



Université de Montréal

**Demi-vérités et vrais mensonges : Une analyse des processus liés à la  
dissimulation dans les questionnaires informatisés**

par

Sébastien Dubois

École de criminologie

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade Maître ès Sciences (M. Sc.)  
en criminologie

Avril 2014

© Sébastien Dubois, 2014

Université de Montréal  
Faculté des études supérieures

Ce mémoire intitulé :  
Demi-vérités et vrais mensonges : Une analyse des processus liés à la dissimulation  
dans les questionnaires informatisés

présenté par :  
Sébastien Dubois

a été évalué par un jury composé des personnes suivantes :

Marc Ouimet  
Président-rapporteur

Jean-Pierre Guay  
Directeur de recherche

Jean Proulx  
Membre du Jury



## Résumé

De nombreux chercheurs et cliniciens sont sceptiques quant à la validité des questionnaires autoadministrés, lorsqu'utilisés auprès d'une population carcérale (Gendreau, Irvine et Knight, 1973), surtout si celle-ci est composée de délinquants sexuels (Marshall et Hall, 1995). La sensibilité des sujets investigués jumelés à la transparence des questions expose l'évaluateur à la possibilité que le participant dissimule et modifie ses réponses (Tierney et McCabe, 2001). L'objectif de ce projet est de comprendre les processus impliqués dans la performance des participants à une évaluation autoadministrée. Les données de 282 délinquants sexuels ayant complétés le *Multidimensional Inventory of Development, Sex, and Agression* (MIDSA) ont été analysées afin de mieux comprendre l'interaction entre les échelles de désirabilité sociale, les temps de latence et les coefficients d'ajustement du modèle de Rasch. La convergence des analyses de temps de latence et des échelles de désirabilité sociale semble indiquer que certains participants dissimuleraient consciemment leurs réponses. Notamment, les participants détectés par les échelles de désirabilité sociales sembleraient répondre de manière de plus lente aux échelles d'évaluations, et certains d'entre eux offriraient des patrons de réponses incohérents à la prescription du modèle de Rasch. Les hypothèses permettant d'expliquer les potentiels mécanismes liés à la dissimulation seront discutées.

**Mots-clés : Désirabilité sociale, Dénier, Délinquants sexuels, Questionnaire autoadministré, Temps de réponse, Modèle de Rasch**

## **Abstract**

Many researchers and clinicians are skeptical towards the validity of self-reported assessment used within a forensic population (Gendreau, Irvine, & Knight, 1973), especially when used within a sexual offender population (Marshall & Hall, 1995). The sensitivity of the topics addressed as well as the transparency of the questions expose the evaluator to possible dissimulations or distortions in the subject responses (Tierney & McCabe, 2001). Responses of 282 sexual offenders who completed the Multidimensional Inventory of Development, Sex, and Aggression (MIDSA) were analyzed in order to understand the interaction between response time, social desirability scales and Rasch person-fit scores in order to account for response distortion. The convergence of latency times and social desirability scales seems to indicate that certain participants consciously manipulate their responses. Notably, participants detected by social desirability scales seem to respond more slowly to particular items. Other participants may answer in ways that are incoherent with the predicted responses according to the Rasch model. Hypotheses as to potential mechanisms underlying this process will be discussed.

**Key-words : Social desirability, Denial, Sexual offenders, Self-report, Response time, Rasch model**

## **Table des matières**

<b>RÉSUMÉ</b>	<b>I</b>
<b>ABSTRACT</b>	<b>II</b>
<b>TABLE DES MATIÈRES</b>	<b>III</b>
<b>LISTE DES TABLEAUX</b>	<b>V</b>
<b>LISTE DES FIGURES</b>	<b>VI</b>
<b>REMERCIEMENTS</b>	<b>VII</b>
<b>INTRODUCTION</b>	<b>1</b>
<b>1. RECENSION DES ÉCRITS</b>	<b>5</b>
<b>1.1 Pourquoi falsifier ses réponses?</b>	<b>5</b>
1.1.1 Caractéristiques contextuelles	5
1.1.2 Caractéristiques individuelles	8
<b>1.2 La désirabilité sociale</b>	<b>10</b>
1.2.1 L'opérationnalisation de la désirabilité sociale	10
1.2.2 La désirabilité sociale et la délinquance	13
1.2.3 Désirabilité sociale et falsification de données : même phénomène?	14
<b>1.3 La recherche sur la falsification de données</b>	<b>17</b>
1.3.1 Protocole avec instructions	17
1.3.2 Protocole Candidat-Titulaire	18
1.3.3 Protocole sans instruction	19
<b>1.4 Les techniques de détection</b>	<b>20</b>
1.4.1 Analyse de contenu	20
1.4.2 Analyse de processus	22
1.4.2.1 Le processus cognitif lié à l'élaboration d'une réponse	22
1.4.2.2 La congruence du schéma cognitif	24
1.4.3 Analyse de cohérence des patrons de réponses	27
1.4.3.1 La courbe caractéristique de l'item	28
1.4.3.2 Les coefficients d'ajustement et les patrons de réponses	32
<b>2. PROBLÉMATIQUE</b>	<b>35</b>
<b>2.1 Objectifs de l'étude</b>	<b>37</b>
2.1.1 Objectif général	37
2.1.2 Objectifs spécifiques	37
<b>3. MÉTHODOLOGIE</b>	<b>39</b>
<b>3.1 L'échantillon</b>	<b>39</b>
<b>3.2 Le MIDSA</b>	<b>40</b>
3.2.1 Les échelles d'évaluations ciblées	41

3.2.2 Les échelles de validité	44
3.2.3 Double standardisation des temps de latence	46
<b>3.4 Préparation des données pour l'analyse de Rasch</b>	<b>48</b>
<b>3.5 Procédure analytique</b>	<b>49</b>
3.5.1 Analyse de contenu	49
3.5.2 Analyse des temps de latence	51
3.5.3 Analyse de cohérence des réponses	51
<b>4. RÉSULTATS</b>	<b>52</b>
<b>4.1 Analyse de contenu</b>	<b>53</b>
4.1.1 Analyse du contenu rapporté	54
<b>4.2 Analyse des temps de latence</b>	<b>58</b>
<b>4.3 Analyse de cohérences des réponses</b>	<b>62</b>
<b>5. INTERPRÉTATION DES RÉSULTATS</b>	<b>67</b>
<b>5.1 Qu'est-ce que nous indique le contenu rapporté ?</b>	<b>67</b>
<b>5.2 L'utilité des temps de latence</b>	<b>68</b>
5.2.1 Gérer son image : une tâche qui requiert un effort	68
5.2.2 L'autoduperie : un phénomène inconscient	69
5.2.3 Le déni : le réel défi des chercheurs	70
<b>5.3 Différents domaines, différentes performances</b>	<b>71</b>
<b>5.4 En tentant d'être crédible, certains se mettent plutôt les pieds dans les plats</b>	<b>71</b>
<b>CONCLUSION</b>	<b>73</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES</b>	<b>76</b>
<b>ANNEXE</b>	<b>VIII</b>



## Liste des tableaux

Tableau I. Exemple d'ajustement entre le score observé et le score attendu à une échelle .....	33
Tableau II. Caractéristiques de l'échantillon .....	39
Tableau III. Statistiques descriptives des échelles d'évaluations.....	42
Tableau IV. Statistiques descriptives des échelles de validité .....	44
Tableau V. Analyse de corrélation entre le contenu rapporté aux échelles d'évaluation et celui aux échelles de validité .....	54
Tableau VI. Comparaison des scores moyens obtenus aux échelles d'évaluation selon le niveau de désirabilité sociale .....	56
Tableau VII. Statistiques descriptives des temps de latence aux quatre échelles d'évaluation.....	58
Tableau VIII. Comparaison du temps moyen pris pour compléter une échelle selon le niveau de désirabilité sociale .....	60
Tableau IX. Répartition des participants selon le niveau d'adéquation entre leurs patrons de réponses et la prescription du modèle .....	63
Tableau X. Test de comparaison de moyennes selon le niveau d'ajustement du patron de réponses du participant à l'échelle d'Impulsivité.....	64
Tableau XI. Test de comparaison de moyennes selon le niveau d'ajustement du patron de réponses du participant à l'échelle de Préoccupation sexuelle.....	64
Tableau XII. Test de comparaison de moyenne selon le niveau d'ajustement du patron de réponse du participant à l'échelle Charme superficiel.....	65
Tableau XIII. Test de comparaison de moyenne selon le niveau d'ajustement du patron de réponse du participant à l'échelle Hostilité envers les femmes.....	65

## Liste des figures

Figure I. Exemple de courbe caractéristique de l'item (CCI) .....	29
Figure III. Représentation graphique d'un écart entre le score attendu et le score observé à une question.....	30

## **Remerciements**

J'aimerais exprimer ma sincère gratitude envers plusieurs personnes qui m'ont permis de me rendre au bout de ce long périple qu'a été la maîtrise. J'aimerais tout d'abord remercier Jean-Pierre, mon directeur de recherche, pour ses précieux conseils, sa patience et son soutien. Ses idées brillantes ont su m'inspirer tout au long de mes études. Merci Jean-Pierre de m'avoir permis d'aller au-delà de mes limites. Merci aussi à Chantal et Catherine de m'avoir fait confiance pour la coordination de leurs projets de recherche. J'en serai reconnaissant toute ma vie.

Je voudrais remercier ma famille pour m'avoir insufflé ma dose quotidienne de courage et de confiance pendant ma rédaction. Avoir le support d'une famille que la mienne est un don du ciel. Merci à mes vieux amis Alex, Ben, Dan, Joey pour votre support moral et pour les blagues à faire pleurer de rire. J'ai les meilleurs amis du monde. Merci à mes collègues et amis de l'école Marc, Julie, Patricia, Camille et Manon avec qui j'ai eu beaucoup trop de plaisir à étudier et travailler. Un merci tout spécial à ma copine pour sa patience indéfectible, sa joie de vivre contagieuse et son amour. Je ne serais pas où je suis maintenant si ce n'était pas de toi.



## Introduction

Les questionnaires autoadministrés servent à la recherche en sciences humaines depuis près de cent ans. Depuis leurs premières utilisations au début du 20<sup>e</sup> siècle (Woodworth, 1917 dans Zicker et Gibby, 2006), un nombre incalculable de ces outils ont vu le jour. Aujourd'hui, les questionnaires autoadministrés, et spécialement ceux qui traitent de la personnalité, sont utilisés dans une multitude de contextes. Que ce soit à l'intérieur d'un processus de recrutement de personnel ou pour simplement compléter une évaluation psychologique, cette formule d'évaluation est encore très répandue. Toutefois, même à l'ère des premiers questionnaires autoadministrés, les chercheurs démontraient un certain scepticisme quant à la validité des réponses obtenues: « *Test-takers may be motivated to fake their response* » (Vernon, 1934). Déjà à cette époque, les chercheurs étaient conscients que l'individu évalué a toujours la liberté de rapporter la réponse qu'il désire, qu'elle soit représentative ou non de sa réalité. Il est donc peu surprenant que l'on s'intéresse à la validité des réponses depuis que l'on tente d'évaluer les gens (Zickar et Gibby, 2006). Avant de comprendre comment et pourquoi la falsification (traduction de *faking*) représente une réelle menace à la validité des évaluations autoadministrées, il est primordial de définir ce phénomène.

### *Définition du phénomène*

Une des plus grandes préoccupations des chercheurs s'intéressant à la falsification des données a été de définir et d'opérationnaliser ce phénomène (MacCann, Ziegler et Roberts, 2011). Même à ce jour, il ne semble pas y avoir de consensus établi sur la définition du phénomène. Ce dernier a été nommé de différentes façons : biais de réponse (*response bias*), style de réponse (*response set*), déformation des réponses (*response distortion*), biais de désirabilité sociale (*social desirability bias*), simulation (*faking-bad*) et dissimulation (*faking-good*) (Ziegler, MacCann et Buehner, 2011). Les termes employés à travers le temps ont été utilisés de manière interchangeable, ce qui en complique considérablement la compréhension. Quoique les auteurs semblent parler

du même problème, certaines divergences dans leurs conceptualisations justifient l'emploi de termes différents. Bien que ces mots constituent des synonymes, leurs interprétations, leurs conceptualisations et les processus sous-jacents peuvent varier (Ellingson, 2011). Il existe toutefois une caractéristique unissant ces synonymes : chacun soutient que la réponse offerte n'est pas une représentation réaliste de la personne évaluée (Paulhus, 2002). La falsification des réponses signifie que la personne évaluée offre une représentation de soi objectivement déformée, marquée par une série de réponses falsifiées.

Pour de ce mémoire, les expressions *falsification des réponses* et *déformations des réponses* seront employées de manière interchangeable, puisque tous deux impliquent une manipulation des réponses par le participant. Cette déformation des réponses aura comme conséquence d'engendrer ce que sera nommé *un patron de réponses erronées*. Inspirée des théories de la mesure, l'expression « patron de réponses erronées » signifie qu'un participant obtient un score à une évaluation qui ne représente pas adéquatement son niveau d'habiletés (Bertrand & Blais, 2004, p.263).

Il existe plusieurs raisons qui entraînent un patron de réponses erronées telles que le manque de concentration, les réponses aléatoires ou la tricherie (Schmitt, Chan, Sacco, Macfarland et Jennings, 1999). Bien qu'il soit vrai que chacune de ces manifestations peut entraîner un patron de réponses erronées, le phénomène d'intérêt du présent mémoire est la déformation des réponses motivée par la volonté de vouloir tromper autrui (Holden et Book, 2011). Nous laisserons donc de côté l'analyse des patrons de réponses aléatoires et des patrons de réponses improbables.

L'action de répondre à un questionnaire est en soi un comportement qui requiert un effort motivé (Leary et Kowalski, 1990; Kuncel, Borneman et Kiger, 2011) et qui sera influencé par différentes caractéristiques contextuelles et individuelles. Puisque certains participants répondent intentionnellement de manière erronée afin de tromper autrui, il est primordial d'être en mesure de détecter et de comprendre les processus sous-jacents à leur prestation.

### *L'intérêt criminologique*

L'importance liée à la validité des réponses est proportionnelle à l'enjeu lié à l'évaluation. En d'autres termes, la validité des réponses n'a de réelle importance que si l'enjeu de l'évaluation est crucial. Or, nulle part ailleurs ne retrouve-t-on une aussi grande importance d'utiliser un processus d'évaluation robuste et valide qu'auprès d'une population délinquante. Puisque certaines de ces évaluations portent sur le progrès en traitement, sur l'acquisition d'habiletés ou sur l'évaluation du risque, ces dernières peuvent avoir ultimement un impact sur la sécurité publique (Tan et Grace, 2008). Comme il existe un réel enjeu dans les conclusions qui seront tirées de ce type d'évaluation, il est crucial de déterminer si l'information que rapporte le délinquant est belle et bien authentique. De plus, les contrevenants évalués peuvent très bien avoir conscience des enjeux liés à leurs évaluations. Ces derniers ont donc avantage à minimiser leurs attitudes antisociales, leurs comportements déviants ou leurs délits. C'est pourquoi, pour de nombreux chercheurs, il est évident que les délinquants modulent leurs réponses afin de mieux paraître (Gendreau, Irvine et Knight, 1973; Hare, 1985; Holden, Kroner, Fekken et Popham, 1992; Mills et Kroner, 2005).

La validité des réponses obtenues aux mesures autoadministrées auprès d'une population contrevenante éveille le scepticisme chez beaucoup de chercheurs, notamment lorsqu'il est question des délinquants sexuels (Marshall et Hall, 1995; Tierney et McCabe, 2001; Mcgrath, Cann et Konopsky, 1995). Le puissant stigma social lié à ce type de délit et à la déviance sexuelle serait un élément non négligeable dans l'évaluation de la validité de leurs réponses (Tierney et McCabe, 2001). Dans notre société, le simple fait de parler de la sexualité peut créer un certain inconfort, mais de tous les types de comportements illégaux confondus, ceux qui sont liés à la sexualité constituent de loin les plus inacceptables (Knight, 1999).

### *Pourquoi éviter le terme « mensonge » ?*

La ligne est très mince entre la dissimulation, la déformation de la réalité, la manipulation et ce que l'on nomme couramment le mensonge. Le mensonge remplit

plusieurs fonctions, il permet de se valoriser, de créer une bonne impression, de dissimuler quelque chose et souvent, il permet d'éviter des conséquences négatives (St-Yves et Patenaude, 2014). Il est vrai que ces motivations sont semblables à celles reconnues chez les participants qui falsifient leurs réponses, toutefois il est préférable d'éviter d'utiliser ce terme. La raison est simple, le mensonge n'existe seulement que lorsque l'on connaît la vérité. En avançant vouloir détecter le mensonge, cela signifie que la vérité est connue. Par déduction, un mensonge n'existe pas sans l'autre. Or, il n'existe aucun moyen ici, à travers l'analyse de performances autoadministrées, d'avoir la certitude qu'un participant ment ou dit la vérité. La prudence dans l'utilisation des termes désignant les performances des participants sera donc adoptée. À travers la démarche méthodologique, différents indices renseignant sur la potentielle falsification de données seront recueillis.

L'objectif du présent mémoire est de comprendre les processus impliqués dans la performance aux questionnaires autoadministrés. Plus précisément, il sera question d'analyser différents indices permettant de déceler si un participant déforme ses réponses. Il est possible de croire que la prestation d'un participant à un test peut renseigner au-delà du contenu que celui-ci rapporte. Le fait de répondre à un questionnaire est en soi un comportement qui comporte divers marqueurs et caractéristiques mesurables, telles que la vitesse des réponses et la logique entre les réponses données. Nous verrons comment il est possible de mieux comprendre et appréhender la falsification des données par les participants.

Afin de comprendre les processus sous-jacents à la performance aux questionnaires autoadministrés, il est important de comprendre quelles sont les caractéristiques contextuelles qui augmentent la propension d'un individu à vouloir dissimuler et falsifier ses réponses. Par la suite, les caractéristiques individuelles qui augmentent la probabilité qu'un individu falsifie ses réponses seront exposées. Ces différentes influences sur la performance des individus aux questionnaires autoadministrés seront traitées dans la prochaine section du mémoire



## **1. Recension des écrits**

### **1.1 Pourquoi falsifier ses réponses?**

Déformer et falsifier ses réponses requiert un effort motivé, ayant pour but de tromper autrui (Holden et Book, 2011). Or, la motivation à offrir une réponse falsifiée nécessite la combinaison de caractéristiques individuelles et de caractéristiques contextuelles (Ellingson, 2011). Le contexte dans lequel se déroule l'évaluation jouera un rôle déterminant dans la propension des participants à offrir une performance authentique ou non. Avant de considérer les caractéristiques individuelles des participants, il est important de comprendre dans quels contextes la falsification des réponses est plus probable de se produire.

#### **1.1.1 Caractéristiques contextuelles**

L'utilisation des tests de personnalité est monnaie courante dans plusieurs domaines connexes aux sciences humaines. Bien que leur utilité puisse être polyvalente, ce type d'évaluation est généralement administré dans l'optique de renseigner sur les caractéristiques personnelles des individus qui sont évalués. En faisant l'inventaire des différentes caractéristiques des participants, l'évaluateur se munit d'une représentation réaliste de ce dernier (Ziegler, MacCann et Roberts, 2011). En sciences de la médecine ainsi qu'en psychologie par exemple, l'évaluation peut être utilisée dans le but d'apposer un diagnostic, alors que dans un milieu de travail, l'évaluation peut être administrée dans l'optique de sélectionner les meilleurs candidats pour une poste vacant (Ziegler et al., 2011). Peu importe l'objectif de l'évaluation, le participant sera toujours conscient des buts visés par celle-ci, et ce facteur pourra venir influencer sa performance. En effet, étant conscient du but de l'évaluation, si le participant a un intérêt dans les résultats de celle-ci, il peut adéquatement modifier ses réponses afin d'obtenir les bénéfices qui y sont associés (Snell, Sydell et Lueke, 1999). Dépendamment si le participant doit déformer à la hausse ou à la baisse ses résultats, les chercheurs parleront de deux phénomènes distincts : la simulation et la dissimulation. Afin d'illustrer la différence entre la simulation (traduction de *faking-*

*bad*) et la dissimulation (traduction de *faking-good*), un exemple de ces deux phénomènes distincts sera présenté.

La simulation fait référence à l'exagération ou la fabrication de symptômes psychopathologiques (Wasyliw, Grossman, Haywood et Cavanaugh, 1988). Elle risque de survenir lorsque le participant a un incitatif à se présenter comme étant malade, par exemple pour un délinquant voulant influencer le cours de son processus judiciaire (Wasyliw et al., 1988). Dans l'éventualité où ce dernier ne souffre d'aucun symptôme mais qu'il en rapporte quand même, il sera considéré que ce participant simule (traduction de *faking-bad* et de *malingering*) (Egeland & Langfjaeran, 2007). Dans ces conditions, le participant offre la réponse qu'il croit être la bonne, sans que celle-ci représente sa réalité (Ziegler et al., 2011). La simulation de détresse psychologique au sein d'une population carcérale a fait l'objet de nombreuses études (Wasyliw et al., 1988; Bagby, Rogers & Buis, 1994; Stefan, Morgan, Lee, Sellbom, 2010) puisqu'il y aurait jusqu'à 32 % des délinquants nouvellement admis à l'intérieur d'institutions carcérales psychiatriques qui simulerait une problématique mentale (Pollock, Quigley, Worley & Bashford, 1997). La simulation représente une forme de falsification de données où le participant « gonfle » les réponses lors de son évaluation. Bien qu'il s'agisse d'un type de falsification de données, celle-ci diffère énormément du phénomène d'intérêt du présent mémoire, soit la dissimulation.

La dissimulation représente le processus opposé à la simulation, où le participant tente d'amoindrir ses réponses afin de paraître le plus « normal » possible. Un lot important d'études se sont intéressées à comprendre comment et pourquoi certains délinquants dissimulent leurs réponses lors d'une évaluation psychologique (Gudjonsson, 1990; Tierney & McCabe, 2001; Curwen, 2003; Mills & Kroner, 2005; Gannon & Polachek, 2005). Pour qu'un participant veuille dissimuler ses réponses, l'influence contextuelle aura aussi un impact crucial. Tout comme pour la simulation, le participant devra évaluer s'il existe un quelconque intérêt pour lui dans les retombées de l'évaluation. Prenons l'exemple d'un individu incarcéré à qui on administre un questionnaire évaluant les attitudes antisociales. Si ce dernier a un intérêt à ne pas être dépeint

comme le plus grand délinquant du pays (par exemple pour bien paraître aux yeux de l'évaluateur), il dissimulera ses traits antisociaux. Encore une fois ici, le participant offre la réponse qu'il croit être la bonne, sans que celle-ci représente sa réalité (Ziegler et al., 2011). D'autre part, plus grand sera l'enjeu lié aux résultats de l'évaluation (traduction de *high-stakes*) plus forte sera la propension de l'individu à offrir une performance non authentique (Eggestad, 2011).

