Demographic Differences and Language Equivalence for the Conditional Reasoning Test for Workplace Psychopathy in a Large Canadian Sample

Ryan Cook[1] & Nicolas Roulin[1]

Paper "in press" in the Canadian Journal of Behavioural Science

Funding: This research was funded by the Social Sciences and Humanities Research Council (SSHRC) via a Canada Graduate Scholarship – Doctoral (767-2020-1448) awarded to the first author.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval for this data collection was provided by the Saint Mary's University Research Ethics Board (SMU REB File #22-034).

Corresponding author: Ryan Cook, PhD, MSc, BScH.

MS 323, Saint Mary's University

T 902-523-1133

ryan.cook@smu.ca

923 Robie Street | Halifax, N.S. Canada | B3H 3C3

[1] Psychology Department, Saint Mary's University, Halifax, Nova Scotia, Canada

Abstract

Psychopathic traits are linked to counterproductive work behaviours, white-collar crime, and unethical decision-making. However, psychopathy remains underassessed in hiring due to concerns about clinical methods and susceptibility to faking in self-report tools. The Conditional Reasoning Test for Workplace Psychopathy (CRT-WP) offers a promising alternative by measuring implicit psychopathic reasoning through scenario-based reasoning tasks. This study evaluated the CRT-WP's fairness and equivalence across subgroups and languages within a Canadian context. Using data from 6,746 English and 2,182 French-speaking Canadian job applicants, Item Response Theory (IRT) and Differential Item Functioning (DIF) analyses examined potential subgroup differences based on sex, ethnicity, Indigeneity, disability, as well as potential differences between an English and French language version. Results showed minimal DIF for visible minorities, with small or negligible differences in overall scores. Although more DIF was observed across sex and language versions, particularly at extreme trait levels, these differences did not substantively alter overall score interpretation. Indigeneity and disability comparisons were explored with traditional mean comparisons and observed DIF methods due to limited sample size. The CRT-WP demonstrated resistance to subgroup bias and potential for use in diverse hiring contexts, though the French version may require refinement to be precisely comparable to the English version. CRT-WP scores were unexpectedly negatively correlated with cognitive ability, but after removing distractor choices, the relationship weakened. Findings suggest that the CRT-WP may be a fair and faking-resistant tool for assessing non-clinical psychopathic traits in hiring contexts, though continued research is needed to further validate its cross-cultural and linguistic applicability.

Keywords: selection, psychopathy, implicit measurement, Conditional Reasoning

Public Significance Statement

The Conditional Reasoning Test for Workplace Psychopathy (CRT-WP) is an implicit personality measure assessing psychopathic tendencies and rationalizations in a work context. Most CRT-WP items showed negligible differential responding based on sex, ethnicity, Indigeneity, or disability status comparisons. There were 10 out of 22 CRT-WP items which did show moderate to large differential responding between English and French versions. The CRT-WP could be used to assess applicant or employee psychopathic tendencies with limited potential for adverse impact, though caution is warranted when comparing English and French scores.

Psychopathic personality includes insensitivity, risky behaviour, deceitfulness, manipulation, remorselessness, egocentricity, and blame externalization (Babiak, 1995; Smith & Lilienfeld, 2013). Employees and managers with higher levels of psychopathic personality are more likely to engage in counterproductive work behaviours such as theft (Thapar & Brar, 2022), interpersonal abuse (Cook et al., 2024), white-collar crime (Karandikar & Jones, 2025), and unethical decision-making (Stevens et al., 2012). Evaluating candidates' levels of psychopathic personality traits could help avoiding hiring individuals more likely to engage in such behaviors, and thus limit the risks of such negative workplace outcomes. This is particularly relevant for positions with power or authority over the public (i.e., law enforcement), which are coveted by psychopathic individuals (Henley, 2002). Although organizations use personality tests in hiring, psychopathic personality is rarely assessed (Cook et al., 2024; Roth & Klehe, 2024). This is largely because existing measures face one of three problems (see Cook et al., 2024, for a review): (1) clinical assessments are resource-intensive and legally questionable for hiring; (2) measures involving third-party raters (e.g., supervisor or colleague) can lead to overestimating psychopathy and are impractical in most hiring contexts; (3) overt self-report measures are highly susceptible to faking and socially desirable responding (Smith & Lilienfeld, 2013).

