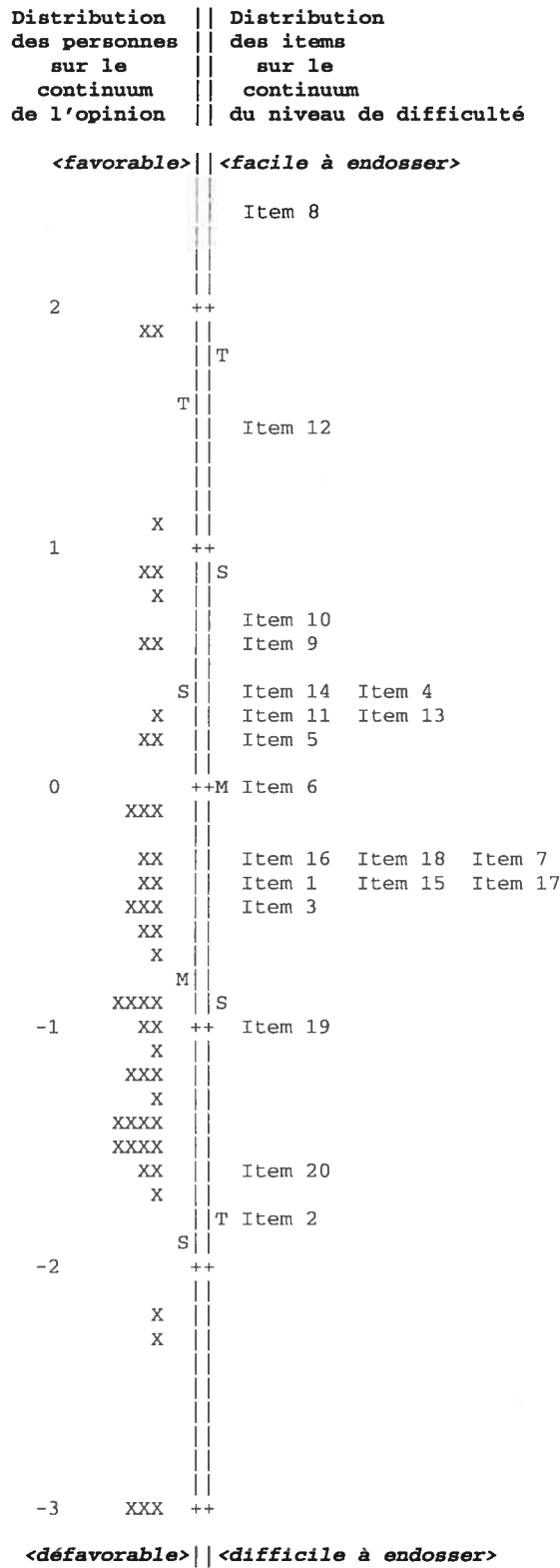


Figure 11 : Distribution des personnes sur le continuum de l'habileté et des items sur le continuum du niveau de difficulté pour la version A du questionnaire 2004



Tableau 30 : Points d'ancrage de la version A du questionnaire 2004

Catégorie	Nb. d'obs.	% d'obs.	Mesure moyenne observée	Mesure moyenne attendue	Infit MNSQ	Outfit MNSQ	Points d'ancrage	Erreur type
1	123	12	-2,60	-2,63	1,07	1,04	NONE	
2	269	26	-1,61	-1,57	0,90	0,90	-2,84	0,12
3	268	26	-0,80	-0,80	0,99	0,96	-1,17	0,08
4	193	19	0,08	0,00	0,89	0,92	-0,08	0,09
5	142	14	0,96	0,96	1,02	1,01	0,77	0,11
6	23	2	1,88	2,09	1,23	1,12	3,32	0,24

Tableau 31 : Points d'ancrage de la version B du questionnaire 2004

Catégorie	Nb. d'obs.	% d'obs.	Mesure moyenne observée	Mesure moyenne attendue	Infit MNSQ	Outfit MNSQ	Points d'ancrage	Erreur type
1	125	12	-2,24	-2,17	0,85	0,87	NONE	
2	190	18	-1,12	-1,22	1,12	1,08	-2,06	0,12
3	250	24	-0,65	-0,57	1,01	1,01	-1,16	0,09
4	251	24	0,16	0,08	0,84	0,89	-0,26	0,08
5	176	17	0,86	0,89	1,11	1,10	0,82	0,10
6	48	5	1,81	1,86	1,05	1,03	2,66	0,17

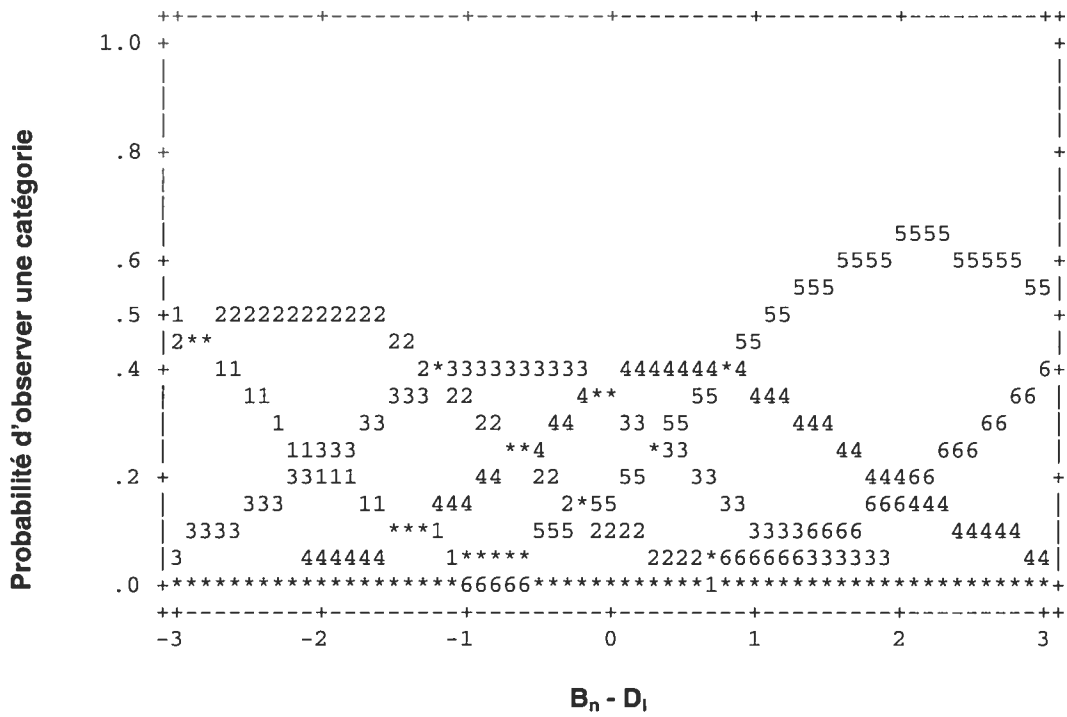


Figure 13 : Courbes de probabilités de chacune des catégories de réponse de la version A du questionnaire 2004



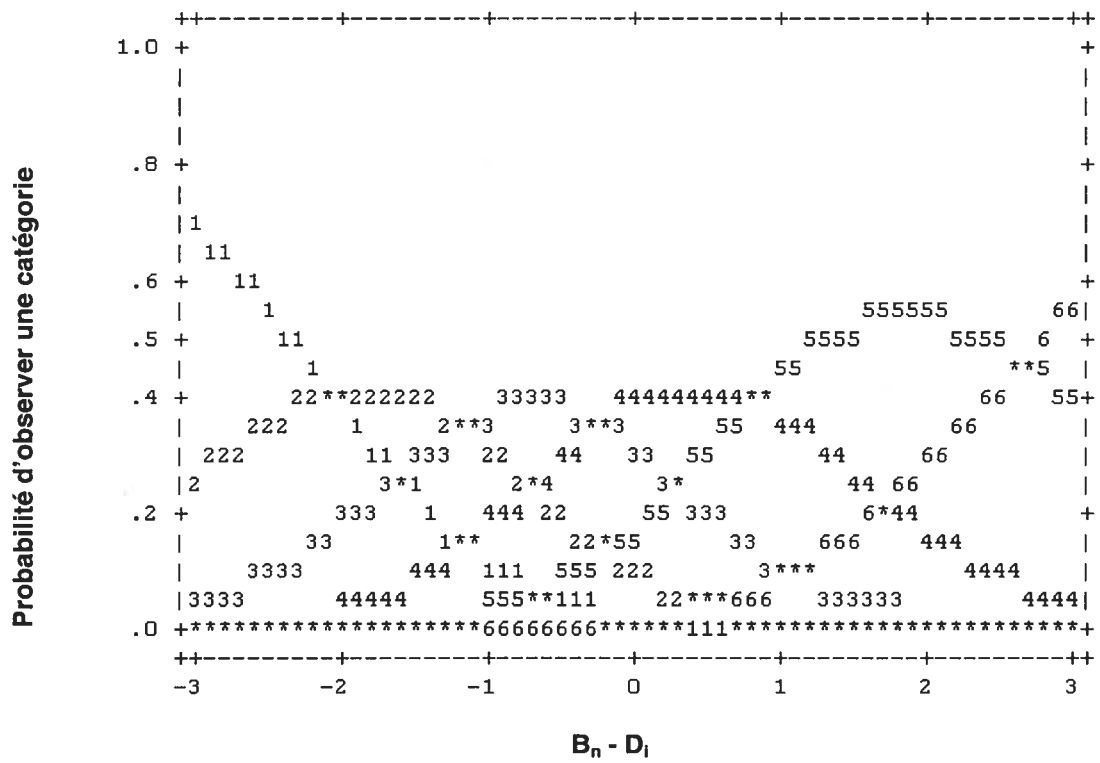
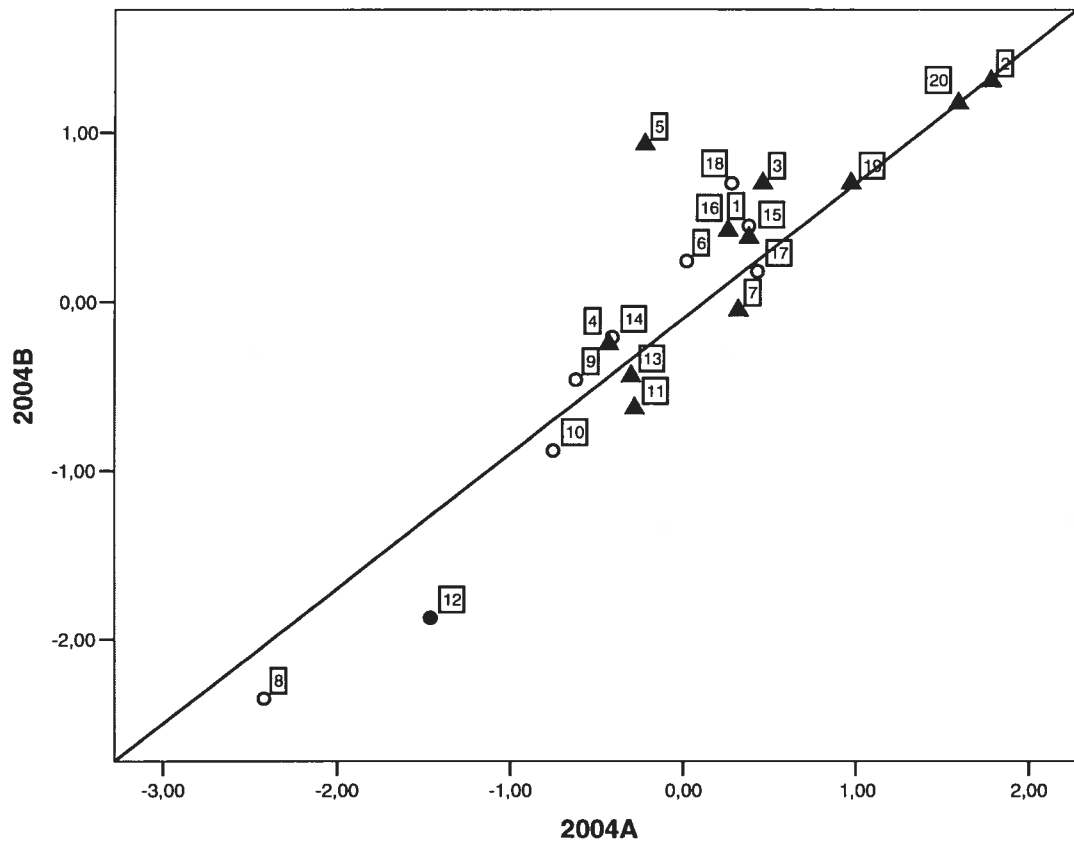


Figure 14 : Courbes de probabilités de chacune des catégories de réponse de la version B du questionnaire 2004



**Figure 15 : Diagramme de dispersion des items des versions A et B du questionnaire 2004 par rapport au niveau de difficulté des items**

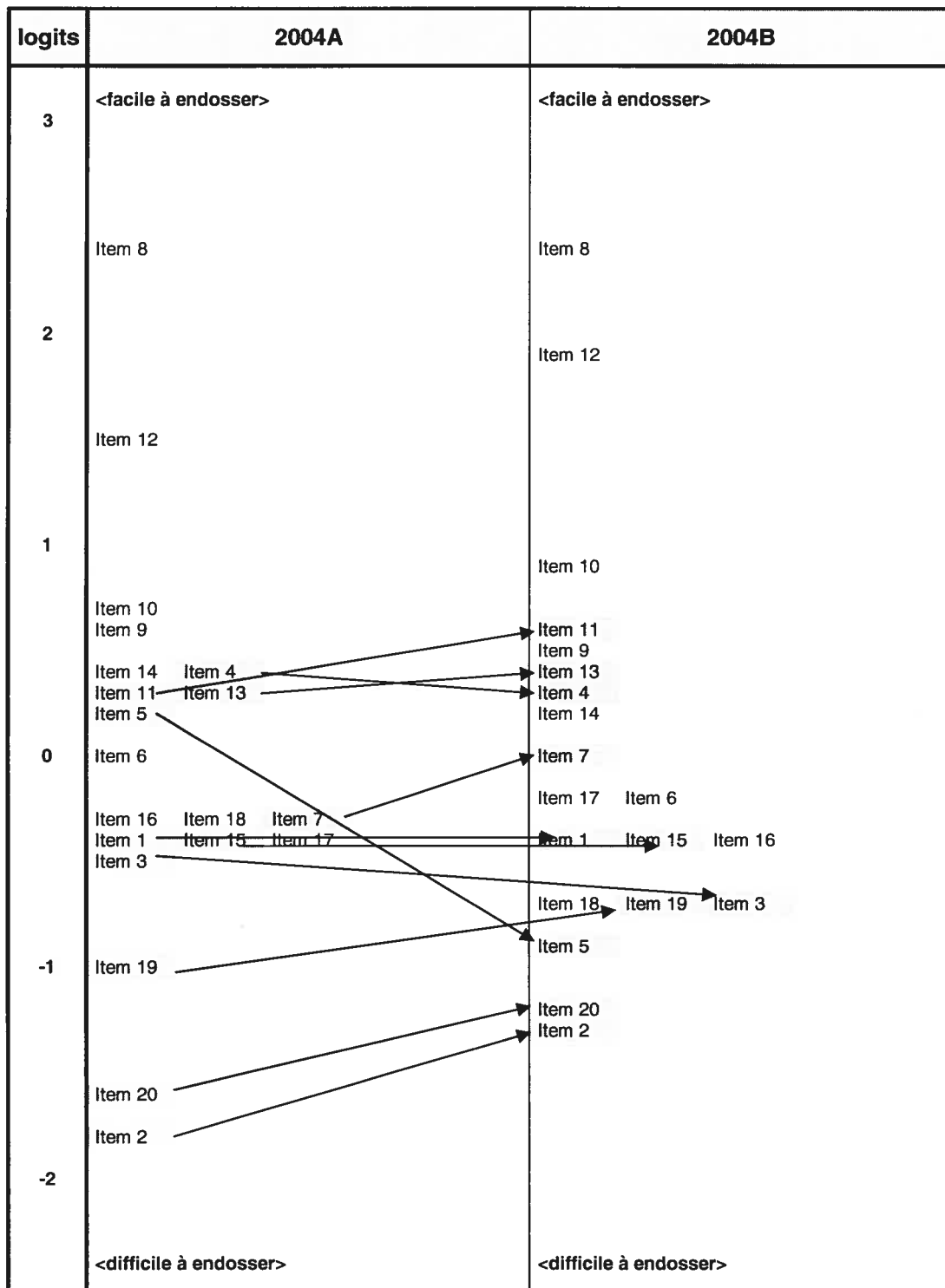


Figure 16 : Diagramme de comparaison montrant la distribution des items sur le continuum du niveau de difficulté pour les versions A et B du questionnaire 2004

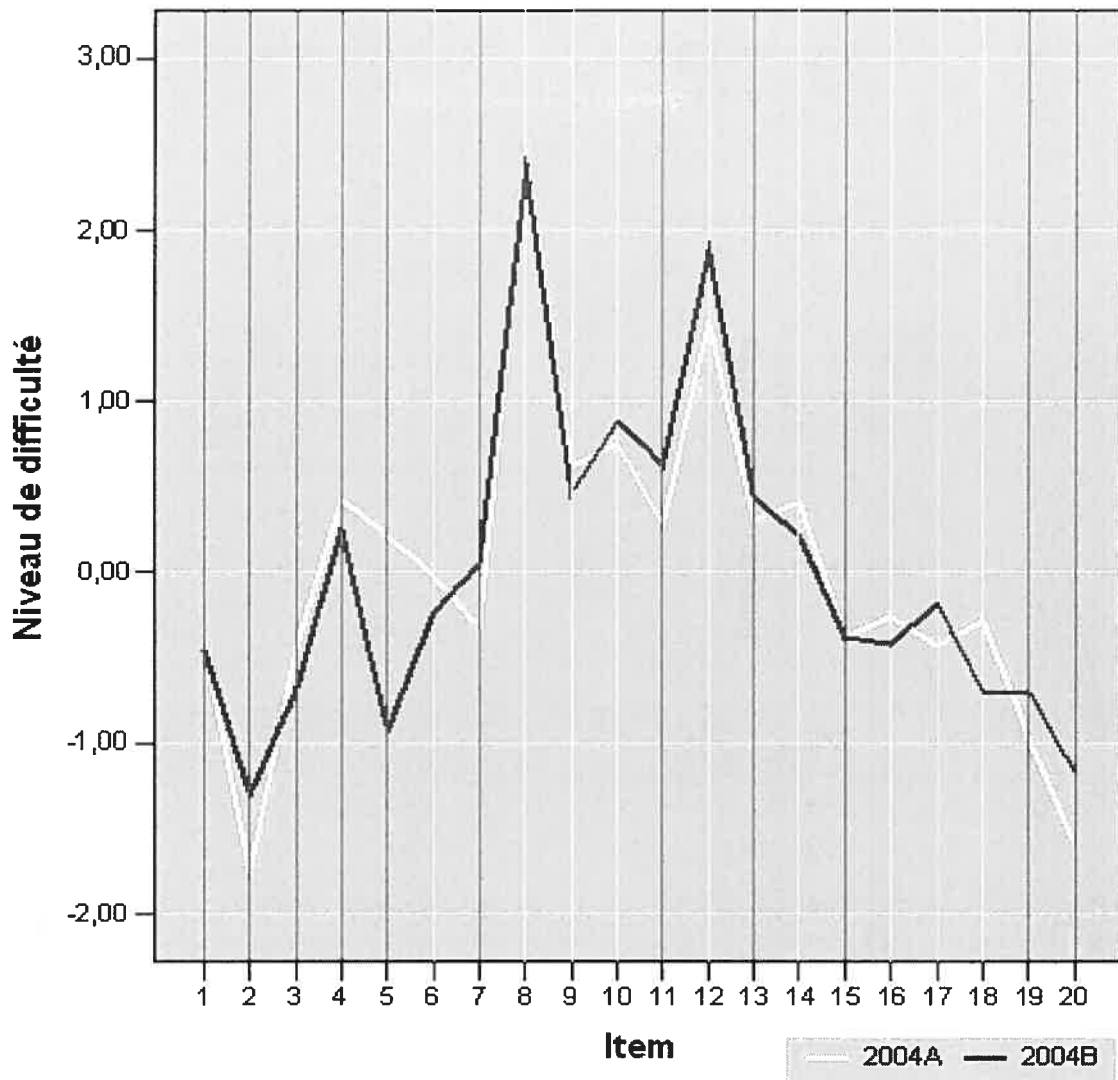


Figure 17 : Courbes de distribution des mesures estimées par le modèle pour le niveau de difficulté des items des versions A et B du questionnaire 2004

