

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉTUDE COMPARATIVE DE NOUVEAUX INDICES DE DÉTECTION DE
RÉPONSES INAPPROPRIÉES

THÈSE PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN ÉDUCATION

PAR
SÉBASTIEN BÉLAND

JUIN 2015

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.07-2011). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

Pas de science = pas de preuve = pas de vérité = pas de démocratie

REMERCIEMENTS

Je dois remercier trois groupes de personnes dans le cadre de ce projet doctoral. Dans un premier temps, je suis extrêmement redevable à ma mère (c'est un grand privilège d'être ton fils) et à mon frère, qui m'ont épaulé dans cette longue aventure. Je vous aime beaucoup et je suis chanceux de vous avoir dans ma vie.

Dans un deuxième temps, je ne peux passer sous le silence l'apport important de mes amis et de mes collègues: Mike et Tartine, Dave, Rox, Cath et Pat, ma cousine Isabelle, Frank, Didi (je suis tellement chanceux de pouvoir travailler avec toi sur l'évaluation des arts et merci pour le soutien), François Pichette (un grand professeur de linguistique doublé d'un sacré farceur et d'un foutu bon gars), Éric Dionne (un gars qui m'inspire beaucoup), la gang de l'ADEESE (je vous adore tellement), Louis-Marc Bourdeau (tu es un intellectuel comme il ne s'en fait plus, mon cher ami), mes précieux collègues du doctorat (Gen, Amé, Math, Séb, Eugène et les autres), mes anciens collègues de l'Université de Sherbrooke (mention toute spéciale à Christelle-que-j'adore, François le Grand mètre, Isabelle et Marie-Pier : vous me manquez), à mes collègues de l'Université de Montréal (quel privilège d'évoluer aux côtés de Nathalie, Jean-Guy, Micheline-Joanne, Joëlle, Max, Angel et toute la gang en mesure et évaluation), les copains du CDAME et les Macédoniens (vous me manquez, même après toutes ces années).

Dans un troisième temps, il faut mentionner des gens extraordinaires qui ont permis la réalisation de cette thèse. Tout d'abord, je dois sincèrement remercier mon

chum David Magis, un gars qui a des valeurs humaines incroyables et qui se place déjà parmi les meilleurs psychométriciens de sa génération. Je dois aussi remercier mon codirecteur Matiin Riopel, qui a l'âme tranquille des grands esprits. Il ne le sait pas encore, mais j'aimerais bien avoir son envergure intellectuelle un jour (on peut bien rêver, non?!?) Enfin, il y a Gilles Raïche, mon cher directeur de thèse, qui est aussi le grand maître de ce projet doctoral. Gilles, on ne pourra pas dire qu'on a chômé ces dernières années. Nos projets communs m'ont littéralement permis de voir le monde et de développer une passion sans limites pour la mesure. C'est probablement cliché de te l'écrire, mais j'ai contracté une dette importante envers toi. Tu l'auras deviné: mes remerciements et ma gratitude à ton égard suivent une fonction exponentielle.

Enfin, je dois mentionner ma jolie Julie chérie, qui n'est rien de moins que l'être humain le plus accompli et le plus extraordinaire que j'ai rencontré. Et vous savez quoi? C'est avec elle que je partage ma couverture quand le temps froid revient... Avouez que je suis chanceux, non?

TABLE DES MATIÈRES

REMERCIEMENTS	iv
LISTE DES FIGURES.....	x
LISTE DES TABLEAUX.....	xii
RÉSUMÉ	xvi
INTRODUCTION	1
PROBLÉMATIQUE	5
1.1 Problématique des réponses inappropriées en évaluation	6
1.1.1 Étude des patrons de réponse inappropriés en évaluation des apprentissages	9
1.2 Solutions pour détecter les patrons de réponses inappropriées	12
1.2.1 Intervention avant la passation de l'épreuve d'évaluation	12
1.2.2 Intervention après la passation de l'épreuve d'évaluation	14
1.3 Les indices de détection de patrons de réponses inappropriés dans les épreuves d'évaluation.....	15
1.3.1 Une difficulté à interpréter le score des indices paramétriques	17
1.4 Synthèse	19
1.5 La question de recherche.....	20
CADRE THÉORIQUE	21
2.1 La théorie de la réponse à l'item	22
2.1.1.1 Les modélisations à réponse dichotomique.....	24
2.1.1.2 La théorie de la réponse à l'item: les postulats de base	26
2.1.1.3 Le modèle de Rasch	28
2.1.1.4 Le modèle à deux paramètres.....	30
2.1.1.5 Le modèle à trois paramètres	32
2.1.2.1 Estimation des paramètres d'item a , b , et c	34
2.1.2.2 L'estimation du paramètre d'habileté de l'étudiant θ	35
2.1.2.3 Estimateur par maximum de vraisemblance	35
2.1.2.4 Estimateur par maximum <i>a posteriori</i>	36

2.1.2.5 Estimateur par vraisemblance maximale pondérée.....	37
2.2 Présentation des principaux indices qui sont basés sur la théorie de la réponse à l'item	38
2.2.1 Les indices de détection pour réponse dichotomique: un mode d'emploi	38
2.2.1.1 Le choix de l'indice.....	39
2.2.1.2 La sélection d'un point de coupure	39
2.2.1.3 La prise de décision.....	41
2.2.2 Les indices paramétriques les plus populaires	43
2.2.2.1 Les indices de type vraisemblance	43
2.2.2.2 Les indices de type carré-moyen	45
2.2.2.3 Les indices de prudence (<i>caution index</i>).....	49
2.2.2.4 Les indices standardisés retenus dans cette thèse de doctorat	52
2.3 Solutions qui ont été proposées pour tenir compte du fait que le niveau d'habileté réel n'est pas connu	52
2.3.1.1 La stratégie de simulation de Raïche et Blais (2005).....	53
2.3.1.2 Le modèle multidimensionnel de Raïche, Magis, Blais et Brochu (2012)	54
2.3.1.3 La correction de Snijders (2001).....	55
2.3.2 Création des indices ZU^* et ZW^*	58
2.4 Synthèse	59
2.5 Objectifs de ce projet doctoral	60
MÉTHODOLOGIE	62
3.1 Informations générales sur les données simulées.....	62
3.1.1 La longueur du test.....	63
3.1.2 Le nombre de patrons de réponses généré	64
3.1.3 Les modèles de réponse à l'item utilisés.....	65
3.1.4 Méthode de génération des patrons de réponses appropriés	67
3.1.5 Les méthodes d'estimation du niveau d'habileté	67
3.2 Étude 1 : analyse descriptive des scores des indices.....	68
3.3 Étude 2 : analyse des erreurs de type I empiriques et théoriques.....	69

3.4 Étude 3 :analyse de la puissance des indices	71
3.4.1 Méthode de génération de patrons de réponses inappropriés.....	71
RÉSULTATS	75
4.1 Étude 1 : analyse descriptive des scores des indices.....	75
4.1.1.1 Fonction de densité des scores des indices : résultats pour 30 items	76
4.1.1.2 Fonction de densité des scores des indices : résultats pour 80 items	91
4.1.1.3 Synthèse	103
4.1.2.1 Corrélations et statistiques descriptives entre les indices : 30 items.....	104
4.1.2.2 Corrélations et statistiques descriptives entre les indices : 80 items.....	115
4.1.2.2 Synthèse	124
4.2 Étude 2 : analyse des erreurs de type I empiriques et théoriques.....	125
4.2.1 Erreurs de type I :résultats pour 30 items.....	126
4.2.2 Erreurs de type I :résultats pour 80 items.....	136
4.2.3 Synthèse	144
4.3 Étude 3 :analyse de la puissance des indices	145
4.3.1 Puissance des indices: résultats pour 30 items.....	145
4.3.1.1 Réponse au hasard.....	146
4.3.1.2 Inattention	169
4.3.2 Puissance des indices: résultats pour 80 items.....	193
4.3.2.1 Réponse au hasard.....	194
4.3.2.2 Inattention	214
4.3.3 Synthèse	233
DISCUSSION	236
5.1 Étude 1: analyse descriptive des scores des indices.....	236
5.1.1 Comparaison des courbes de densité des scores des indices.....	237
5.1.2 Corrélations et statistiques descriptives	239
5.2 Étude 2 :analyse des erreurs de type I empiriques et théoriques.....	243
5.3 Étude 3 : analyse de la puissance des indices	246

5.4 Limites de cette étude.....	251
CONCLUSION.....	252
6.1 Synthèse des résultats.....	252
6.1.1 Étude 1 : analyse descriptive des scores des indices.....	252
6.1.2 Étude 2 : Analyse des erreurs de type I empiriques et théoriques	254
6.1.3 Étude 3 : Analyse de la puissance des indices	255
6.2 Recommandations générales	257
6.3 Études à conduire dans le futur	257
APPENDICE A- CODER UTILISÉ POUR CALCULER LES INDICES	259
RÉFÉRENCES.....	263

LISTE DES FIGURES

2.1	Réponses manifestes et trait latent.....	25
2.2	Courbes caractéristiques d'item selon le modèle de Rasch.....	29
2.3	Courbes caractéristiques d'item selon le modèle à deux paramètres.....	31
2.4	Courbes caractéristiques d'item selon le modèle à trois paramètres.....	33
2.5	Les seuils critiques $Z_{0,01}=-2,33$, $Z_{0,05}=-1,64$ et $Z_{0,1}=-1,28$	41
4.1	Scores de l'indice I_z selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items.....	78
4.2	Scores de l'indice I_z^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items.....	80
4.3	Scores de l'indice ZU selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items.....	83
4.4	Scores de l'indice ZU^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items.....	86
4.5	Scores de l'indice ZW selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items.....	88
4.6	Scores de l'indice ZW^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items.....	90
4.7	Scores de l'indice I_z selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items.....	92
4.8	Scores de l'indice I_z^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items.....	94
4.9	Scores de l'indice ZU selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items.....	96
4.10	Scores de l'indice ZU^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items.....	98
4.11	Scores de l'indice ZW selon l'estimateur de θ par maximum de	

	vraisemblance pondérée pour 80 items.....	100
4.12	Scores de l'indice ZW^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items.....	102
4.13	Erreurs de type I empiriques et théoriques pour 30 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle de Rasch.....	129
4.14	Erreurs de type I empiriques et théoriques pour 30 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle à deux paramètres.....	132
4.15	Erreurs de type I empiriques et théoriques pour 30 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle à trois paramètres.....	135
4.16	Erreurs de type I empiriques et théoriques pour 80 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle de Rasch.....	138
4.17	Erreurs de type I empiriques et théoriques pour 80 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle à deux paramètres.....	141
4.18	Erreurs de type I empiriques et théoriques pour 80 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle à trois paramètres.....	143

LISTE DES TABLEAUX

1.1	Exemples de patrons de réponses inappropriés.....	11
3.1	Synthèse des fonctions de densités de probabilité permettant de générer les paramètres de personnes et d'items.....	67
3.2	Synthèse des manipulations pour l'étude descriptive des scores des indices de détection (estimation selon la méthode de vraisemblance maximale pondérée).....	68
3.3	Indices de détection et interprétation.....	70
3.4	Synthèse des manipulations pour l'étude portant sur les erreurs de type I empiriques et théoriques (estimation selon la méthode de vraisemblance maximale pondérée).....	71
3.5	Synthèse des fonctions de densité permettant de générer les paramètres de personnes et d'items pour l'étude.....	73
3.6	Synthèse des manipulations pour l'étude portant sur la puissance de détection des indices (estimation selon la méthode de vraisemblance maximale pondérée).....	74
4.1	Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle de Rasch (30 items).....	107
4.2	Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle à deux paramètres (30 items).....	111
4.3	Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle à trois paramètres (30 items).....	114
4.4	Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le	

	modèle de Rasch (80items).....	117
4.5	Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle à deux paramètres (80 items).....	120
4.6	Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle à trois paramètres (80 items).....	123
4.7ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z (30 items).....	148
4.8ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z^* (30 items).....	152
4.9ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU (30 items).....	156
4.10ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU^* (30 items).....	160
4.11ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW (30 items).....	164
4.12ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW^* (30 items).....	168
4.13ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice I_z (30 items).....	172
4.14ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice I_z^* (30 items).....	176
4.15ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU (30 items).....	180
4.16ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU^* (30 items).....	184
4.17ab	Pourcentage de détection de l'inattention et statistiques	

	descriptives pour l'indice ZW (30 items).....	188
4.18ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW^* (30 items).....	192
4.19ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z (80 items).....	196
4.20ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z^* (80 items).....	199
4.21ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU (80 items).....	203
4.22ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU^* (80 items).....	207
4.23ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW (80 items).....	210
4.24ab	Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW^* (80 items).....	213
4.25ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice I_z (80 items).....	216
4.26ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice I_z^* (80 items).....	219
4.27ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU (80 items).....	222
4.28ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU^* (80 items).....	226
4.29ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW (80 items).....	229
4.30ab	Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW^* (80 items).....	232
4.31	Indice présentant le plus haut pourcentage de détection de la	

	réponse au hasard.....	235
4.32	Indice présentant le plus haut pourcentage de détection de l'inattention.....	235
5.1	Certains résultats tirés de l'étude des erreurs de type I de I_z^* (De la Torre et Deng, 2008, p. 167-168)	244

RÉSUMÉ

Lors de l'administration d'instruments visant à mesurer le niveau d'habileté d'un étudiant, il est fréquent d'assister à des tentatives de réponses inappropriées telles que la réponse au hasard et l'inattention. Plusieurs approches ont déjà été développées pour détecter ce type de réponses (Zickar et Drasgow, 1996). Parmi celles-ci, l'utilisation d'indices de détection de patrons de réponses inappropriés (*person-fit index*) est certainement l'approche qui est la plus étudiée et qui semble la plus prometteuse (Karabatsos, 2003; Meijer et Sijtsma, 2001).

Dans le cadre de ce projet doctoral, nous nous concentrerons sur trois indices de détection populaires qui présentent des caractéristiques permettant d'en faciliter l'interprétation : l_z (Drasgow, Levine et Williams, 1985), ZU (Wright et Masters, 1982) et ZW (Wright et Masters, 1982). Toutefois, il s'est avéré qu'ils sont tous fortement affectés par le fait que l'habileté d'un étudiant est estimée plutôt que réelle (Li et Olejnik, 1997; Molenaar et Hoijtink, 1990). Voilà pourquoi Snijders (2001) a proposé une version corrigée de l'indice l_z (nommée l_z^*) et qui prend considération de ce problème important.

L'objectif général de ce projet sera de corriger deux autres indices selon l'approche de Snijders (2001) : ZU^* et ZW^* . De plus, nous investiguerons le comportement des indices corrigés l_z^* , ZU^* et ZW^* et de leur version standardisée l_z , ZU et ZW . Pour ce faire, nous effectuerons trois études différentes : une analyse descriptive des scores des indices, une analyse des erreurs de type I et une analyse de leur puissance de détection.

Les analyses ont démontré que c'est l'indice corrigé l_z^* qui est le plus intéressant à utiliser puisque ses scores suivent approximativement la loi $N(0,1)$ et ils permettent de bien détecter la réponse au hasard et l'inattention dans la majorité des situations de simulation. À l'opposé, les scores de ZU et ZU^* ne suivent pas systématiquement les quantiles de leur loi théorique. Pour cette raison, l'utilisation de ces indices n'est pas recommandée en contexte réel.

Mots clés : mesure, théorie de la réponse aux items, indice de détection de patrons de réponses inappropriés, étude de simulation

INTRODUCTION

Il est connu que certains étudiants répondent de façon inappropriée aux épreuves d'évaluation contenant des bonnes et des mauvaises réponses. Par exemple, le Centre pour l'intégrité académique (*Center for Academic Integrity*) rapportait que plus de 75 % des étudiants américains ont admis avoir déjà triché à un test (cité par Hutton, 2006) alors que Christensen Hughes et McCabe (2006) ont montré que 53 % des étudiants canadiens ont avoué avoir triché durant la dernière année. De son côté, Raïche (2002) a déjà étudié le cas du sous-classement intentionnel dans le test de classement en anglais, langue seconde, au niveau collégial francophone.

Cet état de fait n'est pas étonnant puisqu'il existe un climat de plus en plus compétitif au sein des établissements d'enseignement postsecondaires (Potter, 2011). Pensons, par exemple, à l'anxiété qu'éprouvent certains étudiants à vouloir augmenter leur cote R afin d'intégrer un programme d'études universitaires contingenté. Cela entraînerait certains étudiants à éviter d'aider leurs collègues de classe de peur de voir leur cote R diminuer (Périard, 2002). Un autre exemple concerne les étudiants qui souhaitent avoir une moyenne cumulative élevée afin d'être acceptés au sein d'un programme de doctorat ou avoir des chances d'obtenir une bourse de recherche prestigieuse.

Hunt (2003) croit que l'attrait d'un diplôme ou d'une accréditation, comme condition nécessaire à l'obtention d'un emploi, peut pousser certains étudiants à tricher pour augmenter leur note immédiatement ou à se sous-classer

intentionnellement et stratégiquement pour augmenter leur note plus tard. Malheureusement, les exemples cités ci-haut peuvent hypothéquer plusieurs éléments importants de l'apprentissage des étudiants. Ainsi, les individus qui ont l'habitude de se sous-classer de façon intentionnelle courent un plus grand risque d'être démotivés et de développer une image négative d'eux-mêmes (Drasgow, Levine et McLaughlin, 1987).

Les exemples précédents sont des manifestations de patrons de réponses inappropriés. Dans le cadre de cette thèse de doctorat, nous allons plutôt nous concentrer sur deux autres exemples de patrons de réponses inappropriés : la réponse au hasard et l'inattention. Survient alors le problème suivant: que doit faire l'évaluateur pour prévenir ce type de réponse?

Plusieurs approches ont déjà été développées pour détecter les individus qui répondent de façon inappropriée aux épreuves d'évaluation principalement en contexte scolaire et professionnel (Zickar et Drasgow, 1996). Parmi celles-ci, l'utilisation d'indices de détection de patrons de réponses inappropriés (*person-fit index*) est certainement celle qui est la plus étudiée et qui semble la plus prometteuse (Karabatsos, 2003; Meijer et Sijtsma, 2001).

Dans le cadre de ce projet doctoral, nous nous concentrerons sur trois indices de détection populaires qui présentent des caractéristiques permettant d'en faciliter l'interprétation : I_z (Drasgow, Levine et Williams, 1985), ZU (Wright et Masters, 1982) et ZW (Wright et Masters, 1982). Toutefois, il s'est avéré qu'ils sont tous fortement affectés par le fait que l'habileté d'un étudiant est estimée et non pas réelle

(Li et Olejnik, 1997; Molenaar et Hoijtink, 1990). Voilà pourquoi Snijders (2001) a proposé une version corrigée de l'indice l_z (qui se retrouve dans les écrits scientifiques sous le nom de l'indice l_z^*) et qui prend en considération ce problème important.

Aujourd'hui, il faut adapter cette correction à d'autres indices. Ce projet de recherche ira dans ce sens et proposera deux nouveaux indices corrigés selon l'approche de Snijders (2001) : ZU^* et ZW^* . Pour ce faire, nous investiguerons le comportement des indices corrigés l_z^* , ZU^* et ZW^* ainsi que leur version standardisée l_z , ZU et ZW à partir de trois études : une analyse descriptive des scores des indices, une analyse de leurs erreurs de type I empiriques et théoriques et une analyse de la puissance des indices à détecter la réponse au hasard et l'inattention.

Cette thèse de doctorat comprend six chapitres. Dans le premier chapitre, le problème de l'occurrence des patrons de réponse inappropriés est présenté ainsi qu'une recension des différentes avenues proposées par les écrits scientifiques pour en faire la prévention et la détection. Le deuxième chapitre présente les indices de détection paramétriques les plus connus. Il montrera aussi que la correction de Snijders (2001) permet d'améliorer la détection des patrons de réponse inappropriés. Voilà pourquoi cette approche sera retenue par la suite pour créer deux nouveaux indices: ZU^* et ZW^* . Le troisième chapitre décrit les détails méthodologiques de ce projet, et le quatrième présente les résultats de trois études: l'analyse descriptive du score des indices, l'analyse des erreurs de type I empiriques et théoriques et l'analyse

de la puissance des indices. La discussion fait l'objet du chapitre cinq et est suivie par la conclusion de la thèse.

CHAPITRE I

PROBLÉMATIQUE

Imaginons un contexte d'évaluation certificative où une note supérieure ou égale à 60 % représente les acquis nécessaires permettant aux étudiants de passer à un contexte d'apprentissage supérieur. Ainsi, il y aurait une entrave au principe de validité (Kane, 1992; Messick, 1989) si un étudiant éprouvant des difficultés avait obtenu un résultat au-dessus de la note de passage en étant chanceux, alors qu'il répondait au hasard. En effet, même si l'épreuve présentait des qualités psychométriques intéressantes (par exemple, une structure factorielle satisfaisante), l'évaluateur ne pourra pas procéder à une interprétation valide du score de celui-ci sans savoir que cet étudiant présentait un nombre de bonnes réponses artificiellement plus élevé qu'attendu.

De nombreux chercheurs se sont intéressés à l'étude de réponses inappropriées telle que la réponse au hasard. Leurs principaux travaux seront présentés comme suit. Premièrement, nous discuterons du problème des patrons de réponses inappropriés en contexte d'évaluation. Nous continuerons en recensant quelques solutions pertinentes pour détecter ces patrons de réponses inappropriés. La section suivante proposera une introduction portant sur les indices de détection (l'approche retenue dans cette thèse) avant de présenter une synthèse qui aboutira, finalement, à la question de recherche de ce projet.

1.1 Problématique des réponses inappropriées en évaluation

Lors de l'administration d'instruments visant à mesurer le niveau d'habileté d'un étudiant, il est fréquent d'assister à des tentatives de réponses inappropriées. Ainsi, le Centre pour l'intégrité académique (*Center for academic integrity*) rapportait que plus de 75 % des étudiants américains ont admis avoir déjà triché à un test (cité par Hutton, 2006). Dans le même sens, Laurier, Froio, Pearo et Fournier (1998) ont montré que 78 % des enseignants de niveau collégial croient que les étudiants ne répondent pas correctement aux épreuves d'évaluation.

Le phénomène discuté au paragraphe précédent peut aussi s'appliquer aux enseignants (Levitt et Jacob, 2003). Depuis l'imposition, aux États-Unis, de la loi *No child left behind*, les résultats des étudiants américains servent aussi à évaluer les résultats globaux de leurs établissements scolaires. Or, il semblerait que cette situation soit un terreau fertile pour les enseignants et les administrateurs qui souhaitent améliorer artificiellement les résultats de leurs étudiants aux évaluations nationales. En effet, Levitt et Dubner (2005) ont remarqué que des enseignants sélectionnaient certaines copies d'étudiants et y remplaçaient les mauvaises réponses par de bonnes réponses afin de faire augmenter la moyenne de la classe.

Si certains étudiants et enseignants tentent de se surclasser en trichant, d'autres peuvent aussi se sous-classer intentionnellement à une épreuve d'évaluation (Raïche, 2002). Dans une telle situation, un étudiant pourrait, par exemple, chercher à

être inséré au sein d'un groupe plus faible afin d'obtenir des résultats nettement au-dessus de la moyenne de la classe.

Comme nous venons de le constater, le sous-classement intentionnel peut se produire en milieu scolaire (Dodeen 2003; Meijer, 1998). Nous relevons aussi des exemples lors d'évaluations professionnelles pour l'obtention d'un emploi (Meijer, 1998; Zickar et Drasgow, 1996), lors d'évaluations psychologiques (Zickar et Drasgow, 1996) ou lors d'évaluations juridiques (Bénézech, 2007).

La tricherie ou le sous-classement intentionnel ne sont pas les seuls exemples de patrons de réponses inappropriés qui peuvent être donnés par un étudiant à une épreuve d'évaluation. Par exemple, il est reconnu que certains étudiants peuvent tenter de copier les réponses de collègues de classe qu'ils jugent plus habiles qu'eux (Brassard, 2011). L'occurrence de ce type de problème montre toute l'importance d'étudier plus attentivement le phénomène des comportements inappropriés lors d'épreuves d'évaluation. Au moins trois raisons nous poussent à croire que cela est pertinent.

Dans un premier temps, l'observation d'un biais systématique dans la façon de répondre d'un étudiant peut rendre invalide l'interprétation des scores (Messick, 1989) de ce dernier à une épreuve d'évaluation. En effet, le travail de régulation ou de classement peut être biaisé si l'étudiant ne présente pas des réponses qui reflètent son niveau de connaissances réel dans les épreuves d'évaluation.

Dans un deuxième temps, un autre problème a trait aux difficultés logistiques engendrées par le mauvais classement d'un étudiant. Raïche (2002) a déjà montré que

certains cégépiens souhaitaient se sous-classer au test d'anglais-langue seconde (TCALS-II) afin de faire partie d'un groupe plus faible où ils n'auraient pas à investir autant d'efforts d'apprentissage. Or, il serait souhaitable que ces étudiants puissent être adéquatement reclassés, car Fournier (1992) a déjà soulevé que ces individus ont tendance à être plus dérangeants que les autres étudiants. Ainsi, une telle situation engendre un travail de gestion de classe qui demande beaucoup de temps et d'énergie aux enseignants.

Dans un troisième temps, les recherches sur la motivation scolaire (Viau, 1994, 2009) peuvent être intéressantes afin d'étudier l'effet de certains comportements inappropriés lors d'épreuves d'évaluation. Ainsi, certaines études psychométriques démontrent que les individus qui adoptent systématiquement le sous-classement intentionnel comme stratégie peuvent biaiser la perception qu'ils ont d'eux-mêmes. Par exemple, Drasgow, Levine et McLaughlin (1987) ont soulevé l'idée que les étudiants qui ont l'habitude de volontairement présenter de mauvaises réponses à une épreuve d'évaluation risquent de développer une piètre image d'eux-mêmes à long terme. Or, de nombreux chercheurs ont démontré que l'étudiant doit évaluer positivement ses chances de réussite pour s'engager cognitivement (Laveault, 2009). Ainsi, être en mesure de détecter un tel comportement permet d'intervenir rapidement auprès de ces étudiants et d'éviter que leur motivation scolaire se déprécie.

1.1.1 Étude des patrons de réponse inappropriés en évaluation des apprentissages

L'étude de patrons de réponses inappropriés est une préoccupation importante depuis le début de la recherche en mesure et en évaluation. Par exemple, Thurstone (1927) reconnaissait déjà que certaines personnes pouvaient volontairement mal répondre à une épreuve d'évaluation. Cronbach démontrait, en 1946, l'existence de six grands types de patrons de réponses inappropriés: 1) la réponse au hasard (*tendency to gamble*), 2) la différenciation de la signification de la réponse (*definition of judgement categories*), 3) les réponses multiples (*inclusiveness*), 4) le biais engendré par l'acquiescement à un test vrai ou faux (*bias; acquiescence*), 5) la rapidité à répondre et son influence (*speed versus accuracy*) et 6) le style hétérogène de réponses (*miscellaneous*).

L'intérêt pour ce thème d'étude est resté d'actualité dans la deuxième moitié du vingtième siècle et au début des années 2000 (Meijer, 1996; Reise et Flannery, 1996; Trabin et Weiss, 1979; Wright, 1977; Wright et Stone, 1979). Quelques taxonomies ont même été développées pour tenter de comprendre et de classifier ces types de patrons de réponses (Brassard, 2011; Johnson, 1998; Meijer, 1996; Ro, 2001; Smith, 1982). À titre d'exemple, Ro (2001) a fait ressortir l'existence de 29 différents patrons de réponses inappropriés.

Il est important de le comprendre : les patrons de réponses étudiés dans le passé ne sont pas tous pertinents et applicables au monde de l'éducation. Nous

présentons, dans le tableau 1, quelques exemples de patrons de réponses qu'il est possible de rencontrer lors d'une évaluation en contexte scolaire.

Tableau 1.1 : Exemples de patrons de réponses inappropriés

Terminologie	Descripteur	Principaux auteurs
Répondre au hasard	Répondre sans réfléchir ou sans se soucier de connaître la bonne réponse	Glas et Meijer (2003); Hendrawan, Glas et Meijer (2005); Karabatsos (2003); Kogut (1987); Meijer (1994a), Meijer, Muijtjens, van der Vleuten (1996); Raïche (2002); Rogers et Hattie (1987).
Être inattentif	Ne pas porter toute son attention au test ou être inappliqué	Emons, Glas, Meijer et Sijtsma (2003); Karabatsos (2003); Raïche, Magis, Blais et Brochu (2013).
Tricher	Répondre au test en utilisant différentes stratégies (par exemple, copier) qui vont faire augmenter artificiellement l'habileté de l'étudiant	Meijer (1994a) ; Meijer, Muijtjens, van der Vleuten (1996)
Répondre trop rapidement	Répondre tellement rapidement que l'étudiant ne porte pas toute l'attention requise sur l'évaluation	Cronbach (1946); Brassard (2011); Wright et Stone (1979)
Être anxieux	Être tellement anxieux que l'étudiant répond de façon inadéquate au test	Brassard (2011); Hopkins, Stanley et Hopkins (1990)
Faire des choix en tenant compte du niveau de difficulté des questions	Identifier le niveau de difficulté des items, par exemple: facile ou difficile	Brassard (2011); Raïche (2002)
Répondre aux questions les plus faciles seulement	Identifier les questions faciles et décider de bien répondre ou non à cet item	Brassard (2011); Raïche (2002)
Choisir la mauvaise réponse volontairement	Ne répondre qu'aux items faciles	Brassard (2011); Raïche (2002)

Dans le cadre de cette thèse, nous nous intéresserons plus spécifiquement à la détection de la réponse au hasard et de l'inattention. Ce sont donc à ces comportements que nous référerons lorsque nous parlerons de patrons de réponses inappropriés.

Puisque l'existence de ce type de patrons de réponses peut engendrer des problèmes d'interprétation, il est important d'en faire la prévention et, si possible, la détection. Les principales propositions provenant des écrits scientifiques seront présentées dans la prochaine section.

1.2 Solutions pour détecter les patrons de réponses inappropriées

Il existe deux grandes catégories de solutions qui ont été proposées pour faire la prévention et la détection des patrons de réponses inappropriés. Nous les classerons selon qu'ils interviennent *avant* ou *après* la passation de l'épreuve d'évaluation. Ces catégories seront analysées dans les sections suivantes.

1.2.1 Intervention avant la passation de l'épreuve d'évaluation

Il est possible de faire de la prévention en intervenant avant la passation de l'épreuve d'évaluation. Par exemple, nous savons que les tentatives de réponses inappropriées, à une épreuve d'évaluation, peuvent être minimisées par l'intermédiaire de stratégies de prévention intuitives que les administrateurs et les

enseignants connaissent généralement bien. Selon Bertrand et Blais (2004), l'utilisation d'un local bien éclairé, le recours à plusieurs surveillants ou l'utilisation d'échelles de désirabilité sociale peuvent être des avenues à considérer.

De leur côté, Zickar et Drasgow (1996) recensent quelques approches plus sophistiquées qui permettent de prévenir l'occurrence de réponses inappropriées dans les tests. Notons, à titre d'exemple, que certains auteurs conseillent aux évaluateurs d'écrire des items difficiles à falsifier. Ainsi, Edwards (1970) a proposé de favoriser les questions ambiguës et moins explicites que ce que les administrateurs scolaires souhaitent mesurer. De plus, il peut être intéressant de considérer l'inclusion d'une échelle de validité visant à découvrir les réponses auxquelles un individu semble avoir répondu de façon aléatoire, par exemple.

La théorie microéconomique (Bade, Parkin et Gonzalez, 2010) permet aussi d'obtenir des balises permettant de prévenir les comportements inappropriés en situation d'évaluation. L'idée derrière cette théorie consiste à poser l'hypothèse que les individus évaluent les coûts et les bénéfices d'une situation donnée dans le but de poursuivre la voie qui maximisera leur satisfaction personnelle (Hutton, 2006). Ainsi, un étudiant tentera d'être chanceux en répondant au hasard à un examen d'entrée à l'université s'il estime que les coûts de se faire détecter sont peu élevés. Comme Levitt et Dubner (2005, p. 38) le soulignent : « tricher [cela s'applique également à la réponse au hasard] est un acte économique primordial [car] c'est obtenir davantage en donnant moins ». Dans cette situation, il serait possible de prévenir l'occurrence de réponses inappropriées à une épreuve d'évaluation en exposant les sanctions relatives

(par exemple, le renvoi de l'établissement d'enseignement) auxquelles s'expose l'étudiant qui souhaite répondre de façon inappropriée à une épreuve d'évaluation. En ce sens, la théorie microéconomique stipule qu'un étudiant révisera les bénéfices qu'il retire à adopter un comportement inapproprié si on lui indique clairement la nature des sanctions avant la passation de l'examen.

1.2.2 Intervention après la passation de l'épreuve d'évaluation

Il est possible d'intervenir après la passation de l'épreuve d'évaluation. Cette stratégie consiste généralement en deux approches distinctes. Dans un premier temps, un administrateur ou un enseignant peut tenter de rencontrer individuellement, ou en groupe, certains étudiants qui auraient répondu à une épreuve d'évaluation de façon inappropriée. Bien que cette approche semble intéressante, elle nécessite une très bonne connaissance des étudiants qui ont répondu à l'épreuve d'évaluation. De plus, la dimension subjective de l'intervention pourrait aussi poser des problèmes de nature éthique et peut difficilement être applicable à des évaluations contenant un grand nombre de répondants.

Dans un deuxième temps, il est possible d'analyser les réponses des étudiants en utilisant différents outils statistiques. Par exemple, Zijlstra, Van Der Ark et Sijtsma (2011) ont proposé d'utiliser la distance de Mahalanobis pour détecter les étudiants qui présentent des patrons de réponses différents de la majorité des autres individus évalués. De leur côté, Conijn, Emons, Van Assen et Sijtsma (2011)

proposent plutôt d'utiliser un modèle de régression logistique multiniveau pour permettre de détecter les patrons de réponses inappropriés.

Depuis quelques années, l'approche statistique qui retient le plus d'attention consiste à utiliser des indices de détection de patrons de réponses inappropriés (*person-fit indexes*). De nombreux auteurs ont déjà soulevé la flexibilité de ces indices et les résultats prometteurs qu'ils permettent d'obtenir (Emons, 2008; Karabatsos, 2003). Puisque c'est cette approche qui sera retenue dans le cadre de ce projet, nous la présenterons dans la prochaine section.

1.3 Les indices de détection de patrons de réponses inappropriés dans les épreuves d'évaluation

Les indices de détection de patrons de réponses inappropriés ont déjà été proposés comme un outil permettant de faire l'analyse des réponses des étudiants à des épreuves d'évaluation. Par exemple, il est possible de penser aux indices *Iz*, *ZU* ou *ZW*, qui seront formellement présentés au chapitre suivant.

Reise et Flannery (1996) ont écrit que le but d'utiliser de tels indices est de détecter les processus psychologiques qui sous-tendent un type de réponse inapproprié comme la réponse au hasard et l'inattention. Dans le même ordre d'idée, Emons (2008) écrivait que ces indices devraient même être utilisés comme outil de mesure visant à diagnostiquer les étudiants lors de la passation d'examens. Ainsi,

cette approche peut soutenir le bon jugement de l'enseignant qui a des doutes sur la qualité du patron de réponses d'un étudiant.

La majorité des auteurs s'entendent pour déclarer que les indices de détection de patrons de réponses inappropriés peuvent être classés en deux grandes catégories (Meijer et Sijtsma, 2001) : les indices qui ne reposent pas sur les paramètres de la théorie de la réponse à l'item (appelés indices non paramétriques) et les indices qui reposent sur les paramètres de théorie de la réponse à l'item (appelés indices paramétriques). Il est à noter que nous nous concentrerons uniquement sur les indices paramétriques dans le cadre de cette thèse. Trois raisons peuvent expliquer ce choix. Premièrement, leur nature probabiliste permet de procéder à la détection de patrons de réponses inappropriés (*person-fit*), de comparer les réponses de groupes d'étudiants pour chacune des questions (*differential item functioning*), d'administrer des tests adaptatifs (*computerized adaptive testing*), de procéder à la mise à l'échelle des réponses de différents groupes d'étudiants (*scaling et equating*), etc. De leur côté, les indices non paramétriques ont un champ d'application qui est beaucoup plus restreint à cause de leur nature non probabiliste. Deuxièmement, les indices paramétriques ont été plus souvent étudiés par les auteurs en mesure. De plus, ils sont généralement plus faciles à interpréter que les indices non paramétriques (Meijer et Sijtsma, 2001), car la fonction de densité de leurs scores suit souvent les quantiles d'une fonction de densité qui est connue (par exemple, la loi normale). Troisièmement, il est possible de corriger le biais du niveau d'habileté estimé des indices paramétriques (voir Snijders (2001) ou Raïche et Blais (2005)). Ce type de

correction n'est malheureusement pas encore disponible pour les indices non paramétriques.

1.3.1 Une difficulté à interpréter le score des indices paramétriques

À ce jour, de nombreuses recherches ont tenté d'étudier l'efficacité des indices de détection dans le cadre d'épreuves d'évaluation tirées de données réelles ou de données simulées par ordinateur (Dragow, 1982; Dragow, Levine et McLaughlin, 1987; Emons, Glas, Meijer et Sijtsma, 2003; Karabatsos, 2003; Li et Olejnik, 1997; Meijer, 1994; Meijer, Muijtens et van der Vleuten, 1996; Rogers et Hattie, 1987). Les résultats généraux sont déjà très encourageants et démontrent que cette approche semble flexible et efficace dans plusieurs situations d'évaluation différentes. Par exemple, Béland, Brassard et Raïche (2011) ont démontré que ces indices semblent surtout efficaces pour détecter la fluctuation du niveau d'habileté d'un étudiant dans une épreuve d'évaluation.

Un problème de taille subsiste, par contre, dans l'interprétation du score des indices paramétriques. Molenaar et Hoijtink (1990, 1996), Nering (1995, 1997) et Reise (1995) ont démontré que la fonction de densité de l'indice I_z ne suivrait pas exactement les quantiles de la loi normale $N(0,1)$ lorsque le niveau d'habileté θ des étudiants est estimé. De leur côté, Li et Olejnik (1997) ont démontré que les indices ZU et ZW posaient un problème au moment de l'interprétation de ces indices.

Ces conclusions sont très importantes, car elles rapportent le fait que si le point de coupure au seuil α sélectionné n'est pas le bon, l'interprétation du score d'un indice sera erronée. Cela peut, à la marge, créer deux grandes catégories d'erreurs (Molenaar et Hoijsink, 1990, 1996). Premièrement, une erreur de type I est une situation où un étudiant ayant bien répondu à l'épreuve d'évaluation serait classé comme ayant présenté un patron de réponses inapproprié. Deuxièmement, une erreur de type II stipule qu'un étudiant ayant répondu de façon inappropriée à une évaluation aurait été classé comme ayant un patron de réponses approprié.

Il est essentiel de procéder à de bonnes interprétations des patrons de réponses des étudiants, car cela peut avoir des conséquences graves sur celui qui n'aurait pas été correctement détecté. En effet, un évaluateur qui commettrait une erreur de type I (le patron de réponses d'un étudiant est détecté, à tort, comme étant inapproprié) pourrait créer chez l'étudiant une perte de confiance envers le système d'évaluation. Or, cela pourrait mener à des conséquences désastreuses si l'étudiant en vient à percevoir négativement sa probabilité de réussir, à se désengager progressivement et à être moins motivé (Louis, 2004).

Aujourd'hui, la recherche d'un estimateur moins biaisé du niveau d'habileté reste un défi majeur à relever pour améliorer le jugement posé par l'évaluateur à l'égard des étudiants. Dans leur article de référence, Meijer et Sijtsma (2001) ont d'ailleurs soulevé l'importance de trouver rapidement une voie pour dépasser ce problème, car en ce moment, la quasi-totalité des recherches rapporte des résultats sans se soucier de cette limite importante.

1.4 Synthèse

Les études recensées dans la section précédente ont mis en évidence le fait qu'il existe plusieurs indices paramétriques qui sont interprétables, car ils suivent les quantiles d'une fonction de densité connue. Un exemple évocateur est I_z , qui est fort probablement le plus populaire et le plus étudié de tous les indices de détection. Par contre, un problème survient avec l'interprétation des scores de celui-ci lorsqu'on estime le niveau d'habileté de l'étudiant. Ce problème est aussi applicable à plusieurs autres indices paramétriques qui ont un point de coupure connu : mentionnons ZU (Wright et Masters, 1982), ZW (Wright et Masters, 1982) ou les indices de prudence standardisés qui ont été développés par Tatsuoka et Linn (1983). Malheureusement, les études qui comparent les indices entre eux n'ont pas tenu compte de ce problème et il est aujourd'hui important de trouver des méthodes pour corriger ces indices et procéder à des interprétations plus justes des scores obtenus par ces derniers. Rappelons que l'évaluation des apprentissages a comme objectif de permettre à l'enseignant de poser un jugement visant à aider l'étudiant à progresser adéquatement dans ses apprentissages. Ainsi, il est important d'utiliser des indices corrigés pour soutenir le jugement d'un évaluateur.

1.5 La question de recherche

La question de recherche de ce projet doctoral peut s'énoncer comme suit : *comment tenir compte du fait que, dans le calcul des indices paramétriques de détection des patrons de réponses inappropriés, le niveau d'habileté réel d'un étudiant n'est malheureusement pas connu?*

CHAPITRE II

CADRE THÉORIQUE

Dans le cadre de ce projet doctoral, nous nous concentrons à étudier les indices de détection de patrons de réponses inappropriés qui sont basés sur les modélisations issues de la théorie de la réponse à l'item (appelés « indices paramétriques » dans ce chapitre). Nous constatons que ces indices ont été plus souvent investigués par les chercheurs, car plusieurs d'entre eux peuvent être interprétés à partir des quantiles d'une fonction de densité connue (Meijer et Sijtsma, 2001).

Il a déjà été soulevé que ces indices présentent des limites, car le niveau d'habileté réel des étudiants évalués n'est pas connu et que, conséquemment, c'est une estimation de celui-ci qui est utilisée dans le calcul des indices (Molenaar et Hoijtink, 1990, 1996). Pour répondre à ce problème, quelques stratégies ont été proposées ces dernières années. Par exemple, il est possible de penser au modèle multidimensionnel de Raïche, Magis, Blais et Brochu (2012) ou à la correction proposée par Snijders (2001).

Ce chapitre présentera plus en détail les indices paramétriques et les principales stratégies éduométriques de correction du niveau d'habileté. Nous procéderons comme suit. La théorie de la réponse à l'item sera expliquée dans la section suivante alors que les principaux indices de détection paramétriques seront présentés dans la section 2.2. Ensuite, les solutions qui ont été proposées pour tenir

compte du fait que le niveau d'habileté réel n'est pas connu seront discutées. Enfin, nous présenterons une synthèse et les objectifs de cette recherche dans les deux dernières sections de ce chapitre.

2.1 La théorie de la réponse à l'item

Bien que les premières épreuves de classement d'individus remontent à l'an 1115 avant Jésus-Christ (Bertrand et Blais, 2004), l'histoire de la mesure contemporaine débute à la fin du 19^e siècle avec les travaux d'auteurs tels que Francis Galton, Alfred Binet et Charles Spearman (Hambleton et Swaminathan, 1985; Sijtsma et Junker, 2006). Ces premiers travaux s'étendent jusqu'aux années 1950 et sont, aujourd'hui, associés principalement à la théorie classique des tests.

Dans le cadre de cette thèse, nous nous intéressons plus particulièrement à une approche qui s'est développée à la suite de la Deuxième Guerre mondiale : la théorie de la réponse à l'item appliquée à des matrices de réponses dichotomiques (Birnbaum, 1968; Hambleton et Swaminathan, 1985; Lord, 1980). Au moins trois arguments généraux peuvent expliquer l'émergence de cette théorie. Premièrement, la théorie classique des tests est une approche qui est dépendante des résultats des étudiants entre eux. Pour illustrer cela, il peut être pertinent de référer à la difficulté d'un item, qui est définie comme étant la proportion d'étudiants qui répondent correctement à cet item. Imaginons que des experts aient créé un item de difficulté moyenne. Si un groupe d'étudiants est peu habile, cet item sera considéré comme

étant difficile. Par contre, si un autre groupe est très habile, le même item sera considéré comme étant trop facile. Cette illustration témoigne donc d'un problème important : l'analyse des items est différente d'un groupe à l'autre, ce qui complexifie les comparaisons à plus grande échelle.

En ce qui concerne la théorie de la réponse à l'item, l'analyse des items n'est pas dépendante des résultats des groupes d'étudiants à un examen. Cela lui confère donc l'important avantage d'être un outil de mesure invariant : on peut calculer la difficulté d'un item sans se soucier des réponses fournies par l'ensemble des groupes d'étudiants à un examen.

Deuxièmement, le développement de la micro-informatique a beaucoup contribué à faire de cette approche un modèle populaire. Aujourd'hui, il existe toute une variété de logiciels payants (par exemple, BILOG-MG ou RUMM-2030) et de logiciels gratuits (par exemple, R ou ICL) qui permettent d'analyser des résultats d'examen à l'aide de cette théorie.

Troisièmement, le champ d'application de la théorie de la réponse à l'item est très vaste et ouvre la porte à toute une gamme d'analyses pertinentes : détecter les patrons de réponses inappropriés (*person-fit*), comparer les réponses de groupes d'étudiants pour chacune des questions (*differential item functioning*), administrer des tests adaptatifs (*computerized adaptive testing*), procéder à la mise à l'échelle des réponses de différents groupes d'étudiants (*scaling et equating*), etc.

La section suivante sera divisée en deux parties. D'abord, nous présenterons les principales modélisations issues de la théorie de la réponse à l'item à réponse

dichotomique. Ensuite, nous discuterons des méthodes d'estimation qui seront utilisées dans le cadre de cette thèse.

2.1.1.1 Les modélisations à réponse dichotomique

Les modèles de réponse à l'item dichotomique reposent sur l'idée que l'on puisse mesurer le niveau d'habileté d'un étudiant (considéré comme un trait latent¹) à partir des bonnes ou mauvaises réponses qu'il a fournies à une série de questions à une épreuve d'évaluation. Pour comprendre cette idée, il est pertinent de référer à la figure 2.1. Dans ce cas-ci, le trait latent θ , représenté par un cercle, est inféré à partir des réponses manifestes à quatre items, notés x_1 à x_4 et représentées par des carrés. Nous remarquons que les flèches vont du trait latent aux items, ce qui indique que c'est le niveau d'habileté θ qui explique la réponse aux items.

¹ Le lecteur qui souhaite obtenir plus d'information sur la nature théorique du trait latent pourra consulter Borsboom, Mellenbergh et Van Heerden (2003).

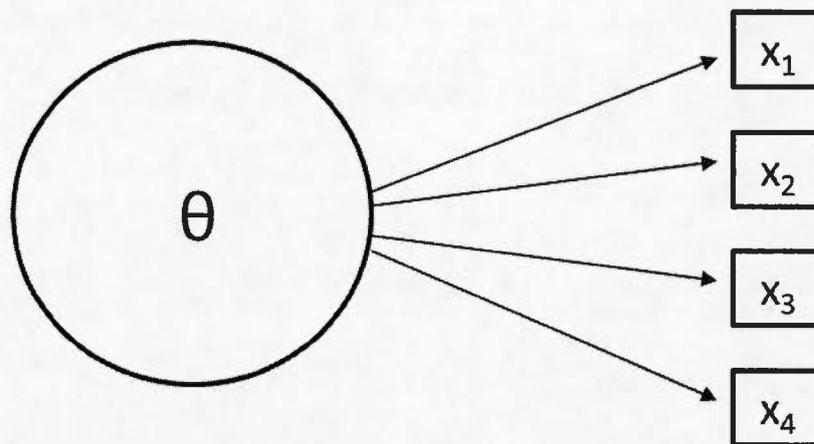


Figure 2.1 Réponses manifestes et trait latent

Imaginons une situation où nous souhaitons mesurer le niveau d'habileté d'un étudiant à partir des bonnes ou mauvaises réponses qu'il a fournies à une série d'items. Dans une telle situation, un item contient une bonne réponse (généralement représentée par « 1 ») et plusieurs leurres qui sont des mauvaises réponses (généralement représentés par « 0 »). Par exemple, la question suivante permet d'observer une manifestation du niveau d'habileté en mathématique :

Paul a 2 guitares et 3 tambours. Ringo a 3 mandolines et 2 harmonicas.

Combien d'instruments à cordes ont Paul et Ringo, ensemble?

- a) 3 instruments à cordes
- b) 4 instruments à cordes
- c) 5 instruments à cordes.

Dans ce cas-ci, il existe une seule bonne réponse (la réponse c) et deux mauvaises réponses. Ainsi, un étudiant qui aurait répondu c) obtiendrait un score d'un alors qu'un étudiant qui aurait répondu a) ou b) obtiendrait le score de zéro.

2.1.1.2 La théorie de la réponse à l'item : les postulats de base

Bien qu'il soit possible de retrouver des références historiques qui remontent aux travaux de Richardson (1936) et à ceux de Lawley (1943, 1944), les modèles de réponse à l'item dichotomique sont issus d'une transformation en fonction logistique du modèle d'ogive normale initialement proposée par Lord (1952). Cela est rendu possible grâce à l'utilisation d'une constante égale à 1,702 (appelée constante de Haley) qui vise à minimiser les différences existant entre ces deux modèles (Camilli, 1994).

Trois postulats doivent être respectés pour que ces modèles puissent être appliqués adéquatement à une matrice de données dichotomiques. Dans un premier temps, il doit y avoir indépendance locale entre les items d'une épreuve d'évaluation. Ainsi, pour un niveau d'habileté θ fixe, la probabilité du patron de réponses aux items A et B est égale au produit des probabilités suivantes :

$$P(x_A \wedge x_B | \theta) = P(x_A | \theta)P(x_B | \theta). \quad (1)$$

Dans un deuxième temps, il doit y avoir un seul niveau d'habileté ou un niveau d'habileté principal mesuré dans l'épreuve d'évaluation (postulat d'unidimensionnalité). Par exemple, dans le cadre d'un test de classement en anglais langue seconde, il ne devrait y avoir que l'habileté en anglais qui est évaluée. La figure 2.1 schématise cette idée : les réponses aux items x_1 à x_4 sont toutes tributaires d'un seul trait latent θ .

Dans un troisième temps, les modèles de réponses à l'item imposent une relation de monotonie. Si le niveau d'habileté estimé d'un étudiant nommé Félix $\theta_{Félix}$ est plus élevé que le niveau d'habileté estimé d'une étudiante nommée Sophie θ_{Sophie} , les probabilités d'obtenir une bonne réponse à l'item i doivent suivre la logique suivante :

$$P(x_i = 1 | \theta_{Félix}) > P(x_i = 1 | \theta_{Sophie}). \quad (2)$$

Ainsi, la probabilité d'obtenir une bonne réponse de Félix doit être plus élevée que la probabilité d'obtenir une bonne réponse de Sophie.

Maintenant, trois modélisations logistiques à réponses dichotomiques seront présentées : le modèle de Rasch (ou modèle à un paramètre), le modèle à deux paramètres et le modèle à trois paramètres.

2.1.1.3 Le modèle de Rasch

Le modèle de Rasch (1960), aussi appelé modèle à un paramètre (Bertrand et Blais, 2004), permet de calculer la probabilité $P_i(\theta)$ qu'un étudiant obtienne une bonne réponse à un item en se basant uniquement sur le niveau de difficulté de l'item. Dans le cadre de ce chapitre, x_i représente la réponse de l'étudiant aux items $i(i=1, \dots, I)$. Mathématiquement, nous écrivons ce modèle comme suit :

$$P_i(\theta) = P(x_i = 1 | \theta, b_i) = \frac{\exp[(\theta - b_i)]}{1 + \exp[(\theta - b_i)]} \quad (3)$$

où θ est un paramètre d'habileté du sujet et b_i un paramètre de difficulté de l'item. Ce modèle présente l'avantage de mettre sur une même unité de mesure le niveau d'habileté de l'étudiant et la difficulté de l'item. Cela facilite l'interprétation de ces paramètres puisqu'ils peuvent être considérés comme des scores standardisés (scores z).

La figure 2.2 présente deux courbes caractéristiques d'items selon la modélisation de Rasch. Il est possible de remarquer que la probabilité d'obtenir une bonne réponse à l'item 1 (voir l'échelle sur l'axe vertical) augmente avec le niveau estimé de l'habileté (voir l'échelle sur l'axe horizontal).

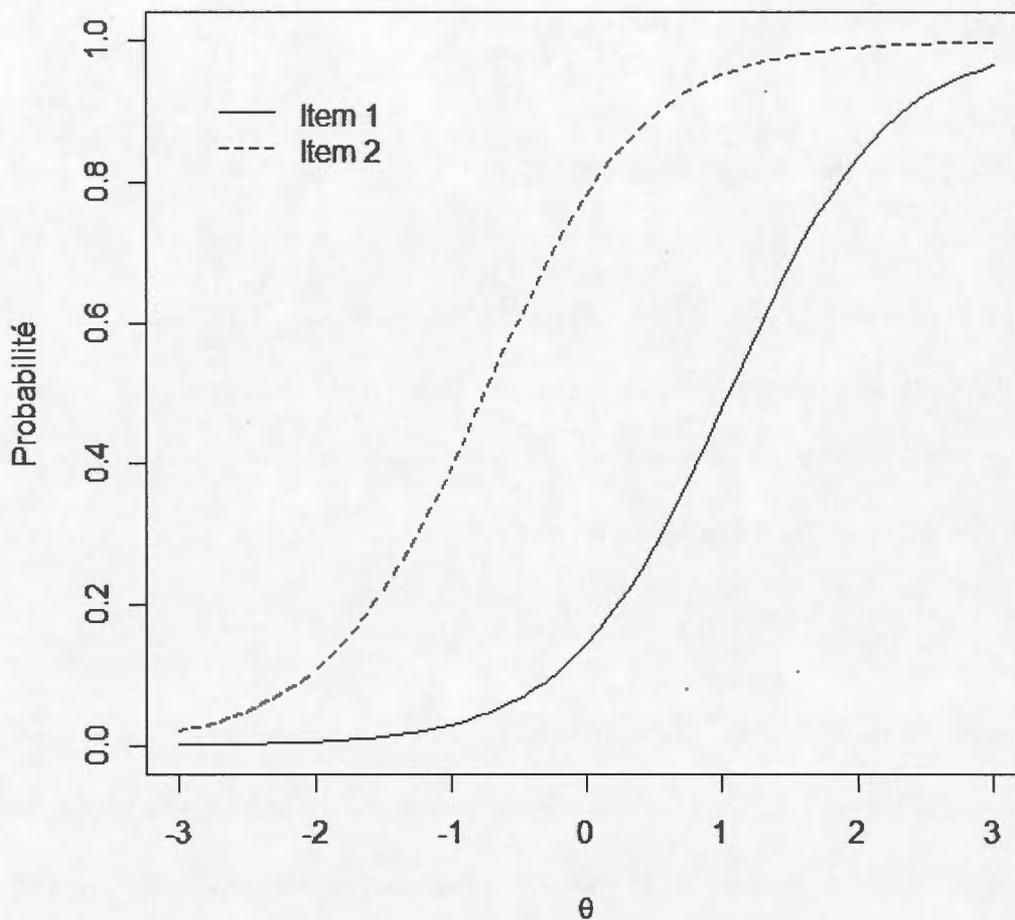


Figure 2.2 Courbes caractéristiques d'items selon le modèle de Rasch

Ainsi, chaque item est représenté par une courbe qui lui est propre. Nous retrouvons une courbe pour l'item 1, qui est l'item le plus difficile, et une courbe pour l'item 2, qui est l'item le plus facile. Par exemple, un étudiant qui aurait un niveau d'habileté estimé à zéro aurait environ une probabilité égale à 0,10 d'obtenir une bonne réponse

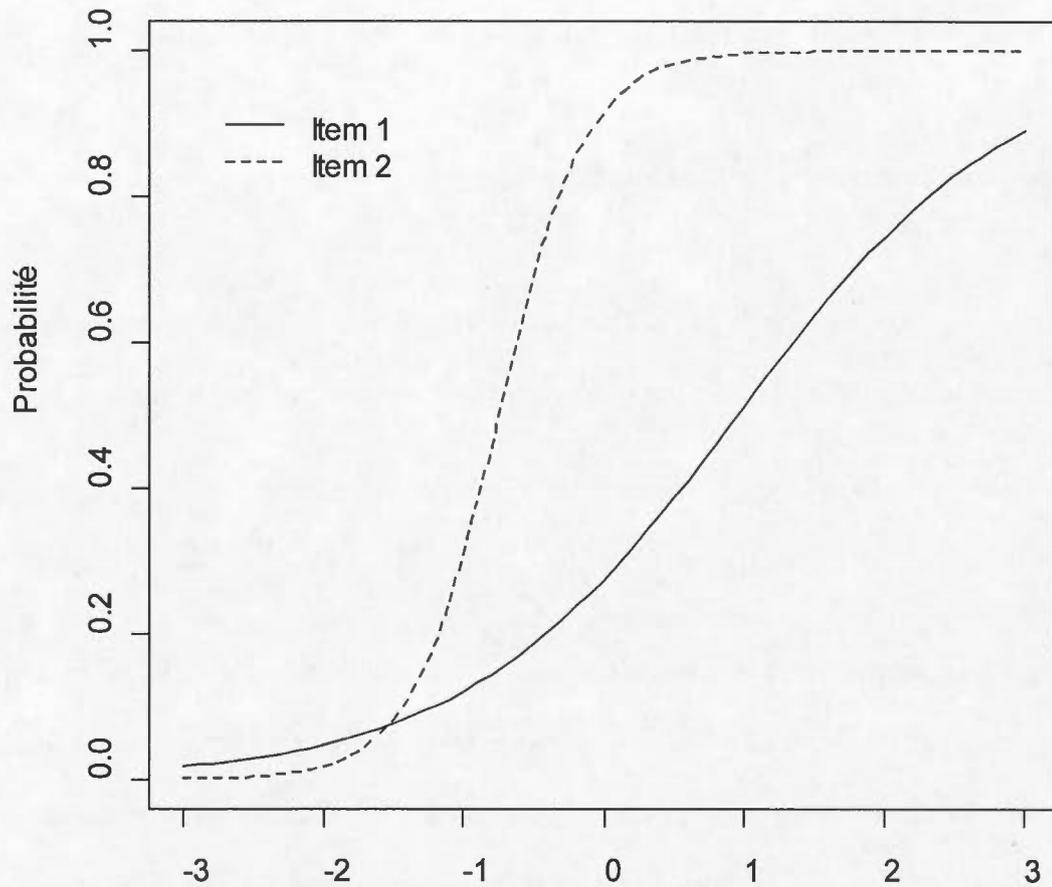
à l'item 1 et de 0,80 d'obtenir une bonne réponse à l'item 2. Pour un étudiant qui aurait un niveau d'habileté estimé à 2,00, cette probabilité est d'un peu plus de 0,80 pour l'item 1 et de près de 1,00 pour l'item 2.

2.1.1.4 Le modèle à deux paramètres

Le modèle de Rasch présente la distinction que toutes les courbes caractéristiques d'items sont parallèles et ne peuvent jamais se croiser. Pour permettre à ce modèle d'être plus flexible, Birnbaum (1968) eut l'idée de modéliser la pente en réécrivant l'équation (3) de la façon suivante :

$$P_i(\theta) = P(x_i = 1 | \theta, a_i, b_i) = \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]} \quad (4)$$

Comme dans le modèle de Rasch, nous retrouvons un paramètre θ , qui représente le niveau d'habileté de l'étudiant et un paramètre b_i , qui représente la difficulté de l'item. L'ajout principal de cette modélisation est un paramètre de discrimination de l'item, noté a_i . L'inclusion de a_i permet de considérer la propriété qu'a un item à discriminer les étudiants dont le niveau d'habileté est élevé et les étudiants dont le niveau d'habileté est plus faible. Ce paramètre est donc un facteur de poids permettant de faire varier la pente de la courbe caractéristique de l'item. Le lecteur intéressé pourra aller consulter l'article de Tuerlinckx et De Boeck (2005) pour obtenir plus d'information sur la nature de a_i .



□

Figure 2.3 Courbes caractéristiques d'items selon le modèle à deux paramètres

La figure 2.3 présente des courbes sensiblement différentes de celles de la figure 2.2, car, dans celles-ci, les courbes ne sont plus parallèles et peuvent se croiser. Pour un niveau d'habileté θ de zéro, la probabilité d'obtenir une bonne réponse est

d'environ 0,25 pour l'item 1 et d'un peu plus de 0,90 pour l'item 2. Par contre, nous remarquons qu'au niveau d'habileté estimé à -2,00, la probabilité d'obtenir une bonne réponse à l'item 1 est légèrement plus élevée que celle obtenue à l'item 2.