In response to calls for better tools for assessing psychopathy in the workplace (O'Boyle et al., 2012; Roth & Klehe, 2024), Cook et al. (2024) developed and validated the Conditional Reasoning Test for Workplace Psychopathy (CRT-WP). This new measure builds on the seminal work of James (1998), who demonstrated that conditional reasoning tests (CRTs) could be adapted to measure implicit personality, by capturing biases, tendencies, and rationalizations behind individuals' choices and actions. It also follows suggestions that CRTs held promise for measuring psychopathy and other dark personality traits at work (O'Boyle et al., 2012). In short, Cook et al. (2024) conducted a series of six studies showing that the CRT-WP was a reliable,

construct-valid (positively correlated with scores on three other measures of psychopathy and negatively correlated with Honesty-Humility), and faking-resistant measure of workplace-specific (and non-clinical) psychopathy that predicts relevant outcomes (e.g., counterproductive behavior, selfish decision making), and they discuss its practicality.

Before the CRT-WP can be used as a selection tool by organizations in a diverse, multi-cultural, and multi-linguistic environment like Canada, it is important to ensure that it does not contribute to adverse impact and can be used in Francophone populations. Cook et al. (2024) found no systematic adverse impact due to sex or ethnicity, but visible minorities scored higher on the CRT-WP in two samples. These preliminary findings highlight the need for further exploring potential demographic differences for the CRT-WP. The first goal of the current study is thus to examine potential sub-group differences in CRT-WP scores (e.g., based on sex or ethnicity). Cook et al.'s (2024) evidence was also based on an English version of the test. We thus compare scores on the English and French versions of the CRT-WP to demonstrate equivalence, similar to prior translations of CRTs for personality (e.g., CRT-A; Galić et al., 2014). We examine subgroup differences and equivalence using Item Response Theory (IRT) analyses, and specifically Differential Item Functioning (DIF). We also examine relationships between CRT-WP scores and cognitive abilities, as a potential indirect source of adverse impact.

Item Response Theory and Differential Item Functioning for the CRT-WP

IRT refers to a set of processes and models used to understand psychological measures (Embretson & Reise, 2000). IRT is not concerned with "overall" test scores but instead assesses how accurate each individual item is at measuring the latent construct (Theta, θ), at all levels of that construct. Theta is a theoretical value estimated based on test-takers' responses on each item, and their likelihood of responding "correctly" to those items. For a personality test with no "correct" answer, as respondents' Theta level increase (e.g., higher psychopathic personality

tendencies) they should be more likely to endorse response options theoretically linked to higher levels of the trait (e.g., more psychopathic responses on the CRT-WP). Moreover, IRT is centered around three parameters: difficulty, discrimination, and guessing. Cook et al. (2024) found the CRT-WP to be best determined using a two-parameter (2-PL) model including both item difficulty and discrimination, but not guessing, in line with previous IRT research focused on other implicit CRTs (DeSimone & James, 2015; Galić et al., 2014; Theriault, 2019).

One of the main applications of IRT analyses is to assess DIF. An item shows differential functioning if individuals from two groups score differently despite having the same level of the construct (Theta). In contrast, the absence of DIF helps demonstrate measurement equivalence between two or more groups or versions of the same scale (Galić et al., 2014), and would provide support for using a scale across contexts (e.g., language) or subgroups (e.g., sex). Previous CRT research indicates that such tests can be vulnerable to DIF. DeSimone and James (2015) showed DIF between student and professional participants for half of the CRT-A (i.e., Aggression) items, although effects sizes were relatively small. Theriault (2019) found DIF between men and women for only one CRT-A item, and no DIF between White and Black participants, although the number of Black participants may have limited the ability to detect differences. Galić et al. (2014) found DIF between U.S. and Croatian language versions for eight items with at least moderate effect sizes. Thus, it is important to examine potential DIF for the CRT-WP.

Subgroup Differences

Human rights legislation in many countries protects demographic subgroups from discrimination in hiring. While there are many protected groups in Canada depending on the federal or provincial jurisdictions, we focus on the four designated groups under the Canadian Employment Equity Act: women (sex), visible minorities (ethnicity), Indigenous peoples, and individuals with a disability. Meaningful DIF exists for the Levenson Self-Report Psychopathy

scale and Triarchic Psychopathy Measure for sex and nationality, respectively (Hauck-Filho & Teixeira, 2014; Shou et al., 2018). A measure of clinical psychopathy demonstrated general invariance between Indigenous and non-Indigenous criminal offenders (Olver et al., 2018), and for prisoners with and without intellectual disabilities (Morrissey et al., 2010). CRT-WP scores did not differ based on sex or ethnicity, but small differences were found for visible minorities, and differences for Indigeneity or disability have not been examined yet (Cook et al., 2024). We thus propose to examine the following:

RQ1: Are CRT-WP items functioning similarly across sub-groups?