En bref, les caractéristiques contextuelles influenceront le choix de réponses du participant puisque ce dernier sera conscient de l'objectif de l'évaluation. De plus, si ce dernier a un intérêt dans les retombés de l'évaluation, il sera motivé à modifier en conséquence ses réponses. Toutefois, le participant s'engagera dans une déformation de ses réponses seulement s'il considère que la « bonne » réponse à donner (la réponse qui augmente ses probabilités d'atteindre son but) diffère de sa « vraie » réponse (la réponse représentative de sa réalité objective) (McFarland et Ryan, 2000). Notamment, la transparence des questions peut rendre la tâche plus facile pour les participants de cerner la « bonne » réponse attendue (McFarland et Ryan, 2000), et ceci serait particulièrement vrai pour une population présentant une problématique en délinquance sexuelle (Tierney et McCabe, 2001). En effet, la plupart des échelles employées afin d'évaluer les différentes problématiques sexuelles sont composées de questions très transparentes (Tierney et McCabe, 2001). Une question transparente est une question où le sujet investigué est très clair. Ainsi, lorsqu'il est facile pour le participant de cerner le sujet investigué par la question, il peut adéquatement s'ajuster et trouver la « bonne » réponse à donner. Par exemple, lorsque confronté à des questions traitant des fantaisies sexuelles pédophiles, il devient évident pour le participant que la « bonne » réponse à donner dans cette situation est la réponse socialement acceptable. Cette réalité, jumelée au fait que les participants condamnés pour un crime sexuel sont conscients du stigma social lié à leurs problématiques, rend la tâche particulièrement difficile aux évaluateurs tentant de minimiser l'impact de la désirabilité sociale sur les résultats des évaluations (Tan et Grace, 2008).

Bien qu'il existe certains contextes d'évaluation qui augmentent la probabilité qu'un

participant dissimule ses réponses, certaines caractéristiques individuelles augmentent la propension à la falsification des réponses. La perception du participant quant à l'acceptabilité de falsifier ses réponses, son évaluation de sa capacité à déformer efficacement ses réponses sans se faire prendre et certains traits de personnalité jouent tous un rôle déterminant dans le choix que fera le participant à s'engager ou non dans la déformation de ses réponses.

### **1.1.2 Caractéristiques individuelles**

En contexte d'évaluation, certaines caractéristiques personnelles augmentent la probabilité qu'un individu falsifie ses réponses. Tout d'abord, la perception quant à l'acceptabilité de falsifier ses réponses et l'éthique personnelle des participants est à prendre en considération (Snell et al., 1999). Si un individu croit qu'offrir des réponses non authentiques est une pratique acceptable et courante, sa propension à offrir des réponses falsifiées augmentera (Snell et al., 1999). Évidemment, si ce dernier croit que l'ensemble des participants modifie ses réponses, il sera davantage tenté de faire de même. De plus, une évaluation positive de ses capacités à falsifier ses réponses sans se faire détecter augmenterait aussi la probabilité qu'un individu s'engage dans la falsification de ses réponses (Mueller-Hanson, Heggstad et Thornton 2006). Donc, percevoir la falsification de données comme étant acceptable et comme étant une pratique répandue, jumelée à une évaluation positive de ses capacités à modifier ses réponses sans se faire détecter augmenterait la probabilité qu'un individu falsifie ses réponses.

Il existerait certains traits de personnalité et attitudes qui augmenteraient la probabilité qu'un individu perçoive la falsification comme étant un comportement acceptable (Griffith et al., 2006; Mueller-Hanson et al., 2006). Les individus qui adoptent des comportements de manipulation seraient plus enclins à déformer leurs réponses. D'ailleurs, on rapporte que les individus ayant un score élevé aux échelles de machiavélisme s'engageraient davantage à produire des patrons de réponses erronées

(Mueller-Hanson et al., 2006). De plus, la délinquance organisationnelle, ou la tendance à ne pas prendre en considération les règles, a aussi été mise en relation avec la probabilité de falsifier ses réponses (Snell et al., 1999).

Il est important de noter que la propension à la manipulation, au mensonge, à l'escroquerie et le non-respect des normes sont toutes des caractéristiques qui se retrouvent souvent auprès d'individus souffrant d'un trouble de la personnalité antisociale (DSM-IV-R, 2000), de même que chez les psychopathes (Hare, 1991). C'est d'ailleurs pourquoi l'on reconnaît aux délinquants une meilleure capacité à manipuler et tromper que chez les individus de la population générale (Schretlen & Arkowitz, 1990). Ce qui explique pourquoi plusieurs chercheurs sont sceptiques quant à la validité des réponses obtenues aux mesures autoadministrées, lorsqu'employées auprès d'une population délinquante (Gendreau, Irvine & Knight, 1973; Hare, 1985; Holden, Kroner, Fekken, & Popham, 1992).

Les cliniciens et chercheurs qui œuvrent auprès d'une population délinquante et qui désirent employer une mesure autoadministrée sont confrontés à une double problématique. Tout d'abord, ces derniers travaillent auprès d'une population que l'on reconnaît avoir plusieurs caractéristiques qui augmentent leur propension à la manipulation, la déformation et la falsification de leurs réponses. Deuxièmement, le contexte d'évaluation peut présenter plusieurs incitatifs pour les délinquants à vouloir bien se représenter (Tan et Grace, 2008). La dissimulation aux mesures autoadministrées a longtemps été associée au concept de désirabilité sociale. Ce concept représente la pierre angulaire de nombreuses études s'intéressant à la validité des mesures autoadministrées chez la population générale (Edwards, 1957; Nederhof, 1985; Paulhus, 1986) et chez les délinquants (Gudjonsson, 1990; Nugent et Kroner, 1996; Kroner et Weekes, 1996; Mills et Kroner, 2005). Dans la prochaine section, il sera présenté en quoi consiste la désirabilité sociale et de quelle manière ce concept peut aider à mieux comprendre la motivation des participants à dissimuler lors d'une passation.

## **1.2 La désirabilité sociale**

Un des biais, menaçant la validité des questionnaires autoadministrés, le plus rapporté depuis les quarante dernières années est sans doute celui lié à la désirabilité sociale (Edwards, 1957; Nederhof, 1985). La désirabilité sociale est motivée par un besoin général d'approbation de l'autre (Crowne & Marlowe, 1960). Concrètement, la désirabilité sociale chez un individu le pousse à s'attribuer des qualités socialement désirables et de rejeter ses qualités indésirables (Edwards, 1957). La grande menace que pose la désirabilité sociale à la validité des données réside dans le fait que le participant répond non seulement au contenu manifeste du questionnaire, mais aussi selon l'implication sociale de sa réponse (Tan & Grace, 2008). Il peut devenir alors difficile de séparer l'objet de la mesure du style de réponse du participant (Paulhus, 1991). Nous verrons tout d'abord de quelle manière la désirabilité sociale a été opérationnalisée à travers le temps, pour ensuite présenter différentes recherches ayant traité de la désirabilité sociale chez les délinquants.

### **1.2.1 L'opérationnalisation de la désirabilité sociale**

De nombreux chercheurs ont travaillé à comprendre et développer une mesure adéquate de la désirabilité sociale (Marlowe & Crowne, 1960; Wiggins, 1964; Sackeim et Gur, 1979; Paulhus, 1984; Helmes & Holden, 2003). Après plus de soixante années d'étude sur ce phénomène, il semblerait que la désirabilité sociale découle de deux dimensions distinctes. Dans son étude, Wiggins (1964) a identifié deux facteurs, *Alpha* et *Gamma*, liés au concept de la désirabilité sociale. Le facteur *Alpha* fait référence à l'aveuglement volontaire du participant, aussi nommé « autoduperie ». Un participant ayant un score élevé à ce facteur aurait tendance à diminuer l'ampleur et l'intensité des moins bons côtés de sa personnalité. La seconde dimension, le facteur *Gamma*, fait référence à la tendance du participant à se représenter aux autres d'une manière plus favorable, aussi nommée « hétéroduperie ». Un participant ayant un score élevé à ce facteur aurait tendance à augmenter l'ampleur et l'intensité des bons côtés de sa personnalité. Les travaux de Sackeim et Gur (1979) ont suivi cette logique

bidimensionnelle en créant deux questionnaires distincts, soit le *Self-Deception questionnaire* et le *Other Deception Questionnaire*. Comme ces travaux semblent l'indiquer, il y aurait une dimension axée sur l'aveuglement involontaire dirigé vers soi et une dimension axée vers la gestion de l'image de soi afin de duper l'autre, ce que Paulhus reprit en 1984.

À travers ses travaux, Paulhus (1984) identifia aussi deux dimensions au concept de désirabilité sociale. De ses travaux a découlé le *Balanced Inventory of Desirable responding* (BIDR)(Paulhus, 1984), qui est aujourd'hui sans doute un des outils les plus utilisés pour évaluer la désirabilité sociale. L'outil est composé de deux dimensions, l'autoduperie (nommé *Self-deception*) et la gestion de l'impression (*Impression management*). Ces deux dimensions font référence respectivement aux facteurs *Alpha* et *Gamma* identifiés vingt ans plus tôt par Wiggins (1964). La dimension de l'autoduperie est associée au fait de nier certaines caractéristiques indésirables, mais probables, de sa personnalité. Quant à elle, la gestion de l'impression fait référence à l'attribution de qualités désirables, mais improbables. Il existerait un aspect volontaire et conscient au fait de vouloir faire bonne impression, toutefois il y aurait un aspect inconscient et involontaire dans le fait de minimiser les caractéristiques négatives de sa personnalité (Paulhus, 1986; Sackeim et Gur, 1979).

En conclusion de son étude, Paulhus (1984) stipula que l'autoduperie ne serait pas un mécanisme de défense conscient, puisqu'elle ne serait pas affectée par le contexte de passation (ex : passation avec le couvert de l'anonymat vs sans le couvert de l'anonymat). En contrepartie, lorsque le chercheur analysa les pointages obtenus à l'échelle de gestion de l'impression, il remarqua que ceux-ci étaient nettement plus élevés lorsque les participants ne bénéficiaient pas du couvert de l'anonymat (Paulhus, 1984). Ce dernier conclut donc que les participants présentant un niveau élevé d'autoduperie seraient en train de se défendre contre des pensées qui représenteraient une menace à leur psyché, et ce, de manière inconsciente.

L'aspect volontaire et dirigé vers autrui associé à la gestion de l'impression a fait de cette dimension la mesure d'intérêt des chercheurs s'intéressant à la falsification des

réponses (Reeder et Ryan, 2011). D'ailleurs, cette dimension semblerait être davantage affectée par l'influence contextuelle, où plus l'importance à bien performer à l'évaluation est grande, plus sont élevés les scores à la mesure de la gestion de l'impression (Zerbe et Paulhus, 1987).

En bref, nous reconnaissons aujourd'hui deux dimensions distinctes à la désirabilité sociale (Paulhus, 2002). La première dimension nommée autoduperie (*Self-deception*) fait référence à un aveuglement involontaire de la part du participant, où celui-ci déforme inconsciemment son image de soi. Le fait de nier certains aspects négatifs de sa personnalité ne serait pas dû à un style de réponse adopté, mais ressemblerait davantage à un trait latent (Paulhus, 2002). L'autoduperie, contrairement à la gestion de l'image, ne serait pas affectée par l'influence contextuelle. La deuxième dimension, nommée gestion de l'impression (*Impression management*), fait référence à l'habileté de l'individu à gérer l'image qu'il projette aux autres. La gestion de l'impression sous-tend que le participant est conscient de l'image qu'il projette, et peut modifier adéquatement celle-ci afin de bien paraître aux yeux de l'évaluateur. Il existerait un aspect volontaire et conscient dans le fait de vouloir faire bonne impression (Paulhus, 1984), et plus l'influence contextuelle incite le participant à bien se présenter, plus grande sera la propension de celui-ci à modifier l'image qu'il projette (Zerbe et Paulhus, 1987). Puisque l'on reconnaît aux délinquants une meilleure capacité à manipuler et à tromper que chez les individus de la population générale (Schretlen et Arkowitz, 1990), il est peu surprenant qu'un nombre important de chercheurs se soient intéressés à l'évaluation de la désirabilité sociale chez les délinquants (Gudjonsson, 1990; Nugent et Kroner, 1996; Kroner et Weekes, 1996; Mills et Kroner, 2005).



### 1.2.2 La désirabilité sociale et la délinquance

Bien que la grande majorité des études traitant de la désirabilité soit issues d'études effectuées auprès de la population générale, la structure dimensionnelle de la désirabilité sociale serait sensiblement la même une fois appliquée à une population délinquante (Kroner et Weekes, 1996). Après avoir conduit une analyse factorielle confirmatoire auprès d'un échantillon d'individus incarcérés, Kroner et Weekes (1996) sont arrivés à la conclusion que le *Balanced Inventory of Desirable Responding* (Paulhus, 1984) serait une mesure appropriée et applicable auprès d'une population délinquante.

Néanmoins, certains auteurs restent sceptiques quant à la parfaite transférabilité du concept de désirabilité sociale chez une population de délinquants sexuels. Puisque bon nombre d'études sur la désirabilité sociale ont été conduites auprès d'une population mixte d'étudiants universitaires, les résultats peuvent être difficilement transférables à une population de délinquants sexuels, puisque ces derniers sont majoritairement des hommes, plus âgés et avec un niveau d'éducation et un QI généralement plus faibles (Beech et al, 2002; Gannon & Polachek, 2005 dans Tan & Grace, 2008). D'ailleurs, le stigma social associé aux délits sexuels serait un puissant incitatif à offrir des réponses prosociales pour ce type de délinquant (Tan & Grace, 2008) qui ne trouverait pas d'égal dans un échantillon de la population générale. L'impact de l'influence contextuelle est donc difficilement comparable entre les études conduites auprès d'une population d'étudiants universitaires et celles conduites auprès de délinquants sexuels.

Le biais lié à la désirabilité sociale pose un problème particulier pour ceux qui désirent évaluer les délinquants sexuels à l'aide de mesures autoadministrées. Le stigma social lié à leurs délits et leurs déviances sexuelles seraient des éléments non négligeables dans l'évaluation de la désirabilité sociale (Tierney et McCabe, 2001; Tan et Grace, 2008), d'autant plus qu'il est reconnu qu'une grande proportion d'hommes accusés ou condamnés d'un crime à caractère sexuel font du déni (Abel, Becker, Mittleman, Cunningham-Rathner, Rouleau et Murphy, 1987). D'ailleurs, le contenu des échelles d'évaluation serait particulièrement transparent, rendant les résultats vulnérables à la dissimulation (Tierney & McCabe, 2001). Par exemple, un délinquant confronté à un

questionnaire où on lui demande de rapporter son attirance sexuelle envers les enfants n'aura aucune difficulté à discriminer la réponse socialement acceptable de celle qui est non acceptable. Comparativement aux délinquants violents, les délinquants sexuels, et plus particulièrement les pédophiles, présenteraient un plus haut taux de réponses teintées de désirabilité sociale (Gudjonsson & Sigurdsson, 2000; Tierney & McCabe, 2001). De plus, ces derniers rejetteraient davantage leurs caractéristiques sexuelles indésirables que les autres types de délinquants (McGrath, Cann & Konopsky, 1998; Tierney & McCabe, 2001). Confrontée à une population ayant de bonnes raisons de dissimuler, l'évaluation criminologique doit composer avec la possibilité que l'individu évalué falsifie et déforme ses réponses.

### **1.2.3 Désirabilité sociale et falsification de données : même phénomène?**

Afin de clarifier l'objet d'étude du présent mémoire, il est important d'apporter certaines précisions afin de départager ce qu'est la falsification de données de la désirabilité sociale. Il existe plusieurs similarités dans les définitions respectives de ces phénomènes, il est toutefois important pour le lecteur de bien distinguer la falsification de données de la désirabilité sociale.

Tout d'abord, voici la définition proposée par Ziegler et ses collègues (2011) sur la falsification de données:

«Faking represents a response set aimed at providing a portrayal of the self that helps a person to achieve personal goals. Faking occurs when this response set is activated by situational demands and person characteristics to produce systematic differences in test scores that are not due to the attribute of interest. » (P.8)

Il existe trois aspects qui caractérisent la falsification de données. La falsification de données est: 1) motivée par un objectif personnel, 2) influencée par des caractéristiques contextuelles et individuelles et 3) produit une différence dans les scores non attribuables au trait mesuré. La falsification de données représente donc un

comportement, soit celui de rapporter des scores à une évaluation avec comme objectif l'autopromotion. Bien que cette définition semble adéquate, elle suppose une motivation sous-jacente sans toutefois l'expliquer. Puisque le fait d'offrir une réponse falsifiée représente un comportement motivé, l'origine de cette motivation doit être clarifiée. La désirabilité sociale, en contexte d'évaluation auprès de délinquants sexuels, semble être la motivation de prédilection pour expliquer la falsification des réponses.

Le biais de désirabilité sociale est généralement défini comme étant la tendance à offrir une autodescription trop positive (Paulhus, 2002). Il existe trois aspects centraux au phénomène de la désirabilité sociale. Premièrement, il existerait un aspect volontaire et conscient dans le fait de vouloir faire bonne impression (Paulhus, 1984). Deuxièmement, plus l'influence contextuelle incite le participant à bien se présenter, plus grande sera la propension de celui-ci à modifier l'image qu'il projette (Zerbe et Paulhus, 1987). Finalement, la désirabilité sociale représente un biais de réponses, où les réponses d'un participant se basent sur des considérations externes à l'objet mesuré (e.g. le fait de vouloir bien paraître versus le fait de donner une réponse juste) qui interfère avec la précision de la mesure (Paulhus, 2002). Donc, tout comme la falsification de données, la désirabilité sociale est à la fois influencée par un objectif personnel, par des caractéristiques contextuelles et individuelles et produit une différence dans les scores observés qui est non attribuable au trait mesuré. Les deux phénomènes partagent des caractéristiques similaires, mais sont malheureusement confondus à tort. Ne pas différencier la désirabilité sociale avec la falsification des données revient à confondre l'action avec sa motivation. La désirabilité sociale sera la force qui motivera et poussera le participant à falsifier ses réponses. En bref, il est important de retenir que les deux phénomènes sont intrinsèquement liés, tout en demeurant distincts.

À l'intérieur de ce mémoire, la falsification des données représente un comportement mesurable. Ce comportement sera influencé par le désir de bien paraître, aussi nommé biais de désirabilité sociale. La désirabilité sociale sera opérationnalisée comme étant l'objectif personnel visé par les participants offrant des patrons de réponses erronées.

Compte tenu du stigma social associé aux crimes sexuels ainsi que du contexte particulier lié à une évaluation psychologique en milieu carcéral, il est adéquat de supposer que ces hommes présenteront un niveau élevé de désirabilité sociale.

Une dernière considération à noter concerne la temporalité du biais de désirabilité sociale. Une des préoccupations des chercheurs qui se sont intéressés à la falsification des données était de déterminer si le fait de falsifier ses réponses s'avérait un style de réponse constant à travers le temps (Jackson et Messick, 1958) ou s'il s'agissait d'un biais de réponse (*response-set*) de courte durée attribuable à une motivation spécifique (Paulhus, 2002). Il est aujourd'hui reconnu que la falsification de données représente un biais non constant à travers le temps et spécifique à certains questionnaires (Paulhus, 2002; Ziegler et al., 2009). Dans son étude évaluant le biais de désirabilité sociale auprès d'une population de délinquants sexuels, Knight (1999) arriva à la conclusion que seulement certains domaines spécifiques étaient influencés par le biais de désirabilité sociale, tels que les domaines évaluant les comportements sexuels, les fantasmes sexuelles, l'impulsivité et l'hypermasculinité. C'est donc dire que la falsification de données, motivée par la désirabilité sociale, sera présente que dans certains domaines d'évaluation, et que même si un participant présente un patron de réponses erronées dans un domaine spécifique, le reste de son évaluation peut être exempté de désirabilité sociale.

Une des questions qui domine le domaine de la recherche sur la falsification de données est de savoir si les participants placés dans un contexte où il existe un incitatif à déformer leurs réponses le font vraiment (Smith et McDaniel, 2011). Afin de répondre à cette question, trois types de *designs* de recherche sont généralement employés afin d'étudier la potentielle falsification de données. Nous examinerons ces dernières dans la prochaine section du mémoire.

### **1.3 La recherche sur la falsification de données**

Évaluer l'impact de la désirabilité sociale sur les données est une tâche assez complexe. Les chercheurs qui s'intéressent à cette problématique sont confrontés à une embûche particulière : comment être certain que les participants déforment et falsifient réellement leurs réponses? On ne peut jamais avoir la certitude à 100 % que quelqu'un dissimule ou déforme ses réponses. Comme Griffith et Converse (2011) l'avancent, en se posant la question à savoir si les participants déforment et falsifient leurs réponses, les chercheurs se sont donnés la tâche de tenter de démontrer l'improvable. Or, pour contourner cette incertitude face à la réelle prévalence de la falsification des données, les chercheurs ont mis sur pieds différents protocoles de recherche où il est possible d'obtenir une mesure concrète de la falsification des données. Nous verrons quels sont les types de méthodologie généralement employés par les chercheurs qui désirent mesurer la falsification des données.

#### **1.3.1 Protocole avec instructions**

L'utilisation d'un protocole de recherche avec instructions (*experimentally induced faking*) est historiquement la stratégie la plus fréquemment utilisée (Smith et McDaniel, 2011). La stratégie consiste à comparer deux groupes de participants à l'aide de mesures indépendantes, où un groupe expérimental reçoit l'instruction de répondre au questionnaire de manière à se présenter le mieux possible et un groupe contrôle qui ne reçoit aucune instruction. Les chercheurs peuvent alors comparer, selon les groupes d'appartenance des participants, les scores obtenus aux échelles d'évaluation et aux échelles de validité. Il est reconnu que les participants ayant obtenu l'instruction de falsifier leurs données le font de manière efficace (Hough, Eaton, Dunnette, Kamp, et McCloy, 1990, dans Holden, Wood et Tomashewski, 2001). Une méta-analyse effectuée par Viswesvaran et Ones (1999) indique que l'influence de la falsification de données induite par une instruction augmente les scores moyens aux dimensions du NEO-PI-R (Costa et McCrea, 1995) d'environ un demi écart-type et augmente les scores aux échelles de désirabilité sociale d'au moins un écart-type. En bref, dans un contexte où

les participants reçoivent l'instruction de se présenter de manière favorable, ces derniers sont aptes à le faire et nous sommes en mesure de voir l'effet de la désirabilité sociale sur leurs données autoadministrées. Il est bien évident que les conclusions tirées d'études où la falsification de données est induite sont difficilement comparables aux études où la falsification de données se produit naturellement. Une critique souvent portée à ce type de recherche avec instructions est que la réelle « motivation » à falsifier ses données est complètement laissée de côté (Smith et Robie 2004). Notamment, Paulhus (2002) est d'avis que la désirabilité sociale se produisant de manière naturelle ne serait pas comparable à celle induite par une instruction. Pour combler cette lacune où le facteur « motivation » est laissé de côté, certains chercheurs emploient un second type de protocole de recherche nommé *Candidat-Titulaire*.