2.1.1.5 Le modèle à trois paramètres

Dans le cadre du modèle logistique à trois paramètres (Birnbaum, 1968), nous pouvons décrire la probabilité qu'un étudiant obtienne une bonne réponse à un item comme suit :

$$P_i(\theta) = P(x_i = 1 | \theta, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]} \quad (5)$$

où θ est un paramètre d'habileté du sujet, b_i un paramètre de difficulté de l'item, a_i un paramètre de discrimination de l'item et c_i un paramètre de pseudo-chance de l'item. L'ajout du paramètre c_i permet de modéliser l'asymptote inférieure de la courbe caractéristique d'item. Pour l'expliquer autrement, cela rend possible la probabilité qu'un étudiant faible obtienne une bonne réponse à un item de diverses façons, par exemple, en répondant au hasard ou par attirance pour la bonne réponse. Par exemple, imaginons qu'un item comporte quatre choix de réponse. Dans ce cas-ci, cet étudiant aurait une probabilité de 0,25 d'obtenir une bonne réponse en répondant de façon aléatoire.

Comparativement aux deux figures précédentes, la figure 2.4 présente une asymptote inférieure qui ne débute pas à zéro sur l'axe des probabilités estimées; ce

qui sous-entend que les sujets dont le niveau d'habileté estimé est faible ont une probabilité non nulle d'obtenir une bonne réponse.

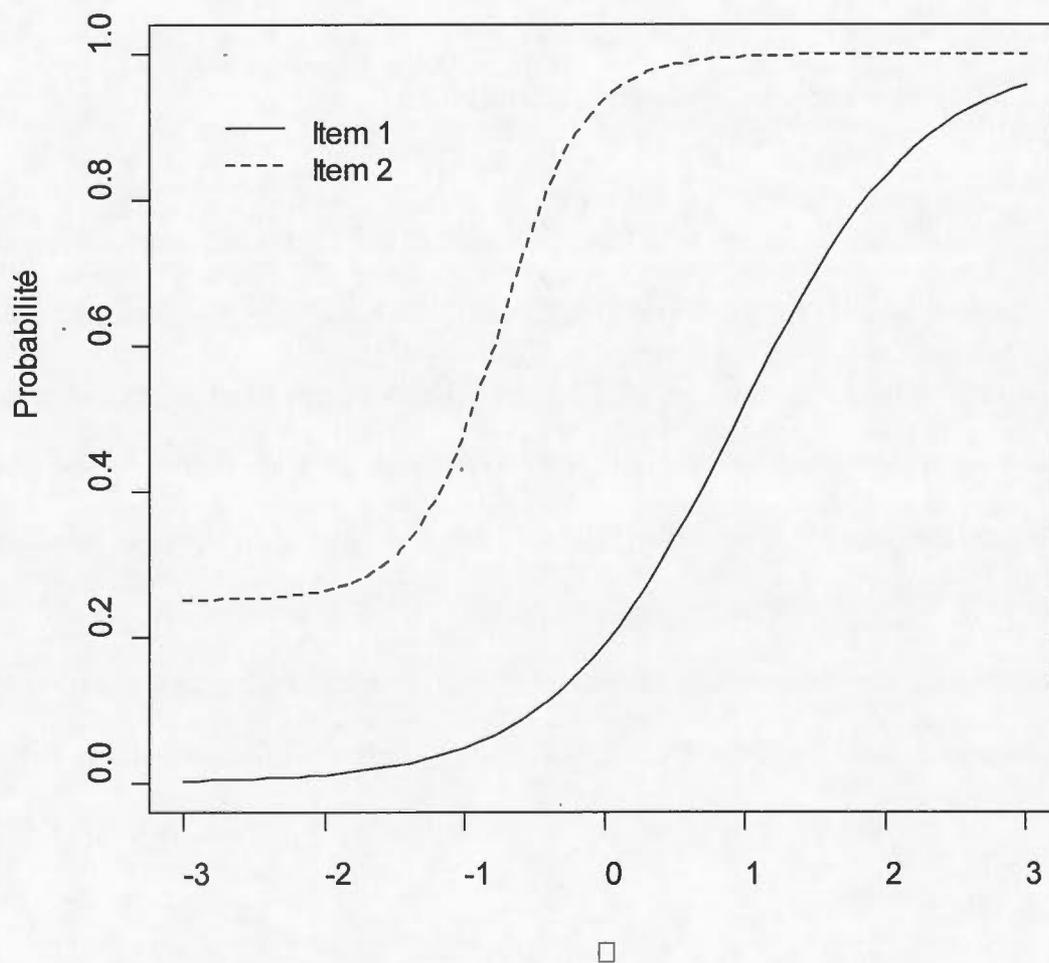


Figure 2.4 Courbes caractéristiques d'items selon le modèle à trois paramètres

Pour un niveau d'habileté θ de -3,00, la probabilité d'obtenir une bonne réponse est de près de 0,00 pour l'item 1 et de 0,25 pour l'item 2. Enfin, il est à noter que le modèle à trois paramètres (équation 5) se réduit au modèle à deux paramètres (équation 4) lorsque le paramètre $c_i = 0$ et au modèle de Rasch (équation 3) lorsque le paramètre $c_i = 0$ et le paramètre $a_i = 1$.

2.1.2.1 Estimation des paramètres d'item a_i , b_i et c_i

Comme le lecteur peut le comprendre, la valeur des paramètres cités au sein des équations (3) à (5) doit être disponible afin de calculer $P_i(\theta)$. Dans ce cas-ci, les paramètres d'items a_i , b_i et c_i peuvent être estimés à partir de plusieurs méthodes qui ont déjà été exposées plus en détail par Baker et Kim (2004), Hambleton et Swaminathan (1985) et Lord (1980). Citons, à titre d'exemple, la méthode du maximum de vraisemblance conjointe, du maximum de vraisemblance marginale, du maximum de vraisemblance conditionnelle ou les méthodes bayésiennes utilisant les chaînes de Markov Monte Carlo (*Monte Carlo Markov Chain*). Considérant l'objet de cette thèse, ces méthodes d'estimation des paramètres d'items ne seront pas décrites plus en détail.

2.1.2.2 L'estimation du paramètre d'habileté de l'étudiant θ

Dans le cadre de ce projet, il est plus important de se concentrer sur le niveau d'habileté d'un étudiant. Dans cette section, nous présenterons succinctement les trois estimateurs du niveau d'habileté qui sont les plus fréquemment cités dans les écrits en mesure.

2.1.2.3 Estimateur par maximum de vraisemblance

La méthode du maximum de vraisemblance θ_{ML} (*maximum likelihood*, ML) détermine la valeur du niveau d'habileté de l'étudiant θ qui maximise la vraisemblance d'un patron de réponses :

$$L(\theta) = \prod_i P_i(\theta)^{x_i} Q_i(\theta)^{1-x_i} \quad (6)$$

où $Q_i(\theta)$ est égal à $1-P_i(\theta)$. Pour le patron de réponses $X=[1,0,0,1,0,1]$, nous cherchons à trouver la valeur de θ qui maximise $L(\theta)=P_1Q_2Q_3P_4Q_5P_6$. Ainsi, le niveau d'habileté estimé d'un étudiant sera la valeur obtenue au moment où la dérivée première de $L(\theta)$ est égale à zéro :

$$\left. \frac{d L(\theta)}{d\theta} \right|_{\theta=\hat{\theta}_{ML}} = 0, \quad (7)$$

qui est la valeur de $L(\theta)$ maximale pour un patron de réponses.

2.1.2.4 Estimateur par maximum *a posteriori*

La méthode d'estimation par maximum *a posteriori* θ_{MAP} (MAP) est une approche qui utilise une information *a priori* pour pondérer la vraisemblance $L(\theta)$ du patron de réponses. Dans ce cas-ci, nous utilisons deux éléments pour obtenir une estimation du niveau d'habileté des étudiants. Dans un premier temps, nous devons utiliser une fonction de densité connue du niveau d'habileté des étudiants, notée $f(\theta)$, qui est généralement la loi normale $N(\mu, \sigma^2)$:

$$f(\theta; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) \quad (8)$$

où μ est la moyenne et σ^2 est l'écart-type. Dans un deuxième temps, nous devons calculer la probabilité $L(\theta)$. Enfin, ces deux éléments permettent d'obtenir la fonction qui maximise la fonction de densité *a posteriori* $g(\theta)$ où $g(\theta) = f(\theta) L(\theta)$. Ainsi, cette méthode d'estimation peut s'écrire :

$$\left. \frac{d g(\theta)}{d\theta} \right|_{\theta=\hat{\theta}_{MAP}} = 0, \quad (9)$$

ce qui correspond à la valeur obtenue lorsque la dérivée première de $g(\theta)$ est égale à zéro.

2.1.2.5 Estimateur par vraisemblance maximale pondérée

La méthode d'estimation par vraisemblance maximale pondérée θ_{WLE} (*weighted likelihood*, WLE) suit un peu la même logique que l'approche selon la méthode d'estimation par maximum *a posteriori*, puisqu'elle intègre aussi une information *a priori*. Par contre, cette information vise plutôt à réduire le biais lié à l'estimateur du niveau d'habileté. Ainsi, Warm (1989) a proposé que :

$$\frac{d \log f(\theta)}{d\theta} = \frac{J(\theta)}{2 I_{warm}(\theta)} \quad (10)$$

où

$$I_{warm}(\theta) = \sum_{i=1}^I \frac{[P_i'(\theta)]^2}{P_i(\theta) Q_i(\theta)} \quad \text{et} \quad J(\theta) = \sum_{i=1}^I \frac{P_i'(\theta) P_i''(\theta)}{P_i(\theta) Q_i(\theta)}. \quad (11)$$

De plus, notez que $P_i'(\theta)$ est la dérivée première de $P_i(\theta)$ et $P_i''(\theta)$ est la dérivée seconde de $P_i(\theta)$.

Les modèles de réponse à l'item présentés aux sections 2.1.1.3 à 2.1.1.5 et les méthodes d'estimation du niveau d'habileté de l'étudiant θ présentées aux sections 2.1.2.3 à 2.1.2.5 sont utilisés dans le calcul des probabilités $P_i(\theta)$, qui sont à la base du calcul des indices paramétriques de détection des patrons de réponses inappropriés. La section suivante décrira la méthode de calcul des indices de détection des patrons de réponses inappropriés les plus populaires.

2.2 Présentation des principaux indices qui sont basés sur la théorie de la réponse à l'item

Il existe toute une variété d'indices paramétriques de détection de patrons de réponses inappropriés qui sont calculés à partir des modèles de la théorie de la réponse à l'item (voir les équations 3 à 5). Dans le cadre de cette section, nous présenterons successivement un mode d'emploi des indices de détection avant de présenter ceux qui sont les plus souvent cités dans les écrits scientifiques et ceux pour lesquels des points de coupure sont connus.

2.2.1 Les indices de détection pour réponse dichotomique: un mode d'emploi

En nous inspirant de Molenaar et Hoijtink (1990, 1996) et de St-Onge (2007), nous pouvons montrer que l'utilisation d'indices de détection de patrons de réponses inappropriés comporte généralement trois grandes séquences successives : le choix de l'indice, la sélection du point de coupure et la prise de décision. Nous les décrirons dans les sections suivantes.

2.2.1.1 Le choix de l'indice

Selon Meijer et Sijtsma (2001), il n'existerait pas de meilleur indice de détection de patron de réponses inapproprié. L'évaluateur doit plutôt sélectionner l'indice qui semble le plus adéquat pour analyser les données découlant d'une épreuve spécifique d'évaluation. Par exemple, si l'épreuve d'évaluation comporte uniquement les réponses de quelques étudiants, l'indice non paramétrique G peut s'avérer fort utile (Meijer, 1994a). Dans une autre situation, un chercheur peut souhaiter utiliser un indice pour détecter un type de réponse aussi précis que la réponse au hasard ou l'inattention. Dans ce cas, il est important de sélectionner un indice qui est connu pour être efficace à détecter un type de réponses inapproprié. Par exemple, Karabatsos (2003) a démontré que l'indice I_z était plus efficace que l'indice G pour détecter la réponse au hasard. De leur côté, Stark, Chernyshenko et Drasgow (2005) ont plutôt élaboré un indice qui semble prometteur pour détecter les étudiants qui sont peu motivés lors d'un examen.

2.2.1.2 La sélection d'un point de coupure

À la base, le score obtenu par un indice ne donne pas beaucoup d'information permettant de juger de la qualité d'un patron de réponses. Pour que ce score soit intelligible, le chercheur doit reconnaître un point de coupure (*cut score*) qui servira à distinguer les patrons de réponses appropriés et les patrons de réponses qui sont

inappropriés. Les écrits scientifiques ont démontré que les indices paramétriques sont plus faciles à interpréter, car ils sont beaucoup plus nombreux à présenter des points de coupure connus (Meijer et Sijtsma, 2001).

Dans une situation où le score d'un indice suit les quantiles d'une loi de probabilité connue, il est possible d'interpréter ce score à l'aide d'un test d'hypothèse classique. Ainsi, nous pouvons stipuler l'hypothèse nulle H_0 , qu'un étudiant présente un patron de réponses approprié à l'égard d'un modèle de mesure de la façon suivante:

H_0 : le patron de réponses est approprié.

À l'opposé, l'hypothèse alternative H_1 qu'un étudiant présente un patron de réponses inapproprié à l'égard d'un modèle de mesure s'écrit:

H_1 : le patron de réponses est inapproprié.

Dans le cas de l'indice I_z , qui devrait se distribuer selon la loi $N(0,1)$ (Drasgow, Levine et Williams, 1985), on dira qu'un patron de réponses est inapproprié lorsqu'il respecte la condition $I_z \leq z_\alpha$, où Z_α correspond à un seuil Z critique. La figure suivante exprime les points de coupure pour les seuils $Z_{0,01} = -2,33$, $Z_{0,05} = -1,64$ et $Z_{0,1} = -1,28$.

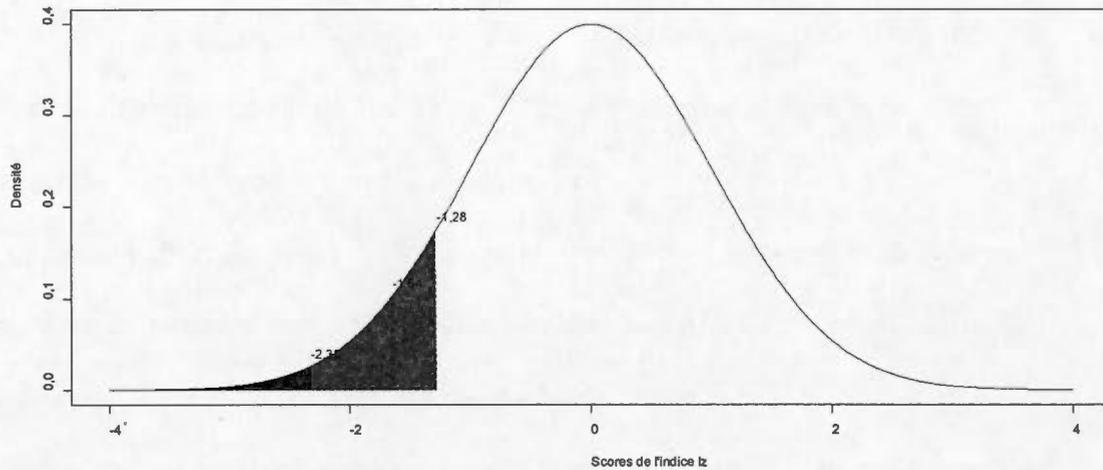


Figure 2.5 Les seuils $Z_{0,01}=-2,33$, $Z_{0,05}=-1,64$ et $Z_{0,1}=-1,28$

Ainsi, au point de coupure sélectionné au seuil α de 0,01, un étudiant présentant un score l_z plus petit ou égal à -2,33 sera considéré comme présentant un patron de réponses inapproprié. Statistiquement, il s'agit d'un rejet de l'hypothèse nulle H_0 . À ce point de coupure, les valeurs supérieures à -2,33 indiquent, au contraire, que le patron de réponses d'un étudiant est approprié.

2.2.1.3 La prise de décision

Dans cette dernière étape, l'évaluateur doit prendre une décision à l'égard des résultats obtenus à l'aide d'un indice de détection. Comme nous l'avons démontré à la section précédente, cette décision est de nature binaire : le score de l'indice est

comparé à un point de coupure qui détermine si le patron de réponses d'un étudiant est « approprié » ou « inapproprié ».

À cause de sa simplicité relative, ce processus n'est pas exempt de mauvaises décisions. Les écrits en mesure démontrent d'ailleurs que deux erreurs peuvent être commises dans une telle situation (Bourque, Blais et Larose, 2009). Premièrement, l'erreur de type I (rejeter H_0 alors qu'elle est vraie) est une situation où un étudiant ayant bien répondu à l'épreuve serait classé comme ayant présenté un patron de réponses inapproprié. Deuxièmement, l'erreur de type II (rejeter H_0 alors qu'elle est fausse) stipule qu'un étudiant ayant répondu de façon inappropriée à une évaluation aurait été classé comme ayant un patron de réponses approprié.

Comme nous l'avons soulevé au chapitre précédent, il est important de minimiser les risques d'occurrence de ces erreurs (par exemple, en augmentant ou en diminuant le point de coupure au seuil α retenu), car elles peuvent avoir des conséquences graves sur les étudiants qui n'auraient pas été correctement détectés.

Comme Molenaar et Hoijtink (1990, 1996) le soulignent, cette dernière étape est la plus délicate, car le calcul des indices comporte un autre problème de taille : nous ne connaissons pas le niveau d'habileté réel de l'étudiant évalué. En effet, nos calculs sont produits à partir de méthodes d'estimation statistiques qui reposent généralement sur la logique du maximum de vraisemblance ou de leur version bayésienne (Baker et Kim, 2005). Ainsi, nous avons seulement un estimateur biaisé du niveau d'habileté réel de l'étudiant. Or, tous les scores des indices de détection

sont basés sur cet estimateur biaisé en faisant comme si ce résultat correspond au niveau d'habileté réel de l'étudiant.

2.2.2 Les indices paramétriques les plus populaires

Il existe un grand nombre d'indices de détection de patrons de réponses inappropriés. Pour ne pas alourdir inutilement la présentation, nous allons présenter uniquement les indices paramétriques les plus populaires.

2.2.2.1 Les indices de type vraisemblance

Initialement présenté par Levine et Drasgow (1979), l'indice l_0 calcule le maximum du logarithme de la vraisemblance d'un patron de réponses tel que défini en (6) :

$$l_0(\theta) = \log L(\theta) = \sum_{i=1}^I \{x_i \log P_i(\theta) + (1 - x_i) \log Q_i(\theta)\} . \quad (12)$$

Cette approche permet de déterminer si un étudiant répond en conformité avec un modèle de réponse à l'item tel que le modèle de Rasch, le modèle à deux paramètres ou le modèle à trois paramètres. Dans la situation où l'étudiant répond conformément à l'un de ces modèles, la fonction de vraisemblance tendra à atteindre sa valeur maximale. À l'opposé, on considère que plus le résultat de l_0 est faible, plus les réponses d'un étudiant peuvent être considérées comme étant inappropriées.

Bien qu'il semble intéressant, cet indice présente un problème de taille : son interprétation est difficile, car il n'est pas indépendant du niveau d'habileté estimé d'un étudiant et il n'existe pas de valeur précise à partir de laquelle on considère qu'un patron de réponses est inapproprié. Pour cette raison, Drasgow, Levine et Williams (1985) ont élaboré une version standardisée de l_0 : l'indice l_z . Mathématiquement, l_z est une mise en scores z des résultats obtenus par l'approche de Levine et Drasgow (1979) :

$$l_z = \frac{l_0 - E(l_0)}{V(l_0)^{1/2}} \quad (13)$$

où $E(l_0)$ et $V(l_0)$ sont respectivement la moyenne et la variance de l_0 :

$$E(l_0) = \sum_{i=1}^I \{P_i(\theta) \log P_i(\theta) + Q_i(\theta) \log Q_i(\theta)\} \quad (14)$$

et

$$V(l_0) = \sum_{i=1}^I P_i(\theta) Q_i(\theta) \left(\log \frac{P_i(\theta)}{Q_i(\theta)} \right)^2. \quad (15)$$

Selon Drasgow et ses collaborateurs, cette transformation permettrait à l_z de se distribuer selon les quantiles d'une loi $N(0,1)$ lorsque les épreuves d'évaluation présentent 80 items et plus. Par exemple, si un chercheur fixe le seuil de signification α à 0,01, une valeur inférieure au point de coupure -2,33 permettra de considérer que les réponses qu'a obtenues un étudiant à une épreuve d'évaluation sont inappropriées.

Bien que cet indice soit le plus cité dans les recherches en mesure et en évaluation en éducation, son interprétation ne serait pas sans problèmes. En effet,

Molenaar et Hoijtink (1990, 1996), Nering (1995, 1997) et Reise (1995) ont déjà démontré que la fonction de densité de cet indice ne suivrait pas exactement les quantiles d'une loi normale lorsque le niveau d'habileté θ d'un étudiant est estimé à l'aide des méthodes suivantes : maximum de vraisemblance, maximum de vraisemblance *a posteriori* ou maximum de vraisemblance pondérée. De plus, cet indice serait moins efficace dans le cadre de tests contenant un nombre limité d'items (Li et Olejnik, 1997) et il serait plus approprié de l'utiliser lorsque $P_i(\theta)$ est estimé à l'aide du modèle logistique à deux et à trois paramètres (Schmitt, Chan, Sacco, MacFarland et Jennings, 1999).

Enfin, l_z semble présenter une bonne puissance de détection. Par exemple, Drasgow, Levine et McLaughlin (1987) et Raïche (2002) ont procédé à la comparaison de plusieurs indices et ils ont démontré que l_z avait la meilleure puissance de détection. De son côté, Karabatsos (2003) a comparé 36 indices et il n'a pas été en mesure de montrer que l_z était un indice présentant un des plus hauts taux de détection.

2.2.2.2 Les indices de type carré-moyen

Les deux prochains indices sont souvent présentés dans les écrits portant sur le modèle de Rasch et sont inspirés des travaux sur les résidus standardisés de Wright et Panchapakesan (1969).

Wright et Stone (1979) ont développé l'indice U , aussi appelé *outfit mean square*, qui peut être écrit comme suit :

$$U = \frac{1}{I-1} \sum_{i=1}^I \frac{[x_i - P_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}, \quad (16)$$

où $P_i(\theta)$ est une moyenne et $P_i(\theta)Q_i(\theta)$, une variance. Il est important de comprendre que U n'est pas une statistique pondérée, ce qui lui confère une grande sensibilité aux scores extrêmes par rapport au niveau de difficulté de chacun des items ou au niveau d'habileté des personnes.

À ce jour, il n'existe pas de consensus sur la façon dont l'indice devrait être interprété. En effet, Wright et Stone (1979) écrivent que la somme des résidus au carré prend la forme mathématique :

$$\sum_{i=1}^I \frac{[x_i - P_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (17)$$

et devrait suivre les quantiles d'une loi chi-carré avec $I-1$ degré de liberté. Or, d'autres auteurs proposent d'utiliser une autre fonction de densité. Par exemple, Karabatsos (2001, p. 158) suggère d'utiliser la valeur de 1.3 pour U mais, en basant l'interprétation du score de l'indice sur une loi log-normale (Karabatsos, 2003, p. 281). Ainsi, une certaine confusion existe sur la nature de cet indice et sa façon de l'interpréter adéquatement.

Pour pallier ce problème, Karabatsos (2003) transforme U en score z de la façon suivante :

$$ZU = \left(U^{1/3} - 1 \right) \frac{3}{S_U} + \frac{S_U}{3} \quad (18)$$

où

$$S_U = \frac{\sqrt{\sum_{i=1}^I P_i(\theta) Q_i(\theta) (P_i(\theta) - Q_i(\theta))^2}}{\sum_{i=1}^I P_i(\theta) Q_i(\theta)}. \quad (19)$$

Avec cette transformation, ZU devrait suivre les quantiles d'une loi $N(0,1)$.

L'indice W (Wright et Masters, 1982), aussi appelé *infit mean square*, est un indice pondéré qui donne plus de poids aux étudiants dont le score est près du niveau de difficulté de chacun des items ou du niveau d'habileté: c'est pourquoi plusieurs le préfèrent à U . Mathématiquement, W s'écrit de la façon suivante :

$$W = \frac{\sum_{i=1}^I [x_i - P_i(\theta)]^2}{\sum_{i=1}^I P_i(\theta) Q_i(\theta)}. \quad (20)$$

Comme dans le cas de U , il ne semble pas exister de consensus permettant d'interpréter adéquatement cet indice de détection. Par contre, Schulz (2002) propose d'utiliser les quantiles d'une loi chi-carré avec $I-1$ degré de liberté pour fixer un point de coupure. Encore ici, il est possible d'appliquer la même standardisation que pour ZU :

$$ZW = \left(W^{1/3} - 1 \right) \frac{3}{S_W} + \frac{S_W}{3} \quad (21)$$

où S_W prend la même forme qu'à l'équation (19).

ZU et ZW ont été investigués dans quelques études. Ainsi, Drasgow, Levine et McLaughlin (1987), Li et Olejnik (1997) et Noonan, Boss et Gessaroli (1992) ont démontré que ZU et ZW ne suivent pas exactement les quantiles d'une loi $N(0,1)$. Par contre, Rogers et Hattie (1987) ont affirmé que ces indices sont relativement robustes à l'hétérogénéité du paramètre de discrimination a_i et à la multidimensionnalité du niveau d'habileté. Enfin, Al-Mahrazi (2003) a démontré que la puissance de détection des indices ZU et ZW était limitée. Il recommande même d'éviter d'utiliser ces indices puisqu'ils sont très fortement affectés par certaines caractéristiques du test (par exemple, le nombre d'items ou la valeur des paramètres d'item estimés).

L'indice *Zeta* (Tatsuoka, 1996) est un peu plus récent et n'a pas encore été souvent cité dans les écrits en mesure. Par contre, cette approche présente des caractéristiques techniques intéressantes. Mathématiquement, *Zeta* peut s'écrire :

$$Zeta = \frac{\sum_{i=1}^I [P_i(\theta) - x_i][P_i(\theta) - T(\theta)]}{\sum_{i=1}^I \{P_i(\theta)(1 - P_i(\theta))[P_i(\theta) - T(\theta)]^2\}^{1/2}} \quad (22)$$

où

$$T(\theta) = \frac{\sum_{i=1}^I P_i(\theta)}{I} \quad (23)$$

est la moyenne des probabilités d'obtention d'une bonne réponse et I est le nombre d'items au test. Ici, le numérateur est la covariance conditionnelle des vecteurs

$P_i(\theta) - x_i$ et $P_i(\theta) - T(\theta)$. De son côté, le dénominateur est l'écart-type conditionnel du numérateur.

Si peu de recherches ont étudié cet indice, Raïche (2002) a tout de même noté que *Zeta* devrait suivre les quantiles d'une loi normale lorsque le nombre d'items est grand.

2.2.2.3 Les indices de prudence (*caution index*)

Les indices de prudence (*extended caution index*) de Tatsuoka et Linn (1983) s'inspirent directement de l'indice non paramétrique de prudence (*caution index*) de Sato (1975). À l'exclusion de leurs versions standardisées, ces indices *ECI* sont tous interprétables de la même façon : plus leur score est élevé, plus l'évaluateur doit interpréter les réponses d'un étudiant avec prudence.

Le premier indice prend la forme :

$$ECI_1 = 1 - \frac{Cov(x_i, RC)}{Cov(P_i(\theta), RC)} \quad (24)$$

où x_i est le patron de réponses d'un étudiant, $P_i(\theta)$ la probabilité d'obtenir une bonne réponse à un item et RC est le vecteur du nombre total de bonnes réponses à chacun des items de l'épreuve d'évaluation.

Le deuxième indice reprend le même aspect que ECI_1 :

$$ECI_2 = 1 - \frac{Cov(x_i, G)}{Cov(P_i(\theta), G)}, \quad (25)$$

mais Tatsuoka (1984) a remplacé RC par le vecteur des probabilités moyennes d'obtenir une bonne réponse à chacun des items

$G = (\frac{1}{I} \sum_{i=1}^I P_1(\theta_i), \frac{1}{I} \sum_{i=1}^I P_2(\theta_i), \frac{1}{I} \sum_{i=1}^I P_3(\theta_i))$. Cet indice est basé sur le groupe, car les réponses données par un étudiant à chacun des items sont dérivées en fonction du groupe dont il fait partie.

L'indice ECI_3 a la même forme que ECI_2 , mais il utilise la corrélation plutôt que la covariance :

$$ECI_3 = 1 - \frac{Cor(x_i, G)}{Cor(P_i(\theta), G)} \quad (26)$$

Les deux indices suivants focalisent sur l'étudiant puisqu'ils calculent l'ampleur du patron de réponses de ce dernier et la fonction de réponse à une épreuve d'évaluation :

$$ECI_4 = 1 - \frac{Cov(x_i, P_i(\theta))}{Cov(P_i(\theta), G)} \quad (27)$$

L'indice ECI_5 , de son côté, calcule la corrélation existant entre les mêmes éléments qu' ECI_4 :

$$ECI_5 = 1 - \frac{Cor(x_i, P_i(\theta))}{Cor(P_i(\theta), G)} \quad (28)$$

Enfin, ECI_6 s'inspire d'un indice non paramétrique présenté par Tatsuoka et Tatsuoka (1983) et peut s'écrire :

$$ECI_6 = 1 - \frac{Cor(x_i, P_i(\theta))}{Var(P_i(\theta))} \quad (29)$$

La famille d'indices de prudence comporte le désavantage de ne pas suivre les quantiles d'une loi $N(0,1)$. Pour cette raison, Tatsuoka et Linn (1983) ont aussi élaboré des versions standardisées de leurs six indices de prudence. En s'inspirant de la notation de Karabatsos (2003), nous pouvons généraliser l'écriture de ces indices standardisés à:

$$ECI_{z\Omega} = ECI_{\Omega} - \left[\frac{E(ECI_{\Omega})}{V(ECI_{\Omega})} \right] \quad (30)$$

où Ω est égal à l'indice 1, 2, 4 ou 6. Notons qu'il existe, en réalité, deux indices standardisés, car les indices ECI_{z1} et ECI_{z2} sont fortement corrélés (Harnisch et Tatsuoka, 1983; Tatsuoka et Tatsuoka, 1982) et les indices ECI_{z4} et ECI_{z6} sont mathématiquement équivalents. Ainsi, dans chacun des cas, ils peuvent être utilisés de façon interchangeable et leur interprétation suit les quantiles de la loi $N(0,1)$.

De récentes études ont démontré que les indices standardisés de la famille ECI semblent efficaces. Par exemple, St-Onge, Valois, Abdous et Germain (2011) ont démontré que ECI_{z2} était plus robuste que l_z à la détection de patrons de réponses inappropriés et Karabatsos (2003) a démontré que les indices ECI_{z2} et ECI_{z4} sont parmi les plus puissants de la famille des indices de détection paramétriques.

Dragow, Levine et McLaughlin (1987) et Li et Olejnik (1987) ont démontré que ECI_{z4} se rapproche plus d'une loi $N(0,1)$ que ECI_{z2} . Noonan, Boss et Gessaroli (1992) abondent dans le même sens et déclarent, en plus, que ECI_{z4} se rapproche plus d'une loi $N(0,1)$ que l_z et ZW . De plus, cet indice est moins affecté par le nombre d'items administrés et par le modèle de réponse à l'item utilisé.

2.2.2.4 Les indices standardisés retenus dans cette thèse de doctorat

Dans le cadre de cette thèse de doctorat, nous retiendrons uniquement l'indice standardisé de type vraisemblance I_z et les indices standardisés de type carré-moyen ZU et ZW . Au moins deux raisons nous poussent à adopter cette stratégie. Premièrement, l'indice I_z est l'indice le plus connu et le plus étudié de tous. De plus, c'est cet indice qui a été corrigé dans l'article de Snijders (2001), qui nous a grandement inspiré pour ce projet doctoral. Deuxièmement, les indices ZU et ZW bénéficient de quelques écrits scientifiques et ils sont très utilisés par les gens qui travaillent avec le modèle de Rasch. Par contre, Drasgow, Levine et McLaughlin (1987), Li et Olejnik (1997) ainsi que Noonan, Boss et Gessaroli (1992) ont démontré que ces indices ne suivent pas parfaitement leur fonction de densité théorique. Ainsi, une correction doit être appliquée à ces derniers afin de les utiliser convenablement. La section suivante présentera quelques solutions proposées par les chercheurs pour tenir compte du fait que le niveau d'habileté θ réel n'est pas connu.

2.3 Solutions qui ont été proposées pour tenir compte du fait que le niveau d'habileté réel n'est pas connu

Dans les écrits scientifiques sur les indices de détection de patrons de réponses inappropriés, Meijer et Sijtsma (2001) ont soulevé l'importance de chercher de nouvelles alternatives afin de compenser le fait que le niveau d'habileté réel d'un

étudiant n'est pas connu. À ce jour, peu d'auteurs se sont penchés sur ce problème, sauf au regard de l'indice l_z .

Dans le cadre de cette section, nous présenterons certaines propositions d'intérêt aux sections 2.3.1.1 et 2.3.1.2 avant de formellement discuter de la correction proposée par Snijders (section 2.3.1.3) et retenue dans cette thèse. Notons, enfin, qu'il existe aussi une approche inspirée de la logique bayésienne (voir Glas et Meijer, 2003), mais nous ne la retiendrons pas dans cette thèse afin de nous centrer uniquement sur les approches fréquentistes.

2.3.1.1 La stratégie de simulation de Raïche et Blais (2005)

Cette stratégie a comme objectif d'utiliser la simulation informatique afin de vérifier si les points de coupure aux seuils de 0,01 et 0,05, d'un indice de détection, peuvent être bien prédits pour différentes valeurs du niveau d'habileté θ .

Dans leur étude, Raïche et Blais (2005) ont généré de nombreux patrons de réponses pour tous les niveaux d'habileté possibles θ fixés entre -3,00 et 3,00, en procédant par sauts de 0,25. Ensuite, ils ont étudié la fonction de densité de l'indice l_z en fonction du niveau d'habileté estimé et non pas réel. Cela a permis aux auteurs d'obtenir des points de coupure en fonction des quantiles correspondants aux seuils de signification de 0,01 et de 0,05 dans un test d'hypothèse unilatéral.

Bien que cette stratégie semble prometteuse, elle est encore à l'état embryonnaire et de nombreuses extensions doivent être investiguées afin de

considérer son plein potentiel. Elle présente aussi le désavantage de reposer uniquement sur des simulations et non pas sur un calcul algébrique des points de coupure.

2.3.1.2 Le modèle multidimensionnel de Raïche, Magis, Blais et Brochu (2012)

Le modèle de Raïche, Magis, Blais et Brochu (2012) est inspiré de Ferrando (2004), qui a élaboré un indice de fluctuation personnelle (s) à l'aide du modèle de Rasch adapté et s'écrit de la façon suivante :

$$P_i = P(x_i = 1 | \theta, S, C, D, s_i, b_i, c_i, d_i) = (C + c_i) \frac{(D + d_i) - (C + c_i)}{1 + e^{\left\{ \frac{-(\theta - b_i)}{\sqrt{s_i^2 + S^2}} \right\}} \quad (31)$$

où la discrimination de l'item devient :

$$a_i = \frac{1}{\sqrt{s_i^2 + S^2}} \quad (32)$$

θ est un paramètre d'habileté de l'étudiant, b_i un paramètre de difficulté de l'item, c_i un paramètre de pseudo-chance de l'item, d_i un paramètre d'inattention de l'item, S un paramètre de fluctuation personnelle, C un paramètre de pseudo-chance personnelle et D est un paramètre d'inattention.

L'attrait de la modélisation présentée à l'équation (31) est qu'elle semble suffisamment efficace pour corriger de façon appréciable le niveau d'habileté d'un

étudiant présentant un patron de réponses inapproprié. Toutefois, l'utilisation de ce modèle fait en sorte que l'erreur-type de l'estimateur du niveau d'habileté augmente (Raïche, Magis et Blais, 2008), ce qui est prévisible puisqu'on tient maintenant compte de facteurs qui affectent la variabilité du niveau d'habileté à l'intérieur du patron de réponses. Par contre, il reste encore plusieurs études à mener afin de bien comprendre le comportement de ce modèle et de s'assurer de l'estimation des paramètres de personne supplémentaires : un défi très important.

2.3.1.3 La correction de Snijders (2001)

Puisque de nombreux chercheurs ont démontré que l_z ne suit pas une loi normale lorsque le nombre d'items administrés est restreint (Molenaar et Hoijtink, 1990, 1996; Nering, 1995), il est risqué d'utiliser les quantiles de cette fonction de densité pour tenter de classer un patron de réponses comme inapproprié ou non. Afin de répondre à ce problème, Snijders (2001) a développé un indice, nommé l_z^* , qui permettrait de corriger la moyenne et la variance de la fonction de densité de l_z .

D'entrée de jeu, Snijders (2001) démontre que l_z peut être réécrit sous la forme simplifiée suivante :

$$\frac{W_I(\theta)}{V[W_I(\theta)]^{1/2}}, \quad (33)$$

où le numérateur est égal à :

$$W_I(\theta) = \sum_{i=1}^I [x_i - P_i(\theta)] w_i(\theta). \quad (34)$$

Dans l'équation (34) $w_i(\theta)$ est un élément permettant de pondérer les écarts entre la réponse à un item et la probabilité d'obtenir une bonne réponse à cet item : $x_i - P_i(\theta)$. Sachant que x_i est une variable de type Bernoulli, $W_I(\theta)$ présente une moyenne et une variance qui sont respectivement égales à :

$$E[W_I(\theta)] = 0 \quad \text{et} \quad V[W_I(\theta)] = I \sigma_I^2(\theta) \quad (35)$$

où

$$\sigma_I^2(\theta) = \frac{1}{I} \sum_{i=1}^I w_i(\theta)^2 P_i(\theta) Q_i(\theta) \quad (36)$$

(Snijders, 2001, p. 332-335). Ainsi, pour établir la forme que prend l_z à l'équation (34), nous devons fixer la pondération $w_i(\theta)$ à :

$$w_i(\theta) = \log \frac{P_i(\theta)}{Q_i(\theta)}. \quad (37)$$

en déduisant que :

$$l_0 - E(l_0) = W_I(\theta) \quad \text{et que} \quad V(l_0) = I \sigma_I^2(\theta) = V[W_I(\theta)]. \quad (38)$$

Enfin, la correction proposée pour corriger l'indice l_z (Snijders, 2001) peut s'écrire :

$$l_z^* = \frac{l_0(\theta) - E[l_0(\theta)] + c_I(\theta) r_0(\theta)}{\tilde{V}[l_0(\theta)]^{1/2}}, \quad (39)$$

où

$$c_i(\hat{\theta}) = \frac{\sum_{i=1}^I P_i'(\theta) w_i(\theta)}{\sum_{i=1}^I P_i'(\theta) r_i(\theta)}, \quad (40)$$

$r_0(\theta)$ dépend de la méthode d'estimation du niveau d'habileté (par exemple, maximum de vraisemblance ou vraisemblance maximale pondérée), $r_i(\theta)$ dépend du modèle de réponse à l'item utilisé (par exemple, le modèle de Rasch ou le modèle à trois paramètres) et où la variance présente au dénominateur peut s'écrire comme suit :

$$\tilde{V}[L_0(\theta)] = \sum_{i=1}^I \tilde{w}_i(\theta)^2 P_i(\theta) Q_i(\theta) \quad (41)$$

sachant que la pondération modifiée est égale à :

$$\tilde{w}_i(\theta) = w_i(\theta) - c_n(\theta) r_i(\theta). \quad (42)$$

Même si l'article de Snijders a été publié il y a plus de douze ans, très peu d'études ont été conduites de façon approfondie sur cette correction. Par exemple, Van Krimpen-Stoop et Meijer (1998) ont démontré que l_z et l_z^* obtiennent des résultats comparables en contexte de testing adaptatif. De leur côté, Meijer et Sijtsma (2001) et de la Torre et Deng (2008) ont discuté de la pertinence de cette approche pour améliorer la détection de patrons de réponses inappropriés. Enfin, Magis, Raïche et Béland (2011) ont présenté un article didactique permettant de faciliter la compréhension de l'article en plus d'en faire une analyse sur un ensemble de données

dans le domaine des langues. Leurs résultats ont démontré que l_z^* est supérieur à sa version non corrigée, l_z .

2.3.2 Création des indices ZU^* et ZW^*

Comme nous l'avons déjà soulevé un peu plus haut, l'indice l_z peut être réécrit sous la forme suivante :

$$\frac{W_I(\theta)}{V[W_I(\theta)]^{1/2}} \quad (43)$$

À l'aide de quelques manipulations algébriques, il est possible de transposer le format présenté à l'équation (34) aux indices U et W en fixant uniquement le poids $w_i(\theta)$ approprié. Dans le cas de U , ce facteur de poids de l'équation devient :

$$w_i(\theta) = \frac{Q_i(\theta) - P_i(\theta)}{I P_i(\theta) Q_i(\theta)} \quad (44)$$

alors qu'il prend la forme suivante pour l'indice de détection W :

$$w_i(\theta) = \frac{Q_i(\theta) - P_i(\theta)}{\sum_{i=1}^I P_i(\theta) Q_i(\theta)} \quad (45)$$

Ensuite, l'indice ZU^* peut se réécrire sous la même forme que l_z^* (équation 39)

$$ZU^* = \frac{U(\theta) - E[U(\theta)] + c_I(\theta) r_0(\theta)}{\tilde{V}[U(\theta)]^{1/2}} \quad (46)$$

et cette transformation est aussi applicable à l'indice ZW^* :

$$ZW^* = \frac{W(\theta) - E[W(\theta)] + c_I(\theta) r_0(\theta)}{\tilde{V}[W(\theta)]^{1/2}}. \quad (47)$$

Tous les indices corrigés par Snijders (2001) peuvent être interprétés à l'aide des quantiles de la loi $N(0,1)$. Le lecteur intéressé trouvera plus de détails sur une application au modèle de Rasch dans l'article de Magis, Béland et Raïche (2014). Leurs résultats démontrent que les indices corrigés sont généralement plus efficaces à détecter la réponse au hasard et l'inattention que leur version non corrigée.

2.4 Synthèse

Dans le cadre de cette thèse, nous avons décidé de retenir les indices de type vraisemblance et les indices de type carré-moyen. Si plusieurs d'entre eux existent en version standardisée (par exemple, I_z , ZU et ZW), de nombreux auteurs ont démontré que leur fonction de densité réelle ne suit pas leur fonction de densité théorique lorsque le nombre d'items administrés est restreint et que le niveau d'habileté est estimé.

À ce jour, il existe peu de méthodes permettant de considérer le fait que le niveau d'habileté est estimé plutôt que réel. Par contre, la correction de Snijders (2001) semble la plus prometteuse, car elle peut être utilisée avec plusieurs types d'estimateurs du niveau d'habileté et avec tous les modèles de réponse à l'item présentés dans ce chapitre. De plus, de la Torre et Deng (2008), Magis, Raïche et Béland (2011), Meijer et Sijtsma (2001) ont déjà souligné que cette approche était

fort prometteuse. Enfin, la flexibilité de l'approche de Snijders lui confère un autre avantage de taille : la forme de l'équation (34) peut être généralisée à d'autres indices paramétriques tels que les indices de vraisemblance ou les indices de type carré-moyen.

2.5 Objectifs de ce projet doctoral

Plusieurs approches ont déjà été développées dans de précédentes recherches (Zickar et Drasgow, 1996) pour détecter les individus qui répondent de manière inappropriée aux épreuves d'évaluation. Parmi celles-ci, l'utilisation d'indices de détection de patrons de réponses inappropriés est certainement celle qui est la plus étudiée et la plus prometteuse (Meijer et Sijtsma, 2001).

Dans ce projet, nous nous concentrerons sur trois indices standardisés populaires qui présentent des caractéristiques permettant d'en faciliter l'interprétation : I_z , ZU et ZW (la justification est présentée à la section 2.2.2.4). Toutefois, il est reconnu que ces indices sont fortement affectés par le fait que le niveau d'habileté d'un étudiant est estimé et non pas réel. C'est pourquoi Snijders (2001) a proposé une correction dans le calcul de I_z qui tient compte de ce problème d'estimation. Aujourd'hui, nous retiendrons aussi cette approche pour créer deux nouveaux indices: ZU^* et ZW^* .

L'objectif général de ce projet est la mise à l'épreuve des indices paramétriques standardisés I_z , ZU et ZW et de leur version corrigée selon la méthode

de Snijders (2001), soit L_z^* , ZU^* et ZW^* , à l'aide de données simulées par ordinateur. Pour ce faire, nous allons poursuivre trois objectifs spécifiques, qui seront présentés comme trois études distinctes dans la section suivante. Premièrement, nous allons conduire une étude des scores des indices afin de savoir s'ils suivent les quantiles de leur loi théorique : la loi normale. Deuxièmement, nous allons étudier les erreurs de type I empiriques et théoriques des six indices de détection. Troisièmement, nous allons étudier la puissance des indices à détecter la réponse au hasard et l'inattention.

CHAPITRE III

MÉTHODOLOGIE

À l'instar de la majorité des études qui ont porté sur les indices de détection de patrons de réponses inappropriés, nous avons sélectionné la simulation assistée par ordinateur (Landry, 2004) pour investiguer le comportement des indices standardisés I_z , ZU , ZW et de leur version corrigée I_z^* , ZU^* et ZW^* . Notez que les considérations éthiques ne s'appliquent pas dans le cadre de ce projet, puisque nous ne faisons pas intervenir de participants humains.

Ce chapitre sera divisé en quatre grandes sections. Dans la première section, nous discuterons des informations générales portant sur les données simulées. Dans la deuxième section, nous traiterons de l'étude descriptive des scores des indices. Dans la troisième section, nous traiterons de l'étude sur les erreurs de type I empiriques et théoriques. Dans la quatrième section, nous traiterons de l'étude de la puissance des indices.

3.1 Informations générales sur les données simulées

La génération des données simulées (Landry, 2004; Laurencelle, 2001; Lemieux, 2009) s'articulera autour de quatre grands éléments : la longueur du test, le nombre de patrons de réponses générés, le modèle de réponse à l'item utilisé et la méthode d'estimation du niveau d'habileté de l'étudiant. Chacun de ces éléments sera

discuté avant de donner des informations précises sur les trois études présentées dans cette section : l'étude descriptive des scores des six indices étudiés dans cette thèse, l'étude portant sur les erreurs de type I et, enfin, l'étude portant sur la puissance des indices de détection.

3.1.1 La longueur du test

Le survol des écrits scientifiques permet de démontrer qu'il n'y a pas de valeur de référence pour caractériser ce qu'est un test court, un test moyen et un test long. En effet, les auteurs ont choisi différentes longueurs de test qui rendent parfois difficile la comparaison des résultats obtenus par les différents chercheurs. Par exemple, Rudner (1983) a fait ses simulations à l'aide d'un test de 45 items et d'un test de 80 items alors que Smith (1985) a fait ses analyses sur des échantillons de 50 items. Meijer (1994a) propose de tester deux longueurs de test (17 items et 33 items) tout comme Emons, Glas, Meijer et Sijtsma (2003) (20 et 40 items). Glas et Meijer (2003) et Hendrawan, Glas et Meijer (2005) ont produit des simulations sur des tests de 30 et 60 items.

Karabatsos (2003) a montré que l'efficacité des indices de détection change selon la longueur du test. En nous inspirant des recherches citées un peu plus haut, nous générerons deux longueurs de test : un test court de 30 items et un test long de 80 items. Mentionnons, enfin, que ces longueurs de tests sont fréquentes dans la pratique. Par exemple, le test de classement en anglais- langue seconde (TCALS-II),

au niveau collégial francophone au Québec, comporte 85 items (Laurier, Froio, Pearo et Fournier, 1998).

3.1.2 Le nombre de patrons de réponses généré

La majorité des chercheurs se sont affairés à générer différentes tailles d'échantillons de patrons de réponses. Ainsi, Drasgow (1982) utilise deux tailles d'échantillons de patrons de réponses : un petit échantillon (500 patrons) et un grand échantillon (3000 patrons). Raïche (2002), de son côté, simule 1000 patrons de réponses différents alors que Glas et Meijer (2003) utilisent trois tailles : 100, 400 et 1000 patrons. Enfin, Hendrawan, Glas et Meijer (2005) utilisent 400 et 1000 patrons alors que St-Onge (2007) simule 100 et 1000 patrons dans un de ses chapitres de thèse de doctorat.

Bien sûr, plus il y a de patrons de réponse simulés, plus les résultats obtenus sont précis et l'erreur-type faible. Dans le cadre de nos trois études, nous allons générer un grand nombre de patrons de réponses, soit 15 000 par niveau d'habileté. Cette stratégie se justifie aisément puisque nous souhaitons avoir suffisamment de données pour procéder à des analyses au regard des quantiles associés à de très petites probabilités, soit des erreurs de type I. Par exemple, avec 15 000 observations, le premier centile correspondrait à 150 patrons de réponses, ce qui est suffisant pour effectuer le calcul de diverses statistiques descriptives.

3.1.3 Les modèles de réponse à l'item utilisés

Nous analyserons les données à l'aide des trois modèles issus de la théorie de la réponse aux items présentés aux sections 2.2.1 à 2.2.3. Rappelons que dans le cas du modèle à trois paramètres, la probabilité qu'un étudiant obtienne une bonne réponse à un item est calculée comme suit :

$$P_i(\theta) = P(x_i = 1 | \theta, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]}.$$

Dans ce modèle, θ est un paramètre d'habileté de l'étudiant, a_i un paramètre de discrimination de l'item, b_i un paramètre de difficulté de l'item et c_i un paramètre de pseudo-chance de l'item. Il est à noter que le modèle à trois paramètres se réduit au modèle à deux paramètres lorsque $c_i = 0$ et au modèle à un paramètre (ou modèle de Rasch) lorsque $c_i = 0$ et $a_i = 1$.

Voici comment nous procéderons pour générer les paramètres des différents modèles de réponse à l'item. Premièrement, nous nous inspirerons de certains éléments de la méthodologie de Van Krimpen-Stoop et Meijer (1999) en générant cinq niveaux de θ allant de moins deux à deux :

$$\theta = -2; -1; 0; 1; 2.$$

Cela nous permettra d'avoir suffisamment de données pour chacun des cinq niveaux de θ estimés, tout spécialement pour les θ les plus extrêmes (par exemple, -2 et 2). Notons que l'utilisation d'une loi $N(0,1)$ ne nous aurait pas permis d'obtenir suffisamment de données pour les niveaux $\theta = -2$ et $\theta = 2$. De plus, cette stratégie

permettra d'étudier le comportement des indices de détection à différentes valeurs de θ .

Deuxièmement, le paramètre de difficulté b_i sera simulé à l'aide de la loi normale $N(0,1)$. Cette stratégie est fréquente et elle a été utilisée par des auteurs tels que Tendeiro et Meijer (2012) ainsi que par Li et Olejnik (1997).

Troisièmement, le paramètre de discrimination a_i sera généré à l'aide de la loi log-normale. Rappelons que cette loi permet de générer des données allant de zéro à l'infini, soit les limites de ce paramètre. De plus, cette loi a aussi été utilisée dans la thèse de Al-Mahrazi (2003) et est décrite par Baker et Kim (2004, p. 186-187) dans leur ouvrage de référence.

Quatrièmement, le paramètre de pseudo-chance c_i sera simulé en utilisant une loi beta (voir Baker et Kim (2004), p. 188-189) qui est bornée entre zéro et un. Bien que ce choix soit moins fréquent, nous utiliserons cette fonction de densité avec une moyenne de 0,25, en assumant une situation où les items comportent quatre choix de réponse. Ce résultat peut être obtenu en générant les données de la façon suivante : beta ($\alpha=6$, $\beta=18$).

Le tableau suivant synthétise les fonctions de densité utilisées pour générer les paramètres de personnes et d'items.

Tableau 3.1 Synthèse des fonctions de densité permettant de générer les paramètres de personnes et d'items

Paramètre	Fonction de densité
θ	aucune puisque nous utiliserons 5 valeurs fixes
a_i	log-normale ($\mu=0, \sigma^2=0,5$)
b_i	normale $N(\mu=0, \sigma^2=1)$
c_i	beta ($\alpha=6, \theta=18$)

Enfin, notons que toutes les données seront générées à l'aide du logiciel R.

3.1.4 Méthode de génération des patrons de réponses appropriés

Pour les deux premières études uniquement, la génération des patrons de réponses appropriés sera réalisée à l'aide de la fonction `sim`, contenue dans la librairie `irtoys` (Partchev, 2012) et disponible via le logiciel R. Il est à noter que les matrices analysées ne contiendront aucune donnée manquante. De plus, nous simulerons des données unidimensionnelles et contenant des items localement indépendants afin de ne pas augmenter inutilement la section présentant les résultats (ces analyses seront plutôt proposées comme piste de recherche à conduire dans le futur).

3.1.5 Les méthodes d'estimation du niveau d'habileté

Un seul estimateur sera retenu dans le cadre de cette thèse : la vraisemblance maximale pondérée. Ce choix est justifié par le fait que cet estimateur est

techniquement celui qui est le moins biaisé de ceux présentés au chapitre précédent (Warm, 1989). De plus, les calculs associés aux indices auxquels est appliquée la correction de Snijders (2001) reposent sur le même ajustement du biais que celui utilisé dans le calcul de cet estimateur.

3.2 Étude 1 : analyse descriptive des scores des indices

Nous étudierons le comportement des indices standardisés I_z , ZU , ZW et de leur version corrigée I_z^* , ZU^* et ZW^* en utilisant deux longueurs de test, une seule taille d'échantillon, trois modèles de réponse à l'item et la méthode d'estimation de θ par vraisemblance maximale pondérée. Le tableau suivant propose une synthèse du plan d'analyse.

Tableau 3.2 Synthèse des manipulations pour l'étude descriptive des scores des indices de détection (estimation selon la méthode de vraisemblance maximale pondérée)

Modélisation	Nombre d'items	Niveau d'habileté
Rasch	30, 80	-2, -1, 0, 1, 2
2 paramètres	30, 80	-2, -1, 0, 1, 2
3 paramètres	30, 80	-2, -1, 0, 1, 2

15 000 patrons de réponses simulés par niveau d'habileté

Dans un premier temps, nous comparerons la fonction de densité de la loi $N(0,1)$ à la fonction de densité des scores des indices calculés selon le modèle de Rasch, le modèle à deux paramètres et le modèle à trois paramètres. Dans un deuxième temps,

les coefficients de corrélation de Pearson entre les six indices seront calculés. Quatre statistiques descriptives (la moyenne, l'écart-type, le coefficient d'asymétrie et le coefficient de kurtose) seront ajoutées afin de vérifier si la fonction de densité de l'indice est en adéquation avec la loi $N(0,1)$. Rappelons que le coefficient d'asymétrie est égal à zéro pour une loi normale. De son côté, le coefficient de kurtose est aussi égal à zéro pour une fonction de densité parfaitement symétrique et contenant un seul mode comme la $N(0,1)$.

3.3 Étude 2 : analyse des erreurs de type I empiriques et théoriques

Comme nous l'avons déjà soulevé dans le chapitre précédent, une erreur de type I est observée lorsque des étudiants ayant répondu de manière appropriée à une épreuve d'évaluation sont faussement identifiés comme ayant présenté un patron de réponses inapproprié. Dans le cadre de cette thèse, nous étudierons plus en détail cet élément en comparant les erreurs de type I empiriques et les erreurs de type I théoriques.

D'abord, l'erreur de type I théorique est calculée à partir des seuils α de la loi normale $N(0,1)$. Ensuite, l'erreur de type I empirique est calculée à partir des seuils α de la fonction de densité des scores des indices obtenus par simulation informatique. Le tableau suivant donne les informations permettant d'interpréter les différents indices à l'étude.

Tableau 3.3 Indices de détection et interprétation

Indice	Règle	Référence
ZU	$ZU \geq q_{z/2}(1-\alpha)$	Wright et Stone (1979), Karabatsos (2001)
ZW	$ZW \geq q_{z/2}(1-\alpha)$	Wright et Masters (1982), Karabatsos (2001)
Lz	$Lz \leq q_z(\alpha)$	Drasgow, Levine et Williams (1985)
U^*	$U^* \geq q_z(1-\alpha)$	Inspiré de Snijders (2001)
W^*	$W^* \geq q_z(1-\alpha)$	Inspiré de Snijders (2001)
Lz^*	$Lz^* \leq q_z(\alpha)$	Snijders (2001)

q = quantile des fonctions de densité de z

Il est reconnu que les indices qui ne sont pas corrigés sont hautement conservateurs, car leur erreur de type I théorique est généralement inférieure à leur erreur de type I empirique. Cette étude permettra de vérifier si les quantiles théoriques et les quantiles empiriques impliqués sont effectivement les mêmes. Pour ce faire, nous utiliserons des valeurs α variant entre 0,005 et 0,10, par sauts de 0,01.

Les analyses seront produites pour deux longueurs de tests, une seule taille d'échantillon, trois modèles de réponse à l'item, cinq valeurs de θ et la méthode d'estimation de θ par vraisemblance maximale pondérée. Le tableau qui suit présente une synthèse des conditions de génération des données.

Tableau 3.4 Synthèse des manipulations pour l'étude portant sur les erreurs de type I empiriques et théoriques (estimation selon la méthode de vraisemblance maximale pondérée)

Modélisation	Items	Niveau d'habileté	Seuils α
Rasch	30, 80	-2, -1, 0, 1, 2	[0,005-0,10]
2 paramètres	30, 80	-2, -1, 0, 1, 2	[0,005-0,10]
3 paramètres	30, 80	-2, -1, 0, 1, 2	[0,005-0,10]

15 000 patrons de réponses simulés par niveau d'habileté

Enfin, les erreurs de type I empiriques et les erreurs de type I théoriques seront calculées pour les indices lz , ZU , ZW , lz^* , ZU^* et ZW^* et comparées à l'aide d'une approche graphique.

3.4 Étude 3 : analyse de la puissance des indices

Cette dernière étude propose d'analyser si les six indices permettent de détecter adéquatement deux types de patrons de réponses inappropriés : la réponse au hasard et l'inattention.

3.4.1 Méthode de génération de patrons de réponses inappropriés

D'entrée de jeu, il est important de comprendre que la conception de la puissance qui est utilisée dans cette thèse est légèrement différente de celle de puissance statistique : il s'agit ici plutôt de la puissance à détecter des patrons de

réponses inappropriés. Nous allons utiliser la modélisation développée par Raïche, Magis, Blais et Brochu (2012) et qui est disponible dans la librairie *irtProb* (Raïche, 2012) afin de générer des patrons de réponses inappropriés. Notons qu'il existe d'autres méthodes telles que celle de Levine et Drasgow (1982), mais elles n'ont pas été retenues à cause de leur caractère artificiel.

Mathématiquement, le modèle de Raïche et de ses collaborateurs (2012) correspond à :

$$P_i = P(x_i = 1 | \theta, S, C, D, s_i, b_i, c_i, d_i) = (C + c_i) \frac{(D + d_i) - (C + c_i) \left\{ \frac{-(\theta - b_i)}{\sqrt{s_i^2 + S^2}} \right\}}{1 + e^{\left\{ \frac{-(\theta - b_i)}{\sqrt{s_i^2 + S^2}} \right\}}}$$

où la discrimination de l'item devient :

$$a_i = \frac{1}{\sqrt{s_i^2 + S^2}}$$

θ est un paramètre d'habileté de l'étudiant, b_i un paramètre de difficulté de l'item, c_i un paramètre de pseudo-chance de l'item, d_i un paramètre d'inattention de l'item, S un paramètre de fluctuation personnelle, C un paramètre de pseudo-chance personnelle et D est un paramètre d'inattention. Les deux derniers paramètres peuvent être utilisés pour générer des réponses au hasard et des réponses inattentives. Dans le cadre de cette thèse, nous allons les utiliser pour générer trois niveaux de pseudo-chance :

$$C=0,1, 0,2 \text{ et } 0,3$$

ainsi que trois niveaux d'inattention :

$$D=0,1, 0,2 \text{ et } 0,3.$$

Les autres paramètres qui se trouvent dans le modèle de Raïche et de ses collaborateurs (2012) seront générés exactement comme dans les deux précédentes études, sauf en ce qui concerne le paramètre S , qui sera fixé à zéro.