English-French Language Equivalence

Two studies have explored differences across languages with the CRT-A scale: IRT analyses showed significant DIF for the English and Croatian versions (Galić et al., 2014), whereas traditional (non-IRT) analyses showed only negligible differences between English and Arabic (Gadelrab, 2019). However, these findings might reflect both language and cultural differences, because they compared U.S. to Croatian or Egyptian samples. CRT-A translations also involved adapting item content. For instance, Galić et al. (2014) changed characters' names, places, or industries to reflect the Croatian culture. Such changes were not made in the French adaptation of the CRT-WP, and the present study examined English- and French-speaking Canadians (i.e., from the same Western country/culture). DIF can thus be interpreted as pure differences between English and French versions.

RQ2: Are CRT-WP items functioning similarly in the English- and French-language versions?

Personality-Based CRTs and Cognitive Ability

Personality-based CRTs (e.g., James, 1998) use the CRT format and test-takers believe that they are completing a test related to intelligence. Examining how strongly personality-based CRTs correlate with cognitive ability is thus relevant. However, in theory, there is not much

room for cognitive ability to influence scores on personality-based CRTs. Indeed, these CRTs include four response options: two clearly illogical options that should be ignored by attentive participants, leaving them with a binary choice between two equally logical options that represent high and low degrees of the personality trait of interest (e.g., psychopathic personality, Cook et al. 2024). Test takers only need enough cognitive ability to eliminate the two illogical response options, and it should not guide their choices between personality-based ones (LeBreton et al., 2007). CRT-A scores and cognitive ability were uncorrelated across five samples (LeBreton et al., 2007), although these samples consisted of university students and thus possible range restriction. Relationships between psychopathy and cognitive ability have been either weak or mixed (e.g., Durand et al., 2023). No research has examined relationships between CRT-WP and cognitive ability scores yet, so we explore this in the current study.

RQ3: What is the relationship between CRT-WP scores and cognitive ability tests scores?

Methods

Participants

Data was collected between March 2022 and August 2023 as part of research collaboration with the partner organization, a branch of the Canadian federal government. All participants were job seekers who completed an online practice cognitive ability test for actively preparing an application for a position in the organization, and could complete the CRT-WP for research purposes. Thus, the collected data was not used for selection. Participants could complete the tests either in English or French. For the English/French version, data was collected from 14,246/7,955 respondents, including 6,746/2,182 usable responses (based on attentiveness and completion). Participant demographics are presented in Table 1. Our samples sizes exceed Hulin et al.'s (1982) recommendation of 500 participants for conducting IRT. However, the sample size was too small to conduct IRT and DIF analyses with confidence for two groups of

interest (Indigenous and persons with disabilities). Analyses based in classical test theory (e.g., mean comparisons) were thus used to compare and describe groups where appropriate, although these results should be interpreted with caution.

Measures

Conditional Reasoning Test for Workplace Psychopathy (CRT-WP). We use the 22-item version of the CRT-WP (Cook et al., 2024) to assess psychopathic tendencies and ways of thinking. Each response is scored +1 for selecting the psychopathic option, -1 for the non-psychopathic option, and 0 for one of the illogical distractors. Participants who select over 25% of illogical responses are considered inattentive and excluded from analyses. CRT-WP scores thus range from -22 to +22, with higher scores representing a higher level of implicit psychopathic personality (see Cook et al., 2024 for more details on the psychometric properties of the CRT-WP). A French language version of the CRT-WP was created via back-translation with three bilingual individuals: one PhD student translating the original items into French, one co-author making slight edits, and another PhD student translating the French version back to English. Differences between the original and back-translated English versions were reviewed, revisions were made to improve clarity, resulting in a 22-item French version. An example CRT-WP item (in both English and French) is included in the Online Supplement. The full list of items is not provided, consistent with previously developed CRTs for implicit personality. The faking-resistance of the measure relies on its content being unknown to test-takers. However, the authors will provide all items and scoring instructions to researchers interested in using the CRT-WP upon request. Reliabilities (KR-20) were .82 for the English and .80 for the French versions.