### **1.3.2 Protocole Candidat-Titulaire**

Une seconde approche utilisée consiste à comparer deux groupes d'individus, un pour qui le chercheur suppose qu'il existe une motivation à bien se présenter (personnes qui postulent pour un emploi convoité) et un groupe pour qui il n'existe peu ou pas de motivation à se bien se présenter (personnes qui détiennent déjà l'emploi). Généralement nommée l'approche *Candidat-Titulaire (Applicant-Incumbent)*, cette technique tente d'évaluer la falsification de données telle quelle devrait se produire naturellement. Dans leurs études comparant des personnes qui postulent pour un poste de moniteur de ski à ceux qui détiennent déjà ce poste, Rosse, Stecher, Miller et Levin (1998) ont trouvé qu'une inflation des scores se produisait chez le groupe candidat. Ces derniers obtenaient des scores plus élevés aux échelles de personnalité et des scores élevés pour les échelles mesurant la désirabilité sociale. L'influence contextuelle, soit le désir d'obtenir le poste, a influencé la manière dont ces gens ont performé aux évaluations autoadministrées. Une critique portée à ce type de protocole est que le postulat d'homogénéité des groupes associé à l'utilisation d'un design à mesure indépendant est rarement respecté. Mount et Barrick (1995) (dans Smith et McDaniel, 2011) avancent que les groupes « candidats » et « titulaires » ne sont pas suffisamment

comparables pour éliminer toutes autres hypothèses alternatives expliquant les différences dans les scores aux tests de personnalité et aux échelles de validité. Face à ces critiques, certains chercheurs ont préféré mesurer la désirabilité sociale à l'aide d'une mesure directe.

### **1.3.3 Protocole sans instruction**

La troisième approche visant à évaluer l'impact de la désirabilité sociale utilise une mesure directe sans instruction, à l'aide d'une ou plusieurs échelles de validité. Cette approche, qui tente de mesurer ce que l'on nomme le *naturally occurring faking* (Holden et Book, 2011) suppose que la déformation de réponse se produit naturellement dans un contexte donné et que les chercheurs peuvent être en mesure de la capturer à l'aide de différents outils d'évaluations. Cette approche n'utilise aucune instruction et ne compare aucun groupe entre eux. Bien évidemment, il existe aussi de nombreuses critiques face à l'emploi de cette stratégie. La capacité des échelles de validité à détecter adéquatement la falsification de données est souvent remise en question. Certaines échelles de validités covariant avec les dimensions du NEO-FFM, notamment les dimensions du névrosisme, d'agréabilité et de conscience, ce qui signifie que ces échelles mesurent des concepts non indépendants de la personnalité (Smith et Ellingson, 2001). La prévalence de la falsification, telle que détectée par les échelles de validité, sera donc affectée par les caractéristiques individuelles des participants et non pas par leurs réelles motivations à falsifier leurs données. Comme il est généralement reconnu dans la communauté des chercheurs, les échelles de validité capturent davantage de « substance » que de « style » (McCrea et Costa, 1983). La prudence est donc encouragée lorsque vient le temps d'interpréter les données issues d'échelles de validité et aucune conclusion hâtive ne doit être tirée.

Même si l'on doit les employer et les interpréter avec prudence, les échelles de validité sont encore aujourd'hui utilisées dans le but de détecter la falsification de données. Toutefois, de nouvelles méthodes alternatives à l'analyse des échelles de validité ont vu

le jour. Parmi celles-ci il existe l'analyse de processus, aussi nommée analyse des temps de latence. Dans la prochaine section du mémoire, les différentes techniques disponibles aux chercheurs désirent détecter et mesurer la falsification de données seront présentées.

#### **1.4 Les techniques de détection**

Lorsqu'un individu répond à un questionnaire, il existe trois marqueurs observables qui en découlent et qui sont à la disposition du chercheur : le contenu rapporté, la vitesse des réponses et la cohérence des réponses. Bien interprétées, chacune de ces variables peuvent renseigner sur la personnalité du répondant, ses motivations et ultimement, sur la validité de sa performance. Il sera question d'évaluer comment l'analyse du contenu rapporté par un participant peut renseigner sur la validité de sa performance.

##### **1.4.1 Analyse de contenu**

Confrontés à l'évidence que les participants ne sont pas toujours honnêtes dans leurs réponses et qu'il existe des situations où l'enjeu lié à l'évaluation augmente la probabilité que la personne évaluée falsifie ses réponses, de nombreux chercheurs ont tenté de développer différentes techniques de détection des réponses falsifiées. Les premières techniques élaborées consistaient généralement à évaluer le contenu manifeste des échelles de validité incluses dans les questionnaires. Un des premiers questionnaires à employer des échelles de validité a été le *MMPI* (Minnesota Multiphasic Inventory) (Hathaway & McKinley, 1942). À ce jour, le *MMPI-2* comprend onze échelles de validité mesurant à la fois la cohérence des réponses du participant, la simulation de symptômes psychopathologiques (*faking-bad*) et la dissimulation (*faking-good*). Notamment, s'y retrouve la *K scale* (Meehl et Hathaway, 1946) mesurant les formes de défenses que les participants peuvent endosser dans leurs patrons de réponses et la *L scale*, qui mesure si le répondant tente de se présenter de manière trop



favorable. Par exemple, un profil au MMPI est considéré invalide et ininterprétable si le score à l'échelle *L* se situe trois écarts-types au-dessus de la moyenne (Butcher, 2006).

Une autre technique employée par les chercheurs qui désirent détecter les participants qui falsifient leurs réponses constitue à inclure dans leur protocole de recherche un questionnaire mesurant la désirabilité sociale. Le *Marlowe-Crowne Desirability Scale* (MCDS; Marlowe & Crowne, 1960) et le *Balanced Inventory of Desirable Responding* (BIDR; Paulhus, 1984) sont tous deux des mesures de désirabilité sociale fréquemment utilisées. La MCDS se subdivise en deux sous-échelles, une mesurant la tendance à s'attribuer des caractéristiques positives, mais improbables, et l'autre mesurant le déni de caractéristiques négatives, mais probables. Tout comme la MCDS, le BIDR est composé de deux dimensions distinctes. La première dimension mesure l'autoduperie (*Self-deception*), soit la tendance à dénier des caractéristiques négatives, mais probables (e.g. Je ne jure jamais). La deuxième dimension mesure la gestion de l'impression (*Impression-Management*), soit la tendance à s'approprier des caractéristiques positives, mais improbables (e.g. Je respecte toujours la limite de vitesse).

Créé plus spécifiquement pour une population de délinquants sexuels, le *Sexual Social Desirability Scale* (SSDS; McGrath, Cann & Konopsky, 1998) permet d'évaluer le biais de désirabilité sociale, avec l'accent mis sur les attitudes, les comportements et les croyances liés au sexe. Tout comme le MCSD et le BIDR, cet outil se sous-divise en deux dimensions. La première échelle nommée *Acquiescence* mesure si le participant tente de s'approprier des qualités improbables (p.ex. Il m'arrive d'acheter des revues pornographiques principalement pour lire les articles). La deuxième échelle, nommée *Denial*, évalue si le participant nie avoir des caractéristiques négatives, même si celles-ci sont probables pour la plupart des gens (p. ex. Il m'arrive d'avoir des pensées sexuelles au travail).

Que ce soit à l'intérieur même du contenu d'un questionnaire ou que ce soit une mesure parallèle comprise dans le protocole de recherche, l'emploi d'outils d'évaluation de la désirabilité sociale fonctionne généralement de la même façon. Le chercheur additionne le pointage obtenu aux échelles de validité et détermine un point de coupure, où, à

partir d'un score prédéterminé, il peut établir qu'un participant falsifie et déforme ses réponses afin de bien paraître. Il est toutefois important de noter que l'interprétation des résultats obtenus aux échelles de validité nécessite prudence. Comme l'avance Holden (2007): « *Although fakers will score at the extreme on scales of socially desirable responding, extreme scorers on scales of socially desirable responding are not necessarily fakers* ». Il peut être donc utile d'employer une seconde stratégie permettant de capturer d'autres indicateurs de la dissimulation, tels que les temps de latence.

#### **1.4.2 Analyse de processus**

Depuis la venue de l'informatisation des questionnaires, il est maintenant possible pour les chercheurs d'avoir accès à une information autrefois inaccessible: le temps pris par un participant pour répondre à une question. Contrairement à l'analyse de contenu qu'offrent les échelles de validité, l'analyse des temps de latence permet aux chercheurs d'effectuer une analyse de processus. En effet, l'action de répondre à une question implique une série de processus cognitifs (Sudman, Bradburn & Schwarz, 1996; Vrij & Heaven, 1999; Vendemia, Buzan et Green, 2005). Les chercheurs s'intéressant au processus cognitif lié à la dissimulation s'entendent pour dire que le temps de réaction serait une mesure directe du processus cognitif (Vendemia et al, 2005).

##### **1.4.2.1 Le processus cognitif lié à l'élaboration d'une réponse**

D'un point de vue strictement cognitif, il existerait trois étapes permettant d'expliquer le processus lié à la génération d'une réponse (Walczyk, Roper, Seemann, et Humphrey, 2003). La première étape est celle de l'activation où le participant entend la question, l'interprète et récupère l'information pertinente dans sa mémoire à long terme. La deuxième étape consiste en la prise de décision pour le participant à offrir une réponse authentique ou non. La dernière étape, si le participant choisit de falsifier sa réponse, consiste à construire une réponse non authentique, où il doit faire appel à sa mémoire à

long terme afin de créer un mensonge plausible. Cette dernière étape liée à la construction d'une réponse non authentique serait celle qui augmenterait le temps de réponse des participants (Walcyk et al., 2003). Peu importe la conceptualisation du processus cognitif employé, les auteurs s'entendent pour dire qu'il existe une étape du processus cognitif où le participant fait appel à sa mémoire pour récupérer l'information pertinente et une étape d'édition où celui-ci décidera de la réponse qu'il rapportera (Holtgraves, 2004).

Holtgraves (2004) propose trois situations où la désirabilité sociale pourrait avoir un impact sur le processus cognitif lié à la génération d'une réponse. La première situation est celle où la désirabilité sociale vient éliminer l'étape de la récupération de l'information ainsi que l'étape d'évaluation. Comme Hsu et ses collègues (1989) le proposent, si la désirabilité sociale élimine le processus de récupération et d'évaluation pour laisser place à une simple évaluation sémantique et littéraire de la question, les temps de réponse devraient être plus rapides. Ces participants ne feraient qu'une analyse superficielle, littéraire et sémantique de la question afin de déterminer rapidement qu'elle est la « bonne » réponse à donner, sans s'impliquer dans un processus cognitif complet (Rogers, Kuiper et Kirker, 1977). Hsu et ses collègues (1989) avancent qu'une réponse honnête requiert une introspection plus approfondie chez les participants, ce qui augmenterait les temps de réponse. La deuxième possibilité est celle où la désirabilité sociale vient biaiser le processus de récupération de l'information. Dans ce cas, les participants récupérerait seulement l'information qui les avantagerait, et omettrait l'information qui les désavantagerait. Si ce mécanisme s'opère, les temps de latence devraient être plus rapides, puisque seulement un processus partiel de récupération d'information est engagé.

La troisième possibilité serait celle où la désirabilité sociale aurait un impact sur le processus d'édition de la réponse. Dans cette situation, le participant récupère l'information pertinente dans sa mémoire, la reformule, et évalue par la suite sa réponse en regard des considérations liées à la désirabilité sociale. Si la désirabilité sociale joue un rôle dans le processus d'édition, ultimement elle fera augmenter les temps de

réponse des participants. McDaniel et Timm (1990) soumettent l'idée que le fait de dissimuler prendrait plus de temps, car le participant doit donner une réponse incohérente avec son schéma de pensées. Le même point de vue est partagé par Vrij et Heaven (1999). Pour ces auteurs, offrir une réponse authentique nécessite un moins grand effort cognitif que de dissimuler, puisque le processus cognitif lié à l'élaboration d'une réponse non authentique nécessite un effort cognitif afin d'inhiber et de modifier la « vraie » réponse.

En bref, le processus cognitif lié à l'élaboration d'une réponse a comme caractéristique d'affecter le temps pris par un participant à répondre à une question. (Vendemia et al., 2005). Lorsque vient le temps d'analyser le lien entre la désirabilité sociale et les temps de latence, deux écoles de pensées coexistent. D'un côté, des temps de latence trop rapides indiqueraient généralement que le participant n'a fait qu'une analyse sémantique afin de déterminer la « bonne » réponse à offrir (Hsu et al., 1998). À l'inverse, des temps de latence trop lents seraient un marqueur d'un processus cognitif plus complexe, tel que demandé lorsque vient le temps d'élaborer une réponse non authentique, mais plausible (Holtgraves, 2004). Il existe un modèle d'analyse permettant d'expliquer comment un temps de latence rapide et un temps de latence lent peuvent être tous deux des marqueurs de dissimulation.

#### **1.4.2.2 La congruence du schéma cognitif**

Afin d'étudier le lien entre les temps de latence et la dissimulation, Holden et ses collègues (1992) ont élaboré un modèle d'analyse qui se base sur le concept psychologique de schéma cognitif. Les schémas cognitifs sont des représentations mentales influencées par nos expériences passées, qui permettent d'organiser de façon structurée les événements et les expériences semblables dans le but de diriger nos comportements futurs (Fekken et Holden, 1992). Elles servent de cadre de référence et facilitent le traitement d'informations nouvelles. Holden et ses collègues (1992) proposent d'ailleurs que les traits de personnalité représentent des composantes

structurelles des schémas cognitifs. Or, ces derniers avancent qu'un schéma cognitif important (e.g. un trait de personnalité dominant) va induire un traitement différencié des informations pertinentes au schéma (*schema-relevant*) de celle non pertinente (*schema-irrelevant*). Les informations cohérentes avec le schéma cognitif d'un participant seront endossées plus rapidement, alors que les informations non cohérentes seront rejetées plus rapidement (Holden, Fekken et Coton, 1991). La relation unissant le temps de latence et le trait mesuré, en respectant la théorie de la congruence schématique, devrait être représentée par un U inversé. En conséquence, un participant se trouvant à l'extrémité supérieure d'un trait de personnalité endossera plus rapidement les items en lien avec la mesure de ce trait et les participants se trouvant à l'extrémité inférieure rejeteront plus rapidement les items en lien avec la mesure de ce trait.

Toutefois, lorsque survient une incohérence entre le schéma cognitif du participant et la réponse qu'il offre, les temps de latence sont généralement plus lents. Donc, le participant se trouvant à l'extrémité supérieure d'un trait de personnalité rejettera plus lentement les items en lien avec la mesure du trait de personnalité en question. Concrètement, un participant voulant dissimuler ses attitudes et ses comportements problématiques lors d'une évaluation autoadministrée peut le faire, toutefois, ses temps de réponse seront plus lents.

Dans son étude avec protocole *Candidat-Titulaire*, Holden (1995) a appliqué sa théorie de la congruence schématique afin d'évaluer cette théorie et son impact sur les temps de latence des participants. En donnant l'instruction de bien se présenter de manière à maximiser leur probabilité d'être engagé, le chercheur a induit un schéma cognitif chez la moitié des participants. Dans ses instructions, le chercheur donnait l'avertissement aux participants de faire attention à la crédibilité de leur performance puisqu'à l'intérieur du questionnaire se trouvaient des échelles de validité mesurant la désirabilité sociale. Du nombre de questions comprises dans l'étude, une partie d'entre elles évaluait la délinquance. Les participants ayant eu l'instruction de se présenter de manière à maximiser leur probabilité d'être engagé ont répondu (en rejetant les

affirmations) aux questions liées à la délinquance plus rapidement que le groupe contrôle. Toutefois, quand venait le temps d'endosser une question liée à la délinquance (afin de rendre leurs patrons de réponses crédibles) les participants du groupe expérimental ont obtenu des temps de latence significativement plus lents.

En bref, le schéma cognitif d'un participant est généralement influencé par sa personnalité. Par exemple, un participant se trouvant à l'extrémité supérieure d'un trait de personnalité endossera plus rapidement les items en lien avec sa personnalité. À l'inverse, ce dernier sera plus lent lorsqu'il décidera de rejeter les items en lien avec sa personnalité. Toutefois, les conclusions d'études qui employaient le protocole *Candidat-Titulaire* ont démontré qu'un participant ayant obtenu l'instruction de se présenter de manière à augmenter sa probabilité d'être engagé adapte son schéma cognitif afin d'atteindre son objectif. On peut donc supposer qu'un participant placé dans un contexte d'évaluation où il existe un incitatif à bien se présenter adaptera son schéma cognitif en conséquence.

Nous avons maintenant présenté deux approches permettant l'analyse de performance aux questionnaires autoadministrés. L'analyse de contenu permet, à travers les échelles de validité, d'évaluer le contenu que rapporte un participant. Les participants présentant de hauts niveaux de désirabilité sociale, telle que mesurée par les échelles, mettent la puce à l'oreille aux chercheurs quant à la validité de leurs réponses. L'analyse de processus renseigne sur le temps pris pour répondre aux questions pour chacun des participants. Des temps de latence trop rapides ou trop lents éveillent le soupçon quant à la possibilité qu'un participant ait déformé ses réponses. Finalement, il existe une technique statistique qui permet d'évaluer la cohérence des réponses offertes par un participant. En effet, le modèle de Rasch présente différents coefficients d'ajustement qui nous permet d'apprécier le patron de réponses d'un participant. Issue de la famille statistique de la théorie de réponse à l'item, cette dernière présente une nouvelle approche de plus en plus employée par les chercheurs qui s'intéressent à la falsification des données (Zickar et Sliter, 2011).

### 1.4.3 Analyse de cohérence des patrons de réponses

Il existe deux grandes familles de modèles statistiques employés en sciences sociales: les modèles issus de la théorie classique des tests (TCT) et le modèle issu de la théorie de réponse à l'item (TRI). Depuis les dernières années, l'utilisation de la théorie de la réponse à l'item devient de plus en plus courante en sciences sociales (Bond & Fox, 2007). Les modèles de la TRI ont été pensés afin de pallier les désavantages des modèles de la TCT. Les modèles de la TCT conceptualisent le score d'un individu comme étant une relation linéaire constante composée du vrai score et de l'erreur de mesure (Hamon, 2000). Dans les modèles de la TRI, la relation entre le score observé et le trait latent mesuré n'est pas nécessairement linéaire. Les réponses observées sont le résultat d'un processus aléatoire dans lequel la probabilité de donner une réponse dépend de la difficulté de l'item et de l'aptitude du participant (Hamon, 2000)

Il existe de nombreux modèles issus de la TRI, parmi lesquels le modèle de Rasch (Rasch, 1960). Ce modèle a été élaboré à la fin des années 1950 par un mathématicien danois du nom de Georg Rasch (1901-1980), au départ pour des formats de réponses dichotomiques et par la suite adaptés pour les questions polytomiques (Penta, 2005). Le modèle de Rasch, aussi nommé modèle à un paramètre logistique (1PLM), a la caractéristique d'offrir une mesure unidimensionnelle objective, c'est-à-dire qui sera valable peu importe le sous-ensemble d'items utilisés et quel que soit l'échantillon de personnes évaluées.

Le modèle de Rasch prend en considération que chaque question qui mesure une dimension unique ne présente pas le même niveau de difficulté et que chaque participant ne présente pas le même niveau d'habiletés. L'utilisation du modèle de Rasch permet à la fois une évaluation au long d'un continuum des items qui composent la dimension mesurée, mais aussi des participants (Bond & Fox, 2007). Le niveau de difficulté d'un item et le niveau d'habiletés des participants sont tous deux exprimés en *logits* (*log-odds unit*). Selon la matrice de données observées, la difficulté d'un item sera représentée par le log naturel du rapport (*odd*) entre le nombre de mauvaises réponses et le nombre de bonnes réponses obtenues (0 et 1 dans le cas d'un modèle

dichotomique). L'habileté d'un participant sera calculée par le log naturel du rapport entre le nombre d'items réussis et le nombre d'items échoués. À l'annexe 1, un cas fictif qui détaille de manière simplifiée le calcul des *logits* pour les items et les participants est présenté.

Comme il est possible de le voir, les questions qui présentent un plus haut niveau de difficulté, dû au plus grand nombre d'échecs, obtiennent des scores *logits* supérieurs à zéro (par exemple, Q5 = 0,40). Nous sommes aussi en mesure d'évaluer l'habileté des participants, ceux ayant réussi un plus grand nombre d'items obtiennent des scores de *logits* supérieurs à zéro (par exemple, Participant 1 = 0,60). L'avantage de la mesure *logit* est que celle-ci représente une réelle échelle à intervalle puisque la distance entre chaque unité reste constante, due à la transformation logarithmique. Il est important de noter que le modèle de Rasch ne peut calculer de *logit* pour les participants présentant des scores observés de zéro et des scores parfaits. La raison est simple, le rapport entre le nombre d'items réussis et échoués introduit dans ces deux cas la valeur de zéro. Alors, quand vient le temps de calculer le rapport, ayant comme numérateur ou dénominateur la valeur zéro, le résultat sera toujours zéro. Comme le *logit* représente le logarithme naturel du rapport entre le nombre d'items réussis et échoués, il est impossible d'effectuer le logarithme naturel de zéro.

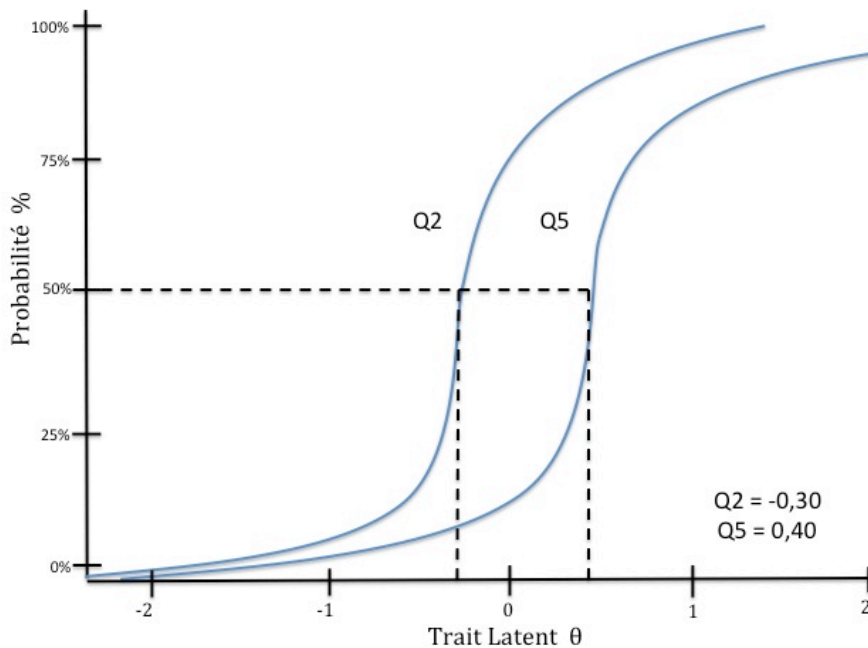
#### **1.4.3.1 La courbe caractéristique de l'item**

Comme il a été présenté, à la base du modèle de Rasch se trouve un simple calcul de probabilités, un pour les items et un pour les participants. Le paramètre associé aux participants, souvent nommé *trait latent*, est représenté par le symbole thêta ( $\Phi$ ). Ce paramètre est distribué à la manière d'un score standardisé avec une moyenne de zéro et un écart-type de la valeur un. La probabilité qu'un participant réussisse une question sera fonction de son niveau d'habiletés et du niveau de difficulté de la question. La relation qui unit le trait latent et la difficulté d'un item est représentée par une fonction monotone croissante qui se représente par une courbe caractéristique d'items.