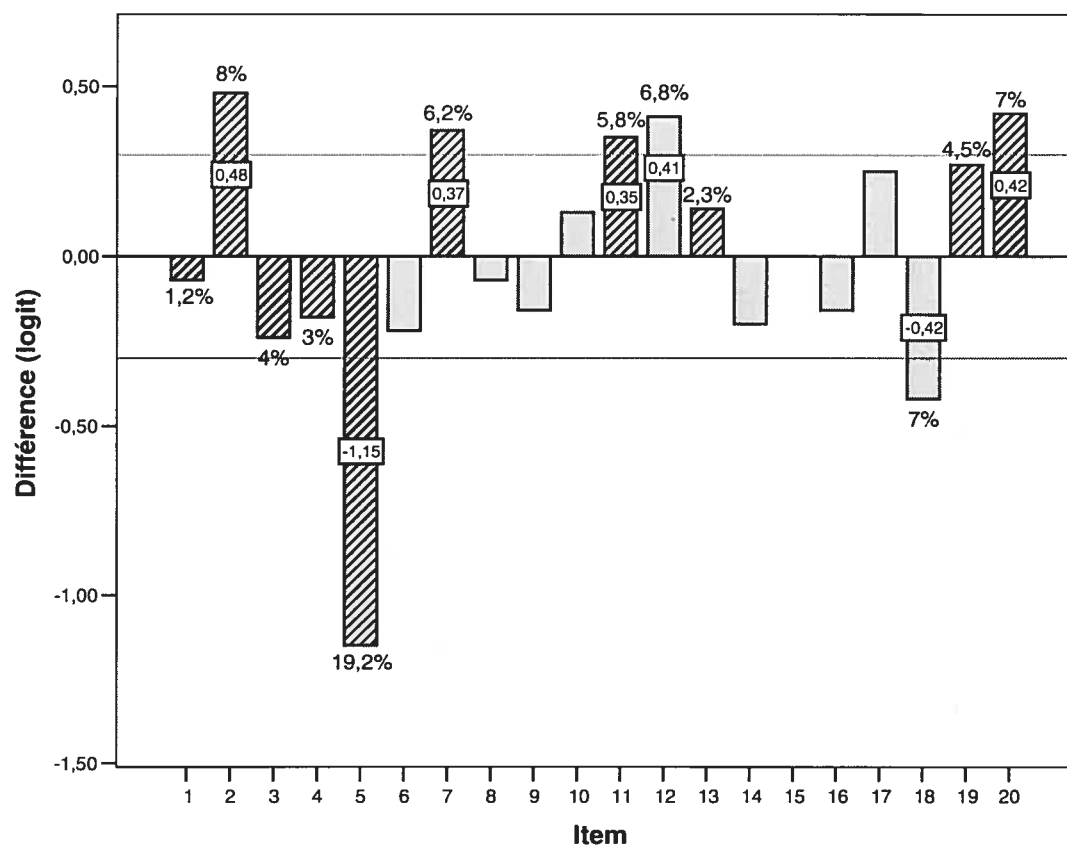
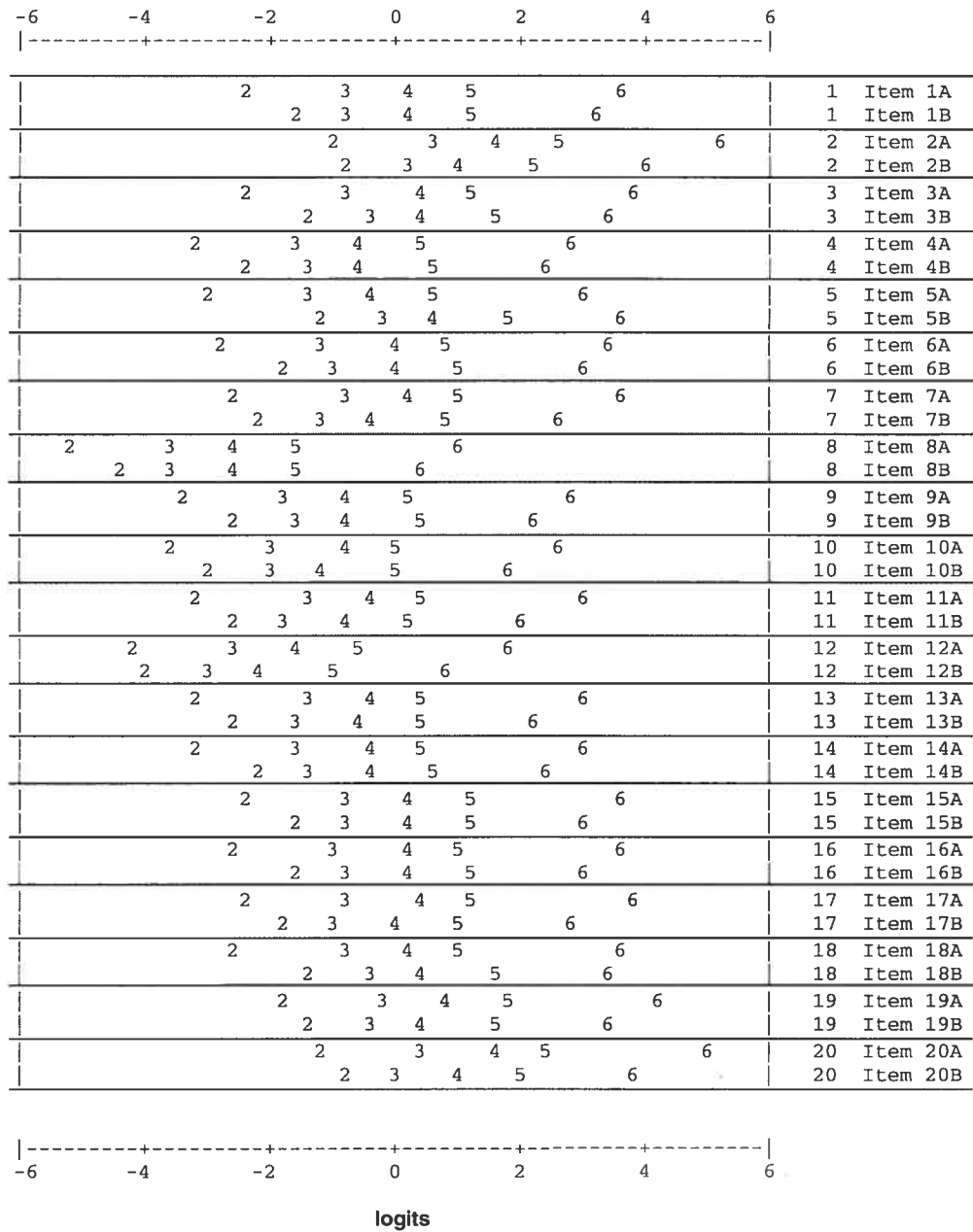


Figure 18 : Diagramme en bâton représentant la valeur des différences entre les mesures estimées par le modèle pour les versions A et B du questionnaire 2004



**Figure 19 : Diagramme de comparaison montrant la position des points d'ancrage estimés par le modèle pour chacun des items des versions A et B du questionnaire 2004**

### C.3. Tableaux de l'analyse des questionnaires 2005

Tableau 32 : Statistiques d'ajustement initiales des personnes pour la version A du questionnaire 2005

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1	1,29	0,28	3,00	4,3	2,91	4,2	0,51
2	1,29	0,28	1,35	1,1	1,27	0,9	0,33
3	-0,09	0,25	1,86	2,4	1,84	2,3	0,35
4	0,99	0,27	1,10	0,4	1,11	0,4	0,32
5	-0,66	0,25	0,52	-1,9	0,52	-1,8	0,74
6	-0,79	0,25	1,34	1,2	1,35	1,2	0,53
7	-0,22	0,25	1,20	0,7	1,23	0,8	0,36
8	-0,03	0,25	1,05	0,3	1,06	0,3	0,68
9	-0,22	0,25	0,71	-1,0	0,77	-0,7	0,05
10	-1,44	0,26	1,65	1,9	1,65	2,0	0,46
11	<b>-2,76</b>	<b>0,33</b>	<b>2,29</b>	<b>3,0</b>	<b>1,94</b>	<b>2,3</b>	<b>0,47</b>
15	-0,53	0,25	1,16	0,6	1,16	0,6	0,36
17	-0,60	0,25	0,81	-0,6	0,81	-0,6	0,38
18	1,87	0,30	0,47	-2,0	0,48	-2,0	0,76
19	0,23	0,26	0,77	-0,7	0,78	-0,7	0,65
20	1,14	0,27	0,54	-1,7	0,54	-1,7	0,45
40	0,63	0,26	0,57	-1,5	0,54	-1,7	0,59
41	2,34	0,32	1,30	1,0	1,43	1,3	0,23
42	1,14	0,27	0,65	-1,2	0,62	-1,3	0,58
43	0,84	0,27	1,11	0,5	1,09	0,4	0,58
44	0,23	0,26	1,05	0,3	1,08	0,4	0,83
45	-0,72	0,25	0,41	-2,5	0,44	-2,3	0,42
46	0,87	0,27	0,75	-0,8	0,75	-0,8	0,52
47	-1,11	0,26	0,59	-1,6	0,59	-1,5	0,68
48	1,45	0,28	1,35	1,1	1,35	1,1	0,54
49	0,17	0,26	0,44	-2,3	0,45	-2,2	0,51
50	<b>1,61</b>	<b>0,29</b>	<b>0,36</b>	<b>-2,6</b>	<b>0,37</b>	<b>-2,6</b>	<b>0,37</b>
51	-1,44	0,26	1,12	0,5	1,07	0,3	0,54
52	-1,37	0,26	0,78	-0,7	0,77	-0,8	0,20
53	-1,11	0,26	0,68	-1,1	0,69	-1,1	0,67
54	-0,09	0,25	0,74	-0,8	0,75	-0,8	0,54
55	-1,51	0,26	0,66	-1,2	0,70	-1,0	0,17
67	<b>-1,04</b>	<b>0,25</b>	<b>1,89</b>	<b>2,5</b>	<b>1,92</b>	<b>2,6</b>	<b>0,15</b>
73	0,17	0,26	1,09	0,4	1,09	0,4	0,58

Suite du tableau 32

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
74	2,24	0,31	0,33	-2,8	0,31	-3,0	0,39
75	0,30	0,26	0,62	-1,3	0,63	-1,3	0,82
76	0,23	0,26	1,27	0,9	1,24	0,8	0,49
77	1,06	0,27	0,57	-1,5	0,57	-1,5	0,54
78	-1,30	0,26	0,98	0,0	1,03	0,2	0,20
79	0,50	0,26	1,02	0,2	1,04	0,2	0,69
80	0,84	0,27	0,37	-2,6	0,46	-2,1	0,38
87	0,92	0,27	0,92	-0,1	0,88	-0,3	0,66
88	0,36	0,26	2,04	2,7	2,00	2,7	0,59
89	0,10	0,25	1,00	0,1	1,00	0,1	0,40
90	-0,28	0,25	0,83	-0,5	0,83	-0,5	0,40
91	-0,34	0,25	0,63	-1,3	0,62	-1,3	0,55
92	-0,66	0,25	0,83	-0,5	0,82	-0,5	0,68
93	0,04	0,25	1,02	0,2	0,99	0,1	0,62
94	-0,28	0,25	1,04	0,2	1,06	0,3	0,60
108	-2,56	0,31	1,50	1,4	1,48	1,4	0,46
109	-0,28	0,25	0,99	0,1	0,98	0,0	0,56
125	0,99	0,27	0,58	-1,5	0,55	-1,6	0,73
126	-0,79	0,25	0,86	-0,4	0,85	-0,4	0,38
127	0,43	0,26	0,32	-3,0	0,33	-2,9	0,44
128	0,10	0,25	0,50	-1,9	0,53	-1,8	0,42
129	-0,15	0,25	2,52	3,7	2,53	3,7	-0,05
130	1,96	0,30	1,38	1,2	1,27	0,9	0,49
131	0,50	0,26	0,22	-3,8	0,22	-3,8	0,73
132	1,29	0,28	1,44	1,3	1,42	1,3	0,62
133	-1,17	0,26	0,75	-0,8	0,71	-1,0	0,56
134	0,77	0,27	0,99	0,1	0,99	0,1	0,30
135	0,70	0,26	1,20	0,7	1,17	0,6	0,59
136	2,14	0,31	0,30	-3,0	0,30	-3,1	0,54
137	2,24	0,31	0,56	-1,5	0,64	-1,2	0,38
138	1,21	0,28	2,64	3,8	2,48	3,5	0,49
147	0,43	0,26	1,17	0,6	1,16	0,6	0,28
148	0,43	0,26	1,07	0,3	1,03	0,2	0,64
149	-0,28	0,25	0,42	-2,4	0,42	-2,4	0,69
150	0,43	0,26	0,80	-0,6	0,78	-0,7	0,21
151	1,53	0,29	0,93	-0,1	0,94	-0,1	0,44
152	0,10	0,25	0,98	0,0	0,98	0,0	0,35
153	0,50	0,26	1,05	0,3	1,02	0,2	0,37
<b>MOY.</b>	<b>0,21</b>	<b>0,27</b>	<b>1,00</b>	<b>-0,2</b>	<b>0,99</b>	<b>-0,2</b>	
<b>É.-T.</b>	<b>1,07</b>	<b>0,02</b>	<b>0,55</b>	<b>1,7</b>	<b>0,53</b>	<b>1,6</b>	



**Tableau 33 : Statistiques d'ajustement initiales des personnes pour la version B du questionnaire 2005**

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
502	0,63	0,23	0,41	-2,4	0,41	-2,5	0,55
503	1,93	0,28	0,77	-0,7	0,86	-0,3	0,51
504	0,63	0,23	0,81	-0,6	1,07	0,3	0,16
<b>523</b>	<b>0,63</b>	<b>0,23</b>	<b>0,40</b>	<b>-2,5</b>	<b>0,39</b>	<b>-2,6</b>	<b>0,70</b>
524	2,45	0,32	0,64	-1,1	0,60	-1,0	0,56
<b>525</b>	<b>0,19</b>	<b>0,23</b>	<b>2,21</b>	<b>3,1</b>	<b>2,19</b>	<b>3,1</b>	<b>0,44</b>
531	-2,19	0,30	0,74	-0,7	0,70	-0,9	0,83
540	-1,03	0,25	0,84	-0,5	0,82	-0,5	0,63
<b>541</b>	<b>-0,13</b>	<b>0,23</b>	<b>2,20</b>	<b>3,1</b>	<b>2,18</b>	<b>3,1</b>	<b>0,52</b>
542	-0,13	0,23	0,66	-1,2	0,71	-1,0	0,17
543	-0,97	0,24	0,69	-1,1	0,65	-1,3	0,76
544	-1,77	0,28	1,21	0,7	1,09	0,4	0,72
545	-0,91	0,24	0,63	-1,3	0,64	-1,3	0,24
546	1,38	0,25	0,69	-1,0	0,73	-0,8	0,65
547	-0,45	0,24	0,58	-1,5	0,58	-1,6	0,52
548	1,93	0,28	0,67	-1,1	0,63	-1,1	0,68
557	-0,68	0,24	0,43	-2,3	0,43	-2,4	0,65
558	-0,79	0,24	1,38	1,2	1,38	1,2	0,50
570	-1,03	0,25	0,50	-1,9	0,48	-2,0	0,35
571	-0,68	0,24	0,54	-1,7	0,56	-1,7	0,55
572	-0,18	0,23	1,66	1,9	1,67	2,0	0,72
573	-1,77	0,28	0,55	-1,6	0,61	-1,3	0,76
574	-1,85	0,28	1,97	2,5	1,82	2,2	0,61
575	-1,85	0,28	1,13	0,5	1,02	0,2	0,65
<b>576</b>	<b>-0,40</b>	<b>0,23</b>	<b>0,34</b>	<b>-2,9</b>	<b>0,34</b>	<b>-2,9</b>	<b>0,89</b>
<b>577</b>	<b>-1,03</b>	<b>0,25</b>	<b>1,98</b>	<b>2,6</b>	<b>1,89</b>	<b>2,5</b>	<b>0,56</b>
578	2,79	0,36	0,59	-1,1	0,68	-0,6	0,50
579	1,08	0,24	1,24	0,8	1,18	0,7	0,32
580	0,91	0,24	0,87	-0,3	0,83	-0,5	0,62
<b>596</b>	<b>0,74</b>	<b>0,24</b>	<b>2,04</b>	<b>2,8</b>	<b>2,33</b>	<b>3,4</b>	<b>0,13</b>
597	1,26	0,25	0,50	-2,0	0,50	-1,9	0,79
598	-0,02	0,23	0,69	-1,1	0,68	-1,1	0,63
607	-1,08	0,25	1,75	2,1	1,79	2,2	0,70
608	-0,34	0,23	0,65	-1,2	0,64	-1,3	0,65

Suite du tableau 33

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
609	-1,09	0,25	1,08	0,4	1,12	0,5	0,29
<b>610</b>	<b>-1,70</b>	<b>0,27</b>	<b>0,24</b>	<b>-3,5</b>	<b>0,25</b>	<b>-3,5</b>	<b>0,59</b>
611	0,25	0,23	0,51	-1,9	0,50	-1,9	0,73
612	-0,62	0,24	0,65	-1,2	0,66	-1,2	0,64
613	-0,79	0,24	0,80	-0,6	0,77	-0,7	0,38
620	0,74	0,24	1,13	0,5	1,14	0,6	0,26
<b>621</b>	<b>0,63</b>	<b>0,23</b>	<b>3,06</b>	<b>4,6</b>	<b>3,02</b>	<b>4,5</b>	<b>0,38</b>
<b>MOY.</b>	<b>-0,13</b>	<b>0,25</b>	<b>0,99</b>	<b>-0,3</b>	<b>0,99</b>	<b>-0,3</b>	
<b>É.-T.</b>	<b>1,21</b>	<b>0,03</b>	<b>0,63</b>	<b>1,9</b>	<b>0,63</b>	<b>1,9</b>	