Cognitive Ability. The proprietary timed cognitive ability test contained 60 multiple choice items which measures *verbal* (e.g., “LETTER is to WORD as SENTENCE is to:” with PARAGRAPH as the correct response), *visuospatial* (e.g., participants view an “unfolded”

version of a figure and select which option would depict it when "folded"), and *problem-solving* (e.g., arithmetic problems and tasks identifying sequences or patterns) types of intelligence. The test demonstrates both construct (e.g., correlations with established cognitive ability measures) and criterion validity (e.g., prediction of training and task performance). The organization used a version of that test for selection purposes, but it offered applicant a practice version of the test (i.e., a parallel form constructed via IRT, and demonstrated as psychometrically similar) which was used in this study. All responses are summed to calculate one overall cognitive ability score.

Procedure

After choosing to complete the tests in English or French and completing an informed consent form, participants read instructions for the CRT-WP. They were told that CRT-WP stands for the Conditional Reasoning Test for Workplace *Problems,* which measures workplace-related reasoning ability. In reality, each problem has two clearly illogical options that should be ignored by attentive participants, leaving them with a binary choice between two equally logical options that represent high and low degrees of psychopathic personality (see Cook et al. 2024). The mild deception is necessary for the CRT to function as intended, so participants are unaware of the true construct being assessed (James, 1998). The 22 CRT-WP items were presented in a randomized order (with the order of the high/low psychopathy response option varying by item). After the CRT-WP, participants completed demographic questions, followed by the verbal, visuospatial, and problem-solving sections of the cognitive ability test, in that order.

Results

Subgroup comparisons were explored using DIF analyses with the English data, due to smaller sample sizes for subgroups in the French data and potential DIF between the language versions. The present study is using the same English-language dataset as Cook et al.'s (2024) Study 4 (but they did not conduct DIF analyses) and they determined that the 2-PL model

showed superior fit indices to the 1-PL ($\Delta\chi 2$ (21) = 642.1, $p < .001$) and the 3-PL model ($\Delta\chi 2$ (21) = 90.4, $p < .001$). All analyses below are thus based on the 2-PL model. DIF analyses were conducted with IRTPro version 6.0 using Bock-Aitkin estimations for sex, ethnicity, and language comparisons. The Bock-Aitkin method implements Marginal Maximum Likelihood (MML) via an Expectation-Maximization (EM) algorithm. This approach integrates over assumed latent ability distributions, allowing for stable item parameter estimation with missing data or complex item formats. All results for these DIF analyses are presented in Table 2.

Overall, seven of 22 items demonstrated omnibus DIF (Total X^2) between men (N = 3916) and women (N = 1738), however, only two of these items showed significant DIF due to item discrimination (X^2a; see Table 2). Of those seven items, women required a lower level of psychopathy to be more likely to select the psychopathic option for five, while the same was true for men for the other two items. Additionally, four more items showed significant DIF due to difficulty, and one due to discrimination, but none of these were significant at the overall level. Only one item showed DIF due to both the discrimination and difficulty parameters. The overall test characteristic curve (Figure S1 of the Online Supplement) shows a 1-point score difference with women scoring higher than men at +3 levels of Theta (i.e., very high psychopathy). NCDIF analyses with Educational Testing Service (ETS) bands were conducted to assess practical significance, and only two items were classified as *moderate* and all others were *negligible*.

Omnibus DIF was present for five items between White (N = 3550) and visible minority (N = 1713) respondents, with one resulting from the discrimination parameter and the other four resulting from item difficulty (see Table 2). Another item showed DIF for the discrimination parameter only, but not the overall item level. For these six items, visible minority participants showed a higher probability to endorse the psychopathic option on three, and White participants did for the other three. This is shown by the test characteristic curves which are near identical

(Figure S2 of the Online Supplement). NCDIF analyses showed only two items having moderate DIF with practical significance while all others were negligible.