Tableau 3.5 Synthèse des fonctions de densité permettant de générer les paramètres de personnes et d'items pour l'étude 3

Paramètre	Fonction de densité
θ	aucune puisque nous utiliserons 5 valeurs fixes
a_i	log-normale ($\mu=0, \sigma^2=0,5$)
b_i	normale $N(\mu=0, \sigma^2=1)$
c_i	beta ($\alpha=6, \theta=18$)
S	aucune puisque ce paramètre est fixé à 0

Cela nous permettra ensuite d'observer dans quelle mesure les six indices à l'étude peuvent bien détecter ces deux types de patrons de réponses inappropriés.

Nous rapporterons le pourcentage de détection de ces réponses inappropriées pour deux seuils de signification, une longueur de test, trois différents modèles de réponse à l'item, la méthode d'estimation de θ par vraisemblance maximale pondérée, cinq valeurs de θ , trois niveaux de pseudo-chance et trois niveaux d'inattention. La synthèse des conditions relatives à cette dernière étude est présentée un peu plus bas.

Tableau 3.6 Synthèse des manipulations pour l'étude portant sur la puissance de détection des indices (estimation selon la méthode de vraisemblance maximale pondérée)

Modèle	Items	Niveaux d'habileté	Seuils α	Niveaux de réponse au hasard C	Niveaux d'inattention D
Rasch	30, 80	-2, -1, 0, 1, 2	0,01; 0,05	0,1; 0,2 et 0,3	0,1; 0,2 et 0,3
2 paramètres	30, 80	-2, -1, 0, 1, 2	0,01; 0,05	0,1; 0,2 et 0,3	0,1; 0,2 et 0,3
3 paramètres	30, 80	-2, -1, 0, 1, 2	0,01; 0,05	0,1; 0,2 et 0,3	0,1; 0,2 et 0,3

15 000 patrons de réponses simulés par niveau d'habileté

Les statistiques descriptives suivantes seront ajoutées : la moyenne des scores des indices, leur écart-type, leur coefficient d'asymétrie et leur coefficient de kurtose. Cela permettra de vérifier si les scores des indices suivent leur loi théorique (la loi $N(0,1)$) et d'obtenir des informations sur la pertinence de nos résultats.

CHAPITRE IV

RÉSULTATS

Les résultats sont présentés dans ce chapitre. Puisque cette étude est surtout exploratoire, nous favoriserons surtout une approche descriptive pour présenter nos résultats.

Le contenu qui suit est divisé en trois sections : l'étude descriptive du score des six indices étudiés dans cette thèse, l'étude de leurs erreurs de type I empiriques et théoriques et l'étude de leur puissance à détecter la réponse au hasard et l'inattention.

4.1 Étude 1 : analyse descriptive des scores des indices

Nous présentons, dans cette première section, deux analyses. D'abord, les résultats d'une étude descriptive de la fonction de densité des scores des indices l_z , ZU , ZW , l_z^* , ZU^* et ZW^* . Ensuite, les tableaux contenant des corrélations de Pearson et des statistiques descriptives calculées à partir des scores des indices de détection investigués dans ce projet.

4.1.1.1 Fonction de densité des scores des indices : résultats pour 30 items

Dans toutes les figures de la section 4.1.1.1 et 4.1.1.2, la valeur θ au-dessus de chacun des quadrants d'une figure est la valeur réelle de ce paramètre. L'axe horizontal de chacun des quadrants d'une figure représente les valeurs estimées de θ obtenues par la simulation informatique. Enfin, l'axe vertical de chacun des quadrants d'une figure est la fonction de densité empirique calculée à partir de la méthode par lissage non paramétrique selon un noyau gaussien.

Trois constats peuvent se dégager de l'analyse des scores de I_z (figure 4.1). Premièrement, la fonction de densité de cet indice se rapproche de la fonction de densité de la loi normale $N(0,1)$ lorsque θ tend vers zéro. Toutefois, l'adéquation entre la fonction de densité de I_z et celle de la loi $N(0,1)$ est beaucoup moins importante lorsque $\theta = -2$ et $\theta = 2$.

Deuxièmement, la fonction de densité du modèle à trois paramètres est celle qui est la plus près de la fonction de densité de la loi $N(0,1)$ lorsque θ est inférieur à zéro. Lorsque $\theta \geq 0$, nous remarquons que c'est la fonction de densité obtenue à l'aide du modèle de Rasch qui présente la meilleure concordance à la fonction de densité de la loi $N(0,1)$.

Pour toutes les valeurs de θ , la fonction de densité de I_z calculée à l'aide du modèle à deux paramètres n'est jamais celle qui se rapproche le plus de la loi $N(0,1)$. Mentionnons toutefois que la fonction de densité obtenue à l'aide du modèle de

Rasch et la fonction de densité obtenue à l'aide du modèle à deux paramètres se rapprochent grandement aux valeurs $\theta = -1$, $\theta = 0$ et $\theta = 1$.

Troisièmement, les fonctions de densité obtenues à l'aide de tous les modèles de réponse à l'item sont légèrement à la droite de la fonction de densité de la loi $N(0,1)$. Cela montre que les scores de I_z présentent une asymétrie négative.

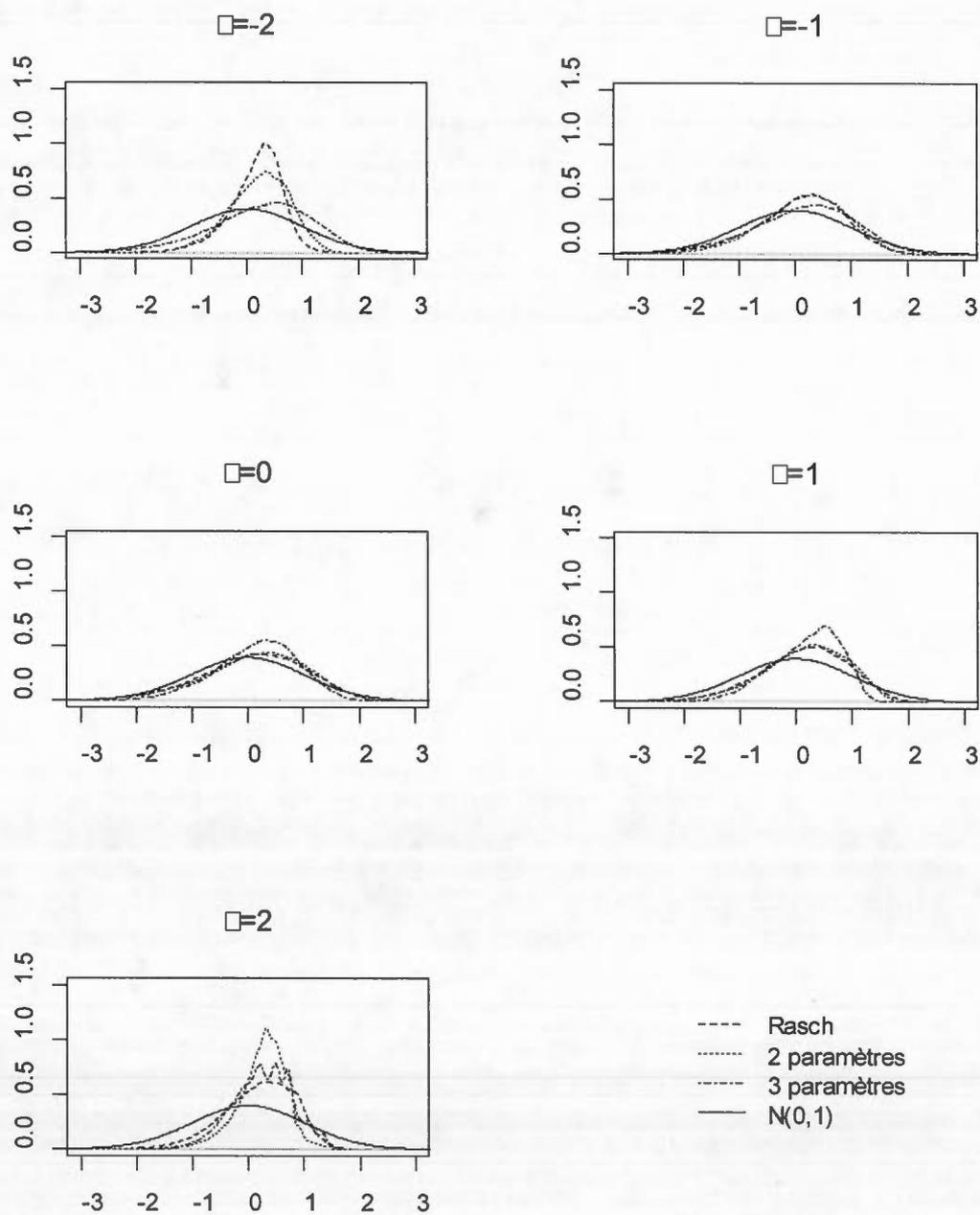


Figure 4.1 Scores de l'indice I_z selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items

Observons les résultats de I_z^* , à la figure 4.2. L'ajustement des scores de cet indice à la loi $N(0,1)$ est généralement bon puisque nous retrouvons très peu de différences entre les courbes de fonction de densité obtenues à l'aide des différents modèles de réponse à l'item pour $\theta \leq 1$. Il n'y a donc pas d'avantage important à utiliser un modèle de réponse à l'item plutôt qu'un autre à ces valeurs de θ . Par contre, l'adéquation entre les courbes de fonction de densité des différents modèles de réponse à l'item et celle de la loi $N(0,1)$ est moins bonne pour $\theta = 2$. Nous observons, en effet, qu'à cette valeur de θ , la fonction de densité du modèle à deux paramètres est légèrement plus en accord avec celle de la loi $N(0,1)$. De plus, les fonctions de densité sont beaucoup plus fluctuantes lorsque $\theta = 2$. Cela est particulièrement le cas lorsque les données sont analysées à l'aide du modèle à trois paramètres.

À l'instar des résultats présentés à la figure 4.1, nous observons que les fonctions de densité sont légèrement à la droite de la fonction de densité de la loi $N(0,1)$. Cela montre que les scores de I_z^* ont une asymétrie négative et que la moyenne des scores de cet indice est inférieure au mode et à la médiane.

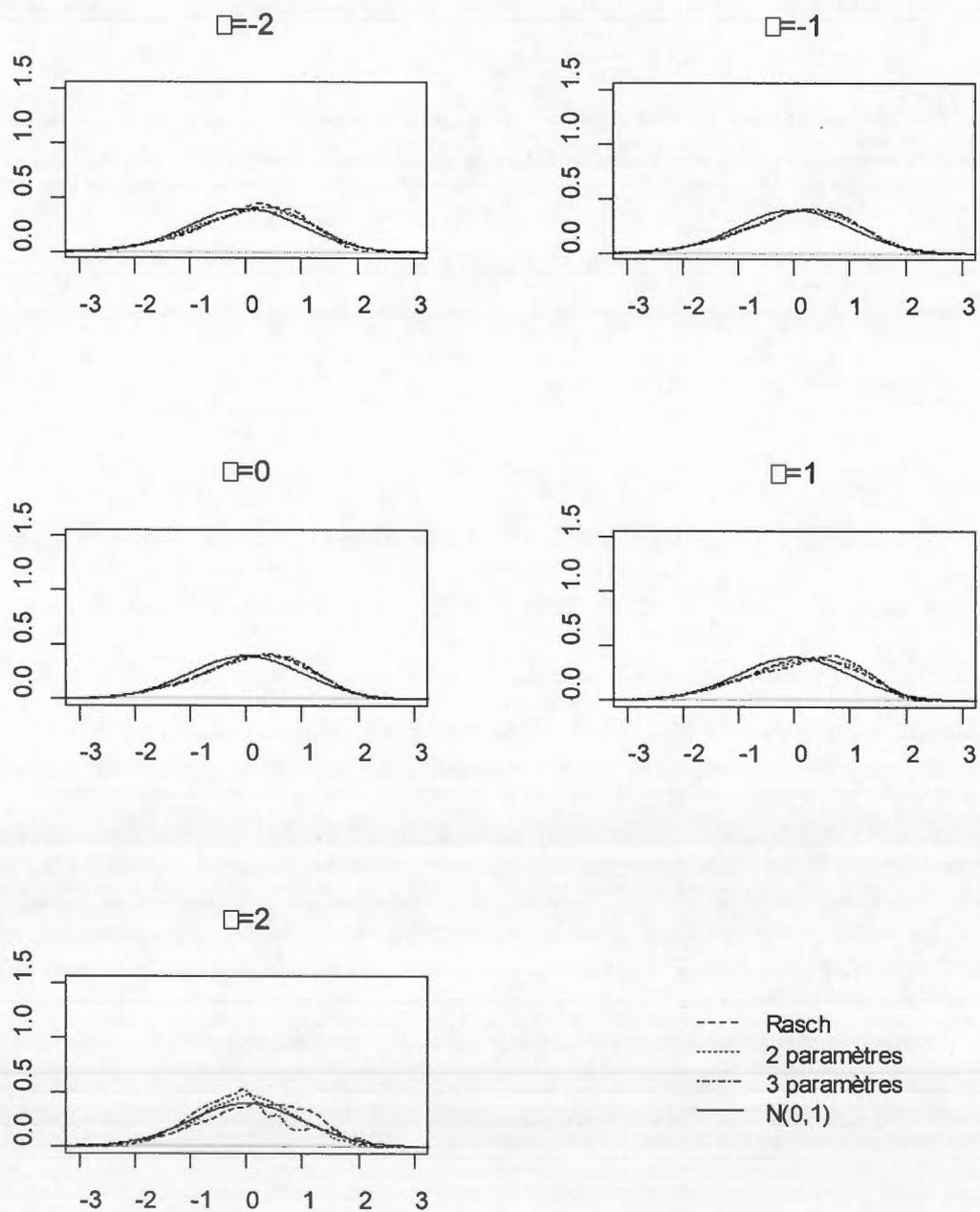


Figure 4.2: Scores de l'indice I_z^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items

Il y a de nombreux commentaires à soulever au sujet de l'indice ZU , qui est présenté à la figure 4.3. Pour la valeur $\theta = -2$, nous observons que la fonction de densité de ZU présente la meilleure adéquation à la loi $N(0,1)$ lorsque nous utilisons le modèle à trois paramètres. Ensuite, les fonctions de densité du modèle de Rasch et du modèle à trois paramètres présentent une asymétrie positive. À l'opposé, la fonction de densité du modèle à deux paramètres présente une asymétrie négative.

Pour la valeur $\theta = -1$, c'est la fonction de densité du modèle à trois paramètres qui présente la meilleure adéquation à la fonction de densité de la loi $N(0,1)$. De plus, les fonctions de densité du modèle de Rasch et du modèle à trois paramètres ont une asymétrie positive, alors que la fonction de densité du modèle à deux paramètres présente une asymétrie négative.

L'écart entre les fonctions de densité des trois modèles de réponse à l'item est moins important à la valeur $\theta = 0$. Dans ce cas-ci, nous observons que c'est plutôt la fonction de densité du modèle de Rasch qui présente la meilleure adéquation à la fonction de densité de la loi $N(0,1)$. De plus, les fonctions de densité du modèle de Rasch et du modèle à deux paramètres présentent une asymétrie positive. De son côté, la fonction de densité du modèle à trois paramètres a une asymétrie négative.

En ce qui concerne la valeur $\theta = 1$, c'est le modèle de Rasch qui présente la fonction de densité la plus près de la fonction de densité de la loi $N(0,1)$. Ensuite, la fonction de densité du modèle à deux paramètres est très éloignée de la $N(0,1)$. De

plus, la largeur de la fonction de densité de ce modèle est beaucoup plus grande que celle de la loi $N(0,1)$. Enfin, la fonction de densité du modèle à trois paramètres est aussi très large et éloignée de la $N(0,1)$. Ces résultats démontrent que ZU a une plus grande propension à offrir des scores élevés avec ces modèles de réponse à l'item, à $\theta = 1$.

Pour la valeur $\theta = 2$, c'est le modèle de Rasch qui présente la fonction de densité la plus près de la fonction de densité de la loi $N(0,1)$. De son côté, le modèle à trois paramètres présente une très mauvaise adéquation à la loi $N(0,1)$. Cela est lié au fait que ZU présente une très grande étendue de scores à cette valeur de θ . Enfin, la fonction de densité du modèle de Rasch présente une asymétrie positive, alors que la fonction de densité du modèle à deux paramètres présente une asymétrie négative.

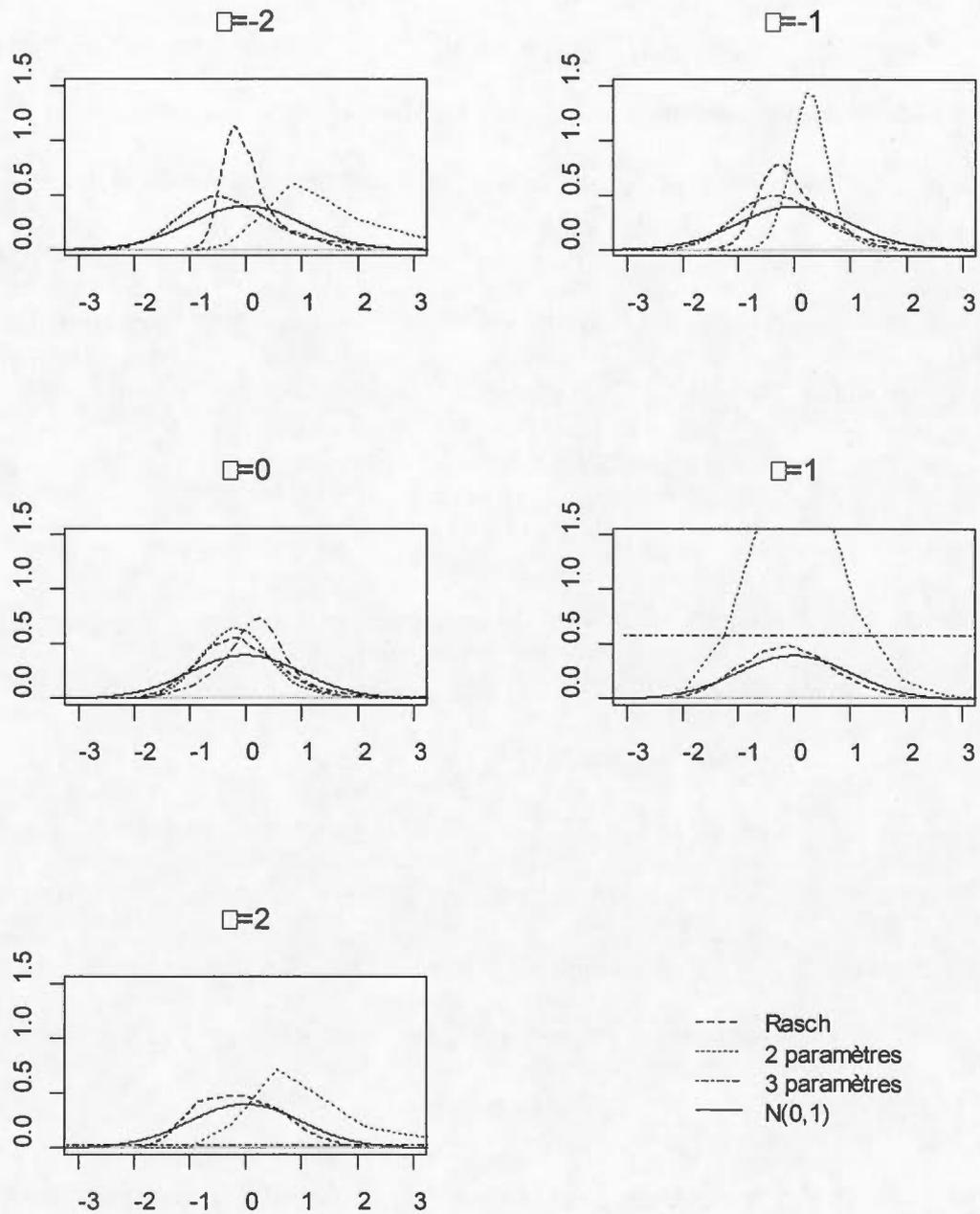


Figure 4.3: Scores de l'indice ZU selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items

Tout comme dans le cas de ZU , il y a beaucoup d'éléments à décrire pour l'indice ZU^* (figure 4.4). Pour la valeur $\theta = -2$, c'est la fonction de densité pour le modèle à trois paramètres qui présente la meilleure adéquation à la loi $N(0,1)$. Ensuite, la fonction de densité du modèle à deux paramètres est celle qui est la plus éloignée de la loi $N(0,1)$, en plus de présenter une étendue de scores plus réduite que celle des autres modèles. Cela peut d'ailleurs créer des problèmes d'interprétation de ZU . Enfin, les fonctions de densité du modèle de Rasch et du modèle à trois paramètres ont une asymétrie positive. À l'opposé, la fonction de densité du modèle à deux paramètres a une asymétrie négative.

Les résultats pour $\theta = -1$ se comparent à ce qui a été présenté dans le paragraphe précédent. C'est la fonction de densité du modèle à trois paramètres qui présente la meilleure adéquation à la fonction de densité de la loi $N(0,1)$. Encore une fois, la fonction de densité du modèle à deux paramètres est problématique, car elle est très étroite en plus d'être celle qui est la plus éloignée de la loi $N(0,1)$. Enfin, les fonctions de densité du modèle de Rasch et du modèle à trois paramètres présentent une asymétrie positive. De son côté, la fonction de densité du modèle à deux paramètres présente une asymétrie négative.

Voici les résultats pour la valeur $\theta = 0$. Dans ce cas-ci, nous observons que c'est la fonction de densité du modèle de Rasch qui présente la meilleure adéquation à la fonction de densité de la loi $N(0,1)$. Ensuite, la fonction de densité du modèle à trois paramètres est celle qui est la plus éloignée de la loi $N(0,1)$ en plus de présenter

une variance plus petite que celle des autres modèles de réponse à l'item. Cela engendre d'ailleurs un problème d'utilisation pour la valeur $\theta = 0$.

Les fonctions de densité du modèle de Rasch et du modèle à deux paramètres ont des asymétries positives, alors que la fonction de densité du modèle à trois paramètres a une asymétrie négative.

Pour $\theta = 1$, c'est le modèle de Rasch qui présente la fonction de densité la plus près de la fonction de densité de la loi $N(0,1)$. Ensuite, la fonction de densité du modèle à trois paramètres est très éloignée de la $N(0,1)$, en plus de présenter une fonction de densité des scores de ZU^* très homogène et qui s'éloigne des valeurs théoriques de la loi $N(0,1)$. Enfin, les fonctions de densité du modèle de Rasch et du modèle à deux paramètres sont plus à gauche de la fonction de densité de la $N(0,1)$, alors que la fonction de densité du modèle à trois paramètres est plus étroite et légèrement à droite de celle de la $N(0,1)$.

En ce qui concerne la valeur $\theta = 2$, c'est le modèle de Rasch qui présente la fonction de densité la plus près de la fonction de densité de la loi $N(0,1)$. Les modèles à deux et à trois paramètres présentent une très mauvaise adéquation à la loi $N(0,1)$. De plus, leurs fonctions de densité sont très étroites, ce qui témoigne de la dispersion limitée des scores de ZU^* pour ces modèles de réponse à l'item. Pour cette valeur de θ , c'est le modèle de Rasch qui est le plus recommandé pour bien interpréter cet indice.

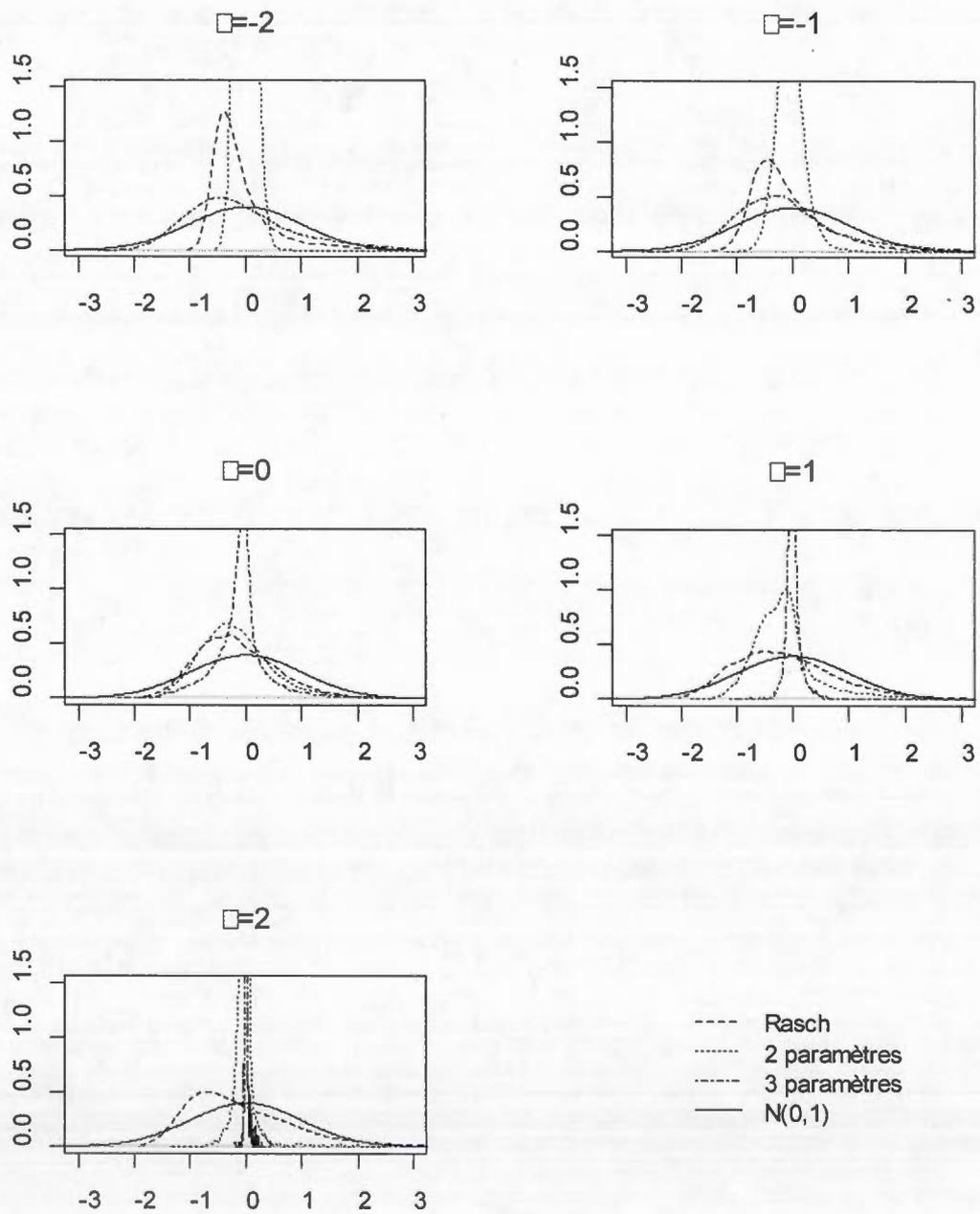


Figure 4.4: Scores de l'indice ZU^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items

Selon les résultats présentés à la figure 4.5, les fonctions de densité de ZW se rapprochent de leur fonction de densité théorique, la loi $N(0,1)$, lorsque θ tend vers zéro. Nous remarquons que le modèle à trois paramètres présente généralement des scores plus près de la $N(0,1)$ lorsque θ est inférieur à zéro. Par contre, aux valeurs $\theta = 0$ à $\theta = 2$, c'est le modèle de Rasch qui présente la meilleure adéquation à la $N(0,1)$.

La fonction de densité obtenue à l'aide du modèle à deux paramètres n'est jamais celle que se rapproche le plus de la loi $N(0,1)$. Par contre, la fonction de densité de cette modélisation se rapproche grandement de la fonction de densité du modèle de Rasch aux valeurs $\theta = -1$, $\theta = 0$ et $\theta = 1$.

Il est important de mentionner qu'à la valeur $\theta = 2$, les fonctions de densité sont beaucoup plus fluctuantes que pour les autres valeurs θ . En effet, pour cette valeur, les courbes sont beaucoup moins lisses.

Enfin, les fonctions de densité obtenue à l'aide de tous les modèles de réponse à l'item présentent une asymétrie légèrement positive. Cela montre que les scores de ZW présentent une moyenne supérieure au mode et à la médiane.

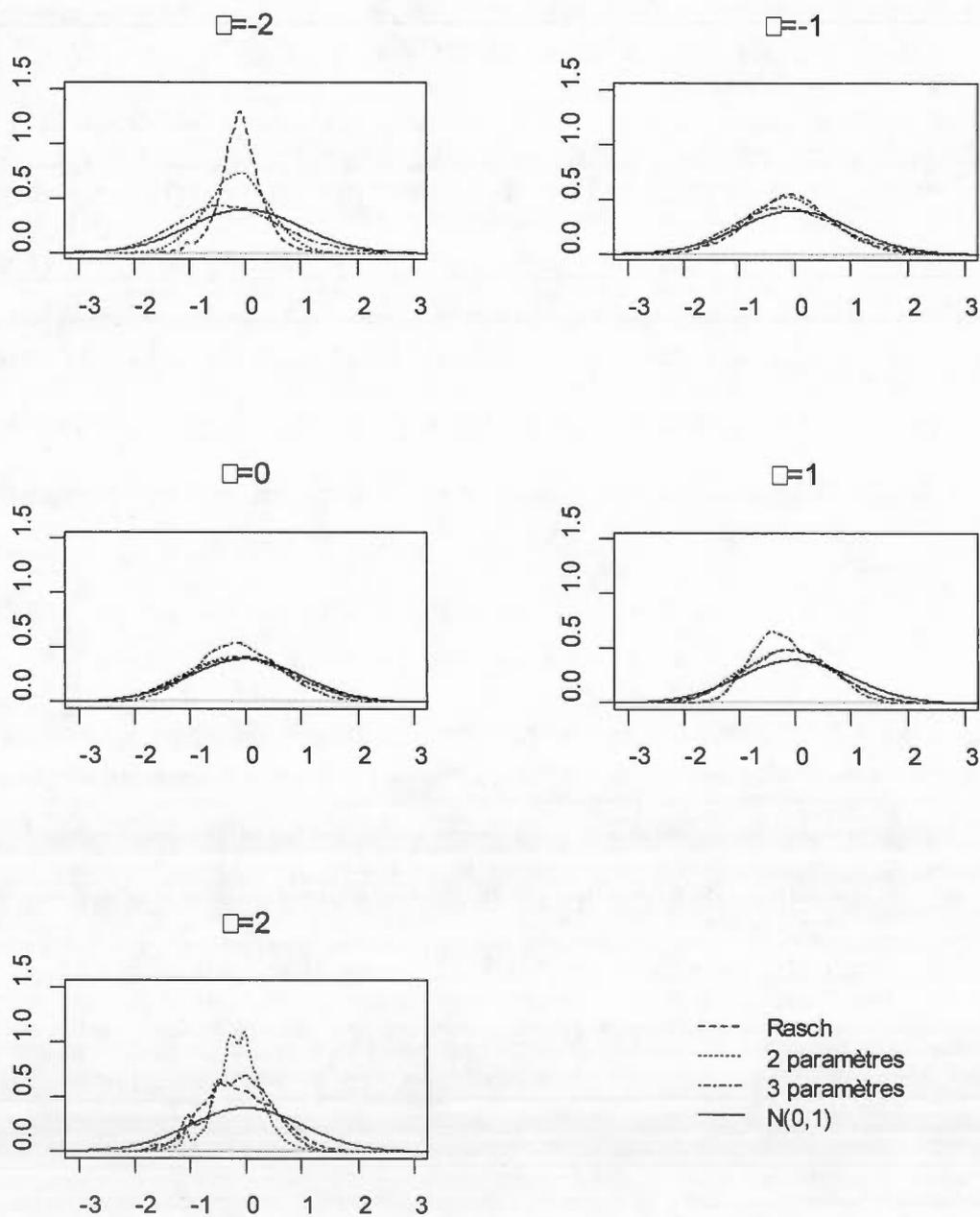


Figure 4.5: Scores de l'indice ZW selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items

À l'instar de I_z^* , la figure 4.6 montre que les scores de ZW^* sont généralement en accord avec la loi $N(0,1)$ pour les valeurs $\theta = -2$ à $\theta = 0$. En effet, nous retrouvons très peu de différences entre les fonctions de densité obtenues à l'aide des différents modèles de réponse à l'item pour $\theta \leq 0$. Par contre, l'adéquation entre les fonctions de densité des différents modèles de réponse à l'item et celle de la loi $N(0,1)$ est moins bonne pour les valeurs $\theta = 1$ et $\theta = 2$. Nous observons que pour ces valeurs de θ la fonction de densité du modèle à deux paramètres est légèrement plus en accord avec celle de la loi $N(0,1)$. De plus, les fonctions de densité sont beaucoup plus fluctuantes lorsque $\theta = 2$.

Enfin, nous observons que les fonctions de densité ont une asymétrie positive.

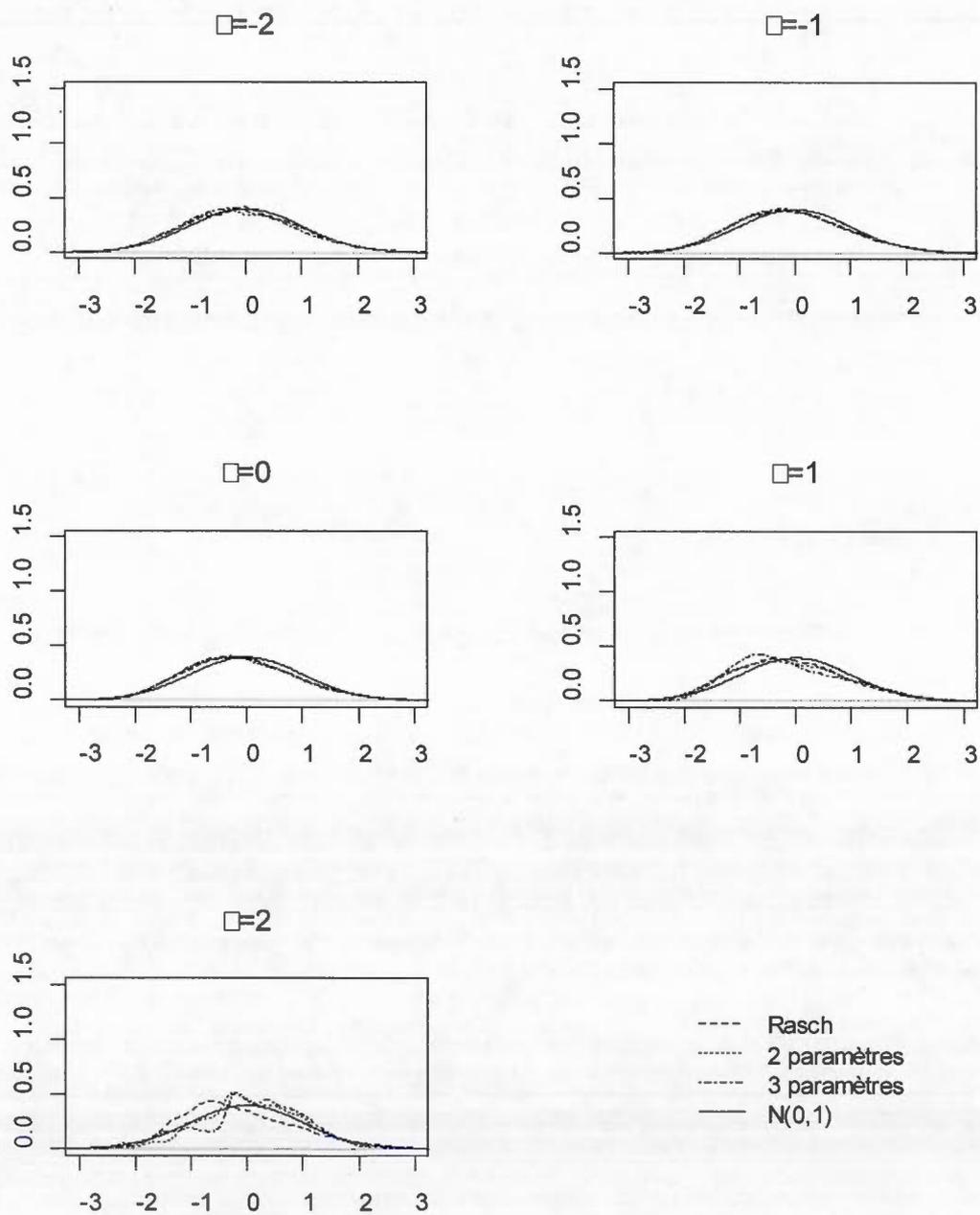


Figure 4.6: Scores de l'indice ZW^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 30 items

4.1.1.2 Fonction de densité des scores des indices : résultats pour 80 items

Cette section présente les résultats pour des tests simulés de 80 items. À l'instar des résultats présentés à la figure 4.1, nous remarquons une asymétrie plus prononcée entre la fonction de densité de I_z et celle de la loi $N(0,1)$ aux valeurs $\theta = -2$ et $\theta = 2$ (voir la figure 4.7).

C'est le modèle à trois paramètres qui présente la fonction de densité la plus près de celle de la $N(0,1)$ lorsque $\theta \leq 0$. Lorsque $\theta \geq 1$, c'est le modèle de Rasch qui présente la fonction de densité la plus près de celle de la $N(0,1)$. La fonction de densité du modèle à deux paramètres est systématiquement celle qui est la plus éloignée de la loi $N(0,1)$.

Enfin, les fonctions de densité ont une asymétrie négative. Encore une fois, cela explicite le fait que la moyenne des scores de I_z est inférieure au mode et à la médiane.

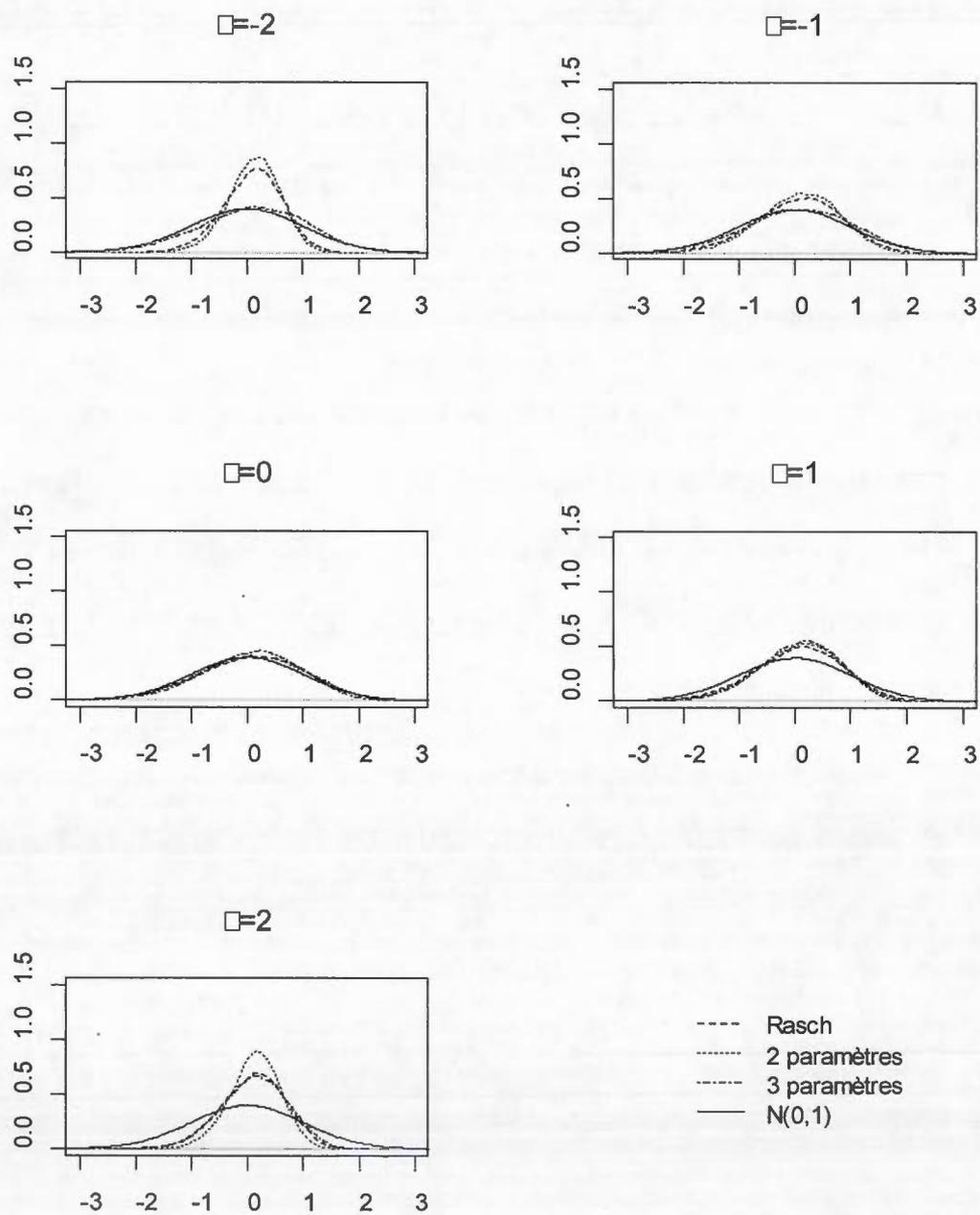


Figure 4.7: Scores de l'indice I_z selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items

La figure 4.8 montre que l'adéquation entre les fonctions de densité de I_z^* et la fonction de densité de la loi $N(0,1)$ est très importante. Ces résultats montrent que les trois modèles de réponse à l'item donnent des résultats similaires pour cet indice.

À l'instar de ce qui a été observé pour I_z , les fonctions de densité de I_z^* ont une asymétrie négative.

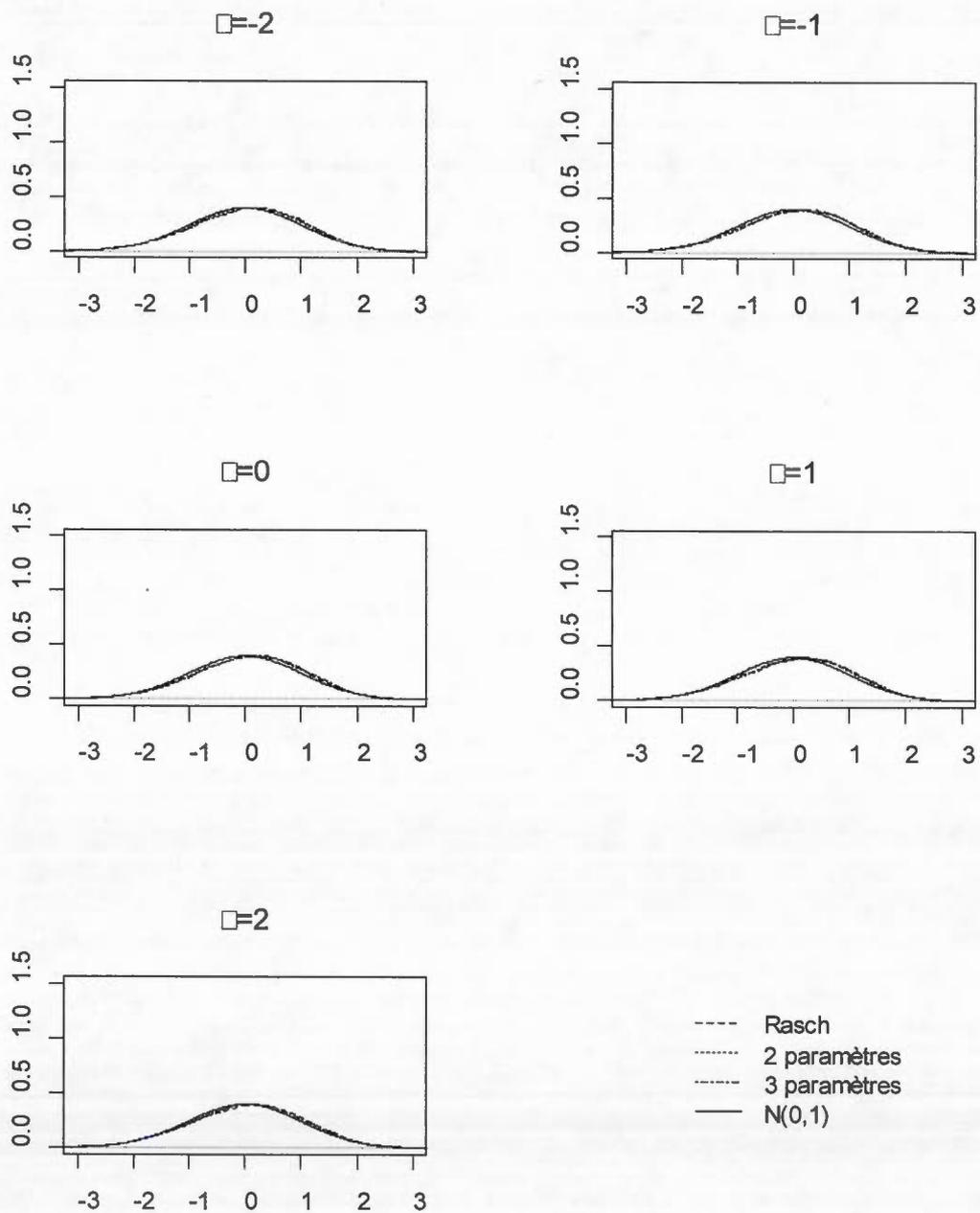


Figure 4.8: Scores de l'indice I_z^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items

La figure 4.9 montre que l'adéquation entre les scores de ZU et la loi $N(0,1)$ est meilleure lorsque nous analysons 80 items plutôt que 30 items. Dans ce cas-ci, plusieurs constats peuvent être soulevés. Premièrement, le modèle à trois paramètres présente généralement les scores les plus près de la $N(0,1)$ lorsque θ est inférieur à zéro. À ces valeurs de θ , c'est la fonction de densité du modèle à deux paramètres qui est la plus éloignée de celle de la loi $N(0,1)$.

Deuxièmement, c'est le modèle de Rasch qui présente la fonction de densité la plus en accord avec celle de la $N(0,1)$ lorsque $\theta \geq 0$. Nous remarquons, de plus, que la fonction de densité selon le modèle à deux paramètres est celle qui est la plus éloignée de la loi $N(0,1)$.

Troisièmement, les scores de ZU présentent une asymétrie positive.

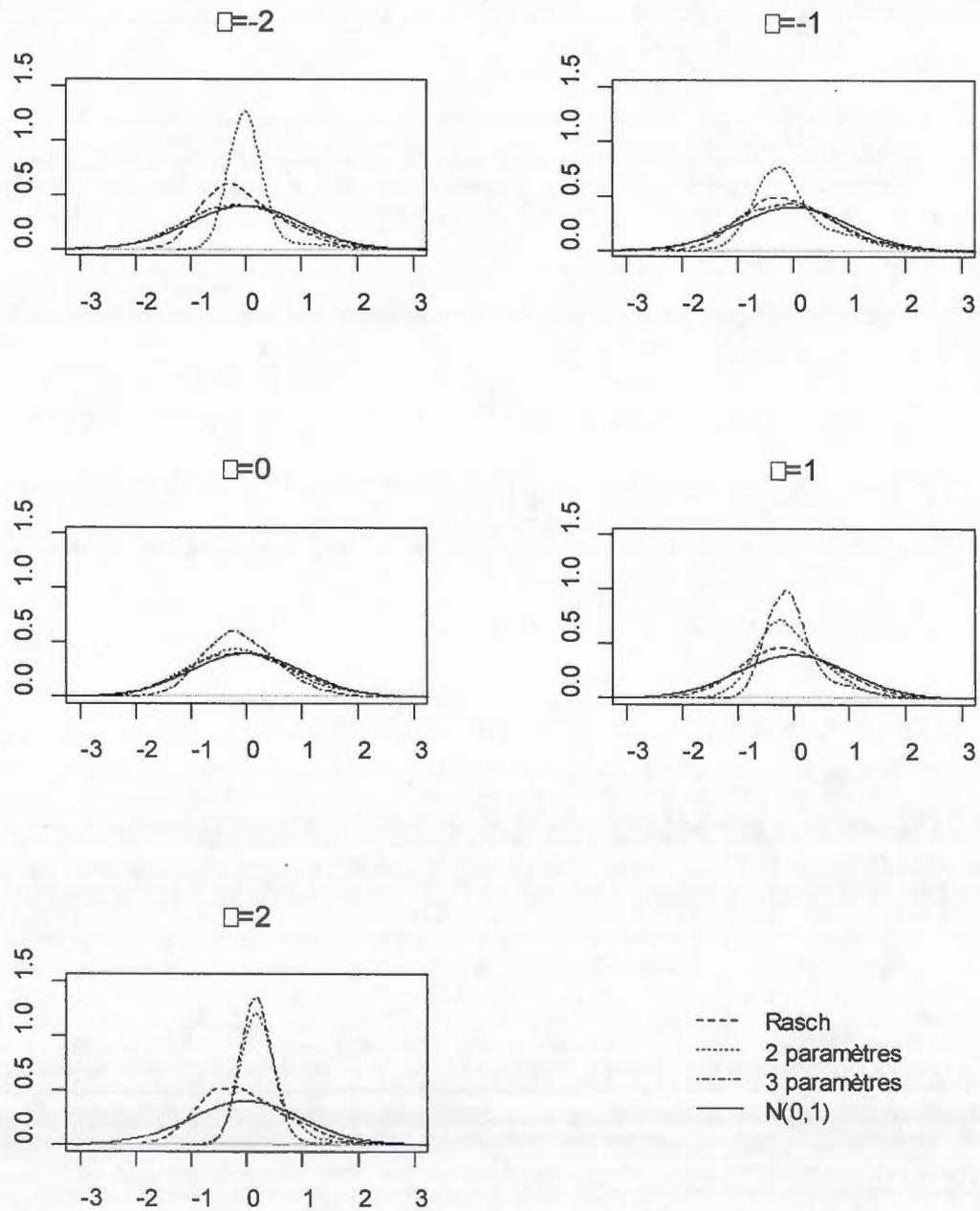


Figure 4.9: Scores de l'indice ZU selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items

Les résultats présentés à la figure 4.10 montrent que l'adéquation des fonctions de densité de ZU^* est plus grande lorsque les données comportent 80 items plutôt que 30 items. D'abord, le modèle à trois paramètres présente généralement les scores les plus près de la $N(0,1)$ lorsque θ est inférieur à zéro. À ces valeurs de θ , c'est la fonction de densité du modèle à deux paramètres qui est la plus en accord avec celle de la loi $N(0,1)$. Ensuite, c'est le modèle de Rasch qui présente la fonction de densité la plus en accord avec celle de la $N(0,1)$ lorsque $\theta \geq 0$. Il est important d'ajouter que la fonction de densité selon le modèle à deux paramètres est celle qui est la plus éloignée de la loi $N(0,1)$. Dans le quadrant $\theta = 2$, les fonctions de densité des modèles à deux et à trois paramètres deviennent particulièrement éloignées de celle de la $N(0,1)$.

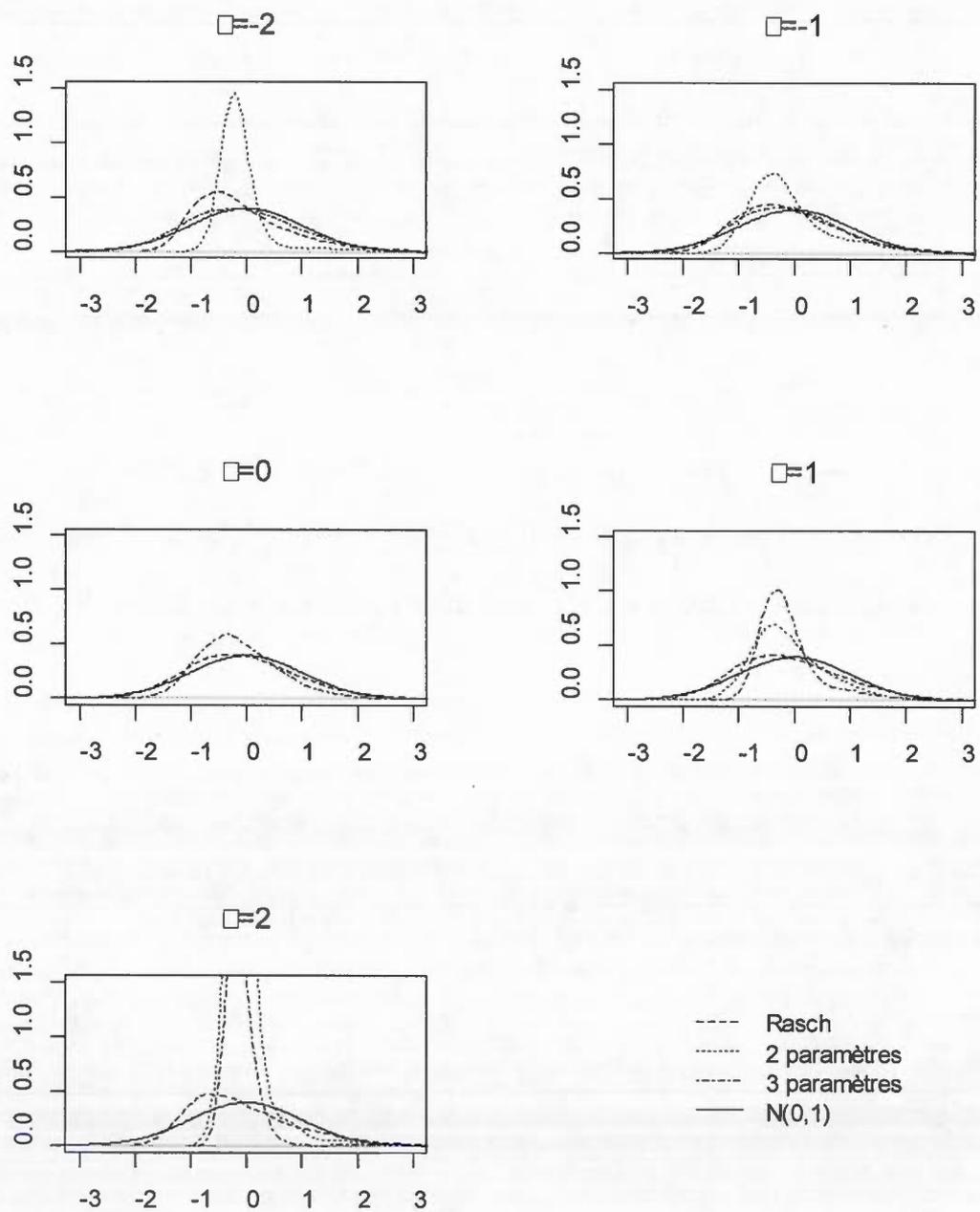


Figure 4.10: Scores de l'indice ZU^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items

Les scores de ZW sont présentés à la figure 4.11. Les fonctions de densité de ZW se rapprochent de la fonction de densité de la loi $N(0,1)$ lorsque θ tend vers zéro. Nous remarquons que le modèle à trois paramètres présente généralement des scores plus près de la $N(0,1)$ lorsque θ est inférieur à zéro. Pour les valeurs $\theta = 0$ à $\theta = 2$, c'est définitivement le modèle de Rasch qui présente la fonction de densité la plus près de celle de la $N(0,1)$. De plus, nous remarquons que la fonction de densité obtenue à l'aide du modèle à deux paramètres n'est jamais celle qui se rapproche le plus de la loi $N(0,1)$.

Enfin, les fonctions de densité obtenues à l'aide de tous les modèles de réponse à l'item présentent une asymétrie légèrement positive.

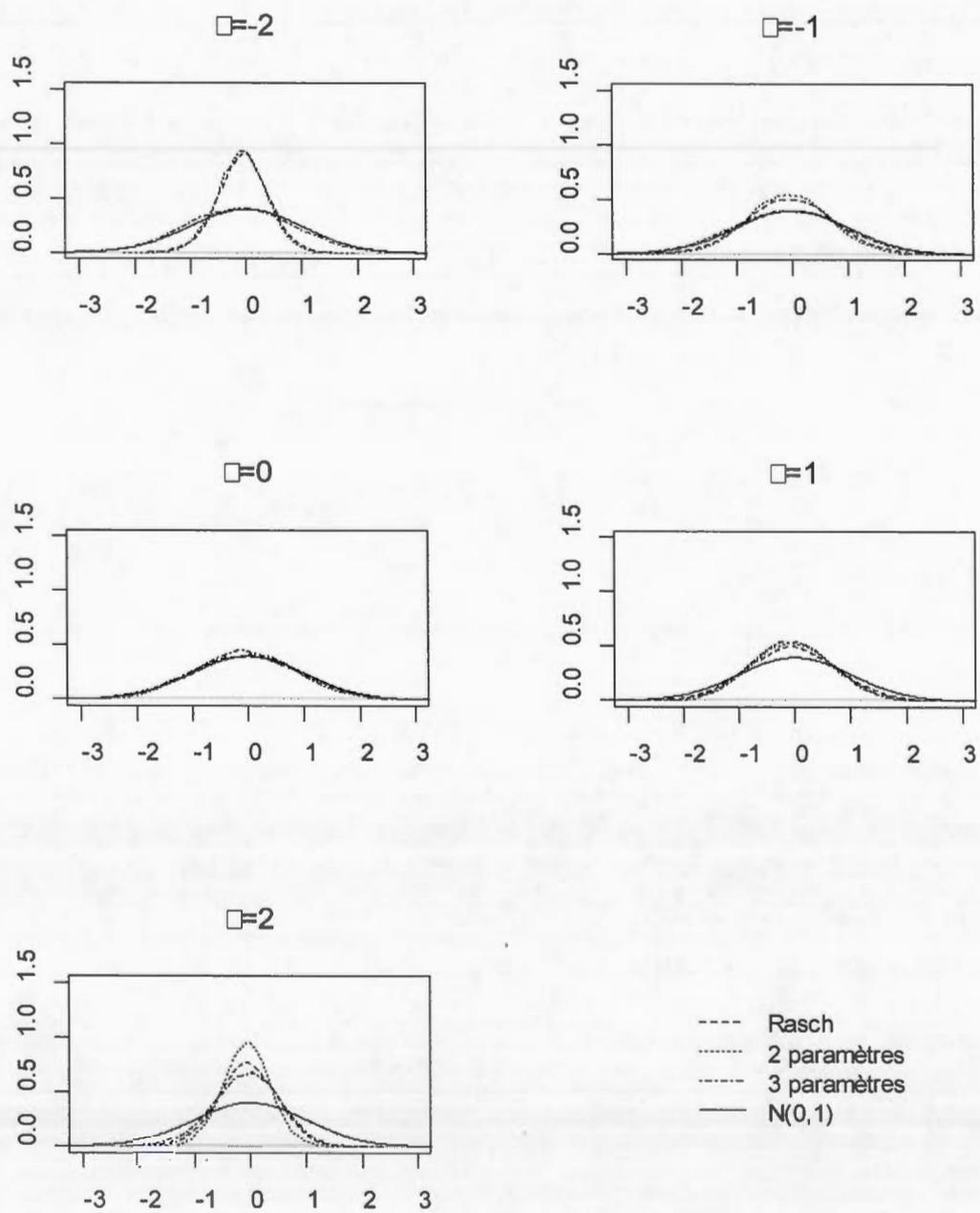


Figure 4.11: Scores de l'indice ZW selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items

Les quadrants de la figure 4.12 montrent que l'adéquation entre les fonctions de densité des scores de ZW^* et la fonction de densité de la loi $N(0,1)$ est très grande. Ce résultat montre qu'il n'existe pas de meilleur modèle de réponse à l'item. De plus, nous remarquons que les fonctions de densité ont une asymétrie légèrement positive. Cela montre que les scores de ZW^* ont une moyenne supérieure à leur médiane et leur mode.