**Tableau 34 : Statistiques d'ajustement finales des personnes pour la version A du questionnaire 2005**

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
2	1,48	0,29	1,47	1,4	1,39	1,2	0,33
4	1,15	0,28	1,22	0,8	1,23	0,8	0,31
5	-0,69	0,27	0,59	-1,5	0,60	-1,5	0,74
6	-0,83	0,27	1,54	1,6	1,56	1,7	0,53
7	-0,19	0,27	1,39	1,3	1,41	1,3	0,33
8	0,03	0,27	1,16	0,6	1,17	0,6	0,68
9	-0,19	0,27	0,80	-0,6	0,84	-0,4	0,06
10	-1,59	0,28	1,94	2,5	1,91	2,5	0,45
15	-0,54	0,27	1,30	1,0	1,32	1,1	0,37
17	-0,61	0,27	0,93	-0,1	0,94	-0,1	0,36
18	2,10	0,31	0,49	-1,9	0,49	-1,9	0,76
19	0,32	0,27	0,86	-0,4	0,87	-0,3	0,65
20	1,31	0,29	0,58	-1,5	0,58	-1,5	0,47
40	0,76	0,28	0,64	-1,2	0,62	-1,3	0,58
41	2,60	0,33	1,39	1,2	1,54	1,6	0,23
42	1,31	0,29	0,72	-0,9	0,70	-1,0	0,57
43	0,99	0,28	1,17	0,6	1,15	0,6	0,61
44	0,32	0,27	1,16	0,6	1,19	0,7	0,83
45	-0,76	0,27	0,47	-2,1	0,49	-2,0	0,42
46	1,02	0,29	0,82	-0,5	0,82	-0,5	0,52
47	-1,20	0,27	0,67	-1,1	0,68	-1,1	0,69
48	1,65	0,29	1,42	1,3	1,41	1,3	0,56
49	0,24	0,27	0,49	-2,0	0,50	-1,9	0,50
50	1,82	0,30	0,42	-2,3	0,43	-2,2	0,35
51	-1,59	0,28	1,30	1,0	1,23	0,8	0,55
52	-1,51	0,28	0,86	-0,4	0,83	-0,5	0,22
53	-1,20	0,27	0,79	-0,6	0,82	-0,5	0,68
54	-0,04	0,27	0,82	-0,5	0,83	-0,5	0,56
55	-1,67	0,28	0,75	-0,8	0,79	-0,6	0,16
73	0,24	0,27	1,20	0,7	1,20	0,7	0,58
75	0,39	0,27	0,68	-1,1	0,69	-1,0	0,82
76	0,32	0,27	1,40	1,3	1,37	1,2	0,50
77	1,23	0,28	0,62	-1,3	0,62	-1,3	0,54
78	-1,43	0,28	1,13	0,5	1,19	0,7	0,18
79	0,61	0,27	1,12	0,5	1,15	0,6	0,68
80	0,99	0,28	0,41	-2,4	0,48	-2,0	0,41
87	1,07	0,28	1,00	0,1	0,96	0,0	0,67

Suite du tableau 34

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
89	0,17	0,27	1,11	0,5	1,12	0,5	0,41
90	-0,26	0,27	0,96	0,0	0,97	0,0	0,37
91	-0,33	0,27	0,71	-1,0	0,71	-1,0	0,56
92	-0,69	0,27	0,95	-0,1	0,94	-0,1	0,69
93	0,10	0,27	1,15	0,6	1,12	0,5	0,62
94	-0,26	0,27	1,20	0,7	1,22	0,8	0,58
108	-2,88	0,33	1,72	2,0	1,65	1,8	0,47
109	-0,26	0,27	1,11	0,5	1,10	0,4	0,56
125	1,15	0,28	0,61	-1,4	0,58	-1,5	0,75
126	-0,83	0,27	0,97	0,0	0,96	0,0	0,39
128	0,17	0,27	0,57	-1,6	0,59	-1,5	0,42
130	2,19	0,31	1,46	1,4	1,36	1,1	0,49
132	1,48	0,29	1,53	1,6	1,51	1,5	0,63
133	-1,28	0,28	0,85	-0,4	0,80	-0,6	0,57
134	0,91	0,28	1,08	0,4	1,08	0,3	0,31
135	0,84	0,28	1,28	0,9	1,25	0,9	0,61
137	2,49	0,32	0,60	-1,4	0,69	-1,0	0,39
147	0,54	0,27	1,31	1,0	1,30	1,0	0,27
148	0,54	0,27	1,16	0,6	1,12	0,5	0,65
149	-0,26	0,27	0,50	-2,0	0,50	-1,9	0,67
150	0,54	0,27	0,88	-0,3	0,86	-0,4	0,21
151	1,73	0,30	0,97	0,0	0,97	0,0	0,48
152	0,17	0,27	1,08	0,4	1,08	0,4	0,36
153	0,61	0,27	1,18	0,6	1,15	0,6	0,36
<b>MOY.</b>	<b>0,24</b>	<b>0,28</b>	<b>0,99</b>	<b>-0,1</b>	<b>0,88</b>	<b>0,0</b>	
<b>É.-T.</b>	<b>1,13</b>	<b>0,02</b>	<b>0,35</b>	<b>1,2</b>	<b>0,34</b>	<b>1,1</b>	

**Tableau 35 : Statistiques d'ajustement finales des personnes pour la version B du questionnaire 2005**

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
502	1,01	0,29	0,50	-1,9	0,51	-1,9	0,64
503	2,89	0,32	1,07	0,3	1,09	0,4	0,49
504	1,01	0,29	1,11	0,5	1,39	1,2	0,25
523	1,01	0,29	0,64	-1,2	0,65	-1,2	0,70
524	3,57	0,36	0,81	-0,5	0,75	-0,5	0,53
531	-3,26	0,35	0,97	0,0	0,87	-0,3	0,82
540	-1,61	0,30	1,24	0,8	1,22	0,8	0,66
542	-0,19	0,29	1,10	0,4	1,10	0,4	0,06
543	-1,52	0,30	0,94	-0,1	0,90	-0,2	0,83
544	-2,69	0,33	1,92	2,5	1,77	2,2	0,62
545	-1,43	0,30	0,76	-0,7	0,74	-0,8	0,35
546	2,13	0,30	0,99	0,1	1,01	0,1	0,64
547	-0,71	0,30	0,76	-0,7	0,76	-0,7	0,59
548	2,89	0,32	0,89	-0,3	0,81	-0,5	0,67
557	-1,07	0,30	0,58	-1,5	0,57	-1,5	0,70
570	-1,62	0,31	0,75	-0,8	0,74	-0,8	0,29
571	-1,07	0,30	0,63	-1,3	0,64	-1,2	0,67
573	-2,69	0,33	0,65	-1,2	0,65	-1,2	0,83
575	-2,79	0,33	1,34	1,1	1,23	0,8	0,73
576	-0,62	0,30	0,62	-1,3	0,61	-1,3	0,89
578	3,99	0,39	0,67	-1,0	0,73	-0,5	0,55
579	1,69	0,29	1,58	1,7	1,57	1,7	0,44
580	1,43	0,29	1,49	1,5	1,45	1,4	0,53
597	1,95	0,30	0,85	-0,4	0,83	-0,5	0,71
598	-0,02	0,29	1,14	0,5	1,15	0,6	0,61
608	-0,54	0,30	0,92	-0,1	0,92	-0,1	0,71
609	-1,70	0,30	1,47	1,5	1,47	1,5	0,36
611	0,41	0,29	0,88	-0,3	0,88	-0,3	0,71
612	-0,98	0,30	0,91	-0,2	0,90	-0,2	0,70
613	-1,24	0,30	1,10	0,4	1,10	0,4	0,42
620	1,18	0,29	1,59	1,7	1,57	1,7	0,31
<b>MOY.</b>	<b>-0,02</b>	<b>0,31</b>	<b>0,99</b>	<b>0,0</b>	<b>0,99</b>	<b>0,0</b>	
<b>É.-T.</b>	<b>1,92</b>	<b>0,02</b>	<b>0,34</b>	<b>1,1</b>	<b>0,33</b>	<b>1,0</b>	

**Tableau 36 : Statistiques d'ajustement des items pour la version A du questionnaire 2005**

Item	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1	0,64	0,16	0,60	-2,6	0,60	-2,6	0,69
2	1,03	0,16	1,42	2,2	1,42	2,2	0,48
3	0,86	0,16	1,60	3,0	1,59	2,9	0,56
4	-0,21	0,16	1,18	1,0	1,16	0,9	0,64
5	-0,20	0,16	1,06	0,4	1,13	0,8	0,64
6	-0,38	0,16	1,08	0,5	1,10	0,6	0,62
7	0,10	0,16	0,90	-0,5	0,91	-0,4	0,73
8	-2,01	0,18	1,20	1,1	1,17	1,0	0,55
9	-0,32	0,16	0,69	-1,9	0,71	-1,8	0,80
10	-0,71	0,16	0,92	-0,4	0,90	-0,5	0,72
11	-0,38	0,16	0,80	-1,2	0,78	-1,3	0,75
12	-0,95	0,16	1,01	0,1	0,98	0,0	0,67
13	-0,03	0,16	0,53	-3,2	0,53	-3,2	0,74
14	-0,10	0,16	1,39	2,1	1,39	2,1	0,68
15	0,76	0,16	1,23	1,3	1,24	1,4	0,66
16	0,56	0,16	0,66	-2,1	0,66	-2,2	0,77
17	0,17	0,16	0,93	-0,3	1,02	0,2	0,67
18	0,86	0,16	0,88	-0,6	0,87	-0,7	0,69
19	0,83	0,16	1,04	0,3	1,01	0,1	0,70
20	-0,53	0,16	0,72	-1,7	0,71	-1,7	0,81
<b>MOY.</b>	<b>0,00</b>	<b>0,16</b>	<b>0,99</b>	<b>-0,1</b>	<b>0,99</b>	<b>-0,10</b>	
<b>É.-T.</b>	<b>0,73</b>	<b>0,01</b>	<b>0,28</b>	<b>1,6</b>	<b>0,28</b>	<b>1,6</b>	

**Tableau 37 : Statistiques d'ajustement des items pour la version B du questionnaire 2005**

Item	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
1	0,37	0,24	0,82	-0,7	0,84	-0,6	0,82
2	0,85	0,25	1,12	0,6	1,12	0,6	0,74
3	1,27	0,25	1,09	0,4	1,06	0,3	0,79
4	-0,27	0,24	1,08	0,4	1,07	0,3	0,77
5	1,03	0,25	1,16	0,7	1,13	0,6	0,80
<b>6</b>	<b>-0,69</b>	<b>0,24</b>	<b>1,78</b>	<b>2,6</b>	<b>1,75</b>	<b>2,5</b>	<b>0,74</b>
7	-0,22	0,24	1,60	2,1	1,60	2,1	0,69
<b>8</b>	<b>-2,61</b>	<b>0,27</b>	<b>1,88</b>	<b>2,7</b>	<b>1,69</b>	<b>1,7</b>	<b>0,61</b>
9	-1,11	0,25	0,74	-1,0	0,71	-1,1	0,86
10	-1,29	0,25	0,70	-1,2	0,70	-1,2	0,85
11	-0,68	0,25	0,48	-2,5	0,47	-2,5	0,93
12	-1,11	0,25	1,18	0,8	1,28	1,1	0,80
13	-0,27	0,24	0,81	-0,7	0,82	-0,7	0,85
14	-0,04	0,24	1,08	0,4	1,06	0,3	0,91
15	0,85	0,25	1,08	0,4	1,04	0,2	0,85
16	0,67	0,24	0,66	-1,5	0,66	-1,5	0,90
17	0,43	0,24	0,68	-1,4	0,69	-1,4	0,85
18	0,73	0,24	0,95	-0,1	0,91	-0,3	0,90
<b>19</b>	<b>1,40</b>	<b>0,25</b>	<b>0,47</b>	<b>-2,6</b>	<b>0,46</b>	<b>-2,7</b>	<b>0,91</b>
20	0,73	0,24	0,68	-1,4	0,68	-1,4	0,94
<b>MOY.</b>	<b>0,00</b>	<b>0,25</b>	<b>1,00</b>	<b>-0,1</b>	<b>0,99</b>	<b>-0,2</b>	
<b>É.-T.</b>	<b>1,00</b>	<b>0,01</b>	<b>0,38</b>	<b>1,5</b>	<b>0,36</b>	<b>1,4</b>	

**Tableau 38 : Statistiques d'ajustement générales des personnes pour la version A du questionnaire 2005**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	73,5	20,0	0,24	0,28	0,99	-0,1	0,99	0,0
<b>É.-T.</b>	14,7	0,1	1,13	0,02	0,35	1,2	0,34	1,1
<b>MAX.</b>	102,0	20,0	2,60	0,33	1,94	2,5	1,91	2,5
<b>MIN.</b>	35,0	19,0	-2,88	0,27	0,41	-2,4	0,43	-2,2

<b>Réel</b>	<b>Séparation</b>	3,62	« Fidélité »	0,93
<b>Modèle</b>	<b>Séparation</b>	3,90	« Fidélité »	0,94

**Tableau 39 : Statistiques d'ajustement générales des personnes pour la version B du questionnaire 2005**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	69,1	20,0	-0,02	0,31	0,99	0,0	0,99	0,0
<b>É.-T.</b>	21,3	0,2	1,92	0,02	0,34	1,1	0,33	1,0
<b>MAX.</b>	111,0	20,0	3,99	0,39	1,92	2,5	1,77	2,2
<b>MIN.</b>	35,0	19,0	-3,26	0,29	0,50	-1,9	0,51	-1,9

<b>Réel</b>	<b>Séparation</b>	5,75	« Fidélité »	0,97
<b>Modèle</b>	<b>Séparation</b>	6,13	« Fidélité »	0,97



**Tableau 40 : Statistiques d'ajustement générales des items pour la version A du questionnaire 2005**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	224,1	61,0	0,00	0,16	0,99	-0,1	0,99	-0,1
<b>É.-T.</b>	28,5	0,2	0,73	0,01	0,28	1,6	0,28	1,6
<b>MAX.</b>	298,0	61,0	1,03	0,18	1,60	3,0	1,59	2,9
<b>MIN.</b>	183,0	60,0	-2,01	0,16	0,53	-3,2	0,53	-3,2

<b>Réel</b>	<b>Séparation</b>	4,24	« Fidélité »	0,95
<b>Modèle</b>	<b>Séparation</b>	4,48	« Fidélité »	0,95

**Tableau 41 : Statistiques d'ajustement générales des items pour la version B du questionnaire 2005**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	107,0	31,0	0,00	0,25	1,00	-0,1	0,99	-0,2
<b>É.-T.</b>	16,4	0,2	1,00	0,01	0,38	1,5	0,36	1,4
<b>MAX.</b>	149,0	31,0	1,40	0,27	1,88	2,7	1,75	2,5
<b>MIN.</b>	84,0	30,0	-2,61	0,24	0,47	-2,6	0,46	-2,7

<b>Réel</b>	<b>Séparation</b>	3,63	« Fidélité »	0,93
<b>Modèle</b>	<b>Séparation</b>	3,93	« Fidélité »	0,94

Distribution des personnes sur le continuum de l'opinion	Distribution des items sur le continuum du niveau de difficulté
<favorable>	<facile à endosser>
X	
X T	
X	
X	
2	++ Item 8
X	
X	
X	
XX	T
S	
XX	
X	
XXX	
1	XXX ++
X	Item 12
XX	
	S Item 10
XX	
XXX	Item 20
X	Item 11 Item 6
XXX	Item 9
XXXXX M	Item 4 Item 5
X	Item 14
0	XX ++M Item 13
	Item 7
XX	Item 17
XXXXX	
X	
X	Item 1 Item 16
XX	S
XXX	Item 15 Item 19
S	Item 18 Item 3
-1	++ Item 2
XX	
X	
X	
X	T
XX	
X	
-2	T++
X	
-3	++
<défavorable>	<difficile à endosser>

Figure 20 : Distribution des personnes sur le continuum de l'habileté et des items sur le continuum du niveau de difficulté pour la version A du questionnaire 2005

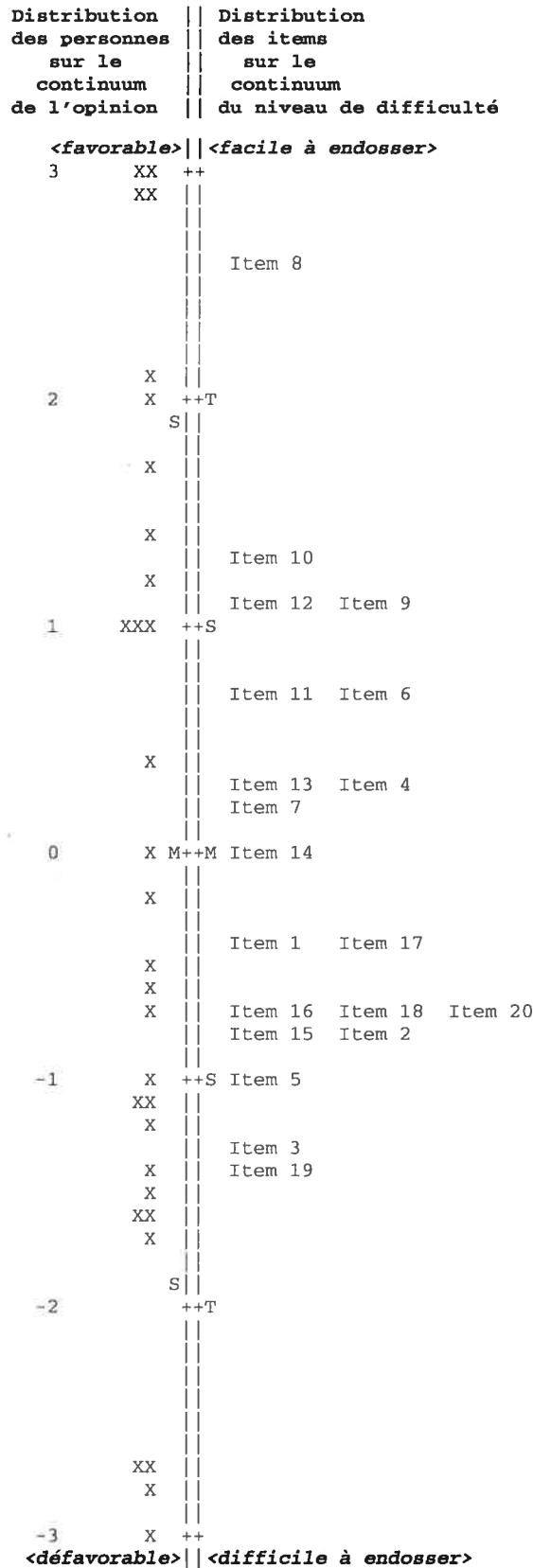


Figure 21 : Distribution des personnes sur le continuum de l'habileté et des items sur le continuum du niveau de difficulté pour la version B du questionnaire 2005