To explore item-level and overall CRT-WP score differences for other subgroups, we used Welch's t-tests given the large differences in sample size and possible differences in group variances (see Table S3 of the Online Supplement). We also conducted logistic regression-based DIF analyses and report Nagelkerke's ΔR^2. Overall, all these results should be interpreted with caution, as there are extreme differences in group size. We found significant differences between Indigenous people ($N = 327$) and non-Indigenous ($N = 4936$) for only three items, as well a significant difference in overall CRT-WP scores. Indigenous people ($M = -11.50$, $SD = 5.93$) scored slightly higher than non-Indigenous participants ($M = -12.40$, $SD = 5.61$), albeit with a small effect size ($d = 0.16$). We found significant differences between participants identified as disabled ($N = 144$) versus not ($N = 5119$) for only three items, but no difference in overall CRT-WP score ($d = 0.10$). Items with DIF were not consistent across these comparisons.

The comparison between the English and French versions of the CRT-WP was conducted using the same DIF analyses. Thirteen of 22 items showed significant omnibus DIF between English ($N = 6746$) and French ($N = 2182$) versions (see Table 2). Six of those showed significant DIF based on the discrimination and difficulty parameters, suggesting that they were markedly different for the English and French versions. Scores were higher at higher levels of psychopathy for four items in the French version, but higher for the remaining two in the English version. Overall, the test characteristic curve shows that French version respondents score about one point higher on the CRT-WP at -3 and +3 levels of Theta, indicating that some items should be reviewed for translation, cultural differences, or other potential contributors to DIF (Galić et al., 2014). NCDIF analyses of practical significance showed four items classified as large, six as moderate, and the remaining twelve were negligible (see Table 2). All ten items with large or

moderate practical significance were also flagged by the Bock-Aitkin DIF analyses, indicating that DIF between the language versions is practically relevant and warrants future investigation.

Finally, English/French CRT-WP scores were significantly negatively correlated with overall cognitive ability scores (rs = -.20/-.20, $p < .001$), verbal ability (rs = -.18/-.14, $p < .001$), spatial ability (rs = -.16/-.21, $p < .001$), and problem solving (rs = -.14/-.13, $p < .001$). There was no significant correlation with age, and only a very small correlation with education (r = -.03), eliminating two other potential sources of bias (see Table S5 of the Online Supplement).

Discussion

In this research, we examined potential subgroup differences and versions equivalence for the CRT-WP using IRT analyses (e.g., DIF) and overall score differences. Addressing RQ1, we found very limited evidence of subgroup differences for ethnicity, Indigenous or disability status. For instance, we found DIF for only for 5 of the 22 items when comparing responses of White and visible Minority participants, with just one item showing DIF based on the discrimination parameter. As a result, the overall test curve showed no difference between those two groups in expected CRT-WP scores at any level of the trait (i.e., psychopathy). This suggests that CRT-WP items or scale can be used, and scores can be interpreted the same way, across groups. Similarly, we found score differences (using Welch's t-tests) for only three items for Indigenous vs. non-Indigenous people or disabled and non-disabled persons, with a significant but very small difference for overall test score for the Indigeneity comparison and no overall difference for the disability comparison. These results suggest that the CRT-WP could be used with limited risk of adverse impact against those protected groups in Canada. Yet, results for Indigeneity and disability status should replicated with larger samples to directly examine DIF.

Our results suggest that at very high levels of trait psychopathy, women and participants completing the French version of the CRT-WP may score slightly higher than their comparison

groups (i.e., men and those completing the English version, respectively). The presence of DIF does not mean the CRT-WP cannot be used across groups, but simply that some items or overall scale measures the latent trait *differently* for those groups. DIF was not systematically associated with higher endorsement of the psychopathic response option for men over women, or French over English, but it varied by item. These results are largely consistent with those found for the CRT-A for both sex (Theriault, 2019) and language (Galić et al., 2014). Thus, in response to RQ2, our findings suggest that the French translation of specific CRT-WP items should be re-evaluated, for instance to identify potential cultural differences as the source of DIF (as in Galić et al., 2014). One item (INS3) exhibited significant omnibus DIF for discrimination and difficulty parameters in both sex and language comparisons. Revising or removing this item may lead to more equal expected scores at high levels of psychopathy, where there were about 1-point score differences in both contrasts. A nominal response model may be useful for future tests of DIF within CRTs, analyzing how the illogical response options are selected between groups.