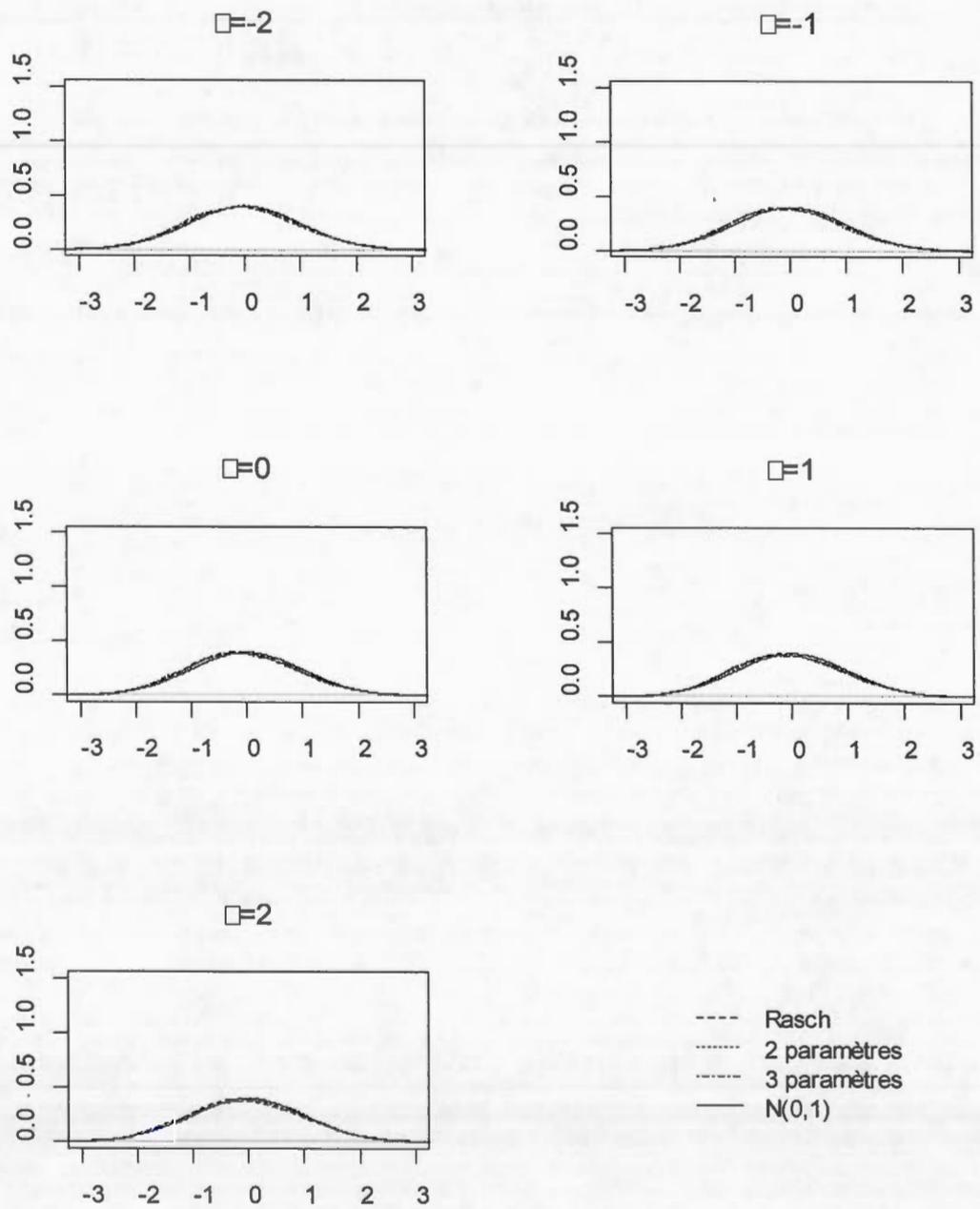


Figure 4.12: Scores de l'indice ZW^* selon l'estimateur de θ par maximum de vraisemblance pondérée pour 80 items

4.1.1.3 Synthèse

Quatre constats généraux découlent de l'analyse visuelle de la fonction de densité des scores des indices étudiés. Premièrement, nous remarquons que l'adéquation des scores des six indices à la loi $N(0,1)$ est plus importante lorsque l'épreuve d'évaluation comporte 80 items plutôt que 30 items. Il est donc préférable d'utiliser des tests longs avec les indices de détection de patrons de réponses inappropriés.

Deuxièmement, les indices I_z^* et ZW^* présentent systématiquement une meilleure adéquation à la loi $N(0,1)$ pour l'ensemble des valeurs de θ . Nous recommandons donc de favoriser ces indices pour la détection de patrons de réponses inappropriés.

Troisièmement, nos analyses ont montré que les indices ZU et ZU^* présentent une mauvaise adéquation à la loi $N(0,1)$ pour plusieurs valeurs de θ . Par exemple, pour 30 items, l'étendue des scores de ZU est parfois très importante, alors qu'elle est parfois limitée pour ZU^* . Ces résultats suggèrent d'ailleurs d'utiliser ces indices avec beaucoup de précautions, sinon de les oublier.

Quatrièmement, aux valeurs θ inférieures à zéro, il existe généralement une plus grande adéquation à la loi $N(0,1)$ lorsque les scores des indices ont été obtenus à l'aide du modèle à trois paramètres. Aux valeurs $\theta \geq 0$, il existe généralement une plus grande adéquation à la loi $N(0,1)$ lorsque les scores des indices ont été obtenus à l'aide du modèle de Rasch.

4.1.2.1 Corrélations et statistiques descriptives entre les indices : 30 items

Nous présentons, au tableau 4.1, les résultats pour le modèle de Rasch. Premièrement, les coefficients de corrélation diffèrent en fonction du θ estimé. Par exemple, nous observons un r de 0,50 entre ZU^* et ZW^* pour $\theta = -2$ comparativement à un r de 0,89 lorsque $\theta = 1$. Ce résultat est très important puisqu'il met en exergue le fait que certains indices ne se comportent pas de la même façon pour les différentes valeurs de θ .

Deuxièmement, nos analyses montrent qu'il existe une corrélation positive et très forte entre l_z et l_z^* (par exemple, $r = 1,00$ pour $\theta = 0$), entre ZU et ZU^* (par exemple, $r = 0,99$ pour $\theta = 0$) et entre ZW et ZW^* (par exemple, $r = 1,00$ pour $\theta = 0$). Ces résultats illustrent que les indices standardisés sont bien corrélés avec leur version corrigée.

Troisièmement, il est possible d'observer un coefficient de corrélation positif et généralement élevé entre les indices de type carré-moyen ZU , ZU^* , ZW et ZW^* . Ce résultat peut être expliqué par le fait que ces indices doivent présenter des scores élevés pour détecter des patrons de réponses inappropriés. Notons, néanmoins, l'existence d'un coefficient de corrélation positif et moyen entre les indices ZU^* et ZW^* pour les niveaux $\theta = -2$ et $\theta = -1$. De plus, ZU , ZU^* , ZW et ZW^* sont négativement corrélés avec l_z et l_z^* . Ce résultat est prévisible puisqu'un patron de réponses inapproprié est détecté lorsque le score des indices de type vraisemblance

est négatif, alors que ce score doit être positif et élevé pour les indices de type carré-moyen.

Quatrièmement, nous remarquons que la moyenne et l'écart-type des indices corrigés l_z^* , ZU^* et ZW^* est généralement près des valeurs de la courbe de densité normale $N(0,1)$. Ce résultat est prévisible puisque les fonctions de densité de ces indices, respectivement présentés aux figures 4.2, 4.4 et 4.6, allaient aussi dans ce sens. De leur côté, l_z , ZU et ZW présentent généralement des moyennes et des écarts-types plus faibles que ceux des indices corrigés. Les résultats de ces indices, respectivement présentés aux figures 4.1, 4.3 et 4.5, ont montré l'existence d'un écart avec la fonction de densité de la loi $N(0,1)$. Par contre, il est important de souligner que les indices standardisés se rapprochent grandement des valeurs théoriques de la $N(0,1)$ lorsque $\theta = 0$.

Cinquièmement, il est important de noter que les coefficients d'asymétrie et de kurtose des indices ZU et ZU^* sont beaucoup plus élevés que ceux des autres indices aux valeurs $\theta = -2$, $\theta = -1$ et $\theta = 0$. Ainsi, pour l'indice ZU^* , le coefficient de kurtose est égal à 111,29 pour $\theta = -2$. Ce résultat extrême n'est pas directement observé à la figure 4.4 puisque nous avons borné l'axe horizontal entre -3 et 3. Par contre, nos résultats démontrent que les scores de cet indice sont bornés entre -1,11 et 24,3. Toujours pour ZU^* , nous avons fait le même constat à la valeur $\theta = -1$: le coefficient de kurtose égal à 35,13 peut être expliqué par le fait que les scores de cet indice sont bornés entre -1,64 et 11,8. ZU et ZU^* ont donc une plus grande

propension à présenter des scores extrêmes que les quatre autres indices à l'étude.

Cela pourra donc créer des problèmes d'interprétation pour ces valeurs θ .

Enfin, le coefficient d'asymétrie des indices l_z et l_z^* est systématiquement négatif, alors qu'il est généralement positif pour ZU , ZU^* , ZW et ZW^* . Ce résultat a aussi été observé aux figures de la section 4.1.1.1 et 4.1.1.2.

Tableau 4.1 : Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle de Rasch (30 items)

	lz	lz*	ZU	ZU*	ZW	ZW*	M.(É.-T.)	A.(K.)
$\theta=-2$								
lz	1	0,96	-0,85	-0,65	-0,94	-0,91	0,16(0,48)	-0,85(1,60)
lz*	0,96	1	-0,87	-0,68	-0,87	-0,95	0,04(0,97)	-0,75(0,89)
ZU	-0,85	-0,87	1	0,91	0,67	0,71	0,08(0,58)	2,89(16,61)
ZU*	-0,65	-0,68	0,91	1	0,46	0,50	-0,01(1,01)	8,66(111,29)
ZW	-0,94	-0,87	0,67	0,46	1	0,9	-0,08(0,39)	0,04(1,54)
ZW*	-0,91	-0,95	0,71	0,50	0,90	1	-0,05(0,98)	0,18(-0,24)
$\theta=-1$								
lz	1	0,98	-0,87	-0,75	-0,97	-0,96	0,08(0,80)	-0,62(0,91)
lz*	0,98	1	-0,86	-0,75	-0,95	-0,98	0,05(1,00)	-0,59(0,44)
ZU	-0,87	-0,86	1	0,96	0,76	0,76	0,02(0,72)	2,28(10,18)
ZU*	-0,75	-0,75	0,96	1	0,62	0,63	-0,03(0,98)	4,78(35,13)
ZW	-0,97	-0,95	0,76	0,62	1	0,97	-0,05(0,78)	0,12(0,38)
ZW*	-0,96	-0,98	0,76	0,63	0,97	1	-0,05(1,01)	0,34(-0,01)
$\theta=0$								
lz	1	1	-0,92	-0,86	-0,99	-0,99	0,09(0,95)	-0,46(0,27)
lz*	1	1	-0,92	-0,87	-0,99	-0,99	0,08(0,99)	-0,45(0,24)
ZU	-0,92	-0,92	1	0,99	0,86	0,86	-0,05(0,86)	1,23(3,64)
ZU*	-0,86	-0,87	0,99	1	0,79	0,8	-0,07(0,98)	2,32(9,74)
ZW	-0,99	-0,99	0,86	0,79	1	1	-0,08(0,95)	0,05(-0,11)
ZW*	-0,99	-0,99	0,86	0,8	1	1	-0,08(1,00)	0,34(0,05)
$\theta=1$								
lz	1	0,99	-0,95	-0,94	-0,99	-0,98	0,12(0,80)	-0,50(0,45)
lz*	0,99	1	-0,95	-0,94	-0,98	-0,99	0,10(1,00)	-0,46(0,20)
ZU	-0,95	-0,95	1	0,99	0,91	0,91	-0,08(0,83)	0,38(0,37)
ZU*	-0,94	-0,94	0,99	1	0,88	0,89	-0,09(0,98)	0,92(1,43)
ZW	-0,99	-0,98	0,91	0,88	1	0,99	-0,09(0,81)	0,06(-0,04)
ZW*	-0,98	-0,99	0,91	0,89	0,99	1	-0,10(1,01)	0,38(0,05)
$\theta=2$								
lz	1	0,99	-0,92	-0,9	-0,97	-0,97	0,18(0,64)	-0,59(0,35)
lz*	0,99	1	-0,92	-0,9	-0,96	-0,97	0,11(0,98)	-0,56(0,15)
ZU	-0,92	-0,92	1	0,98	0,82	0,81	-0,09(0,76)	0,29(-0,32)
ZU*	-0,9	-0,9	0,98	1	0,78	0,78	-0,09(0,95)	1,03(1,42)
ZW	-0,97	-0,96	0,82	0,78	1	0,99	-0,12(0,64)	0,20(-0,14)
ZW*	-0,97	-0,97	0,81	0,78	0,99	1	-0,12(1,00)	0,56(0,00)

Le tableau 4.2 présente les résultats pour les scores des indices obtenus à l'aide du modèle à deux paramètres. Nous observons généralement des résultats comparables au tableau 4.1 pour les indices l_z , l_z^* , ZW et ZW^* . Par contre, les résultats pour les indices ZU et ZU^* sont légèrement différents de ceux obtenus au tableau précédent, spécialement en ce qui concerne $\theta = -2$, $\theta = -1$, $\theta = 1$ et $\theta = 2$. Par exemple, la corrélation entre ZU^* et ZW^* est de 0,13 ($\theta = -2$) dans le tableau 4.2, alors que cette corrélation était de 0,50 ($\theta = -2$) dans le tableau 4.1. Cette situation montre que ces indices sont sensibles au modèle de réponse à l'item utilisé qu'ils remettent en question leur utilisation avec le modèle à deux paramètres. Rappelons que les résultats présentés aux sections 4.1.1.1 et 4.1.1.2 ont aussi montré que le modèle de réponse à l'item utilisé avait un effet sur l'adéquation des fonctions de densité des indices à la fonction de densité de la loi $N(0,1)$.

Encore une fois, nous observons des coefficients de corrélation qui varient en fonction du θ estimation. Par exemple, il existe un $r = 0,13$ entre ZU^* et ZW^* pour $\theta = -2$ comparativement à un $r = 0,61$ lorsque $\theta = 1$. Évidemment, cet écart n'est pas aussi grand pour tous les coefficients de corrélation calculés entre les autres indices. Ensuite, nos analyses montrent qu'il existe une corrélation positive et très élevée entre l_z et l_z^* (par exemple, $r = 1,00$ pour $\theta = 0$) ainsi qu'entre ZW et ZW^* (par exemple, $r = 1,00$ pour $\theta = 0$). Par contre, la corrélation entre ZU et ZU^* varie grandement en fonction de θ : nous obtenons un $r = 0,01$ pour $\theta = -2$ et un $r = 0,97$ pour $\theta = 0$. Finalement, les indices de type carré-moyen ZU , ZU^* , ZW et ZW^* sont

généralement négativement corrélés avec les indices de type vraisemblance l_z et l_z^* . Ce résultat est assez prévisible puisqu'un patron de réponses inapproprié est détecté lorsque le score des indices de type vraisemblance est négatif et faible, alors que ce score doit être positif et élevé pour les indices de type carré-moyen.

En ce qui a trait aux statistiques présentées aux deux dernières colonnes du tableau 4.2, nous obtenons des moyennes et des écarts-types relativement similaires et en accord avec la $N(0,1)$ pour les indices l_z , l_z^* , ZW et ZW^* . De son côté, l'indice ZU présente des moyennes et des écarts-types beaucoup plus élevés que les autres indices aux valeurs $\theta = -2$ et $\theta = 2$. Par exemple, nous calculons une moyenne de 3,84 et un écart-type de 19,58 pour $\theta = -2$ et une moyenne de 9,28 et un écart-type de 50,38 pour $\theta = 2$. En ce qui concerne ZU^* , nous avons calculé des moyennes près des valeurs de la $N(0,1)$. Par contre, les écarts-types de cet indice sont les plus petits de tous (ce qui explique la fonction de densité des scores plutôt étroite que nous avons observée à la figure 4.4). Par exemple, l'écart-type de cet indice est égal à 0,50 pour $\theta = -2$ et à 0,25 pour $\theta = 2$. Ces résultats complexifient grandement l'utilisation de ces deux indices en contexte réel.

À l'instar des résultats présentés aux sections 4.1.1.1 et 4.1.1.2, l'analyse des coefficients d'asymétrie et de kurtose permet de montrer que les indices ZU et ZU^* s'éloignent de la loi $N(0,1)$. Par exemple, ZU présente un coefficient d'asymétrie de 128,17 et un coefficient de kurtose de 17 523,96 à la valeur $\theta = 1$. Ces résultats ne sont pas surprenants puisque la fonction de densité des scores de cet indice différerait

beaucoup de celle de la $N(0,1)$, à la figure 4.3. Il est à noter que des coefficients élevés ont aussi été trouvés aux valeurs $\theta = -2$ et $\theta = -1$. De son côté, ZU^* présente aussi quelques résultats extrêmes. Par exemple, nous calculons un coefficient d'asymétrie de 95,77 et un coefficient de kurtose de 9739,26 à la valeur $\theta = -2$. À la valeur $\theta = -1$, le coefficient d'asymétrie est égal à 33,60 et le coefficient de kurtose à 1738,16. Ainsi, les deux quadrants du haut de la figure 4.4 montrent explicitement que la fonction de densité de cet indice remet en question la possibilité d'utiliser cet indice de façon adéquate.

Tableau 4.2 : Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle à deux paramètres (30 items)

	lz	lz*	ZU	ZU*	ZW	ZW*	M.(É.-T.)	A.(K.)
$\theta=-2$								
lz	1	0,98	0,07	-0,2	-0,96	-0,94	0,22(0,55)	-0,40(0,15)
lz*	0,98	1	-0,01	-0,18	-0,93	-0,95	0,07(1,00)	-0,30(-0,21)
ZU	0,07	-0,01	1	0,01	-0,11	0	3,84(19,58)	13,91(200,50)
ZU*	-0,2	-0,18	0,01	1	0,14	0,13	-0,01(0,50)	95,77(9739,26)
ZW	-0,96	-0,93	-0,11	0,14	1	0,97	-0,14(0,56)	-0,04(0,01)
ZW*	-0,94	-0,95	0	0,13	0,97	1	-0,04(1,01)	0,17(-0,35)
$\theta=-1$								
lz	1	0,99	-0,33	-0,44	-0,99	-0,98	0,15(0,75)	-0,45(0,35)
lz*	0,99	1	-0,33	-0,42	-0,98	-0,99	0,06(1,00)	-0,38(0,02)
ZU	-0,33	-0,33	1	0,38	0,31	0,3	0,32(0,66)	37,51(2962,38)
ZU*	-0,44	-0,42	0,38	1	0,37	0,35	-0,03(0,58)	33,60(1738,16)
ZW	-0,99	-0,98	0,31	0,37	1	0,99	-0,12(0,77)	0,11(-0,03)
ZW*	-0,98	-0,99	0,3	0,35	0,99	1	-0,05(1,01)	0,32(-0,06)
$\theta=0$								
lz	1	1	-0,89	-0,81	-0,99	-0,99	0,12(0,94)	-0,61(0,64)
lz*	1	1	-0,89	-0,81	-0,98	-0,99	0,10(0,99)	-0,61(0,60)
ZU	-0,89	-0,89	1	0,97	0,82	0,83	-0,07(0,78)	1,80(7,79)
ZU*	-0,81	-0,81	0,97	1	0,72	0,73	-0,09(0,93)	4,00(33,35)
ZW	-0,99	-0,98	0,82	0,72	1	1	-0,11(0,95)	0,19(-0,04)
ZW*	-0,99	-0,99	0,83	0,73	1	1	-0,09(1,00)	0,44(0,23)
$\theta=1$								
lz	1	0,99	-0,16	-0,76	-0,97	-0,96	0,15(0,78)	-0,62(0,42)
lz*	0,99	1	-0,17	-0,73	-0,96	-0,97	0,10(1,02)	-0,52(0,06)
ZU	-0,16	-0,17	1	0,19	0,14	0,14	0,12(2,92)	128,17(17523,96)
ZU*	-0,76	-0,73	0,19	1	0,64	0,61	-0,11(0,72)	5,89(138,29)
ZW	-0,97	-0,96	0,14	0,64	1	0,99	-0,11(0,79)	0,19(-0,16)
ZW*	-0,96	-0,97	0,14	0,61	0,99	1	-0,08(1,04)	0,33(-0,25)
$\theta=2$								
lz	1	0,93	0,15	-0,61	-0,92	-0,86	0,25(0,45)	-0,81(2,14)
lz*	0,93	1	0,01	-0,55	-0,84	-0,9	-0,02(0,88)	-0,35(0,29)
ZU	0,15	0,01	1	0,01	-0,27	-0,04	9,28(50,38)	7,52(55,28)
ZU*	-0,61	-0,55	0,01	1	0,4	0,36	-0,01(0,25)	5,21(42,06)
ZW	-0,92	-0,84	-0,27	0,4	1	0,93	-0,15(0,44)	-0,11(1,51)
ZW*	-0,86	-0,9	-0,04	0,36	0,93	1	0,07(0,85)	-0,03(-0,01)

Les résultats pour le modèle à trois paramètres sont présentés au tableau 4.3. Tout comme dans les tableaux 4.1 et 4.2, il existe une corrélation positive et très forte entre l_z et l_z^* (par exemple, $r = 0,99$ pour $\theta = 0$) ainsi qu'entre ZW et ZW^* (par exemple, $r = 0,98$ pour $\theta = 0$). À l'instar des résultats pour le modèle à deux paramètres, la corrélation entre ZU et ZU^* varie grandement en fonction de θ . À titre d'exemple, nous obtenons un $r = 0,98$ pour $\theta = -2$ et un $r = 0,01$ pour $\theta = 2$.

Pour les valeurs $\theta = -2$ et $\theta = -1$, les corrélations entre les indices de type carré moyen et les indices de type vraisemblance tendent généralement vers moins un. Il est toutefois important de soulever la nuance suivante: aux valeurs $\theta = 1$ et $\theta = 2$, nous calculons un coefficient de corrélation de Pearson faible entre ZU et les indices de type vraisemblance l_z et l_z^* .

Nous remarquons que la moyenne et l'écart-type de l_z^* et de ZW^* sont généralement près des valeurs de la $N(0,1)$. De leur côté, les indices l_z et ZW s'éloignent légèrement de la $N(0,1)$ lorsque θ est au-dessus de zéro. Par exemple, l'écart-type de ZW est égal à 0,93 lorsque $\theta = -1$ et il descend à 0,58 lorsque $\theta = 2$. De son côté, ZU présente des moyennes et des écarts-types qui sont parfois très loin des valeurs théoriques de la $N(0,1)$ lorsque $\theta > 0$. Ainsi, la moyenne et l'écart-type de l'indice sont respectivement égaux à 537,78 et à 19721,75 pour la valeur $\theta = 1$ et à 19205,37 et 117281,32 pour la valeur $\theta = 2$. Ces résultats s'expliquent par le fait que les scores de ZU sont très étendus. Par exemple, les scores de ZU sont bornés entre -1,00 et 745 700 pour $\theta = 1$ (voir les derniers quadrants de la figure 4.3). Enfin, les

scores de ZU^* présentent des moyennes près de la $N(0,1)$, mais les écarts-types de cet indice deviennent très petits lorsque $\theta = 1$ (écart-type de 0,09) et $\theta = 2$ (écart-type de 0,01). Cela n'est pas surprenant puisque les fonctions de densité de cet indice sont parfois très étroites dans les deux derniers quadrants de la figure 4.4. Encore une fois, ces résultats montrent que ZU et ZU^* sont difficiles à utiliser dans la réalité.

L'étude des coefficients d'asymétrie et de kurtose permet de mettre en exergue le fait que les indices ZW^* et ZW sont légèrement plus près de la $N(0,1)$ que l_z et l_z^* . Pour les niveaux $\theta \leq 1$, les coefficients d'asymétrie des indices de type carré moyens sont généralement positifs alors qu'ils sont tous négatifs pour les indices de type vraisemblance. De plus, ces coefficients sont beaucoup plus élevés pour les indices ZU et ZU^* ; surtout lorsque θ est situé entre zéro et deux. Pour ZU , les coefficients d'asymétrie et de kurtose calculés sont respectivement égaux à 18,51 et 625,54 lorsque $\theta = 0$ et 37,75 et 1423,30 lorsque $\theta = 1$. Ce constat est aussi observé dans la figure 4.3. Certains résultats sont aussi élevés pour ZU^* : les coefficients d'asymétrie et de kurtose calculés sont respectivement égaux à 9,69 et 355,53 lorsque $\theta = 1$ et 7,34 et 153,56 lorsque $\theta = 2$.

Tableau 4.3 : Coefficients de corrélation de Pearson et statistiques descriptives selon

l'estimateur par maximum de vraisemblance pondérée et le modèle à trois paramètres (30 items)

	lz	lz*	ZU	ZU*	ZW	ZW*	M.(É.-T.)	A.(K.)
$\theta=-2$								
lz	1	0,97	-0,97	-0,94	-0,99	-0,97	0,25(0,91)	-0,54(0,40)
lz*	0,97	1	-0,95	-0,96	-0,96	-0,99	0,08(1,00)	-0,44(0,18)
ZU	-0,97	-0,95	1	0,98	0,94	0,92	-0,21(0,93)	0,67(0,52)
ZU*	-0,94	-0,96	0,98	1	0,9	0,93	-0,07(0,99)	0,91(0,95)
ZW	-0,99	-0,96	0,94	0,9	1	0,97	-0,26(0,92)	0,23(0,04)
ZW*	-0,97	-0,99	0,92	0,93	0,97	1	-0,09(1,00)	0,33(0,04)
$\theta=-1$								
lz	1	0,99	-0,96	-0,95	-0,99	-0,99	0,13(0,93)	-0,47(0,25)
lz*	0,99	1	-0,95	-0,95	-0,98	-0,99	0,07(0,99)	-0,46(0,18)
ZU	-0,96	-0,95	1	0,99	0,93	0,92	-0,11(0,90)	0,69(0,64)
ZU*	-0,95	-0,95	0,99	1	0,92	0,92	-0,07(0,95)	0,96(1,21)
ZW	-0,99	-0,98	0,93	0,92	1	0,99	-0,12(0,93)	0,18(-0,06)
ZW*	-0,99	-0,99	0,92	0,92	0,99	1	-0,07(1,00)	0,35(0,04)
$\theta=0$								
lz	1	0,99	-0,26	-0,82	-0,99	-0,98	0,15(0,75)	-0,44(0,31)
lz*	0,99	1	-0,25	-0,75	-0,97	-0,99	0,11(0,98)	-0,49(0,13)
ZU	-0,26	-0,25	1	0,32	0,26	0,24	0,36(1,66)	18,51(625,54)
ZU*	-0,82	-0,75	0,32	1	0,79	0,72	-0,08(0,54)	1,17(6,35)
ZW	-0,99	-0,97	0,26	0,79	1	0,98	-0,12(0,76)	0,11(0,10)
ZW*	-0,98	-0,99	0,24	0,72	0,98	1	-0,10(0,99)	0,40(0,04)
$\theta=1$								
lz	1	0,99	0,02	-0,5	-0,97	-0,97	0,18(0,61)	-0,62(0,14)
lz*	0,99	1	0	-0,46	-0,97	-0,97	0,10(1,00)	-0,55(-0,11)
ZU	0,02	0	1	0	-0,04	0	537,78(19721,75)	37,75(1423,30)
ZU*	-0,5	-0,46	0	1	0,42	0,4	-0,01(0,09)	9,69(355,53)
ZW	-0,97	-0,97	-0,04	0,42	1	0,99	-0,12(0,62)	0,32(-0,29)
ZW*	-0,97	-0,97	0	0,4	0,99	1	-0,09(1,00)	0,53(-0,18)
$\theta=2$								
lz	1	0,95	0,15	-0,45	-0,95	-0,9	0,25(0,52)	-0,25(-0,18)
lz*	0,95	1	0	-0,46	-0,91	-0,94	-0,04(0,90)	0,27(-0,24)
ZU	0,15	0	1	0,01	-0,25	-0,06	19205,37(117281,32)	6,03(34,39)
ZU*	-0,45	-0,46	0,01	1	0,33	0,35	0,00(0,01)	7,34(153,56)
ZW	-0,95	-0,91	-0,25	0,33	1	0,96	-0,16(0,58)	-0,21(-0,26)
ZW*	-0,9	-0,94	-0,06	0,35	0,96	1	0,08(0,92)	-0,29(-0,25)

4.1.2.2 Corrélations et statistiques descriptives entre les indices : 80 items

Nous présentons, dans le tableau 4.4, les résultats des scores des indices analysés selon le modèle de Rasch et 80 items. Nous pouvons procéder à plusieurs constats qui abondent dans le même sens que ceux obtenus pour 30 items (voir le tableau 4.1).

Premièrement, les coefficients de corrélations diffèrent en fonction du θ estimé. En effet, nous calculons un $r = 0,65$ entre ZU^* et ZW^* au niveau $\theta = -2$ comparativement à un $r = 0,94$ lorsque $\theta = 0$. Cela indique que les indices ne se comportent pas de la même façon à toutes les valeurs de θ .

Deuxièmement, nos analyses montrent qu'il existe une corrélation positive et très forte entre l_z et l_z^* (par exemple, $r = 1,00$ pour $\theta = 0$), entre ZU et ZU^* (par exemple, $r = 1,00$ pour $\theta = 0$) ainsi qu'entre ZW et ZW^* (par exemple, $r = 1,00$ pour $\theta = 0$). Ainsi, les indices standardisés sont fortement corrélés avec leur version corrigée.

Troisièmement, il est possible d'observer un coefficient de corrélation positif et généralement élevé entre les indices de type carré-moyen ZU , ZU^* , ZW et ZW^* . À l'instar des analyses pour 30 items, les coefficients de corrélation entre les indices de type carré-moyen et les indices l_z et l_z^* sont bornés entre $-0,81$ et $-0,99$. De tels résultats sont attendus puisqu'un patron de réponses inapproprié est détecté lorsque le

score des indices de type vraisemblance est négatif, alors que ce score doit être positif et élevé pour les indices de type carré-moyen.

Quatrièmement, la moyenne et l'écart-type des scores des indices de détection sont généralement plus près des valeurs de la loi $N(0,1)$ pour les résultats analysés sur 80 items plutôt que sur 30 items. Ce résultat a aussi été observé aux figures de la section 4.1.1.2. Par contre, les indices l_z et ZW présentent des écarts-types plus petits que ceux de ZU pour $\theta = -2$ et $\theta = 2$.

Cinquièmement, les coefficients d'asymétrie des indices l_z et l_z^* sont systématiquement négatifs, alors qu'ils sont généralement positifs pour tous les indices de type carré-moyens. Il est important de mentionner que ce sont les indices ZU et ZU^* qui présentent les coefficients d'asymétrie les moins en accord avec ceux de la loi $N(0,1)$. De plus, ZU^* présente les coefficients de kurtose le plus éloignés de la loi $N(0,1)$, par exemple 4,22 à $\theta = -2$, alors que les autres indices présentent des coefficients de kurtose relativement près de la $N(0,1)$. Notons que ces résultats sont tous cohérents avec les fonctions de densité des scores des indices présentés à la section 4.1.1.2.

Tableau 4.4 : Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle de Rasch (80 items)

	lz	lz*	ZU	ZU*	ZW	ZW*	M.(É.-T.)	A.(K.)
$\theta=-2$								
lz	1	0,99	-0,85	-0,81	-0,96	-0,95	0,10(0,53)	-0,27(0,24)
lz*	0,99	1	-0,86	-0,82	-0,94	-0,96	0,05(1,00)	-0,27(0,05)
ZU	-0,85	-0,86	1	0,99	0,69	0,7	-0,02(0,75)	0,80(0,97)
ZU*	-0,81	-0,82	0,99	1	0,64	0,65	-0,03(0,98)	1,56(4,22)
ZW	-0,96	-0,94	0,69	0,64	1	0,98	-0,05(0,46)	-0,10(0,45)
ZW*	-0,95	-0,96	0,7	0,65	0,98	1	-0,05(1,00)	0,10(-0,10)
$\theta=-1$								
lz	1	0,99	-0,9	-0,88	-0,98	-0,98	0,06(0,79)	-0,28(0,17)
lz*	0,99	1	-0,9	-0,88	-0,97	-0,98	0,04(1,00)	-0,27(0,06)
ZU	-0,9	-0,9	1	1	0,81	0,8	-0,03(0,84)	0,52(0,45)
ZU*	-0,88	-0,88	1	1	0,78	0,78	-0,04(0,99)	0,94(1,57)
ZW	-0,98	-0,97	0,81	0,78	1	0,99	-0,04(0,79)	0,01(0,08)
ZW*	-0,98	-0,98	0,8	0,78	0,99	1	-0,05(1,00)	0,19(0,00)
$\theta=0$								
lz	1	1	-0,97	-0,96	-0,99	-0,99	0,05(0,98)	-0,26(0,08)
lz*	1	1	-0,97	-0,97	-0,99	-0,99	0,05(1,00)	-0,26(0,07)
ZU	-0,97	-0,97	1	1	0,94	0,94	-0,05(0,97)	0,18(0,02)
ZU*	-0,96	-0,97	1	1	0,93	0,94	-0,05(0,99)	0,46(0,35)
ZW	-0,99	-0,99	0,94	0,93	1	1	-0,05(0,98)	0,03(-0,03)
ZW*	-0,99	-0,99	0,94	0,94	1	1	-0,05(1,00)	0,22(0,05)
$\theta=1$								
lz	1	1	-0,93	-0,92	-0,99	-0,98	0,06(0,81)	-0,26(0,09)
lz*	1	1	-0,92	-0,92	-0,98	-0,99	0,05(1,00)	-0,25(0,00)
ZU	-0,93	-0,92	1	1	0,86	0,85	-0,04(0,85)	0,25(-0,04)
ZU*	-0,92	-0,92	1	1	0,84	0,84	-0,04(1,00)	0,58(0,40)
ZW	-0,99	-0,98	0,86	0,84	1	0,99	-0,05(0,81)	0,01(-0,01)
ZW*	-0,98	-0,99	0,85	0,84	0,99	1	-0,05(1,00)	0,20(-0,05)
$\theta=2$								
lz	1	0,99	-0,87	-0,85	-0,96	-0,96	0,10(0,58)	-0,28(0,15)
lz*	0,99	1	-0,87	-0,85	-0,95	-0,96	0,05(1,00)	-0,25(-0,02)
ZU	-0,87	-0,87	1	0,99	0,73	0,73	-0,04(0,77)	0,48(0,04)
ZU*	-0,85	-0,85	0,99	1	0,69	0,7	-0,04(0,99)	1,01(1,37)
ZW	-0,96	-0,95	0,73	0,69	1	0,99	-0,05(0,53)	-0,02(0,20)
ZW*	-0,96	-0,96	0,73	0,7	0,99	1	-0,05(1,01)	0,15(-0,12)

Le tableau 4.5 présente les résultats pour le modèle à deux paramètres. Encore une fois, nos observations vont dans le même sens que celles des analyses portant sur des patrons de réponses constitués de 30 items. Ainsi, les coefficients de corrélation entre les indices varient en fonction des valeurs de θ . Par exemple, nous remarquons un $r = 0,94$ entre ZU^* et ZW^* pour $\theta = 0$ comparativement à un $r = 0,14$ lorsque $\theta = 2$. Ensuite, nos analyses montrent généralement l'existence d'une corrélation positive et très élevée entre l_z et l_z^* (par exemple, $r = 1,00$ pour $\theta = 0$), entre ZU et ZU^* (par exemple, $r = 1,00$ pour $\theta = 0$) ainsi qu'entre ZW et ZW^* (par exemple, $r = 1,00$ pour $\theta = 0$).

Les indices de type carré-moyen ZU , ZU^* , ZW et ZW^* sont négativement corrélés avec les indices de type vraisemblance l_z et l_z^* . Nous remarquons aussi que cette corrélation diminue pour les indices ZU et ZU^* aux valeurs $\theta = -2$ et $\theta = 2$.

Comme nous l'avons observé aux figures de la section 4.1.1.2, les moyennes (bornées entre -0,08 et 0,29) et les écarts-types (bornés entre 0,43 et 1,00) sont généralement près des valeurs de la $N(0,1)$. Par contre, les moyennes et les écarts-types des indices corrigés sont généralement plus en accord avec la $N(0,1)$ que ceux des indices standardisés. En effet, nous remarquons que les indices l_z , ZU et ZW présentent des écarts-types légèrement inférieurs à $\theta = -2$ et $\theta = 2$.

En ce qui a trait à l'asymétrie, nous observons que les indices l_z et l_z^* présentent des coefficients systématiquement négatifs, alors qu'ils sont généralement

positifs pour les indices de type carré-moyen. De plus, dans les situations où θ s'éloigne de zéro, les coefficients d'asymétrie et de kurtose de ZU et ZU^* deviennent de plus en plus élevés, ce qui témoigne d'un écart avec la loi $N(0,1)$. Par exemple, pour ZU , les coefficients d'asymétrie et de kurtose sont respectivement égaux à 3,24 et 21,17 pour $\theta = -2$ et à 7,16 et 170,10 pour $\theta = 2$. Ces résultats peuvent être observés au premier quadrant et au dernier quadrant de la figure 4.9. Dans le cas de ZU^* , les coefficients d'asymétrie et de kurtose sont respectivement égaux à 10,28 et 195,61 pour $\theta = -2$ et 58,55 et 4870,87 pour $\theta = 2$. Nous obtenons aussi des résultats qui vont dans ce sens au sein de la figure 4.10. En ce qui a trait aux autres indices, ils présentent tous des coefficients d'asymétrie et de kurtose relativement près de ceux de la loi $N(0,1)$. Ainsi, il est difficile de recommander l'utilisation de ZU et de ZU^* lors de l'analyse de données à l'aide du modèle à deux paramètres. Ce constat est aussi confirmé par les analyses présentées aux sections 4.1.1.1 et 4.1.1.2.

Tableau 4.5 : Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle à deux paramètres (80 items)

	lz	lz*	ZU	ZU*	ZW	ZW*	M.(É.-T.)	A.(K.)
$\theta=-2$								
lz	1	0,99	-0,59	-0,44	-0,92	-0,9	0,13(0,46)	-0,27(0,27)
lz*	0,99	1	-0,59	-0,44	-0,9	-0,91	0,01(0,99)	-0,22(0,01)
ZU	-0,59	-0,59	1	0,91	0,31	0,3	0,09(0,52)	3,24(21,17)
ZU*	-0,44	-0,44	0,91	1	0,19	0,17	-0,03(0,87)	10,28(195,61)
ZW	-0,92	-0,9	0,31	0,19	1	0,99	-0,07(0,44)	-0,03(0,21)
ZW*	-0,9	-0,91	0,3	0,17	0,99	1	0,02(0,99)	0,10(-0,05)
$\theta=-1$								
lz	1	0,99	-0,77	-0,69	-0,97	-0,97	0,09(0,73)	-0,26(0,16)
lz*	0,99	1	-0,77	-0,69	-0,96	-0,97	0,04(0,99)	-0,23(0,00)
ZU	-0,77	-0,77	1	0,98	0,63	0,62	-0,03(0,76)	1,96(8,14)
ZU*	-0,69	-0,69	0,98	1	0,55	0,54	-0,03(0,98)	3,62(25,45)
ZW	-0,97	-0,96	0,63	0,55	1	0,99	-0,06(0,72)	0,02(0,09)
ZW*	-0,97	-0,97	0,62	0,54	0,99	1	-0,04(0,99)	0,16(-0,04)
$\theta=0$								
lz	1	1	-0,97	-0,96	-0,99	-0,99	0,06(0,97)	-0,26(0,04)
lz*	1	1	-0,97	-0,97	-0,99	-0,99	0,06(0,99)	-0,26(0,03)
ZU	-0,97	-0,97	1	1	0,94	0,94	-0,06(0,95)	0,44(0,41)
ZU*	-0,96	-0,97	1	1	0,93	0,94	-0,07(0,96)	2,05(18,45)
ZW	-0,99	-0,99	0,94	0,93	1	1	-0,06(0,98)	0,05(-0,08)
ZW*	-0,99	-0,99	0,94	0,94	1	1	-0,06(0,99)	0,20(0,00)
$\theta=1$								
lz	1	0,99	-0,81	-0,7	-0,98	-0,97	0,08(0,75)	-0,28(0,14)
lz*	0,99	1	-0,8	-0,69	-0,97	-0,97	0,05(1,00)	-0,27(0,00)
ZU	-0,81	-0,8	1	0,97	0,69	0,67	-0,02(0,69)	2,06(14,25)
ZU*	-0,7	-0,69	0,97	1	0,58	0,57	-0,05(0,90)	6,28(97,91)
ZW	-0,98	-0,97	0,69	0,58	1	0,99	-0,06(0,75)	0,04(0,06)
ZW*	-0,97	-0,97	0,67	0,57	0,99	1	-0,05(1,00)	0,18(-0,07)
$\theta=2$								
lz	1	0,98	-0,45	-0,32	-0,92	-0,9	0,13(0,46)	-0,26(0,28)
lz*	0,98	1	-0,47	-0,3	-0,9	-0,91	0,01(0,99)	-0,19(-0,05)
ZU	-0,45	-0,47	1	0,54	0,28	0,28	0,29(0,49)	7,16(170,10)
ZU*	-0,32	-0,3	0,54	1	0,16	0,14	-0,01(0,87)	58,55(4870,87)
ZW	-0,92	-0,9	0,28	0,16	1	0,98	-0,08(0,43)	-0,06(0,24)
ZW*	-0,9	-0,91	0,28	0,14	0,98	1	0,01(1,00)	0,24(-0,14)

Finalement, les résultats pour les scores obtenus à l'aide du modèle à trois paramètres sont présentés au tableau 4.6. Encore une fois, nous observons l'existence d'une corrélation positive et très forte entre l_z et l_z^* (par exemple, $r = 1,00$ pour $\theta = 0$), entre ZU et ZU^* (par exemple, $r = 0,99$ pour $\theta = 0$) ainsi qu'entre ZW et ZW^* (par exemple, $r = 1,00$ pour $\theta = 0$). Par contre cette corrélation a tendance à fluctuer selon les différentes valeurs de θ . À l'instar de certains résultats soulevés un peu plus haut et aux sections 4.1.1.1 et 4.1.1.2, les indices ZU et ZU^* corréleront moins bien avec les autres indices à l'étude, surtout lorsque $\theta = 1$ et $\theta = 2$. Enfin, ZU , ZU^* , ZW et ZW^* sont encore négativement corrélés avec les indices de type vraisemblance l_z et l_z^* . Ce résultat est prévisible puisqu'un patron de réponses inapproprié est détecté lorsque le score des indices de type vraisemblance est négatif, alors que ce score doit être positif et élevé pour les indices de type carré-moyen.

Nous remarquons que la moyenne et l'écart-type des trois indices corrigés sont généralement près des valeurs de la courbe de densité normale $N(0,1)$. De leur côté, les indices ZU et ZW s'éloignent légèrement plus de la loi $N(0,1)$ lorsque θ est au-dessus de zéro. Par exemple, l'écart-type de ZU est égal à 0,96 lorsque $\theta = -1$ et il descend à 0,44 lorsque $\theta = 2$. De son côté, l'écart-type de ZW est égal à 0,99 lorsque $\theta = -1$ et il descend à 0,59 lorsque $\theta = 2$.

L'étude des coefficients d'asymétrie et de kurtose permet de mettre en évidence le fait que les indices ZW^* et ZW sont légèrement plus près de la $N(0,1)$ que les indices l_z et l_z^* . Pour $\theta \leq 1$, les coefficients d'asymétrie des indices de type carré

moyens sont généralement positifs alors qu'ils sont tous négatifs pour les indices de type vraisemblance. De plus, ces coefficients sont beaucoup plus élevés pour les indices ZU et ZU^* , surtout lorsque θ augmente. Pour ZU , les coefficients d'asymétrie et de kurtose calculés sont respectivement égaux à 0,30 et 0,40 lorsque $\theta = -1$ et 3,58 et 32,48 lorsque $\theta = 2$. Les résultats sont aussi étonnants pour ZU^* : les coefficients d'asymétrie et de kurtose calculés sont respectivement égaux à 0,57 et 0,99 lorsque $\theta = -1$ et 21,92 et 833,08 lorsque $\theta = 2$. Comme nous l'avons observé aux figures 4.9 et 4.10, l'adéquation des scores des indices à la loi $N(0,1)$ est moins bonne pour $\theta = 1$ et $\theta = 2$. Ces résultats sont importants puisqu'ils montrent que les fonctions de densité de ZU et de ZU^* s'éloignent considérablement de la fonction de densité de la loi $N(0,1)$.

Tableau 4.6 : Coefficients de corrélation de Pearson et statistiques descriptives selon l'estimateur par maximum de vraisemblance pondérée et le modèle à trois paramètres (80 items)

	lz	lz*	ZU	ZU*	ZW	ZW*	M.(É.-T.)	A.(K.)
$\theta=-2$								
lz	1	1	-0,98	-0,98	-1	-0,99	0,10(0,96)	-0,24(0,06)
lz*	1	1	-0,98	-0,98	-0,99	-1	0,06(1,01)	-0,23(0,04)
ZU	-0,98	-0,98	1	1	0,96	0,96	-0,09(0,97)	0,15(0,00)
ZU*	-0,98	-0,98	1	1	0,96	0,96	-0,05(1,00)	0,35(0,15)
ZW	-1	-0,99	0,96	0,96	1	1	-0,10(0,97)	0,04(-0,01)
ZW*	-0,99	-1	0,96	0,96	1	1	-0,06(1,01)	0,20(0,02)
$\theta=-1$								
lz	1	1	-0,96	-0,95	-0,99	-1	0,08(0,98)	-0,18(0,00)
lz*	1	1	-0,96	-0,95	-0,99	-1	0,09(1,00)	-0,19(0,00)
ZU	-0,96	-0,96	1	1	0,93	0,93	-0,08(0,96)	0,30(0,40)
ZU*	-0,95	-0,95	1	1	0,92	0,92	-0,09(0,98)	0,57(0,99)
ZW	-0,99	-0,99	0,93	0,92	1	1	-0,08(0,99)	0,00(-0,04)
ZW*	-1	-1	0,93	0,92	1	1	-0,08(1,00)	0,16(0,01)
$\theta=0$								
lz	1	1	-0,85	-0,79	-0,99	-0,99	0,07(0,88)	-0,29(0,16)
lz*	1	1	-0,85	-0,78	-0,99	-0,99	0,08(0,99)	-0,30(0,14)
ZU	-0,85	-0,85	1	0,99	0,78	0,77	-0,05(0,81)	1,75(8,62)
ZU*	-0,79	-0,78	0,99	1	0,71	0,7	-0,07(0,96)	3,34(25,04)
ZW	-0,99	-0,99	0,78	0,71	1	1	-0,06(0,89)	0,07(0,01)
ZW*	-0,99	-0,99	0,77	0,7	1	1	-0,07(1,00)	0,23(0,08)
$\theta=1$								
lz	1	1	-0,71	-0,55	-0,97	-0,97	0,08(0,72)	-0,33(0,13)
lz*	1	1	-0,71	-0,55	-0,97	-0,97	0,06(1,00)	-0,31(0,08)
ZU	-0,71	-0,71	1	0,94	0,55	0,55	0,03(0,62)	3,11(23,65)
ZU*	-0,55	-0,55	0,94	1	0,41	0,4	-0,04(0,91)	9,66(180,54)
ZW	-0,97	-0,97	0,55	0,41	1	1	-0,05(0,71)	0,08(-0,07)
ZW*	-0,97	-0,97	0,55	0,4	1	1	-0,05(1,00)	0,24(0,00)
$\theta=2$								
lz	1	0,99	-0,58	-0,36	-0,95	-0,95	0,12(0,60)	-0,23(0,04)
lz*	0,99	1	-0,59	-0,36	-0,94	-0,95	0,04(1,00)	-0,19(-0,10)
ZU	-0,58	-0,59	1	0,81	0,4	0,4	0,24(0,44)	3,58(32,48)
ZU*	-0,36	-0,36	0,81	1	0,2	0,19	-0,03(0,84)	21,92(833,08)
ZW	-0,95	-0,94	0,4	0,2	1	0,99	-0,07(0,59)	-0,08(-0,04)
ZW*	-0,95	-0,95	0,4	0,19	0,99	1	-0,02(1,01)	0,11(-0,17)

4.1.2.2 Synthèse

Il est possible de regrouper les constats présentés dans cette section en six grands éléments. Premièrement, l'augmentation du nombre d'items a tendance à faire augmenter légèrement la valeur des coefficients de corrélation. De plus, l'augmentation du nombre d'items fait accroître l'adéquation de la moyenne, de l'écart-type, de l'asymétrie et de la kurtose aux valeurs de la $N(0,1)$.

Deuxièmement, il existe généralement une corrélation positive et forte entre l_z et l_z^* . Ensuite, la corrélation entre les indices ZU , ZU^* , ZW et ZW^* fluctue en fonction de θ et du modèle de réponse à l'item utilisé. De plus, ZU , ZU^* , ZW et ZW^* sont presque toujours négativement corrélés avec l_z et l_z^* . Il n'y a que l'indice ZU qui présente une corrélation positive et très faible avec l_z dans certaines situations.

Troisièmement, les indices standardisés sont bien corrélés avec leur version corrigée selon l'approche de Snijders (2001). En effet, nous retrouvons une corrélation positive et élevée entre l_z et l_z^* , une corrélation positive et élevée entre ZU et ZU^* et une corrélation positive et élevée entre ZW et ZW^* .

Quatrièmement, les coefficients de corrélation sont plus élevés lorsque θ tend vers zéro. À l'instar des résultats rapportés à l'étude des fonctions de densité des indices, cela suggère que les indices ne se comportent pas de la même façon aux différentes valeurs de θ . Il est donc important de sélectionner le bon indice lorsque nous souhaitons procéder à la détection de réponses inappropriées d'étudiants faibles ou d'étudiants forts.

Cinquièmement, la moyenne et l'écart-type des indices corrigés sont généralement plus près des valeurs de la loi $N(0,1)$. Ainsi, il semble qu'il serait plus approprié d'utiliser les indices corrigés l_z^* et ZW^* puisque leur fonction de densité est plus en accord avec les quantiles de la loi $N(0,1)$. De leur côté, les indices standardisés présentent des moyennes et des écarts-types qui s'éloignent de cette loi théorique lorsque $\theta = -2$ et $\theta = 2$. Cette catégorie d'indices semble donc plus appropriée pour détecter les patrons de réponses inappropriés des étudiants dont l'habileté est moyenne.

Sixièmement, les coefficients d'asymétrie sont négatifs pour les indices l_z et l_z^* , alors qu'ils sont généralement positifs pour ZU , ZU^* , ZW et ZW^* . Soulevons que l'asymétrie et la kurtose des indices de type carré-moyen ZU et ZU^* sont parfois très élevées ce qui nous fait croire qu'il faut être très prudent avec l'utilisation de ceux-ci. En ce qui a trait aux indices l_z , l_z^* , ZW et ZW^* , nous observons des résultats beaucoup plus encourageants puisque les coefficients de kurtose sont en accord avec la loi $N(0,1)$.

4.2 Étude 2 : analyse des erreurs de type I empiriques et théoriques

Nous présentons, dans cette section, une étude descriptive des erreurs de type I empiriques et théoriques pour les six indices investigués dans le cadre de cette thèse.

Les résultats seront successivement décrits pour deux longueurs de test : 30 items et 80 items.

4.2.1 Erreurs de type I : résultats pour 30 items

Dans toutes les figures des sections 4.2.1 et 4.2.2, la ligne continue diagonale représente la situation où les erreurs de type I théoriques et les erreurs de type I empiriques sont les mêmes. Toutes les valeurs au-dessus de cette ligne diagonale indiquent que les erreurs de type I empiriques surestiment les valeurs théoriques. À l'opposé, les valeurs en dessous de cette ligne diagonale indiquent que les erreurs de type I empiriques sous-estiment les erreurs de type I théoriques.

La figure 4.13 présente les résultats pour des données de 30 items analysées à l'aide du modèle de Rasch. Tout d'abord, nous observons qu'à la valeur $\theta = -2$, l'indice ZW^* présente une très légère sous-estimation des erreurs empiriques. C'est aussi cet indice qui s'approche le plus de la diagonale représentant l'égalité entre les erreurs de type I empiriques et les erreurs de type I théoriques. Ensuite, I_z^* surestime les erreurs empiriques aux seuils α bornés entre 0,00 et 0,08. De son côté, l'indice corrigé ZU^* présente une sous-estimation croissante des erreurs empiriques lorsque les seuils α passent de 0,02 à 0,10. Finalement, nous observons que les indices standardisés sous-estiment grandement les erreurs de type I empiriques. La supériorité des indices corrigés par la méthode de Snijders n'est pas étonnante,

puisque ces derniers présentaient aussi les moyennes et les écarts-types les plus en accord avec les valeurs théoriques de la $N(0,1)$ (voir le tableau 4.1).

Pour $\theta = -1$, l_z^* et ZW^* présentent une légère surestimation des erreurs de type I empiriques. Ce sont encore une fois ces indices qui présentent le plus d'équivalences entre les deux types d'erreurs analysés. Rappelons que l_z^* et ZW^* présentaient aussi une grande similitude avec les distributions théoriques au tableau 4.1. ZU^* , de son côté, sous-estime les erreurs empiriques. Il est à noter que cette sous-estimation s'accroît au fur et à mesure que le seuil α augmente. Les indices standardisés sous-estiment systématiquement les erreurs de type I empiriques, mais la différence est moins grande que dans la situation $\theta = -2$. Cette fois-ci, c'est l_z qui présente la moins grande disparité entre les erreurs empiriques et théoriques.

Dans le quadrant $\theta = 0$, les indices l_z^* et ZW^* sont ceux qui présentent le moins de différences entre les erreurs de type I empiriques et théoriques. Ensuite, les indices l_z et ZW sous-estiment systématiquement les erreurs de type I empiriques. De leur côté, ZU^* et ZU surestiment les erreurs empiriques entre les seuils α 0,00 à 0,025 et ils sous-estiment les erreurs empiriques aux seuils α supérieurs à 0,025.

Si nous observons les résultats pour le quadrant $\theta = 1$, c'est encore la supériorité des indices corrigés qui est mise en exergue. Ainsi, aux seuils $0 \leq \alpha \leq 0,065$, l_z^* , ZW^* et ZU^* surestiment légèrement les erreurs empiriques, alors qu'aux seuils $0,065 \leq \alpha \leq 0,1$, ils sous-estiment ce type d'erreur. Enfin, les indices

standardisés sous-estiment systématiquement les erreurs de type I empiriques. Il est important de rapporter que ZU est l'indice qui s'approche le plus de la droite diagonale d'égalité. C'est aussi l'indice standardisé qui avait la moyenne et l'écart-type le plus en accord avec la $N(0,1)$ dans le tableau 4.1.

À la valeur $\theta = 2$, nous observons que l_z^* présente des erreurs empiriques et théoriques relativement équivalentes aux seuils $0,0 \leq \alpha \leq 0,7$. Les deux autres indices corrigés surestiment généralement les erreurs empiriques. En ce qui a trait aux indices standardisés, se sont respectivement ZU , l_z et ZW qui présentent le moins de différences entre les erreurs empiriques et les erreurs théoriques. Notons, de plus, que ces indices sous-estiment systématiquement les erreurs de type I empiriques.

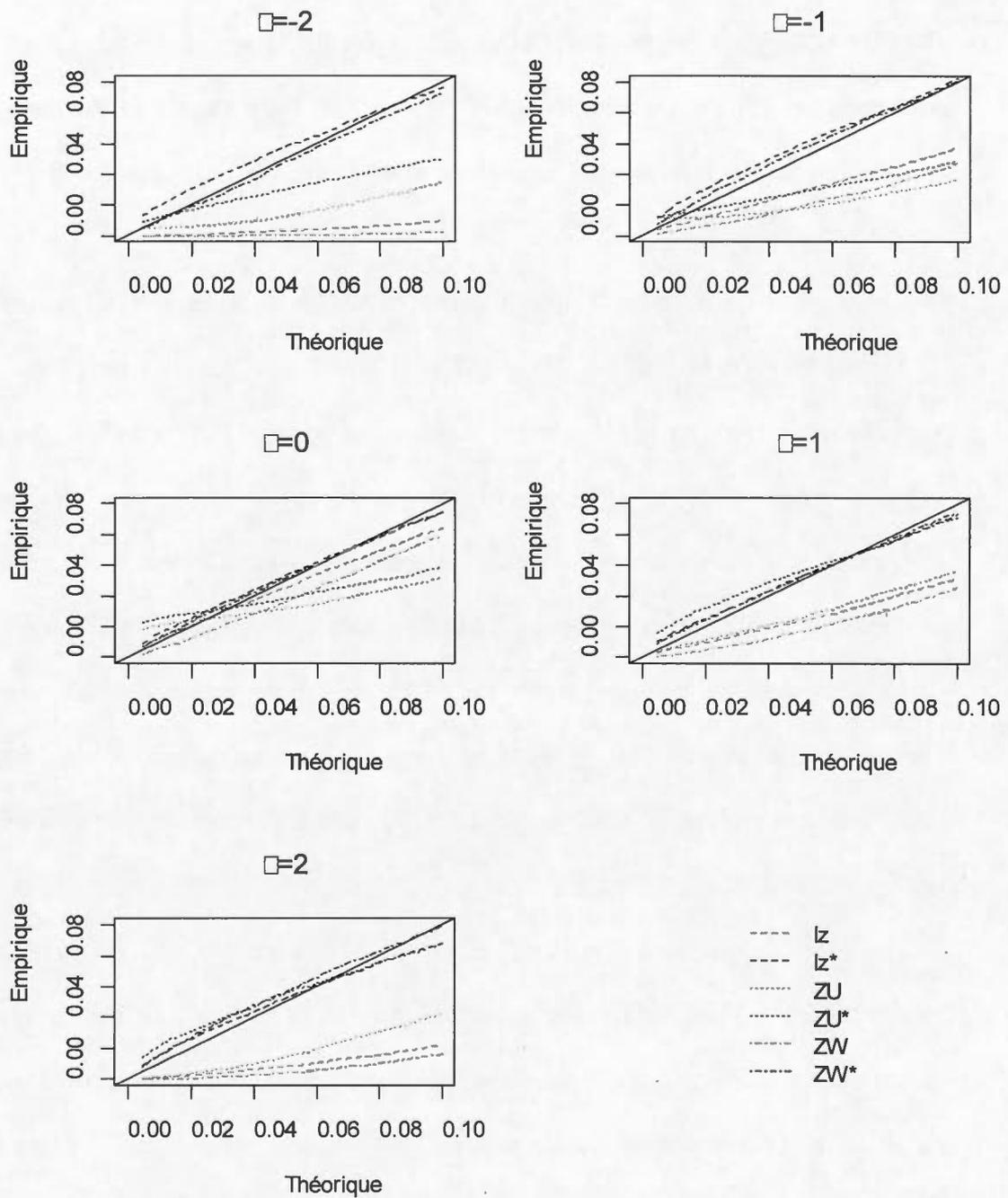


Figure 4.13 : Erreurs de type I empiriques et théoriques pour 30 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle de Rasch

La figure 4.14 présente les résultats pour les données analysées selon le modèle à deux paramètres. Nous observons que pour la valeur $\theta = -2$, ce sont les indices l_z^* et ZW^* qui présentent les erreurs de type I empiriques et théoriques les plus similaires. Les quatre autres indices sous-estiment grandement les erreurs de type I empiriques.

Pour $\theta = -1$, l_z^* et ZW^* surestiment légèrement les erreurs de type I empiriques et théoriques entre les seuils α 0,00 à 0,07. De son côté, ZU^* sous-estime les erreurs empiriques à tous les seuils α . En ce qui a trait aux indices standardisés, nous observons aussi une sous-estimation des erreurs empiriques. Par contre, la différence est moins grande qu'à $\theta = -2$. Enfin, notons que l_z est l'indice standardisé qui s'approche le plus de la droite diagonale d'égalité entre les erreurs empiriques et théoriques. Dans la première étude, les scores de ZU ne présentaient pas une très bonne adéquation à la $N(0,1)$. Nous ne sommes donc pas surpris de voir que cet indice est très éloigné de la diagonale d'égalité entre les erreurs de type I théoriques et empiriques.

Dans le quadrant $\theta = 0$, l_z^* et ZW^* sont encore une fois les indices qui s'approchent le plus de la droite d'égalité entre les erreurs de type I empiriques et théoriques. Ensuite, l_z et ZW sont les indices standardisés qui s'approchent le plus de la droite d'égalité entre les erreurs de type I empiriques et théoriques. À l'instar des résultats présentés au quadrant $\theta = -1$, ZU^* et ZU sont les indices qui s'éloignent le plus de la droite d'égalité aux seuils α supérieurs à 0,04.