Tableau 42 : Points d'ancrage de la version A du questionnaire 2005

Catégorie	Nb. d'obs.	% d'obs.	Mesure moyenne observée	Mesure moyenne attendue	Infit MNSQ	Outfit MNSQ	Points d'ancrage	Erreur type
1	38	3	-2,03	-1,99	0,92	0,93	NONE	
2	170	14	-1,10	-1,13	1,01	1,00	-3,04	0,18
3	316	26	-0,30	-0,32	1,00	1,00	-1,35	0,09
4	381	31	0,41	0,49	1,11	1,11	-0,11	0,07
5	252	21	1,42	1,33	0,87	0,89	1,32	0,08
6	62	5	2,19	2,25	1,04	1,03	3,18	0,15

Tableau 43 : Points d'ancrage de la version B du questionnaire 2005

Catégorie	Nb. d'obs.	% d'obs.	Mesure moyenne observée	Mesure moyenne attendue	Infit MNSQ	Outfit MNSQ	Points d'ancrage	Erreur type
1	46	7	-3,17	-2,94	0,79	0,82	NONE	
2	120	19	-1,94	-1,95	0,82	0,81	-3,41	0,18
3	166	27	-0,59	-0,75	1,24	1,26	-1,70	0,13
4	137	22	0,53	0,66	0,98	0,93	0,14	0,13
5	91	15	2,05	2,07	1,02	0,99	1,78	0,14
6	59	10	3,51	3,49	1,04	1,04	3,19	0,18

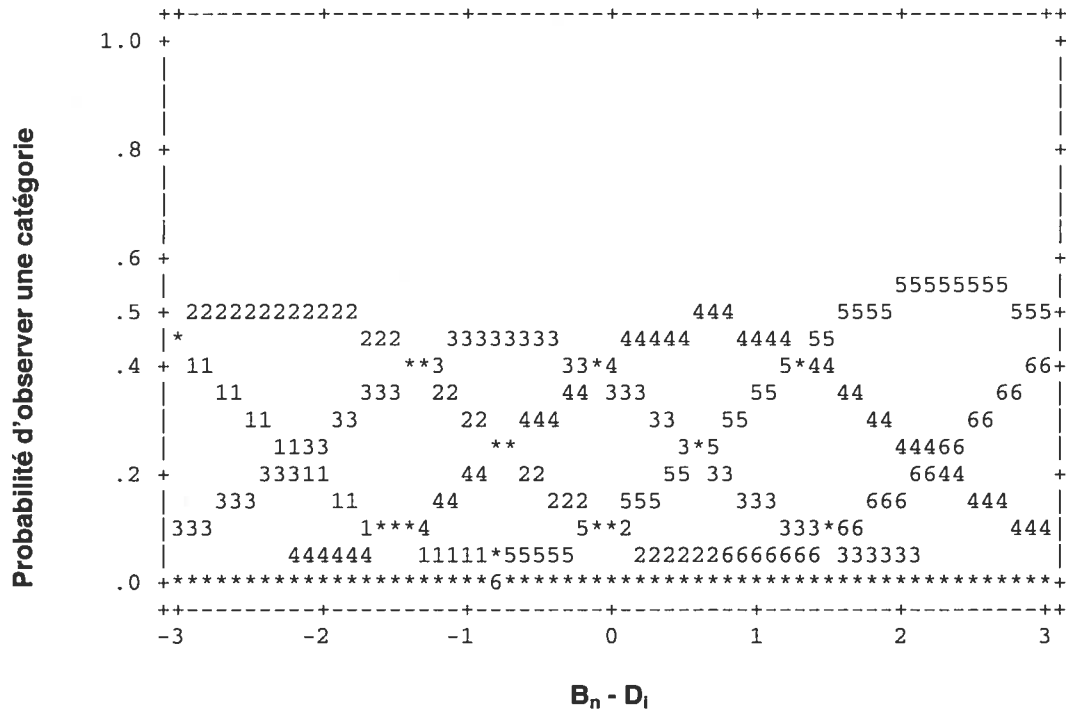


Figure 22 : Courbes de probabilités de chacune des catégories de réponse de la version A du questionnaire 2005

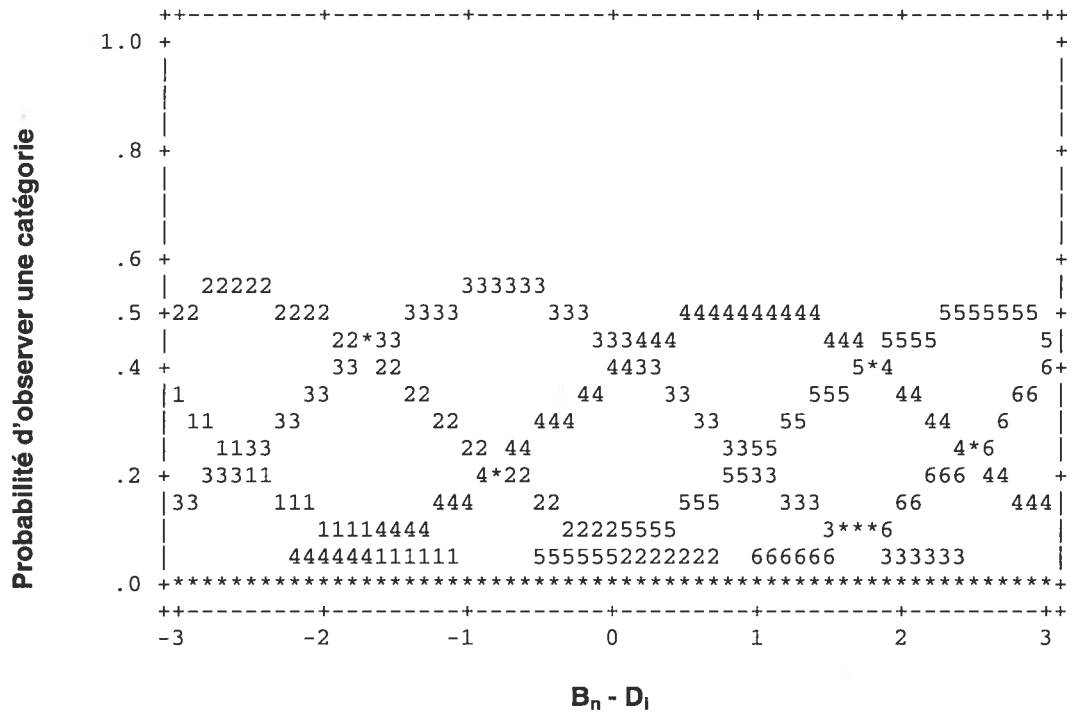


Figure 23 : Courbes de probabilités de chacune des catégories de réponse de la version B du questionnaire 2005

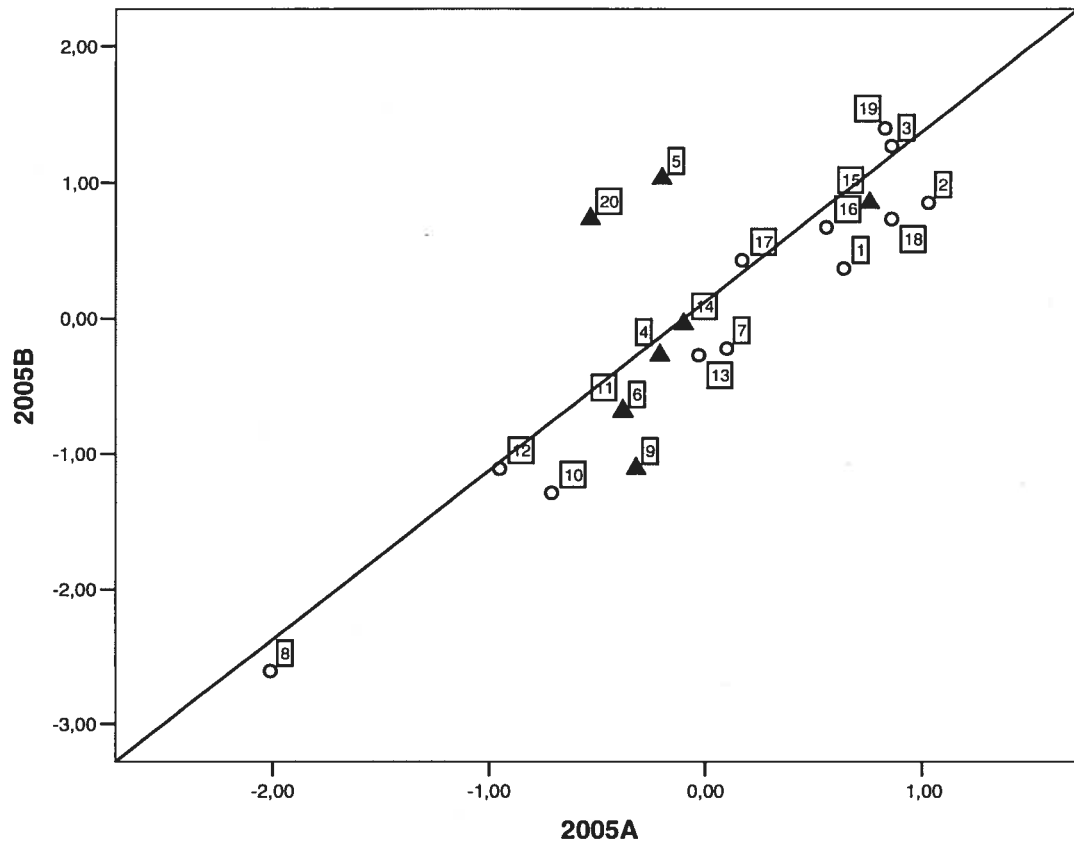


Figure 24 : Diagramme de dispersion des items des versions A et B du questionnaire 2005 par rapport au niveau de difficulté des items

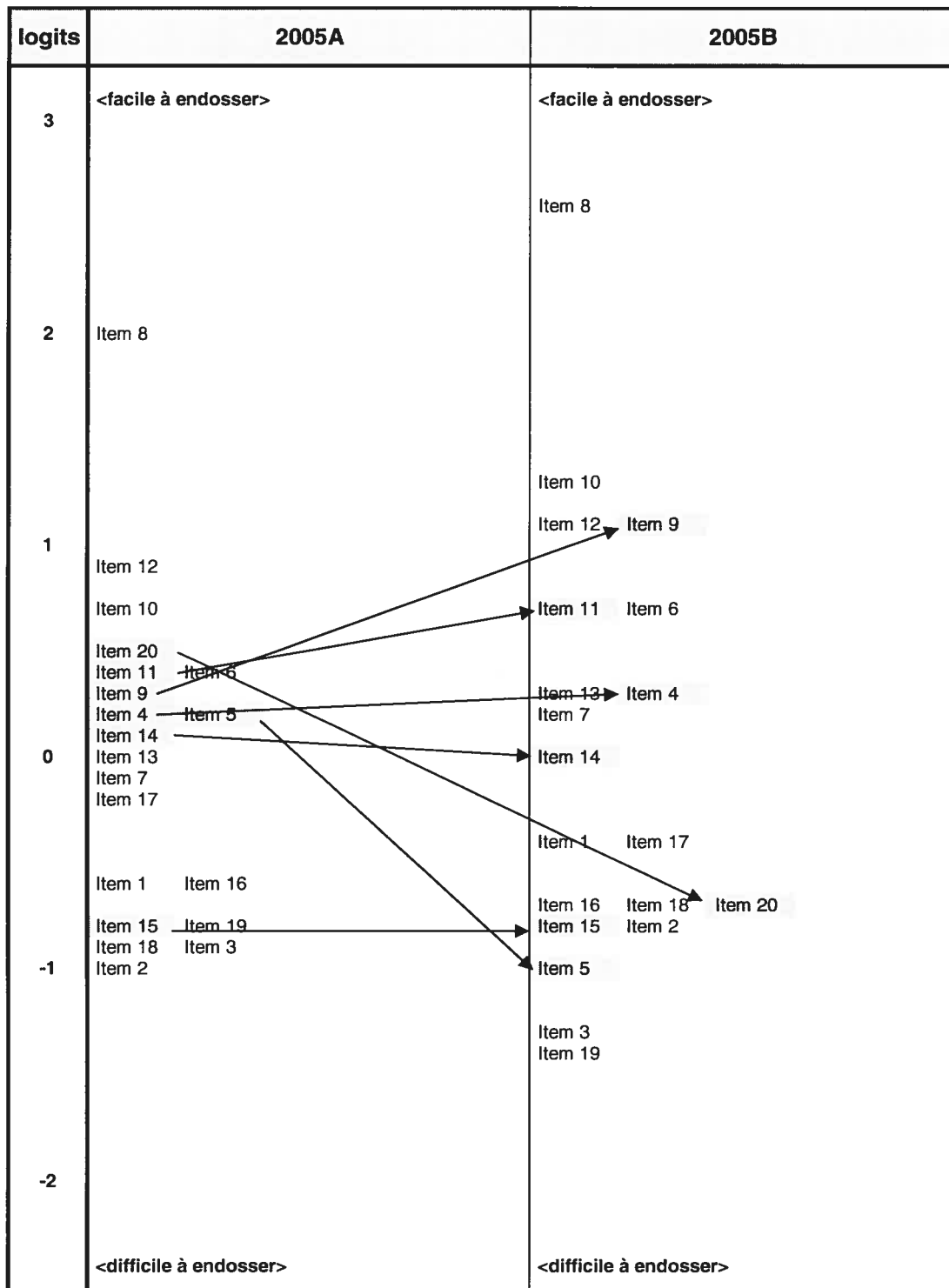


Figure 25 : Diagramme de comparaison montrant la distribution des versions A et B du questionnaire 2005 sur le continuum du niveau de difficulté



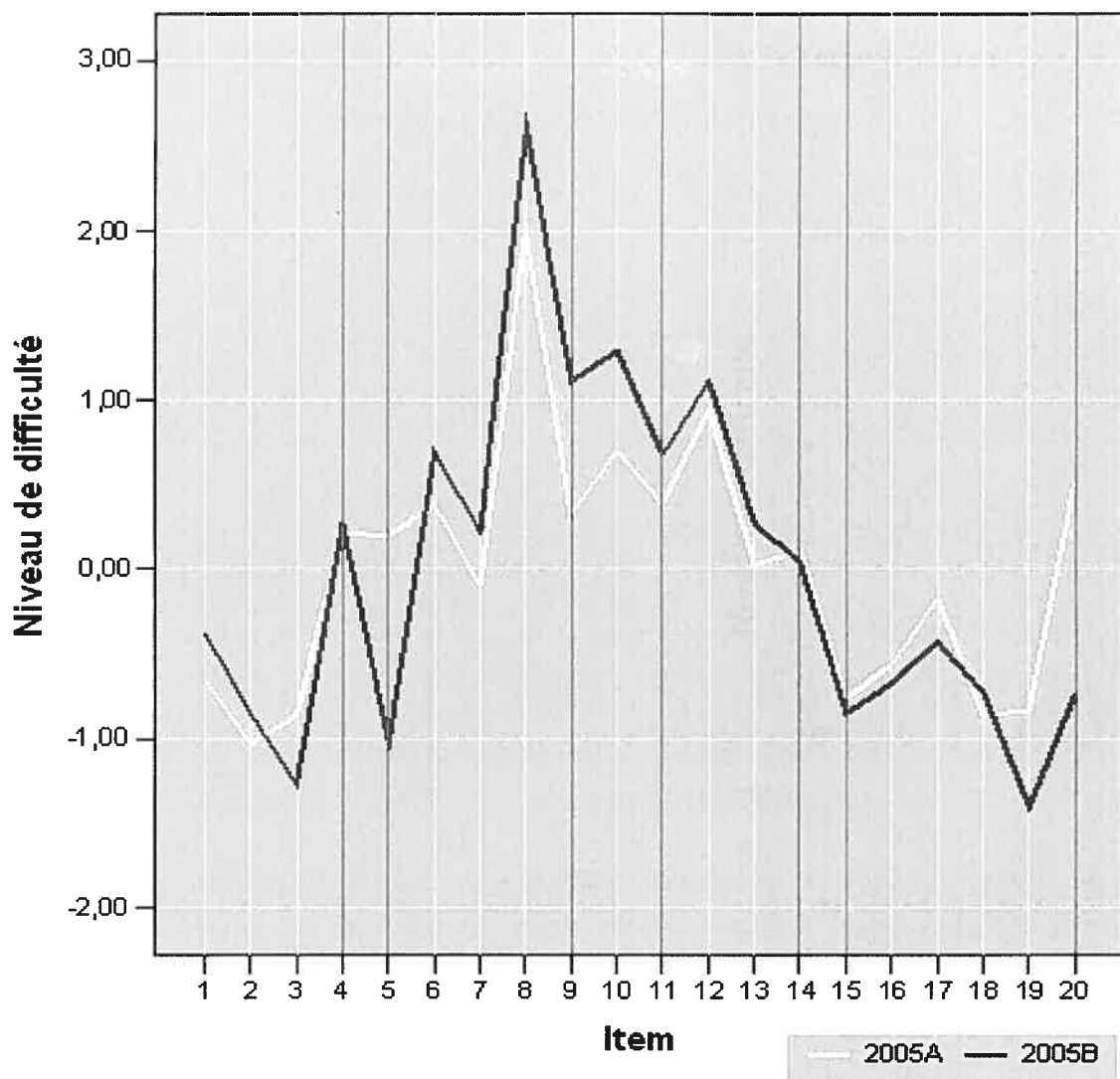


Figure 26 : Courbes de distribution des mesures estimées par le modèle pour le niveau de difficulté des items des versions A et B du questionnaire 2005