Finally, addressing RQ3, there were negative correlations between CRT-WP scores and cognitive ability (rs = -.20 for both the English and French versions). Theoretically, although the CRT-WP is presented as a test of logic and problem-solving, each scenario should result in a choice between two equally logical answers associated with high vs. low implicit personality rationales, and thus test scores *should* be unrelated to cognitive ability. Our analyses revealed that cognitive ability scores were negatively correlated with the number of illogical response options selected on the CRT-WP (r = -.23/-.22, $p < .001$ for English/French versions). Further, when the number of illogical responses was controlled for, the relationship between CRT-WP and cognitive ability scores decreased to r = -.08/-.11 (English/French). Thus, our findings suggest that test-takers higher on cognitive ability are less likely to select illogical response options. This leads to two implications: On the one hand, pure CRT-WP scores were largely

unrelated to cognitive ability, which reduces risks of adverse impact. On the other hand, the relationship between cognitive ability and illogical responses raises questions about the use of CRTs in less-educated populations, who might find it more difficult to identify these responses. For instance, only 26.3% of our sample was university-educated, much less than previous CRT-WP studies where illogical responses were rarely endorsed (Cook et al., 2024).

Overall, our findings show promising evidence that the CRT-WP could be used by organizations interested in assessing psychopathic or dark personality tendencies with limited potential for adverse impact. Combined with prior evidence about the CRT-WP's strong psychometric properties and faking-resistance (Cook et al., 2024), this answers the call for a reliable, valid, and evidence-based measure of psychopathy for use in workplace contexts (Lilienfeld et al., 2015; O'Boyle et al., 2012). Although the CRT-WP is more time-consuming (estimated 18 minutes) than overt self-report measures, the confidence in the scores not being influenced by faking is likely a positive trade-off, especially for positions of power or leadership. However, future research may explore whether IRT analyses could be used to create an adaptive testing version of the CRT-WP (or other CRTs). Adaptive tests involve first presenting a set of items identified to be of average difficulty, followed by more or less difficult items depending on participant responses. This would result in an assessment with fewer items and shorter completion time. Future research should also examine applicant reactions to the CRT-WP (e.g., fairness perceptions) and how to best debrief applicants about the true nature of the test. Additionally, the current study was conducted in a practice context, and participants completed the CRT-WP knowing that their scores were not used for hiring decisions. Thus, future research should explore the CRT-WP within a high-stakes hiring process. More research providing evidence for the criterion-related validity of the CRT-WP (e.g., task performance or relationship with subordinate well-being) would also be beneficial.

References

Babiak, P. (1995). When psychopaths go to work: A case study of an industrial psychopath. *Applied Psychology: An International Review*, *44*(2), 171–188. https://doi.org/10.1111/j.1464-0597.1995.tb01073.x

Cook, R., Roulin, N., & Joy, K. (2024). Development, Validation, and Faking-Resistance of an Implicit Measure of Psychopathy in the Workplace. *Human Performance*, *37*(5), 245–279. https://doi.org/10.1080/08959285.2024.2422341

DeSimone, J. A., & James, L. R. (2015). An item analysis of the Conditional Reasoning Test of Aggression. *Journal of Applied Psychology*, *100*(6), 1872–1886. https://doi.org/10.1037/apl0000026

Durand, G., Rutten, B. P. F., & Lobbestael, J. (2023). Exploring the Relationship Between Cognitive Abilities and Adaptive Components of Psychopathic Traits. *Sage Open*, *13*(2), 21582440231173823. https://doi.org/10.1177/21582440231173823

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* (pp. xi, 371). Lawrence Erlbaum Associates Publishers.

Gadelrab, H. F. (2019). An investigation of differential relationships of implicit and explicit aggression: Validation of an Arabic version of the Conditional Reasoning Test for Aggression. *Journal of Personality Assessment*, *101*(6), 609–620.

Galić, Z., Scherer, K. T., & LeBreton, J. M. (2014). Validity evidence for a Croatian version of the Conditional Reasoning Test for Aggression. *International Journal of Selection and Assessment*, *22*(4), 343–354. https://doi.org/10.1111/ijsa.12082

Hauck-Filho, N., & Teixeira, M. A. P. (2014). Revisiting the psychometric properties of the Levenson Self-Report Psychopathy Scale. *Journal of Personality Assessment*, *96*(4), 459–464. https://doi.org/10.1080/00223891.2013.865196

Henley, A. G. (2002). *Psychopathy and career interest in a noncriminal population* (2002-95012-130; Issues 12-B). The University of Texas at Austin.

Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement*, *6*(3), 249–260. https://doi.org/10.1177/014662168200600301

James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods*, *1*(2), 131–163. https://doi.org/10.1177/109442819812001

Karandikar, S., & Jones, D. N. (2025). From Embezzlement to Antitrust: White Collar Crime Preferences of the Dark Triad. *Journal of White Collar and Corporate Crime*, *6*(2), 88–97. https://doi.org/10.1177/2631309X231210846

LeBreton, J. M., Barksdale, C. D., Robin, J., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *Journal of Applied Psychology*, *92*(1), 1–16. https://doi.org/10.1037/0021-9010.92.1.1

Lilienfeld, S. O., Watts, A. L., & Smith, S. F. (2015). Successful psychopathy: A scientific status report. *Current Directions in Psychological Science*, *24*(4), 298–303. https://doi.org/10.1177/0963721415580297

Morrissey, C., Cooke, D., Michie, C., Hollin, C., Hogue, T., Lindsay, W. R., & Taylor, J. L. (2010). Structural, item, and test generalizability of the Psychopathy Checklist—Revised to offenders with intellectual disabilities. *Assessment*, *17*(1), 16–29. https://doi.org/10.1177/1073191109344052

O'Boyle, E. H. Jr., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behavior: A social exchange perspective. *Journal of Applied Psychology*, *97*(3), 557–579. https://doi.org/10.1037/a0025679

Olver, M. E., Neumann, C. S., Sewall, L. A., Lewis, K., Hare, R. D., & Wong, S. C. P. (2018). A comprehensive examination of the psychometric properties of the Hare Psychopathy Checklist-Revised in a Canadian multisite sample of indigenous and non-indigenous offenders. *Psychological Assessment*, *30*(6), 779–792. https://doi.org/10.1037/pas0000533

Roth, L., & Klehe, U.-C. (2024). The enemy within one's own ranks: Meta-analysis on the effects of psychopathy on workplace-related behavior. *Journal of Applied Psychology*, *Advance online publication*. https://doi.org/10.1037/apl0001248

Shou, Y., Sellbom, M., & Xu, J. (2018). Psychometric properties of the Triarchic Psychopathy Measure: An item response theory approach. *Personality Disorders: Theory, Research, and Treatment*, *9*(3), 217–227. https://doi.org/10.1037/per0000241

Smith, S. F., & Lilienfeld, S. O. (2013). Psychopathy in the workplace: The knowns and unknowns. *Aggression and Violent Behavior*, *18*(2), 204–218. https://doi.org/10.1016/j.avb.2012.11.007

Stevens, G. W., Deuling, J. K., & Armenakis, A. A. (2012). Successful psychopaths: Are they unethical decision-makers and why? *Journal of Business Ethics*, *105*(2), 139–149. https://doi.org/10.1007/s10551-011-0963-1

Thapar, R., & Brar, S. (2022). A Study of Counterproductive Work Behavior in Relation to Personality amongst Police Personnel. *International Journal of Education and Management Studies*, *12*(2), 118–126. ProQuest Central Premium

Theriault, C. M. (2019). *An analysis of statistically & practically significant dif in the CRT-A by gender & race* [Pennsylvania State University]. https://etda.libraries.psu.edu/files/final_submissions/20644

Table 1.

Demographic Information for the English- and French-speaking Samples

Demographic characteristic	*English*	*French*
Sex		
Men	58.0%	62.9%
Women	25.8%	24.1%
Other	2.5%	1.7%
Missing	13.7%	11.3%
Race/Ethnicity		
White/Caucasian	50.5%	61.2%
Visible Minority	25.4%	21.4%
Indigenous	4.8%	1.5%
Person with a Disability	2.1%	0.2%

Note. These race, ethnicity, and disability status demographics were presented so that participants could select all (or none) that apply. As a result, 22.0%/16.7% of participants were coded as missing – e.g., selected nothing or "prefer not to say").

Table 2.