Si nous observons les résultats pour $\theta = 1$, c'est la supériorité des indices corrigés qui est mise de l'avant. Dans ce cas-ci, l_z^* et ZW^* surestiment légèrement les erreurs de type I empiriques. Tous les autres indices sous-estiment les erreurs de type I empiriques.

Tous les indices sous-estiment les erreurs de type I empiriques au sein du quadrant $\theta = 2$. Dans cette situation d'analyse, nous observons que ce sont respectivement les indices l_z^* , ZW^* , ZU^* , l_z , ZW et ZU qui s'approchent le plus de la droite d'égalité entre les erreurs de type I empiriques et les erreurs de type I théoriques.

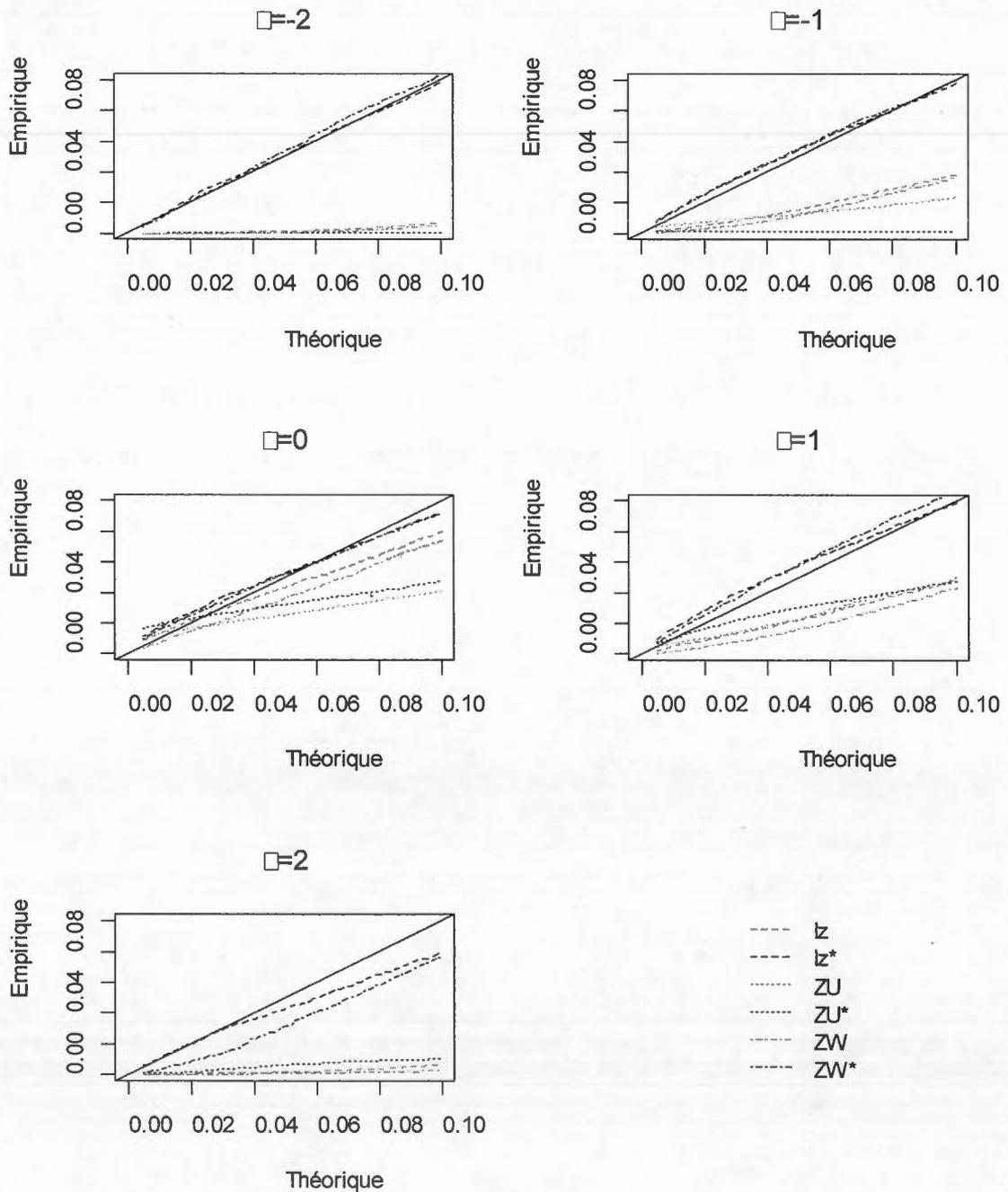


Figure 4.14: Erreurs de type I empiriques et théoriques pour 30 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle à deux paramètres

La figure 4.15 présente les résultats pour le modèle à trois paramètres. Pour la valeur $\theta = -2$, ce sont les indices l_z^* et ZW^* qui offrent les erreurs de type I empiriques et théoriques les plus similaires. L'indice ZU^* présente une surestimation systématique des erreurs empiriques. Du côté des indices standardisés, nous observons une sous-estimation systématique des erreurs de type I empiriques. Mentionnons que ZU est l'indice standardisé qui est le plus près de la droite d'égalité entre les erreurs empiriques et théoriques. Cela est attendu puisque, pour cette valeur de θ , cet indice présente une moyenne et un écart-type s'approchant de façon appréciable de ceux de la $N(0,1)$.

À la valeur $\theta = -1$, l_z^* , ZW^* et ZU^* présentent une légère surestimation des erreurs de type I empirique aux seuils $0,0 \leq \alpha \leq 0,8$. En ce qui a trait aux indices standardisés, nous observons plutôt qu'ils ont tendance à sous-estimer les erreurs empiriques. Par contre, la différence est moins grande que dans la situation présentée au quadrant $\theta = -2$. Notons, enfin, que ZU est l'indice standardisé qui s'approche le plus de la droite diagonale d'égalité entre les erreurs empiriques et théoriques. Cela est peu surprenant puisque, pour cette valeur de θ , l'indice présente une moyenne et un écart-type s'approchant de façon appréciable de ceux de la $N(0,1)$.

Dans le quadrant $\theta = 0$, les indices l_z^* et ZW^* sont encore ceux qui s'approchent le plus de la diagonale d'égalité entre les erreurs de type I empiriques et théoriques. Tous les autres indices sous-estiment progressivement les erreurs empiriques lorsque le seuil α augmente.

Si nous observons les résultats pour $\theta = 1$, c'est la supériorité des indices corrigés qui est mise de l'avant. Dans ce cas-ci, l_z^* et ZW^* surestiment légèrement les erreurs de type I empiriques pour tous les seuils α . Enfin, les autres indices sous-estiment grandement les erreurs empiriques et ZU est l'indice le plus éloigné de la droite diagonale d'égalité entre les erreurs.

Tous les indices sous-estiment les erreurs empiriques au sein du quadrant $\theta = 2$. Dans cette situation, nous observons que ce sont respectivement les indices ZW^* et l_z^* qui présentent le plus de similitudes entre les erreurs empiriques et théoriques. Les autres indices sous-estiment grandement les erreurs de type I empiriques.

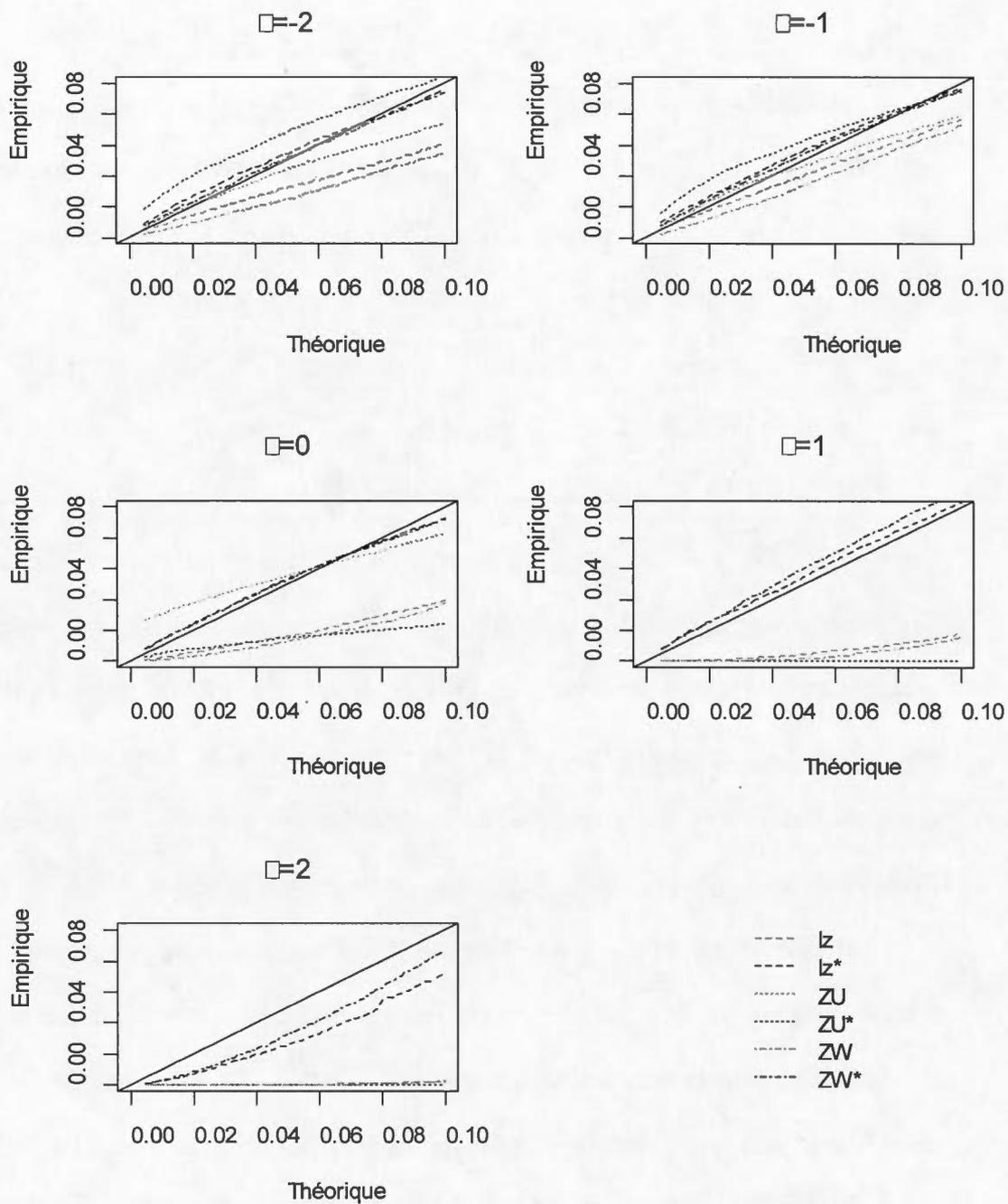


Figure 4.15 : Erreurs de type I empiriques et théoriques pour 30 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle à trois paramètres

4.2.2 Erreurs de type I : résultats pour 80 items

La figure 4.16 présente les résultats pour 80 items et des données qui ont été analysées à l'aide du modèle de Rasch. Avant de formellement entreprendre la description de cette figure, nous devons souligner que les tendances générales qui ont été observées à la figure 4.13 se répètent dans les lignes qui suivent.

Nous observons qu'au quadrant $\theta = -2$, ce sont les indices l_z^* et ZW^* qui s'approchent le plus de la ligne continue diagonale qui représente la situation où les erreurs de type I théoriques et les erreurs de type I empiriques sont les mêmes. En ce qui concerne l'indice ZU^* , nous remarquons qu'il surestime les erreurs empiriques aux seuils α inférieurs à 0,08. De leur côté, les indices standardisés sous-estiment grandement les erreurs de type I empiriques. Il est à noter que ZU est l'indice standardisé qui s'approche le plus de la droite diagonale d'égalité entre les deux types d'erreurs analysés. C'est aussi l'indice standardisé qui présentait la moyenne et l'écart-type les plus près des valeurs de la loi normale (voir le tableau 4.4).

Pour $\theta = -1$, l_z^* et ZW^* présentent une légère surestimation des erreurs de type I empiriques aux seuils α inférieurs à 0,06 et une légère sous-estimation des erreurs de type I empiriques aux seuils α supérieurs à 0,06. ZU^* surestime les erreurs empiriques pour presque tous les seuils α étudiés. De leur côté, les indices standardisés sous-estiment encore une fois les erreurs empiriques. À l'instar des résultats pour $\theta = -2$, ZU est l'indice standardisé qui s'approche le plus de la droite diagonale d'égalité entre les erreurs de type I empiriques et les erreurs de type I

théoriques. Cela n'est pas surprenant puisque les scores de cet indice présentent généralement des résultats en accord avec la $N(0,1)$ lorsque les probabilités sont estimées à l'aide du modèle de Rasch (voir les résultats à l'étude 1).

Dans le quadrant $\theta = 0$, les indices corrigés et les indices standardisés s'approchent de façon appréciable de la droite diagonale d'égalité entre les erreurs de type I empiriques et théoriques. Notons que parmi tous les indices, ce sont l_z^* et ZW^* qui présentent le plus de similitudes entre ces deux types d'erreurs.

Voici les résultats pour $\theta = 1$. Ce sont les indices l_z^* et ZW^* qui présentent le plus de similitudes entre leurs erreurs de type I empiriques et théoriques. ZU^* , de son côté, surestime systématiquement les erreurs empiriques. En ce qui concerne les indices standardisés, il est possible d'observer une sous-estimation systématique des erreurs de type I empiriques. Enfin, il est à noter que ZU est l'indice standardisé qui s'approche le plus de la droite diagonale d'égalité entre les deux types d'erreurs. Ce résultat est en accord avec les moyennes et les écarts-types près des valeurs de la $N(0,1)$ présentés au tableau 4.4.

Les résultats généraux pour $\theta = 2$ ressemblent beaucoup à ceux de $\theta = 1$. Ce sont encore les indices l_z^* et ZW^* qui présentent le moins de différences entre les erreurs de type I empiriques et théoriques. À l'instar des résultats présentés pour les autres quadrants de la figure 4.16, ZU^* surestime les erreurs empiriques. De leur côté, les indices standardisés sous-estiment systématiquement les erreurs de type I empiriques.

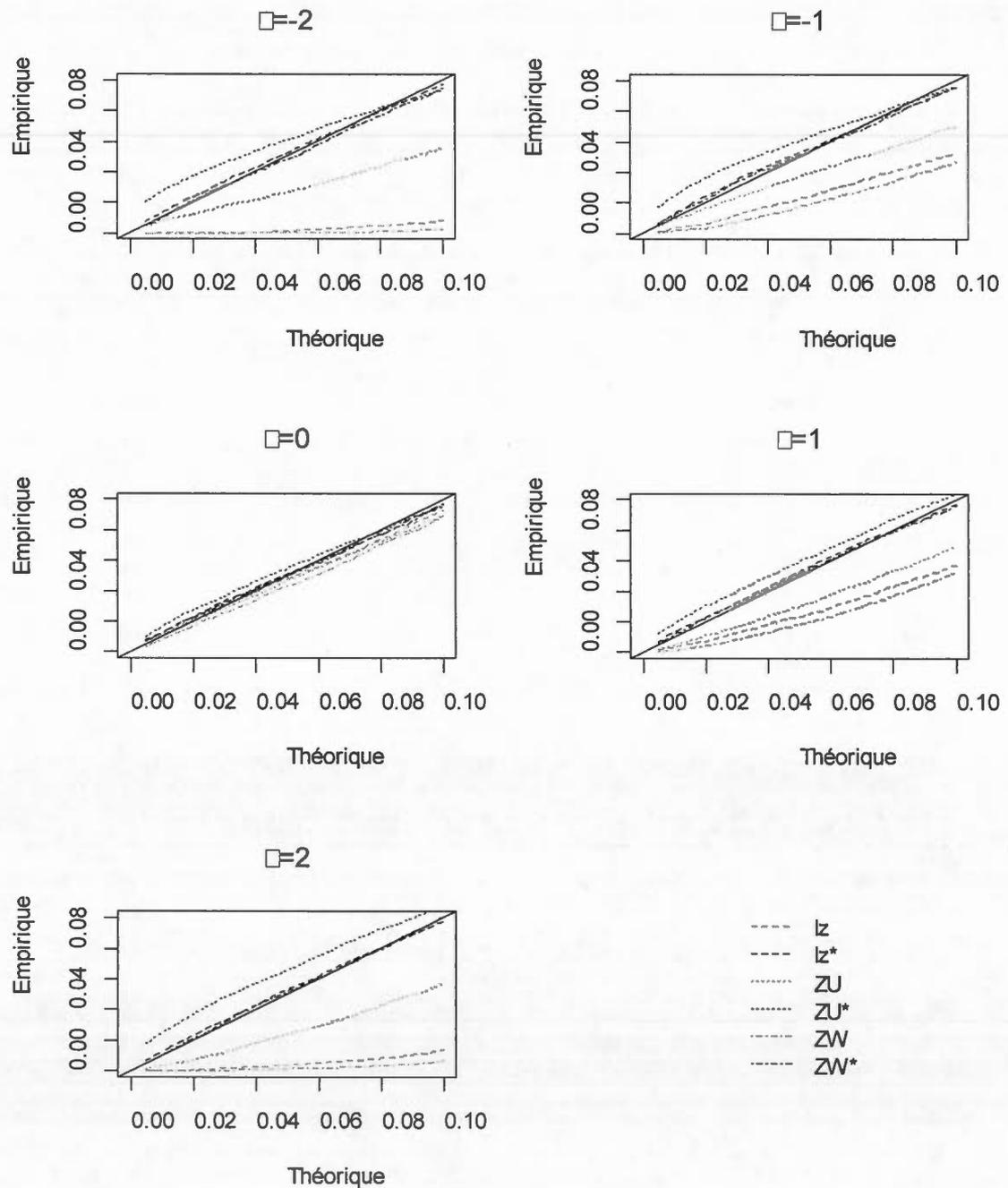


Figure 4.16: Erreurs de type I empiriques et théoriques pour 80 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle de Rasch

Nous présentons maintenant les résultats pour le modèle à deux paramètres (figure 4.17). Tout d'abord, il est important de souligner que les principaux résultats présentés à la figure 4.14 sont aussi observés dans les lignes qui suivent.

À la valeur $\theta = -2$, ce sont les indices I_z^* et ZW^* qui s'approchent le plus de la droite d'égalité entre les erreurs de type I empiriques et théoriques. ZU présentent une sous-estimation croissante des erreurs empiriques lorsque α augmente. Tous les autres indices sous-estiment aussi les erreurs empiriques.

Pour $\theta = -1$, ce sont les indices corrigés qui présentent la meilleure adéquation entre les valeurs des erreurs théoriques et des erreurs empiriques. En ce qui a trait aux indices standardisés, nous observons une sous-estimation des erreurs empiriques. Enfin, ZU est l'indice standardisé qui s'approche le plus de la droite diagonale d'égalité des erreurs empiriques et théoriques. C'est aussi l'indice standardisé qui présentait la moyenne et l'écart-type les plus près des valeurs de la loi normale (voir le tableau 4.5).

Dans le quadrant $\theta = 0$, les indices s'approchent généralement tous de la droite diagonale d'égalité entre les erreurs de type I empiriques et théoriques. Par contre, nous remarquons que les indices standardisés ont tendance à sous-estimer légèrement les erreurs de type I empiriques.

Les résultats pour $\theta = 1$ suivent le schéma suivant : c'est encore la supériorité des indices corrigés I_z^* et ZW^* qui est mise de l'avant. En effet, tous les autres indices affichent une sous-estimation des erreurs de type I empiriques. Notons, ici, que ce

sont respectivement ZU^* , l_z , ZW et ZU qui s'approchent le plus de la droite diagonale d'égalité entre les erreurs de type I empiriques et théoriques.

Enfin, ce sont les indices l_z^* et ZW^* qui présentent le plus de similitudes entre les erreurs de type I empiriques et théoriques au sein du quadrant $\theta = 2$. Viennent ensuite tous les autres indices, qui sous-estiment les erreurs de type I empiriques.

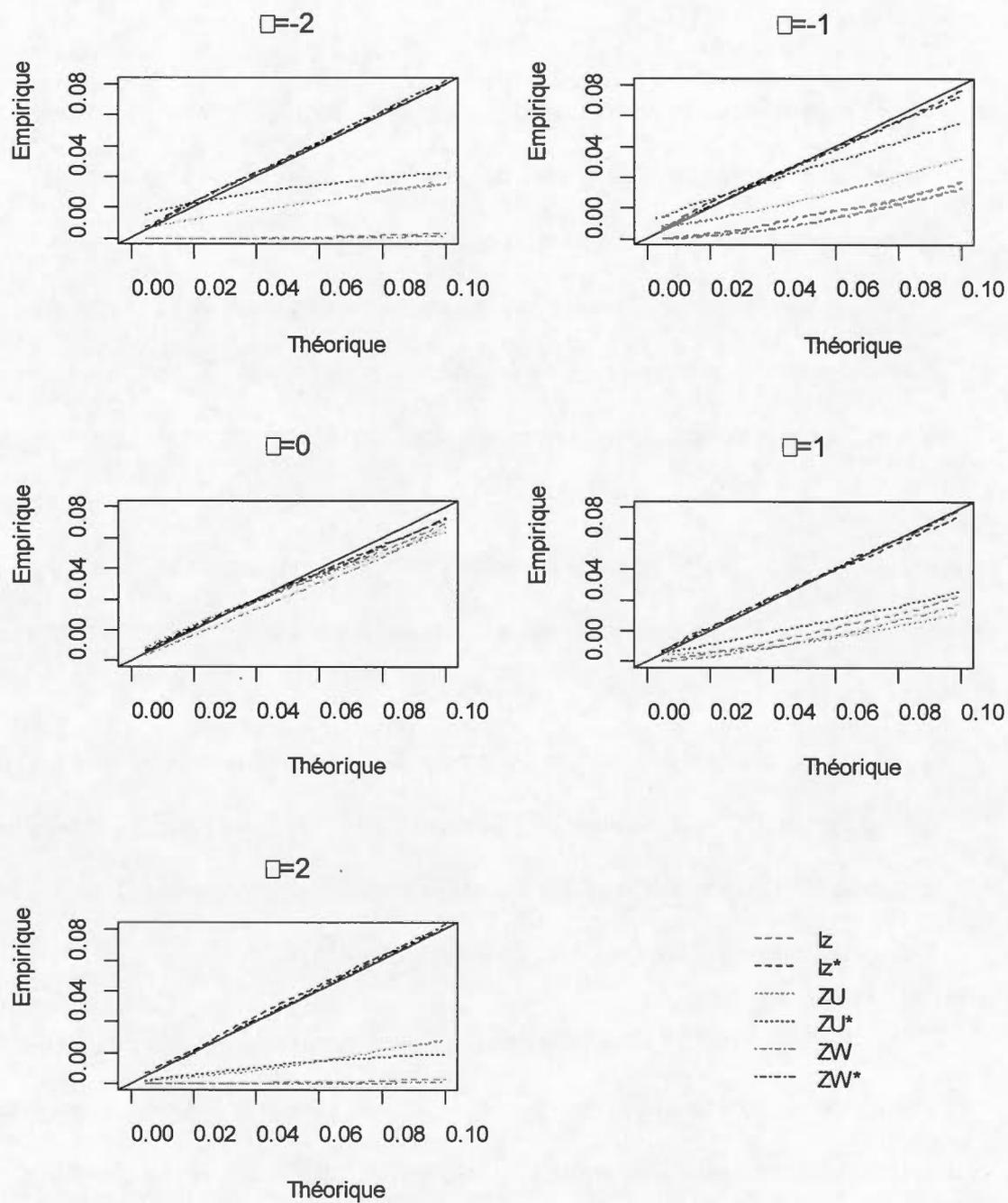


Figure 4.17: Erreurs de type I empiriques et théoriques pour 80 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle à deux paramètres

La figure 4.18 présente les résultats de données analysées selon le modèle à trois paramètres.

Pour $\theta = -2$, ce sont les indices corrigés qui sont le plus près de la droite diagonale d'égalité entre les erreurs de type I empiriques et théoriques. Du côté des indices standardisés, les erreurs empiriques sont légèrement sous-estimées.

À la valeur $\theta = -1$, tous les indices se rapprochent de la droite diagonale d'égalité entre les erreurs de type I empiriques et théoriques. Notons que les indices sous-estiment les erreurs de type I empiriques lorsque α est plus grand que 0,04.

Dans le quadrant $\theta = 0$, l_z^* et ZW^* sont encore les indices qui s'approchent le plus de la droite d'égalité entre les erreurs de type I empiriques et théoriques. Tous les autres indices sous-estiment progressivement les erreurs empiriques lorsque le seuil α augmente.

Si nous observons les résultats pour $\theta = 1$, c'est la supériorité des indices corrigés l_z^* et ZW^* qui est mise en évidence (ils présentent une meilleure adéquation à la droite diagonale d'égalité des erreurs empiriques et théoriques). Tous les autres indices sous-estiment les erreurs de type I empiriques.

Enfin, l_z^* et ZW^* sont les indices qui présentent le plus de similitudes entre les erreurs de type I empiriques et théoriques au quadrant $\theta = 2$. Dans cette situation, nous observons que ce sont ensuite les indices ZU^* et ZU qui s'approchent le plus de la droite diagonale d'égalité entre les erreurs empiriques et théoriques. Les deux autres indices sous-estiment largement les erreurs de type I empiriques.

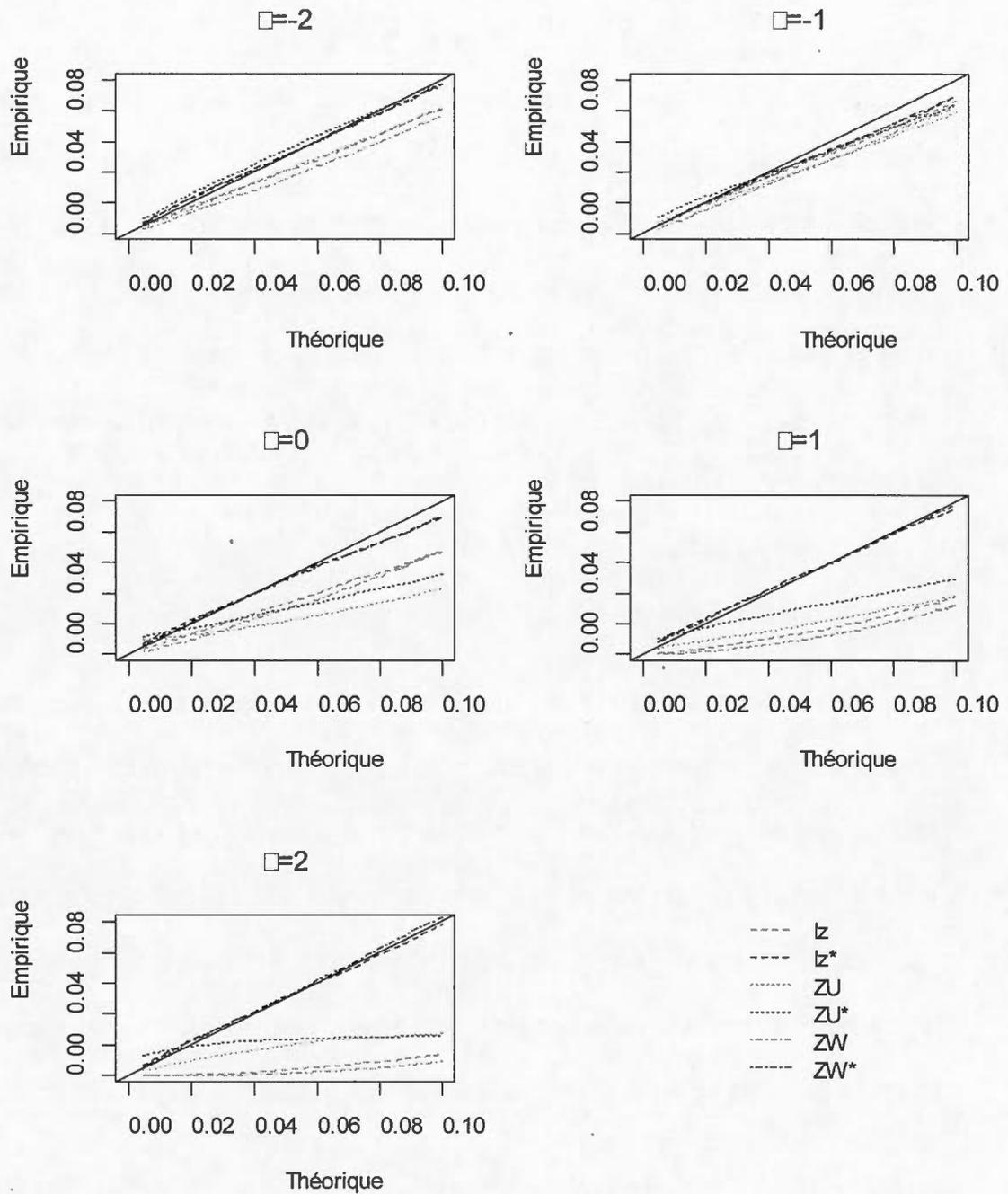


Figure 4.18 : Erreurs de type I empiriques et théoriques pour 80 items, selon l'estimateur par maximum de vraisemblance pondérée et le modèle à trois paramètres

4.2.3 Synthèse

Plusieurs constats émergent de la section 4.2. Premièrement, dans la quasi-totalité des situations étudiées, ce sont les indices corrigés I_z^* et ZW^* qui présentent le plus de similitudes entre leurs erreurs de type I empiriques et théoriques. Ces indices sont aussi ceux qui étaient le plus en accord avec la fonction de densité de loi $N(0,1)$, dans la section 4.1. Ils présentent donc des propriétés favorables à leur utilisation. Ensuite, l'indice ZU^* présente généralement des erreurs de type I empiriques qui s'approchent des erreurs de type I théoriques lorsque les données sont analysées à l'aide du modèle de Rasch. Par contre, la similitude entre ces deux types d'erreurs est légèrement moins importante pour les modèles à deux et à trois paramètres.

Deuxièmement, l'étude des indices standardisés montre que ZU présente des erreurs de type I empiriques et des erreurs de type I théoriques relativement similaires dans plusieurs situations d'analyse uniquement. Cela n'est pas surprenant puisque nous avons observé des moyennes et des écarts-types près de ceux de la $N(0,1)$ dans plusieurs situations d'analyses de la première étude. Par contre, l'observation de coefficients d'asymétrie et de kurtose élevés montre que cet indice a une propension élevée à avoir des scores extrêmes; ce qui peut compliquer son utilisation.

Troisièmement, nous observons que lorsque les données sont analysées à l'aide du modèle de Rasch (figures 4.13 pour 30 items et 4.16 pour 80 items) et du modèle à deux paramètres (figures 4.14 pour 30 items et 4.17 pour 80 items), les

indices ont tendance à se rapprocher grandement de la droite diagonale d'égalité entre leurs erreurs empiriques et leurs erreurs théoriques pour $\theta = 0$. Dans le cas du modèle à trois paramètres (figures 4.15 et 4.18), nous remarquons plutôt que les indices se rapprochent surtout de la droite diagonale d'égalité entre leurs erreurs empiriques et théoriques aux valeurs $\theta = -2$ et $\theta = -1$.

4.3 Étude 3 : analyse de la puissance des indices

Dans cette dernière section, nous présentons une étude de la puissance des indices. Les résultats sont exposés successivement pour une longueur de test de 30 items et de 80 items.

4.3.1 Puissance des indices : résultats pour 30 items

Cette section présente les résultats pour l'analyse des scores des indices selon 30 items. Les résultats pour la détection de la réponse au hasard sont présentés avant les résultats pour la détection de l'inattention.

4.3.1.1 Réponse au hasard

Les tableaux 4.7a et 4.7b présentent les résultats pour la détection de la réponse au hasard à l'aide de l'indice I_z . D'entrée de jeu, nous observons que lorsque le niveau de réponse au hasard C augmente, le pourcentage de détection de patrons de réponses au hasard augmente lui aussi. Par exemple, pour le modèle de Rasch, $\alpha = 0,01$ et la valeur $\theta = -1$, I_z détecte 34,42% des patrons à $C = 0,10$, 64,48% à $C = 0,20$ et 81,59 à $C = 0,30$. Ce résultat est prévisible puisqu'augmenter la valeur de C augmente aussi la probabilité d'en faire la détection. Fait intéressant, les pourcentages de détection sont plus élevés lorsque le modèle à deux paramètres est utilisé alors qu'ils sont moins élevés lorsque le modèle à trois paramètres est utilisé. Ensuite, lorsque θ augmente, nous remarquons que le pourcentage de patrons détectés tend à diminuer. Cela est logique puisque les individus ayant une valeur θ élevée n'ont pas besoin de tricher : ils connaissent la bonne réponse à un item et ainsi l'impact du paramètre de pseudo-chance personnel C est moins important. Comme nous pouvions l'anticiper, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$.

En ce qui a trait aux statistiques descriptives, il est possible de remarquer que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de I_z est négativement élevée et plus l'écart-type est grand. Ce constat est cohérent puisque le score de I_z est négatif lorsqu'un patron de réponses est

inapproprié. Ainsi, plus une matrice de données va contenir un grand nombre de patrons de réponses inappropriés, plus la moyenne des scores de cet indice va être négativement élevée.

Dans le cas des analyses selon le modèle de Rasch et selon le modèle à deux paramètres, nous observons que la moyenne et l'écart-type se rapprochent des valeurs de la loi $N(0,1)$ au fur et à mesure que les valeurs θ augmentent. Cela peut être expliqué par le fait que l'impact du paramètre C diminue avec l'augmentation de θ puisqu'avec des valeurs élevées, l'effet du hasard est moins important. Notons, de plus, que les moyennes et les écarts-types selon le modèle à trois paramètres sont ceux qui sont les plus près des valeurs de la $N(0,1)$.

Nous observons, pour le modèle de Rasch et pour le modèle à deux paramètres, que lorsque le niveau de réponse au hasard C augmente, le coefficient d'asymétrie a tendance à s'approcher de zéro, soit, la valeur prévue pour la loi $N(0,1)$. Ce constat ne s'applique cependant pas au modèle à trois paramètres lorsque θ est au-dessus de zéro. De plus, notons que l'augmentation de θ entraîne généralement une augmentation de la valeur du coefficient de kurtose, qui est relativement importante avec le modèle à trois paramètres.

Tableau 4.7a: Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	57,47	68,36	-2,97	2,4	-0,46	0
	C=0,2	85,55	91,23	-5,16	2,64	-0,26	0
	C=0,3	94,97	97,37	-6,6	2,67	-0,17	-0,06
$\theta=-1$	C=0,1	34,42	48,17	-1,74	1,83	-0,57	0,23
	C=0,2	64,48	76,01	-3,22	2,08	-0,36	0,03
	C=0,3	81,59	89,57	-4,29	2,15	-0,26	-0,01
$\theta=0$	C=0,1	16,61	28,33	-0,91	1,43	-0,64	0,34
	C=0,2	33,07	48,83	-1,75	1,6	-0,5	0,14
	C=0,3	47,09	63,54	-2,34	1,69	-0,45	0,25
$\theta=1$	C=0,1	5,83	13,35	-0,32	1,16	-0,74	0,51
	C=0,2	10,15	21,27	-0,7	1,24	-0,59	0,24
	C=0,3	14,46	27,81	-1	1,29	-0,64	0,56
$\theta=2$	C=0,1	2,23	5,92	0,03	0,95	-0,98	1,29
	C=0,2	2,87	7,27	-0,1	0,97	-1	1,43
	C=0,3	2,95	7,78	-0,18	0,96	-0,93	1,11
2 paramètres							
$\theta=-2$	C=0,1	58,36	67,74	-3,6	3,27	-0,88	0,57
	C=0,2	85,79	90,9	-6,46	3,84	-0,49	-0,13
	C=0,3	95,38	97,47	-8,53	3,94	-0,29	-0,29
$\theta=-1$	C=0,1	40,38	51,3	-2,24	2,56	-0,98	0,82
	C=0,2	68,03	77,02	-4,09	3,01	-0,6	-0,02
	C=0,3	83,72	89,69	-5,44	3,07	-0,38	-0,25
$\theta=0$	C=0,1	25,09	34,79	-1,29	1,99	-1,04	0,99
	C=0,2	45,51	57,13	-2,43	2,3	-0,69	0,11
	C=0,3	60,54	71,47	-3,25	2,41	-0,49	-0,15
$\theta=1$	C=0,1	12,21	19,74	-0,53	1,47	-1,08	1,19
	C=0,2	22,23	32,89	-1,12	1,68	-0,83	0,46
	C=0,3	29,69	42,22	-1,5	1,74	-0,63	0,06
$\theta=2$	C=0,1	4,11	9,15	-0,06	1,08	-1,14	1,64
	C=0,2	5,9	12,59	-0,24	1,17	-1,04	1,15
	C=0,3	6,49	14,44	-0,34	1,18	-0,92	0,85

Tableau 4.7b: Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	C=0,1	7,45	18,91	-0,71	1,08	-0,3	0,04
	C=0,2	23,29	44,14	-1,52	1,16	-0,26	0,03
	C=0,3	45,27	67,2	-2,22	1,21	-0,22	0,02
$\theta=-1$	C=0,1	4,39	12,83	-0,48	1,01	-0,35	0,1
	C=0,2	13,09	29,23	-1,09	1,09	-0,3	0,13
	C=0,3	25,64	47,53	-1,63	1,15	-0,33	0,22
$\theta=0$	C=0,1	2,19	7,73	-0,19	0,95	-0,54	0,4
	C=0,2	5,71	15,2	-0,6	1,02	-0,53	0,4
	C=0,3	10,53	24,28	-0,94	1,07	-0,5	0,3
$\theta=1$	C=0,1	1,46	4,59	0,02	0,88	-0,83	1,15
	C=0,2	2,35	6,73	-0,16	0,91	-0,83	0,98
	C=0,3	3,21	8,82	-0,3	0,93	-0,94	1,49
$\theta=2$	C=0,1	0,74	2,63	0,13	0,76	-1,02	1,82
	C=0,2	0,74	2,46	0,12	0,73	-1,11	2,03
	C=0,3	0,56	2,09	0,14	0,69	-1,18	2,38

Les tableaux 4.8a et 4.8b présentent les résultats pour la détection de la réponse au hasard à l'aide de l'indice I_z^* . Notons, d'abord, que les pourcentages de détection de I_z^* sont légèrement plus élevés que ceux de I_z . De plus, lorsque θ augmente, nous observons que le pourcentage de détection de I_z^* devient moins important. Comme nous l'avions expliqué précédemment, cela est logique puisque les individus ayant une valeur θ élevée n'ont pas besoin de tricher, car ils connaissent la bonne réponse à un item. Ainsi l'impact du paramètre de pseudo-chance personnel C est moins important

Nous examinons encore une fois que lorsque le niveau de réponse au hasard C augmente, le pourcentage de détection de patrons de réponses au hasard augmente lui

aussi. Il est, par contre, important de nuancer le fait que ce pourcentage diffère selon le modèle de réponse à l'item utilisé. En effet, le pourcentage de patrons de réponses détectés est à son maximum lorsque le modèle à deux paramètres est utilisé alors qu'il est à son minimum lorsque c'est le modèle à trois paramètres qui est utilisé. Ensuite, lorsque la valeur θ augmente, nous remarquons que le pourcentage de patrons détectés tend à diminuer. Cela peut être expliqué par le fait que la réponse au hasard touche surtout les individus qui ont une valeur θ plus faible.

Sans surprise, nous observons que le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$. Ce dernier constat est prévisible puisque la détection au niveau $\alpha = 0,05$ est moins limitative que la détection au niveau $\alpha = 0,01$.

Il est possible de remarquer que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de I_z^* est négativement élevée et plus l'écart-type est grand. Ainsi, plus une matrice de données va contenir un grand nombre de patrons de réponses inappropriés, plus la moyenne des scores de cet indice va être négativement élevée et plus les scores vont être étendus (ce qui va faire augmenter l'écart-type).

Dans le cas des analyses selon le modèle de Rasch et le modèle à deux paramètres, nous observons que la moyenne et l'écart-type se rapprochent des valeurs de la loi $N(0,1)$ au fur et à mesure que les valeurs θ augmentent. Notons, de plus, que

les moyennes et les écarts-types obtenus selon le modèle à trois paramètres sont ceux qui sont les plus près de la $N(0,1)$.

Nous observons, pour le modèle de Rasch et pour le modèle à deux paramètres, que lorsque le niveau de réponse au hasard C augmente, le coefficient d'asymétrie a tendance à diminuer. Ce constat ne s'applique toutefois pas au modèle à trois paramètres lorsque θ est au-dessus de zéro. Enfin, ajoutons que l'augmentation des valeurs de θ génère une augmentation de la valeur du coefficient d'asymétrie et du coefficient de kurtose.

Tableau 4.8a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z^* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta = -2$	C=0,1	58,45	68,85	-3,02	2,43	-0,45	-0,01
	C=0,2	85,77	91,31	-5,18	2,65	-0,26	0
	C=0,3	95,01	97,38	-6,61	2,67	-0,17	-0,06
$\theta = -1$	C=0,1	34,42	48,17	-1,75	1,83	-0,57	0,23
	C=0,2	64,49	76,01	-3,22	2,08	-0,36	0,03
	C=0,3	81,61	89,58	-4,3	2,16	-0,26	-0,01
$\theta = 0$	C=0,1	16,61	28,33	-0,91	1,43	-0,64	0,34
	C=0,2	33,13	48,85	-1,75	1,61	-0,5	0,14
	C=0,3	47,41	63,65	-2,35	1,7	-0,45	0,25
$\theta = 1$	C=0,1	6,01	13,46	-0,33	1,17	-0,74	0,51
	C=0,2	10,82	21,79	-0,72	1,26	-0,59	0,23
	C=0,3	15,93	29,4	-1,05	1,34	-0,65	0,58
$\theta = 2$	C=0,1	2,99	7,27	0	1,04	-0,99	1,33
	C=0,2	3,95	10,03	-0,16	1,09	-0,98	1,33
	C=0,3	4,7	11,43	-0,27	1,1	-0,9	0,99
2 paramètres							
$\theta = -2$	C=0,1	58,87	68,08	-3,64	3,29	-0,87	0,55
	C=0,2	85,93	90,99	-6,49	3,85	-0,49	-0,13
	C=0,3	95,41	97,48	-8,56	3,95	-0,29	-0,29
$\theta = -1$	C=0,1	40,43	51,38	-2,24	2,56	-0,98	0,82
	C=0,2	68,1	77,08	-4,1	3,02	-0,6	-0,01
	C=0,3	83,74	89,72	-5,46	3,09	-0,38	-0,25
$\theta = 0$	C=0,1	25,17	34,83	-1,3	1,99	-1,05	1
	C=0,2	45,67	57,24	-2,45	2,32	-0,7	0,11
	C=0,3	60,83	71,71	-3,3	2,45	-0,49	-0,15
$\theta = 1$	C=0,1	12,65	20,03	-0,54	1,5	-1,1	1,25
	C=0,2	23,25	33,59	-1,16	1,74	-0,85	0,47
	C=0,3	31,41	43,51	-1,58	1,83	-0,65	0,06
$\theta = 2$	C=0,1	5,03	10,32	-0,08	1,16	-1,15	1,64
	C=0,2	7,41	14,34	-0,28	1,26	-1,03	1,14
	C=0,3	8,44	17	-0,41	1,27	-0,9	0,79

Tableau 4.8b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z^* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	C=0,1	8,11	20,1	-0,73	1,11	-0,29	0,03
	C=0,2	25,74	46,33	-1,58	1,21	-0,26	0,05
	C=0,3	49,54	70,19	-2,36	1,27	-0,22	0,05
$\theta=-1$	C=0,1	6,08	15,62	-0,55	1,08	-0,34	0,09
	C=0,2	17,07	34,29	-1,22	1,17	-0,31	0,17
	C=0,3	32,02	53,53	-1,82	1,25	-0,36	0,31
$\theta=0$	C=0,1	3,55	10,03	-0,26	1,04	-0,55	0,43
	C=0,2	8,31	19,15	-0,71	1,12	-0,55	0,48
	C=0,3	15,16	30,55	-1,11	1,19	-0,53	0,47
$\theta=1$	C=0,1	2,77	7,05	-0,05	1,01	-0,86	1,22
	C=0,2	4,45	10,39	-0,26	1,07	-0,83	0,99
	C=0,3	6,32	13,99	-0,45	1,13	-0,91	1,44
$\theta=2$	C=0,1	2,25	6,01	0,06	0,98	-0,99	1,61
	C=0,2	2,47	6,25	0,01	0,98	-1,03	1,67
	C=0,3	2,38	6,13	-0,03	0,96	-1	1,61

Les tableaux 4.9a et 4.9b présentent les résultats pour ZU . Comme nous l'avons observé à la première étude, les coefficients d'asymétrie et de kurtose de cet indice ne sont pas toujours en harmonie avec les quantiles de la loi $N(0,1)$. Il faut donc être vigilant avec l'utilisation de ZU .

Pour des données analysées selon le modèle de Rasch, nous remarquons que le taux de détection de I_z et de I_z^* est plus élevé que celui de ZU . Pour le modèle à deux paramètres, ZU présente des pourcentages de détection excessivement élevés et qui sont au-dessus de ceux des indices de type vraisemblance. Rappelons que l'étendue des scores de cet indice était très grande dans l'étude 1. L'amplitude des moyennes, des écarts-types et des coefficients de kurtose va dans le même sens dans

le tableau 4.9a : ce qui démontre encore une fois que cet indice a une propension élevée à avoir des scores extrêmes. En ce qui concerne le modèle à trois paramètres, I_z et I_z^* présentent uniquement des pourcentages de détection plus élevés pour $\theta = -2$ et $\theta = -1$. C'est donc ZU qui présente les pourcentages de détection les plus élevés aux valeurs $\theta = 0$, $\theta = 1$ et $\theta = 2$.

En général, C augmente de pair avec le pourcentage de détection. L'explication est la même que précédemment : augmenter la valeur de C augmente aussi la probabilité d'en faire la détection. Le pourcentage de détection est encore à son niveau le plus élevé lorsque le modèle à deux paramètres est utilisé. De son côté, c'est le modèle de Rasch qui présente les pourcentages de détection les plus faibles. Ce résultat est expliqué par le fait que c'est ce modèle qui engendre le moins de scores extrêmes. De plus, nous observons que le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$.

Il est possible de remarquer que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne et l'écart-type des scores de ZU sont grands. Cela est particulièrement vrai dans le cas du modèle à deux paramètres qui présente des moyennes et des écarts-types excessivement élevés (un résultat aussi observé à la valeur $\theta = 1$ de la figure 4.3). Par exemple, pour la situation où $\theta = 2$ et $C = 0,3$, nous obtenons respectivement une moyenne et un écart-type de 5444,5 et de 13 102,38. Ces résultats sont très loin de ceux de la loi $N(0,1)$ et indiquent que ZU semble peu approprié pour être utilisé dans un contexte réel.

Nous observons, pour le modèle de Rasch, des coefficients d'asymétrie et de kurtose près de la loi $N(0,1)$ lorsque le niveau θ est plus petit que zéro. Lorsque le niveau θ est au-dessus de zéro, les coefficients d'asymétrie et de kurtose ont tendance à s'éloigner des valeurs de la loi $N(0,1)$. Les coefficients d'asymétrie et de kurtose calculés à partir du modèle à deux paramètres et du modèle à trois paramètres sont très élevés et éloignés de la loi $N(0,1)$. Par exemple, les coefficients d'asymétrie pour le modèle à deux paramètres sont bornés entre 2,04 et 36,55. De leur côté, les coefficients d'asymétrie pour le modèle à trois paramètres sont bornés entre 0,33 et 17,74. Les coefficients de kurtose pour le modèle à deux paramètres et le modèle à trois paramètres sont respectivement bornés entre 11,38 et 2408,35 et entre 0,44 et 643,62, soit, des valeurs très élevées. Enfin, de tels résultats ont aussi été observés à la figure 4.3.

Tableau 4.9a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	41,49	55,95	2,4	1,24	0,34	-0,97
	C=0,2	62,81	76,43	3	1,24	-0,14	-0,84
	C=0,3	73,75	86,96	3,32	1,12	-0,27	-0,47
$\theta=-1$	C=0,1	25,09	36,5	1,6	1,28	0,48	-0,76
	C=0,2	35,78	53,27	2,11	1,19	0,11	-0,65
	C=0,3	32,79	53,87	2,16	1,04	0,26	-0,33
$\theta=0$	C=0,1	4,31	11	0,76	0,91	0,72	0,41
	C=0,2	4,35	11,17	1,04	0,79	0,7	1,25
	C=0,3	3,79	10,63	1,19	0,66	0,97	2,58
$\theta=1$	C=0,1	0,71	1,85	0,71	0,52	1,07	4,55
	C=0,2	1,06	3,45	0,95	0,52	1,27	7,58
	C=0,3	1,91	7,79	1,2	0,55	1,17	3,73
$\theta=2$	C=0,1	5,91	20,45	1,61	0,67	1,46	5,56
	C=0,2	11,16	32,55	1,85	0,8	1,94	9,85
	C=0,3	19,14	46,48	2,15	0,98	2,44	11,93
2 paramètres							
$\theta=-2$	C=0,1	99,99	100	23,6	16,57	5,9	115,16
	C=0,2	99,7	100	13,65	9,69	2,04	11,38
	C=0,3	99,35	100	9,08	10,56	27,4	1338,45
$\theta=-1$	C=0,1	94,15	100	5,32	2,83	4,02	37,01
	C=0,2	92,67	100	7,03	25,92	32,09	1247,1
	C=0,3	94,07	100	20,56	108,47	36,55	2408,35
$\theta=0$	C=0,1	86,03	100	17,85	57,12	17,05	455,95
	C=0,2	95,8	100	64,89	194,05	8,37	104,03
	C=0,3	98,85	100	192,66	460,29	6,05	58,06
$\theta=1$	C=0,1	99,96	100	290,33	546,27	9,87	247,01
	C=0,2	99,99	100	609,27	1262,15	19,99	846,88
	C=0,3	100	100	1182,33	2277,63	14,33	373,64
$\theta=2$	C=0,1	100	100	2166,75	4390,46	28,31	1395,68
	C=0,2	100	100	3498,19	7678,07	15,4	390,43
	C=0,3	100	100	5444,55	13102,38	11,76	191,38

Tableau 4.9b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	C=0,1	1,53	5,95	0,48	0,9	0,33	0,44
	C=0,2	4,79	14,33	0,96	0,92	0,71	1,15
	C=0,3	13,06	26,35	1,43	1,54	10,51	279,09
$\theta=-1$	C=0,1	0,95	2,22	0,31	0,68	1,58	8,46
	C=0,2	4,42	7,21	0,73	1,05	6,35	95,41
	C=0,3	15,1	19,47	1,52	2,99	17,74	643,62
$\theta=0$	C=0,1	8,13	10,94	0,96	1,27	6,7	102,01
	C=0,2	20,32	24,95	1,7	2,43	6,73	78,12
	C=0,3	39,08	44,41	3,25	5,69	9,33	166,7
$\theta=1$	C=0,1	52,18	63,16	3,38	3,89	7,96	146,29
	C=0,2	67,61	76,52	5,06	6,94	7,22	96,96
	C=0,3	82,49	88,14	8,24	11,82	5,53	47,83
$\theta=2$	C=0,1	98,18	98,93	12,58	13,53	4,9	37,07
	C=0,2	99,11	99,46	17,73	20,21	3,94	19,57
	C=0,3	99,63	99,8	24,67	26,72	2,92	9,49

Les tableaux 4.10a et 4.10b présentent les résultats pour ZU^* . Pour le modèle de Rasch et le modèle à deux paramètres, cet indice présente des taux de détection moins élevés que ceux de I_z , I_z^* et ZU . Par contre, pour les données analysées à l'aide du modèle à trois paramètres, le pourcentage de détection de ZU^* est plus élevé que celui des autres indices aux valeurs $\theta = -2$ et $\theta = -1$.

Dans une situation où les données sont analysées à l'aide du modèle de Rasch, le niveau de réponse au hasard C augmente généralement de pair avec le pourcentage de détection de patrons de réponses au hasard. Par contre, cette relation n'est pas systématiquement observée pour les modèles à deux et à trois paramètres. Par exemple, le pourcentage de détection diminue lorsque la valeur de C augmente dans

le cadre des analyses selon le modèle à trois paramètres. Il est à noter que cette relation a aussi été observée pour les valeurs $\theta = 1$ et $\theta = 2$ lorsque les données sont analysées à l'aide du modèle à deux paramètres. Rappelons que c'est sous cette modélisation que l'indice ZU^* présentait les moyennes et les écarts-types les moins en accord avec ceux de la $N(0,1)$. Ensuite, le pourcentage de détection est à son niveau le plus élevé lorsque le modèle à trois paramètres est utilisé. À l'opposé, c'est le modèle à deux paramètres qui présente les pourcentages de détection les plus faibles (ce modèle est aussi celui qui présentait les moyennes et les écarts-types les moins en accord avec la $N(0,1)$, au tableau 4.2). Ajoutons, de plus, que nous observons que le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$. Ce résultat est évidemment prévisible puisque le seuil $\alpha = 0,05$ est moins restrictif.

Observons maintenant les statistiques descriptives. Nous remarquons que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne et l'écart-type des scores de ZU^* sont aussi élevés. Ce résultat est logique puisque les patrons de réponses inappropriés ont des valeurs positives avec cet indice. De plus, les résultats sont très élevés pour $\theta = -2$ et $\theta = -1$, ce qui montre que cet indice ne suit pas exactement les quantiles de la loi $N(0,1)$.

Nous observons, pour cet indice, des coefficients d'asymétrie et de kurtose très élevés et éloignés de ceux prescrits pour la loi $N(0,1)$. Par exemple, pour le modèle à deux paramètres, la valeur $\theta = 2$ et $C = 0,3$, le coefficient d'asymétrie est

égal à 63,34 et le coefficient de kurtose à 4734,75. Pour le modèle à trois paramètres, la valeur $\theta = 2$ et $C = 0,1$, le coefficient d'asymétrie est égal à 19,65 et le coefficient de kurtose à 537,16. Ces résultats ne sont pas étonnants au regard de la moyenne et de l'écart-type de cet indice pour cette situation de simulation. Encore une fois, ces résultats montrent de façon explicite que cet indice est difficilement interprétable à partir des quantiles d'une loi $N(0,1)$.

Tableau 4.10a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	34,81	34,81	194,73	729,68	6,64	80,56
	C=0,2	58,57	58,71	195,83	526,03	4,75	36,48
	C=0,3	75,93	76,41	159,78	361,29	4,27	29,03
$\theta=-1$	C=0,1	34,68	35,46	43,28	161,09	5,8	44,83
	C=0,2	55,65	58,03	54,38	147,65	4,52	28,14
	C=0,3	66,17	70,15	41,96	98,35	4,48	32,59
$\theta=0$	C=0,1	23,77	27,67	7,79	32,53	7,42	70,68
	C=0,2	30,51	35,45	7,93	26,77	6,77	68,15
	C=0,3	29,75	34,95	7,49	27,31	8,82	114,04
$\theta=1$	C=0,1	6,36	7,62	1,47	11,35	20,92	711,33
	C=0,2	6,76	8,06	2,39	22,08	20,75	602,57
	C=0,3	7,13	8,21	2,9	23,43	14,16	243,41
$\theta=2$	C=0,1	2,19	2,33	1,6	25,41	24,63	732,68
	C=0,2	2,11	2,2	2,13	36,58	31,19	1326,54
	C=0,3	2,04	2,09	1,77	30,29	27,37	890,51
2 paramètres							
$\theta=-2$	C=0,1	34,47	34,48	195,7	697,57	5,12	36,79
	C=0,2	58,93	59,07	208,47	541,51	4,1	25,04
	C=0,3	75,78	76,09	163,47	352,88	3,57	17,9
$\theta=-1$	C=0,1	33,69	34,5	40,79	142,07	5,48	45,88
	C=0,2	53,67	56,23	44,16	115,9	4,13	23,34
	C=0,3	62,94	67,87	34,26	82,27	4,31	28,57
$\theta=0$	C=0,1	14,89	18,33	3,6	16,74	8,35	96,17
	C=0,2	18,27	22,41	3,72	16,04	8,88	102,22
	C=0,3	15,77	19,7	2,63	12	10,78	158,29
$\theta=1$	C=0,1	0,79	1,06	0,06	0,87	21,98	587,27
	C=0,2	0,71	1	0,08	1,05	49,75	3525,99
	C=0,3	0,66	0,88	0,09	1,54	61,89	4405,2
$\theta=2$	C=0,1	0,08	0,08	0	0,19	39,49	1782,77
	C=0,2	0,05	0,05	0	0,13	44,41	2246,29
	C=0,3	0,05	0,05	0	0,25	63,34	4734,75

Tableau 4.10b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU^* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	C=0,1	82,03	82,99	114,25	250,9	4,3	29,29
	C=0,2	87,84	90,08	75,26	145,14	4,14	27,4
	C=0,3	86,35	89,1	52,7	97,03	4,75	36,01
$\theta=-1$	C=0,1	66,31	72,04	29,23	70,35	5,04	39,25
	C=0,2	62,16	68,18	22,5	50,12	4,64	31,78
	C=0,3	55,55	60,47	23,06	64,03	6,92	74,67
$\theta=0$	C=0,1	30,15	35,18	8,3	27,67	7,83	102,17
	C=0,2	27,51	31,46	9,79	41,13	10,74	190,87
	C=0,3	25,49	27,52	12,6	55,58	9,69	146,12
$\theta=1$	C=0,1	13,52	15,09	5,71	35,71	16,01	435,04
	C=0,2	12,91	13,8	7,4	45,9	12,03	206,04
	C=0,3	11,15	11,61	7,54	49,91	11,87	184,76
$\theta=2$	C=0,1	6,02	6,11	4,65	44,12	19,65	537,16
	C=0,2	5,62	5,7	5,22	47,34	18,06	477,74
	C=0,3	4,69	4,71	5,03	50,35	18,13	440,89

Les résultats pour l'indice ZW sont présentés aux tableaux 4.11a et 4.11b.

Premièrement, le pourcentage de détection de ZW est moins élevé que celui des indices ZU , ZU^* , I_z et I_z^* .

Deuxièmement, nous remarquons que l'augmentation de C est accompagnée par une augmentation du pourcentage de détection. Le pourcentage présenté aux deux premières colonnes de ces tableaux peut, par contre, différer selon le modèle de réponse à l'item. Par exemple, nous remarquons que le pourcentage de patrons de réponses détectés est plus élevé lorsque le modèle à deux paramètres est utilisé alors que c'est le modèle à trois paramètres qui est associé aux pourcentages de détection les moins élevés.

Troisièmement, lorsque θ augmente, le pourcentage de patrons détectés tend à diminuer. Cela est logique puisque la réponse au hasard touche surtout les sujets dont l'habileté est faible (dans cette situation, l'impact du paramètre C est moins important).

Quatrièmement, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$. Ce résultat est aussi prévisible puisque le seuil $\alpha = 0,05$ est moins restrictif.

Cinquièmement, plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de ZW est élevée. Ce résultat est logique puisque les patrons de réponses inappropriés ont des valeurs positives avec cet indice. Le lecteur comprendra donc qu'un plus grand nombre de patrons de réponses inappropriés fera ainsi augmenter la moyenne de ZW . Fait intéressant, nous observons que l'écart-type de ZW est généralement plus près de la valeur théorique de la $N(0,1)$ que celui de l_z ou de l_z^* . Par contre, la valeur de l'écart-type tend à diminuer au fur et à mesure que la valeur de θ augmente.

Sixièmement, le coefficient d'asymétrie est généralement négatif et relativement près des valeurs de la loi $N(0,1)$. Nous observons aussi des coefficients de kurtose dont la valeur est relativement en accord avec la $N(0,1)$. En ce qui a trait à ce coefficient, il n'y a pas de distinction importante à soulever entre le modèle de

Rasch et le modèle à deux paramètres. Par contre, c'est le modèle à trois paramètres qui présente les coefficients de kurtose les plus faibles.