La figure I représente une courbe caractéristique d'items pour un modèle dichotomique, où l'abscisse représente la distribution du trait latent et l'ordonnée la probabilité de réussite à l'item.

**Figure I. Exemple de courbe caractéristique de l'item (CCI)**

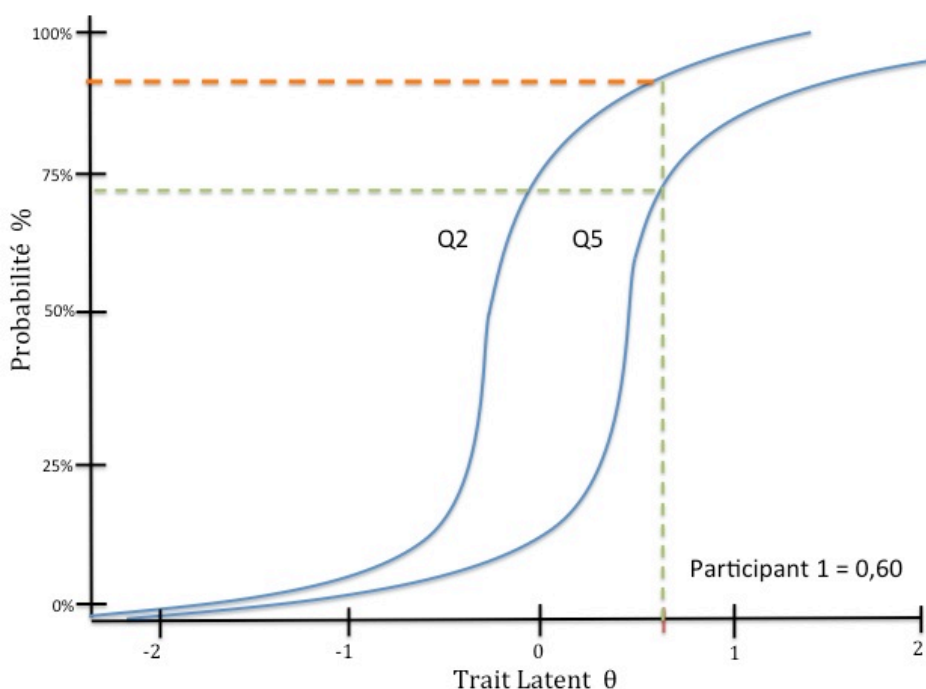


En traçant une ligne en partant de l'ordonnée à 50 % jusqu'à la courbe caractéristique, il est possible d'arriver au point d'inflexion de celle-ci, où la courbe passe de convexe à concave. Le point d'inflexion détermine le niveau de difficulté de la question. Par exemple, la question 2 (telle que présentée à la figure I) présente un niveau de difficulté de -0,30, c'est-à-dire qu'un participant ayant lui aussi un niveau d'habiletés de -0,30 aura la bonne réponse à cette question dans 50 % des cas. C'est donc dire qu'un participant ayant une plus grande habileté détient une plus grande probabilité de réussir une question qu'un participant avec moins d'habiletés. De même, les items

faciles ont plus de probabilité d'être répondus par un plus grand nombre de participants que les items difficiles.

L'avantage principal du modèle de Rasch, lorsque vient le temps d'évaluer les patrons de réponses des participants, repose sur ces coefficients d'ajustements. Ce modèle statistique permet de sélectionner un ensemble d'items et d'évaluer si ces derniers sont bien ajustés au modèle unidimensionnel mesuré. Le Rasch permet de calculer le score attendu selon la matrice des données observées. La comparaison entre les scores observés et les scores attendus permet de quantifier l'ajustement des données observées. En poursuivant avec l'exemple présenté à l'annexe 1, la figure II illustre un exemple où le score observé d'un participant diffère du score attendu.

**Figure III. Représentation graphique d'un écart entre le score attendu et le score observé à une question**



Tel que présenté à la figure II, le participant 1 a échoué la question 2. Toutefois, selon la relation unissant la difficulté de l'item et son niveau d'habiletés, ce dernier aurait dû vraisemblablement réussir cette question. Cet écart entre le score obtenu et observé est nommé *résidu*.

Le degré d'ajustement entre les réponses observées et attendues indique dans quelle mesure les réponses observées de chaque participant à chacune des questions se conforment aux prescriptions du modèle. En additionnant les résidus, il est possible de calculer un indice d'ajustement pour chaque personne (*person fit*). Il existe deux indices d'ajustement, l'*Outfit* (*Outlier sensitive fit statistic*) et l'*Infit* (*Information weighted fit statistic*). Ces deux indices peuvent être présentés soit sous la forme des carrés moyens (*Mean Square*), soit sous la forme des carrés moyens standardisés (ZSTD). Plus un participant aura de valeur résiduelle entre le score attendu et le score observé, plus élevé sera son indice d'ajustement.

Pour ce mémoire, le coefficient *Outfit* sous sa forme standardisée (ZSTD) utilisée dans le logiciel Winstep (Linacre, 2004) sera employé. L'indice d'ajustement *Outfit* est calculé en faisant la moyenne des carrés des résidus standardisés et révèle les réponses inattendues des personnes ayant une localisation éloignée de celle de l'item (Linacre, 2004). L'indice *Outfit* ZSTD a une valeur attendue de 0, pour les patrons de réponses s'ajustant parfaitement au modèle prescrit, avec un écart-type de 1.

Avec le modèle de Rasch, l'indice d'ajustement permet de mesurer à quel point le paramètre de l'item (son *Logit* calculé selon son niveau de gravité) est cohérent avec le *Logit* du répondant (son niveau d'implication à la dimension). Théoriquement, les items faciles seront endossés par une grande proportion de répondants, de même qu'un répondant ayant un haut niveau d'aptitudes réussira un plus grand nombre d'items. Il arrive certaines fois qu'un répondant présentant un *Logit* élevé ne parvienne pas à répondre aux items ayant un *Logit* faible. Ce sont ces cas de figure qui éveillent un soupçon quant à la validité des réponses obtenues par ce répondant. Dans ce cas, le coefficient *Outfit* du répondant sera élevé. Selon Smith, Shumacker et Bush, (1998) l'intervalle des valeurs standardisées des coefficients d'ajustement est acceptable s'il se

situe entre -2 et 2. Un coefficient de fit sous le seuil de -2 signifie que la mesure du participant *Overfit* (colle trop parfaitement à la prescription du modèle) et un coefficient de fit au-dessus de 2 signifie que la mesure du participant *Underfit* (ne concorde pas avec le modèle prescrit).

Le cas de figure intéressant pour ce mémoire sera l'*Underfit*, où le patron de réponses du participant ne correspond pas à la prescription du modèle. Nous vous présenterons pourquoi l'*Underfit* représente un marqueur de la falsification des données.

#### **1.4.3.2 Les coefficients d'ajustement et les patrons de réponses**

Les coefficients d'ajustement issus des modèles de la TRI ont tout d'abord été développés afin de détecter les tricheurs aux épreuves cognitives (Meijer, 2004). Un participant présentant un faible niveau d'habiletés, mais qui parvient à copier certaines réponses de son voisin obtiendra un coefficient d'ajustement aberrant. En échouant la plupart des questions, mais en réussissant une série d'entre elles, ce tricheur présentera un patron de réponses incohérent. L'indice d'ajustement sera donc influencé par le niveau d'adéquation entre le trait de participant et le niveau de difficulté des réponses qu'il parviendra à réussir. De mauvais scores d'ajustement signifient soit que le participant a triché, qu'il a répondu aléatoirement aux questions, qu'il a offert une prestation désintéressée ou qu'il a simplement intentionnellement offert une prestation non représentative de son trait latent (Schmitt, Chan, Sacco, McFarland et Jennings, 1999). Être en mesure de détecter les participants qui adoptent un patron de réponses aberrant représente un atout intéressant lorsque vient le temps d'analyser leur performance à un questionnaire autoadministré. Voici comment il est possible d'adopter le modèle de Rasch à l'objectif principal de ce mémoire.

Tout d'abord, il est primordial d'ajuster le vocabulaire employé par les théoriciens du modèle de Rasch. À l'intérieur de ce mémoire, les termes *difficultés* et *habileté* prennent un tout autre sens. Par exemple, en appliquant le modèle de Rasch à une échelle d'*Impulsivité*, la difficulté de l'item se traduit par son niveau de gravité. Comme il

n'existe pas de « bonnes » ou de « mauvaises » réponses à une échelle d'impulsivité, il est préférable de parler de gravité. Alors, un item sera endossé (réussi ou répondu) par moins de participants s'il présente un niveau de gravité plus élevé. À l'inverse, un item sera endossé par une plus grande proportion de participants si son niveau de gravité est moins important. En respectant la prescription du modèle, nous devrions nous attendre à avoir des patrons de réponses cohérents avec la gravité des questions. Or, il arrive que certains participants présentent des patrons de réponses aberrants. Au tableau 1 les coefficients *logits* des questions issues de l'échelle d'*Impulsivité* (échelle likert 0-5) se trouvant à l'intérieur du MIDSA (Multidimensional Inventory of Development, Sex and Aggression, Augur Enterprises, 2011) ainsi que deux exemples fictifs de patrons de réponses sont présentés.

**Tableau I. Exemple d'ajustement entre le score observé et le score attendu à une échelle**

Échelle d'évaluation	Difficulté (logit)	Participant1	Participant2
<b>Impulsivité</b>		Score observé	
J'ai blessé quelqu'un avec mes paroles	-0,46	4	0
Mon humeur change subitement	-0,31	3	4
J'ai déjà agi sur l'impulsion	-0,25	2	0
J'ai perdu contrôle de moi-même	0,10	1	0
Même si ce n'était de ma faute, eu des problèmes	0,27	1	0
Fais actions qui m'ont fait sentir coupable	0,31	1	4
Sentiments effrayants que je ne peux comprendre	0,34	0	4
	Total	12	12

Théoriquement, plus une question est difficile, plus la probabilité de l'endosser est faible. Dans notre exemple, la question *J'ai blessé quelqu'un avec mes paroles* représente la question avec le *logit* le plus faible (-0,46) et la question *Sentiments effrayants que je ne peux comprendre* représente la question avec le *logit* plus élevé (0,34). Or, un patron de réponses respectant la prescription du modèle présentera une certaine logique entre

ces réponses. Comme il est possible de le voir au tableau I, le participant 1 présente un patron de réponses cohérent avec la prescription du modèle. Toutefois, le patron de réponses du participant 2 ne semble pas respecter la prescription du modèle. Nous pouvons donc supposer que l'indice d'ajustement participant respectera la prescription du modèle et celui du participant 2 présentera un *Underfit* du modèle, signifiant une mauvaise adéquation entre les réponses observées et les réponses attendues.

Lorsqu'un participant est confronté à une évaluation autoadministrée dans laquelle il ne désire pas rapporter certaines attitudes ou comportements problématiques, il performera d'une manière atypique. Bien que certains vont nier en bloc chacune des questions, d'autres, dans l'optique d'offrir une performance plausible, vont varier leurs réponses afin de paraître le moins suspect possible. Dans son étude, Holden (1992) a observé que les participants qui avaient reçu l'instruction de se présenter le mieux possible à une évaluation autoadministrée, tout en faisant attention à ne pas se faire détecter par les échelles de validité, parvenaient à modifier leurs patrons de réponses afin de paraître le moins suspect possible. Nous croyons que ce procédé s'enclenche automatiquement chez les gens voulant dissimuler mais qui désirent offrir une performance qui semble plausible. D'ailleurs, admettre certaines problématiques est nécessaire pour que le participant démontre un certain niveau d'honnêteté. Certains auteurs ont rapporté ce « désir » de crédibilité chez les délinquants (Nagayama Hall, 1989 ; Mills et Kroner, 2005). La volonté de dissimuler motivera les participants à rapporter très peu de problématiques, mais dans le l'optique d'offrir une performance plausible, ces derniers vont endosser certaines questions qui ne seront pas toujours cohérentes avec la prescription du modèle. Ce processus entrainera donc en un patron de réponses qui présentera un mauvais ajustement (*Underfit*) avec le modèle prescrit.

## 2. Problématique

Lorsque vient le temps d'étudier l'étiologie des problématiques chez les délinquants sexuels, les cliniciens et chercheurs qui emploient des mesures autoadministrées sont confrontés à un défi de taille. En raison du puissant stigma social lié aux délits et aux déviances sexuelles de ces participants, le biais de désirabilité sociale pose une menace à la validité de l'information recueillie. En effet, cette menace à la validité réside dans le fait que les participants répondraient non seulement au contenu manifeste du questionnaire, mais aussi selon l'implication sociale de leurs réponses (Tan et Grace, 2008). De plus, les questions contenues à l'intérieur des échelles d'évaluation sont particulièrement transparentes, rendant les résultats vulnérables à la dissimulation (Tierney et McCabe, 2001). La transparence des questions, ajoutée au fait qu'une grande proportion d'hommes accusés ou condamnés d'un crime à caractère sexuel font du déni (Abel et al., 1987) explique bien tout le scepticisme entourant la validité et l'utilité d'employer des mesures autoadministrées auprès de délinquants sexuels (Gendreau, Irvine et Knight, 1973). La question est de savoir si l'utilisation de mesures autoadministrées auprès de délinquants sexuels permet de renseigner réellement sur leurs attitudes et leurs opinions, ou bien simplement sur leur désir de bien paraître.

Il existe trois stratégies permettant d'évaluer la prévalence de la désirabilité sociale. Une des stratégies les plus fréquemment employées consiste à intégrer une échelle de désirabilité sociale au protocole de recherche. Toutefois, une telle technique requiert énormément de prudence dans l'interprétation des données, puisque ces échelles capturent non seulement un biais potentiel, mais aussi un trait covariant avec certaines dimensions de la personnalité (Smith et Ellingson, 2011). Les participants qui dissimulent et présentent un haut niveau de désirabilité sociale se trouveront à l'extrémité positive de ces échelles. Par contre, la totalité des participants se trouvant à l'extrémité positive de ces échelles n'est pas nécessairement en train de dissimuler (Holden, 2007). C'est pourquoi il est avisé d'employer plus d'une stratégie de détection.

Une seconde stratégie, qui consiste à analyser les temps latence, permet d'évaluer le processus cognitif entrepris par le participant lorsqu'il est confronté à un questionnaire

autoadministré. Trois écoles de pensées coexistent. Certains chercheurs croient que les participants voulant se présenter de manière favorable présenteront des temps de latence plus rapides, puisque la tâche de déterminer la bonne réponse à donner ne nécessite qu'une analyse littéraire et linguistique de la question (Hsu et al., 1998). À l'inverse, certains chercheurs croient que d'offrir une performance non authentique à un questionnaire autoadministré nécessite un plus grand effort cognitif. Notamment, le fait de devoir inhiber son désir de dire la vérité et d'inventer une fausse réponse qui reste plausible augmenterait les temps de réponse des participants désirant offrir une performance non authentique (Holtgraves, 2004). Finalement, d'autres chercheurs croient que le temps de latence requis est fonction de la cohérence entre la réponse donnée et le schéma cognitif du participant (Holden, 1995). Un participant offrant une réponse incohérente avec son schéma cognitif prendra plus de temps à répondre.

Finalement, le modèle de Rasch permet d'obtenir un indice d'ajustement qui représente le niveau d'adéquation entre les scores observés et les scores attendus pour chacun des participants. Des participants présentant des patrons de réponses aberrants, qui ne respectent pas la prescription du modèle, obtiennent des scores d'*Outfit* qui indique qu'ils ont offert une performance hors normes. Il est reconnu que les indices d'ajustement hors normes peuvent être dus à des performances où le participant a intentionnellement déformé ses réponses (Zickar et Drasgow, 1996).

Ce mémoire s'intéresse donc à la manière dont les délinquants sexuels répondent aux questionnaires autoadministrés, ainsi qu'à la potentielle utilité d'employer différentes stratégies de détection de réponses falsifiées. Puisqu'il existe un impératif pour les cliniciens et chercheurs travaillant auprès de délinquants sexuels à statuer sur la validité des mesures autoadministrées, ceux-ci se doivent d'être en mesure de comprendre et d'analyser les performances de ces derniers.

Bien qu'il soit impossible de connaître les réelles intentions des participants, et que la quête de la vérité en est une impossible à accomplir, nous tiendrons pour acquis que certains participants déformeront leurs réponses. Une approche axiomatique sera donc privilégiée, puisque l'évaluation autoadministrée aborde des thèmes particulièrement



sensibles, il sera pris pour acquis que certains participants falsifieront leurs réponses à un moment ou un autre.

À travers une analyse de contenu, du temps de latence et de cohérence des réponses, le potentiel impact de la désirabilité sociale sur la performance au questionnaire autoadministré sera documenté.

## **2.1 Objectifs de l'étude**

### **2.1.1 Objectif général**

L'objectif principal du présent mémoire sera d'explorer les processus liés aux prestations à un questionnaire autoadministré à l'aide de trois indicateurs.

### **2.1.2 Objectifs spécifiques**

- Évaluer l'impact de la désirabilité sociale sur le contenu rapporté à l'évaluation à l'aide d'échelles de validité;
- Déterminer le lien unissant les temps de latence et les scores aux échelles de désirabilité sociale;
- Évaluer la potentielle convergence des analyses de Rasch avec les échelles de désirabilité sociale et les temps de latence.

Puisque les échelles de validité représentent l'outil le plus fréquemment employé pour juger de la prestation des participants, à l'intérieur de ce mémoire, nous déterminerons si les participants signalés par cette mesure performant de façon différente. Si les échelles de validité mesurent bel et bien un désir de vouloir bien paraître, nous croyons que les participants qui présentent les plus hauts niveaux de désirabilité sociale performeront de manière différente au test autoadministré. L'analyse des temps de latence et de la cohérence de leurs patrons de réponses vont fournir des indications

quant au processus entrepris par ces participants. Compte tenu des limites concernant l'utilisation de mesures autoadministrées auprès de délinquants sexuels, ce mémoire permettra de mieux comprendre l'utilité d'employer différentes stratégies de détection de patrons de réponses erronés. Le tout permettra de vérifier si le scepticisme entourant la validité des questionnaires autoadministrés a une réelle raison d'être.

### 3. Méthodologie

Dans le présent chapitre, la méthodologie employée afin de répondre aux différents objectifs du mémoire sera présentée. De plus, une brève description de l'échantillon évalué ainsi que l'outil employé à la collecte de données, le *Multidimensional Inventory of Development, Sex, and Agression* (MIDSA)(Augur Enterprises, 2011) seront présentés.

#### 3.1 L'échantillon

Les données du présent mémoire ont été colligées auprès de 282 hommes condamnés d'une infraction sexuelle, rencontrés entre le mois d'avril 1998 et le mois de mars 2000. Ces derniers étaient mis sous garde dans deux différents établissements situés dans les états du Massachusetts (Massachusetts Treatment Center) et du Minnesota (Lino Lakes prison). Au tableau 2 sont présentées les différentes caractéristiques des participants.

**Tableau II. Caractéristiques de l'échantillon**

<b>Caractéristiques</b>	Moyenne	É.-T.
<b>Âge</b>	40,16	10,76
<b>Nb.Arrestations</b>	8,18	17,10
	Effectifs	%
<b>Origine ethnique</b>		
Caucasien	220	78,0
Afro-Américain	33	11,7
Hispanique	4	1,4
Asiatique	1	0,4
Autre	24	8,5
<b>Incarcération dernier crime</b>		
< 1 an	18	6,4
1 à 2 ans	27	9,6
2 à 5 ans	76	27,0
5 à 7 ans	27	9,6
7 à 10 ans	33	11,7
>10 ans	95	33,7

L'âge moyen des répondants était de 40,16 ans (É.-T. = 10,76) et 78 % de l'échantillon était composé d'hommes caucasiens. La moyenne d'arrestations à vie est de 8,18 (É.-T. = 17,1) et 33,7 % des participants ont eu une peine de plus de 10 ans pour le dernier crime commis.

Le MIDSA a été administré sur une base volontaire, et chacun des participants a reçu une compensation financière. Ces derniers ont été assurés de la totale confidentialité de leurs résultats. Les passations se sont déroulées de manière individuelle. Après avoir signé les formulaires de consentement, le participant se voyait désigner un ordinateur et il pouvait débiter le questionnaire.

### **3.2 Le MIDSA**

Le *Multidimensional Inventory of Development, Sex, and Agression* (MIDSA) est un outil d'évaluation psychologique spécialement conçu afin d'être administré auprès de délinquants sexuels juvéniles ou adultes (Knight et Cerce, 1999). Ce questionnaire prend la forme d'une évaluation informatisée, où le participant répond par lui-même via un ordinateur. L'administration du questionnaire prend généralement trois heures, toutefois on y retrouve l'option de pouvoir mettre la passation sur pause si le participant le désire. Le MIDSA a été conçu dans le but d'identifier les domaines importants à cibler lors d'intervention thérapeutique auprès d'individus ayant été sexuellement coercitifs. À la fin du protocole d'évaluation, le MIDSA offre la possibilité de générer un rapport incluant les scores du participant à chacune des échelles, ce qui permet aux cliniciens de mieux orienter leur intervention auprès du contrevenant évalué.

Deux fonctions incluses à l'intérieur du protocole MIDSA assurent un certain contrôle sur la validité des réponses obtenues par les participants. Premièrement, un test de lecture est inclus au début du test, s'assurant ainsi que le participant détient les habiletés de lecture minimale pour comprendre les questions. Deuxièmement, à l'intérieur du protocole se trouve une mesure de rapidité (*speeding-test*). Si les réponses

d'un participant sont plus rapides que la limite permise (2 secondes pour un écran avec une question et 6 secondes pour un écran avec multiples questions), le test donnera un avertissement au participant. Après trois avertissements, le protocole s'arrête et l'assistant de recherche doit intervenir. Cette fonction s'assure que les participants lisent bel et bien chacune des questions.

À l'intérieur de ce questionnaire se retrouve un éventail impressionnant de sujets investigués. En plus de faire l'historique développemental du participant, le MIDSA évalue les difficultés vécues durant l'enfance, l'historique des comportements antisociaux et l'historique des problématiques liées à la consommation de drogues et d'alcool. On y retrouve une section s'intéressant particulièrement à l'historique sexuel du participant. On y évalue, les fantasmes, les attitudes et les comportements sexuels en plus de poser différentes questions en lien avec les caractéristiques des infractions sexuelles commises par le participant. Puisque ce questionnaire est conçu spécialement pour les délinquants sexuels, les concepteurs ont pris en considération les différences liées aux préférences sexuelles et les différences liées aux infractions commises par les participants. Ainsi, un participant n'ayant pas rapporté ni d'infraction ni de préférence sexuelle envers un type de victime n'aura pas à répondre au bloc de questions en lien avec ce type de victime.