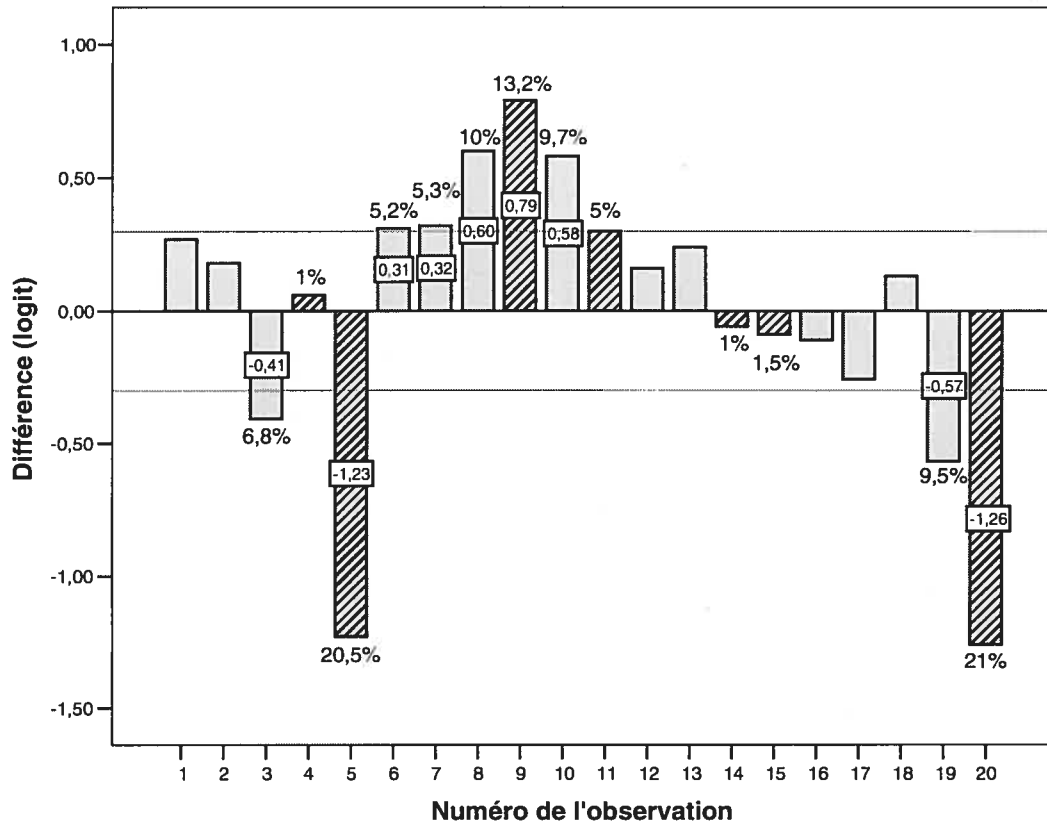
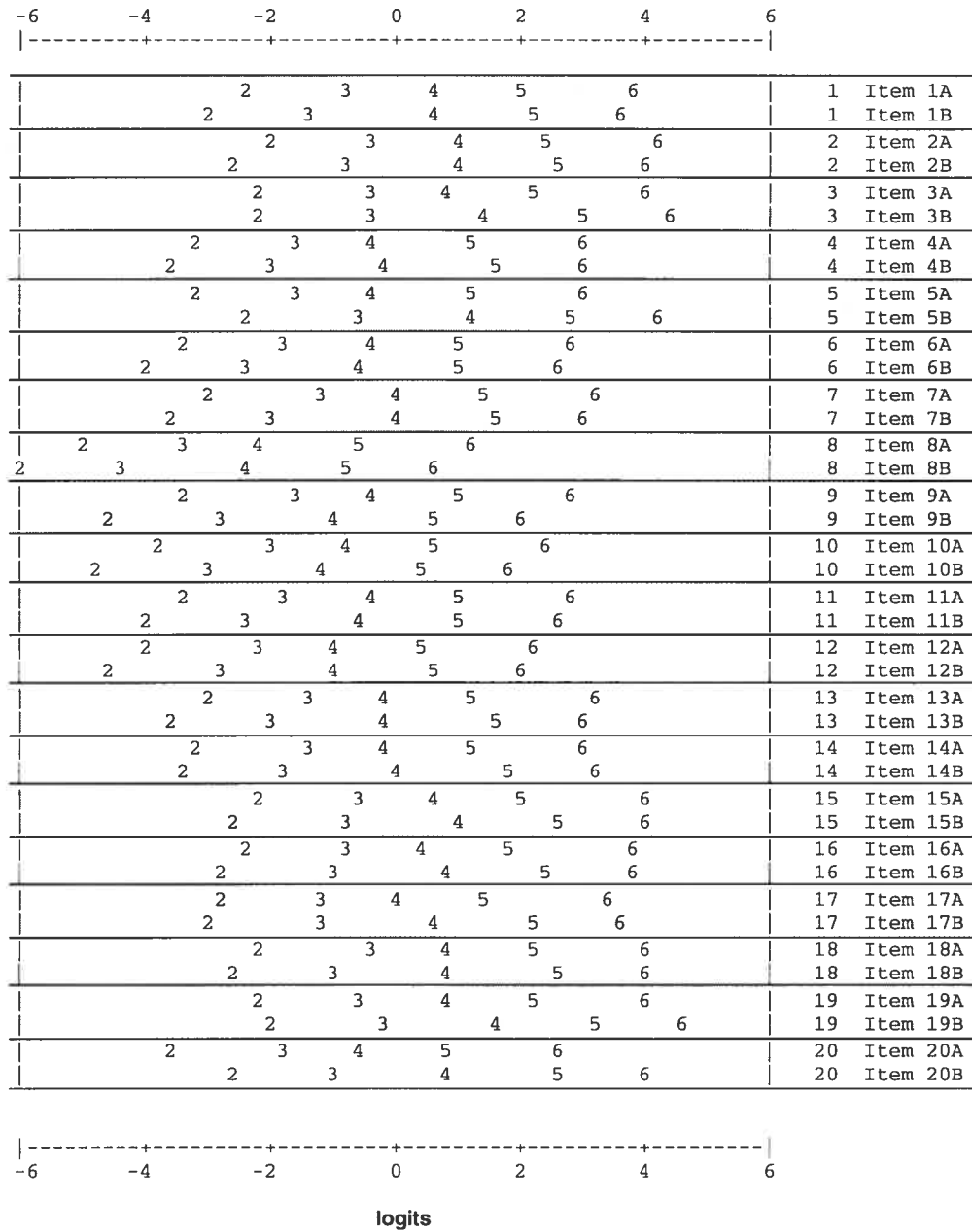


Figure 27 : Diagramme en bâton représentant la valeur des différences entre les mesures estimées par le modèle pour les versions A et B du questionnaire 2005



**Figure 28 : Diagramme de comparaison montrant la position des points d'ancrage estimés par le modèle pour chacun des items des versions A et B du questionnaire 2005**

## **Annexe D : Tableaux des analyses complémentaires**

### **D.1. Version A du questionnaire 2004**

**Tableau 44 : Statistiques d'ajustement finales des personnes pour la version A du questionnaire 2004**

<b>Sujet</b>	<b>Mesure estimée</b>	<b>Erreur type</b>	<b>Infit MNSQ</b>	<b>Infit ZSTD</b>	<b>Outfit MNSQ</b>	<b>Outfit ZSTD</b>	<b>Corr.</b>
1	-0,66	0,26	1,33	1,1	1,29	1,0	0,80
2	-1,00	0,27	1,80	2,2	1,70	2,0	0,60
3	1,15	0,29	1,00	0,1	1,00	0,1	0,69
4	-1,00	0,27	0,55	-1,6	0,51	-1,8	0,72
5	-1,30	0,27	1,12	0,5	1,14	0,5	0,40
7	0,69	0,27	0,56	-1,6	0,51	-1,8	0,67
8	0,69	0,27	1,07	0,3	1,03	0,2	0,45
10	-0,46	0,26	1,62	1,8	1,58	1,7	0,67
11	-0,66	0,26	1,63	1,8	1,58	1,7	0,78
12	-1,22	0,27	1,02	0,2	0,97	0,0	0,76
33	-1,00	0,27	1,39	1,2	1,38	1,2	0,81
34	-1,08	0,27	0,85	-0,4	0,81	-0,5	0,59
35	-0,93	0,27	0,78	-0,6	0,77	-0,7	0,76
61	-1,53	0,28	0,46	-2,0	0,48	-1,9	0,70
62	-1,53	0,28	1,10	0,4	0,98	0,0	0,68
63	0,14	0,26	0,48	-2,1	0,57	-1,6	0,48
64	-0,11	0,26	1,21	0,8	1,26	0,9	0,82
65	-0,33	0,26	1,41	1,3	1,46	1,4	0,80
66	-0,46	0,26	1,08	0,4	1,04	0,2	0,76
67	-2,22	0,31	0,58	-1,4	0,62	-1,2	0,51
68	-1,45	0,28	0,59	-1,4	0,53	-1,7	0,66
69	1,89	0,32	1,19	0,6	1,16	0,6	0,64
70	-0,66	0,26	0,66	-1,1	0,85	-0,4	0,48
71	-0,46	0,26	0,46	-2,1	0,54	-1,7	0,55
74	-0,86	0,26	0,94	-0,1	1,00	0,1	0,79
75	0,92	0,28	1,01	0,2	0,90	-0,2	0,50
76	-0,26	0,26	1,02	0,2	1,06	0,3	0,72
77	-1,45	0,28	0,83	-0,4	0,82	-0,5	0,65
78	-1,15	0,27	1,35	1,1	1,33	1,1	0,77

Suite du tableau 44

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
79	0,92	0,28	0,54	-1,6	0,62	-1,3	0,42
80	-3,13	0,37	1,86	2,0	1,58	1,4	0,55
81	-1,53	0,28	0,38	-2,4	0,39	-2,4	0,64
82	-1,37	0,28	1,27	0,9	1,30	1,0	0,79
83	-1,30	0,27	0,86	-0,4	0,92	-0,1	0,15
124	-2,13	0,31	0,80	-0,5	0,69	-1,0	0,73
125	-0,19	0,26	0,77	-0,7	0,78	-0,7	0,56
126	-0,13	0,26	1,09	0,4	1,10	0,4	0,34
127	-1,69	0,29	1,97	2,5	1,80	2,1	0,64
144	-1,15	0,27	0,66	-1,1	0,64	-1,2	0,57
145	0,20	0,26	1,32	1,1	1,32	1,1	0,49
146	-0,33	0,26	1,36	1,2	1,41	1,3	0,58
147	1,89	0,32	0,69	-0,9	0,60	-1,2	0,35
148	-1,37	0,28	0,63	-1,2	0,60	-1,4	0,56
149	0,20	0,26	0,79	-0,6	0,90	-0,2	0,48
150	0,55	0,27	0,48	-2,0	0,46	-2,1	0,72
153	-0,66	0,26	1,20	0,7	1,15	0,6	0,61
161	-1,61	0,28	0,58	-1,4	0,59	-1,4	0,50
162	-1,45	0,28	0,83	-0,4	0,88	-0,3	0,49
163	-3,60	0,42	0,87	-0,2	0,78	-0,4	0,52
164	-3,43	0,40	0,51	-1,4	0,57	-1,1	0,69
165	-0,93	0,27	1,44	1,4	1,68	1,9	0,35
<b>MOY.</b>	<b>-0,75</b>	<b>0,28</b>	<b>0,98</b>	<b>-0,1</b>	<b>0,97</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>1,14</b>	<b>0,03</b>	<b>0,40</b>	<b>1,2</b>	<b>0,37</b>	<b>1,2</b>	

**Tableau 45 : Statistiques d'ajustement finales des items pour la version A du questionnaire 2004**

<b>Item</b>	<b>Mesure estimée</b>	<b>Erreur type</b>	<b>Infit MNSQ</b>	<b>Infit ZSTD</b>	<b>Outfit MNSQ</b>	<b>Outfit ZSTD</b>	<b>Corr.</b>
1	0,35	0,17	0,86	-0,7	0,86	-0,7	0,69
2	1,73	0,20	1,08	0,5	1,07	0,4	0,63
3	0,44	0,17	1,23	1,2	1,25	1,3	0,63
4	-0,44	0,16	1,23	1,2	1,17	0,9	0,64
5	-0,23	0,16	1,08	0,5	1,02	0,2	0,68
6	0,00	0,16	0,79	-1,1	0,77	-1,2	0,75
7	0,30	0,17	0,60	-2,4	0,61	-2,3	0,76
8	-2,38	0,19	0,97	-0,1	1,01	0,1	0,65
9	-0,62	0,16	0,69	-1,8	0,71	-1,6	0,81
10	-0,75	0,16	1,00	0,1	1,02	0,2	0,66
11	-0,29	0,16	1,04	0,3	1,01	0,1	0,66
12	-1,44	0,17	1,55	2,5	1,51	2,3	0,54
14	-0,41	0,16	1,24	1,3	1,29	1,4	0,60
15	0,35	0,17	1,06	0,4	0,98	0,0	0,71
16	0,24	0,17	0,61	-2,3	0,60	-2,4	0,76
17	0,41	0,17	0,99	0,0	0,98	0,0	0,67
18	0,26	0,17	0,97	-0,1	0,99	0,0	0,67
19	0,94	0,18	0,75	-1,3	0,69	-1,7	0,76
20	1,54	0,19	1,02	0,2	0,96	-0,1	0,64
<b>MOY.</b>	<b>0,00</b>	<b>0,17</b>	<b>0,99</b>	<b>-0,1</b>	<b>0,97</b>	<b>-0,2</b>	
<b>É.-T.</b>	<b>0,92</b>	<b>0,01</b>	<b>0,23</b>	<b>1,2</b>	<b>0,23</b>	<b>1,2</b>	

**Tableau 46 : Statistiques d'ajustement générales des personnes pour la version A du questionnaire 2004**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	57,3	19,0	-0,75	0,28	0,98	-0,1	0,97	-0,1
<b>É.-T.</b>	14,8	0,2	1,14	0,03	0,40	1,2	0,37	1,2
<b>MAX.</b>	92,0	19,0	1,89	0,42	1,97	2,5	1,80	2,1
<b>MIN.</b>	27,0	18,0	-3,60	0,26	0,38	-2,4	0,39	-2,4

<b>Réel</b>	<b>Séparation</b>	3,61	« Fidélité »	0,93
<b>Modèle</b>	<b>Séparation</b>	3,90	« Fidélité »	0,94

**Tableau 47 : Statistiques d'ajustement générales des items pour la version A du questionnaire 2004**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	153,7	50,9	0,00	0,17	0,99	-0,1	0,97	-0,2
<b>É.-T.</b>	32,7	0,3	0,92	0,01	0,23	1,2	0,23	1,2
<b>MAX.</b>	238,0	51,0	1,73	0,20	1,55	2,5	1,51	2,3
<b>MIN.</b>	97,0	50,0	-2,38	0,16	0,60	-2,4	0,60	-2,4

<b>Réel</b>	<b>Séparation</b>	5,15	« Fidélité »	0,96
<b>Modèle</b>	<b>Séparation</b>	5,35	« Fidélité »	0,97





Tableau 48 : Points d'ancrage de la version A du questionnaire 2004

Catégorie	Nb. d'obs.	% d'obs.	Mesure moyenne observée	Mesure moyenne attendue	Infit MNSQ	Outfit MNSQ	Points d'ancrage	Erreur type
1	121	12	-2,52	-2,55	1,07	1,04	NONE	
2	257	27	-1,58	-1,54	0,91	0,90	-2,76	0,12
3	252	26	-0,80	-0,79	0,99	0,96	-1,14	0,09
4	179	18	0,07	0,00	0,90	0,93	-0,07	0,09
5	135	14	0,94	0,94	1,01	0,99	0,74	0,11
6	23	2	1,84	2,04	1,23	1,12	3,24	0,24

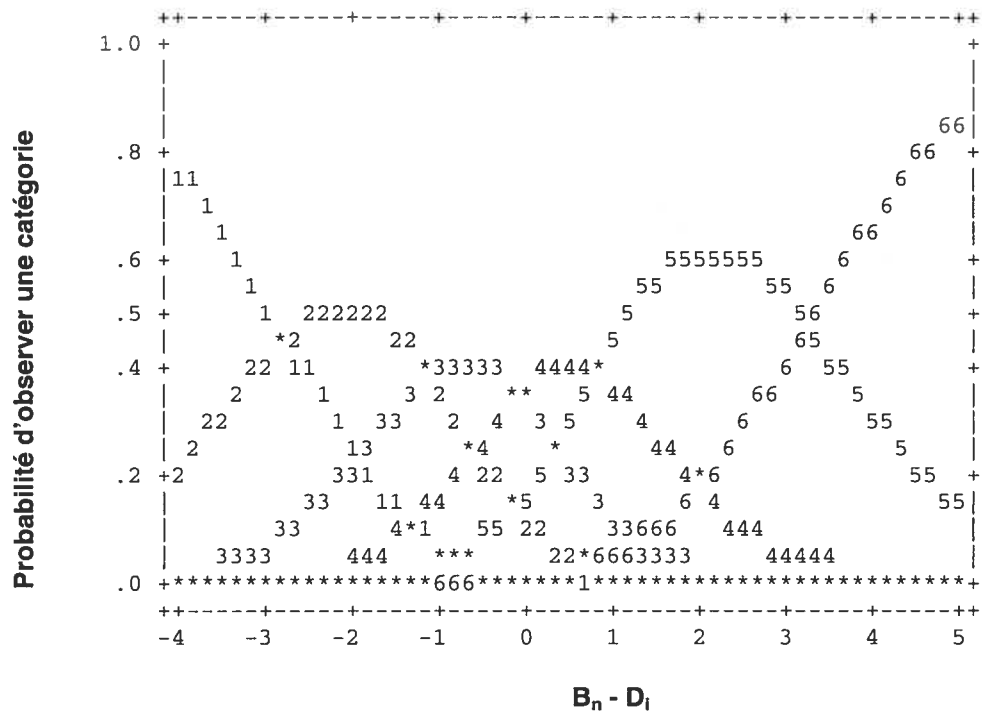


Figure 30 : Courbes de probabilités de chacune des catégories de réponse de la version A du questionnaire 2004

## D.2. Comparaison entre les versions A et B du questionnaire 2004

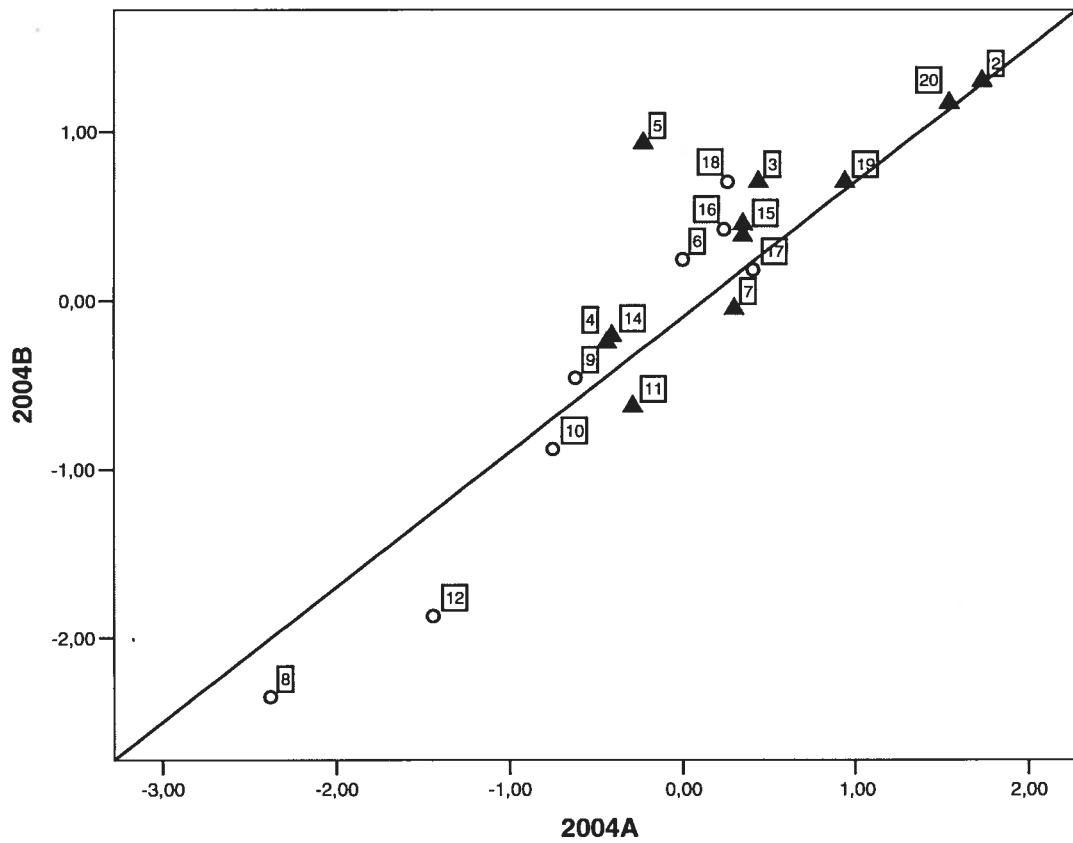


Figure 31 : Diagramme de dispersion des items des versions A et B du questionnaire 2004 par rapport au niveau de difficulté des items