Tableau 4.11a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	20,9	38,87	1,54	1,27	-0,23	-0,09
	C=0,2	58,89	75,99	2,78	1,23	-0,37	0,22
	C=0,3	82,52	92,02	3,61	1,13	-0,42	0,35
$\theta=-1$	C=0,1	10,73	24,56	1,1	1,21	-0,19	-0,06
	C=0,2	35,69	56,33	2,08	1,2	-0,32	0,12
	C=0,3	59,38	77,98	2,78	1,12	-0,41	0,32
$\theta=0$	C=0,1	4,63	14,07	0,7	1,16	-0,15	-0,17
	C=0,2	14,71	31,75	1,37	1,14	-0,25	-0,03
	C=0,3	25,98	47,8	1,83	1,12	-0,36	0,21
$\theta=1$	C=0,1	1,23	6,18	0,29	1,09	-0,1	-0,32
	C=0,2	3,05	11,58	0,67	1,09	-0,22	-0,21
	C=0,3	5,27	16,85	0,94	1,06	-0,25	-0,04
$\theta=2$	C=0,1	0,29	1,99	-0,02	0,98	0	-0,3
	C=0,2	0,53	2,37	0,12	0,95	-0,01	-0,2
	C=0,3	0,33	2,81	0,21	0,91	-0,05	-0,15
2 paramètres							
$\theta=-2$	C=0,1	21,28	38,43	1,55	1,28	-0,12	0,03
	C=0,2	58,74	75,37	2,83	1,3	-0,21	0,02
	C=0,3	82,91	91,9	3,74	1,22	-0,33	0,07
$\theta=-1$	C=0,1	12,69	26,19	1,12	1,28	-0,06	-0,12
	C=0,2	38	56,63	2,14	1,33	-0,19	-0,02
	C=0,3	61,69	77,84	2,9	1,25	-0,3	-0,03
$\theta=0$	C=0,1	7,33	17,41	0,77	1,24	-0,01	-0,25
	C=0,2	20,56	37,67	1,52	1,27	-0,17	-0,15
	C=0,3	35	55,01	2,05	1,25	-0,3	0,01
$\theta=1$	C=0,1	2,67	8,43	0,36	1,16	0,02	-0,36
	C=0,2	6,96	18,17	0,83	1,2	-0,1	-0,35
	C=0,3	11,37	26,17	1,14	1,19	-0,22	-0,32
$\theta=2$	C=0,1	0,63	4,09	0,05	1,04	0,17	-0,4
	C=0,2	1,16	5,8	0,21	1,08	0,13	-0,51
	C=0,3	1,23	6,29	0,31	1,07	0,06	-0,6

Tableau 4.11b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	C=0,1	2,51	9,84	0,7	0,99	-0,14	0
	C=0,2	11,39	29,61	1,43	0,97	-0,17	-0,01
	C=0,3	27,5	53,37	2	0,92	-0,22	0,04
$\theta=-1$	C=0,1	1,18	6,5	0,52	0,97	-0,17	-0,04
	C=0,2	5,17	18,45	1,1	0,95	-0,25	0,07
	C=0,3	12,48	33,47	1,55	0,91	-0,27	0,14
$\theta=0$	C=0,1	0,52	3,38	0,26	0,96	-0,14	-0,14
	C=0,2	1,73	8,79	0,67	0,95	-0,16	-0,07
	C=0,3	3,28	14,39	0,98	0,92	-0,19	-0,03
$\theta=1$	C=0,1	0,16	1,37	0,02	0,92	-0,02	-0,27
	C=0,2	0,38	2,51	0,22	0,9	-0,02	-0,17
	C=0,3	0,41	3,09	0,35	0,85	0,02	-0,04
$\theta=2$	C=0,1	0,06	0,44	-0,09	0,76	0,1	-0,04
	C=0,2	0,03	0,45	-0,07	0,72	0,18	0,04
	C=0,3	0,03	0,29	-0,08	0,67	0,25	0,07

Les résultats pour l'indice ZW^* sont présentés aux tableaux 4.12a et 4.12b. Tout d'abord, si les données sont analysées à l'aide du modèle de Rasch, l_z et l_z^* présentent des pourcentages de détection plus élevés que ZW^* . À l'opposé, ZU , ZU^* et ZW présentent des pourcentages de détection moins élevés que ZW^* pour ce modèle de réponse à l'item. En ce qui concerne le modèle à deux paramètres, l_z , l_z^* et ZU présentent des pourcentages de détection plus élevés que ZW^* alors que ZU^* et ZW présentent des pourcentages de détection moins élevés. Les résultats tirés des analyses selon le modèle à trois paramètres montrent que l_z , l_z^* , ZU (pour $\theta = 0$ à $\theta = 2$ uniquement) et ZU^* présentent des pourcentages de détection plus élevés que ZW^* et ZW présente les pourcentages de détection moins élevés. Ensuite,

l'augmentation du niveau C est accompagnée par une hausse du pourcentage de détection des patrons de réponses au hasard. Ce constat a aussi été soulevé pour d'autres indices. De plus, nous avons observé que le pourcentage de patrons de réponses détectés est plus élevé lorsque le modèle à deux paramètres est utilisé alors que c'est le modèle à trois paramètres qui engendre les pourcentages de détection les moins élevés.

Lorsque la valeur de θ augmente, le pourcentage de patrons détectant la réponse au hasard tend à diminuer. Ce constat semble logique dans le sens où le fait d'obtenir une bonne réponse au hasard a un effet plus important pour les étudiants dont le niveau d'habileté θ est faible. En effet, les étudiants dont le niveau d'habileté est élevé ont moins besoin de recourir à la réponse au hasard pour obtenir une bonne réponse. Ensuite, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$. Ce résultat est évidemment prévisible puisque le seuil $\alpha = 0,01$ est plus restrictif que $\alpha = 0,05$.

Tout comme pour l'indice ZW , nous remarquons que plus le pourcentage de patrons de réponses détecté est élevé, plus la moyenne des scores de ZW^* est élevée. Il est intéressant de remarquer que la valeur de l'écart-type tend à diminuer au fur et à mesure que la valeur de θ augmente. Ce résultat peut être expliqué par le fait que les scores de l'indice sont moins élevés lorsque l'habileté des sujets est élevée (l'impact du paramètre de pseudo-chance personnel C est moins important). Ensuite, le coefficient d'asymétrie est positif et est plus élevé pour le modèle à deux paramètres

que pour les deux autres modèles de réponse à l'item. Nous observons aussi des coefficients de kurtose dont la valeur est relativement près de la $N(0,1)$. Par contre, le coefficient de kurtose est plus élevé pour les données analysées selon le modèle à deux paramètres.

Tableau 4.12a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW^* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	37,88	53,12	1,9	1,72	0,42	0,07
	C=0,2	75,05	84,89	3,77	2,05	0,27	0,04
	C=0,3	91,55	95,73	5,26	2,16	0,14	-0,08
$\theta=-1$	C=0,1	23,33	37,87	1,29	1,51	0,5	0,27
	C=0,2	54,78	69,71	2,64	1,78	0,32	-0,01
	C=0,3	76,81	87,05	3,74	1,87	0,18	-0,08
$\theta=0$	C=0,1	13,09	24,99	0,79	1,33	0,5	0,16
	C=0,2	30,3	46,75	1,62	1,49	0,38	-0,01
	C=0,3	46,11	63,31	2,24	1,57	0,26	-0,02
$\theta=1$	C=0,1	5,75	13,64	0,32	1,16	0,5	0,07
	C=0,2	11,12	23,35	0,74	1,25	0,37	-0,11
	C=0,3	16,79	32,01	1,09	1,29	0,34	-0,02
$\theta=2$	C=0,1	2,52	7,59	0,01	1,06	0,54	0,13
	C=0,2	3,5	10,24	0,18	1,1	0,49	0,05
	C=0,3	4,46	12,38	0,31	1,11	0,37	-0,19
2 paramètres							
$\theta=-2$	C=0,1	37,12	52,67	1,9	1,74	0,56	0,51
	C=0,2	74,28	84,52	3,82	2,14	0,35	0
	C=0,3	91,33	95,75	5,37	2,25	0,16	-0,16
$\theta=-1$	C=0,1	24,59	38,51	1,32	1,59	0,6	0,46
	C=0,2	54,53	68,5	2,7	1,94	0,43	0,12
	C=0,3	76,41	86	3,87	2,04	0,23	-0,21
$\theta=0$	C=0,1	15,9	27,79	0,88	1,45	0,61	0,31
	C=0,2	35,75	51,47	1,83	1,69	0,45	0,05
	C=0,3	53,32	68,02	2,58	1,81	0,32	-0,09
$\theta=1$	C=0,1	8,08	16,64	0,41	1,28	0,64	0,34
	C=0,2	17,8	30,27	0,98	1,46	0,52	0,05
	C=0,3	27,03	40,93	1,4	1,56	0,4	-0,18
$\theta=2$	C=0,1	4,78	10,49	0,08	1,15	0,75	0,38
	C=0,2	6,97	14,88	0,28	1,24	0,69	0,19
	C=0,3	7,8	17,47	0,41	1,25	0,57	-0,05

Tableau 4.12b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW^* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	C=0,1	8,22	20,79	0,75	1,1	0,23	-0,02
	C=0,2	26,41	48,23	1,61	1,18	0,17	-0,06
	C=0,3	51,11	72,29	2,38	1,21	0,08	-0,11
$\theta=-1$	C=0,1	6,33	16,87	0,6	1,08	0,22	-0,05
	C=0,2	18,49	37,49	1,29	1,15	0,13	-0,08
	C=0,3	34,02	56,85	1,86	1,18	0,09	-0,06
$\theta=0$	C=0,1	3,49	11,14	0,32	1,05	0,27	-0,09
	C=0,2	9,29	22,03	0,8	1,11	0,23	-0,1
	C=0,3	15,89	33,92	1,18	1,13	0,14	-0,12
$\theta=1$	C=0,1	2,15	7,59	0,08	1,03	0,39	-0,05
	C=0,2	3,79	11,72	0,31	1,07	0,35	-0,1
	C=0,3	5,44	14,77	0,49	1,09	0,32	-0,09
$\theta=2$	C=0,1	1,53	5,89	-0,05	1	0,45	0,01
	C=0,2	1,9	6,35	0	1	0,49	0,06
	C=0,3	1,71	5,89	0,03	0,97	0,49	0,11

4.3.1.2 Inattention

Les tableaux 4.13a et 4.13b présentent les résultats pour la détection de l'inattention à l'aide de l'indice I_z . D'entrée de jeu, nous observons que lorsque le niveau d'inattention D augmente, le pourcentage de détection de patrons de réponses inattentives augmente lui aussi. Par exemple, pour le modèle de Rasch, $\alpha = 0,01$ et $\theta = -1$, I_z détecte 5,59% des patrons à $D = 0,10$, 9,92% à $D = 0,20$ et 14,85% à $D = 0,30$. Ce résultat est prévisible puisqu'augmenter la valeur de D augmente aussi la probabilité d'en faire la détection. Il est à noter que nous avons observé une tendance

comparable pour la détection de la réponse au hasard à l'aide de cet indice (tableaux 4.7a et 4.7b).

Le pourcentage de patrons de réponses détectés est plus élevé lorsque le modèle à deux paramètres est utilisé et c'est le modèle à trois paramètres qui engendre les pourcentages de détection les moins élevés.

Lorsque la valeur de θ augmente, nous remarquons que le pourcentage de patrons détectés tend à augmenter. Ce résultat est opposé à celui observé pour la réponse au hasard puisque nous avons examiné une baisse du pourcentage de patrons de réponses détectés lorsque θ augmentait. Il est important de comprendre que ce constat est logique puisque l'inattention est modélisée de telle façon qu'elle représente les étudiants ayant une habileté élevée qui ne répondraient pas correctement à un item. Encore une fois, le pourcentage de détection de l'inattention est plus élevé lorsque $\alpha = 0,05$.

En ce qui a trait aux statistiques descriptives, il est possible de remarquer que plus le pourcentage de patrons de réponses détecté est élevé, plus la moyenne des scores de l_z est négative et plus l'écart-type est grand. Ce constat est cohérent puisque le score de l_z est négatif lorsqu'un patron de réponses est inapproprié. Ainsi, plus une matrice de données va contenir un grand nombre de patrons de réponses inappropriés, plus la moyenne des scores de cet indice va être négativement élevée. En ce qui concerne l'écart-type, il est attendu qu'il augmente en présence d'un plus grand

nombre de scores négativement élevés. Contrairement aux analyses portant sur la réponse au hasard, nous observons que la moyenne et l'écart-type se rapprochent des valeurs de la loi $N(0,1)$ lorsque θ tend vers moins deux.

Nous observons que lorsque θ augmente, les coefficients d'asymétrie et de kurtose ont tendance à s'approcher de zéro, soit, la valeur de la loi $N(0,1)$. Il est à noter que nous obtenions la relation inverse dans le cas de l'analyse de la réponse au hasard. Enfin, les coefficients d'asymétrie et de kurtose sont moins élevés pour le modèle à trois paramètres, ce qui témoigne d'une plus grande proximité avec la loi $N(0,1)$.

Tableau 4.13a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice I_z (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	2,2	5,86	0,01	0,95	-0,96	1,23
	D=0,2	2,83	6,93	-0,09	0,97	-1	1,5
	D=0,3	3,32	8,11	-0,18	0,96	-0,97	1,15
$\theta=-1$	D=0,1	5,59	13,09	-0,32	1,15	-0,73	0,54
	D=0,2	9,92	20,69	-0,69	1,24	-0,65	0,45
	D=0,3	14,85	28,09	-0,99	1,29	-0,6	0,35
$\theta=0$	D=0,1	15,12	27,21	-0,88	1,42	-0,66	0,45
	D=0,2	33,17	49,09	-1,75	1,62	-0,52	0,2
	D=0,3	47,23	63,98	-2,34	1,68	-0,39	0,06
$\theta=1$	D=0,1	34,77	49,08	-1,76	1,84	-0,57	0,22
	D=0,2	65,11	76,62	-3,23	2,08	-0,34	0,01
	D=0,3	80,69	88,79	-4,25	2,17	-0,29	0
$\theta=2$	D=0,1	56,79	68,13	-2,96	2,4	-0,5	0,1
	D=0,2	85,92	91,61	-5,19	2,62	-0,24	-0,08
	D=0,3	95,22	97,51	-6,68	2,68	-0,17	-0,06
2 paramètres							
$\theta=-2$	D=0,1	1,91	5,95	0,03	0,95	-0,89	1,08
	D=0,2	2,05	6,39	-0,04	0,93	-0,82	0,74
	D=0,3	2,6	7,06	-0,11	0,94	-0,97	1,22
$\theta=-1$	D=0,1	5,56	12,83	-0,29	1,15	-0,73	0,56
	D=0,2	9,69	20,05	-0,64	1,24	-0,64	0,4
	D=0,3	13,25	26,15	-0,89	1,28	-0,57	0,31
$\theta=0$	D=0,1	19,35	31,4	-1,01	1,54	-0,64	0,2
	D=0,2	36,94	52,43	-1,88	1,69	-0,48	0,2
	D=0,3	50,37	66,2	-2,44	1,74	-0,4	0,18
$\theta=1$	D=0,1	38,74	50,75	-1,93	2,05	-0,67	0,35
	D=0,2	66,91	77,34	-3,49	2,34	-0,41	0,03
	D=0,3	82,63	89,71	-4,62	2,41	-0,33	0,15
$\theta=2$	D=0,1	59,09	68,72	-3,31	2,78	-0,62	0,18
	D=0,2	87,09	91,67	-5,87	3,14	-0,33	-0,07
	D=0,3	95,74	97,73	-7,63	3,2	-0,21	-0,1

Tableau 4.13b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice l_z (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta = -2$	D=0,1	1,39	5,31	0,06	0,98	-0,39	0,07
	D=0,2	1,76	6,25	-0,02	1	-0,37	0,06
	D=0,3	1,86	6,55	-0,03	1,01	-0,38	0,12
$\theta = -1$	D=0,1	4,34	11,71	-0,28	1,09	-0,42	0,03
	D=0,2	6,99	17,78	-0,57	1,14	-0,31	-0,11
	D=0,3	8,14	19,61	-0,68	1,14	-0,29	-0,03
$\theta = 0$	D=0,1	13,23	25,01	-0,8	1,3	-0,42	-0,08
	D=0,2	22,23	39,58	-1,32	1,3	-0,21	-0,2
	D=0,3	25,11	44,57	-1,5	1,23	-0,14	-0,14
$\theta = 1$	D=0,1	33,93	49,04	-1,66	1,66	-0,29	-0,36
	D=0,2	49,76	67,61	-2,33	1,47	-0,08	-0,09
	D=0,3	52,13	71,11	-2,42	1,32	-0,08	-0,15
$\theta = 2$	D=0,1	57,59	67,93	-2,68	2,03	-0,12	-0,5
	D=0,2	75,64	86,33	-3,41	1,62	0,08	-0,02
	D=0,3	77,78	89,17	-3,42	1,42	-0,04	-0,08

Les tableaux 4.14a et 4.14b présentent les résultats pour la détection de l'inattention à l'aide de l'indice l_z^* . Dans ce cas-ci, nous obtenons des pourcentages de détection légèrement plus élevés que ceux de l_z .

Premièrement, nous observons que lorsque le niveau d'inattention D augmente, le pourcentage de détection de patrons de réponses inattentives augmente lui aussi. L'explication à cela est identique à celle présentée pour l'indice l_z .

Deuxièmement, le modèle à deux paramètres engendre les pourcentages de détection les plus élevés. À l'opposé, c'est le modèle à trois paramètres qui présente

les pourcentages de détection les moins élevés. Il est à noter que ce constat a aussi été fait pour l'indice l_z .

Troisièmement, lorsque θ augmente, nous remarquons que le pourcentage de réponses détectées tend à augmenter. Il est important de comprendre que ce constat est logique puisque l'inattention est modélisée de telle façon qu'elle représente les étudiants ayant une habileté élevée qui ne répondraient pas correctement à un item. Évidemment, ce résultat est à l'opposé de celui observé pour la réponse au hasard.

Quatrièmement, le pourcentage de détection de l'inattention est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$. Ce résultat est attendu puisque $\alpha = 0,05$ est moins restrictif que $\alpha = 0,01$.

Cinquièmement, il est possible de remarquer que plus le pourcentage de patrons de réponses détecté est élevé, plus la moyenne des scores de l_z est négativement élevée et plus l'écart-type est élevé. De plus, contrairement aux analyses portant sur la réponse au hasard, nous observons que la moyenne et l'écart-type se rapprochent des valeurs de la loi $N(0,1)$ au fur et à mesure que la valeur de θ diminue. Cela fait du sens puisque le score de l_z est négatif lorsqu'un patron de réponses est inapproprié. Ainsi, plus une matrice de données va contenir un grand nombre de patrons de réponses inappropriés (ce qui est le cas aux valeurs $\theta < 0$), plus la moyenne des scores de cet indice va être négativement élevée et plus l'écart-type va augmenter.

Enfin, nous observons, que lorsque θ augmente, les coefficients d'asymétrie et de kurtose ont tendance à s'approcher de zéro, soit, la valeur théorique pour la loi $N(0,1)$.

Tableau 4.14a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice I_z^* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	2,89	7,39	-0,02	1,03	-0,97	1,26
	D=0,2	3,95	9,51	-0,15	1,08	-0,98	1,38
	D=0,3	5,23	11,35	-0,28	1,1	-0,93	1,05
$\theta=-1$	D=0,1	5,77	13,19	-0,32	1,16	-0,73	0,54
	D=0,2	10,57	21,33	-0,71	1,26	-0,65	0,47
	D=0,3	16,35	29,65	-1,04	1,33	-0,6	0,35
$\theta=0$	D=0,1	15,12	27,21	-0,88	1,42	-0,67	0,45
	D=0,2	33,2	49,09	-1,76	1,62	-0,52	0,21
	D=0,3	47,43	64,07	-2,36	1,69	-0,39	0,05
$\theta=1$	D=0,1	34,79	49,08	-1,76	1,85	-0,56	0,22
	D=0,2	65,11	76,63	-3,24	2,08	-0,34	0,01
	D=0,3	80,7	88,79	-4,26	2,17	-0,29	0
$\theta=2$	D=0,1	57,79	68,69	-3,01	2,43	-0,48	0,08
	D=0,2	86,13	91,72	-5,21	2,62	-0,24	-0,08
	D=0,3	95,25	97,51	-6,68	2,68	-0,17	-0,06
2 paramètres							
$\theta=-2$	D=0,1	2,6	7,31	-0,01	1,02	-0,86	0,92
	D=0,2	3,06	8,59	-0,11	1,02	-0,8	0,6
	D=0,3	4,31	9,83	-0,2	1,06	-0,92	1,06
$\theta=-1$	D=0,1	5,67	12,96	-0,29	1,16	-0,73	0,56
	D=0,2	10,02	20,43	-0,65	1,25	-0,64	0,39
	D=0,3	14,16	27,15	-0,92	1,3	-0,57	0,31
$\theta=0$	D=0,1	19,42	31,47	-1,01	1,54	-0,64	0,2
	D=0,2	37,09	52,63	-1,89	1,69	-0,48	0,2
	D=0,3	50,57	66,45	-2,46	1,74	-0,4	0,18
$\theta=1$	D=0,1	38,87	50,83	-1,94	2,06	-0,66	0,35
	D=0,2	67,01	77,35	-3,5	2,35	-0,41	0,03
	D=0,3	82,71	89,73	-4,63	2,41	-0,33	0,15
$\theta=2$	D=0,1	60,07	69,39	-3,38	2,82	-0,6	0,15
	D=0,2	87,33	91,8	-5,92	3,15	-0,32	-0,07
	D=0,3	95,77	97,76	-7,66	3,21	-0,21	-0,1

Tableau 4.14b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice l_z^* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	1,54	5,79	0,05	1	-0,39	0,06
	D=0,2	1,94	6,68	-0,04	1,01	-0,37	0,06
	D=0,3	2,15	7,19	-0,06	1,02	-0,38	0,11
$\theta=-1$	D=0,1	5,06	13	-0,31	1,14	-0,41	0,02
	D=0,2	7,85	18,95	-0,6	1,18	-0,3	-0,12
	D=0,3	9,03	20,85	-0,71	1,16	-0,28	-0,05
$\theta=0$	D=0,1	15,85	28,53	-0,9	1,39	-0,39	-0,11
	D=0,2	24,99	42,71	-1,41	1,36	-0,19	-0,18
	D=0,3	27,25	46,89	-1,57	1,27	-0,13	-0,13
$\theta=1$	D=0,1	38,73	53,16	-1,85	1,78	-0,27	-0,35
	D=0,2	54,33	70,97	-2,5	1,54	-0,08	-0,01
	D=0,3	55,69	73,53	-2,53	1,37	-0,09	-0,09
$\theta=2$	D=0,1	62,23	71,42	-3,02	2,23	-0,1	-0,38
	D=0,2	79,31	88,5	-3,68	1,71	0,06	0,11
	D=0,3	80,45	90,57	-3,59	1,47	-0,06	0,03

Les tableaux 4.15a et 4.15b présentent les résultats pour l'indice ZU . Pour commencer, le pourcentage de détection de cet indice est inférieur à ceux de l_z et l_z^* pour le modèle de Rasch et pour le modèle à trois paramètres. En ce qui concerne les données analysées selon le modèle à deux paramètres, le pourcentage de détection de ZU est plus élevé que celui des indices de type vraisemblance. Il est à noter que nous avons obtenu une relation comparable lors de l'analyse de la réponse au hasard (tableau 4.9ab).

Selon les résultats présentés pour le modèle de Rasch, l'augmentation du niveau d'inattention D est accompagnée par une hausse du pourcentage de détection

des patrons de réponses représentant l'inattention. Par contre, ce constat n'est pas observé pour les deux autres modèles de réponse à l'item.

Le pourcentage de patrons de réponses détectés est plus élevé pour le modèle à deux paramètres (rappelons que cet indice présentait une plus grande propension à avoir des scores extrêmes, au tableau 4.2, ce qui a pour effet d'augmenter les patrons détectés). Cela est d'ailleurs explicite lorsqu'on observe les pourcentages de détection au seuil $\alpha = 0,05$. C'est lorsque le modèle à trois paramètres est utilisé que les pourcentages de détection sont les moins élevés.

Pour le modèle de Rasch et pour le modèle à trois paramètres uniquement, l'augmentation de θ engendre généralement une augmentation du pourcentage de patrons de réponses détectés.

Plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne et l'écart-type des scores de ZU sont aussi élevés. Il est important de mentionner que les moyennes et les écarts-types obtenus à l'aide du modèle à deux paramètres sont beaucoup plus élevés. Par exemple, pour la situation où $\theta = 2$ et $D = 0,1$, nous obtenons une moyenne et un écart-type qui sont respectivement égaux à 384,02 et 849,01. De son côté, l'asymétrie et la kurtose sont généralement positives et très élevées dans certaines situations d'analyse. Ainsi, l'étude des données selon le modèle à deux paramètres supporte bien ce dernier argument. Par exemple, pour la situation où $\theta=2$ et $D = 0,3$, nous obtenons un coefficient d'asymétrie de 39,72 et un

coefficient de kurtose de 2772,8. Ces valeurs sont très importantes puisqu'elles soutiennent les résultats présentés aux sections 4.1.1.1 et 4.1.1.2 et témoignent de la difficulté d'utiliser *ZU* dans la réalité. De plus, ces coefficients d'asymétrie présentent une certaine cohérence avec les résultats présentés au tableau 4.2.

Tableau 4.15a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	5,75	20,77	1,6	0,68	1,57	7,09
	D=0,2	11,43	32,41	1,86	0,81	2,06	10,87
	D=0,3	20,01	46,73	2,15	0,97	2,32	11,43
$\theta=-1$	D=0,1	0,65	1,95	0,72	0,51	0,99	4,09
	D=0,2	0,94	3,63	0,95	0,52	1,03	4,16
	D=0,3	1,86	8,23	1,2	0,55	1,28	5,86
$\theta=0$	D=0,1	4,52	11	0,76	0,91	0,73	0,37
	D=0,2	4,58	11,75	1,05	0,79	0,69	1,12
	D=0,3	3,97	10,9	1,2	0,66	0,98	2,57
$\theta=1$	D=0,1	25,13	36,87	1,6	1,29	0,46	-0,77
	D=0,2	35,73	53,62	2,12	1,19	0,1	-0,68
	D=0,3	33,18	54,35	2,17	1,02	0,25	-0,3
$\theta=2$	D=0,1	42,03	56,38	2,41	1,24	0,34	-0,96
	D=0,2	63,51	77,41	3,03	1,23	-0,17	-0,82
	D=0,3	74,24	86,93	3,32	1,12	-0,27	-0,47
2 paramètres							
$\theta=-2$	D=0,1	100	100	53,03	60,36	14,68	377,63
	D=0,2	100	100	70,35	93,52	9,89	156,69
	D=0,3	100	100	97,29	148,01	7,73	80,6
$\theta=-1$	D=0,1	99,2	100	10,63	7,24	1,75	4,82
	D=0,2	99,83	100	15,77	11,18	3,64	41,13
	D=0,3	99,94	100	22,65	18,87	9,73	237,73
$\theta=0$	D=0,1	75,39	100	4,06	3,59	24,94	1505,99
	D=0,2	87,85	100	5,03	3,85	3,43	15,36
	D=0,3	95,96	100	8,09	7,12	2,58	11,18
$\theta=1$	D=0,1	96,9	100	29,02	65,97	11,43	236,64
	D=0,2	94,71	100	10,75	25,64	16,55	427,71
	D=0,3	95,46	100	6,23	9,09	24,13	989,78
$\theta=2$	D=0,1	99,87	100	384,02	849,01	16	565,72
	D=0,2	99	100	98,46	323,58	12,48	342,73
	D=0,3	98,31	100	22,2	115,76	39,72	2772,8

Tableau 4.15b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice $ZU(30 \text{ items})$

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	0,26	1,63	-0,08	0,93	0,12	-0,11
	D=0,2	0,34	1,82	-0,05	0,95	0,08	-0,07
	D=0,3	0,32	1,98	-0,05	0,96	0,1	-0,13
$\theta=-1$	D=0,1	0,33	1,88	0,08	0,82	0,38	0,85
	D=0,2	0,49	2,64	0,15	0,9	0,17	0,13
	D=0,3	0,6	3,21	0,18	0,94	0,13	-0,06
$\theta=0$	D=0,1	3,27	8,77	0,69	0,9	1,17	3,28
	D=0,2	2,71	9,89	0,74	0,93	0,46	0,92
	D=0,3	2,5	9,69	0,7	0,94	0,16	0,06
$\theta=1$	D=0,1	23,67	38,35	1,85	1,43	2,75	23,91
	D=0,2	15,02	32,57	1,56	1,07	0,81	1,21
	D=0,3	11,63	29,52	1,43	0,98	0,28	0,42
$\theta=2$	D=0,1	66,83	77,73	4,14	3,88	5,04	46,14
	D=0,2	41,37	61,69	2,52	1,76	4,87	64,92
	D=0,3	29,96	53,92	2,06	1,05	0,64	4,08

Les tableaux 4.16a et 4.16b présentent les résultats pour l'indice ZU^* . Les constats effectués à la première étude ont montré que cet indice doit être utilisé avec précaution puisque la fonction de densité de ses scores ne suit pas les quantiles d'une loi $N(0,1)$.

Pour débiter, ZU^* présente un pourcentage de détection inférieur à l_z , l_z^* et ZU lorsque les données sont analysées à l'aide du modèle de Rasch et du modèle à deux paramètres. Par contre, le pourcentage de détection de ZU^* est supérieur à celui de tous ces indices lorsque le modèle à trois paramètres est utilisé.

Le pourcentage de patrons de réponses détectés de cet indice est plus élevé pour le modèle à trois paramètres. À l’opposé, c’est lorsque le modèle à deux paramètres est utilisé que les pourcentages de détection sont les moins élevés. Notons que ces résultats ont aussi été observés pour la détection de la réponse au hasard.

L’augmentation de θ engendre généralement une augmentation du pourcentage de patrons de réponses détectés. Cela est logique puisque l’inattention est modélisée pour les sujets dont la valeur θ est élevée. De plus, l’augmentation du pourcentage de patrons de réponses détectés s’accompagne généralement de moyennes et d’écart-types plus élevés. Cela est prévisible puisqu’augmenter la valeur de D augmente aussi la probabilité d’en faire la détection.

Il est important de mentionner que les moyennes et les écart-types obtenus à l’aide du modèle à trois paramètres sont beaucoup plus élevés. Par exemple, pour la situation où $\theta = -2$ et $D = 0,3$, nous obtenons une moyenne et un écart-type qui sont respectivement égaux à 296,66 et à 770,54. De leur côté, les coefficients d’asymétrie et de kurtose sont généralement positifs et très élevés. Ainsi, l’étude des scores obtenus à l’aide du modèle à deux paramètres montre bien cet argument. Par exemple, pour la situation où $\theta = 0$ et $D = 0,2$, nous obtenons un coefficient d’asymétrie de 115,6 et un coefficient de kurtose de 13 870,88. De plus, les coefficients d’asymétrie et de kurtose présentés pour ZU^* abondent dans le même sens que ceux de ZU : la fonction de densité de ces indices n’est généralement pas en

accord avec les quantiles de la loi $N(0,1)$. Cela complique donc considérablement l'utilisation de ZU et ZU^* .

Tableau 4.16a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	0,15	0,15	1,1	65,49	75,16	6351,35
	D=0,2	0,11	0,11	0,98	88,21	112,86	13272,82
	D=0,3	0,09	0,11	0,63	71,83	122,38	14981,63
$\theta=-1$	D=0,1	0,92	0,94	1,95	48,68	29,23	915,42
	D=0,2	1,01	1,01	2,37	96,54	81,57	7979,17
	D=0,3	0,77	0,77	3,19	84,7	32,9	1229,84
$\theta=0$	D=0,1	7,8	9,92	2,12	22,92	19,54	435,74
	D=0,2	6,63	8,12	4,03	51,13	22,26	628,17
	D=0,3	4,55	5,16	5,84	88,42	24,74	785,03
$\theta=1$	D=0,1	38,63	43,29	8,03	21	7,02	85,37
	D=0,2	40,55	46,41	10,08	44,44	15,42	412,78
	D=0,3	28,92	33,81	11,33	67,14	12,37	209,75
$\theta=2$	D=0,1	55,88	57,91	38,7	85,27	5,11	40,95
	D=0,2	75,29	78,43	33,87	60,13	4,66	35,94
	D=0,3	72,9	76,63	31,49	77,85	7,23	95,17
2 paramètres							
$\theta=-2$	D=0,1	0,01	0,01	0	0,08	117,31	14143,91
	D=0,2	0,01	0,01	0	0,18	121,47	14831,84
	D=0,3	0	0	0	0,01	28,77	1100,24
$\theta=-1$	D=0,1	0,05	0,05	0,01	0,16	59,32	4710,61
	D=0,2	0,07	0,07	0,02	0,14	33,56	1272,91
	D=0,3	0,05	0,06	0,02	0,33	79,8	7339,31
$\theta=0$	D=0,1	5,79	8,3	0,59	3,46	71,28	7099,34
	D=0,2	4,15	6,1	0,54	7,73	115,6	13870,88
	D=0,3	1,46	2,32	0,27	0,94	15,55	344,77
$\theta=1$	D=0,1	35,71	40,35	6,03	13,25	4,1	26,34
	D=0,2	35,87	42,63	4,49	10,13	7,88	129,29
	D=0,3	21,85	28	2,51	10,72	48,22	3558,11
$\theta=2$	D=0,1	53,89	55,62	29,7	62,37	4,78	38,96
	D=0,2	71,58	75,27	23,92	43,36	5,77	69,71
	D=0,3	65,96	70,83	17,4	43,17	7,61	84,78

Tableau 4.16b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU^* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	71,02	71,21	190,02	493,44	5,44	58,82
	D=0,2	71,77	71,87	234,1	601,31	4,64	35,97
	D=0,3	70,89	70,98	296,66	770,54	4,78	41,59
$\theta=-1$	D=0,1	69,41	71,6	61,14	162,35	5,56	51,8
	D=0,2	71,09	72,21	92,1	240,86	4,81	34,56
	D=0,3	71,93	72,48	138,97	367,96	5,29	52,39
$\theta=0$	D=0,1	61,61	66,16	24,49	60,96	5,25	38,99
	D=0,2	72,07	74,88	45,94	122,21	7,08	106,74
	D=0,3	74,21	75,69	80,87	207,5	4,48	26,76
$\theta=1$	D=0,1	58,13	61,49	26,25	59,05	6,49	82,51
	D=0,2	76,82	80,23	35,32	71,56	5,89	65,05
	D=0,3	79,49	81,73	54,01	133,43	6,61	73,09
$\theta=2$	D=0,1	60,91	62,5	63,52	135,36	4,54	30,15
	D=0,2	82	84,23	57,02	97,37	4,72	37,45
	D=0,3	88,29	90,17	54,86	95,76	4,69	36,01

Les résultats pour l'indice ZW sont présentés aux tableaux 4.17a et 4.17b. Tout d'abord, le pourcentage de détection de cet indice est inférieur à ceux de l_z , l_z^* et ZU^* pour le modèle de Rasch. En ce qui concerne le modèle à deux paramètres, ZW présente des pourcentages de détection inférieurs à ceux de l_z , l_z^* et ZU , mais un pourcentage de détection supérieur à celui de ZU^* aux valeurs $\theta = -2$ à $\theta = 0$. Les résultats selon le modèle à trois paramètres montrent que ZW a un pourcentage de détection inférieur à l_z , l_z^* et ZU^* , mais un pourcentage de détection supérieur à celui de ZU .

L'augmentation du niveau d'inattention D est systématiquement accompagnée par une hausse du pourcentage de détection des patrons de réponses représentant l'inattention. Ce constat logique (augmenter la valeur de D augmente aussi la probabilité d'en faire la détection) a aussi été observé avec les indices de type vraisemblance. De plus, nous avons remarqué que le pourcentage de patrons de réponses détectés est légèrement plus élevé pour le modèle à deux paramètres. Lorsque c'est le modèle à trois paramètres qui est utilisé, les pourcentages de détection sont à leurs valeurs minimales.

Lorsque la valeur de θ augmente, le pourcentage de patrons de réponses détectant l'inattention augmente lui aussi (dans le cas de la réponse au hasard, nous obtenions la relation inverse). Cela suit une certaine logique puisque le fait d'obtenir une mauvaise réponse par inattention a un effet plus important pour les étudiants dont le niveau d'habileté est élevé. Ensuite, nous avons découvert, sans trop de surprise, que le pourcentage de détection est plus élevé lorsque $\alpha = 0,05$.

Nous remarquons que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de ZW est élevée. Il est, par contre, intéressant de remarquer que la valeur de l'écart-type reste relativement constante et près de un lorsque θ augmente. Cela est lié au fait que les quantiles des scores de ZW suivent approximativement ceux de la $N(0,1)$.

De son côté, l'asymétrie est généralement négative et augmente légèrement lorsque θ augmente. Notons que c'est le modèle à trois paramètres qui présente les coefficients d'asymétrie les plus près de ceux d'une loi $N(0,1)$. Enfin, les coefficients de kurtose sont près des valeurs de la loi $N(0,1)$ et légèrement moins élevés pour les résultats obtenus à l'aide du modèle à trois paramètres.

Tableau 4.17a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW(30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	0,29	2,1	0	0,98	0	-0,31
	D=0,2	0,36	2,44	0,12	0,95	-0,04	-0,22
	D=0,3	0,35	2,69	0,22	0,91	-0,03	-0,14
$\theta=-1$	D=0,1	1,3	5,72	0,29	1,08	-0,1	-0,31
	D=0,2	2,92	10,69	0,65	1,09	-0,21	-0,13
	D=0,3	4,95	17	0,93	1,06	-0,26	-0,09
$\theta=0$	D=0,1	4,31	13,03	0,68	1,15	-0,15	-0,18
	D=0,2	15,01	31,78	1,38	1,15	-0,25	-0,01
	D=0,3	25,87	47,67	1,83	1,11	-0,36	0,17
$\theta=1$	D=0,1	11,15	25,17	1,11	1,22	-0,2	-0,14
	D=0,2	36,15	57,19	2,1	1,2	-0,33	0,09
	D=0,3	58,73	76,98	2,76	1,13	-0,36	0,19
$\theta=2$	D=0,1	21,16	38,38	1,54	1,26	-0,2	-0,07
	D=0,2	58,93	76,18	2,79	1,22	-0,33	0,16
	D=0,3	82,71	92,25	3,63	1,13	-0,4	0,27
2 paramètres							
$\theta=-2$	D=0,1	0,29	1,92	-0,02	0,97	0	-0,28
	D=0,2	0,26	1,97	0,08	0,94	-0,02	-0,25
	D=0,3	0,37	2,51	0,15	0,91	0,06	-0,09
$\theta=-1$	D=0,1	1,23	5,59	0,27	1,09	-0,08	-0,32
	D=0,2	2,9	10,28	0,6	1,09	-0,14	-0,18
	D=0,3	4,07	14,33	0,83	1,07	-0,23	-0,08
$\theta=0$	D=0,1	5,81	15,54	0,72	1,2	-0,11	-0,26
	D=0,2	16,29	33,58	1,42	1,18	-0,25	-0,08
	D=0,3	26,83	48,51	1,85	1,13	-0,36	0,12
$\theta=1$	D=0,1	11,57	25,06	1,1	1,24	-0,11	-0,14
	D=0,2	35,4	55,27	2,07	1,25	-0,25	0
	D=0,3	59,58	76,86	2,79	1,18	-0,38	0,2
$\theta=2$	D=0,1	21,93	38,85	1,56	1,29	-0,11	-0,15
	D=0,2	59,84	76,55	2,84	1,28	-0,3	0,09
	D=0,3	83,47	92,1	3,71	1,18	-0,39	0,16

Tableau 4.17b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	0,3	1,92	-0,06	1	-0,03	-0,18
	D=0,2	0,37	2,39	0,02	1	-0,04	-0,13
	D=0,3	0,48	2,6	0,03	1,01	-0,05	-0,09
$\theta=-1$	D=0,1	0,74	4,05	0,2	1,02	-0,06	-0,13
	D=0,2	1,68	6,85	0,46	1,03	-0,12	-0,09
	D=0,3	1,97	8,17	0,57	1,01	-0,13	-0,01
$\theta=0$	D=0,1	2,42	8,49	0,52	1,07	-0,09	-0,04
	D=0,2	6,15	18,35	1	1,06	-0,16	-0,01
	D=0,3	7,93	23,94	1,22	1,01	-0,2	0,03
$\theta=1$	D=0,1	8,85	21,43	1,02	1,16	-0,11	-0,15
	D=0,2	21,06	41,93	1,69	1,08	-0,26	0,07
	D=0,3	26,51	50,83	1,93	0,99	-0,23	0,03
$\theta=2$	D=0,1	22,14	39,6	1,56	1,27	-0,11	-0,33
	D=0,2	46,29	68,13	2,41	1,11	-0,39	0,28
	D=0,3	54,57	76,74	2,65	0,98	-0,29	0,13

Les résultats pour l'indice ZW^* sont présentés aux tableaux 4.18a et 4.18b. Pour débiter, le pourcentage de détection de cet indice est inférieur à ceux de l_z et l_z^* lorsque le modèle de Rasch est utilisé. En ce qui concerne le modèle à deux paramètres, ZW^* présente des pourcentages de détection inférieurs à ceux de l_z , l_z^* et ZU , mais un pourcentage de détection supérieur à celui de ZU^* et ZW . Les résultats selon le modèle à trois paramètres montrent que ZW^* a un pourcentage de détection inférieur à l_z , l_z^* et ZU^* , mais un pourcentage de détection supérieur à celui de ZU et ZW .

L'augmentation du niveau d'inattention D est accompagnée par une hausse du pourcentage de détection de l'inattention. Ce constat a aussi été observé et expliqué pour certains indices de cette section. Comparativement à ZW , nous avons remarqué que le pourcentage de patrons de réponses détectés est légèrement plus élevé lorsque le modèle de Rasch est utilisé. Par contre, c'est le modèle à trois paramètres qui engendre les pourcentages de détection les moins élevés pour ZW et ZW^* .

Dans un autre ordre d'idée, lorsque θ augmente, le pourcentage de patrons détectant l'inattention augmente lui aussi. Cela est logique puisque l'inattention est modélisée pour les sujets dont l'habileté est élevée. Ensuite, le pourcentage de détection de la réponse inattentive est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$. Ce résultat est évidemment prévisible puisque le niveau $\alpha = 0,01$ est plus restrictif que le niveau $\alpha = 0,05$.

Tout comme pour l'indice ZW , nous remarquons que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de ZW^* est élevée. Ce constat est cohérent puisque le score de cet est positif lorsqu'un patron de réponses est inapproprié. Ainsi, plus une matrice de données va contenir un grand nombre de patrons de réponses inappropriés, plus la moyenne des scores de cet indice va être élevée. Ensuite, nous remarquons que la valeur de l'écart-type tend à légèrement augmenter au fur et à mesure que θ augmente.

En ce qui concerne le coefficient d'asymétrie, nous observons qu'il est généralement positif et moins élevé pour le modèle à trois paramètres. Nous remarquons, finalement, des coefficients de kurtose dont la valeur est relativement près de la $N(0,1)$. Ces résultats sont tous en accord avec ceux présentés à la première étude.

Tableau 4.18a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	2,56	7,6	0,03	1,06	0,53	0,08
	D=0,2	3,5	10,02	0,17	1,09	0,45	-0,05
	D=0,3	4,45	12,14	0,31	1,1	0,38	-0,16
$\theta=-1$	D=0,1	5,23	13,45	0,32	1,15	0,51	0,13
	D=0,2	10,29	22,39	0,73	1,24	0,41	0,07
	D=0,3	17,04	32,19	1,08	1,29	0,31	-0,11
$\theta=0$	D=0,1	12,11	24,06	0,76	1,31	0,5	0,22
	D=0,2	30,33	46,82	1,62	1,5	0,38	0,05
	D=0,3	46,11	63,36	2,25	1,57	0,24	-0,09
$\theta=1$	D=0,1	23,86	38,45	1,31	1,51	0,47	0,14
	D=0,2	55,69	70,12	2,67	1,78	0,3	-0,04
	D=0,3	75,79	86,48	3,71	1,88	0,22	-0,06
$\theta=2$	D=0,1	37,45	52,89	1,89	1,71	0,45	0,18
	D=0,2	75,15	85,27	3,8	2,05	0,29	-0,02
	D=0,3	91,64	95,85	5,31	2,18	0,16	-0,09
2 paramètres							
$\theta=-2$	D=0,1	2,27	6,93	0,01	1,04	0,52	0,13
	D=0,2	2,75	8,91	0,13	1,05	0,46	-0,06
	D=0,3	3,79	10,35	0,24	1,07	0,52	0,22
$\theta=-1$	D=0,1	5,07	13,07	0,29	1,15	0,52	0,13
	D=0,2	9,37	20,93	0,66	1,23	0,46	0,11
	D=0,3	13,77	28,01	0,94	1,26	0,35	0
$\theta=0$	D=0,1	14,47	26,15	0,82	1,38	0,49	0,09
	D=0,2	31,8	48,87	1,68	1,53	0,36	0,03
	D=0,3	46,72	64,31	2,27	1,58	0,24	-0,04
$\theta=1$	D=0,1	23,21	38,01	1,28	1,53	0,52	0,3
	D=0,2	52,99	68,41	2,58	1,8	0,35	0,01
	D=0,3	75,33	85,6	3,69	1,91	0,2	-0,05
$\theta=2$	D=0,1	37,88	52,69	1,91	1,73	0,46	0,11
	D=0,2	75,23	85,11	3,8	2,06	0,27	0
	D=0,3	91,35	95,8	5,28	2,15	0,11	-0,14

Tableau 4.18b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW^* (30 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	1,45	5,61	-0,05	1,01	0,32	-0,01
	D=0,2	1,81	6,53	0,04	1,01	0,3	0
	D=0,3	2,02	7,06	0,07	1,02	0,31	0,06
$\theta=-1$	D=0,1	3,33	10,09	0,22	1,08	0,32	-0,01
	D=0,2	5,69	15,53	0,5	1,11	0,25	-0,06
	D=0,3	6,53	17,71	0,62	1,1	0,23	0
$\theta=0$	D=0,1	8,1	19,17	0,62	1,19	0,3	0,03
	D=0,2	16,56	33,06	1,13	1,23	0,19	-0,08
	D=0,3	21,03	40,27	1,36	1,18	0,13	-0,09
$\theta=1$	D=0,1	21,77	37,29	1,23	1,4	0,26	-0,17
	D=0,2	40,82	60,45	2,02	1,36	0,06	-0,17
	D=0,3	47,77	68,29	2,27	1,27	0,06	-0,13
$\theta=2$	D=0,1	41,54	56,57	1,97	1,65	0,14	-0,35
	D=0,2	68,22	82,03	3,01	1,49	-0,11	-0,08
	D=0,3	75,21	88,33	3,25	1,34	-0,03	-0,11

4.3.2 Puissance des indices : résultats pour 80 items

Les résultats pour des tests simulés de 80 items sont présentés dans cette section. Encore une fois, nous décrivons successivement les tableaux pour la réponse au hasard et pour l'inattention. Il est important de noter que les explications des résultats pour 80 items sont exactement les mêmes que pour les analyses sur 30 items. En conséquence, nous allons uniquement nous concentrer sur la description des résultats dans la présente section.

4.3.2.1 Réponse au hasard

Les résultats présentés dans cette section vont généralement dans le même sens que ceux abordés à la section portant sur l'analyse de données simulées de 30 items. Tout d'abord, les tableaux 4.19a et 4.19b présentent les résultats pour l'indice I_z . Nous observons que les pourcentages de détection sont légèrement plus élevés lorsque cet indice analyse des tests de 80 items plutôt que de 30 items (voir les tableaux 4.7a et 4.7b). Ensuite, l'augmentation du niveau de réponse au hasard C engendre une augmentation du pourcentage de détection. Ainsi, pour le modèle de Rasch, la valeur $\theta = 0$ et le seuil $\alpha = 0,01$, I_z détecte 29,30% des patrons à $C = 0,10$, 63,89% à $C = 0,20$ et 83,75% à $C = 0,30$. Cette tendance a aussi été trouvée lorsque cet indice analyse 30 items.

Aux valeurs $\theta = -1$, $\theta = 0$ et $\theta = 1$, nous remarquons que le modèle à deux paramètres présente les pourcentages de détection les plus élevés. Pour les valeurs $\theta = -2$ et $\theta = 2$, le pourcentage de détection est généralement légèrement plus élevé pour le modèle de Rasch. Enfin, les pourcentages de détection du modèle à trois paramètres sont systématiquement plus faibles.

Tout comme nous avons pu l'observer dans les résultats portant sur l'analyse de données de 30 items, nous observons que le pourcentage de détection de la réponse au hasard est plus élevé lorsque le seuil $\alpha = 0,05$.

En ce qui concerne les statistiques descriptives, nous observons que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de

I_z s'éloigne de zéro. De son côté, les écarts-types sont situés entre 3,01 et 0,7, soit, des valeurs relativement près de l'écart-type de la loi $N(0,1)$. À l'instar des résultats pour 30 items, nous observons aussi que les écarts-types sont plus près de la loi $N(0,1)$ lorsque la valeur de θ approche de deux.

Les résultats montrent que, pour le modèle de Rasch et pour le modèle à deux paramètres, une augmentation du niveau de réponse au hasard C fait diminuer le coefficient d'asymétrie vers zéro, soit, la valeur prévue pour la loi $N(0,1)$. Par contre, ce constat ne s'applique pas à la valeur $\theta = 2$. De plus, cette relation n'est pas observée pour le modèle à trois paramètres. Enfin, plus θ augmente, plus le coefficient de kurtose augmente aussi. Par contre, cet accroissement est modeste puisque les coefficients sont bornés entre -0,08 et 0,65.

Tableau 4.19a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	87,65	93,08	-5,07	2,39	-0,26	-0,01
	C=0,2	99,52	99,81	-8,64	2,6	-0,12	-0,08
	C=0,3	99,97	99,99	-10,95	2,58	-0,09	-0,02
$\theta=-1$	C=0,1	63,73	77,63	-3,05	1,82	-0,31	0,06
	C=0,2	93,82	97,51	-5,41	2,04	-0,19	0
	C=0,3	99,13	99,81	-7,1	2,11	-0,16	0
$\theta=0$	C=0,1	29,3	47,27	-1,62	1,41	-0,39	0,23
	C=0,2	63,89	79,21	-2,97	1,6	-0,3	0,04
	C=0,3	83,75	92,42	-3,99	1,69	-0,28	0,09
$\theta=1$	C=0,1	8,47	19,47	-0,67	1,15	-0,46	0,25
	C=0,2	19,68	36,99	-1,29	1,25	-0,38	0,17
	C=0,3	31,78	52,12	-1,79	1,28	-0,35	0,09
$\theta=2$	C=0,1	2,11	7,15	-0,13	0,96	-0,55	0,41
	C=0,2	2,75	9,11	-0,29	0,95	-0,52	0,31
	C=0,3	3,9	11,31	-0,46	0,96	-0,61	0,62
2 paramètres							
$\theta=-2$	C=0,1	87,51	92,63	-5,46	2,75	-0,39	0,06
	C=0,2	99,39	99,79	-9,48	3,01	-0,13	-0,07
	C=0,3	99,97	99,99	-12,01	2,98	-0,11	0
$\theta=-1$	C=0,1	66,21	78,31	-3,32	2,05	-0,43	0,17
	C=0,2	94,26	97,21	-5,82	2,31	-0,22	-0,05
	C=0,3	99,19	99,71	-7,55	2,34	-0,17	-0,02
$\theta=0$	C=0,1	32,77	49,11	-1,74	1,56	-0,46	0,18
	C=0,2	67,1	80,97	-3,19	1,75	-0,34	0,13
	C=0,3	86,89	93,83	-4,29	1,79	-0,27	0,12
$\theta=1$	C=0,1	10,89	22,91	-0,77	1,23	-0,48	0,22
	C=0,2	24,11	41,72	-1,44	1,31	-0,36	0,03
	C=0,3	37,33	56,51	-1,95	1,36	-0,33	0,1
$\theta=2$	C=0,1	2,23	7,29	-0,08	0,99	-0,57	0,44
	C=0,2	2,75	8,65	-0,22	0,99	-0,55	0,35
	C=0,3	3,09	9,87	-0,3	0,98	-0,55	0,35

Tableau 4.19b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	C=0,1	15,01	34,22	-1,23	1,06	-0,13	0,05
	C=0,2	52,85	76,05	-2,45	1,12	-0,14	-0,02
	C=0,3	85,03	95,15	-3,55	1,17	-0,1	-0,03
$\theta=-1$	C=0,1	7,33	21,03	-0,84	1	-0,21	0,02
	C=0,2	27,01	50,53	-1,7	1,05	-0,2	0,03
	C=0,3	54,39	77,99	-2,49	1,09	-0,21	0,05
$\theta=0$	C=0,1	3,33	11,55	-0,48	0,94	-0,32	0,08
	C=0,2	10,01	26,62	-1,05	0,98	-0,26	0,1
	C=0,3	22,55	44,26	-1,54	1,05	-0,29	0,04
$\theta=1$	C=0,1	1,59	5,96	-0,19	0,88	-0,51	0,41
	C=0,2	2,88	9,69	-0,44	0,9	-0,47	0,29
	C=0,3	4,09	13,65	-0,63	0,91	-0,43	0,15
$\theta=2$	C=0,1	0,3	2,09	0,05	0,74	-0,54	0,35
	C=0,2	0,36	2,16	0,03	0,72	-0,59	0,47
	C=0,3	0,36	1,87	0,05	0,7	-0,67	0,65

Les tableaux 4.20a et 4.20b présentent les résultats pour l'indice I_z^* . D'entrée de jeu, nous remarquons que les pourcentages de détection sont légèrement plus élevés lorsque cet indice analyse des tests de 80 items plutôt que de 30 items (veuillez consulter les tableaux 4.8a et 4.8b). De plus, I_z^* présente des pourcentages de détection légèrement plus élevés que ceux de I_z pour 80 items.

C'est sans grande surprise que nous avons remarqué que l'augmentation de C s'accompagne d'une augmentation du pourcentage de détection de l'indice. Tout comme pour I_z , c'est le modèle à deux paramètres qui présente les plus hauts pourcentages de détection aux valeurs $\theta = -1$, $\theta = 0$ et $\theta = 1$. Par contre, c'est généralement le modèle de Rasch qui présente les plus hauts taux de détection pour

les valeurs $\theta = -2$ et $\theta = 2$. Enfin, les pourcentages de détection selon le modèle à trois paramètres sont systématiquement les moins élevés.

Tout comme nous avons pu l'observer pour les résultats portant sur les données de 30 items, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$. Tel que mentionné précédemment, ce résultat est prévisible puisque le seuil $\alpha = 0,01$ est plus restrictif que le seuil $\alpha = 0,05$.

L'élément suivant est en accord avec les résultats présentés à la section précédente : plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de l_z^* est négative et basse. De plus, l'écart-type est presque toujours près des valeurs de la $N(0,1)$.

Les résultats montrent que, pour le modèle de Rasch et pour le modèle à deux paramètres, une augmentation de C fait généralement baisser la valeur du coefficient d'asymétrie. Ajoutons aussi que cette relation n'est pas observée pour le modèle à trois paramètres. Enfin, les coefficients de kurtose sont bornés entre $-0,08$ et $0,50$, soit, des valeurs qui s'approchent de celles de la loi $N(0,1)$.

Tableau 4.20a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	88,12	93,39	-5,14	2,41	-0,25	0
	C=0,2	99,54	99,81	-8,67	2,6	-0,12	-0,08
	C=0,3	99,97	99,99	-10,96	2,58	-0,09	-0,02
$\theta=-1$	C=0,1	63,73	77,65	-3,06	1,82	-0,31	0,06
	C=0,2	93,82	97,51	-5,41	2,04	-0,19	0
	C=0,3	99,13	99,81	-7,1	2,11	-0,16	0
$\theta=0$	C=0,1	29,3	47,27	-1,62	1,41	-0,39	0,23
	C=0,2	63,89	79,23	-2,97	1,6	-0,3	0,04
	C=0,3	83,89	92,5	-4,01	1,7	-0,28	0,09
$\theta=1$	C=0,1	8,73	19,91	-0,68	1,16	-0,46	0,25
	C=0,2	20,46	38,02	-1,31	1,27	-0,39	0,18
	C=0,3	33,74	53,63	-1,84	1,32	-0,35	0,09
$\theta=2$	C=0,1	3,09	8,96	-0,16	1,05	-0,56	0,4
	C=0,2	4,76	12,16	-0,36	1,07	-0,52	0,3
	C=0,3	7,17	16,33	-0,57	1,12	-0,58	0,49
2 paramètres							
$\theta=-2$	C=0,1	88,07	92,99	-5,55	2,77	-0,37	0,05
	C=0,2	99,39	99,79	-9,51	3	-0,13	-0,07
	C=0,3	99,97	99,99	-12,02	2,99	-0,11	0
$\theta=-1$	C=0,1	66,24	78,34	-3,32	2,05	-0,43	0,17
	C=0,2	94,28	97,22	-5,82	2,31	-0,22	-0,05
	C=0,3	99,19	99,71	-7,55	2,35	-0,17	-0,02
$\theta=0$	C=0,1	32,79	49,16	-1,74	1,56	-0,46	0,18
	C=0,2	67,17	81	-3,2	1,75	-0,35	0,14
	C=0,3	87,09	93,92	-4,33	1,81	-0,28	0,12
$\theta=1$	C=0,1	11,48	23,63	-0,79	1,25	-0,49	0,23
	C=0,2	25,12	42,84	-1,47	1,34	-0,36	0,03
	C=0,3	38,91	57,82	-2,01	1,4	-0,34	0,1
$\theta=2$	C=0,1	2,76	8,31	-0,1	1,05	-0,58	0,46
	C=0,2	3,84	10,4	-0,25	1,07	-0,56	0,36
	C=0,3	4,87	12,88	-0,36	1,09	-0,54	0,3

Tableau 4.20b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice I_z^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
3 paramètres							
$\theta=-2$	C=0,1	18,4	37,9	-1,32	1,12	-0,12	0,05
	C=0,2	58,47	79,33	-2,63	1,19	-0,13	-0,02
	C=0,3	88,61	96,32	-3,85	1,26	-0,11	0,01
$\theta=-1$	C=0,1	10,45	24,86	-0,93	1,09	-0,22	0,02
	C=0,2	34,32	57,01	-1,89	1,16	-0,21	0,06
	C=0,3	64,31	83,41	-2,82	1,22	-0,22	0,08
$\theta=0$	C=0,1	5,97	16,12	-0,58	1,07	-0,32	0,08
	C=0,2	16,36	34,69	-1,24	1,13	-0,27	0,14
	C=0,3	32,87	54,19	-1,83	1,22	-0,31	0,1
$\theta=1$	C=0,1	3,66	10,39	-0,28	1,05	-0,52	0,4
	C=0,2	6,69	16,21	-0,59	1,09	-0,47	0,28
	C=0,3	10,36	23,22	-0,84	1,14	-0,43	0,17
$\theta=2$	C=0,1	2,03	6,61	-0,01	1,01	-0,55	0,32
	C=0,2	2,48	7,25	-0,07	1,01	-0,57	0,35
	C=0,3	2,58	7,7	-0,07	1,03	-0,64	0,5

Les tableaux 4.21a et 4.21b présentent les résultats pour l'indice ZU . Premièrement, le pourcentage de détection de cet indice n'est pas systématiquement plus élevé lorsque nous analysons 80 items plutôt que 30 items. Par exemple, les pourcentages de détection obtenus selon le modèle à deux paramètres sont généralement plus élevés lorsque les données analysées comportent 30 items. Ce résultat est donc différent de celui des indices de type vraisemblance. De plus, les pourcentages de détection de ZU sont moins élevés que ceux de I_z et I_z^* pour les données analysées selon le modèle de Rasch. En ce qui a trait au modèle à deux paramètres, le pourcentage de détection de ZU est plus élevé que ceux des indices de type vraisemblance aux valeurs $\theta = -2$, $\theta = 1$ et $\theta = 2$. Enfin, les résultats pour le

modèle à trois paramètres montrent que le pourcentage de détection de ZU est plus élevé que celui des indices l_z et l_z^* aux valeurs $\theta \geq 0$ seulement. Deuxièmement, nous avons remarqué que l'augmentation du niveau de réponse au hasard C engendre aussi une augmentation du pourcentage de détection. Rappelons que ce résultat est aussi observé pour les indices de type vraisemblance. Troisièmement, pour $\theta = -2$, $\theta = -1$ et $\theta = 0$, c'est le modèle à deux paramètres qui présente les pourcentages de détection les plus élevés. Pour $\theta = 1$ et $\theta = 2$, les modèles à deux et à trois paramètres présentent, sans distinction notable, les plus hauts pourcentages de détection. Enfin, le pourcentage de détection selon le modèle de Rasch est celui qui est le plus faible. Rappelons que c'est le modèle à trois paramètres qui engendrait les pourcentages de détection les plus faibles avec les indices l_z et l_z^* . Quatrièmement, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$. Cinquièmement, plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de ZU est positive et élevée. Notons, de plus, que pour les modèles à deux et à trois paramètres, la moyenne augmente considérablement aux valeurs $\theta = 1$ et $\theta = 2$. De leur côté, les écarts-types de ZU sont généralement près des valeurs de la loi $N(0,1)$ aux valeurs $\theta = -2$ et $\theta = -1$. Par contre, lorsque les données sont analysées à l'aide du modèle à deux paramètres ou du modèle à trois paramètres, les écarts-types deviennent très élevés aux valeurs $\theta = 1$ et $\theta = 2$. Ces résultats soutiennent d'ailleurs les observations effectuées à la première étude puisqu'ils montrent la difficulté à utiliser cet indice en contexte pratique. Sixièmement, les coefficients d'asymétrie calculés à partir du modèle de Rasch et du modèle à deux

paramètres sont bornés entre -0,42 et 2,44. De leur côté, les coefficients d'asymétrie obtenus lors des analyses de données à l'aide du modèle à trois paramètres sont beaucoup plus élevés puisqu'ils sont bornés entre 1,00 et 25,05.