### **3.2.1 Les échelles d'évaluations ciblées**

Comme mentionné précédemment, le MIDSA évalue une multitude de facettes de la personnalité et de la vie du répondant. Afin d'effectuer les analyses, quatre échelles d'évaluations ont été ciblées. Celles-ci distinguent entre-elle selon les dimensions évaluées et selon le niveau de transparence des items la composant. Rappelons qu'un item transparent est un item où il est facile de détecter la réponse socialement acceptable à donner (McGrath et al., 1998) et il est reconnu qu'un item transparent est plus susceptible à la dissimulation (Tierney et McCabe, 2001). De plus, afin de pouvoir conduire les analyses sur l'ensemble de l'échantillon, toutes échelles traitant des

préférences sexuelles ou des infractions commises ont été ignorées. Bien que nous soyons conscients que ces échelles présentent les plus hauts niveaux de transparence, pour les besoins du présent mémoire, nous avons choisi d'effectuer nos analyses sur des échelles mesurant quatre dimensions qui peuvent concerner autant les pédophiles que les violeurs.

Nous avons donc choisi d'effectuer nos analyses sur les échelles mesurant l'*Impulsivité*, le *Charme superficiel*, la *Préoccupation sexuelle* et l'*Hostilité envers les femmes*. Le tableau 3 présente les statistiques descriptives de chacune de ces quatre échelles<sup>1</sup>

**Tableau III. Statistiques descriptives des échelles d'évaluations**

<b>Échelles d'évaluation</b>	Moy	Méd	É-T	Min	Max	Alpha
<b>Impulsivité</b>	2,42	2,41	0,99	0	5	0,79
<b>Préoccupation sexuelle</b>	2,63	2,58	1,32	0	5	0,9
<b>Charme superficiel</b>	3,12	3,25	1,1	0	5	0,79
<b>Hostilité envers les femmes</b>	0,94	0,63	0,96	0	4,38	0,88

Le premier domaine ciblé provient de la section évaluant les tendances psychopathiques du participant. L'échelle d'évaluation mesurant l'*Impulsivité* est composée de sept items mesurant la propension du participant à perdre le contrôle de soi. Voici un exemple de question : *J'ai déjà agi sur l'impulsion du moment ou sans penser*

<sup>1</sup> À l'annexe 3 se retrouve un histogramme des résultats obtenus au quatre échelles d'évaluations. Les différences observées dans les distributions des résultats indiquent que la prévalence de chacune de ces problématiques varie énormément. L'échelle évaluant l'hostilité envers les femmes présente une distribution asymétrique positive, il semblerait que cette problématique soit celle la moins rapporté par les participants. On peut toutefois se questionner à savoir si ces distinctions dans les distributions sont dus à une réelle différence dans la prévalence des problématique ou si par le contenu évalué, ces échelles se distinguent tellement quant au niveau de transparence des questions posés que la distribution des résultats en est affectée.

*d'avance*. Un participant présentant un score élevé à cette échelle rapportera avoir perdu fréquemment le contrôle de soi et d'avoir agi sur l'impulsion du moment.

Le second domaine ciblé provient de la section évaluant les attitudes sexuelles du participant. L'échelle d'évaluation mesurant la *Préoccupation sexuelle* du participant est composée de six questions mesurant la fréquence et l'intensité des pensées liées au sexe que le participant peut avoir. Voici un exemple de question : *Lorsque je suis au travail, il m'arrive de penser au sexe*. Un participant présentant un score élevé à cette échelle rapportera avoir plusieurs pensées liées à la sexualité durant une journée, et ce dans différents contextes.

Le troisième domaine ciblé provient de la section évaluant les tendances psychopathiques du participant. L'échelle d'évaluation mesurant le charme superficiel est composée de six questions évaluant la propension du participant à user de charme afin de manipuler les gens de son entourage. Voici un exemple de question : *J'ai déjà manipulé quelqu'un pour qu'il fasse ce que je lui demande*. Un participant présentant un score élevé à cette échelle rapportera avoir à plusieurs reprises manipulé et menti pour atteindre ses objectifs.

Finalement, le dernier domaine ciblé provient de la section du questionnaire évaluant l'hypermasculinité du participant. L'échelle d'évaluation mesurant l'*Hostilité envers les femmes* est composée de huit questions mesurant les attitudes négatives envers les femmes que certains participants peuvent avoir. Voici un exemple de question : *Les femmes victimes de viol le méritent probablement*. Un participant présentant un score élevé à cette échelle rapportera plusieurs attitudes et croyances négatives envers les femmes.

Chacune de ces quatre échelles prend la forme de questions avec choix de réponses de type *likert*, où le participant rapportera à quel point il est d'accord vis-à-vis une affirmation donnée ou la fréquence à laquelle il s'engage dans le comportement énoncé. Ce seront les performances aux questions qui composent ces quatre échelles qui seront

évaluées tout au long du mémoire. La première étape sera de comparer le contenu rapporté à chacune de ces échelles avec les scores obtenus aux échelles de validité.

### 3.2.2 Les échelles de validité

À l'intérieur du MIDSA se trouvent trois échelles mesurant la validité de performance des participants. Ces échelles permettent de cibler les participants qui présentent un trop haut niveau de désirabilité sociale dans le contenu rapporté. Le tableau IV présente les statistiques descriptives de chacune de ces trois mesures.

**Tableau IV. Statistiques descriptives des échelles de validité**

<b>Échelles de validité</b>						
	Moy	Méd	É-T	Min	Max	Alpha
<b>Impression</b>	1,12	1,11	0,65	0	3,67	0,69
<b>Autoduperie</b>	1,31	1,25	0,8	0	4	0,73
<b>Deni sexuel</b>	2,28	2,29	1,05	0	4,67	0,74

Deux de ces échelles ont été construites d'après les travaux de Paulhus (1984) et représentent chacune des dimensions de la désirabilité sociale. La première échelle de validité nommée *Positive image scale* fait référence à la dimension liée à la gestion de l'impression identifiée par Paulhus (1984). À l'intérieur de ce mémoire, cette échelle sera nommée *Gestion de l'Impression* (traduction de *Impression Management*). L'échelle *Gestion de l'impression* est composée de neuf items évaluant si le participant tente de se présenter de manière trop positive face aux autres. Voici un exemple de question : *Je suis toujours poli, même avec les gens qui sont insolents.* . Un score élevé à cette échelle



suggère que le participant manipule consciemment le contenu qu'il rapporte afin de bien paraître.

La seconde échelle nommée *Negative emotion denial scale* fait référence à la dimension de la désirabilité sociale nommée *Self-deception* identifiée par Paulhus (1984). À l'intérieur de ce mémoire, cette échelle sera nommée *Autoduperie* (traduction de *Self-deception*). L'échelle d'*Autoduperie* est composée de neuf questions évaluant la tendance du participant à nier l'influence négative que peuvent avoir ses émotions sur ses comportements. Voici un exemple de question : *Il m'arrive de me fâcher quand les choses ne fonctionnent pas à ma façon* (inversé). Un score élevé à cette échelle reflète la tendance du participant à rejeter l'influence des émotions négatives qu'il peut vivre.

Finalement, la dernière échelle de validité nommée *Sexual denial scale* est composée de six questions qui décrivent certains comportements et pensées liés à la sexualité. À l'intérieur de ce mémoire, cette échelle sera nommée *Déni sexuel*, voici un exemple de question : *Il m'arrive de me masturber* (inversé). Cette échelle permet d'évaluer si le participant nie avoir des comportements ou pensées liés à la sexualité, même si ceux-ci représentent des comportements normaux. Un participant désirant nier en bloc toutes questions liées à la sexualité obtiendra un score élevé à cette échelle. L'ensemble des questions composants les quatre échelles d'évaluations et les trois échelles de validités sont présentés à l'annexe 4.

Les créateurs du MIDSA stipulent que des scores dépassant un écart-type et demi (T scores  $\geq 65$ ) pour chacune de ces échelles devraient être considérés comme étant élevés. En bref, des scores élevés à chacune de ces échelles représentent différents types de patrons de réponses défensifs. Des scores élevés aux échelles *Gestion de l'impression* et d'*Autoduperie* suggèrent que le participant tente de se dépeindre, consciemment ou inconsciemment, d'une manière trop positive. Ces participants seraient donc moins disposés à rapporter certains aspects plus négatifs de leur personnalité à l'intérieur d'autres domaines. Des scores élevés à l'échelle de *Déni sexuel* indiqueraient que le participant a soit très peu de libido ou qu'il n'est pas prêt à

reconnaître quoi que ce soit au sujet de ses comportements et fantasmes sexuels. (MIDSA Clinical Manual, 2011).

Le contenu rapporté par un participant représente l'étape finale d'un processus cognitif complexe. Avant d'arrêter son choix sur sa réponse définitive, ce dernier va entreprendre une période de réflexion, qui se traduira par un temps de latence. La prochaine section présente comment il est possible de préparer et d'utiliser le temps de latence qui découle du processus cognitif lié à l'élaboration d'une réponse.

### **3.2.3 Double standardisation des temps de latence**

À l'intérieur du logiciel d'exploitation du MIDSA se trouve une fonction calculant le temps pris par le participant à répondre à chacune des questions. Afin de rendre ces données utilisables, la méthodologie employée par Holden et Hibbs (1995), afin de contrôler pour les caractéristiques des questions et les caractéristiques des répondants qui peuvent venir affecter les temps de réponse, a été suivie.

Les temps de réponse sont affectés par une multitude de facteurs qui ne sont pas tous en lien avec le processus cognitif entrepris par le participant. Il existe certains facteurs inhérents aux questions (Rogers, 1974) et certains facteurs liés aux caractéristiques du répondant (Holden, Fekken et Coton, 1991) qui doivent être pris en considération. Rogers (1974) a identifié de nombreuses caractéristiques des questions qui viennent influencer les temps de latence. Par exemple, la longueur de la question, l'ambiguïté de celle-ci et le nombre de choix de réponses ont tous été identifiés comme étant des facteurs influençant le temps de réponse. De leurs côtés, Holden et ses collègues (1991) ont identifié différentes caractéristiques individuelles qui viennent influencer les temps de réponse. Par exemple, les habiletés en lecture, les habiletés verbales, l'aisance avec les ordinateurs et la vitesse motrice ont toutes été identifiées comme étant des facteurs pouvant influencer la vitesse de réponse des participants.

Conséquemment, les chercheurs voulant employer les temps de latence comme mesure du traitement cognitif doivent s'assurer de contrôler l'influence des caractéristiques des items et des caractéristiques des participants. Afin d'éviter la confusion entre l'effet des caractéristiques individuelles et des items avec le réel traitement cognitif entrepris par les participants, une double standardisation des temps de latence bruts (Holden et Hibbs, 1995) doit être conduite. Ce processus s'effectue en cinq étapes.

Afin d'effectuer une double standardisation, le tout doit débiter par un resserrement de la distribution des temps de réponse. Les temps de réponse dépassant 40 secondes ont été recodés à 40 secondes et les temps inférieurs à 0,5 seconde ont été recodés à 0,5 seconde. Un temps de réponse dépassant les 40 secondes peut être expliqué soit par la prise d'une pause ou par une distraction extérieure qui est venue perturber le protocole d'évaluation. Les temps de réponse sous le seuil de 0,5 seconde ont été très peu présents, puisqu'à l'intérieur du MIDSA se trouve un procédé détectant les participants qui répondent trop rapidement de façon consécutive. Un participant ayant répondu à trois reprises en moins de deux secondes verra son écran gelé et l'intervention de l'évaluateur sera nécessaire afin de poursuivre le protocole d'évaluation. En somme, tout temps de réponse qui ne semble découler du processus cognitif du participant doit être éliminé.

La deuxième étape consiste à standardiser en score Z les temps bruts de chaque participant. En transformant les temps de chacun des participants de manière à ce qu'ils partagent tous la même moyenne (0) et le même écart-type (1), l'influence des facteurs individuels sur les temps de réponse (telles que les habiletés de lectures et l'aisance avec les ordinateurs) est éliminée. Ainsi, un participant particulièrement rapide à répondre aura le même temps moyen qu'un participant particulièrement lent.

La troisième étape consiste à standardiser les données issues de la deuxième étape de sorte que chaque question présente le même temps moyen (0) et le même écart-type (1). Cette étape permet d'éliminer l'influence des caractéristiques des questions sur les temps de réponse, telle que la longueur de la question. Donc, peu importe le nombre de

mots composant la question, chacune d'entre elles obtiendra le même temps moyen de réponse.

Quatrièmement, tous scores se trouvant à l'extérieur d'un intervalle de plus ou moins trois écarts-types doivent être éliminés. Finalement, les scores issus de la double standardisation ont été transformés en secondes. Nous croyons qu'il est plus facile d'interpréter les temps de latence si ces derniers sont représentés en secondes. Pour ce faire, chacun des scores à chacune des questions a été multiplié par l'écart-type de sa question et additionné à sa moyenne. Puisqu'il s'agit de temps issus d'une double standardisation, les temps négatifs ont été codifiés afin qu'ils deviennent des temps de zéro seconde. À l'annexe 2, vous trouvez une démonstration à l'aide de données fictives du processus de double standardisation.

Les temps de latence issus de la double standardisation ont été utilisés comme marqueur du processus cognitif entrepris par les participants. Plus spécifiquement, les temps de latence de chacune des questions composant les quatre échelles d'évaluation ont été analysés.

### **3.4 Préparation des données pour l'analyse de Rasch**

Afin d'analyser la cohérence des patrons de réponses des participants, logiciel Winsteps© (Linacre, 2004) a été employé. Ce logiciel permet d'estimer les paramètres de difficulté des items, les paramètres des sujets ainsi que les coefficients d'ajustement nécessaires à l'interprétation de la cohérence des patrons de réponses. Puisque l'analyse de Rasch se veut une analyse unidimensionnelle, une base de données pour chacune des échelles d'évaluation a été créée. Un exemple de base permettant l'utilisation du logiciel Winsteps© (Linacre, 2004) est présenté à l'annexe 5. L'analyse de Rasch permet à la fois d'évaluer les coefficients logits des participants, mais aussi ceux des questions. Le logit se veut une unité permettant de comparer les questions entre elles selon leur niveau de difficulté. Les statistiques descriptives issues des analyses de Rasch pour chacune des questions des quatre échelles d'évaluation analysées dans ce mémoire sont présentées à l'annexe 6.

Le coefficient d'ajustement utilisé a été l'*Outfit ZSTD*. Cet indice d'ajustement détermine si le patron de réponse d'un participant respecte le modèle prescrit par l'analyse Rasch. Selon Smith, Shumacker et Bush, (1998) l'intervalle des valeurs standardisées est acceptable s'il se situe entre -2 et 2. Nous avons donc divisé les coefficients d'*Outfit ZSTD* des participants en trois catégories. Nous avons créé une catégorie pour les coefficients sous le seuil de -2 *ZSTD*, une catégorie pour ceux se trouvant entre -2 et 2 *ZSTD* et finalement une autre catégorie pour les coefficients au-dessus de 2 *ZSTD*. Cette catégorisation a été utilisée afin de conduire nos analyses en lien avec l'évaluation des patrons de réponses des participants.

### **3.5 Procédure analytique**

Puisqu'il existe un scepticisme entourant la validité des évaluations autoadministrées employées auprès d'une population de délinquants sexuels, l'objectif du présent mémoire est d'explorer les processus liés à la prestation à un questionnaire autoadministré. Si certains délinquants dissimulent et falsifient leurs réponses, il devrait être possible de capturer certains indices découlant de leur performance qui indiqueraient que ces derniers ont bel et bien falsifié leurs réponses. L'analyse de processus se décompose en trois parties : une analyse de contenu, une analyse des temps de latence et une analyse de cohérence des réponses. Les analyses de contenu et des temps de latence ont été conduites à l'aide logiciel SPSS version (SPSS IBM, New York, É.U.) et l'analyse de cohérence de patrons de réponses à l'aide du logiciel Winstep ©(Linacre, 2004).

#### **3.5.1 Analyse de contenu**

La première stratégie employée sera d'évaluer les scores obtenus aux trois échelles de validité incluses dans le MIDSA. Le contenu rapporté à la fois aux échelles de validité et aux échelles d'évaluation sera le premier indicateur de performance analysé. Il est

possible de croire que la dissimulation aura deux marqueurs présents à l'intérieur de l'analyse de contenu. Premièrement, les participants qui dissimulent obtiendront des pointages aux échelles d'évaluation (e.g. Impulsivité) plus bas que la norme. En effet, les participants ne voulant pas rapporter certaines attitudes ou certains comportements problématiques, afin de bien paraître, auront des scores faibles à ces échelles. Deuxièmement, les participants présentant un haut niveau de désirabilité sociale obtiendront des scores aux échelles de validité plus élevés. Donc, il est possible de croire que le lien unissant le contenu rapporté aux échelles d'évaluation et aux échelles de validité sera proportionnellement négatif. Moins un individu rapportera de problématiques, plus il présentera des scores élevés aux échelles de validité.

Si cette hypothèse se confirme, les participants présentant les plus hauts niveaux de désirabilité sociale, telle que mesurée par les échelles de validité, se distingueront par rapport au reste de l'échantillon quant aux scores moyens rapportés aux échelles de validité. À l'intérieur du manuel d'utilisation du MIDSA, les auteurs avancent que des scores supérieurs à un écart-type et demi ( $T \geq 65$ ) à chacune de ces échelles doivent être considérés comme suspects. Les participants obtenant des scores supérieurs à cette limite seront signalés comme présentant un niveau de désirabilité sociale élevé. Nous effectuerons un test de comparaison de moyennes afin de déterminer si ces derniers obtiennent des scores significativement différents aux échelles d'évaluation.

En bref, ce niveau d'analyse permettra d'observer le lien unissant les scores obtenus aux échelles d'évaluation et ceux obtenus aux échelles de validité. Puisque les échelles de validité ont été conçues dans le but d'évaluer la propension d'un participant à vouloir bien se représenter, des scores trop élevés à ces échelles deviennent donc un indice de potentielle falsification de données. Les analyses subséquentes tenteront de déterminer si ces échelles de validité capturent réellement un désir de bien paraître ou si elles mesurent simplement un trait qui informe peu sur la validité de la performance du participant. Il sera d'abord question d'évaluer si les participants ciblés comme présentant un contenu teinté de désirabilité sociale se distinguent quant au temps requis pour répondre aux questions.

### **3.5.2 Analyse des temps de latence**

La deuxième série d'analyses vise à évaluer la plus-value de l'utilisation des temps de latence à la détection de la dissimulation et la falsification de données. Les temps de réponse représentent un marqueur du processus cognitif entrepris par les participants confrontés à une évaluation autoadministrée.

L'analyse des temps de latence s'effectuera donc en deux étapes. Tout d'abord, les statistiques descriptives du temps moyen pris pour compléter chacune des échelles d'évaluation seront présentées. Comme mentionné précédemment, les temps de latence employés dans ce mémoire sont issus d'un processus de double standardisation. Ce procédé permet de contrôler l'impact sur les temps de latence que peuvent avoir les différences individuelles des participants (e.g habileté de lecture différente) et l'impact que peuvent avoir certaines caractéristiques des questions (e.g. longueur de la question) sur les temps de latence des participants.

La deuxième étape visera à observer le lien unissant le temps pris pour répondre à une question et le score obtenu aux échelles de validité. Plus spécifiquement, cette étape permettra d'évaluer si les participants signalés comme présentant des scores trop élevés aux échelles de désirabilité sociale ( $T \geq 65$ ) effectuent un traitement cognitif différent, marqué par un temps de latence distinct, lorsque vient le temps de répondre aux différentes échelles d'évaluation. Puisque le temps pris pour répondre à une question est un marqueur du processus cognitif entrepris par le participant, et que les participants qui falsifient leurs réponses effectuent un traitement cognitif différent de ceux qui répondent honnêtement, il devrait être possible d'observer des différences quant au temps pris pour répondre aux échelles d'évaluation.

### **3.5.3 Analyse de cohérence des réponses**

La troisième et dernière série d'analyses permettra d'évaluer si les participants présentant des patrons de réponses aberrants, tels que détectés par l'analyse de Rasch,

se distinguent selon le contenu qu'ils ont rapporté et selon le temps pris pour répondre. L'analyse de Rasch permet d'évaluer la cohérence entre le niveau de difficulté d'un item et le niveau d'habiletés d'un participant. Ce modèle statistique permet d'obtenir un indice d'ajustement qui représente le niveau d'adéquation entre le patron de réponses d'un participant et la prescription du modèle. Un mauvais coefficient d'ajustement signifie qu'un participant a offert une performance non représentative de son trait latent (Schmitt et al, 1999). Selon Nagayama Hall (1989), la volonté de dissimuler motiverait certains participants à rapporter très peu de problématiques, mais dans l'optique d'offrir une performance crédible, ces derniers endosseraient quelques questions. Comme il est impossible pour un participant d'évaluer adéquatement le paramètre de difficulté des questions selon le modèle de Rasch, cette stratégie se reflèterait dans ses pointages de cohérence de réponses. Par exemple, si un participant rejette une question présentant un paramètre de difficulté bas, mais endosse une question avec un paramètre élevé (dans l'objectif d'être crédible), il sera détecté par l'analyse de Rasch comme présentant un patron de réponse erroné. Ce processus entrainera donc en un patron de réponses qui présentera un mauvais ajustement (*Underfit*) avec le modèle prescrit.

#### **4. Résultats**

Puisque trois marqueurs renseignant sur les processus liés aux prestations sur questionnaire autoadministré ont été identifiés, le présent chapitre sera divisé en trois parties. Premièrement, les analyses ayant pour objectif d'évaluer le contenu rapporté par les participants seront présentées. À l'aide des échelles de validité, il sera question d'évaluer si la désirabilité sociale, telle que captée par ces mesures, représente bel et bien un marqueur de dissimulation. Deuxièmement, les analyses en liens avec les temps de latence seront effectuées. Il sera question d'évaluer si le temps pris pour répondre à une question est un marqueur de dissimulation qui converge avec l'analyse de contenu. Finalement, des analyses permettant d'apprécier l'apport du modèle de Rasch quant à l'évaluation des patrons de réponses des participants seront présentées.



#### 4.1 Analyse de contenu

À l'intérieur du MIDSA se trouvent trois échelles permettant d'évaluer la validité de performance des participants. L'échelle de *Gestion d'impression* permet d'évaluer si le participant tente de se présenter de manière trop positive. L'échelle d'*Autoduperie* évalue si le participant nie la présence et l'influence de ses affects négatifs sur ses comportements. Finalement, l'échelle mesurant le *Déni sexuel*, évalue si le participant refuse d'avouer ses pensées et comportements sexuels. À l'intérieur du manuel d'utilisation du MIDSA, les auteurs avancent que les participants ayant des pointages d'un écart-type et demi au-dessus de la moyenne ( $T \geq 65$ ) tentent vraisemblablement de modifier leurs réponses afin de mieux paraître.