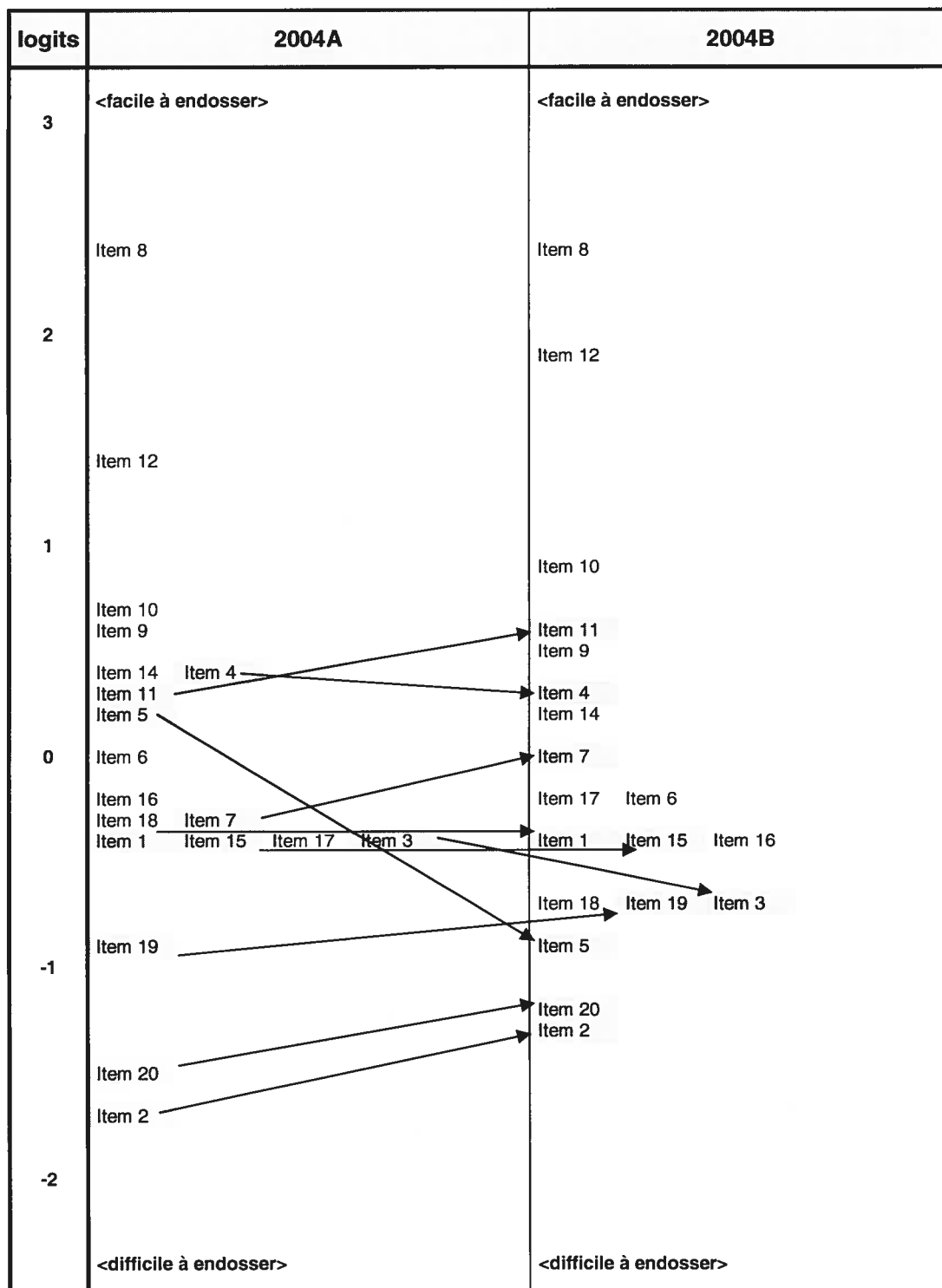


Figure 32 : Diagramme de comparaison montrant la distribution des versions A et B du questionnaire 2004 sur le continuum du niveau de difficulté

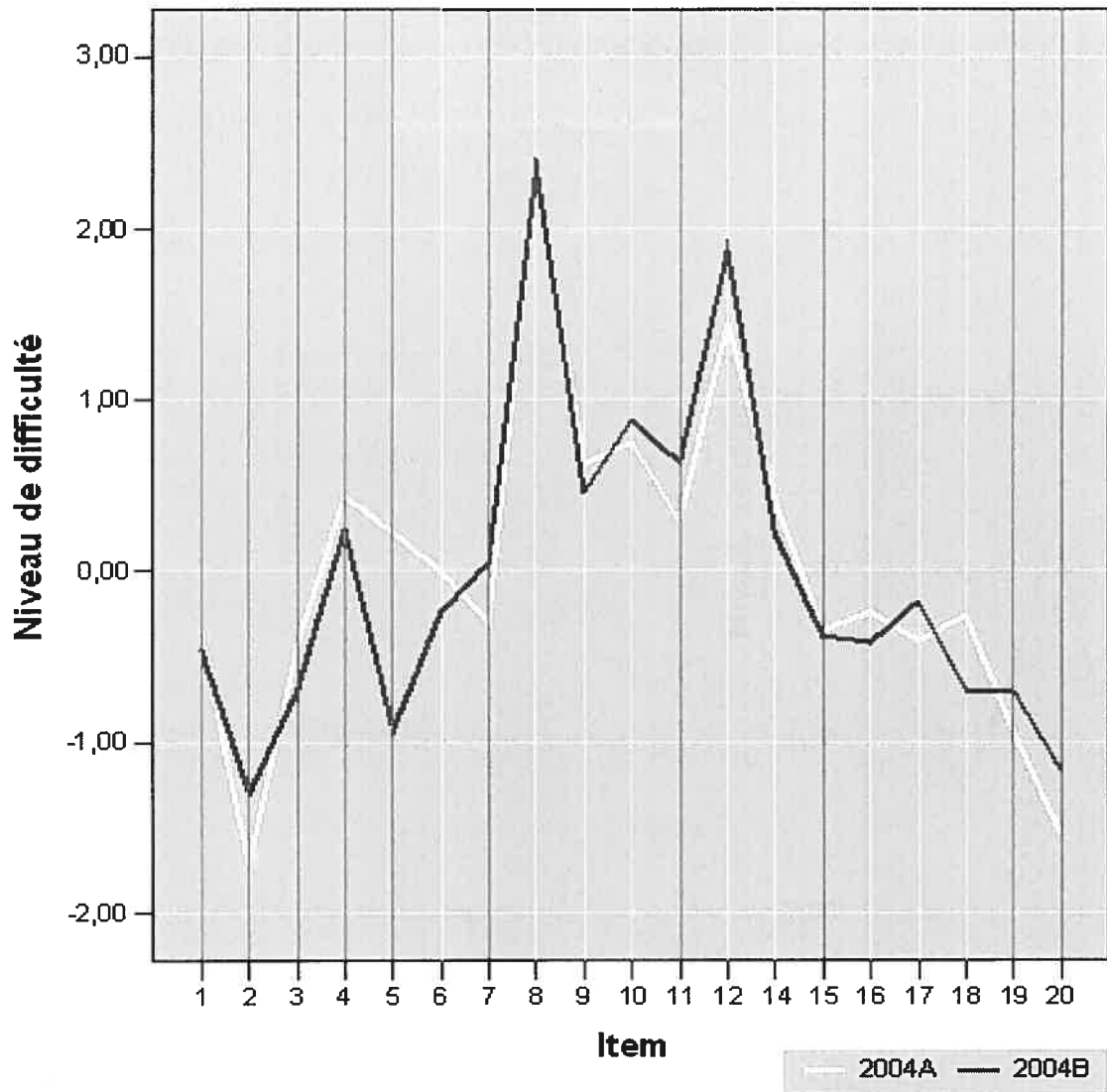


Figure 33 : Courbes de distribution des mesures estimées par le modèle pour le niveau de difficulté des items des versions A et B du questionnaire 2004

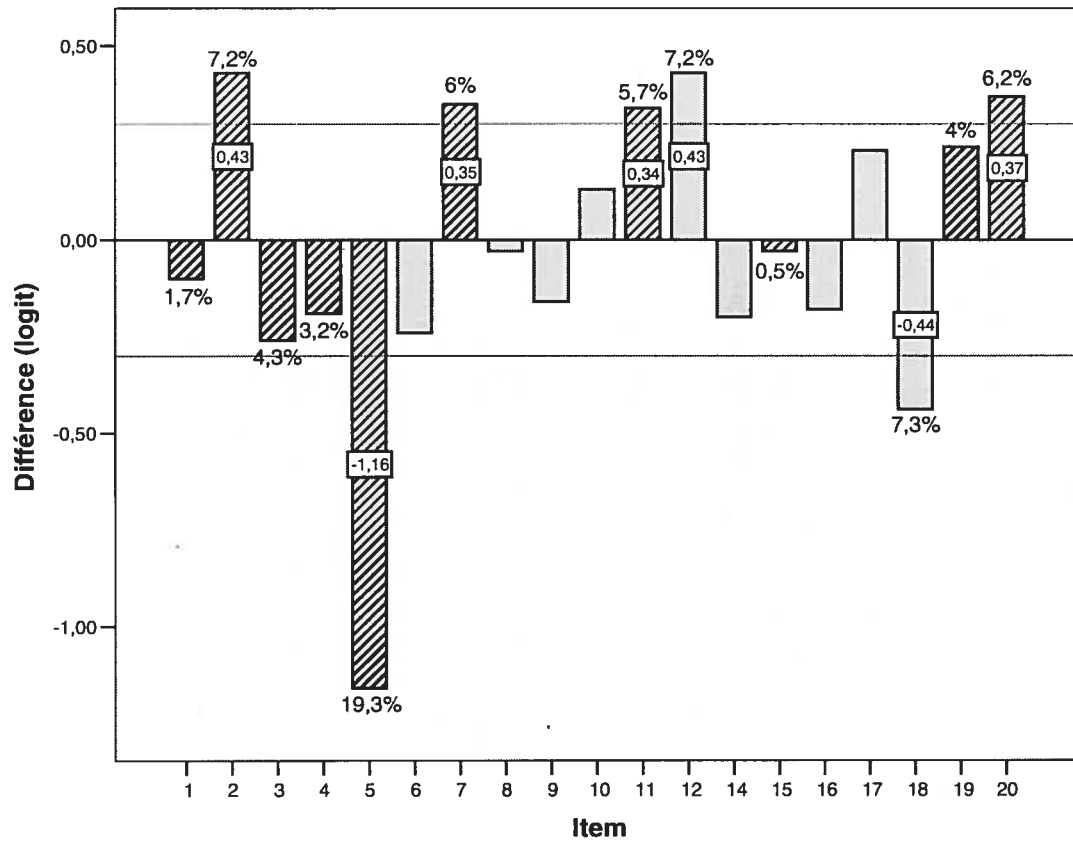
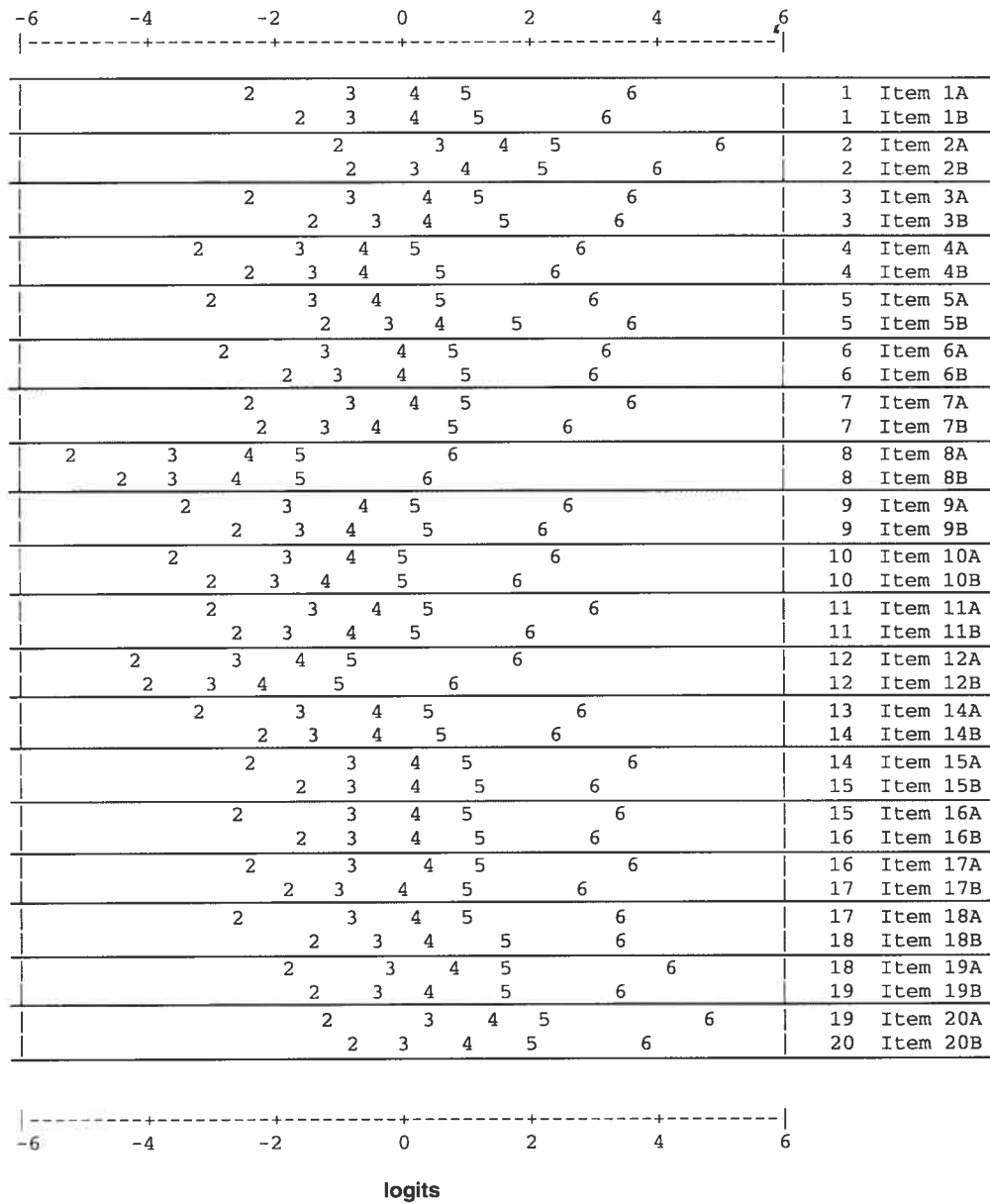


Figure 34 : Diagramme en bâton représentant la valeur des différences entre les mesures estimées par le modèle pour les versions A et B du questionnaire 2004



**Figure 35 : Diagramme de comparaison montrant la position des points d'ancrage estimés par le modèle pour chacun des items des versions A et B du questionnaire 2004**

### D.3. Version A du questionnaire 2005

Tableau 49 : Statistiques d'ajustement finales des personnes pour la version A du questionnaire 2005

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
2	1,41	0,30	1,19	0,7	1,14	0,5	0,48
4	1,22	0,30	1,28	0,9	1,29	0,9	0,25
5	-0,52	0,29	0,48	-1,9	0,48	-1,9	0,73
6	-0,68	0,29	1,58	1,7	1,58	1,7	0,48
7	-0,11	0,28	1,50	1,5	1,51	1,5	0,27
8	0,13	0,29	1,22	0,8	1,24	0,8	0,66
9	-0,19	0,28	0,87	-0,3	0,91	-0,2	0,02
10	-1,71	0,30	1,81	2,2	1,70	1,9	0,57
15	-0,36	0,28	1,26	0,9	1,27	0,9	0,28
17	-0,76	0,29	0,78	-0,6	0,80	-0,6	0,48
18	2,20	0,33	0,53	-1,6	0,54	-1,6	0,75
19	0,46	0,29	0,86	-0,3	0,87	-0,3	0,61
20	1,41	0,30	0,63	-1,2	0,63	-1,2	0,46
40	0,87	0,29	0,47	-1,9	0,45	-2,0	0,56
41	2,77	0,35	1,55	1,5	1,73	1,9	0,16
42	1,41	0,30	0,80	-0,6	0,76	-0,7	0,56
43	1,14	0,30	1,28	0,9	1,24	0,8	0,59
44	0,46	0,29	1,20	0,7	1,22	0,8	0,81
45	-0,76	0,29	0,44	-2,1	0,45	-2,1	0,39
46	0,99	0,30	0,73	-0,8	0,74	-0,7	0,61
47	-1,18	0,29	0,76	-0,7	0,76	-0,7	0,68
48	1,79	0,31	1,57	1,6	1,55	1,5	0,56
49	0,21	0,29	0,48	-1,9	0,48	-1,9	0,56
50	1,89	0,32	0,46	-1,9	0,48	-1,9	0,29
51	-1,62	0,30	1,45	1,3	1,38	1,2	0,55
52	-1,53	0,30	0,85	-0,4	0,80	-0,6	0,29
53	-1,27	0,29	0,75	-0,7	0,79	-0,6	0,70
54	0,13	0,29	0,83	-0,4	0,84	-0,4	0,52
55	-1,71	0,30	0,71	-0,9	0,75	-0,8	0,24

Suite du tableau 49

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
73	0,37	0,29	1,31	1,0	1,31	1,0	0,58
75	0,54	0,29	0,65	-1,1	0,66	-1,1	0,81
76	0,21	0,29	1,26	0,9	1,24	0,8	0,61
77	1,14	0,30	0,51	-1,7	0,52	-1,7	0,65
78	-1,44	0,30	1,16	0,6	1,22	0,8	0,23
79	0,71	0,29	1,15	0,6	1,17	0,6	0,73
80	1,14	0,30	0,41	-2,2	0,51	-1,7	0,33
87	1,14	0,30	1,03	0,2	0,98	0,1	0,65
89	0,29	0,29	1,16	0,6	1,16	0,6	0,34
90	-0,28	0,28	0,95	-0,1	0,94	-0,1	0,44
91	-0,36	0,28	0,65	-1,1	0,65	-1,1	0,64
92	-0,60	0,29	1,06	0,3	1,05	0,3	0,67
93	0,29	0,29	0,96	0,0	0,93	-0,1	0,58
94	-0,19	0,28	1,28	0,9	1,30	1,0	0,55
108	-2,86	0,35	1,86	2,2	1,80	2,1	0,44
109	-0,11	0,28	1,20	0,7	1,19	0,7	0,53
125	1,22	0,30	0,59	-1,4	0,56	-1,5	0,74
126	-0,85	0,29	1,06	0,3	1,06	0,3	0,41
128	0,21	0,29	0,64	-1,2	0,66	-1,1	0,38
130	2,31	0,33	1,64	1,7	1,50	1,4	0,46
132	1,59	0,31	1,70	1,9	1,65	1,8	0,63
133	-1,18	0,29	0,92	-0,1	0,88	-0,3	0,56
134	1,05	0,30	1,09	0,4	1,07	0,3	0,33
135	0,87	0,29	1,40	1,2	1,37	1,1	0,63
137	2,53	0,34	0,66	-1,1	0,74	-0,7	0,40
147	0,62	0,29	1,37	1,1	1,35	1,1	0,30
148	0,62	0,29	1,07	0,3	1,04	0,2	0,64
149	-0,28	0,28	0,52	-1,7	0,53	-1,7	0,69
150	0,46	0,29	0,74	-0,8	0,72	-0,8	0,34
151	1,89	0,32	0,74	-0,8	0,75	-0,7	0,58
152	0,29	0,29	1,12	0,5	1,12	0,5	0,28
153	0,54	0,29	0,64	-1,2	0,65	-1,1	0,61
<b>MOY.</b>	<b>0,29</b>	<b>0,30</b>	<b>1,00</b>	<b>-0,1</b>	<b>1,00</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>1,16</b>	<b>0,02</b>	<b>0,38</b>	<b>1,2</b>	<b>0,37</b>	<b>1,2</b>	