Les coefficients de kurtose montrent que la fonction de densité des scores de cet indice s'éloigne parfois grandement de la loi $N(0,1)$. Par exemple, pour des données analysées selon le modèle à trois paramètres, nous calculons des coefficients de kurtose bornés entre 3,12 et 797,22.

Tableau 4.21a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	73,36	83,73	3,48	1,38	-0,24	-0,6
	C=0,2	94,6	97,82	4,79	1,24	-0,42	0,03
	C=0,3	99,17	99,81	5,45	1,06	-0,38	0,15
$\theta=-1$	C=0,1	57,01	70,55	2,76	1,42	-0,14	-0,57
	C=0,2	80,42	90,87	3,58	1,17	-0,19	-0,08
	C=0,3	79,23	92,36	3,49	1,07	0,1	-0,23
$\theta=0$	C=0,1	8,31	20,32	1,16	0,98	0,28	-0,04
	C=0,2	9,75	23,71	1,43	0,84	0,64	0,59
	C=0,3	8,31	19,82	1,43	0,75	1,11	1,56
$\theta=1$	C=0,1	1,31	2,75	0,47	0,57	2,17	8,02
	C=0,2	1,5	3,35	0,65	0,54	2,3	8,38
	C=0,3	1,51	3,55	0,82	0,51	2,44	9,58
$\theta=2$	C=0,1	0,62	1,66	0,85	0,38	2,36	12,54
	C=0,2	0,6	1,97	1	0,38	1,94	9,47
	C=0,3	0,9	3,66	1,18	0,4	1,55	6,27
2 paramètres							
$\theta=-2$	C=0,1	94,44	99,1	4,76	1,8	0,99	0,84
	C=0,2	85,89	96,83	4,46	1,94	0,66	-0,81
	C=0,3	83,57	96,52	4,27	1,78	0,66	-0,61
$\theta=-1$	C=0,1	32,18	52,01	2,58	1,73	1,44	1,31
	C=0,2	41,48	65,97	2,72	1,38	1,33	1,62
	C=0,3	43,96	80,19	2,59	0,82	1,7	5,84
$\theta=0$	C=0,1	10,29	34,38	1,76	0,67	0,86	2,04
	C=0,2	42,89	75,61	2,61	0,96	1,59	5,75
	C=0,3	81,91	95,81	4,05	1,87	2,39	12,97
$\theta=1$	C=0,1	98,8	99,83	6,33	2,66	1,59	4,56
	C=0,2	99,87	99,99	9,27	4,51	2	8,03
	C=0,3	100	100	13,8	6,98	1,61	4,26
$\theta=2$	C=0,1	100	100	29,73	12,31	1,42	3,97
	C=0,2	100	100	37,81	16,2	1,79	7,66
	C=0,3	100	100	48,25	22,13	2,04	8,66

Tableau 4.21b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	C=0,1	4,51	10,49	0,85	0,9	1	3,12
	C=0,2	11,03	20,39	1,36	1	1,85	6,06
	C=0,3	12,86	20,29	1,53	1,16	4,46	73,12
$\theta=-1$	C=0,1	0,56	0,69	0,42	0,44	4,97	62,14
	C=0,2	1,19	2,03	0,71	0,51	4,76	51,82
	C=0,3	7,69	12,13	1,23	1,26	8,32	135,56
$\theta=0$	C=0,1	18,55	33,4	1,83	1,19	2,5	11,96
	C=0,2	45,34	62,25	3,08	2,71	5	48,34
	C=0,3	75,19	85,74	6,1	7,09	5,41	51,82
$\theta=1$	C=0,1	99,44	99,92	11,87	8,51	2,94	17,88
	C=0,2	99,85	99,97	18,89	14,56	2,96	20,48
	C=0,3	99,99	100	33,18	27,8	3,3	21,75
$\theta=2$	C=0,1	100	100	69,65	53,24	6,03	102,14
	C=0,2	100	100	99,57	91,65	9,71	212,09
	C=0,3	100	100	159,98	312,42	25,05	797,22

Les tableaux 4.22a et 4.22b présentent les résultats pour l'indice ZU^* . D'abord, il est important de souligner que l'analyse de tests contenant 80 items (plutôt que 30 items) ne fait pas systématiquement augmenter le pourcentage de détection de cet indice. Par exemple, nous obtenons des pourcentages de détection moins élevés pour 80 items lorsque les données sont analysées à l'aide du modèle à trois paramètres. Ajoutons que généralement, les pourcentages de détection de ZU^* sont globalement moins élevés que ceux de I_z , I_z^* et ZU . Ensuite, nous avons remarqué que l'augmentation du niveau de réponse au hasard C n'engendre pas systématiquement une augmentation du pourcentage de détection.

C'est le modèle de Rasch qui présente les pourcentages de détection les plus élevés. Ce résultat est différent dans le cas de ZU puisque c'est le modèle de Rasch qui présente les pourcentages de détection les moins élevés. Notons, de plus, que les pourcentages de détection pour le modèle à trois paramètres sont les moins élevés pour 80 items, alors que c'est ce modèle qui présentait les pourcentages les plus élevés pour la situation à 30 items (voir le tableau 4.10b).

Plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de ZU^* est positive et élevée. Notons que la moyenne diminue vers zéro au fur et à mesure que la valeur de θ augmente. De leur côté, les écarts-types sont généralement beaucoup plus élevés que la valeur de la loi $N(0,1)$ aux valeurs $\theta = -2$ et $\theta = -1$ et beaucoup plus faible que la valeur de la loi $N(0,1)$ lorsque θ augmente vers une valeur de deux. Ensuite, les coefficients d'asymétrie augmentent avec la valeur de θ . Les résultats montrent que la fonction de densité des scores de cet indice ne suit pas la loi $N(0,1)$. Par exemple, ces coefficients sont bornés entre 0,58 et 17,26 pour le modèle de Rasch, entre 3,04 et 121,23 pour le modèle à deux paramètres et entre 6,90 et 120,74 pour le modèle à trois paramètres. La valeur des coefficients de kurtose est aussi très élevée. Par exemple, ces coefficients sont bornés entre 0,49 et 412,72 pour le modèle de Rasch, entre 14,20 et 14 787,53 pour le modèle à deux paramètres et entre 190,20 et 14 707,11 pour le modèle à trois paramètres.

Tableau 4.22a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	85,77	89,15	12,17	9,12	0,86	0,76
	C=0,2	98,36	99,16	16,07	7,88	0,63	0,63
	C=0,3	99,87	99,95	15,73	5,93	0,58	0,49
$\theta=-1$	C=0,1	72,93	79,26	5,45	4,24	0,96	1,2
	C=0,2	92,12	95,64	7,15	3,81	0,78	0,89
	C=0,3	93,94	97,55	6,7	3,3	0,8	0,85
$\theta=0$	C=0,1	22,74	35,35	1,41	1,62	1,33	2,46
	C=0,2	27,26	42,2	1,8	1,69	1,75	4,44
	C=0,3	24,57	38,37	1,79	1,84	2,54	9,34
$\theta=1$	C=0,1	3,43	5,05	0,2	1,15	6,19	54,05
	C=0,2	4,08	5,94	0,33	1,28	6,32	54,89
	C=0,3	4,3	6,33	0,41	1,48	7,58	77,68
$\theta=2$	C=0,1	1,54	2,13	0,03	1,15	15,95	352,58
	C=0,2	1,29	1,94	0,04	1,13	17,26	412,72
	C=0,3	1,34	1,91	0,08	1,2	16,69	366,08
2 paramètres							
$\theta=-2$	C=0,1	63,33	66,73	62,6	158,9	3,91	18,95
	C=0,2	88,02	90,99	60,09	99,4	3,04	14,62
	C=0,3	95,67	97,5	35,93	49,79	3,1	14,2
$\theta=-1$	C=0,1	50,43	57,73	13,95	31,59	4	20,59
	C=0,2	63,85	73,37	10,55	18,69	4,31	28,49
	C=0,3	51,43	63,49	4,87	10,34	12,67	271,15
$\theta=0$	C=0,1	9,22	13,58	0,84	2,1	9,16	169,41
	C=0,2	5,01	8,57	0,62	1,27	16,53	687,59
	C=0,3	2	3,21	0,37	1,37	58,28	4534,18
$\theta=1$	C=0,1	0,07	0,09	0,02	0,34	82,96	8016,79
	C=0,2	0,11	0,12	0,02	0,19	51,55	3726,88
	C=0,3	0,11	0,11	0,03	0,22	40,72	2106,38
$\theta=2$	C=0,1	0,01	0,01	0	0,64	121,23	14787,53
	C=0,2	0	0	0	0,03	26,76	892,02
	C=0,3	0,01	0,01	0	0,06	79,47	7798,67

Tableau 4.22b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZU^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
3 paramètres							
$\theta=-2$	C=0,1	9,47	20,61	0,94	1,23	6,9	201,4
	C=0,2	20,01	34,05	1,59	1,94	10,47	229,68
	C=0,3	20,16	33,09	1,93	3,79	11,38	190,2
$\theta=-1$	C=0,1	0,61	0,79	0,16	0,84	19,73	646,48
	C=0,2	0,57	0,88	0,29	1,34	45,29	2980,01
	C=0,3	0,47	0,99	0,36	2,11	35,18	1537,29
$\theta=0$	C=0,1	0,03	0,06	0,03	2,08	119,25	14456,68
	C=0,2	0,01	0,11	0,04	0,3	56,5	4179,05
	C=0,3	0,02	0,24	0,08	2,81	120,74	14707,11
$\theta=1$	C=0,1	0,02	0,07	0	0,26	106,09	12312,29
	C=0,2	0,01	0,07	0	0,08	16,62	360,63
	C=0,3	0,03	0,07	0,01	0,09	25,82	1032,3
$\theta=2$	C=0,1	0,02	0,03	0	0,06	59,9	4106,42
	C=0,2	0,01	0,03	0	0,05	63,97	5004,66
	C=0,3	0,01	0,01	0	0,04	57,79	3929,86

Les résultats pour l'indice ZW sont présentés aux tableaux 4.23a et 4.23b.

Premièrement, nous observons globalement des pourcentages de détection plus élevés lorsque nous analysons 80 items plutôt que 30 items. Ensuite, nous remarquons que l'augmentation de C est généralement accompagnée par une augmentation du pourcentage de détection des patrons de réponses au hasard. De plus, le pourcentage de détection de ZW est globalement moins élevé que celui de l_z, l_z^* et ZU .

Deuxièmement, c'est le modèle à deux paramètres qui présente les pourcentages de patrons de réponses détectés les plus élevés. À l'opposé, c'est le modèle à trois paramètres qui engendre les pourcentages de détection les moins élevés.

Troisièmement, lorsque la valeur θ augmente, le pourcentage de patrons détectés tend à diminuer. Notons qu'un résultat comparable a aussi été soulevé pour les indices l_z et l_z^* .

Quatrièmement, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$.

Cinquièmement, plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de ZW est élevée.

Sixièmement, le coefficient d'asymétrie est généralement négatif et relativement près des valeurs de la loi $N(0,1)$ pour tous les modèles de réponse à l'item. Nous observons aussi des coefficients de kurtose dont la valeur est relativement en accord avec les quantiles de la loi $N(0,1)$.

Tableau 4.23a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	50,3	68,36	2,51	1,23	-0,11	-0,09
	C=0,2	94,96	98,01	4,1	1,20	-0,29	-0,04
	C=0,3	98,43	98,17	6,01	1,05	-0,31	0,18
$\theta=-1$	C=0,1	28,16	48,77	1,69	1,11	-0,13	-0,04
	C=0,2	75,28	89,55	3,02	1,05	-0,29	0,13
	C=0,3	92,97	98,54	3,92	1,03	-0,28	0,04
$\theta=0$	C=0,1	11,58	24,74	1,01	1,11	-0,12	-0,09
	C=0,2	37,31	56,32	1,96	1,06	-0,18	0,12
	C=0,3	64,31	78,51	2,33	1,01	-0,19	0,12
$\theta=1$	C=0,1	1,84	9,3	0,46	1,03	-0,05	-0,18
	C=0,2	7,59	22,75	1,01	1,03	-0,17	-0,14
	C=0,3	11,82	34,7	0,93	0,99	-0,16	-0,09
$\theta=2$	C=0,1	0,12	1,38	0,03	0,96	-0,07	-0,12
	C=0,2	0,29	1,59	0,08	0,92	-0,06	-0,11
	C=0,3	0,19	1,02	0,11	0,83	-0,03	0,06
2 paramètres							
$\theta=-2$	C=0,1	51,17	69,38	2,59	1,26	-0,06	-0,03
	C=0,2	94,73	98,11	4,67	1,24	-0,23	0,06
	C=0,3	99,75	99,94	6,01	1,11	-0,26	0,15
$\theta=-1$	C=0,1	30,17	50,37	1,96	1,19	-0,08	0,01
	C=0,2	78,19	89,87	3,48	1,18	-0,21	0,09
	C=0,3	95,37	98,59	4,51	1,1	-0,2	0,01
$\theta=0$	C=0,1	11,58	26,15	1,22	1,14	-0,04	-0,06
	C=0,2	40,36	61,85	2,29	1,14	-0,1	0,08
	C=0,3	69,81	85,54	3,12	1,1	-0,15	0,08
$\theta=1$	C=0,1	3,98	12,37	0,69	1,09	-0,03	-0,13
	C=0,2	12,09	27,91	1,31	1,09	-0,11	-0,1
	C=0,3	22,45	43,43	1,76	1,06	-0,14	-0,02
$\theta=2$	C=0,1	0,41	2,94	0,1	0,99	-0,04	-0,06
	C=0,2	0,63	3,37	0,23	0,96	-0,05	-0,06
	C=0,3	0,67	3,57	0,31	0,92	-0,03	0,04

Tableau 4.23b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	C=0,1	7,85	22,43	1,21	0,98	-0,11	0,06
	C=0,2	38,79	64,53	2,3	0,94	-0,09	-0,03
	C=0,3	74,91	91,02	3,18	0,9	-0,13	-0,02
$\theta=-1$	C=0,1	3,64	13,05	0,89	0,95	-0,07	0
	C=0,2	17,33	39,63	1,71	0,93	-0,09	-0,01
	C=0,3	41,39	68,77	2,37	0,88	-0,11	0,01
$\theta=0$	C=0,1	1,2	6,16	0,54	0,92	-0,03	-0,04
	C=0,2	4,55	17,11	1,1	0,9	-0,12	-0,01
	C=0,3	11,16	30,49	1,5	0,88	-0,09	-0,08
$\theta=1$	C=0,1	0,3	2,25	0,24	0,86	0,03	-0,12
	C=0,2	0,57	3,64	0,47	0,84	0,01	-0,09
	C=0,3	0,48	4,27	0,58	0,79	0	-0,08
$\theta=2$	C=0,1	0,01	0,22	-0,03	0,71	0,02	-0,13
	C=0,2	0,01	0,21	-0,01	0,67	0,06	-0,07
	C=0,3	0,01	0,11	-0,05	0,63	0,12	-0,15

Les résultats pour l'indice ZW^* sont présentés dans les tableaux 4.24a et 4.24b. Tout d'abord, les pourcentages de détection sont légèrement plus élevés pour 80 items que pour 30 items. De plus, le pourcentage de détection de cet indice est plus élevé que celui de ZW , ZU^* et l_z . Par contre, il est moins élevé que celui de ZU et de l_z^* .

L'augmentation du niveau C est généralement accompagnée par une hausse du pourcentage de détection de patrons de réponses au hasard. Ce résultat est aussi valable pour les indices l_z , l_z^* , ZU et ZW . De plus, nous avons observé que le pourcentage de patrons de réponses détectés est généralement à son niveau le plus élevé lorsque le modèle de Rasch est utilisé alors que c'est le modèle à trois

paramètres qui engendre les pourcentages de détection les moins élevés. Ces résultats ont aussi été obtenus par d'autres indices analysés dans cette section et dans la section portant sur l'analyse de données contenant 30 items.

À l'instar des indices de vraisemblance et de ZW , nous observons que lorsque la valeur de θ augmente, le pourcentage de patrons détectant le hasard tend à diminuer. De plus, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$.

Nous remarquons que plus le pourcentage de patrons de réponses détecté est élevé, plus la moyenne des scores de ZW^* est élevée. Ce résultat a aussi été mis en exergue pour les autres indices de type carré-moyen analysé dans cette thèse. Pour le modèle de Rasch et pour le modèle à deux paramètres, il est intéressant de remarquer que la valeur de l'écart-type tend à diminuer au fur et à mesure que la valeur de θ augmente. En ce qui a trait au modèle à trois paramètres, les écarts-types sont près de l'unité, soit, la valeur de la loi $N(0,1)$. Ensuite, les coefficients d'asymétrie sont positifs et plus élevés pour le modèle à deux paramètres que pour les autres modèles de réponse à l'item. Finalement, nous observons des coefficients de kurtose dont la valeur est relativement près de ceux de la loi $N(0,1)$.

Tableau 4.24a : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	C=0,1	68,74	82,54	3,23	1,7	0,29	0,14
	C=0,2	97,97	99,31	6,32	2,06	0,19	-0,02
	C=0,3	99,89	99,96	8,8	2,15	0,09	-0,04
$\theta=-1$	C=0,1	46,96	65,05	2,27	1,5	0,29	0,09
	C=0,2	88,91	95,01	4,46	1,76	0,18	-0,02
	C=0,3	98,64	99,6	6,25	1,85	0,11	-0,08
$\theta=0$	C=0,1	23,61	41,77	1,42	1,3	0,29	0,13
	C=0,2	59,99	76,86	2,76	1,49	0,24	0,02
	C=0,3	83,24	92,51	3,86	1,57	0,18	0,03
$\theta=1$	C=0,1	8,2	20,1	0,68	1,15	0,32	0,07
	C=0,2	21,21	40,05	1,35	1,24	0,24	0
	C=0,3	36,29	57,11	1,91	1,27	0,19	-0,04
$\theta=2$	C=0,1	2,83	8,89	0,19	1,05	0,3	0,04
	C=0,2	4,37	12,69	0,41	1,07	0,24	-0,07
	C=0,3	6,75	17,54	0,64	1,1	0,23	0,05
2 paramètres							
$\theta=-2$	C=0,1	66,01	80,17	3,11	1,7	0,31	0,12
	C=0,2	97,62	99,16	6,19	2,05	0,14	-0,04
	C=0,3	99,89	99,97	8,59	2,1	0,05	-0,04
$\theta=-1$	C=0,1	45,28	63,82	2,24	1,5	0,33	0,12
	C=0,2	87,63	94,35	4,32	1,74	0,16	-0,01
	C=0,3	98,07	99,43	5,92	1,8	0,12	-0,05
$\theta=0$	C=0,1	21,99	38,57	1,33	1,32	0,35	0,11
	C=0,2	56,91	74,43	2,65	1,49	0,29	0,16
	C=0,3	82,32	91,85	3,81	1,59	0,21	0,05
$\theta=1$	C=0,1	10,11	22,13	0,75	1,2	0,35	0,07
	C=0,2	24,31	42,94	1,45	1,29	0,26	-0,04
	C=0,3	39,16	59,43	2,01	1,33	0,22	0,03
$\theta=2$	C=0,1	2,41	8,07	0,11	1,04	0,32	0,06
	C=0,2	3,24	10,15	0,26	1,05	0,29	-0,01
	C=0,3	4,05	12,04	0,37	1,07	0,28	0,05

Tableau 4.24b : Pourcentage de détection de la réponse au hasard et statistiques descriptives pour l'indice ZW^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
3 paramètres							
$\theta=-2$	C=0,1	18,42	38,41	1,34	1,11	0,09	0,04
	C=0,2	59,51	80,24	2,64	1,16	0,08	-0,07
	C=0,3	89,14	96,53	3,8	1,19	0,03	-0,07
$\theta=-1$	C=0,1	11,31	27,04	0,99	1,09	0,15	-0,01
	C=0,2	37,34	61,58	1,99	1,14	0,11	-0,02
	C=0,3	67,98	86,03	2,88	1,15	0,07	-0,03
$\theta=0$	C=0,1	6,21	17,44	0,64	1,07	0,2	-0,01
	C=0,2	17,77	38	1,32	1,1	0,1	-0,04
	C=0,3	33,17	56,13	1,84	1,13	0,09	-0,13
$\theta=1$	C=0,1	3,31	10,84	0,32	1,04	0,26	-0,04
	C=0,2	5,87	16,75	0,61	1,06	0,24	-0,02
	C=0,3	8,06	21,19	0,78	1,07	0,19	-0,11
$\theta=2$	C=0,1	1,49	5,94	0	1,01	0,29	-0,11
	C=0,2	1,57	6,37	0,04	1,01	0,3	-0,1
	C=0,3	1,39	5,84	-0,01	1	0,33	-0,11

4.3.2.2 Inattention

Les tableaux 4.25a et 4.25b présentent les résultats pour la détection de l'inattention à l'aide de l'indice I_z . Pour commencer, nous observons que les pourcentages de détection augmentent lorsque nous analysons 80 items. De plus, lorsque le niveau d'inattention D augmente, le pourcentage de détection de patrons de réponses inattentives augmente lui aussi.

Le pourcentage de patrons de réponses détectés est à son niveau maximal lorsque le modèle à deux paramètres est utilisé et c'est le modèle à trois paramètres

qui engendre les pourcentages de détection les moins élevés. Ce constat a aussi été observé dans la section portant sur l'analyse de données de 30 items.

Lorsque θ augmente, nous remarquons que le pourcentage de détection tend aussi à augmenter. Ce résultat est opposé à celui observé pour la réponse au hasard puisque nous avons examiné une baisse du pourcentage de patrons de réponses détectés lorsque θ augmente. De plus, cela est en accord avec les résultats présentés dans la section portant sur l'analyse de données de 30 items.

Sans surprise, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$.

En ce qui a trait aux statistiques descriptives, il est possible de remarquer que plus le pourcentage de patrons de réponses détecté est élevé, plus la moyenne des scores de l_z s'éloigne négativement de zéro et plus l'écart-type est positif et élevé.

Nous observons que lorsque la valeur de θ augmente, les coefficients d'asymétrie et de kurtose ont tendance à s'approcher de zéro, soit, la valeur de la loi $N(0,1)$. De plus ces coefficients sont bornés entre -0,64 et 0,69. Enfin, c'est le modèle à trois paramètres qui présente les résultats les plus près de la loi $N(0,1)$.

Tableau 4.25a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice I_z (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	2,13	6,85	-0,12	0,95	-0,56	0,36
	D=0,2	3,02	9,25	-0,31	0,96	-0,58	0,55
	D=0,3	3,83	10,86	-0,43	0,96	-0,64	0,69
$\theta=-1$	D=0,1	8,12	19,45	-0,67	1,15	-0,51	0,42
	D=0,2	19,74	36,37	-1,28	1,25	-0,38	0,07
	D=0,3	31,08	51,4	-1,77	1,29	-0,37	0,19
$\theta=0$	D=0,1	29,3	46,36	-1,61	1,42	-0,39	0,14
	D=0,2	63,97	79,63	-2,98	1,59	-0,3	0,06
	D=0,3	84,49	92,96	-3,99	1,66	-0,25	0,04
$\theta=1$	D=0,1	63,18	76,94	-3,03	1,82	-0,33	0,08
	D=0,2	94,09	97,47	-5,41	2,06	-0,24	0,05
	D=0,3	99,17	99,73	-7,08	2,11	-0,19	0,07
$\theta=2$	D=0,1	87,62	93,41	-5,06	2,36	-0,27	-0,03
	D=0,2	99,55	99,86	-8,61	2,6	-0,13	-0,08
	D=0,3	99,99	100	-10,97	2,61	-0,11	-0,04
2 paramètres							
$\theta=-2$	D=0,1	2,57	7,84	-0,16	0,99	-0,54	0,36
	D=0,2	4,12	11,29	-0,36	1,03	-0,55	0,4
	D=0,3	5,19	13,52	-0,47	1,04	-0,51	0,26
$\theta=-1$	D=0,1	12,24	24,33	-0,82	1,27	-0,52	0,33
	D=0,2	27,96	44,38	-1,55	1,42	-0,42	0,06
	D=0,3	40,94	59,91	-2,09	1,45	-0,37	0,16
$\theta=0$	D=0,1	37,54	53,79	-1,94	1,66	-0,48	0,13
	D=0,2	71,22	83,22	-3,51	1,92	-0,36	0,09
	D=0,3	87,38	93,8	-4,61	1,99	-0,26	0,07
$\theta=1$	D=0,1	66,63	78,73	-3,44	2,18	-0,47	0,12
	D=0,2	94,83	97,67	-6,22	2,51	-0,28	-0,01
	D=0,3	99,44	99,81	-8,26	2,57	-0,18	-0,08
$\theta=2$	D=0,1	87,68	92,84	-5,45	2,76	-0,4	0,05
	D=0,2	99,42	99,81	-9,41	3,14	-0,23	-0,06
	D=0,3	99,99	100	-12,18	3,16	-0,14	-0,05

Tableau 4.25b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice l_z (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	1,39	6,01	-0,06	0,99	-0,19	-0,02
	D=0,2	1,78	7,43	-0,17	0,99	-0,21	-0,04
	D=0,3	2,3	8,13	-0,21	1	-0,22	-0,01
$\theta=-1$	D=0,1	4,87	14,75	-0,54	1,04	-0,23	0,09
	D=0,2	8,45	22,25	-0,84	1,05	-0,2	-0,03
	D=0,3	9,11	24,31	-0,92	1,04	-0,18	0
$\theta=0$	D=0,1	27,15	46,49	-1,54	1,31	-0,17	-0,18
	D=0,2	45,35	67,69	-2,2	1,19	-0,07	-0,1
	D=0,3	44,34	68,77	-2,19	1,11	-0,09	0,02
$\theta=1$	D=0,1	62,45	76,77	-2,9	1,64	-0,16	-0,21
	D=0,2	86,14	94,28	-3,8	1,35	0,07	0,06
	D=0,3	86,82	95,58	-3,67	1,19	-0,04	-0,05
$\theta=2$	D=0,1	85,99	92,23	-4,52	1,99	0	-0,26
	D=0,2	98,17	99,37	-5,51	1,48	0,12	-0,01
	D=0,3	98,99	99,79	-5,17	1,25	-0,01	-0,01

Les tableaux 4.26a et 4.26b présentent les résultats pour la détection de l'inattention à l'aide de l'indice l_z^* . Premièrement, les pourcentages de détection sont plus élevés lorsque nous analysons 80 items. Ensuite, nous observons que le pourcentage de détection de l_z est moins élevé que celui de l_z^* . De plus, lorsque le niveau d'inattention D augmente, le pourcentage de détection de patrons de réponses inattentives augmente lui aussi. Il est à noter que nous avons obtenu une tendance comparable avec l_z .

Deuxièmement, le pourcentage de patrons de réponses détectés est plus élevé lorsque le modèle à deux paramètres est utilisé. À l'opposé, c'est le modèle à trois

paramètres qui engendre les pourcentages de détection les moins élevés. Ce constat a aussi été observé dans la section portant sur l'utilisation de l_z^* sur des tests de 30 items.

Troisièmement, lorsque θ augmente, nous remarquons que le pourcentage de patrons détectés tend à augmenter.

Quatrièmement, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$.

Cinquièmement, il est possible de remarquer que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de l_z est négative et faible et plus l'écart-type est positif et élevé.

Finalement, nous observons, que lorsque θ augmente, les coefficients d'asymétrie et de kurtose ont tendance à s'approcher de la valeur de la loi $N(0,1)$. Par contre, il est important de soulever la nuance suivante : c'est le modèle à trois paramètres qui présente les moyennes les plus près de la loi $N(0,1)$. Enfin, l'asymétrie et la kurtose sont aussi près des valeurs de la $N(0,1)$. De plus, ce sont les scores obtenus à l'aide du modèle à trois paramètres qui présentent la meilleure adéquation à la fonction de densité de la loi $N(0,1)$.

Tableau 4.26a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice I_z^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	3,08	8,54	-0,15	1,04	-0,56	0,35
	D=0,2	4,81	12,52	-0,38	1,08	-0,57	0,48
	D=0,3	6,99	15,84	-0,55	1,13	-0,62	0,62
$\theta=-1$	D=0,1	8,39	19,95	-0,68	1,16	-0,51	0,42
	D=0,2	20,56	37,35	-1,31	1,27	-0,38	0,07
	D=0,3	32,93	52,99	-1,82	1,32	-0,37	0,18
$\theta=0$	D=0,1	29,30	46,36	-1,61	1,42	-0,39	0,14
	D=0,2	63,99	79,64	-2,98	1,59	-0,30	0,06
	D=0,3	84,61	93,01	-4,01	1,66	-0,25	0,04
$\theta=1$	D=0,1	63,19	76,96	-3,04	1,82	-0,33	0,08
	D=0,2	94,09	97,47	-5,41	2,06	-0,24	0,05
	D=0,3	99,17	99,73	-7,08	2,11	-0,19	0,07
$\theta=2$	D=0,1	88,29	93,68	-5,13	2,38	-0,26	-0,04
	D=0,2	99,57	99,87	-8,64	2,60	-0,12	-0,07
	D=0,3	99,99	100,00	-10,97	2,61	-0,11	-0,04
2 paramètres							
$\theta=-2$	D=0,1	3,29	9,23	-0,18	1,06	-0,54	0,35
	D=0,2	5,67	13,86	-0,41	1,12	-0,55	0,39
	D=0,3	7,48	17,2	-0,54	1,16	-0,50	0,22
$\theta=-1$	D=0,1	12,61	24,73	-0,83	1,29	-0,53	0,34
	D=0,2	29,06	45,18	-1,59	1,45	-0,43	0,07
	D=0,3	43,19	61,5	-2,17	1,50	-0,38	0,14
$\theta=0$	D=0,1	37,59	53,86	-1,94	1,66	-0,48	0,13
	D=0,2	71,27	83,24	-3,51	1,93	-0,36	0,09
	D=0,3	87,49	93,85	-4,63	2,00	-0,26	0,07
$\theta=1$	D=0,1	66,74	78,79	-3,45	2,18	-0,47	0,12
	D=0,2	94,85	97,67	-6,23	2,51	-0,28	-0,01
	D=0,3	99,44	99,81	-8,27	2,58	-0,18	-0,08
$\theta=2$	D=0,1	88,16	93,05	-5,55	2,79	-0,38	0,04
	D=0,2	99,46	99,82	-9,48	3,14	-0,22	-0,06
	D=0,3	99,99	100,00	-12,2	3,16	-0,14	-0,05

Tableau 4.26b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice I_z^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	1,71	6,70	-0,07	1,01	-0,19	-0,03
	D=0,2	2,11	7,99	-0,19	1,01	-0,21	-0,05
	D=0,3	2,49	8,65	-0,24	1,02	-0,21	-0,01
$\theta=-1$	D=0,1	6,25	17,18	-0,59	1,11	-0,22	0,08
	D=0,2	10,16	24,58	-0,89	1,10	-0,18	-0,04
	D=0,3	10,37	25,86	-0,96	1,07	-0,17	0,01
$\theta=0$	D=0,1	33,12	51,71	-1,72	1,43	-0,15	-0,18
	D=0,2	50,98	71,27	-2,36	1,26	-0,06	-0,07
	D=0,3	48,19	71,49	-2,30	1,14	-0,07	0,02
$\theta=1$	D=0,1	68,99	81,15	-3,31	1,83	-0,13	-0,21
	D=0,2	89,69	95,54	-4,16	1,46	0,06	0,13
	D=0,3	89,22	96,39	-3,87	1,24	-0,04	0,02
$\theta=2$	D=0,1	89,97	94,37	-5,28	2,26	0,03	-0,18
	D=0,2	98,97	99,59	-6,09	1,60	0,05	0,13
	D=0,3	99,38	99,84	-5,47	1,30	-0,05	0,09

Les résultats pour ZU sont présentés aux tableaux 4.27a et 4.27b. D'abord, le fait d'analyser 80 items plutôt que 30 items ne fait pas systématiquement augmenter le pourcentage de détection. Par exemple, les pourcentages de détection selon le modèle à deux paramètres sont légèrement plus élevés pour 30 items (figure 4.15a). Ensuite, le pourcentage de détection de ZU est généralement moins élevé que celui des indices de type vraisemblance. De plus, l'augmentation du niveau d'inattention D est généralement accompagnée par une hausse du pourcentage de détection pour le modèle de Rasch et pour le modèle à deux paramètres. Par contre, nous n'avons pas trouvé cette relation entre D et le pourcentage de détection pour le modèle à trois paramètres.

Le pourcentage de patrons de réponses détectés est généralement plus élevé pour le modèle à deux paramètres. Par contre, il faut être prudent avec ces résultats puisque nous avons démontré dans l'étude 1 que la moyenne et l'écart-type des scores de ZU sont assez éloignés des valeurs attendues de la loi $N(0,1)$.

Pour le modèle de Rasch et pour le modèle à trois paramètres, nous observons que le pourcentage de patrons de réponses détectant l'inattention augmente lorsque la valeur de θ augmente aussi. Ce résultat a aussi été décrit pour d'autres indices tels que l_z et l_z^* . Ensuite, nous avons découvert, sans surprise, que le pourcentage de détection est plus élevé lorsque $\alpha = 0,05$.

Nous remarquons que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne et l'écart-type des scores de ZU sont élevés. De leur côté, les coefficients d'asymétrie sont bornés entre -0,39 et 3,38, des valeurs relativement près de celles d'une loi $N(0,1)$. Les coefficients de kurtose sont surtout élevés aux valeurs $\theta = -2$ et $\theta = -1$ pour le modèle de Rasch et le modèle à deux paramètres. Pour le modèle à trois paramètres, ce coefficient est surtout élevé pour la valeur $\theta = 2$. De plus, les résultats montrent que la fonction de densité des scores de ZU est moins en accord avec la loi $N(0,1)$ que la fonction de densité des scores des indices de type vraisemblance.

Tableau 4.27a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	0,48	1,48	0,85	0,37	2,02	9,58
	D=0,2	0,61	2,14	1,00	0,38	1,80	7,94
	D=0,3	0,89	3,91	1,18	0,41	1,68	7,18
$\theta=-1$	D=0,1	1,33	2,79	0,47	0,57	2,19	8,21
	D=0,2	1,43	3,20	0,65	0,54	2,38	9,22
	D=0,3	1,49	3,40	0,81	0,50	2,35	8,95
$\theta=0$	D=0,1	8,41	20,20	1,15	0,99	0,29	0,01
	D=0,2	9,77	23,84	1,44	0,84	0,66	0,64
	D=0,3	8,13	19,87	1,43	0,75	1,12	1,63
$\theta=1$	D=0,1	55,89	69,65	2,73	1,43	-0,12	-0,63
	D=0,2	79,93	90,71	3,57	1,19	-0,21	-0,10
	D=0,3	79,78	92,88	3,50	1,06	0,08	-0,26
$\theta=2$	D=0,1	73,53	83,91	3,48	1,36	-0,22	-0,56
	D=0,2	94,43	97,81	4,76	1,25	-0,39	-0,02
	D=0,3	99,11	99,80	5,43	1,07	-0,36	0,11
2 paramètres							
$\theta=-2$	D=0,1	99,98	100,00	9,57	3,95	1,47	4,04
	D=0,2	100,00	100,00	12,28	5,29	1,44	4,17
	D=0,3	100,00	100,00	15,87	7,09	1,58	4,73
$\theta=-1$	D=0,1	27,87	59,37	2,25	0,86	1,34	3,95
	D=0,2	65,47	88,10	3,25	1,35	1,70	5,42
	D=0,3	89,63	97,73	4,70	2,17	1,71	4,91
$\theta=0$	D=0,1	17,85	24,67	1,75	1,49	2,08	3,99
	D=0,2	21,52	41,54	1,99	1,12	1,71	3,88
	D=0,3	28,37	64,07	2,27	0,82	1,30	4,13
$\theta=1$	D=0,1	68,43	89,91	3,71	1,87	1,24	0,65
	D=0,2	57,75	80,55	3,84	2,23	0,83	-0,71
	D=0,3	60,10	78,82	3,77	2,06	0,75	-0,65
$\theta=2$	D=0,1	99,98	100,00	10,19	4,55	1,85	6,55
	D=0,2	98,22	99,89	6,11	2,37	1,05	3,23
	D=0,3	91,10	99,04	5,23	2,19	0,26	-1,33

Tableau 4.27b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	0,87	3,20	0,07	0,97	0,25	0,44
	D=0,2	0,92	3,82	0,17	0,98	0,15	0,06
	D=0,3	0,83	3,93	0,20	0,99	0,08	-0,07
$\theta=-1$	D=0,1	13,17	18,09	0,92	1,55	2,07	5,07
	D=0,2	12,43	21,48	1,07	1,35	1,36	3,27
	D=0,3	7,40	17,50	0,98	1,11	0,66	1,69
$\theta=0$	D=0,1	26,47	28,02	2,12	2,28	1,79	2,33
	D=0,2	38,30	45,6	2,40	2,03	1,40	1,74
	D=0,3	32,96	49,63	2,11	1,40	1,21	3,10
$\theta=1$	D=0,1	55,53	68,17	3,59	2,42	1,6	5,76
	D=0,2	50,18	55,90	3,22	2,28	1,07	0,43
	D=0,3	61,00	71,39	3,09	1,64	0,90	1,54
$\theta=2$	D=0,1	88,46	92,72	9,21	8,95	3,38	20,37
	D=0,2	60,61	68,27	3,98	2,65	1,66	8,70
	D=0,3	72,37	79,25	3,80	1,79	0,46	0,48

Voici les résultats pour ZU^* (tableaux 4.27a et 4.27b). Premièrement, les pourcentages de détection sont moins élevés que ceux des indices I_z , I_z^* et ZU . Dans ce cas-ci, nous obtenons des pourcentages de détection plus élevés pour le modèle de Rasch et le modèle à deux paramètres lorsque 80 items sont analysés. Par contre, les pourcentages de détection étaient plus élevés lorsque des tests de 30 items étaient analysés à l'aide du modèle à trois paramètres.

Deuxièmement, il n'existe pas de relation entre le niveau d'inattention D et le pourcentage de détection de cet indice. Rappelons que ce résultat est différent de

celui obtenu à l'aide des indices l_z et l_z^* puisque leur pourcentage de détection augmentait lorsque la valeur de D s'élevait aussi.

Troisièmement, le pourcentage de patrons de réponses détectés est généralement plus élevé pour le modèle à deux paramètres que pour le modèle de Rasch. Lorsque c'est le modèle à trois paramètres qui est utilisé, les pourcentages de détection sont les moins élevés (alors que c'est ce modèle qui présentait les pourcentages le plus élevés dans la situation d'analyse à 30 items).

Quatrièmement, nous observons que le pourcentage de patrons de réponses détectant l'inattention augmente lorsque θ augmente aussi. Ensuite, nous avons découvert, sans surprise, que le pourcentage de détection est plus élevé lorsque $\alpha=0,05$.

Cinquièmement, nous remarquons que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de ZU^* est élevée. Il est important d'ajouter que des moyennes très élevées ont été calculées pour $\theta = 1$ et $\theta = 2$. Par exemple, pour le modèle à deux paramètres, nous observons une moyenne de 175,06 à la valeur $\theta = 2$ (situation $D = 0,1$). Pour le modèle à trois paramètres, nous observons une moyenne de 94,52 à la valeur $\theta = 2$ (valeur $D = 0,1$). Des écarts-types très élevés ont aussi été calculés. Par exemple, pour le modèle à deux paramètres, nous observons un écart-type de 472,88 à la valeur $\theta = 2$ (valeur $D = 0,1$).

Pour le modèle à trois paramètres, nous observons un écart-type de 449,66 à la valeur $\theta = 2$ (valeur $D = 0,1$).

Sixièmement, l'asymétrie atteint parfois des valeurs qui sont très éloignées de la loi $N(0,1)$ et, cela, pour tous les modèles de réponse à l'item étudiés. Par exemple, pour le modèle à deux paramètres, nous observons un coefficient d'asymétrie de 118,76 à $\theta = -1$ (valeur $D = 0,2$). De plus, les coefficients de kurtose présentent parfois des valeurs excessivement élevées lorsque le niveau θ tend vers moins deux pour le modèle de Rasch et pour le modèle à deux paramètres. Pour le modèle à trois paramètres, le coefficient de kurtose devient très élevé lorsque $\theta = 1$ et $\theta = 2$. Enfin, ces résultats suggèrent d'utiliser ZU^* avec précaution puisque les scores de cet indice présentent une mauvaise adéquation à la loi $N(0,1)$.

Tableau 4.28a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	1,30	1,97	0,01	0,92	14,01	272,23
	D=0,2	1,47	2,05	0,04	1,03	14,96	307,10
	D=0,3	1,49	2,16	0,09	1,34	16,37	340,19
$\theta=-1$	D=0,1	3,43	4,91	0,19	1,12	6,04	50,87
	D=0,2	4,04	5,70	0,32	1,29	6,59	58,24
	D=0,3	4,19	6,25	0,39	1,35	6,95	67,63
$\theta=0$	D=0,1	22,81	35,19	1,40	1,63	1,43	3,49
	D=0,2	27,66	42,54	1,82	1,7	1,85	5,43
	D=0,3	23,99	38,49	1,79	1,84	2,55	9,48
$\theta=1$	D=0,1	72,05	78,56	5,37	4,20	0,91	0,96
	D=0,2	91,93	95,39	7,12	3,81	0,74	0,77
	D=0,3	94,32	97,72	6,72	3,26	0,78	0,78
$\theta=2$	D=0,1	85,98	89,27	12,06	8,96	0,86	0,65
	D=0,2	98,24	99,06	15,84	7,83	0,64	0,71
	D=0,3	99,84	99,94	15,66	6,01	0,65	0,90
2 paramètres							
$\theta=-2$	D=0,1	0,03	0,06	0,00	0,16	64,75	5609,93
	D=0,2	0,06	0,08	0,01	0,43	104,70	11968,39
	D=0,3	0,03	0,05	0,01	0,14	71,15	6891,49
$\theta=-1$	D=0,1	1,23	2,13	0,16	1,32	82,11	8576,75
	D=0,2	0,53	1,02	0,17	3,62	118,76	14384,07
	D=0,3	0,43	0,64	0,13	1,13	57,30	4437,51
$\theta=0$	D=0,1	24,61	28,85	6,19	17,69	4,89	30,36
	D=0,2	30,99	37,50	4,39	10,42	5,43	42,05
	D=0,3	22,23	29,65	2,27	6,09	15,04	451,60
$\theta=1$	D=0,1	43,24	52,85	44,96	115,69	3,55	15,06
	D=0,2	66,37	76,21	46,50	83,93	2,66	8,89
	D=0,3	77,81	85,91	30,86	48,52	2,81	12,66
$\theta=2$	D=0,1	57,31	57,78	175,06	472,88	3,66	15,87
	D=0,2	81,59	83,46	155,28	289,11	2,79	10,90
	D=0,3	92,23	94,74	109,91	157,68	2,22	7,42

Tableau 4.28b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZU^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	2,43	7,21	0,09	1,03	0,69	2,17
	D=0,2	2,75	8,27	0,19	1,03	0,47	0,74
	D=0,3	2,62	8,68	0,22	1,03	0,33	0,22
$\theta=-1$	D=0,1	17,79	22,09	1,30	3,43	8,02	126,32
	D=0,2	20,12	30,39	1,33	2,25	5,60	73,77
	D=0,3	14,95	28,83	1,10	1,44	3,03	37,41
$\theta=0$	D=0,1	27,05	30,08	7,06	19,81	5,97	57,02
	D=0,2	45,10	53,75	4,85	10,25	6,21	55,55
	D=0,3	46,65	62,65	2,78	3,72	10,14	201,22
$\theta=1$	D=0,1	28,99	35,72	26,29	92,77	7,19	71,88
	D=0,2	52,63	61,99	11,42	29,78	8,11	121,1
	D=0,3	70,88	80,97	5,03	8,75	9,76	143,08
$\theta=2$	D=0,1	33,57	44,71	94,52	449,66	18,87	790,91
	D=0,2	57,21	68,03	23,12	95,59	24,71	1144,84
	D=0,3	79,51	87,17	7,09	15,83	18,56	608,78

Les tableaux 4.29a et 4.29b présentent les résultats pour ZW . Pour commencer, nous observons que le pourcentage de détection de ZW est généralement plus élevé pour 80 items. Ensuite, le taux de détection est moins élevé que celui de I_z , I_z^* et ZU . Tout comme pour les indices de type vraisemblance, nous remarquons que lorsque le niveau d'inattention D augmente, le pourcentage de détection de l'inattention augmente lui aussi.

Aux valeurs $\theta = -2$ et $\theta = -1$, le pourcentage de patrons de réponses détectés est plus élevé lorsque le modèle à deux paramètres est utilisé. Aux valeurs $\theta = 0$ à $\theta = 2$, c'est le modèle de Rasch qui engendre généralement les pourcentages de

détection les plus élevés. Enfin, le modèle à trois paramètres présente les pourcentages de détection les moins élevés pour cet indice.

Lorsque θ augmente, nous remarquons que le pourcentage de patrons détectés tend à augmenter. De plus, le pourcentage de détection de l'inattention est plus élevé lorsque $\alpha = 0,05$ que lorsque $\alpha = 0,01$.

Tout comme pour l_z et l_z^* , il est possible de remarquer que plus le pourcentage de patrons de réponses détectés est élevé, plus la moyenne des scores de ZW est négative et faible et plus l'écart-type est positif et élevé. De plus, nous observons que les coefficients d'asymétrie et de kurtose sont relativement près des valeurs théoriques de la loi $N(0,1)$.

Tableau 4.29a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	0,49	2,49	0,16	0,95	-0,05	-0,11
	D=0,2	0,65	3,77	0,36	0,91	-0,06	0,01
	D=0,3	0,82	4,28	0,47	0,87	-0,02	0,06
$\theta=-1$	D=0,1	3,13	10,11	0,65	1,05	-0,05	-0,03
	D=0,2	9,89	25,15	1,23	1,06	-0,15	-0,07
	D=0,3	18,56	39,57	1,66	1,02	-0,17	0,03
$\theta=0$	D=0,1	11,87	27,75	1,28	1,11	-0,10	-0,01
	D=0,2	43,07	64,6	2,36	1,11	-0,17	0,00
	D=0,3	71,11	86,71	3,14	1,06	-0,20	0,04
$\theta=1$	D=0,1	30,35	51,18	1,96	1,17	-0,12	-0,02
	D=0,2	79,87	91,08	3,53	1,15	-0,15	0,07
	D=0,3	96,55	98,95	4,62	1,08	-0,20	0,08
$\theta=2$	D=0,1	53,13	72,27	2,66	1,22	-0,11	-0,06
	D=0,2	95,47	98,39	4,66	1,19	-0,19	-0,02
	D=0,3	99,84	99,95	6,02	1,09	-0,20	0,03
2 paramètres							
$\theta=-2$	D=0,1	0,54	3,15	0,17	0,97	-0,01	-0,13
	D=0,2	1,10	5,27	0,38	0,98	-0,01	-0,11
	D=0,3	1,26	6,09	0,49	0,96	-0,04	-0,13
$\theta=-1$	D=0,1	3,58	11,05	0,66	1,07	-0,05	-0,05
	D=0,2	12,2	27,59	1,28	1,11	-0,09	-0,09
	D=0,3	22,37	43,43	1,76	1,07	-0,14	-0,01
$\theta=0$	D=0,1	13,35	28,69	1,31	1,14	-0,07	-0,13
	D=0,2	44,32	65,25	2,40	1,16	-0,12	0,00
	D=0,3	70,54	85,57	3,15	1,12	-0,19	0,10
$\theta=1$	D=0,1	28,81	48,09	1,90	1,18	-0,05	-0,10
	D=0,2	77,72	89,71	3,50	1,21	-0,12	-0,05
	D=0,3	96,23	98,79	4,67	1,14	-0,18	0,01
$\theta=2$	D=0,1	51,43	70,67	2,61	1,22	-0,09	0,03
	D=0,2	94,28	97,95	4,53	1,20	-0,17	0,04
	D=0,3	99,67	99,94	5,87	1,12	-0,21	0,02

Tableau 4.29b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	0,49	2,59	0,06	0,99	-0,02	-0,06
	D=0,2	0,59	3,28	0,17	0,98	0,00	-0,12
	D=0,3	0,78	4,09	0,22	0,99	0,01	-0,09
$\theta=-1$	D=0,1	1,29	5,33	0,41	0,97	-0,02	0,02
	D=0,2	2,93	10,54	0,73	0,98	0,00	-0,04
	D=0,3	3,88	12,97	0,86	0,98	-0,03	-0,03
$\theta=0$	D=0,1	5,36	15,95	0,95	1,02	0,00	0,00
	D=0,2	19,87	40,61	1,70	1,03	-0,05	-0,11
	D=0,3	25,91	49,71	1,94	0,97	-0,09	0,02
$\theta=1$	D=0,1	19,25	38,74	1,65	1,09	0,05	0,10
	D=0,2	56,69	77,19	2,75	1,08	-0,08	0,01
	D=0,3	70,57	87,33	3,09	0,98	-0,11	-0,06
$\theta=2$	D=0,1	41,5	61,25	2,33	1,22	0,09	-0,07
	D=0,2	87,79	95,53	3,86	1,10	-0,10	0,04
	D=0,3	95,26	98,98	4,22	0,95	-0,12	0,01

Enfin, le tableau 4.30a et le tableau 4.30b présentent les résultats pour ZW^* .

Pour commencer, nous observons que les pourcentages de détection sont légèrement plus élevés lorsque les données analysées comportent 80 items. Ensuite, le pourcentage de détection de ZW^* est moins élevé que celui de l_z , l_z^* , ZW et ZU^* . Par contre, il est plus élevé que celui de ZU pour des données analysées selon le modèle à trois paramètres. Tout comme pour les indices de type vraisemblance et pour l'indice ZW , nous remarquons que lorsque le niveau d'inattention D augmente, le pourcentage de détection de l'inattention augmente lui aussi.

Aux valeurs $\theta = -2$ à $\theta = 0$, le pourcentage de patrons de réponses détectés est généralement plus élevé lorsque le modèle à deux paramètres est utilisé. Aux valeurs

$\theta = 0$ à $\theta = 2$, c'est le modèle Rasch qui présente les pourcentages de détection les plus élevés. Enfin, le modèle à trois paramètres présente les pourcentages les moins élevés pour cet indice.

Lorsque θ augmente, nous remarquons que le pourcentage de patrons détectés tend à augmenter. Encore une fois, le pourcentage de détection de la réponse au hasard est plus élevé lorsque $\alpha = 0,05$.

Tout comme pour l_z et l_z^* et ZW , nous remarquons que plus le pourcentage de patrons de réponses détectés est élevé, plus les moyennes et les écarts-types des scores de ZW sont positifs et élevés. De plus, nous observons que les coefficients d'asymétrie et de kurtose sont relativement près des valeurs théoriques de la loi $N(0,1)$.

Tableau 4.30a : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
Rasch							
$\theta=-2$	D=0,1	2,59	8,64	0,18	1,04	0,3	-0,02
	D=0,2	4,56	13,07	0,43	1,07	0,26	0,00
	D=0,3	6,35	17,4	0,6	1,1	0,26	0,01
$\theta=-1$	D=0,1	8,05	19,71	0,68	1,14	0,36	0,19
	D=0,2	21,42	39,29	1,35	1,24	0,23	-0,10
	D=0,3	35,43	56,6	1,89	1,27	0,19	-0,02
$\theta=0$	D=0,1	23,39	41,18	1,41	1,31	0,32	0,10
	D=0,2	59,65	77,11	2,77	1,48	0,21	-0,02
	D=0,3	83,85	92,82	3,86	1,55	0,15	-0,03
$\theta=1$	D=0,1	46,27	64,11	2,26	1,49	0,3	0,06
	D=0,2	89,16	95,34	4,46	1,76	0,23	0,03
	D=0,3	98,63	99,55	6,22	1,85	0,14	0,03
$\theta=2$	D=0,1	68,85	82,17	3,22	1,69	0,28	0,01
	D=0,2	97,96	99,33	6,32	2,05	0,17	-0,08
	D=0,3	99,95	99,99	8,83	2,16	0,12	-0,05
2 paramètres							
$\theta=-2$	D=0,1	2,95	9,37	0,19	1,06	0,36	0,08
	D=0,2	5,49	14,36	0,44	1,12	0,37	0,11
	D=0,3	7,03	17,69	0,58	1,14	0,31	-0,04
$\theta=-1$	D=0,1	8,8	20,39	0,7	1,17	0,33	0,09
	D=0,2	23,93	41,57	1,43	1,31	0,29	-0,01
	D=0,3	39,94	60,07	2,04	1,36	0,24	0,00
$\theta=0$	D=0,1	24,13	41,77	1,43	1,33	0,3	-0,03
	D=0,2	60,24	76,37	2,79	1,53	0,25	0,02
	D=0,3	82,39	91,61	3,82	1,6	0,17	0,06
$\theta=1$	D=0,1	43,14	61,46	2,16	1,48	0,32	0,03
	D=0,2	87,27	94,23	4,36	1,8	0,24	-0,01
	D=0,3	98,39	99,56	6,2	1,89	0,14	-0,08
$\theta=2$	D=0,1	67,15	81,4	3,15	1,67	0,29	0,09
	D=0,2	97,45	99,09	6,01	1,97	0,18	0,03
	D=0,3	99,93	99,98	8,33	2,08	0,1	-0,06

Tableau 4.30b : Pourcentage de détection de l'inattention et statistiques descriptives pour l'indice ZW^* (80 items)

		% p. d. ($\alpha=0,01$)	% p. d. ($\alpha=0,05$)	Moyenne	Écart-type	Asymétrie	Kurtose
		3 paramètres					
$\theta=-2$	D=0,1	1,69	6,54	0,07	1,01	0,17	-0,03
	D=0,2	2,03	7,97	0,19	1,01	0,18	-0,06
	D=0,3	2,55	8,79	0,24	1,02	0,2	-0,02
$\theta=-1$	D=0,1	4,13	12,94	0,45	1,05	0,18	0,07
	D=0,2	8,05	21,15	0,79	1,07	0,17	-0,03
	D=0,3	9,42	24,48	0,92	1,06	0,15	0,00
$\theta=0$	D=0,1	14,42	30,95	1,08	1,18	0,18	0,02
	D=0,2	36,42	58,08	1,91	1,21	0,1	-0,15
	D=0,3	43,33	67,29	2,15	1,13	0,06	-0,02
$\theta=1$	D=0,1	38,79	58,99	1,99	1,35	0,21	0,05
	D=0,2	75,73	88,97	3,28	1,35	0,02	-0,05
	D=0,3	84,73	94,31	3,58	1,21	0	-0,09
$\theta=2$	D=0,1	64,47	79,15	2,98	1,6	0,18	-0,09
	D=0,2	95,72	98,52	4,84	1,45	-0,07	-0,05
	D=0,3	98,71	99,72	5,1	1,25	-0,03	-0,04

4.3.3 Synthèse

Quatre constats généraux peuvent être soulevés suite à la présentation des résultats de la troisième étude. Premièrement, plus il y a un grand nombre d'items dans le test, plus les indices l_z , l_z^* , ZW et ZW^* engendrent des pourcentages de détection élevés. Ainsi, les évaluateurs ont tout intérêt à favoriser les tests de 80 items. Notons, par contre, que cette relation ne s'observe pas systématiquement pour les indices ZU et ZU^* .

Deuxièmement, nous avons observé dans plusieurs situations de simulation différentes que le modèle à deux paramètres présente les pourcentages de détection

les plus élevés. À l'opposé, c'est le modèle à trois paramètres qui présente généralement les pourcentages de détection les moins élevés.

Troisièmement, plus le niveau de réponse au hasard C est élevé, plus le pourcentage de détection des indices est aussi généralement élevé. Notons, de plus, que ce constat peut être observé pour l'inattention.

Quatrièmement, le tableau 4.31 présente un panorama des indices qui ont les plus hauts pourcentages de détection. Ainsi, nous observons que I_z^* est l'indice à retenir dans la majorité des situations. Par contre, si les données sont analysées à l'aide du modèle de Rasch, c'est l'indice ZW qui présente les plus hauts pourcentages de détection aux valeurs $\theta = 1$ et $\theta = 2$. Lorsque le modèle à deux paramètres est utilisé, il est préférable d'utiliser I_z^* plutôt que ZU , qui présentait parfois pourtant des pourcentages de détection plus élevés (par exemple, au tableau 4.21a). L'explication de cette décision est liée à la mauvaise adéquation des fonctions de densité de ZU à celle de la loi $N(0,1)$. En ce qui concerne les données selon le modèle à trois paramètres, il faut favoriser l'indice ZW^* aux valeurs $\theta = -2$ à $\theta = 0$ et I_z^* aux valeurs $\theta = 1$ et $\theta = 2$.

Tableau 4.31 : Indice présentant le plus haut pourcentage de détection de la réponse au hasard

	Rasch	2 paramètres	3 paramètres
$\theta = -2$	l_z^*	l_z^*	ZW^*
$\theta = -1$	l_z^*	l_z^*	ZW^*
$\theta = 0$	l_z^*	l_z^*	ZW^*
$\theta = 1$	ZW	l_z^*	l_z^*
$\theta = 2$	ZW	l_z^*	l_z^*

Le tableau 4.32 synthétise les indices à utiliser pour la détection de l'inattention. Dans ce cas-ci, l_z^* est l'indice qui présente les plus hauts pourcentages de détection.

Tableau 4.32 : Indice présentant le plus haut pourcentage de détection de l'inattention

	Rasch	2 paramètres	3 paramètres
$\theta = -2$	l_z^*	l_z^*	l_z^*
$\theta = -1$	l_z^*	l_z^*	l_z^*
$\theta = 0$	l_z^*	l_z^*	l_z^*
$\theta = 1$	l_z^*	l_z^*	l_z^*
$\theta = 2$	l_z^*	l_z^*	l_z^*

Enfin, les indices ZU et ZU^* ont présenté des pourcentages de détection élevés dans plusieurs situations de simulation. Par contre, l'analyse de leurs moyennes, de leurs écarts-types, de leurs coefficients d'asymétrie ainsi que de leurs coefficients de kurtose a montré qu'ils ne suivaient pas les quantiles de la loi $N(0,1)$. Ce résultat a aussi été observé lors de la première étude. Pour cette raison, nous ne recommandons pas leur utilisation pour la détection de patrons de réponses inappropriés.

CHAPITRE V DISCUSSION

L'objectif spécifique de ce projet doctoral consiste à utiliser la simulation informatique pour mettre à l'épreuve les indices standardisés l_z , ZU et ZW et leur version corrigée, soit les indices l_z^* , ZU^* et ZW^* . Dans la section portant sur les résultats, nous avons fait plusieurs constats qui seront maintenant discutés dans le chapitre qui suit.

Afin de rester cohérente avec la structure de cette thèse, la discussion se déclinera pour l'étude descriptive du score des six indices sélectionnés, l'étude de leurs erreurs de type I empiriques et théoriques ainsi que l'étude de leur puissance à détecter la réponse au hasard et l'inattention.

5.1 Étude 1: analyse descriptive des scores des indices

Dans cette section, nous discutons les analyses portant sur la comparaison des fonctions de densité des scores des indices et l'analyse des corrélations et des statistiques descriptives.