L'analyse de contenu se divisera donc en deux parties. Tout d'abord, à l'aide d'une analyse de corrélation, le lien unissant les scores aux échelles d'évaluations et les échelles de validités sera testé afin de déterminer s'il abonde dans le sens attendu. C'est-à-dire, si les échelles de validités mesurent une forme de désirabilité sociale, les participants présentant des scores élevés devraient rapporter moins de problématiques aux échelles d'évaluation. Logiquement, si la désirabilité sociale motive les participants à dissimuler leurs réponses, ceux présentant les plus hauts niveaux de désirabilité sociale devraient être ceux rapportant le moins de problématique aux échelles d'évaluation.

Deuxièmement, comme mentionné dans le manuel d'utilisation du MIDSA, les participants présentant des scores élevés aux échelles de validité ( $T \geq 65$ ) devraient se distinguer des autres par rapport au score moyen rapporté aux échelles d'évaluation. L'hypothèse est que les participants présentant les plus hauts niveaux de désirabilité sociale seront ceux avec les performances présentant le plus haut niveau de dissimulation, c'est-à-dire les performances où l'on rapportera le moins de problématiques aux échelles d'évaluation.

#### 4.1.1 Analyse du contenu rapporté

Cette première série d'analyses permet d'évaluer si les échelles de validité captent ce que l'on pourrait estimer être de la dissimulation. Au tableau V sont présentées les analyses de corrélation entre les scores obtenus aux échelles d'évaluation et les scores aux échelles de validité.

**Tableau V. Analyse de corrélation entre le contenu rapporté aux échelles d'évaluation et celui aux échelles de validité**

Échelles d'évaluation	Échelles de validité		
	Impression	Autoduperie	Déni sexuel
Impulsivité	-0,335**	-0,615**	-0,521**
Préoccupations sexuelle	-0,365**	-0,489**	-0,821**
Charme superficiel	-0,507**	-0,568**	-0,575**
Hostilité envers les femmes	-0,260**	-0,465**	-0,453**

\*  $p < 0,05$  ; \*\*  $p < 0,01$

Comme le présente le tableau V, les corrélations entre le contenu rapporté aux échelles d'évaluation et les échelles de validité vont toutes dans le même sens. Plus le contenu d'un participant est teinté de désirabilité sociale, moins ce dernier rapportera de problématiques aux échelles d'évaluation. C'est donc dire qu'un participant n'ayant aucune retenue dans le contenu qu'il rapporte aux échelles d'évaluation aura moins tendance à se présenter de manière trop positive.

Comme on l'indique dans le manuel d'utilisation du MIDSA, les participants pour qui il existerait un doute sur les motivations sous-jacentes à leurs prestations sont ceux présentant les scores les plus élevés aux échelles de validité. On y indique d'ailleurs que

tout pointage au-dessus d'un écart-type et demi ( $T \geq 65$ ) devrait être considéré comme suspect.

La deuxième étape de la démarche analytique consiste à cibler les participants présentant des scores supérieurs à la limite établie ( $T \geq 65$ ) dans le manuel du MIDSA à chacune des échelles de validité. Pour l'échelle *Gestion de l'Impression*, tous participants ayant obtenu un score de 2,09 et plus ont été identifiés comme présentant un score *élevé*. Pour cette échelle, 20 participants (7,1 %) ont obtenu un pointage supérieur à 2,09. Pour l'échelle d'*Autoduperie*, tous participants ayant obtenu un score de 2,51 et plus ont été identifiés comme présentant un score *élevé*. Pour cette échelle, 18 participants (6,4 %) ont obtenu un pointage supérieur à 2,51. Finalement, pour l'échelle de *Déni sexuel*, tous participants ayant obtenu un score de 3,86 et plus ont été identifiés comme présentant un score *élevé*. Pour cette échelle, 20 participants (7,1 %) ont obtenu un pointage supérieur à 3,86. Nous vous présentons au tableau VI une analyse de comparaison de moyennes distinguant les participants selon leurs scores aux échelles de validité.

**Tableau VI. Comparaison des scores moyens obtenus aux échelles d'évaluation selon le niveau de désirabilité sociale**

	Échelles de validité		Échelles d'évaluation									
	Impulsivité		Préoccupation sexuelle		Charme superficiel		Hostilité envers femmes					
	Moy	É-t	Moy	É-t	Moy	É-t	Moy	É-t	t			
<b>Gestion de l'Impression</b>												
Ok	2,47	0,98	3,27***	2,7	1,3	3,27***	3,22	1,06	5,42***	0,96	0,98	NS
Élevé	1,73	0,87		1,72	1,13		1,89	0,97		0,64	0,67	
<b>Autoduperie</b>												
Ok	2,5	0,95	6,03***	2,72	1,29	4,57***	3,22	1,03	6,56***	0,99	0,97	3,31***
Élevé	1,13	0,63		1,31	0,82		1,58	1,03		0,23	0,3	
<b>Déni Sexuel</b>												
Ok	2,49	0,97	4,76***	2,77	1,24	7,12***	3,23	1,06	6,17***	0,98	0,97	2,70**
Élevé	1,43	0,71		0,77	0,61		1,74	0,8		0,39	0,53	

Ok = T<sub>≤65</sub> ; Élevé = T<sub>≥65</sub> \* p<0,05 ; \*\* p<0,01 ; \*\*\*p<0,001

Comme le présente les résultats au tableau VI, les participants ciblés comme présentant des scores trop élevés aux échelles de validité se distinguent du reste de l'échantillon à ce qui a trait au contenu rapporté aux échelles d'évaluation. Les participants ciblés rapportent généralement moins de problématiques aux échelles d'évaluation. Ce fait est cohérent avec l'hypothèse selon laquelle les échelles de validité mesurent le désir de bien paraître des participants. Les participants présentant les plus hauts niveaux de désirabilité sociale, telle que mesurée par les échelles, sont ceux rapportant le moins de problématiques. Néanmoins, il semblerait que les participants présentant des scores élevés à l'échelle de gestion de l'impression ne se distinguent pas du reste de l'échantillon quant au score moyen obtenu à l'échelle d'hostilité envers les femmes ( $t=1,492$ ;  $p>0,05$ ).

À la lumière des résultats issus de l'analyse de contenu, il est possible de statuer sur deux faits. Premièrement, les corrélations entre les scores aux échelles de validité et les scores aux échelles d'évaluation vont dans le sens attendu. Puisque les échelles de désirabilité sociale mesurent le désir de vouloir bien se présenter, il est normal qu'un participant y obtenant un pointage élevé rapporte peu d'attitudes et de comportements problématiques.

Deuxièmement, les échelles de validité permettent de détecter les participants présentant un niveau « trop » élevé de désirabilité sociale. En effet, les participants ciblés par ces échelles se distinguent du reste de l'échantillon quant au contenu qu'ils rapportent aux différentes échelles d'évaluation. La question suivante est de savoir s'il existe une différence dans les marqueurs de performance de ces participants ciblés. Ces derniers rapportent un contenu distinct, mais performant-ils de manière distincte?

## 4.2 Analyse des temps de latence

La présente série d'analyses aura comme objectif d'évaluer si les participants identifiés comme présentant des scores élevés aux échelles de validité se distinguent du reste de l'échantillon quant au temps pris pour répondre aux questions. Puisque le temps de réponse est un marqueur du processus cognitif entrepris par un participant confronté à une question, l'objectif était d'évaluer si les échelles de validité identifient des participants qui présentent un traitement cognitif distinct. Les différentes statistiques descriptives des temps de latence obtenus aux quatre échelles d'évaluation sont présentées au tableau VII.

**Tableau VII. Statistiques descriptives des temps de latence aux quatre échelles d'évaluation**

<b>Échelles d'évaluation</b>	<b>Temps en secondes</b>			
	<b>Moy</b>	<b>Med</b>	<b>É-t</b>	<b>Temps / Question</b>
<b>Impulsivité</b>	55,19	54,23	16,08	7,88 sec
<b>Préoccupation sexuelle</b>	40,11	38,84	11,67	6,89 sec
<b>Charme superficiel</b>	43,73	42,25	11,75	7,28 sec
<b>Hostilité envers les femmes</b>	70,73	79,67	19,33	8,84 sec

Puisque le temps total requis pour compléter ces échelles est d'abord influencé par le nombre d'items composant l'échelle, le temps moyen requis pour répondre à une question a été inclus à la dernière colonne du tableau VII. En moyenne, ce sont les questions de l'échelle évaluant la préoccupation sexuelle qui ont nécessité le moins de temps à répondre ( $\bar{x}$ = 6,89 secondes) alors que ce sont les questions provenant de l'échelle évaluant l'hostilité envers les femmes qui ont nécessité le plus de temps à répondre ( $\bar{x}$ =8,84 secondes). Il est important de se rappeler que ces temps sont issus

d'un processus de double standardisation, ce qui élimine l'impact des caractéristiques des questions sur le temps requis pour leur répondre.

Afin d'évaluer si les participants ciblés par les échelles de validité performant différemment aux échelles d'évaluation, un test de comparaison de moyennes a été effectué afin de déterminer si ces derniers obtiennent des temps de latence distincts du reste de l'échantillon. Les résultats provenant de cette analyse bivariée sont présentés au tableau VIII.

**Tableau VIII. Comparaison du temps moyen pris pour compléter une échelle selon le niveau de désirabilité sociale**

Échelles de validité	Temps moyens pour compléter l'échelle												
	Impulsivité		Préoccupation sexuelle		Charme superficiel		Hostilité envers femmes						
	Moy	É-t	t	Moy	É-t	t	Moy	É-t	t				
<b>Gestion de l'Impression</b>	Ok	55,4	16,32	NS	40,2	11,68	NS	43,35	11,83	-1,988*	70,14	18,97	-1,864*
	Élevé	52,42	12,51		38,82	11,69		48,75	9,55		78,47	22,63	
<b>Autoduperie</b>	Ok	55,15	15,78	NS	40,02	11,56	NS	43,72	11,48	NS	70,96	19,51	NS
	Élevé	55,75	20,61		41,37	13,53		43,99	15,63		67,36	16,60	
<b>Déni Sexuel</b>	Ok	54,61	15,72	-2,223*	40,51	11,54	2,095*	43,77	11,63	NS	71,05	19,39	NS
	Élevé	62,84	19,14		34,86	12,37		43,32	13,51		66,58	18,52	

Ok = T≤65 ; Élevé = T≥65

\* p<0,05 ; \*\* p<0,01 ; \*\*\*p<0,001



D'après les résultats présentés au tableau VIII, les participants ciblés par les échelles de gestion de l'impression se distinguaient quant au temps moyen pris pour répondre aux questions de l'échelle mesurant le charme superficiel ( $t=-1,988$ ,  $p<0,05$ ) et aux questions de l'échelle mesurant l'hostilité envers les femmes ( $t=-1,864$ ,  $p<0,05$ ). À l'échelle mesurant le charme superficiel, les participants ciblés répondent plus lentement ( $\bar{x}=48,75$ ;  $\acute{E}-T=9,55$ ) que le reste de l'échantillon ( $\bar{x}=43,35$ ;  $\acute{E}-T=11,83$ ), ainsi qu'à l'échelle mesurant l'hostilité envers les femmes où leurs temps de réponses sont significativement plus lent ( $\bar{x}=78,47$ ;  $\acute{E}-T=22,63$ ) que pour le reste de l'échantillon ( $\bar{x}=70,14$ ;  $\acute{E}-T=18,97$ ).

Quant aux participants ciblés par l'échelle mesurant le déni sexuel, ces derniers se distinguent quant au temps moyen pris pour répondre aux questions de l'échelle mesurant l'impulsivité ( $t=-2,223$ ;  $p<0,05$ ) ainsi qu'aux questions de l'échelle mesurant la préoccupation sexuelle ( $t=2,095$ ;  $p<0,05$ ). À l'échelle mesurant l'impulsivité, les participants ciblés ont obtenu des temps de réponse plus lents ( $\bar{x}=62,84$ ;  $\acute{E}-T=19,14$ ) que le reste de l'échantillon ( $\bar{x}=54,61$ ;  $\acute{E}-T=15,72$ ). Toutefois, à l'échelle évaluant la préoccupation sexuelle, les participants détectés ont obtenu en moyenne des temps de réponse plus rapides ( $\bar{x}=34,86$ ;  $\acute{E}-T=12,37$ ) que le reste de l'échantillon ( $\bar{x}=40,51$ ;  $\acute{E}-T=11,54$ ). Il semblerait y avoir un processus différentiel selon le domaine exploré par les questions posées.

Finalement, les participants ayant été signalés et ciblés comme présentant un haut niveau d'autoduperie ne se sont pas distingués du reste de l'échantillon quant au temps requis pour répondre à chacune de quatre échelles d'évaluation.

En bref, les participants ciblés par les échelles de gestion de l'impression et l'échelle de déni sexuel, en plus de se distinguer du reste de l'échantillon quant au contenu rapporté, offrent une performance distincte marquée par des temps latence différents. L'analyse des temps de latence offre donc un deuxième indice permettant de juger de la prestation offerte par ces participants.

### 4.3 Analyse de cohérences des réponses

La troisième et dernière série d'analyses vise à évaluer la potentielle utilité d'employer le coefficient d'ajustement *Outfit ZTD* à la détection de patrons aberrants. Ce coefficient, qui provient du modèle d'analyse de Rasch, évalue l'adéquation entre le niveau d'habiletés d'un participant et son patron de réponses. Selon Smith, Shumacker et Bush, (1998) l'intervalle des valeurs standardisées des coefficients d'ajustement est acceptable s'il se situe entre -2 et 2, donc tout coefficient se trouvant à l'extérieur de cet intervalle indique que le participant présente un patron de réponses aberrant.

Nous croyons qu'un participant qui dissimule lors d'une évaluation autoadministrée performera d'une manière atypique. Bien que certains vont nier en bloc chacune des questions, d'autres, dans l'optique d'offrir une performance plausible, vont varier leurs réponses afin de paraître le moins suspect possible. C'est en tentant de préserver un équilibre entre le fait de se présenter positivement et le fait de vouloir offrir une performance plausible que certains participants vont présenter des patrons de réponses aberrants.

D'ailleurs, Zicker et Slitavancent11) avance l'idée que certains participants répondraient de manière honnête aux questions présentant une faible menace à la désirabilité sociale, mais modifieraient leurs réponses aux questions présentant une menace trop grande en terme de désirabilité sociale. Ce processus entraînera donc un patron de réponses qui présentera un mauvais ajustement (*Underfit*) avec le modèle prescrit.

Puisque le modèle de Rasch se veut une analyse unidimensionnelle, les analyses sur chacune des échelles d'évaluation ont été conduites à tour de rôle. La distribution des participants selon leurs coefficients d'ajustement *Outfit ZSTD* à chacune des échelles d'évaluation est présentée au tableau IX.

**Tableau IX. Répartition des participants selon le niveau d'adéquation entre leurs patrons de réponses et la prescription du modèle**

Échelles d'évaluation	Total	Outfit ZSTD					
		ZSTD ≤ -2		ZSTD -2 à 2		ZSTD ≥ 2	
		n	%	n	%	n	%
<b>Impulsivité</b>	280	23	8,21%	244	87,14%	13	4,64%
<b>Préoccupation sexuelle</b>	273	15	5,49%	248	90,84%	10	3,66%
<b>Charme Superficiel</b>	273	13	4,76%	250	91,58%	10	3,66%
<b>Hostilité envers les femmes</b>	228	6	2,63%	219	96,05%	3	1,32%

ZSTD ≤ -2 = Overfit ; ZSTD -2 à 2 = Respect de la prescription du modèle ; ZSTD ≥ 2 = Underfit

La variable *Outfit ZSTD* a été catégorisée selon les recommandations de Smith et ses collègues(1998). Ce sont les patrons de réponses à l'échelle d'impulsivité qui présentent les plus faibles niveaux d'adéquation, où 87,14% des répondants présentent un patron de réponses compris dans l'intervalle.<sup>2</sup> Les prochaines analyses permettront d'observer si les participants présentant un patron de réponses aberrant se distinguent du reste de l'échantillon quant au score obtenu aux échelles de validité ainsi qu'au temps pris pour répondre aux questions. Un test de comparaison de moyennes distinguant les participants selon leurs patrons de réponses à l'échelle d'impulsivité est présenté au tableau X.

<sup>2</sup> Il est important de noter qu'à l'échelle mesurant l'hostilité envers les femmes, 51 participants ont obtenu un score total de zéro, ce qui rend impossible le calcul du coefficient *Outfit*. Malheureusement, ceci représente une limite des analyses de Rasch lorsque vient le temps d'analyser la dissimulation à l'intérieur de questionnaires autoadministrés.

**Tableau X. Test de comparaison de moyennes selon le niveau d'ajustement du patron de réponses du participant à l'échelle d'Impulsivité**

Échelles de validité	Impulsivité Outfit ZSTD			f
	≤-2	-2 à 2	≥2	
<b>Impression</b>	1,17	1,10	1,29	NS
<b>Autoduperie</b>	1,26	1,32	1,11	NS
<b>Déni sexuel</b>	2,24	2,28	2,30	NS
Temps de latence moyen pour compléter l'échelle	56,68 sec	55,51 sec	47,38 sec	NS

Comme l'indiquent les résultats au tableau X, il n'existe aucune différence pour les scores obtenus aux échelles de validités, ni pour le temps requis pour compléter l'échelle lorsque l'on compare les participants selon leurs patrons de réponses. Le même exercice a été conduit pour l'échelle évaluant la préoccupation sexuelle présenté au tableau XI.

**Tableau XI. Test de comparaison de moyennes selon le niveau d'ajustement du patron de réponses du participant à l'échelle de Préoccupation sexuelle**

Échelles de validité	Préoccupation sexuelle Outfit ZSTD			f
	≤-2	-2 à 2	≥2	
<b>Impression</b>	0,82	1,10	1,72	6,202**
<b>Autoduperie</b>	1,42	1,27	1,76	NS
<b>Déni sexuel</b>	2,24	2,26	2,69	NS
Temps de latence moyen pour compléter l'échelle	38,03 sec	40,81 sec	39,07 sec	NS

\*\*p<0,01

Les participants ayant un patron d'ajustement aberrant (*Outfit* ≥ 2) ont obtenu en moyenne des scores à l'échelle de gestion de l'impression plus élevés que le reste de l'échantillon (f=6,202; p<0,01). En effet, ces derniers présentent en moyenne des scores

de 1,72 (É-T=0,36) comparativement à 1,10 (É-T=0,64) et 0,82 (É-T=0,60) pour le reste de l'échantillon. Au tabletretrouvente retrouve les résultats du même exercice conduits sur l'échelle évaluant le charme superficiel.

**Tableau XII. Test de comparaison de moyenne selon le niveau d'ajustement du patron de réponse du participant à l'échelle Charme superficiel**

Échelles de validité	Charme superficiel Outfit ZSTD			
	≤-2	-2 à 2	≥2	f
<b>Impression</b>	1,20	1,10	1,60	3,021*
<b>Autoduperie</b>	1,47	1,30	1,80	NS
<b>Déni sexuel</b>	2,89	2,26	2,79	NS
Temps de latence moyen pour compléter l'échelle	46,20 sec	43,85 sec	46,13 sec	NS

\*p<0,05

Les participants présentant un patron deaberrants aberrant (*Outfit* ≥ 2) ont obtenu en moyenne des scores à l'échelle de gestion de l'impression plus élevés que les participants présentant un patron de réponses adéquat (*Outfit* -2 à2) (f=3,021; p<0,05). En effet, ces derniers présentent en moyenne des scores de 1,60 (É-t=0,48) comparativement à 1,10 (É-t=0,65). Finalement, au tableau XIII, les dernières analyses conduites sur l'échelle évaluant l'hostilité envers les feprésentées présentés.

**Tableau XIII. Test de comparaison de moyenne selon le niveau d'ajustement du patron de réponse du participant à l'échelle Hostilité envers les femmes**

Échelles de validité	Hostilité envers les femmes Outfit ZSTD			
	≤-2	-2 à 2	≥2	f
<b>Impression</b>	1,07	1,08	0,81	NS
<b>Autoduperie</b>	1,10	1,23	0,83	NS
<b>Déni sexuel</b>	1,36	2,21	1,96	NS
Temps de latence moyen pour compléter l'échelle	90,05 sec	73,57 sec	81,93 sec	NS

Comme l'indiquent les résultats au tableau XII, il n'existe aucune différence pour les scores obtenus aux échelles de validités, ni pour le temps requis pour compléter l'échelle lorsque l'on compare les participants selon leurs patrons de réponses. Il est toutefois primordial de prendre en considération que 18% (n=51) de l'échantillon a obtenu un score total de zéro à l'échelle. Ceci pose un problème lorsque l'on veut analyser les patrons de réponses, puisque l'analyse de Rasch ne peut générer de coefficient d'ajustement *Outfit* pour ces participants.

En bref, ce sont les participants présentant des patrons de réponses aberrants aux échelles mesurant la préoccupation sexuelle et le charme superficiel qui ont présenté des scores divergents aux échelles de validité. Aucune distinction n'a pu être établie quant au temps moyen requis pour compléter chacune des échelles.

## **5. Interprétation des résultats**

Dans l'optique de mieux comprendre la manière avec laquelle les délinquants sexuels abordent les questionnaires autoadministrés, l'objectif principal de ce mémoire se voulait d'explorer les processus liés aux prestations sur questionnaire autoadministré. Puisqu'il existe un doute dans l'esprit de plusieurs quant à la véracité des données auto-rapportées par les délinquants sexuels, l'objectif principal du présent mémoire était de détecter différents indices pouvant nous renseigner sur la potentielle falsification de données. À travers les différentes analyses, quelques marqueurs peuvent éveiller les soupçons quant à la validité de certaines performances ont pu être mis en évidence.

### **5.1 Qu'est-ce que nous indique le contenu rapporté ?**

Le premier sous-objectif du mémoire était de déterminer l'impact de la désirabilité sociale sur le contenu rapporté par les participants. Il est important de se rappeler que la désirabilité sociale est motivée par un besoin général d'approbation de l'autre (Crowne & Marlowe, 1960) et qu'elle pousse généralement les individus à s'attribuer des qualités socialement désirables et de rejeter leurs caractéristiques indésirables (Edwards, 1957). Lorsqu'un participant présente un haut niveau de désirabilité sociale, il répondrait non seulement en regard du contenu de la question, mais aussi selon l'implication sociale de sa réponse (Tan et Grace, 2008).

Afin d'évaluer l'impact de ce potentiel biais, les trois échelles de validité contenues dans le MIDSA ont été utilisées. Avant d'établir que ces échelles évaluaient bel et bien un biais de réponse, nous devons nous assurer que ces mesures covariaient dans le sens attendu avec les échelles d'évaluation. Suite aux analyses de corrélations qui ont été effectuées, il a été possible d'observer qu'en effet, plus un participant présente des scores élevés aux échelles de validité, moins ils ont tendance à rapporter de problématiques. Toutefois, un indice pris isolément renseigne peu quant à la performance d'un individu. Le contenu rapporté par un participant ne représente que le résultat final d'un processus cognitif élaboré.