**Tableau 50 : Statistiques d'ajustement finales des items pour la version A du questionnaire 2005**

<b>Item</b>	<b>Mesure estimée</b>	<b>Erreur type</b>	<b>Infit MNSQ</b>	<b>Infit ZSTD</b>	<b>Outfit MNSQ</b>	<b>Outfit ZSTD</b>	<b>Corr.</b>
1	0,69	0,16	0,62	-2,5	0,61	-2,5	0,69
2	1,09	0,16	1,42	2,2	1,43	2,2	0,49
4	-0,16	0,16	1,21	1,2	1,18	1,0	0,63
5	-0,16	0,16	1,06	0,4	1,14	0,8	0,64
6	-0,33	0,16	1,09	0,5	1,11	0,7	0,62
7	0,14	0,16	0,91	-0,5	0,93	-0,4	0,73
8	-1,98	0,18	1,22	1,2	1,20	1,1	0,55
9	-0,28	0,16	0,71	-1,8	0,72	-1,7	0,80
10	-0,67	0,16	0,90	-0,5	0,88	-0,6	0,73
11	-0,33	0,16	0,80	-1,2	0,78	-1,3	0,75
12	-0,91	0,17	0,99	0,0	0,96	-0,2	0,68
14	-0,06	0,16	1,38	2,0	1,38	2,0	0,69
15	0,82	0,16	1,33	1,8	1,34	1,8	0,63
16	0,62	0,16	0,68	-2,0	0,67	-2,1	0,77
17	0,22	0,16	0,91	-0,5	1,00	0,1	0,69
18	0,92	0,16	0,87	-0,7	0,86	-0,8	0,70
19	0,89	0,16	1,06	0,4	1,04	0,3	0,70
20	-0,49	0,16	0,69	-1,9	0,69	-1,9	0,82
<b>MOY.</b>	<b>0,00</b>	<b>0,16</b>	<b>0,99</b>	<b>-0,1</b>	<b>1,00</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>0,75</b>	<b>0,01</b>	<b>0,24</b>	<b>1,4</b>	<b>0,24</b>	<b>1,4</b>	

**Tableau 51 : Statistiques d'ajustement générales des personnes pour la version A du questionnaire 2005**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	66,7	18,0	0,29	0,30	1,00	-0,1	1,00	-0,1
<b>É.-T.</b>	13,3	0,1	1,16	0,02	0,38	1,2	0,37	1,2
<b>MAX.</b>	93,0	18,0	2,77	0,35	1,86	2,2	1,80	2,1
<b>MIN.</b>	32,0	17,0	-2,86	0,28	0,41	-2,2	0,45	-2,1

<b>Réel</b>	<b>Séparation</b>	3,45	« Fidélité »	0,92
<b>Modèle</b>	<b>Séparation</b>	3,76	« Fidélité »	0,93

**Tableau 52 : Statistiques d'ajustement générales des items pour la version A du questionnaire 2005**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	225,9	60,9	0,00	0,16	0,99	-0,1	1,00	-0,1
<b>É.-T.</b>	28,9	0,2	0,75	0,01	0,24	1,4	0,24	1,4
<b>MAX.</b>	298,0	61,0	1,09	0,18	1,42	2,2	1,43	2,2
<b>MIN.</b>	183,0	60,0	-1,98	0,16	0,62	-2,5	0,61	-2,5

<b>Réel</b>	<b>Séparation</b>	4,35	« Fidélité »	0,95
<b>Modèle</b>	<b>Séparation</b>	4,58	« Fidélité »	0,95

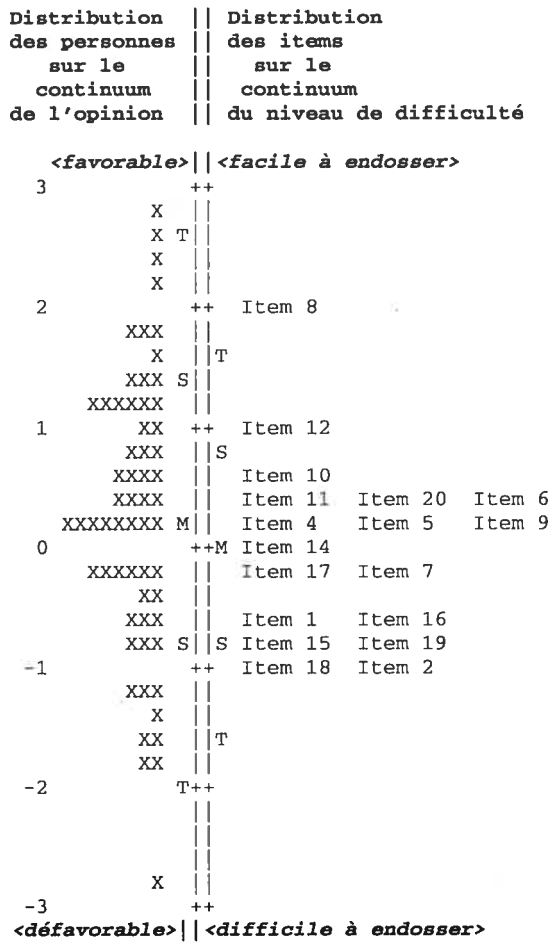


Figure 36 : Distribution des personnes sur le continuum de l'habileté et des items sur le continuum du niveau de difficulté pour la version A du questionnaire 2005

Tableau 53 : Points d'ancrage de la version A du questionnaire 2005

Catégorie	Nb. d'obs.	% d'obs.	Mesure moyenne observée	Mesure moyenne attendue	Infit MNSQ	Outfit MNSQ	Points d'ancrage	Erreur type
1	33	3	-2,11	-2,05	0,89	0,91	NONE	
2	146	13	-1,10	-1,16	1,08	1,07	-3,08	0,20
3	286	26	-0,30	-0,31	0,98	0,98	-1,40	0,10
4	337	31	0,45	0,53	1,12	1,11	-0,05	0,08
5	235	21	1,48	1,39	0,87	0,89	1,32	0,09
6	60	5	2,31	2,33	0,97	0,98	3,22	0,15

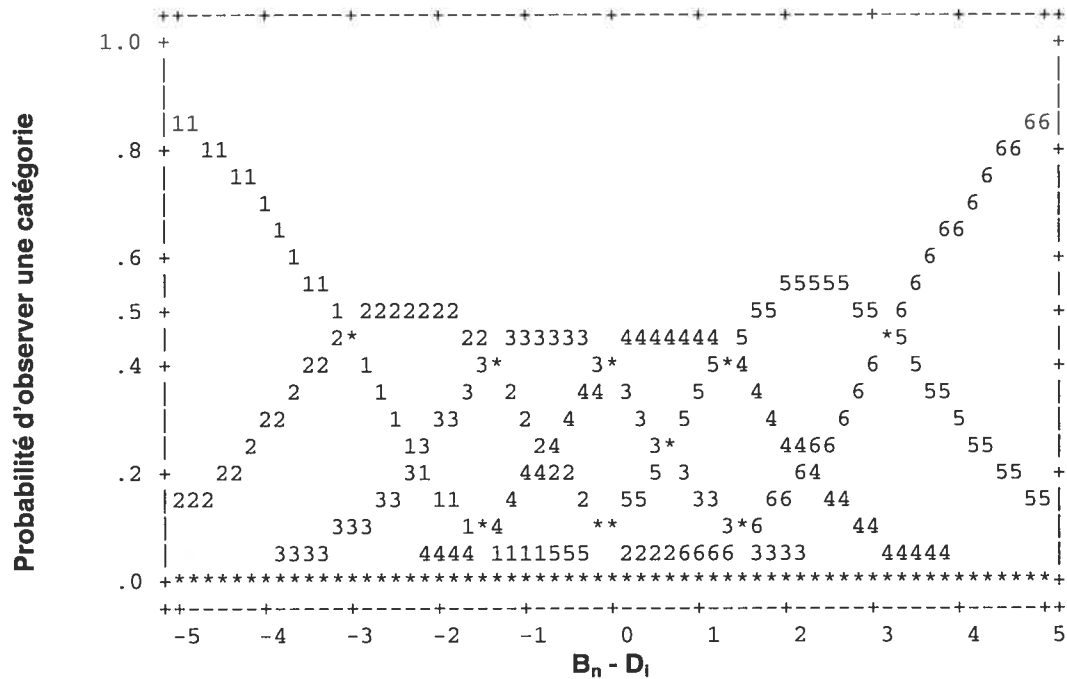


Figure 37 : Courbes de probabilités de chacune des catégories de réponse de la version A du questionnaire 2005

#### D.4. Version B du questionnaire 2005

Tableau 54 : Statistiques d'ajustement finales des personnes pour la version B du questionnaire 2005

Sujet	Mesure estimée	Erreur type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Corr.
502	1,07	0,37	0,66	-1,0	0,67	-1,0	0,56
503	3,91	0,39	0,66	-1,1	0,77	-0,6	0,66
504	1,47	0,36	0,63	-1,1	0,64	-1,1	0,64
523	0,94	0,37	0,88	-0,3	0,89	-0,2	0,67
524	4,38	0,41	0,99	0,1	0,91	-0,1	0,53
531	-4,63	0,44	0,99	0,1	0,86	-0,3	0,75
540	-2,17	0,37	1,05	0,3	1,05	0,3	0,70
542	-0,14	0,37	1,14	0,5	1,13	0,5	0,17
545	-2,04	0,37	0,91	-0,2	0,90	-0,2	0,35
546	2,93	0,37	0,74	-0,8	0,74	-0,7	0,80
547	-1,23	0,37	1,10	0,4	1,10	0,4	0,40
548	3,48	0,37	1,18	0,6	1,08	0,3	0,64
557	-1,90	0,37	0,51	-1,7	0,50	-1,7	0,72
570	-2,17	0,38	0,77	-0,6	0,78	-0,6	0,32
571	-1,90	0,37	0,64	-1,1	0,64	-1,1	0,64
573	-3,76	0,40	0,80	-0,6	0,78	-0,6	0,80
575	-4,08	0,41	1,66	1,8	1,44	1,3	0,70
576	-1,36	0,37	0,52	-1,6	0,52	-1,6	0,86
578	4,72	0,43	0,67	-1,0	0,81	-0,3	0,60
579	2,13	0,36	1,77	2,0	1,83	2,1	0,52
597	2,27	0,36	1,33	1,0	1,30	0,9	0,63
598	-0,28	0,37	0,98	0,1	0,96	0,0	0,71
608	-1,49	0,37	0,59	-1,3	0,59	-1,3	0,70
609	-2,17	0,37	1,57	1,6	1,56	1,5	0,45
611	0,40	0,37	1,45	1,3	1,46	1,3	0,66
612	-1,76	0,37	1,07	0,3	1,07	0,3	0,46
613	-2,17	0,37	1,37	1,1	1,37	1,1	0,20
<b>MOY.</b>	<b>-0,21</b>	<b>0,38</b>	<b>0,99</b>	<b>0,0</b>	<b>0,98</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>2,59</b>	<b>0,02</b>	<b>0,35</b>	<b>1,0</b>	<b>0,33</b>	<b>1,0</b>	

**Tableau 55 : Statistiques d'ajustement des items pour la version B du questionnaire 2005**

<b>Item</b>	<b>Mesure estimée</b>	<b>Erreur type</b>	<b>Infit MNSQ</b>	<b>Infit ZSTD</b>	<b>Outfit MNSQ</b>	<b>Outfit ZSTD</b>	<b>Corr.</b>
1	-0,03	0,29	0,88	-0,4	0,89	-0,3	0,85
2	0,75	0,30	1,45	1,6	1,43	1,5	0,76
3	1,55	0,30	1,25	1,0	1,21	0,8	0,80
4	-0,81	0,30	1,27	1,0	1,24	0,9	0,78
5	1,01	0,30	1,56	1,9	1,55	1,9	0,80
9	-1,90	0,31	0,88	-0,4	0,80	-0,6	0,87
10	-1,90	0,31	0,85	-0,5	0,88	-0,3	0,88
11	-1,17	0,31	0,61	-1,5	0,60	-1,5	0,93
12	-1,44	0,30	0,96	0,0	1,08	0,4	0,89
13	-0,81	0,30	1,05	0,3	1,02	0,2	0,84
14	-0,29	0,29	0,98	0,0	0,97	0,0	0,95
15	1,01	0,30	1,15	0,6	1,12	0,5	0,85
16	0,40	0,29	0,76	-0,9	0,77	-0,9	0,91
17	0,49	0,29	0,78	-0,8	0,80	-0,8	0,87
18	0,92	0,30	0,86	-0,5	0,83	-0,6	0,93
19	1,55	0,30	0,62	-1,5	0,61	-1,6	0,91
20	0,66	0,29	0,81	-0,7	0,79	-0,8	0,93
<b>MOY.</b>	<b>0,00</b>	<b>0,30</b>	<b>0,98</b>	<b>0,0</b>	<b>0,98</b>	<b>-0,1</b>	
<b>É.-T.</b>	<b>1,12</b>	<b>0,00</b>	<b>0,26</b>	<b>1,0</b>	<b>0,26</b>	<b>1,0</b>	

**Tableau 56 : Statistiques d'ajustement générales des personnes pour la version B du questionnaire 2005**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	57,6	17,0	-0,21	0,38	0,99	0,0	0,98	-0,1
<b>É.-T.</b>	19,0	0,2	2,59	0,02	0,35	1,0	0,33	1,0
<b>MAX.</b>	93,0	17,0	4,72	0,44	1,77	2,0	1,83	2,1
<b>MIN.</b>	27,0	16,0	-4,63	0,36	0,51	-1,7	0,50	-1,7

<b>Réel</b>	<b>Séparation</b>	6,35	« Fidélité »	0,98
<b>Modèle</b>	<b>Séparation</b>	6,78	« Fidélité »	0,98

**Tableau 57 : Statistiques d'ajustement générales des items pour la version B du questionnaire 2005**

	Score brut	Nb réponses	Mesure estimée	Écart-type	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
<b>MOY.</b>	91,4	26,9	0,00	0,30	0,98	0,0	0,98	-0,1
<b>É.-T.</b>	12,5	0,2	1,12	0,00	0,26	1,0	0,26	1,0
<b>MAX.</b>	113,0	27,0	1,55	0,31	1,56	1,9	1,55	1,9
<b>MIN.</b>	74,0	26,0	-1,90	0,29	0,61	-1,5	0,60	-1,6

<b>Réel</b>	<b>Séparation</b>	3,42	« Fidélité »	0,92
<b>Modèle</b>	<b>Séparation</b>	3,60	« Fidélité »	0,93

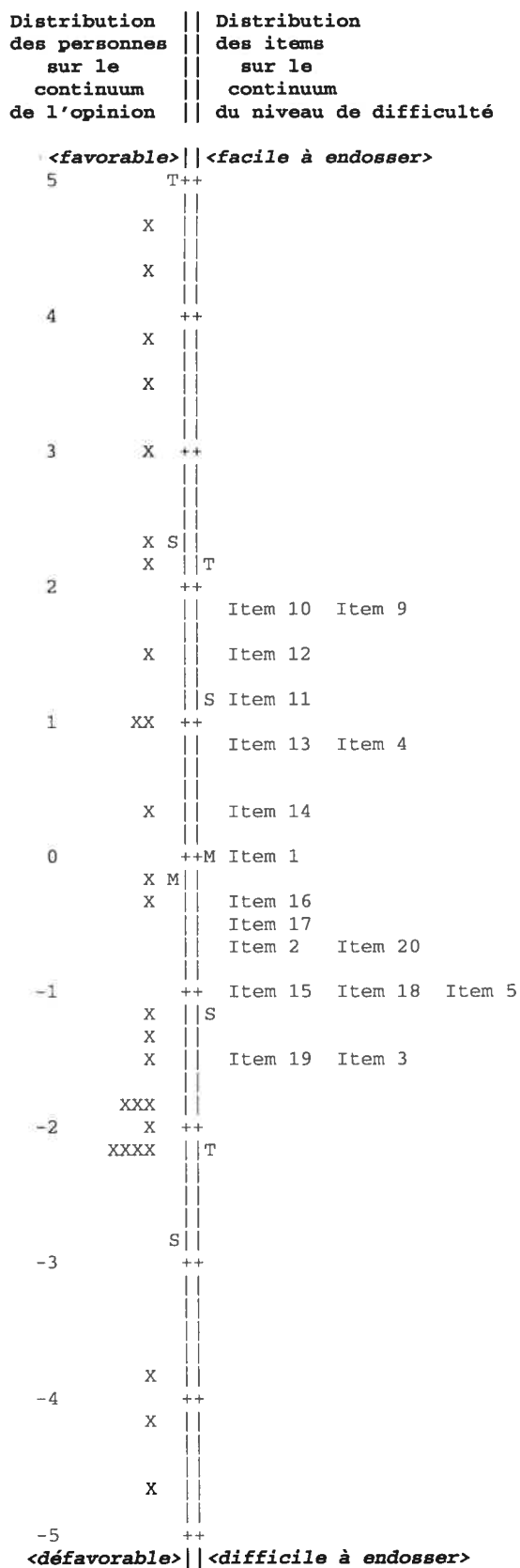


Figure 38 : Distribution des personnes sur le continuum de l'habileté et des items sur le continuum du niveau de difficulté pour la version B du questionnaire 2005



Tableau 58 : Points d'ancrage de la version B du questionnaire 2005

Catégorie	Nb. d'obs.	% d'obs.	Mesure moyenne observée	Mesure moyenne attendue	Infit MNSQ	Outfit MNSQ	Points d'ancrage	Erreur type
1	33	7	-4,44	-4,19	0,79	0,81	NONE	
2	97	21	-2,57	-2,72	1,07	1,02	-4,51	0,23
3	123	27	-1,14	-1,15	1,04	1,02	-2,21	0,15
4	106	23	0,77	0,90	1,11	1,09	-0,01	0,16
5	60	13	3,00	2,95	0,79	0,81	2,53	0,19
6	39	8	4,63	4,57	0,93	0,97	4,20	0,23