5.1.1 Comparaison des courbes de densité des scores des indices

L'analyse des fonctions de densité des indices a permis d'appuyer quelques résultats observés dans les écrits scientifiques. À l'instar de Molenaar et Hoijtink (1990, 1996), Nering (1995, 1997) et Reise (1995), nos résultats présentés aux figures 4.1 et 4.7 ont démontré que la fonction de densité des scores de I_z ne suivrait pas exactement les quantiles de la loi $N(0,1)$ lorsque le niveau d'habileté θ d'un étudiant est estimé.

En ce qui concerne l'indice I_z^* , Snijders (2001) écrivait: "looking more precisely at the simulation results (not elaborated in this paper) gives the impression that the distribution of person fit statistic has [...] a variance that is very close indeed to 1, and a mean that differs slightly from the asymptotic value of 0 (p. 340)". Ces résultats sont cohérents avec ceux présentés aux figures 4.2 et 4.8 de cette thèse, où nous avons démontré que la fonction de densité des scores de l'indice I_z^* suit approximativement la fonction de densité de la loi $N(0,1)$.

Il faut consulter l'étude de Van Krimpen-Stoop et Meijer (1999) pour obtenir les premiers résultats détaillés concernant le fait que la fonction de densité des scores de I_z^* est légèrement plus en accord avec la loi $N(0,1)$ que la fonction de densité des scores de l'indice I_z . Ainsi, pour un test de 20 items, ces auteurs ont montré que la moyenne et l'écart-type de I_z sont respectivement de 0,16 et 0,56 à la valeur $\theta = -2$, alors que la moyenne et l'écart-type de I_z^* sont respectivement de 0,09 et 0,98 à la valeur $\theta = -2$.

Ces auteurs obtiennent des résultats similaires avec 80 items. Par exemple, à la valeur $\theta = -2$, la moyenne et l'écart-type de l_z sont respectivement de 0,08 et 0,55 et respectivement de 0,05 et 0,99 pour l'indice l_z^* . Dans tous ces exemples, nous remarquons que les scores de l_z^* sont légèrement plus près de ceux de la loi $N(0,1)$ que les scores de l_z . Il est à noter que des résultats comparables ont été observés aux figures 4.1, 4.2, 4.7 et 4.8 de cette thèse.

En ce qui a trait à la fonction de densité des scores de l'indice ZW^* , nos résultats ont montré que l'intuition de Snijders (2001) pour l_z^* peut aussi être observée lorsqu'on corrige cet indice de type carré-moyen. Par contre, les figures 4.4 et 4.10 montrent que l'indice ZU^* présente une moins bonne adéquation à la fonction de densité de la loi $N(0,1)$.

Finalement, l'existence d'une légère différence entre la fonction de densité des scores de ZU et de ZW et la fonction de densité de la loi $N(0,1)$ ont aussi été observées aux figures présentées aux sections 4.1.1.1 et 4.1.1.2. Rappelons que Li et Olejnik (1997) avaient aussi obtenu un résultat semblable dans certaines sections de leur étude. En effet, pour des tests de 30 items, ces auteurs ont calculé une moyenne de 0,17 et un écart-type de 0,83 pour ZU . En ce qui concerne l'indice ZW , ils ont obtenu une moyenne de 0,02 et un écart-type de 0,861. Ces résultats sont près des valeurs de la moyenne et de l'écart-type de la loi $N(0,1)$.

5.1.2 Corrélations et statistiques descriptives

L'analyse des corrélations de Pearson entre les scores des indices a permis de faire émerger quatre grands constats. Premièrement, il existe des coefficients de corrélation positifs et élevés entre les indices de type vraisemblance l_z et l_z^* . Le même constat n'a pas été observé pour les coefficients de corrélation existant entre les indices de type carré-moyen ZU , ZU^* , ZW et ZW^* . Par exemple, si nous observons les résultats du tableau 4.6, un $r = 0,96$ a été calculé entre ZU et ZW^* au niveau $\theta = -2$, alors que nous avons obtenu un $r = 0,40$ au niveau $\theta = 2$.

Deuxièmement, peu importe le modèle de réponse à l'item utilisé, les analyses ont permis de montrer l'existence d'un coefficient de corrélation positif et élevé entre l_z et l_z^* et un coefficient de corrélation positif et élevé entre ZW et ZW^* . Nous devons toutefois soulever l'existence d'un coefficient de corrélation positif et faible entre ZU et ZU^* lorsque les données sont analysées à l'aide du modèle à deux paramètres ou du modèle à trois paramètres aux niveaux $\theta = 1$ et $\theta = 2$.

Troisièmement, plusieurs auteurs ont étudié les moyennes et les écarts-types des scores des indices. Ainsi, Van Krimpen-Stoop et Meijer (1999) ont utilisé le modèle à deux paramètres pour étudier l_z . Il est intéressant de constater que ces derniers ont obtenu plusieurs résultats qui vont dans le même sens que les nôtres. Par exemple, pour la valeur $\theta = 2$, ces auteurs ont obtenu une moyenne de 0,24 et un écart-type de 0,48 pour un test simulé de 20 items, alors que nous avons obtenu une moyenne de 0,25 et un écart-type de 0,45 pour un test simulé de 30 items.

De leur côté, Li et Olejnik (1997) ont utilisé le modèle de Rasch pour analyser les scores de l'indice I_z . Pour une longueur de test de 30 items, ces auteurs calculent une moyenne de -0,008 et un écart-type de 0,875. Il est à noter que ces résultats se rapprochent de ceux que nous avons obtenus à la valeur $\theta = 0$ (moyenne = 0,09, écart-type = 0,95). En ce qui concerne l'analyse de tests plus longs (60 items), ces auteurs obtiennent une moyenne de -0,13 et un écart-type égal à 0,853. Nos analyses sur des tests longs (80 items, selon notre devis méthodologique) ont permis d'obtenir des moyennes légèrement plus élevées (bornées entre 0,05 et 0,10) et des écarts-types similaires (bornés entre 0,53 et 0,98).

En ce qui a trait à l'indice I_z^* ($\theta = 2$), Van Krimpen-Stoop et Meijer (1999) ont obtenu une moyenne de 0,16 et un écart-type de 0,94, alors que nous avons obtenu une moyenne de -0,02 et un écart-type de 0,88. Ainsi, les valeurs mentionnées dans cette thèse sont légèrement inférieures à celles de ces auteurs.

À notre connaissance, Li et Olejnik (1997) sont les seuls à avoir étudié les moyennes et les écarts-types de ZU et de ZW . Pour des tests simulés de 30 items, ils ont calculé une moyenne de 0,17 et un écart-type de 0,83 pour l'indice ZU . De notre côté, et pour cet indice, les moyennes que nous avons calculées sont bornées entre -0,09 et 0,02 et les écarts-types entre 0,58 et 0,86. En ce qui concerne ZW , Li et Olejnik ont calculé une moyenne de 0,02 et un écart-type de 0,861. Dans le cadre de nos analyses, les moyennes de ZW sont bornées entre -0,12 et -0,05 et les écarts-types entre 0,39 et 0,95. Ainsi, nous observons que les moyennes que nous avons

calculées pour ces deux indices sont plus faibles que celles de ces auteurs. Par contre, les écarts-types semblent compatibles avec leurs résultats.

Quatrièmement, quelques commentaires doivent être soulevés au sujet des coefficients d'asymétrie et des coefficients de kurtose présentés dans cette thèse. D'abord, Snijders (2001) rapporte que la fonction de densité de l_z^* « has considerable skewness (p. 340) ». Par contre, ce chercheur ne donne pas plus de détails sur la taille de ces coefficients dans son article de 2001. Dans le cadre de nos analyses, il serait probablement excessif de déclarer que nous avons obtenu des coefficients d'asymétrie « considérables » puisqu'ils sont tous bornés entre -0,85 et -0,18, soit des valeurs qui ne sont pas excessivement éloignées de celle de la loi $N(0,1)$.

Les coefficients d'asymétrie de l_z et de l_z^* obtenus par Van Krimpen-Stoop et Meijer (1999) et ceux obtenus dans cette thèse sont tous négatifs. Pour l'indice l_z analysé sur un test de 80 items, les coefficients d'asymétrie calculés par ces auteurs sont bornés entre -0,36 et -0,31 alors que les nôtres sont bornés entre -0,26 et -0,28. Nos résultats sont donc assez similaires aux leurs. En ce qui concerne l_z^* , ces auteurs ont trouvé des coefficients d'asymétrie bornés entre -0,37 et -0,31 alors que les nôtres sont bornés entre -0,19 et -0,27. Encore une fois, l'écart entre les résultats présentés dans leur étude et les résultats présentés dans cette thèse est plutôt modeste.

Les coefficients de kurtose de l_z et l_z^* obtenus par Van Krimpen-Stoop et Meijer (1999) et ceux que nous avons calculés sont presque tous positifs. Ainsi, pour l'indice l_z , les coefficients obtenus par ces auteurs sont bornés entre -0,01 et 0,27

alors que les nôtres sont bornés entre 0,04 et 0,28. En ce qui concerne l'indice l_z^* , ces auteurs ont trouvé des coefficients bornés entre -0,02 et 0,17, alors que les nôtres sont plus faibles puisqu'ils sont bornés entre -0,05 et 0,03.

Les coefficients d'asymétrie et de kurtose calculés dans cette thèse sont légèrement différents de ceux de Li et Olejnik (1997). Pour l_z , ils ont calculé un coefficient d'asymétrie de 0,65 et un coefficient de kurtose de 0,04 (résultats pour un test court). Dans ce projet doctoral, les coefficients d'asymétrie sont bornés entre -0,46 et -0,85 et les coefficients de kurtose sont bornés entre 0,35 et 1,60. Dans la situation où ils analysaient les données d'un test long, ces auteurs ont présenté un coefficient d'asymétrie de 0,51 et un coefficient de kurtose de 0,03. Nos coefficients d'asymétrie sont tous bornés entre -0,26 et -0,28 et nos coefficients de kurtose entre 0,08 et 0,24 pour un test de 80 items.

En ce qui concerne les indices de type carré-moyen, Li et Olejnik (1997) obtiennent un coefficient d'asymétrie pour ZU (30 items) qui est égal à 5,09 et un coefficient de kurtose qui est égal à 75,17. Dans le cadre de nos analyses, nous avons obtenu des coefficients d'asymétrie bornés entre 0,29 et 2,89 et des coefficients de kurtose bornés entre -0,32 et 16,61. Nous observons donc, pour ces auteurs, des résultats beaucoup plus élevés que ceux obtenus dans cette thèse. Néanmoins, le résultat élevé du coefficient de kurtose est extrêmement important puisqu'il permet de valider la mauvaise adéquation des scores de l'indice ZU à la loi $N(0,1)$.

Pour l'indice ZW (30 items), Li et Olejnik (1997) ont calculé un coefficient d'asymétrie égal à 0,78 et un coefficient de kurtose égal à 1,36. Dans le cadre de nos

analyses, nous avons obtenu des coefficients d'asymétrie bornés entre 0,04 et 0,20 et des coefficients de kurtose bornés entre -0,14 et 1,54. Encore une fois, ces résultats sont relativement comparables.

5.2 Étude 2 : analyse des erreurs de type I empiriques et théoriques

Quelques comparaisons peuvent être effectuées avec d'autres études au sujet de l'étude des erreurs de type I empiriques et théoriques. Tout d'abord, Snijders (2001) a utilisé le modèle à deux paramètres pour montrer que les erreurs de type I étaient conformes aux quantiles de la loi $N(0,1)$ aux seuils $\alpha = 0,05$ et $\alpha = 0,10$ pour des tests de 8 et de 15 items. Ce résultat, tiré de l'article décrivant la méthode de correction investiguée dans cette thèse, met en exergue la faculté des indices corrigés à minimiser la différence existant entre les erreurs de type I empiriques et les erreurs de type I théoriques. Il est à noter que nous avons aussi observé des résultats similaires à ceux de Snijders (2001) pour l'indice l_z^* . Par exemple, si nous observons les figures 4.14 et 4.17 (qui analysent aussi des données à l'aide du modèle à deux paramètres), nous observons que l_z^* s'approche grandement de la droite diagonale d'égalité entre les erreurs de type I empiriques et théoriques.

Les résultats de Van Krimpen-Stoop et Meijer (1999) vont globalement dans le même sens que ceux observés dans cette thèse. À titre d'exemple, nous rapportons les résultats de ces derniers lorsque le paramètre de discrimination a_i est distribué selon la loi $N(1, 0,2)$. Pour des tests de 20 items, les erreurs de type I de l_z sont

bornées entre 0,04 et 0,07 et les erreurs de type I de l_z^* sont bornées entre 0,07 et 0,09 au seuil $\alpha = 0,05$. Ces résultats montrent que les erreurs de type I s'approchent généralement du seuil α .

De la Torre et Deng (2008, p. 167-168) ont produit une étude de comparaison de diverses approches au sein de laquelle se trouvait l_z^* . Certains éléments de leur étude portant sur les erreurs de type I de cet indice sont rapportés au tableau 5.1.

Tableau 5.1 : certains résultats tirés de l'étude des erreurs de type I de l_z^* (De la Torre et Deng, 2008, p. 167-168)

		$\alpha=0,01$	$\alpha=0,05$	$\alpha=0,10$
30	$\theta=-2$	0,013	0,052	0,097
	$\theta=0$	0,011	0,047	0,086
	$\theta=2$	0,013	0,050	0,066
50	$\theta=-2$	0,012	0,051	0,093
	$\theta=0$	0,012	0,048	0,094
	$\theta=2$	0,014	0,054	0,024

Ces résultats montrent globalement qu'il existe des similitudes entre les erreurs de type I empiriques et théoriques de l_z^* . Ce résultat général a aussi été observé aux figures 4.13 à 4.18, où l_z^* a généralement tendance à s'approcher de la droite diagonale d'égalité entre les erreurs de type I empiriques et théoriques.

Magis, Béland et Raïche (2014) ont aussi montré que l'indice ZW^* présentait des erreurs de type I qui s'approchent du seuil α . Par exemple, pour des ensembles de

données de 80 items analysés à l'aide du modèle de Rasch, ces auteurs ont calculé des erreurs de type I égales à 0,013 pour $\alpha = 0,01$, 0,052 pour $\alpha = 0,05$ et à 0,097 pour $\alpha = 0,10$. Les résultats que nous avons obtenus dans des conditions comparables (voir la figure 4.16 de cette thèse) nous montrent que ZW^* s'approche aussi considérablement de la droite diagonale d'égalité entre les erreurs de type I empiriques et théoriques.

Dans le cas de ZU^* , Magis, Béland et Raïche (2014) ont obtenu des résultats comparables à ceux de ZW^* puisque les erreurs de type I de cet indice sont égales à 0,021 pour $\alpha = 0,01$, 0,059 pour $\alpha = 0,05$ et 0,097 pour $\alpha = 0,10$. Dans ce cas-ci, nos résultats n'ont pas permis d'obtenir des erreurs de type I autant en accord avec leurs seuils α . En effet, la figure 4.16 montre qu'aux valeurs $\theta = -2$ et $\theta = 2$, ZU^* ne s'approche pas systématiquement de la droite d'égalité entre les erreurs de type I empiriques et théoriques. Une piste d'explication permettant de comprendre ce résultat est lié au fait que la stratégie de génération de données adoptée par Magis, Béland et Raïche (2014) est différente de celle utilisée dans cette thèse.

Enfin, Li et Olejnik (1997, p. 223) ont utilisé le modèle de Rasch pour investiguer les erreurs de type I des indices l_z , ZU et ZW . Dans tous les cas, ces auteurs ont montré que c'est l'indice ZU qui s'approche le plus de l'erreur de type I théorique. Par exemple, dans une situation où la simulation informatique est utilisée pour modéliser le comportement des étudiants qui sont gagnants lorsqu'ils répondent au hasard (ils obtiennent plus de bonnes réponses que de mauvaises réponses), ces auteurs ont remarqué que pour le seuil $\alpha = 0,05$, l'erreur de type I de l_z est 0,29, celle

de ZU de 0,39 et celle de ZW est 0,21. Pour des tests de 60 items et au seuil $\alpha = 0,05$, l'erreur de type I de l_z est 0,22, celle de ZU est 0,34 et celle de ZW est 0,17. Ces résultats sont consistants avec ceux présentés aux valeurs $\theta = -2$, $\theta = 1$ et $\theta = 2$ de la figure 4.13 et les résultats présentés aux valeurs $\theta = -2$, $\theta = -1$, $\theta = 1$ et $\theta = 2$ de la figure 4.16. Rappelons que dans toutes ces situations, ZU était l'indice standardisé qui s'approchait le plus de la droite d'égalité entre les erreurs de type I empiriques et théoriques.

5.3 Étude 3 : analyse de la puissance des indices

Tout d'abord, il est important de rappeler que la façon de simuler la réponse au hasard et l'inattention, plus réaliste ici, est différente de la méthode adoptée dans les écrits antérieurs. Dans ce contexte, trois commentaires généraux peuvent être formulés. Pour faciliter les comparaisons entre les études citées plus bas, nous avons remplacé la proportion de patrons de réponses détectée rapportée dans certaines études par un pourcentage.

Premièrement, nos analyses ont généralement montré que plus il y a d'items dans un test, plus les indices l_z , l_z^* , ZW et ZW^* ont tendance à détecter un grand nombre de patrons de réponses inappropriés. Ce résultat a déjà été rapporté par Karabatsos (2003, p. 288), puisqu'il a montré que le taux de détection des indices l_z , ZU et ZW était systématiquement plus élevé lorsque le test comportait 65 items plutôt que 17 items. En ce qui concerne ZU , nos résultats sont contraires à ceux de cet

auteur, puisque nous avons montré qu'augmenter le nombre d'items dans un test ne fait pas systématiquement augmenter le pourcentage de détection de cet indice. Une piste d'explication pour expliquer ce résultat est que la méthode de génération de données inappropriées est différente de la nôtre.

De leur côté, Li et Olejnik (1997, p. 224) ont montré que dans la situation où la simulation informatique est utilisée pour remplacer artificiellement de mauvaises réponses par de bonnes réponses, I_2 a un pourcentage de détection moyen de 40 % pour 30 items comparativement à 54 % pour 60 items. Même constat pour l'indice ZU qui présente un pourcentage de détection de 40 % pour 30 items et de 55 % pour 60 items, ainsi que pour ZW qui présente respectivement un pourcentage de détection de 33 % pour 30 items et de 43 % pour 60 items.

Deuxièmement, nos résultats tirés des analyses selon le modèle de Rasch ont montré que l'augmentation du niveau de réponse au hasard C et l'augmentation du niveau d'inattention D étaient généralement accompagnées d'une hausse du pourcentage de détection de ces patrons de réponses inappropriés. Karabatsos (2003), qui a aussi utilisé le modèle de Rasch pour faire ses analyses, a plutôt montré le contraire : le taux de détection des indices I_2 , ZU et ZW est moins élevé lorsqu'il y a 50 % de patrons de réponses inappropriés et il est plus élevé lorsqu'il y a seulement 5 % de patrons de réponses inappropriés. Ce résultat peut être expliqué par le fait que la méthode de génération de la réponse au hasard et de l'inattention de ce dernier est différente de celle utilisée dans cette thèse.

Dans le cadre des analyses produites à l'aide du modèle à deux paramètres et du modèle à trois paramètres, nos résultats n'ont pas pu montrer que l'augmentation de C et de D étaient systématiquement accompagnées d'une hausse du pourcentage de détection de ces patrons de réponses inappropriés. Ce résultat est en contradiction avec celui de Drasgow, Levine et McLaughlin (1987) qui ont montré que le taux de détection était plus élevé lorsqu'on utilise la simulation informatique pour transformer le pourcentage voulu de mauvaises réponses en bonnes réponses. Ainsi, I_z présente un taux de détection de 69 % lorsque 15 % des mauvaises réponses sont transformées en bonnes réponses contre un taux de détection de 96 % lorsque 30 % des mauvaises réponses sont transformées en bonnes réponses.

Troisièmement, il est pertinent de comparer nos pourcentages de détection avec ceux trouvés dans d'autres études. Van Krimpen-Stoop et Meijer (1999) ont utilisé le modèle à deux paramètres pour étudier la détection de la réponse au hasard. Pour des tests de 20 items, ils ont calculé un pourcentage de détection de 45 % pour I_z . Nous avons obtenu le même résultat pour la situation $\theta = 0$ et $C = 0,2$ (tableau 4.7a), où nous analysions plutôt 30 items. En ce qui concerne I_z^* , Van Krimpen-Stoop et Meijer (1999) ont obtenu un pourcentage de détection de 65 %. De notre côté, nous avons obtenu un résultat similaire à $\theta = 0$ et $C = 0,3$ (tableau 4.8a).

Karabatsos (2003, p. 286) a étudié le taux de détection de la réponse au hasard des indices I_z , ZU et ZW . Ses résultats ont démontré que ces trois indices présentent un taux de détection avoisinant 90 %. Nos résultats ont démontré que I_z présente un résultat qui va dans ce sens à la valeur $\theta = -2$ (30 items) et à la valeur $\theta = -2$ (80

items) et $\theta = -1$ (80 items). Par contre, nous n'avons pas obtenu des résultats aussi élevés que Karabatsos (2003) pour les indices ZU et ZW pour 30 items. Lorsque nous analysons 80 items, il est possible d'obtenir des résultats comparables à ceux de Karabatsos (2003) uniquement à la valeur $\theta = -2$ pour ZU et à la valeur $\theta = 1$ pour ZW . Dans toutes les autres situations, nos pourcentages de détection sont systématiquement plus faibles.

De leur côté, Magis, Béland et Raïche (2014) ont utilisé le modèle de Rasch pour analyser la détection de la réponse au hasard et de l'inattention. Observons premièrement leurs résultats pour la réponse au hasard. Pour l'indice ZU^* et 40 items, les taux de détection sont bornés entre 8 % et 48 %. Dans une situation comparable, nous avons obtenu des pourcentages de détection plus élevés puisqu'ils étaient bornés entre 2,05 % et 75,93 %. En ce qui concerne l'indice ZW^* , les taux de détection de ces auteurs sont bornés entre 4 % et 57 % (40 items). Encore une fois, nous avons obtenu des pourcentages de détection plus élevés puisqu'ils sont bornés entre 2,52 % et 91,55 % dans cette thèse.

Dans la situation où nous analysons la détection de l'inattention selon ZU^* (40 items), les taux de détection sont bornés entre 3 % et 13 % (Magis, Béland et Raïche, 2014). Nos résultats pour 30 items sont beaucoup plus élevés puisque les pourcentages de détection sont bornés entre 0,77 % et 75,29 %. En ce qui concerne l'indice ZW^* , les taux de détection de Magis, Béland et Raïche (2014) sont bornés entre 1 % et 15 %. Nos pourcentages de détection sont encore une fois plus élevés puisqu'ils sont bornés entre 2,56 % et 91,64 %.

St-Onge, Valois, Abdous et Germain (2011) ont trouvé des taux de détection de la réponse au hasard qui ressemblent à certains pourcentages obtenus dans cette thèse. À un niveau d'aberrance de 0,30 (qui équivaut à $C = 0,30$ selon notre devis méthodologique), le taux de détection de I_z est borné entre 25 % et 90 % lorsque les items analysés sont difficiles et entre 12 % et 70 % lorsque les items sont faciles. Dans le cadre de notre étude, les taux de détection de cet indice sont bornés entre 4,11 % et 99,97 %.

L'étude de la détection de l'inattention produite par St-Onge et ses collaborateurs présente aussi des résultats comparables à ceux obtenus dans cette thèse. À un niveau d'aberrance de 0,30, ($D = 0,30$ dans notre thèse) le taux de détection de I_z est borné entre 20 % et 90 % lorsque les items analysés sont difficiles et à 12 % et 65 % lorsque les items sont faciles. Dans le cadre de notre étude, les taux de détection de cet indice sont bornés entre 1,91 % et 99,99 %.

Enfin, il est très important de souligner le fait que les indices de détection n'ont pas le même comportement pour différentes valeurs de θ . Cela n'a malheureusement pas été observé par les autres auteurs pour l'étude de la réponse au hasard et de l'inattention puisqu'ils généraient les valeurs de θ selon une loi $N(0,1)$ ou une loi uniforme, ce qui ne permettait pas de faire les vérifications efficaces à différentes valeurs de θ .

5.4 Limites de cette étude

Au moins deux limites peuvent se dégager de cette étude. Premièrement, la méthode de génération de la réponse au hasard et de l'inattention dans cette thèse diffère de celle des autres études. En effet, nous n'avons pas reproduit la même méthode de génération de patrons de réponses inappropriés que celles utilisées par Levine et Drasgow (1982) ou par Karabatsos (2003) à cause de leurs caractères artificiels et difficiles à modéliser un contexte réel. C'est plutôt l'approche de Raïche, Magis, Blais et Brochu (2012) qui nous a semblé être la plus intéressante puisqu'elle permet de générer des données inappropriées plus près de la réalité. Ainsi, il est important de comprendre que nos résultats ne sont pas tous parfaitement comparables puisque c'est la première fois que cette modélisation est utilisée pour générer des patrons de réponses inappropriés.

Deuxièmement, l'approche retenue dans cette thèse est strictement descriptive. Cette stratégie nous semble justifiée, car nous avons l'obligation d'explorer les données en décrivant le comportement des indices dans différentes situations de simulation. Néanmoins, il faudra raffiner certains résultats pour mieux comprendre le comportement des indices à l'étude (cela fera l'objet d'une piste de recherche à envisager dans le futur).

CHAPITRE VI

CONCLUSION

Cette thèse vise à comparer six indices de détection de patrons de réponses inappropriés à l'aide d'une méthodologie quantitative par simulation informatique. La synthèse des résultats et les études à conduire dans le futur seront successivement présentées dans ce chapitre.

6.1 Synthèse des résultats

Nos analyses ont fait ressortir de nombreux éléments qui seront explicités dans les sections suivantes.

6.1.1 Étude 1 : analyse descriptive des scores des indices

L'objectif de cette première étude était d'examiner la fonction de densité de six indices de détection. Tout d'abord, nous avons conduit une série de simulations informatiques afin de comparer la fonction de densité de la loi $N(0,1)$ à la fonction de densité des scores des indices calculés selon trois modèles de réponse à l'item. Ensuite, nous avons calculé les coefficients de corrélation de Pearson entre les scores de ces indices et certaines statistiques descriptives.

Plusieurs constats découlent de nos analyses. Premièrement, nous remarquons que l'adéquation entre la fonction de densité des indices et la fonction de densité de la

loi $N(0,1)$ est plus importante lorsque l'épreuve comporte 80 items plutôt que 30 items. Il est donc préférable d'utiliser des tests plus longs avec les indices de détection de patrons de réponses inappropriés.

Deuxièmement, les indices l_z^* et ZW^* présentent une meilleure adéquation à la loi $N(0,1)$ pour toutes les valeurs de θ . En effet, l'analyse des moyennes, des écarts-types, des coefficients d'asymétrie et des coefficients de kurtose a montré que les scores de ces deux indices sont moins affectés par la longueur du test, la valeur de θ et le modèle de réponse à l'item utilisé. Nous recommandons donc de favoriser ces derniers pour la détection de patrons de réponses inappropriés. De leur côté, les indices l_z et ZW présentent une moins bonne adéquation à la loi $N(0,1)$ que leur version corrigée.

Troisièmement, nos analyses ont démontré que ZU et ZU^* présentent généralement une mauvaise adéquation à la loi $N(0,1)$. En nous basant sur les valeurs des coefficients d'asymétrie et de kurtose obtenus aux sections 4.1.1.1 et 4.1.1.2, nous suggérons d'ailleurs d'éviter d'utiliser ces deux indices.

Quatrièmement, aux valeurs de θ inférieures à zéro, il existe généralement une plus grande adéquation à la loi $N(0,1)$ lorsque les scores des indices ont été obtenus à l'aide du modèle à trois paramètres. Aux valeurs θ supérieures à zéro, il existe généralement une plus grande adéquation à la loi $N(0,1)$ lorsque les scores des indices ont été analysés à l'aide du modèle de Rasch.

6.1.2 Étude 2 : Analyse des erreurs de type I empiriques et théoriques

Dans cette deuxième étude, nous avons utilisé la simulation informatique pour comparer les erreurs de type I empiriques et les erreurs de type I théoriques des indices l_z , ZU , ZW , l_z^* , ZU^* et ZW^* . Dans ce cas-ci, trois grands constats émergent des analyses.

Premièrement, dans la majorité des situations de simulation étudiées, ce sont les indices corrigés l_z^* et ZW^* qui présentent le moins de différences entre leurs erreurs de type I empiriques et théoriques. Ces indices sont aussi ceux qui étaient le plus en accord avec la fonction de densité de la loi $N(0,1)$, dans la section 4.1. Ils présentent donc des propriétés favorables à leur utilisation pratique. Ensuite, l'indice ZU^* présente plus souvent des erreurs de type I empiriques qui s'approchent des erreurs de type I théoriques lorsque les données sont analysées à l'aide du modèle de Rasch. Par contre, la similitude entre ces deux types d'erreurs est moins importante pour les modèles à deux et à trois paramètres.

Deuxièmement, ZU est l'indice standardisé qui s'approche le plus souvent de la droite diagonale d'égalité entre les erreurs de type I empiriques et les erreurs de type I théoriques. Même si les coefficients d'asymétrie et de kurtose de cet indice sont très élevés (ses scores ont tendance à avoir des valeurs extrêmes), la moyenne et l'écart-type sont généralement en accord avec ceux de la $N(0,1)$. De leur côté, les indices l_z et ZW performant un peu moins bien que ZU . Enfin, les indices standardisés l_z , ZU et ZW ont tendance à sous-estimer les erreurs de type I empiriques.

Troisièmement, aux valeurs $\theta = -1$, $\theta = 0$ et $\theta = 1$, nous observons que lorsque les données sont analysées à l'aide du modèle de Rasch et du modèle à deux paramètres, les indices de détection présentent des erreurs empiriques et théoriques qui se ressemblent. Dans le cas du modèle à trois paramètres, nous remarquons que les indices de détection présentent moins de différences entre ces deux types d'erreurs aux niveaux $\theta = -2$ et $\theta = -1$ uniquement.

6.1.3 Étude 3 : Analyse de la puissance des indices

Dans la troisième étude, nous avons utilisé le modèle de Raïche, Magis, Blais et Brochu (2012) pour générer des patrons de réponses au hasard et des patrons de réponses inattentives. Ensuite, nous avons calculé le pourcentage de détection de chacun des indices à l'étude dans cette thèse.

Quatre éléments ressortent de ces analyses. Premièrement, plus il y a d'items dans l'épreuve d'évaluation, plus les indices I_z , I_z^* , ZW et ZW^* présentent des pourcentages de détection élevés. Tout comme nous l'avons observé à la première étude, les évaluateurs ont intérêt à favoriser les tests longs lorsqu'ils souhaitent maximiser la performance des indices de détection.

Deuxièmement, nous avons observé que dans plusieurs situations différentes, c'est le modèle à deux paramètres qui présente les pourcentages de détection les plus élevés. À l'opposé, c'est généralement le modèle à trois paramètres qui présente les pourcentages de détection les moins élevés.

Troisièmement, si les données sont analysées à l'aide du modèle de Rasch, nous remarquons que plus le niveau de réponse au hasard C ou d'inattention D est élevé, plus le pourcentage de détection des indices est aussi élevé. Par contre, il est important de souligner que cette relation ne s'observe pas systématiquement pour les modèles à deux et à trois paramètres.

Quatrièmement, ce sont globalement les indices I_z^* et ZW^* qui présentent les pourcentages de détection les plus élevés. Notons que ces mêmes indices présentaient la meilleure adéquation à la fonction de densité de la loi $N(0,1)$ à la première étude en plus d'être les indices qui ont obtenu le moins de différences entre les valeurs de leurs erreurs de type I empiriques et théoriques à la deuxième étude. Ensuite, les indices ZU et ZU^* présentent des pourcentages de détection très élevés dans certaines situations de simulation. Par exemple, l'indice ZU présente des pourcentages très élevés lorsqu'il est calculé à partir du modèle à deux paramètres. Néanmoins, l'analyse des moyennes, des écarts-types, des coefficients d'asymétrie et des coefficients de kurtose a montré que leurs scores ne suivaient pas les quantiles de la loi $N(0,1)$. Il est à noter que ce résultat a aussi été observé lors de la première étude. Tout comme nous l'avons soulevé aux sections précédentes, nous ne recommandons pas l'utilisation de ces indices pour l'analyse de patrons de réponses inappropriés.

6.2 Recommandations générales

Deux recommandations émergent des analyses produites dans cette thèse. Premièrement, il est plus approprié d'utiliser les indices de détection de patrons de réponses inappropriés sur des évaluations contenant 80 items plutôt que 30 items. En effet, la première étude a montré que la fonction de densité des scores des indices est plus en accord avec la loi $N(0,1)$ dans des tests longs.

Deuxièmement, si l'évaluateur doit sélectionner un seul indice de détection pour analyser les résultats à un test, nos analyses ont montré que le meilleur choix est définitivement I_z^* . Dans ce cas-ci, la troisième étude a montré que c'est cet indice qui présente les plus hauts pourcentages de détection de la réponse au hasard et de l'inattention.

6.3 Études à conduire dans le futur

De nombreuses études devront être conduites afin de mieux comprendre le comportement des indices de détection investigués dans cette thèse. Dans ce qui suit, nous avons regroupé les études à conduire en quatre grandes catégories.

Premièrement, il sera important de comparer le comportement des différentes solutions qui ont été proposées pour tenir compte du fait que le niveau d'habileté réel de l'étudiant n'est pas connu (Raïche et Blais, 2005; Raïche, Magis, Blais et Brochu, 2012; Snijders, 2001). À cet effet, comme Raïche et Blais (2005) l'ont déjà fait

antérieurement, les analyses pourraient être réalisées selon différentes valeurs du niveau d'habileté estimé plutôt que réel.

Deuxièmement, nous avons uniquement concentré nos analyses sur trois indices standardisés et trois indices corrigés. Il serait pertinent d'étendre la méthode de Snijders (2001) à d'autres indices tels que les indices de prudence (*extended caution indices*) de Tatsuoka et Linn (1983) et l'indice *Zeta* de Tatsuoka (1996).

Troisièmement, nous avons généré des matrices de données qui devaient respecter les postulats de base des modèles de réponse à l'item. Or, nous savons que dans la réalité, il est fréquent que ces postulats soient violés. Ainsi, il serait pertinent d'étudier la puissance des indices analysés en présence de multidimensionnalité de θ ou de dépendance locale entre les items.

Quatrièmement, les résultats d'épreuves d'évaluation et de questionnaires de recherche en éducation comportent souvent des non-réponses de la part des étudiants. Ce problème peut survenir lorsqu'un étudiant a omis de répondre à une section d'un test ou lorsqu'il a volontairement décidé de ne pas répondre à certaines questions. À l'exception de Zhang et Walker (2007), très peu d'études ont tenté d'analyser la puissance des indices en présence de réponses manquantes. Ainsi, il serait pertinent de vérifier le comportement des indices corrigés dans un tel contexte.

APPENDICE A - CODE R UTILISÉ POUR CALCULER LES INDICES

```

#-----
# CALCULS DES 6 INDICES

# lz, lz*, U* et W*: code de David Magis

### fonction de probabilité du modèle 3 paramètres avec dérivées
première et seconde

Pi<-function(th,it){
res1<-res2<-res3<-NULL
for (i in 1:nrow(it)) {
res1[i]<-it[i,3]+(1-it[i,3])*exp(it[i,1]*(th-
it[i,2]))/(1+exp(it[i,1]*(th-it[i,2])))
res2[i]<-(1-it[i,3])*it[i,1]*exp(it[i,1]*(th-
it[i,2]))/(1+exp(it[i,1]*(th-it[i,2])))^2
res3[i]<-(1-it[i,3])*it[i,1]^2*exp(it[i,1]*(th-it[i,2]))*(1-
exp(it[i,1]*(th-it[i,2]))) / (1+exp(it[i,1]*(th-it[i,2])))^3
}
RES<-list(Pi=res1,dPi=res2,d2Pi=res3)
return(RES)}

### fonctions ri et r0 (Snijders)

ri<-function(th,it) Pi(th,it)$dPi/(Pi(th,it)$Pi*(1-Pi(th,it)$Pi))

r0<-function(method="ML",th,it,mu=0,sigma=1){
res<-switch(method,
ML=0,
BM=(mu-th)/sigma^2,
WL=sum(ri(th,it)*Pi(th,it)$d2Pi)/(2*sum(ri(th,it)*Pi(th,it)$dPi))
return(res)}

### estimation de theta par ML, MAP ou WLE (appel 'irtoys' pour gain
de vitesse)

thetaEst<-function(x,it,method="ML",mu=0,sigma=1){
if (method=="ML" | method=="BM") res<-
mlebme(x,it,method=method,mu=mu,sigma=sigma)[,1]
else res<-wle(x,it)[,1]
return(res)}

### fonction wi (Snijders)

wi<-function(th,it,index="lz",N=NULL){
res<- switch(index,
lz=log(Pi(th,it)$Pi/(1-Pi(th,it)$Pi)),
U=(1-2*Pi(th,it)$Pi)/(nrow(it)*Pi(th,it)$Pi*(1-Pi(th,it)$Pi)),

```

```

W =(1-2*Pi(th,it)$Pi)/sum(Pi(th,it)$Pi*(1-Pi(th,it)$Pi))
return(res)}

### Wn(theta) (Snijders)

Wn<-function(x,th,it,index="lz",N=NULL) sum((x-
Pi(th,it)$Pi)*wi(th,it,index=index,N=N))

### sig2n (Snijders)

sig2n<-function(th,it,index="lz",N=NULL)
sum(wi(th,it,index=index,N=N)^2*Pi(th,it)$Pi*(1-
Pi(th,it)$Pi))/nrow(it)

### cn (Snijders)

cn<-function(th,it,index="lz",N=NULL)
sum(Pi(th,it)$dPi*wi(th,it,index=index,N=N))/sum(Pi(th,it)$dPi*ri(th
,it))

### wi tilde (Snijders)

wiTilde<-function(th,it,index="lz",N=NULL)
wi(th,it,index=index,N=N)-cn(th,it,index=index,N=N)*ri(th,it)

### tau2n (Snijders)

tau2n<-function(th,it,index="lz",N=NULL)
sum(wiTilde(th,it,index=index,N=N)^2*Pi(th,it)$Pi*(1-
Pi(th,it)$Pi))/nrow(it)

### deux indices côte à côte
## lz ou U ou W

bothIndex<-function(data,it,index="lz",method="ML",mu=0,sigma=1){
res0<-res1<-rep(NA,nrow(data))
N<-colSums(data)
theta<-thetaEst(x=data,it=it,method=method,mu=mu,sigma=sigma)
for(t in 1:nrow(data)){
th<-theta[t]
i<-t
EWn1<--
cn(th,it,index=index,N=N)*r0(method=method,th=th,it=it,mu=mu,sigma=s
igma)
VWn1<-nrow(it)*tau2n(th,it,index=index,N=N)
EWn0<-0
# EWn0<-EWn1<-0

```

```

VWn0<-nrow(it)*sig2n(th,it,index=index,N=N)
res0[t]<-switch(index,
lz=(Wn(data[i,],th,it,index=index,N=N)-EWn0)/sqrt(VWn0),
U=Wn(data[i,],th,it,index=index,N=N)+1,
W=Wn(data[i,],th,it,index=index,N=N)+1)
res1[t]<-(Wn(data[i,],th,it,index=index,N=N)-EWn1)/sqrt(VWn1)
}
RES<-cbind(theta,res0,res1)
colnames(RES)<-switch(index,
lz=c("theta","lz","lz*"),
U=c("theta","U","U*"),
W=c("theta","W","W*")
return(RES)}

#-----

# Code de Sébastien Béland (inspire de Karabatsos, 2003)

Uz <- function(x, theta, itIrt, verify=FALSE) {

  if (!(is.data.frame(x) | is.matrix(x) | is.vector(x)))
stop("Argument x must be a data frame, a vector or a matrix")
  if (is.matrix(x)) x <- data.frame(x)
  if (is.vector(x)) x<- data.frame(t(data.frame(x)))

  if (!(is.data.frame(itIrt) | is.matrix(itIrt))) stop("Argument
itIrt must be a data frame or a matrix")
  if (is.matrix(itIrt)) itIrt <- data.frame(itIrt)
  nItems <- dim(x)[2]
  if (dim(x)[1] != length(theta)) stop("Number of observations !=
length(theta).")
  if (nItems != dim(itIrt)[1]) stop("Number of items is not
compatible with the data matrix.")

  p <- NULL
  for (j in 1:dim(x)[1]) p <- rbind(p,p4p1(theta=theta[j],
a=itIrt[,1], b=itIrt[,2], c=itIrt[,3]) )

  variance <- p*(1-p)

  IRT.resid <- (x-p)
  IRT.sq.resid <- (x-p)^2
  IRT.sqstan.resid <- IRT.sq.resid / variance
  inv.IRT.var <- 1/variance
  U <- rowMeans(IRT.sqstan.resid)
  U.sd <- sqrt((rowSums(inv.IRT.var) - (4*nItems)) / (nItems^2))

  ZU <- ((U^(1/3))-1)*(3/(U.sd)) + (U.sd/3)

  result <- as.numeric(ZU)
  if (verify == TRUE) {print(data.frame(x)); print(data.frame(theta,
p=p, ZU))} # Ligne pour vérifier si les résultats sont plausibles

```

```

#class(result) <- c("personFitR", "numeric")
return(result)
}
#-----
# Code de Sébastien Béland (inspire de Karabatsos, 2003)

Wz <- function(x, theta, itIrt, verify=FALSE) {
  if (!(is.data.frame(x) | is.matrix(x) | is.vector(x)))
stop("Argument x must be a data frame, a vector or a matrix")
  if (is.matrix(x)) x <- data.frame(x)
  if (is.vector(x)) x<- data.frame(t(data.frame(x)))

  if (!(is.data.frame(itIrt) | is.matrix(itIrt))) stop("Argument
itIrt must be a data frame or a matrix")
  if (is.matrix(itIrt)) itIrt <- data.frame(itIrt)
  nItems <- dim(x)[2]
  if (dim(x)[1] != length(theta)) stop("Number of observations !=
length(theta).")
  if (nItems != dim(itIrt)[1]) stop("Number of items is not
compatible with the data matrix.")

  p      <- NULL
  for (j in 1:dim(x)[1]) p <- rbind(p, p4pl(theta=theta[j],
a=itIrt[,1], b=itIrt[,2], c=itIrt[,3]) )

  IRT.var <- p*(1-p)
  IRT.resid <- (x-p)
  IRT.sq.resid <- (x-p)^2
  IRT.sqstan.resid <- IRT.sq.resid / IRT.var
  W.sd <- (sqrt((rowSums(IRT.var)) - (4*(rowSums(IRT.var^2)))) /
(rowSums(IRT.var)))
  W <- rowSums(IRT.sq.resid) / rowSums(IRT.var)

  ZW <- ((W^(1/3))-1)*(3/(W.sd)) + (W.sd/3)

  result      <- as.numeric(ZW)
  if (verify == TRUE) {print(data.frame(x)); print(data.frame(theta,
p=p, ZW))} # Ligne pour vérifier si les résultats sont plausibles
  #class(result) <- c("personFitR", "numeric")
  return(result)
}

```

RÉFÉRENCES

- Al-Mahrazi, R. (2003). *Investigating a new modification of the residual-based person fit index and its relationship with other indices in dichotomous item response theory* (Thèse de doctorat non publiée). University of Iowa, Iowa City.
- Bade, B., Parkin, R. et Gonzalez, P. (2010). *Introduction à la microéconomie moderne*. Montréal, Québec : ERPI.
- Baker, F. B. et Kim, S.-H. (2004). *Item response theory: parameter estimation techniques* (2^e édition). New York, New Jersey: Dekker.
- Béland, S., Brassard, P. D. et Raïche, G. (2011). Étude du comportement de 14 indices de détection de patrons de réponses inappropriés paramétriques et non paramétriques à partir d'une analyse par corrélations canoniques. Dans G. Raïche, K. Paquette-Côté, et D. Magis (dir.), *Des mécanismes pour assurer la validité de l'interprétation de la mesure en éducation*. Ste-Foy, Québec : Presses de l'Université du Québec.
- Bénézech, M. (2007). Vérité et mensonge : l'évaluation de la crédibilité en psychiatrie et en pratique judiciaire. *Annales médico psychologiques*, 165, 351-364.
- Bertrand, R. et Blais, J.-G. (2004). *Modèle de mesure: l'apport de la Théorie de la réponse aux items*. Sainte-Foy, Québec : Presses de l'Université du Québec.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. Dans F. M. Lord et M.R. Novick (dir.), *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley.
- Blais, J.-G., Raïche, G. et Magis, D. (2009). La détection des patrons de réponses problématiques dans le contexte des tests informatisés, Dans J.-G. Blais (dir.), *Évaluation des apprentissages et technologies de l'information et de la communication*. Ste-Foy, Québec : Presses de l'Université Laval.
- Borsbooms, D, Mellenbergh, G. J. et van Heerden, J. (2003). The theoretical status of latent variables, *Psychological review*, 110, 203-219.
- Bourque, J., Blais, J.-G. et Larose, F. (2009). Avoir des difficultés scolaires importantes à l'école : quelles formules, quel avenir ? *Revue des sciences de l'éducation*, 35, 211-226.

- Brassard, P. D. (2011). *Identification des stratégies de sous-classement intentionnel aux tests déclassement en anglais, langue seconde, au collégial* (Mémoire de maîtrise non publié) Université du Québec à Montréal, Montréal.
- Camilli, G. (1994). Origin of the scaling constant $d = 1.7$ in item response theory. *Journal of educational and behavioral statistics*, 3, 293-295.
- Christensen Hughes, J. M. et McCabe, D. L. (2006). Academic misconduct within higher education in Canada. *The Canadian journal of higher education*, 36, 1-21.
- Cole, G. (1993). *A Monte Carlo Study of the effects of four factors on the effectiveness of the lz and ZEC14 appropriateness indices* (Thèse de doctorat non publiée). Université d'Ottawa, Ottawa.
- Conijn, J. M., Emons, W. H. M., Van Assen, M. A. L. M. et Sijtsma, K. (2011). On the usefulness of a multilevel logistic regression approach to person-fit analysis. *Multivariate behavioral research*, 46, 365-388.
- Cronbach, L. J. (1946). Response set and test validity. *Educational and psychological measurement*, 6, 475-494.
- Cureton, E. E. (1951). Validity. Dans Lindquist, E. F. (dir.): *Educational Measurement (1st ed)*. Washington, DC: American Council on Education.
- De la Torre, J. et Deng, W. (2008). Improving person fit assessment by correcting the ability estimate and its reference distribution. *Journal of educational measurement*, 45, 159-177.
- Dodeen, H. (2003). *The use of person-fit statistics to analyse placement tests*. Actes du 2003 Proceeding of the meeting of the American educational research association. Chicago, Illinois: AERA.
- Drasgow, F. (1982). Choice of test models for appropriateness measurement. *Applied psychological measurement*, 6, 297-308.
- Drasgow, F., Levine, M. V. et Williams, E. A. (1985). Appropriateness measurement with polytomous item response models and standardized indices. *British journal of mathematical and statistical psychology*, 38, 67-86.
- Drasgow, F., Levine, M. V. et McLaughlin, M. E. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. *Applied psychological measurement*, 11, 59-79.

- Edwards, A. L. (1970). *The measurement of personality traits by scales and inventory*. New York, New York: Rinehart and Winston.
- Emons, W. H. M. (2009). Detection and diagnosis of person misfit from patterns of summed polytomous item scores. *Applied psychological measurement*, 33, 599-619.
- Emons, W. H. M., Glas, C. A. W., Meijer, R. R. et Sijtsma, K. (2003). Person fit in order restricted latent class. *Applied psychological measurement*, 27, 459-478.
- Ferrando, P. J. (2004). Person reliability in personality measurement: an item response theory analysis. *Applied psychological measurement*, 9, 47-64.
- Fournier, P. (1992). *Pour un test incontestable, rapport de recherche sur les test de classement en anglais (langue seconde) au collégial*. Direction générale de l'enseignement collégial, Ministère de l'enseignement supérieur et de la science.
- Glas, C. A. W. et Meijer, R. R. (2003). A Bayesian approach to person-fit analysis in item response theory models. *Applied psychological measurement*, 26, 217-233.
- Hambleton, R. K. et Swaminathan, H. (1985). *Fundamentals of Item response theory*. Newbury Park, Californie: Sage.
- Harnisch, D. L. et Tatsuoka, K. K. (1983). A comparison of appropriateness indices based on item response theory. Dans R.K. Hambleton (dir.), *Applications of item response theory*. Vancouver, Colombie-Britannique: Kluwer academic publishers.
- Hendrawan, I. Glas, C. A. W. et Meijer, R. R. (2005). The effect of person misfit on classification decisions, *Applied psychological measurement*, 29, 26-44.
- Hopkins, K. D., Stanley, J. C. et Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation*. New York, New Jersey: Prentice Hall.
- Hunt, R. (2003). Let's hear it for Internet plagiarism. *Teaching and learning bridges*, 2, 2-5.
- Hutton, P. (2006), Understanding student cheating and what educators can do about it. *College teaching*, 54, 171-176.
- Johnson, E. M. (1998). *A taxonomy of person misfit on affective measures* (Thèse de doctorat non publiée). University of Denver, Denver.

- Kane, M. T. (1992). An argument-based approach to validation. *Psychological bulletin*, 112, 527-535.
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied measurement in education*, 16, 277-298.
- Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of applied measurement*, 2, 389-423.
- Kogut, J. (1987). *Detecting aberrant item response patterns in the Rasch model*. Rapport de recherche 87-3. Enschede, Hollande: University of Twente.
- Landry, R. (2004). La simulation par ordinateurs. In : Recherche sociale- De la problématique à la collecte de données. Dans B. Gauthier (dir.), *Recherche sociale : de la problématique à la collecte de données*. Ste-Foy, Québec : Presses de l'Université du Québec.
- Laurencelle, L. (2001). *Hasard, nombres aléatoires et méthode Monte Carlo*. Ste-Foy, Québec : Presses de l'Université du Québec.
- Laurier, M., Froio, L., Pearo, C. et Fournier, M. (1998). *L'élaboration d'un test provincial pour le classement des étudiants en anglais langue seconde, au collégial*. Québec, Québec : Direction générale de l'enseignement collégial-Gouvernement du Québec.
- Laveault, D. (2009). L'évaluation en classe: des politiques aux pratiques. *Mesure et évaluation en éducation*, 32, 1-22.
- Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal society of Edinburgh*, 61A, 273-287.
- Lawley, D. N. (1944). The factorial analysis of multiple item tests. *Proceedings of the Royal society of Edinburgh (Series A)*, 62, 74-82.
- Lemieux, C. (2009). *Monte carlo and quasi-monte carlo sampling*. New York, New York: Springer.
- Levine, M. V. et Drasgow, F. (1982). Appropriateness measurement: review, critique and validating studies. *British journal of mathematical and statistical psychology*, 35, 42-56.

- Levine, M. V. et Rubin, D. B. (1979). Measuring the appropriateness of multiple choice test scores. *Journal of educational statistics*, 4, 269–290.
- Levitt, S. et Jacob, J. B. (2003). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *Quarterly journal of economics*, 3, 843-77.
- Levitt, S. et Dubner, S. (2005). *Freakonomics: a rogue economist explores the hidden side of everything*, New York, New York: William Morrow.
- Li, M. F. et Olejnik, S. (1997). The power of Rasch person-fit statistics in detecting unusual response patterns. *Applied psychological measurement*, 21, 215-231.
- Lord, F. M. (1952). A theory of test scores. *Psychometric monographs*, 7.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York, New Jersey: Lawrence Erlbaum.
- Lord, F. M. et Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley Publishing Company Reading.
- Louis, R. (2004). *L'évaluation des apprentissages en classe- théorie et pratique*. Montréal, Québec : Beauchemin.
- Magis, D., Béland, S. et Raïche, G. (2014). Snijders's correction of infit and outfit indexes with estimated ability level: an analysis with the Rasch model. *Journal of applied measurement*. 15, 82-93
- Magis, D., Raïche, G. et Béland, S. (2011). A didactic presentation of Snijders' l_2^* index of person fit with emphasis on response model selection and ability estimation. *Journal of educational and behavioral statistics*. 37, 57-81.
- Meijer, R. R. (1994a). The number of Guttman errors as a simple and powerful person-Fit statistic. *Applied psychological measurement*, 18, 311-314.
- Meijer, R. R. (1994b). *Nonparametric and group-based person-fit statistics: a validity study and an empirical example*. Rapport de recherche 94-12. Twente, Pays-Bas: University of Twente.
- Meijer, R. R. (1996). The influence of the presence of deviant score patterns on the power of a person-fit statistic. *Applied psychological measurement*, 20, 141-154.

- Meijer, R. R. (1998). Consistency of test behaviour and individual difference in precision of prediction. *Journal of occupational and organizational psychology*, 71, 147-160.
- Meijer, R. R. et Sijtsma, K. (2001). Methodology review: evaluating person fit. *Applied psychological measurement*, 25, 107-135.
- Meijer, R. R., Muijtjens, A. M. M. et van der Vleuten, C. P. M. (1996). Nonparametric person-fit research: Some theoretical issues and an empirical example. *Applied measurement in education*, 9, 77-89.
- Messick, S. (1989). Validity. Dans R.L. Linn (Dir.), *Educational measurement*. New York, New York: Macmillan Publishing Company.
- Molenaar, I. W. et Hoijtink, H. (1990). The many null distributions of person fit indices. *Psychometrika*, 55, 75-106.
- Molenaar, I. W. et Hoijtink, H. (1996). Person-fit and the Rasch model, with an application to knowledge of logical quantors. *Applied measurement in education*, 9, 27-45.
- Nering, M. L. (1995). The distribution of person fit using true and estimated person parameters. *Applied psychological measurement*, 19, 121-129.
- Nering, M. L. (1997). The distribution of indexes of person-fit within the computerized adaptive testing environment. *Applied psychological measurement*, 21, 115-127.
- Nering, M. L. et Meijer, M. D. (1998). A comparison of the person response function and the L_z person fit statistic. *Applied psychological measurement*, 22, 53-69.
- Noonan, B. W., Boss, M. W. et Gessaroli, M. E. (1992). The effect of test length and IRT model on the distribution and stability of three appropriateness indices. *Applied psychological measurement*, 16, 345-352.
- Partchev, I. (2012). irtoys: Simple interface to the estimation and plotting of IRT models. R package version 0.1.6. <http://CRAN.R-project.org/package=irtoys>.
- Périard, M. (2002). Aider un autre élève peut-il faire baisser ma cote R? *Pédagogie collégiale*, 16, 26-28.
- Potter, A. (2011). *Je suis vrai- Tomber dans le piège de l'authenticité*. Montréal, Québec : les Éditions logiques.

- Raïche, G. (2002). *Le dépistage du sous-classement aux tests de classement en anglais, langue seconde, au collégial*. Gatineau, Québec: Collège de l'Outaouais.
- Raïche, G. et Blais, J.-G. (2005). *Characterization of the distribution of the Lz index of person fit according to the estimated proficiency level*. Actes du 70^e congrès annuel de la Psychometric Society, Tilburg, Pays-Bas: Psychometric Society.
- Raïche, G., Magis, D., et Blais, J.-G. (2008). *Multidimensional item response theory models integrating additional inattention, pseudo-guessing, and discrimination person parameters*. Actes du 73^e congrès annuel de la Psychometric Society, Durham, New Hampshire: Psychometric Society.
- Raïche, G., Magis, D., Blais, J.-G. et Brochu, P. (2012). Taking atypical response patterns into account. Dans Simon, M., Ercikan, K. et M. Rousseau (dir.), *Improving large scale assessment in education: theory, issues and practice*. New York, New York: Taylor and Francis.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago, Illinois: University of Chicago Press.
- Reise, S. P. (1990). A comparison of item- and person-fit methods of assessing model data fit in IRT. *Applied psychological measurement, 14*, 127-137.
- Reise, S. P. et Due, A. M. (1991). Test characteristics and their influence on the detection of aberrant response patterns. *Applied psychological measurement, 15*, 217-226.
- Reise, S. P. et Flannery, W. M. P. (1996). Assessing person-fit on measures of typical performance. *Applied measurement in education, 9*, 9-26.
- Richardson, M. W. (1936). The relationship between difficulty and the differential validity of a test. *Psychometrika, 1*, 33-49.
- Ro, S. (2001). *Characteristics of a likelihood-based person-fit index under the graded response model* (Thèse de doctorat non publiée). University of Minnesota, Minnesota.
- Rogers, H. J. et Hattie, J. A. (1987). A Monte Carlo investigation of several person and item fit statistics for item response models. *Applied psychological measurement, 11*, 47-57.

- Rudner, L. M. (1983). Individual assessment accuracy. *Journal of educational measurement, 20*, 207-219.
- Schmitt, N., Chan, D., Sacco, J. M., McFarland, L. A. et Jennings, D. (1999). Correlates of person fit and effect of person fit on test validity. *Applied psychological measurement, 23*, 41-53.
- Sijtsma, K. et Junker, B. W. (2006). Item response theory: past performance, present developments, and future expectations. *Behaviormetrika, 33*, 75-102.
- Sijtsma, K. et Meijer, R. R. (2001). The person responses function as a tool in person-fit research. *Psychometrika, 66*, 191-208.
- Smith, R. M. (1982). *Detecting measurement disturbances with the Rasch model* (Thèse de doctorat non publiée). University of Chicago, Chicago.
- Smith, R. M. (1985). A comparison of Rasch person analysis and robust estimators. *Educational and psychological measurement, 46*, 433-444.
- Snijders, T. A. B. (2001). Asymptotic null distribution of person fit statistics with estimated person parameter. *Psychometrika, 66*, 331-342.
- Stark, S., Chernyshenko, O. S. et Drasgow, F. (2005). Identifying and understanding the effects of unmotivated examinees on test dimensionality: application of optimal appropriateness measurement. *International journal of testing, 5*, 247-263.
- St-Onge, C. (2007). *La vraisemblance de patrons de réponses : étude de la précision des indices d'ajustement des scores individuels, de leurs points critiques et du taux optimal d'aberrance* (Thèse de doctorat non publiée). Université Laval, Ste-Foy.
- St-Onge, C., Valois, P., Abdous, B. et Germain, S. (2011). Accuracy of person-fit statistics: a monte carlo study of the influence of aberrance rates. *Applied psychological measurement, 35*, 419-432.
- Tatsuoka, K. et Linn, R. L. (1983). Indices for detecting unusual patterns: links between two general approaches and potential applications. *Applied psychological measurement, 7*, 81-96.
- Tatsuoka, K. et Tatsuoka, M. M. (1984). Detection of aberrant response patterns and their effect on dimensionality. *Journal of educational statistics, 7*, 215-231.
- Tatsuoka, K. (1996). Use of generalized person-fit indexes, zetas for statistical pattern classification. *Applied measurement in education, 9*, 65-76.

- Tendeiro, J. N. et Meijer, R. R. (2012). A CUSUM to detect person misfit : a discussion and some alternatives for existing procedures. *Applied psychological measurement, 36*, 420-442.
- Thurstone, L. L. (1927). A law of comparative judgement. *Psychological review, 34*, 273-286.
- Trabin, T. E. et Weiss, D. J. (1979). *The person response curve: fit of individuals to item characteristic curve models*. Rapport de recherche 79-9, Minnesota, Minnesota: University of Minnesota.
- Tuerlinckx, F., et De Boeck, P. (2005). Two interpretations of the discrimination parameter. *Psychometrika, 70*, 629-650.
- Viau, R. (1994). *La motivation en contexte scolaire*. Montréal, Québec : Éditions du Renouveau pédagogique.
- Viau, R. (2009). *La motivation à apprendre en milieu scolaire*. Montréal, Québec : Éditions du Renouveau Pédagogique.
- Wright, B. D. (1977). Solving measurement problems with the Rasch model, *Journal of educational measurement, 14*, 97-116.
- Wright, B. D. et Masters, G. N. (1982). *Rating scale analysis*. Chicago, Illinois: MESA Press.
- Wright, B. D. et Panchapakesan, N. A. (1969). A procedure for sample free item analysis. *Educational and psychological measurement, 29*, 23-48.
- Wright, B. D. et Stone, M. H. (1979). *Best test design*. Chicago, Illinois: MESA press.
- Zhang, B. et Walker, C. M. (2008). Impact of missing data on *person model fit* and *person trait estimation*. *Applied psychological measurement, 32*, 466-480.
- Zickar, M. J. et Dragow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied psychological measurement, 20*, 71-87.