Si l'intention du participant est bel et bien de dissimuler, il sera possible de l'observer dans son contenu rapporté, toutefois il devrait être aussi possible de l'observer à l'intérieur de ses temps de réponses. Le temps de réaction nécessaire à répondre à une question constitue un marqueur du processus cognitif entrepris par le participant (Vendemia et al., 2005) et le fait d'offrir une réponse non authentique requiert un plus grand effort cognitif (McDaniel et Timm, 1990 ; Vrij et Heaven, 1999). Le point de départ était donc de cibler un nombre restreint de participants présentant les pointages le plus élevés à chacune des trois mesures de validité et d'observer les potentielles différences aux temps de latence.

## **5.2 L'utilité des temps de latence**

Le second sous-objectif du mémoire visait à déterminer le lien unissant les temps de latence et les scores aux échelles de désirabilité sociale. L'hypothèse était que les participants ciblés par les échelles de validité, en plus de rapporter moins de problématiques, auraient des temps de latence atypiques. Leurs temps de latence seraient un marqueur d'un processus cognitif propre à une performance non authentique. Trois grands constats découlent de ces résultats.

### **5.2.1 Gérer son image : une tâche qui requiert un effort**

Nous sommes en mesure d'observer que les hommes qui présentent les plus hauts niveaux de gestion d'impression se différencient du reste de l'échantillon quant au temps pris pour répondre aux échelles *Charme superficiel* et *Hostilité envers les femmes*. Faire une bonne gestion de l'image que l'on projette de soi-même est reconnu depuis longtemps comme nécessitant un effort conscient de la part du participant (Paulhus, 1984 ; Sackeim et Gur, 1979). Il est donc normal que l'effort cognitif nécessaire à produire une réponse où le participant se dépeint de manière positive augmente les temps de latence. D'ailleurs, il est reconnu que la désirabilité sociale vient influencer le



processus cognitif associé à l'évaluation de la réponse à donner (Holtgraves, 2004). Plus un participant est préoccupé par la façon dont ses réponses le feront paraître, plus il prendra de temps à évaluer la « bonne » réponse à offrir. Ce résultat permet d'appuyer l'idée selon laquelle le désir de bien paraître, en plus d'avoir des répercussions sur le contenu rapporté, affecte la manière avec laquelle les participants vont répondre.

### **5.2.2 L'autoduperie : un phénomène inconscient**

Le deuxième constat que l'on peut tirer des analyses des temps latence est qu'il existe probablement un aspect inconscient lié au processus d'autoduperie. Cette deuxième dimension de la désirabilité sociale a été longtemps conceptualisée comme étant un processus inconscient et involontaire (Paulhus, 1984 ; Sackeim et Gur, 1979). L'autoduperie fait référence au déni de pensées qui sont psychologiquement menaçantes. Les participants qui en présentent un haut niveau vont généralement nier certaines caractéristiques négatives de leur personnalité, tout en croyant à l'image qu'ils projettent. Ceci constitue une forme de description biaisée de soi-même, mais sans aucune intention de tromper autrui.

Comme il est observé dans les analyses de contenu, les participants présentant un haut niveau d'autoduperie se distinguent du reste de l'échantillon puisqu'ils rapportent significativement moins de problématiques aux échelles d'évaluations. Comme l'autoduperie est marquée par une description de soi biaisée positivement, il est normal qu'elle ait un impact sur le contenu rapporté. Toutefois, lorsque l'on regarde le temps requis pour répondre aux questions, les participants détectés par cette mesure ne présentent aucun indice d'un traitement cognitif qui serait un marqueur d'une déformation volontaire et consciente de leurs réponses. Or, les résultats vont dans le même sens : l'autoduperie serait un processus involontaire et inconscient. Puisque l'on s'intéresse à la falsification des réponses motivées par un désir de vouloir tromper autrui, l'autoduperie ne devrait pas faire partie de nos échelles de détection. Cette échelle semble mesurer une tendance à diminuer certains aspects négatifs de la

personnalité, toutefois la motivation derrière ce style de réponse ne serait pas générée motivé par un désir de bien paraître.

### **5.2.3 Le déni : le réel défi des chercheurs**

En observant les temps latences des participants ayant été ciblés comme présentant un trop haut niveau de déni lié à la sexualité, nous avons été en mesure de faire un constat important. Puisque le déni est reconnu comme étant une forme de défense souvent employée par les délinquants sexuels (Abel et al., 1987 ; McGrath et al., 1998), tout portait à croire que les participants présentant les plus hauts scores au déni performeraient de manière atypique à l'échelle évaluant la préoccupation sexuelle. La forte corrélation entre le niveau de déni et le contenu rapporté à l'échelle mesurant la préoccupation sexuelle n'a rien de surprenant ( $r=-0,821$  ;  $p < 0,001$ ). Ce qui est surprenant par contre, c'est que les participants présentant le plus haut niveau de déni ont répondu de manière plus rapide à cette échelle. Il semblerait que ces participants n'ont fait qu'une évaluation sémantique de la question afin de déterminer rapidement la « bonne » réponse à donner (Hsu et al., 1989). La rapidité avec laquelle ces derniers ont répondu dénote d'une absence d'introspection.

Le désir de vouloir nier certaines caractéristiques en lien avec sa sexualité ne représente pas une forme de désirabilité sociale. Il est raisonnable de penser que le but visé par le déni n'est pas de bien se présenter, mais plutôt d'éviter de s'incriminer. En se préparant à répondre au questionnaire autoadministré et en adaptant son schéma cognitif (Holden et al., 1991), il devient quasiment automatique pour ces participants de nier tout ce qui est en lien avec la sexualité. L'automatisme de ce mécanisme témoigne du fait que le temps requis pour répondre à ces questions et raccourcit raccourci.

Ce constat soulève donc un problème dominant lorsque vient le temps d'évaluer les délinquants sexuels à l'aide d'outils autoadministrés: comment évaluer les domaines les plus « socialement » sensibles si certains vont nier en bloc les questions qui y sont rattachées? De plus, certaines problématiques sont tellement rares (e.g. certaines

paraphilies) qu'il devient aisé pour un participant de dissimuler et de passer sous le radar. Nous croyons qu'une partie de la réponse se trouve dans l'analyse des temps de latence.

### **5.3 Différents domaines, différentes performances**

Comme il l'a été présenté, certains domaines évaluer présente un haut niveau de « sensibilité sociale », où la transparence des questions expose l'évaluateur à la possibilité que le participant dissimule (Tierney et McCabe, 2001). Dans ce mémoire, l'échelle d'évaluation qui répond le mieux au critère de transparence des items est sans doute l'échelle d'*Hostilité envers les femmes*. Cette échelle représente le domaine d'évaluation où l'on retrouve le moins de contenu rapporté. Comment alors distinguer les participants qui ne présentent réellement pas d'hostilité envers les femmes de ceux qui tentent de se présenter de manière positive ? La simple analyse de contenu informe très peu, puisque les participants qui présentent le plus haut niveau de gestion d'impression ne se distinguent pas du reste de l'échantillon quant au contenu rapporté à cette échelle (tableau VI). Sauf qu'à contenu égal, les hommes identifiés comme présentant un trop haut niveau de gestion d'impression prennent significativement plus de temps à répondre à cette échelle.

C'est donc dire qu'ils ont rapporté le même contenu que le reste de l'échantillon, mais d'une façon différente. Ceci représente en soi une information cruciale. Encore une fois, ce n'est pas tant le contenu rapporté qui importe, mais la combinaison du contenu et de la manière avec laquelle celui-ci est livré. Les temps de latence permettent donc de détecter certains participants qui dissimulent, même à l'intérieur d'échelle où la prévalence des problématiques est faible.

### **5.4 En tentant d'être crédibles, certains se mettent plutôt les pieds dans les plats**

Le dernier sous-objectif du présent mémoire constituait à évaluer la potentielle utilité d'employer les coefficients d'ajustement du modèle de Rasch à la détection de patrons de

réponses erronés. L'hypothèse était que certains participants, dans l'optique d'offrir une performance plausible, vont varier leurs réponses afin de paraître le moins suspects que possible. Contrairement aux participants qui nient en bloc, ces derniers vont tenter de naviguer à travers le questionnaire sans trop faire de vagues. Ces participants seront ceux qui voudront trouver un équilibre en le fait de rapporter peu de problématiques, tout en essayant de rester le plus crédibles possible. D'ailleurs, admettre certaines problématiques est nécessaire pour que le participant démontre un certain niveau d'honnêteté. Certains auteurs ont rapporté ce « désir » de crédibilité chez les délinquants (Nagayama Hall, 1989 ; Mills et Kroner, 2005). Après avoir établi cette base de crédibilité, ces derniers peuvent par la suite l'exploiter afin de gérer l'image qu'ils projettent (Mills et Kroner, 2005). Or, cette tentative d'être crédible a été évaluée afin de déterminer si elle aurait des répercussions sur la cohérence du patron de réponse des participants. Les résultats montrent qu'aux échelles mesurant la préoccupation sexuelle et le charme superficiel, les participants présentant un patron de réponse incohérent (*Outfit*  $\geq 2$ ) s'avèrent être ceux qui présentent les plus hauts niveaux de gestion d'impression. Ce résultat corrobore l'idée selon laquelle certains participants vont manipuler leurs réponses afin de bien paraître, tout en essayant de rester le plus crédible possible. En performant de manière atypique dans le but indétectés indétecté, ces participants offrent des patrons de réponses qui ne collent pas à la prescription du modèle.

Néanmoins, la relation entre le coefficient d'ajustement et la performance des participants nécessite davantage d'exploration. Il existe plusieurs types patrons de réponses qui engendrent des coefficients d'ajustement hors normes. Ces patrons de réponses nommés "non-informatifs", peuvent découler de différentes motivations. Par exemple, les patrons de réponses aléatoires peuvent engendrer des d'indices d'*Outfit* supérieurs à 2 ZSTD (Smith, 1996). Les réponses aléatoires démontrent généralement un manque de motivation et d'un désintérêt de la part du participant à répondre au questionnaire autoadministré. Il existe donc plusieurs expossibles possible entraînant le même résultat.

## **Conclusion**

L'objectif du présent mémoire était de mieux comprendre la manière avec laquelle les délinquants sexuels abordent les questionnaires autoadministrés. Puisqu'il existe un doute dans l'esprit de plusieurs quant à la validité des réponses qu'offrent ces hommes, il était primordial d'approfondir nos connaissances sur les potentiels techniques de détection de patrons de réponses. En regard des études qui ont été mené sur la falsification de donnée, trois techniques de détections ont été adoptées dans le présent mémoire. Dans un premier temps, une analyse de contenu a été conduite afin de déterminer si certains participants s'engageaient dans un patron de réponse teinté de désirabilité sociale. Il a été possible d'identifier certains participants tentant de se présenter de manière trop positive, soulevant donc le doute sur la validité de leurs réponses. Dans un deuxième temps, une analyse du temps requis pour répondre aux questions a été conduite. Il a été possible d'observer que certains participants répondaient de manière atypique à certains domaines évalués. Notamment, en répondant plus lentement, certains participants semblent avoir effectués un traitement cognitif différent lorsqu'ils ont été confrontés à certaines questions. Finalement, à l'aide du modèle de Rasch, la cohérence des réponses des participants a été analysée. Comme il a été proposé par différents auteurs, les délinquants ont un souci de crédibilité lorsqu'ils répondent aux questionnaires autoadministrés. Nous avons donc voulu évaluer si certains d'entre eux, en tentant d'être crédible, ont offert des patrons de réponses incohérents. Toutefois, les conclusions que l'on peut tirer de son utilité sont plutôt minces. Certes, le Rasch semble être une technique de détection prometteuse, toutefois de plus amples études l'employant seront nécessaire afin d'appréhender adéquatement son réel potentiel de détection.

Bien que le présent mémoire ait permis de réfléchir sur la question de la désirabilité sociale de façon novatrice, il reposait principalement sur différents axiomes. D'une part, compte tenu du contexte très particulier d'évaluation, nous avons dû assumer que la majorité des participants avaient tendance à falsifier une partie de leurs réponses. Puisque le test aborde des thèmes particulièrement sensibles, nous avons pris pour acquis qu'à un moment ou à un autre, les participants ont été tentés de dissimuler

certaines informations qui ne les mettaient pas en valeur, ou tout simplement nier certains faits. Nous avons assumé que c'est le désir de bien paraître qui influence en partie leurs actions. Bien que de nombreux chercheurs soient sceptiques quant à l'utilité des mesures auto-administrées auprès de délinquants, il nous était impossible d'autre part d'établir à quel point les réponses fournies par les participants étaient véridiques. En d'autres termes, nous n'étions pas en mesure de distinguer le vrai du faux.

Cela étant dit, face à cette difficulté, nous avons proposé une triple stratégie permettant d'informer l'évaluateur sur le caractère plausible des réponses. Les informations recueillies dans le présent mémoire peuvent toutefois être lues de différentes façons. Une première façon est de considérer les réponses vraies comme teintées jusqu'à un certain point, de désirabilité sociale. À la manière de la Théorie de la détection du signal (Swets, 1964), le bruit cache à divers degrés le signal. Nos efforts visent à isoler le signal du bruit, et une évaluation valide des caractéristiques des délinquants passerait par l'élimination du bruit des mesures. Le test valide tirerait le meilleur parti des échelles de mensonges, des patrons de réponses et des temps de latence pour éliminer les questions posant un certain nombre de problèmes, et ainsi offrir à l'évaluateur des mesures pour ainsi dire purifiées.

Deuxièmement, nous avons été en mesure d'identifier les domaines les plus susceptibles de faire objet de dissimulation de la part des participants. La dissimulation et la désirabilité sociale ne sont pas, pour ainsi dire, distribuées au hasard. Certains domaines plus sensibles méritent d'être formulés de manière telle que les participants se sentent moins menacés par le contenu. En modifiant la formulation de ces questions ou en les éliminant simplement, nous pourrions espérer développer une meilleure mesure exempte de dissimulation.

Finalement, l'analogie de la Théorie du signal peut s'arrêter là. Ce que certains appellent du bruit peut être facilement conçu comme une information pertinente. D'ailleurs plusieurs, comme Cronbach et Meehl (1955), rappellent que la performance à un test est bien plus qu'une mesure imparfaite d'un construit latent, comme l'anxiété, l'impulsivité ou la dépression. La performance à un test est elle aussi un comportement dirigé,

comportement produit par le participant dans un contexte très particulier. Le présent mémoire a permis d'étudier ce bruit et a permis, croit-on, de le traiter comme le signal.

## Références bibliographiques

- Abel, G. G., Becker, J. V., Mittelman, M., Cunningham-Rathner, J., Rouleau, J. L., & Murphy, W. D. (1987). Self-reported sex crimes of nonincarcerated paraphiliacs. *Journal of Interpersonal Violence, 2*, 3-25.
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4<sup>th</sup> ed. Text rev.). Washington, DC: Author.
- Augur Enterprises Inc. (2011). Multidimensional Inventory of Development, Sex, and Aggression (MIDSA), Logiciel informatique et manuel d'utilisation disponibles au [www.midsa.us](http://www.midsa.us)
- Bagby, R.M., Rogers, R., Buis, T., (1994). Detecting Malingered and Defensive Responding on the MMPI-2 in a Forensic Inpatient Sample. *Journal of personality assessment, 1994, 62*(2), 191-203
- Beech, A. R., Friendship, C., Erikson, M., et Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse: A Journal of Research and Treatment, 14*, 155-167.
- Bertrand, R., et Blais, J. G., (2004). *Modèles de mesure : l'apport de la théorie des réponses aux items*. Presse de l'Université du Québec.
- Butcher, J.N (2006). *MMPI-2 : A practitioner's guide*, Washington, DC, US: American Psychological Association. xiii 632 pp. doi: [10.1037/11287-000](https://doi.org/10.1037/11287-000)
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.
- Couturat, P.L. (2012). *Troubles de l'acquisition des coordinations à l'école maternelle : validation d'une échelle d'hétéroévaluation*. Thèse de doctorat non publiée. Université Paul Valéry – Montpellier III



- Crowne, D. P., et Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal Consult Psychology*, 24(4), 349-354.
- Cronbach, L.J., Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Curwen, T. (2003). The importance of offense characteristics, victimization history, hostility and social desirability in assessing empathy of male adolescent sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 15, 347-364.
- Dubois, A., Pry, R. (2009). Évaluation de la douleur chez le bébé de moins de 36 mois: application du modèle de Rasch. *Revue européenne de psychologie appliquée*, 59, 163-175
- Ellingson, J. E. (2011). People fake only when they need to fake. In M. Ziegler, C. MacCann, & R.D. Roberts (dir.) *New perspectives on faking in personality assessment* (pp. 3-16). New York : Oxford University Press.
- Fekken, G. C., & Holden, R. R. (1992). Response latency evidence for viewing personality traits as schema indicators. *Journal of Research in Personality*, 26, 103-120.
- Gendreau, P, Irvine, M, et Knight, S. (1973). Evaluating response sets styles on the MMPI with prisoners: Faking good adjustment maladjustment. *Canadian Journal of Behavioral Science*, 5, 183-193.
- Gudjonsson, G. H. (1990). Self-deception and other-deception in forensic assessment. *Personality and Individual Differences*, 11, 219-225.
- Gudjonsson, G. H., & Sigurdsson, J. F. (2000). Differences and similarities between violent offenders and sex offenders. *Child Abuse and Neglect*, 24, 363-372.

- Griffith, R. L., Malm, T., Yoshita, Y., English, A., & Gujar, A. (2006). Applicant faking behavior: Teasing apart the influence of situational variance, cognitive biases, and individual differences. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 149–176). Greenwich, CT: Information Age Publishing.
- Griffith, R.L., Peterson, M.H., (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology*, 1, 308-311.
- Griffith, R. L., & Converse, P. D. (2011). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. MacCann, & R.D. Roberts (dir.) *New perspectives on faking in personality assessment* (pp. 34-52). New York : Oxford University Press.
- Hamon, A. (2000). *Modèle de Rasch et validation de questionnaires de qualité de vie*. Thèse de doctorat non publiée, Université de Bretagne Sud, Vannes, France.
- Hare, R. D. (1985). A checklist for the assessment of psychopathy in criminal populations. In M. H. Ben-Aron, S. J. Hucker, et C. D. Webster (dir.), *Clinical criminology : The assessment and treatment of criminal behavior* (pp. 157-168). Toronto, Ontario, Canada : M & M Graphics.
- Hare, R. D. (1991). *Manual for the Hare Psychopathy Checklist-Revised*. Toronto, ON: Multi-Health Systems.
- Helmes, E., & Holden, R. R. (2003). The construct of social desirability: One or two dimensions? *Personality and Individual Differences*, 34, 1015-1023.
- Holden, R. R., Fekken, G. C., & Cotton, D. H. G. (1991). Assessing psychopathology using structured test-item response latencies. *Psychological Assessment: A Journal of Counselling and Clinical Psychology*, 3, 111-118.

- Holden, R. R., Kroner, D. G., Fekken, G. C., et Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology*, 63, 272-279
- Holden, R. R., (1995). Response latency detection of fakers on personnel tests. *Canadian Journal of Behavioural Science*, 27, 343-355.
- Holden, R.R, Wood, L., et Tomashewski, L. (2001). Do Response Time Limitations Counteract the Effect of Faking on Personality Inventory Validity? *Journal of Personality and Social Psychology*, 81(1), 160-169.
- Holden, R.R., (2007). Socially desirable responding does moderate scale validity both in experimental and non-experimental contexts. *Canadian Journal of Behavioural Science*, 39, 184-201.
- Holden, R.R., Book, A.S., (2011). Faking does distort self-report personality assessment. In M. Ziegler, C. MacCann, & R.D. Roberts (dir.) *New perspectives on faking in personality assessment* (pp. 71-84). New York : Oxford University Press.
- Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, 30, 161-172.
- Hsu, L.M., Santelli, J., et Hsu, J. (1989). Faking Detection Validity and Incremental Validity of Response Latencies to MMPI Subtle and Obvious Items. *Journal of Personality Assessment*, 53(2), 278-295.
- Jackson, D. N., & Messick, S. (1962). Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology*, 65, 285-299.
- Knight, R. A. (1999). *Assessing duplicity in the self-report of sexual offenders*. Présentation effectuée dans le cadre du 18th Annual Meeting of the Association for the Treatment of Sexual Abusers (ATSA), Septembre, Orlando, Florida.

- Kroner, D. G., & Weekes, J. R. (1996). Balanced Inventory of Desirable Responding: Factor structure, reliability, and validity with an offender sample. *Personality and Individual Differences, 21*, 323-333.
- Leary, M.R., Kowalski, R., M., (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*, 34-47
- Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment, 14*, 131-140.
- Linacre, J.M. (2004a). A user's guide and manual to winsteps, ministeps rasch-model computer programs. Chicago : <http://www.winsteps.com>. Consulté en janvier 2014
- Linacre, J.M. (2004). Rasch model estimation: Further topics. *Journal of Applied Measurement, 5*, 95-110.
- Marshall, W.L, et Hall, G.C.N. (1995). The value of the MMPI in deciding forensic issues in accused sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 7*, 205-219.
- McCrae, R. R., & Costa, P. T., Jr. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882-888
- McGrath, M., Cann, S. & Konopasky, R. (1998). New measures of defensiveness, empathy, and cognitive distortions for sexual offenders against children. *Sexual Abuse: A Journal of Research and Treatment, 10*, 25-36
- Meijer, R. R. (2004). Using patterns of summed scores in paper-and-pencil tests and computer-adaptive tests to detect misfitting item score patterns. *Journal of Educational Measurement, 41*(2), 119-136.
- Mills, J. F., & Kroner, D. G. (2005). Impression management and self-report among violent offenders. *Journal of Interpersonal Violence, 21*, 178-192.