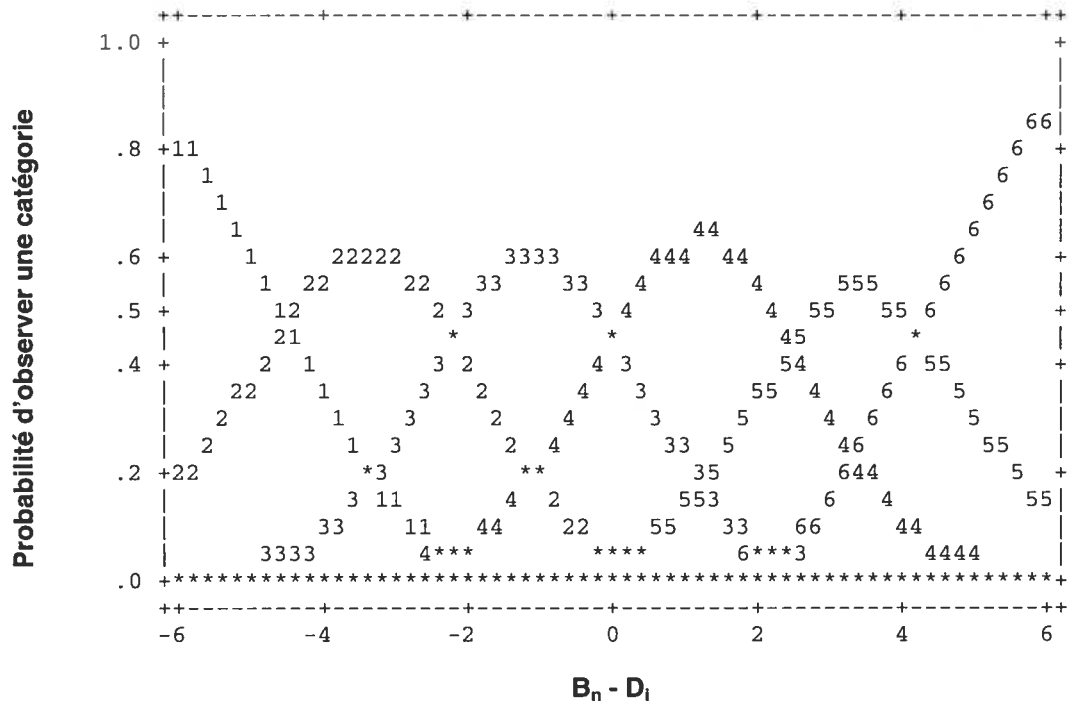


Figure 39 : Courbes de probabilités de chacune des catégories de réponse de la version B du questionnaire 2005

### D.5. Comparaison entre les versions A et B du questionnaire 2005

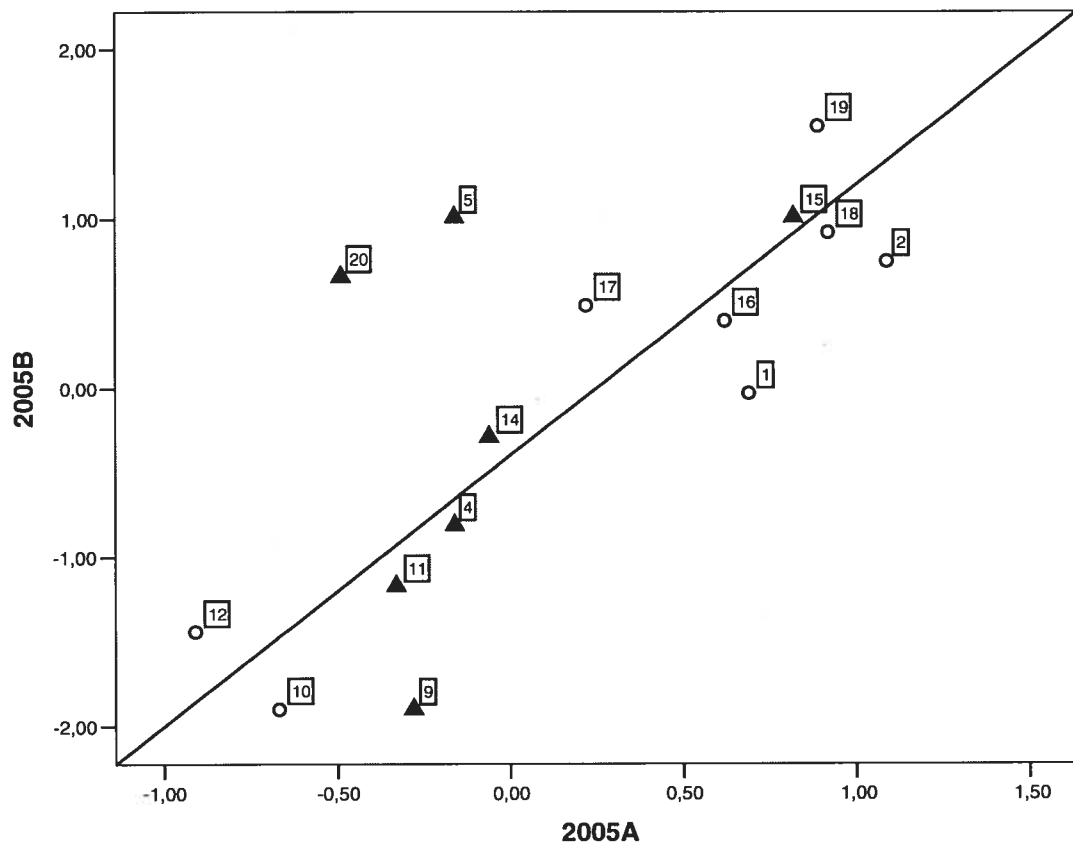


Figure 40 : Diagramme de dispersion des items des versions A et B du questionnaire 2005 par rapport au niveau de difficulté des items

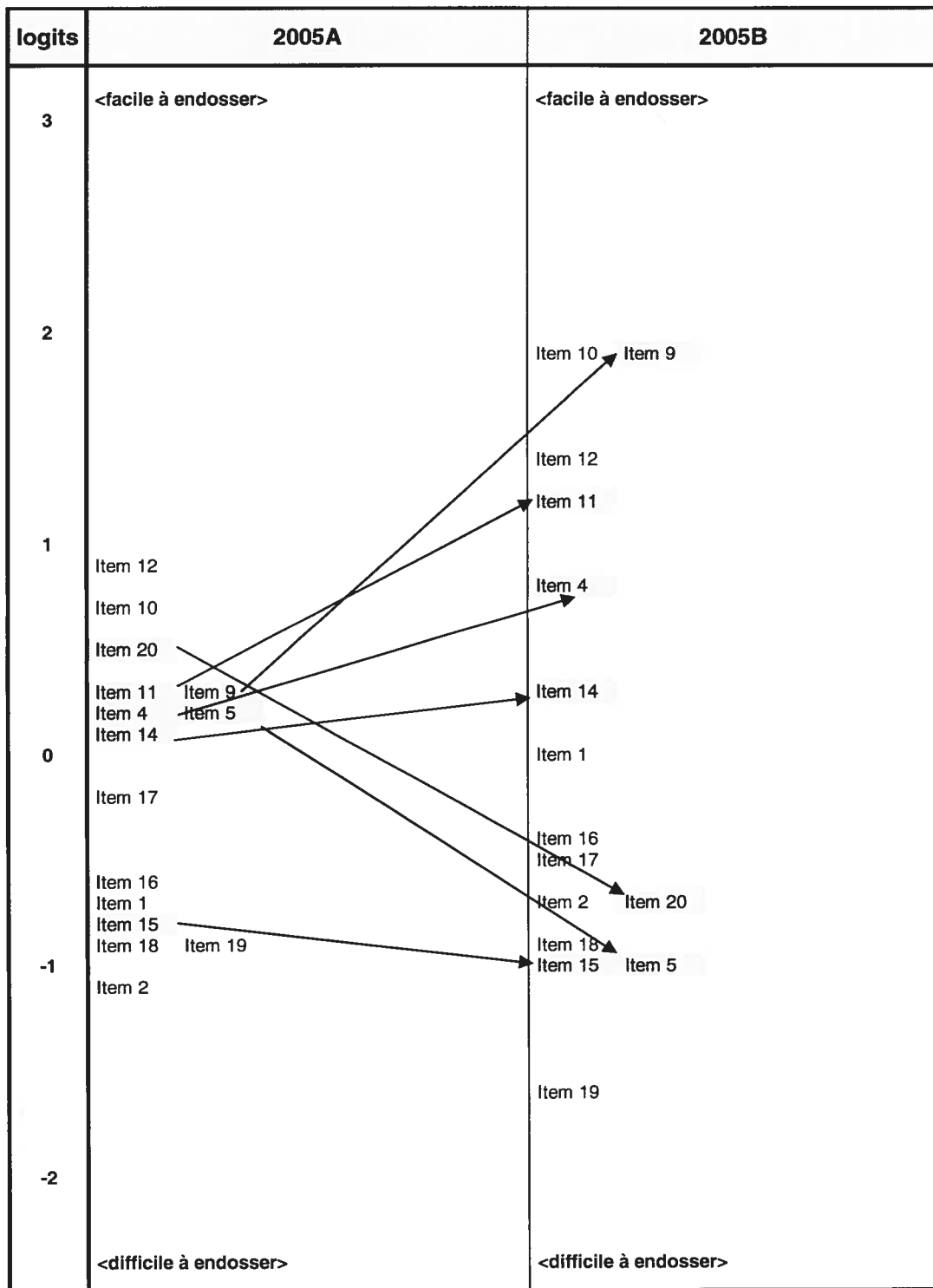


Figure 41 : Diagramme de comparaison montrant la distribution des versions A et B du questionnaire 2005 sur le continuum du niveau de difficulté

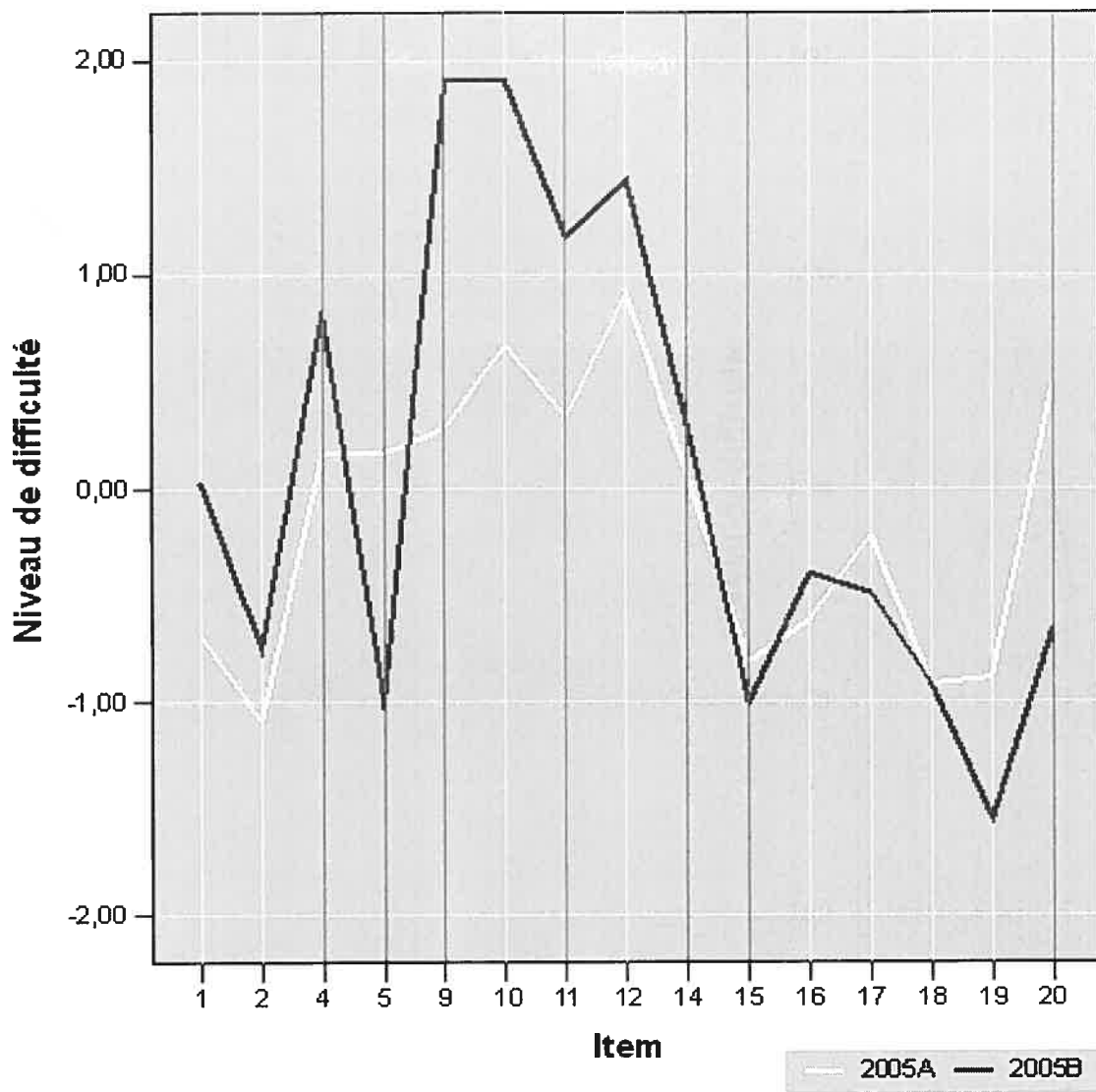


Figure 42 : Courbes de distribution des mesures estimées par le modèle pour le niveau de difficulté des items des versions A et B du questionnaire 2005

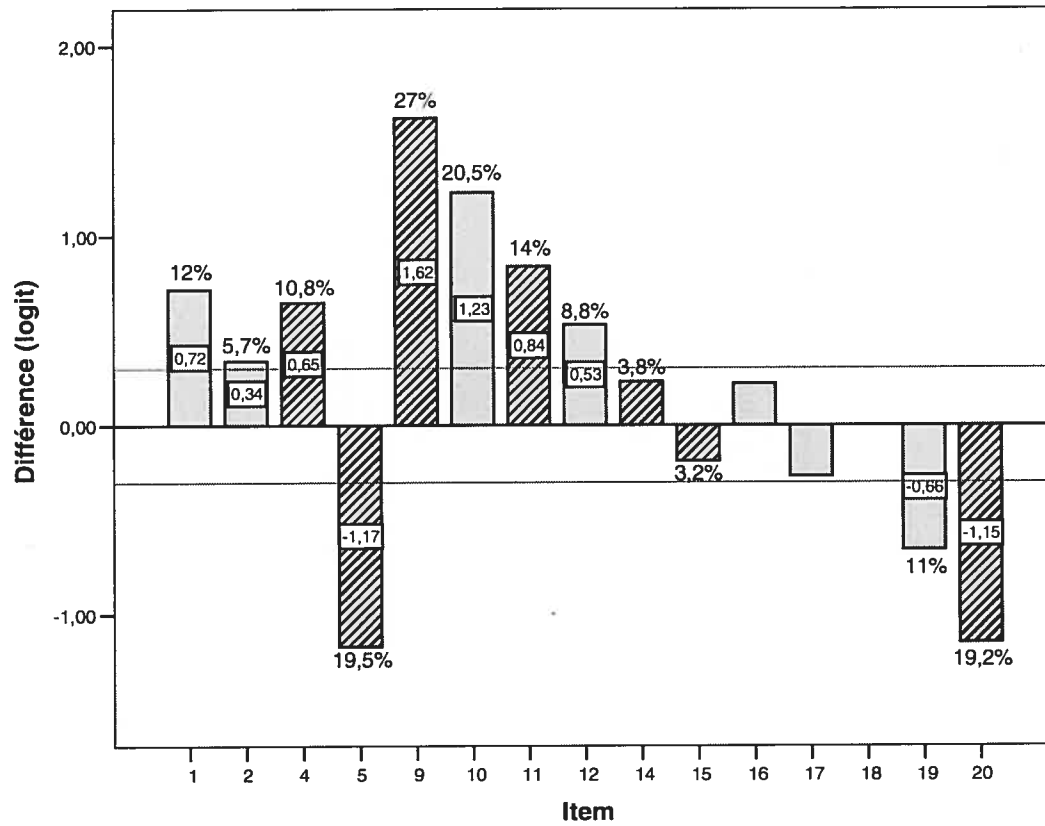
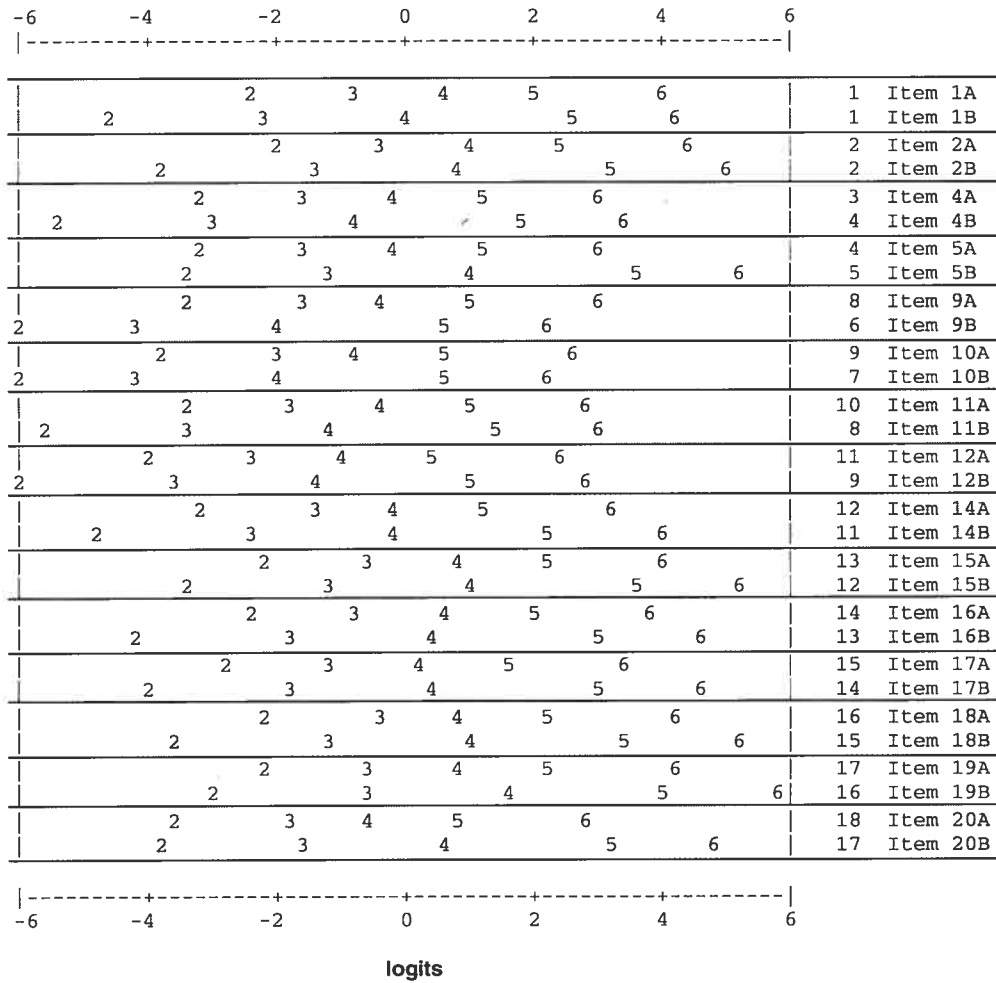


Figure 43 : Diagramme en bâton représentant la valeur des différences entre les mesures estimées par le modèle pour les versions A et B du questionnaire 2005



**Figure 44 : Diagramme de comparaison montrant la position des points d'ancrage estimés par le modèle pour chacun des items des versions A et B du questionnaire 2005**