- Mueller-Hanson, R.A., Heggstad, E.D., Thornton III, G.C., (2006). Individual differences in impression management: an exploration of the psychological processes underlying faking. *Psychology Science*, 48(3), 288-312.
- Nagayama Hall, G.C. (1989). Self-reported hostility as a function of offense characteristics and response style in a sexual offender population. *Journal of Consulting and Clinical Psychology*, 57, 306-308.
- Nederhof, Anton J. (1985). Methods of coping with social desirability bias: a review. *European Journal of Social Psychology*, 15, 263-280.
- Nugent, P. M., & Kroner, D. G. (1996). Denial, response styles, and admittance of offenses among child molesters and rapists. *Journal of Interpersonal Violence*, 11, 475-486.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660-679.
- Paulhus, D. L. (1984). Two-Component Models of Socially Desirable Responding. *Journal of Personality and Social Psychology*, 46(3), 598-609.
- Paulhus, D. (1986). Self-deception and impression management in test responses. In A. Angleitner et J. S. Wiggins (Eds.), *Personality assessment via questionnaires: Current issues in theory and measurement* (pp. 143-165). New York: Springer-Verlag.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp.17-59). San Diego: Academic Press.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Lawrence Erlbaum

- Paulhus, D. L., & Holden, R. R. (2010). Measuring self-enhancement: From self-report to concrete behavior. In C. R. Agnew, D. E. Carlston, W. G. Graziano, & J. R. Kelly (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 227-246). New York: Oxford University Press.
- Penta, M., Arnould, C., Decruynaere, C. (2005). *Développer et interpréter une échelle de mesure. Applications du modèle de Rasch*. Sprimont, Belgique : Mardaga.
- Pollock, P. H., Quigley, B., Worley, K. O., & Bashford, C. (1997). Feigned mental illness in prisoners referred to forensic mental health services. *Journal of Psychiatric and Mental Health Nursing*, 4, 9-15.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago Press (IL) University of Chicago.
- Reeder, M.C., Ryan, A.M., (2011). Methods for correcting for faking. In M. Ziegler, C. MacCann, & R.D. Roberts (dir.) *New perspectives on faking in personality assessment* (pp. 131-150). New York : Oxford University Press.
- Rogers, T.B., (1974). An analysis of the stages underlying the process of responding to personality items. *Acta Psychologica*, 38, 205-213
- Rogers, T. B.; Kuiper, N. A.; Kirker, W. S, (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, Vol 35(9), 677-688
- Sackeim, H. A., & Gur, R. C. (1979). Self-deception, other-deception, and self-reported psychopathology. *Journal of Consulting and Clinical Psychology*, 47, 213-215.
- Schretlen, D., & Arkowitz, H. (1990). A psychological test battery to detect prison inmates who fake insanity or mental retardation. *Behavioral Sciences and the Law*, 8, 75-84.

- 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.
- Smith, R.M., Schumackher, R.E., Bush, M.J. (1998). Using item mean squares to evaluate fit to Rasch model. *Journal of outcome measurement*, 2(1), 66-78.
- Smith, R.M., (1996), Polytomous mean-square fit statistics. *Rasch measurement transactions*, 10 (3), 516-517.
- Snell, A. F., Sydell, E.J., Lueke, S.B., (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Ressource Management Review*, 9(2), 219-242
- Steffan,J.S., Morgan, R.D., Lee, J., Sellbom, M., (2010). A comparative analysis of MMPI-2 malingering detection models among inmates. *Assessment*, 17(2), 185-196. DOI: 10.1177/1073191109359382
- Sudman, S., Bradburn, N.M., et Schwarz, N. (1996). *Thinking about answers: The application off cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Swets, J.A. (1964). *Signal detection and recognition by human observers*. New York: Wiley.
- Tan, L., et Grace, R. C. (2008). Social desirability and sexual offenders: a review. *Sex Abuse*, 20(1), 61-87. doi: 10.1177/1079063208314820
- Tierney, D.W, et McCabe, M.P. (2001). An Evaluation of Self-Report Measures of Cognitive Distortions and Empathy Among Australian Sex Offenders. *Archives of Sexual Behavior*, 30(5), 495-519.
- Vernon, P. E. (1934). The attitude of the subject in personnality testing. *Journal of Applied Psychology*, 18, 165-177

- Vrij, A., & Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology Crime and Law*, 5, 203-215.
- Walczyk, J.J., Mahoney, K.T., Doverspike, D., Griffith-Ross, D.A., (2003). Cognitive lie detection: Response time and consistency of answers as cue of deception. *Journal of Business Psychology*, 24, 33-49.
- Wasylwiw, O.E., Grossman, L.S., Haywood, T.W., et Cavanaugh Jr, J.L. (1988). The Detection of Malingering in Criminal Forensic Groups: MMPI Validity Scales. *Journal of Personality Assessment*, 52(2), 321-333
- Wiggins, J. S. (1964). Convergences Among Stylistic Response Measures from Objective Personality Tests. *Educational and Psychological Measurement*, 24(3), 551-562.  
doi: 10.1177/001316446402400310
- Wright, B.D, et Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch measurement transaction*, 8, 370.
- Zickar, M. J. et Drasgow, F., (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, 20(1), 71-87
- Zickar, M. J., et Gubby, R. E. (2006). *A history of faking and socially desirable responding on personality tests*. In R. L. Griffith & M. H. Peterson (dir.), *A closer examination of applicant faking behavior* (pp. 21-42). Greenwich, CT: Information Age Publishing.
- Zickar, M. J., et Sliter, K. A. (2011). Searching for unicorns : Item response theory-based solutions to the faking problem. In M. Ziegler, C. MacCann, & R.D. Roberts (dir.) *New perspectives on faking in personality assessment* (pp. 3-16). New York : Oxford University Press.
- Ziegler, M., MacCann, C., et Roberts, R. D. (2011). Faking : knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R.D. Roberts (dir.) *New perspectives on faking in personality assessment* (pp. 3-16). New York : Oxford University Press.



Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review*, 12, 250-264.

## Annexe

Annexe 1. Matrice fictive de scores observés permettant le calcul du coefficient *Logit* en Rasch

Participants	Questions					Rapport	Logit (P)
	Q1	Q2	Q3	Q4	Q5		
Participant 1	1	0	1	1	1	4/1=4	$\log(4)=0,60$
Participant 2	1	1	1	1	0	4/1=4	$\log(4)=0,60$
Participant 3	1	1	1	0	0	3/2=1,5	$\log(1,5)=0,17$
Participant 4	1	1	0	0	0	2/3=0,66	$\log(0,66)=-0,17$
Participant 5	0	1	0	0	0	1/4=0,25	$\log(0,25)=-0,60$
Participant 6	1	0	0	0	0	1/4=0,25	$\log(0,25)=-0,60$
Rapport	1/5=0,20	2/4=0,5	3/3=1	4/2=2	5/2=2,5		
Logit (Q)	$\log(0,20)=-0,70$	$\log(0,5)=-0,30$	$\log(1)=0$	$\log(2)=0,30$	$\log(2,5)=0,40$		

1 = Question réussi    0 = Question échoué

## Annexe 2. Un exemple de double standardisation des temps de latence en 5 étapes

Voici des temps de latences brutes d'un cas fictif afin de démontrer les 5 étapes nécessaires à la double standardisation.

Participant	Questions					Moyenne	É-T
	1	2	3	4	5		
1	6,24	7,20	7,05	8,34	3,43	6,45	1,85
2	5,14	7,52	8,77	7,77	4,15	6,67	1,94
3	7,44	5,54	4,42	6,91	5,44	5,95	1,22
4	5,21	4,82	4,61	6,21	4,11	4,99	0,79
5	8,16	53,00	18,90	10,18	22,10	22,47	18,03
Moyenne	6,44	15,62	8,75	7,88	7,85		
É-T	1,34	20,93	5,95	1,52	8,00		

Étape 1 : La première étape consiste à resserrer la distribution en transformant les valeurs extrêmes. Les questions présentant des temps de latence de plus de 40 secondes sont rapportées à 40 secondes, et les questions ayant des temps de latence inférieurs à 0,5 seconde sont rapportées à 0,5 seconde. Les valeurs à l'extérieur de cette étendue sont considérées comme étant des valeurs aberrantes qui viennent influencer incorrectement les analyses.

Participant	Questions					Moyenne	É-T
	1	2	3	4	5		
1	6,24	7,20	7,05	8,34	3,43	6,45	1,85
2	5,14	7,52	8,77	7,77	4,15	6,67	1,94
3	7,44	5,54	4,42	6,91	5,44	5,95	1,22
4	5,21	4,82	4,61	6,21	4,11	4,99	0,79
5	8,16	40,00	18,90	10,18	22,10	19,87	12,67
Moyenne	6,44	13,02	8,75	7,88	7,85		
É-T	1,34	15,13	5,95	1,52	8,00		

Étape 2 : La deuxième étape consiste à standardiser les temps bruts pour chaque participant. En transformant le temps de moyen de chacun des participants de manière à ce qu'ils aillent une moyenne de 0 et un écart-type de 1 vient éliminer l'influence des facteurs individuels sur les temps de réponse, telles les habiletés de lectures.

Participant	Questions					Moyenne	É-T
	1	2	3	4	5		
1	-0,11	0,40	0,32	1,02	-1,64	0,00	1,00
2	-0,79	0,44	1,08	0,57	-1,30	0,00	1,00
3	1,23	-0,34	-1,26	0,79	-0,42	0,00	1,00
4	0,28	-0,22	-0,48	1,55	-1,12	0,00	1,00
5	-0,92	1,59	-0,08	-0,76	0,18	0,00	1,00
Moyenne	-0,07	0,38	-0,08	0,63	-0,86		
É-T	0,87	0,76	0,88	0,86	0,73		

Étape 3 : La troisième étape consiste à standardiser les données issues de l'étape 2 de sorte que chaque question présente le même temps moyen de 0 et un écart-type de 1. Cette étape permet d'éliminer l'influence des caractéristiques des questions sur les temps de réponse, telle la longueur de la question.

Participant	Questions					Moyenne	É-T
	1	2	3	4	5		
1	-0,06	0,04	0,46	0,45	-1,06	-0,03	0,62
2	-0,83	0,08	1,33	-0,07	-0,60	-0,02	0,84
3	1,48	-0,93	-1,34	0,18	0,60	0,00	1,15
4	0,39	-0,78	-0,46	1,06	-0,36	-0,03	0,74
5	-0,98	1,59	0,01	-1,62	1,42	0,08	1,42
Moyenne	0,00	0,00	0,00	0,00	0,00		
É-T	1,00	1,00	1,00	1,00	1,00		

Étape 4 : Cette étape consiste à transformer les données supérieures à 3 écarts-types et inférieures à 3 écarts-types afin que l'étendue des données reste entre -3 et 3 écarts-types (non applicable à notre exemple).

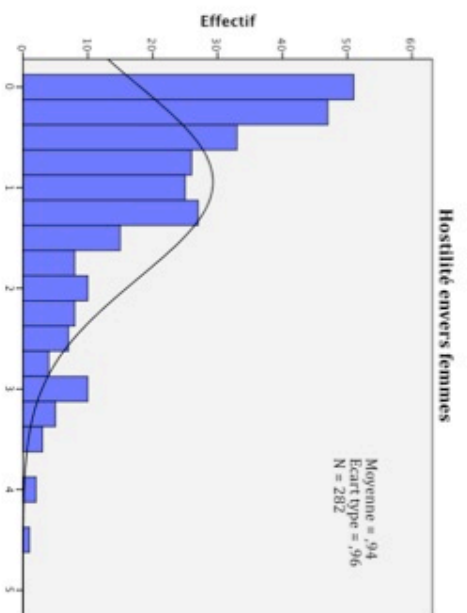
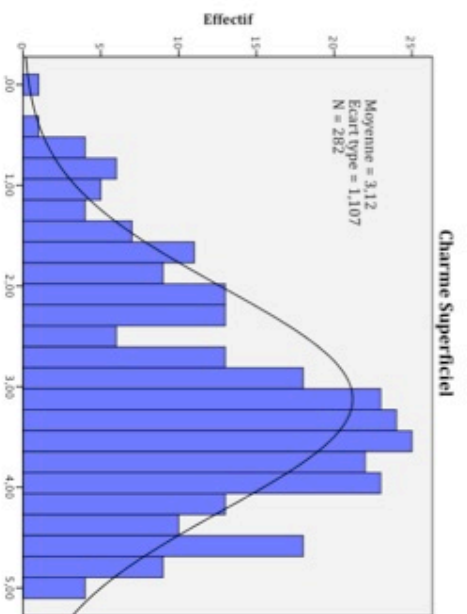
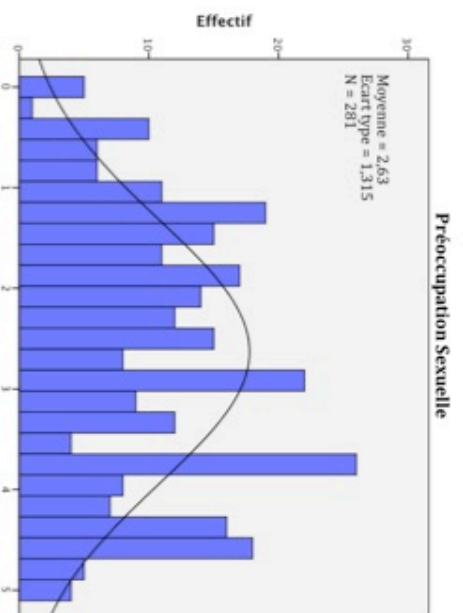
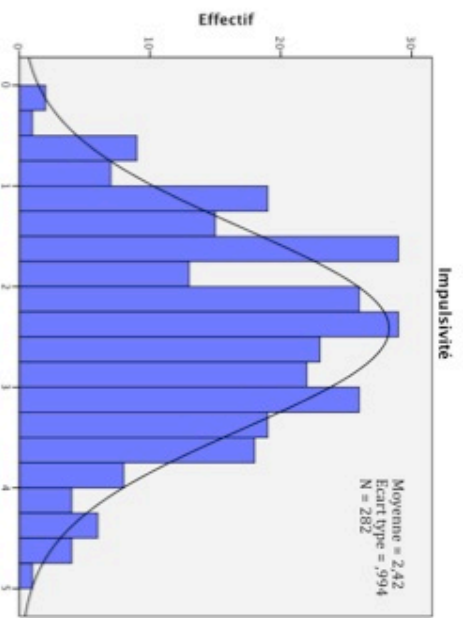
Participant	Questions					Moyenne	É-T
	1	2	3	4	5		
1	6,35	6,52	7,31	7,29	4,49	6,39	1,15
2	5,06	6,83	9,25	6,52	5,50	6,63	1,63
3	7,75	4,82	4,32	6,17	6,68	5,95	1,39
4	5,30	4,38	4,63	5,83	4,71	4,97	0,59
5	7,40	39,98	19,96	-0,68	37,85	20,90	18,03
Moyenne	6,37	12,51	9,09	5,03	11,85		
É-T	1,21	15,40	6,40	3,24	14,56		

Étape 5 : Finalement, pour le mémoire, les données issues de la double standardisation ont été retransformées en secondes. Nous croyons qu'il est plus simple d'analyser et de comprendre les données de latence si celles-ci représentent des secondes. Pour chaque participant, nous avons multiplié le score standardisé par l'écart-type et additionné à la moyenne.

Participant	Questions					Moyenne	É-T
	1	2	3	4	5		
1	6,35	6,52	7,31	7,29	4,49	6,39	1,15
2	5,06	6,83	9,25	6,52	5,50	6,63	1,63
3	7,75	4,82	4,32	6,17	6,68	5,95	1,39
4	5,30	4,38	4,63	5,83	4,71	4,97	0,59
5	7,40	39,98	19,96	0,00	37,85	21,04	17,83
Moyenne	6,37	12,51	9,09	5,16	11,85		
É-T	1,21	15,40	6,40	2,94	14,56		

Puisque ces données proviennent d'une double standardisation, il pouvait y avoir des données en secondes sous zéro. Bien que ce cas de figure était peu fréquent, nous avons recodé ces données à zéro.

### Annexe 3. Distribution des résultats aux quatre échelles d'évaluations



## Annexe 4. Questions contenues dans les échelles d'évaluation et de validité

Échelles d'évaluation		Échelles de validité	
<b>Impulsivité</b>	<p>J'ai déjà agi sur l'impulsion</p> <p>J'ai commi des actes qui m'ont fait sentir coupable</p> <p>J'ai blessé quelqu'un avec mes paroles</p> <p>J'ai perdu le contrôle de moi-même</p> <p>Même si ce n'était de ma faute, j'ai eu des problèmes</p> <p>Mon humeur change subitement</p> <p>J'éprouve des sentiments ébourrés que je ne peux comprendre</p>	<b>Gestion de l'impression</b>	<p>Je suis toujours poli, même avec les gens qui sont insolents</p> <p>Peu importe à qui je parle, j'ai toujours une bonne écoute</p> <p>Je n'ai jamais pris quelque chose qui ne m'appartenait pas</p> <p>J'admets toujours mes erreurs</p> <p>Même s'il est peu probable que je me fasse prendre, j'obéis toujours aux lois</p> <p>Je ne jure jamais</p> <p>Je n'ai jamais lu de livres ou des revues érotiques</p> <p>Je n'ai jamais lu de livres ou des revues érotiques</p> <p>Je n'ai jamais jeté de déchets dans la rue</p> <p>Je n'ai jamais pris avantage de quelqu'un</p>
<b>Préoccupations sexuelles</b>	<p>Au travail je pense souvent au sexe</p> <p>J'ai des rêves sexuels</p> <p>Avant de m'endormir, je pense au sexe</p> <p>Il m'est arrivé de penser sans cesse au sexe</p> <p>J'ai eu des pensées sexuelles</p> <p>Quand je m'ennuie, je pense au sexe</p>	<b>Autoduperie</b>	<p>Lorsque les choses ne fonctionnent pas à ma façon, il m'arrive de me fâcher (1)</p> <p>Lorsque je me fâche, je ne réfléchis plus de la même façon (1)</p> <p>Il m'arrive de ne pas être totalement honnête avec moi-même (1)</p> <p>Je ne comprends pas toujours certaines de mes actions (1)</p> <p>Au lieu de pardonner et d'oublier, il m'arrive de vouloir me venger (1)</p> <p>Quand mes parents me punissaient, ce n'était pas toujours mérité (1)</p> <p>J'ai déjà souhaité que quelque chose de mal arrive à quelqu'un (1)</p> <p>Je n'aime pas lorsque les gens me critiquent (1)</p>
<b>Charme Superficiel</b>	<p>J'ai déjà pris avantage de quelqu'un</p> <p>J'ai déjà manipulé quelqu'un</p> <p>Je n'ai jamais pris avantage de quelqu'un (1)</p> <p>J'ai déjà menti pour arriver à mes buts</p> <p>J'utilise mon charme pour être remarqué</p> <p>Je peux facilement charmer quelqu'un</p>	<b>Déni sexuel</b>	<p>Je pense au sexe</p> <p>Lorsque je pense au sexe, je deviens excité</p> <p>J'ai déjà eu une relation sexuelle avec une autre personne</p> <p>Je me masturbe</p> <p>J'ai fantasmé à l'idée de voir quelqu'un se déshabiller sans qu'elle ne le sache</p> <p>J'ai fantasmé à l'idée de forcer quelqu'un à avoir une relation sexuelle</p>
<b>Hostilité envers les femme</b>	<p>Les femmes victimes de viol le méritent probablement</p> <p>Les femmes saoules sont responsables si l'on prend avantage d'elles</p> <p>Les femmes victimes de viol ont souvent mauvaise réputation</p> <p>Il est moins grave de forcer une prostituée à avoir du sexe</p> <p>Si la femme ne résiste pas assez, cela veut dire qu'elle veut</p> <p>Les hommes ont besoin de sexe tous les jours</p> <p>La plupart des femmes sont froides</p> <p>L'homme devrait être le "boss" dans la relation</p>		

\*Traduction libre

(1) = Pointage inversé

## Annexe 5. Base de donnée permettant l'exploitation du logiciel Winstep© : *Specification control file* pour l'échelle mesurant l'impulsivité

```

&INST
TITLE = "Impulsivité"

:Input Data Format
NAME1 = 1 : column of start of person information
NAMEL = 3 : maximum length of person information
ITEM1 = 4 : column of first item-level response
NI = 7 : number of items = test length
XWIDE = 1 : number of columns per response
PERSON = 0:offender;
ITEM = Item ;
GROUPS=0
CODES = "012345"

&END
My moods change suddenly
I have acted impulsively or without thinking
I do things that make me feel really bad about myself
I have hurt SOMEONE'S feeling by saying something without thinking
Even though I did not want it, I have lost control of myself
I have had frightening feelings that I could not understand
I have gotten in trouble for things that were not my fault

END LABELS
0013220315
0021122101
0034525453
0044445315
0055125303
0064544334
0073404000
0080111001
0090221003
010342333
0112204003
0121112011
0133342314
0152324310
0162224000
0173344411
0182224343
0195554453
0205145444
0214200330
0224554431
0232222000
0242324301
0252025343
0263122301
0272342513
0285204100
0294312304
0300042000
0312404131
0322000333
```



## Annexe 6. Statistiques descriptives issues des analyses de Rasch pour les items des échelles d'évaluation : Items placés en ordre croissant de difficulté

<b>Échelles d'évaluations</b>		<b>Difficulté (logit)</b>	<b>S.E (logit)</b>	<b>Infit ZSTD</b>	<b>Outfit ZSTD</b>	<b>Score observé</b>
<b>Impulsivité</b>						
J'ai blessé quelqu'un avec mes paroles	-0,46	0,06	0,00	-0,30	830	
Mon humeur change subitement	-0,31	0,05	-1,40	-2,10	811	
J'ai déjà agi sur l'impulsion	-0,25	0,05	-1,10	-1,10	762	
J'ai perdu contrôle de moi-même	0,10	0,05	-0,50	-0,60	677	
Même si ce n'était de ma faute, eu des problèmes	0,27	0,05	3,20	4,20	658	
Fais actions qui m'ont fait sentir coupable	0,31	0,05	0,80	0,80	613	
Sentiments ébourrants que je ne peux comprendre	0,34	0,05	-1,10	-1,50	572	
<b>Préoccupations sexuelle</b>						
J'ai eu des pensées sexuelles	-1,74	0,06	-1,80	-1,50	1102	
Avant de m'endormir, je pense au sexe	-0,10	0,06	-1,90	-1,20	780	
Il m'est arrivé de penser constamment au sexe	0,28	0,06	2,50	2,50	654	
J'ai des rêves sexuels	0,32	0,07	4,80	4,80	663	
Quand je m'ennuie, je pense au sexe	0,51	0,06	-3,50	-3,30	606	
Au travail je pense souvent au sexe	0,74	0,06	-0,80	-1,20	561	
<b>Charme superficiel</b>						
Je n'ai jamais pris avantage de quelqu'un	-1,44	0,09	1,90	0,90	940	
J'ai déjà pris avantage de quelqu'un	-0,77	0,08	-3,00	-1,00	876	
J'utilise mon charme pour être remarqué	0,26	0,06	2,00	1,40	638	
J'ai déjà manipulé quelqu'un	0,41	0,07	-2,90	-2,60	605	
J'ai déjà menti pour arriver à mes buts	0,68	0,07	-0,80	-0,80	551	
Je peux facilement charmer quelqu'un	0,87	0,07	2,30	2,60	541	
<b>Hostilité envers les femmes</b>						
Si femme ne résiste pas assez, veut dire qu'elle veut	-0,90	0,07	-0,70	-1,20	462	
Homme devrait être "boss" dans la relation	-0,41	0,06	2,00	1,30	331	
Hommes ont besoin de sexe tous les jours	-0,10	0,07	-0,20	0,30	258	
Moins grave de forcer une prostituée avoir sexe	0,02	0,07	0,90	-0,40	206	
Femmes saoules responsables si prend avantage d'elle	0,02	0,07	-0,60	-0,70	215	
Plupart des femmes sont froides	0,20	0,07	2,50	1,60	234	
Femmes victimes de viols ont mauvaises réputations	0,53	0,07	-0,80	0,00	197	
Femmes victimes de viols le méritent probablement	0,64	0,08	-1,90	-2,30	122	