DIF for 22 CRT-WP Items Across Sex, Ethnicity, and Language Comparisons

	Men-Women					*White-Visible Minority*					*English-French Language*				
Item	*Total X^2*	*X^2a*	*$X^2c\|a$*	*NCDIF*	*ETS Band*	*Total X^2*	*X^2a*	*$X^2c\|a$*	*NCDIF*	*ETS Band*	*Total X^2*	*X^2a*	*$X^2c\|a$*	*NCDIF*	*ETS Band*
1. EXT2	8.1	0.1	8.1*	.000	Negligible	8.4	8.2*	0.2	.007	Moderate	0.4	0.4	0.0	.001	Negligible
2. EXT9	2.8	1.0	1.8	.001	Negligible	0.9	0.4	0.5	.000	Negligible	9.0	0.0	9.0*	.000	Negligible
3. EXT10	17.0**	2.3	14.7**	.002	Negligible	21.7**	0.0	21.7**	.002	Negligible	51.2**	5.8	45.4**	.130	Large
4. CI1	12.1	3.8	8.3	.001	Negligible	6.0	2.0	4.0	.000	Negligible	1.7	0.9	0.8	.001	Negligible
5. CI6	23.3**	1.5	21.8**	.004	Negligible	7.1	2.6	4.5	.000	Negligible	0.4	0.1	0.3	.002	Negligible
6. CI11	6.1	3.8	2.3	.002	Negligible	7.7	1.3	6.4	.004	Negligible	112.2**	0.0	112.2**	.010	Moderate
7. SS1	1.6	0.3	1.3	.000	Negligible	11.5*	6.6*	4.9	.004	Negligible	22.8**	7.1*	15.7**	.006	Moderate
8. SS2	11.3*	7.2*	4.1	.003	Negligible	0.3	0.1	0.2	.000	Negligible	4.7	3.3	1.4	.000	Negligible
9. SS3	4.8	0.8	4.0	.002	Negligible	1.2	0.1	1.1	.002	Negligible	53.1**	2.3	50.8**	.010	Moderate
10. SS9	0.3	0.0	0.3	.000	Negligible	4.3	0.1	4.2	.002	Negligible	12.2*	0.2	12.0**	.006	Negligible
11. FLN1	11.1*	0.1	11**	.002	Negligible	1.0	0.4	0.6	.001	Negligible	2.3	2.1	0.2	.001	Negligible
12. FLN2	1.1	0.0	1.1	.000	Negligible	3.0	0.0	3.0	.001	Negligible	0.3	0.0	0.3	.002	Negligible
13. FLN10	7.4	6.9*	0.6	.006	Moderate	70.7**	1.2	69.5**	.007	Moderate	19.5**	2.1	17.4**	.007	Moderate
14. RSI1	8.6	1.2	7.4	.001	Negligible	0.8	0.7	0.0	.000	Negligible	6.9	0.6	6.3	.000	Negligible
15. RSI6	1.8	0.3	1.5	.001	Negligible	2.7	0.0	2.7	.000	Negligible	23.8**	8.0*	15.8**	.002	Negligible
16. RSI7	12.1*	0.3	11.8*	.005	Negligible	0.5	0.4	0.1	.000	Negligible	49.2**	15.5**	33.6**	.012	Moderate
17. RSI8	2.0	0.1	2.0	.000	Negligible	3.7	0.6	3.2	.000	Negligible	3.0	0.2	2.7	.004	Negligible
18. RSI10	8.7	1.0	7.7*	.006	Moderate	8.3	3.9	4.3	.003	Negligible	14.1*	1.1	13.0**	.002	Negligible
19. RSI12	7.1	0.0	7.1*	.000	Negligible	5.0	1.6	3.4	.001	Negligible	11.7*	1.7	10.0*	.008	Moderate
20. INS2	0.6	0.3	0.3	.000	Negligible	5.0	0.1	4.9	.002	Negligible	104.5**	15.5**	89.1**	.028	Large
21. INS3	38.0**	6.6*	31.4**	.006	Negligible	26.2**	2.2	24.0**	.004	Negligible	526.0**	48.4**	477.6**	.100	Large
22. INS9	7.5	0.1	7.4	.003	Negligible	57.6**	2.0	55.6**	.006	Negligible	121.3**	12.8**	108.5**	.037	Large

Note. Total X^2 = Omnibus DIF, X^2a = DIF due to discrimination parameter, $X^2c|a$ = remaining DIF with discrimination parameter removed. NCDIF and ETS Band based on practical significance analyses using the difR package in R software. ETS = Educational Testing Service, bands are Negligible (NCDIF <.006), Moderate (.006 ≤ NCDIF < .024), and Large (NCDIF ≥ .024). * $p < .01$; ** $p < .001$.

Table S1

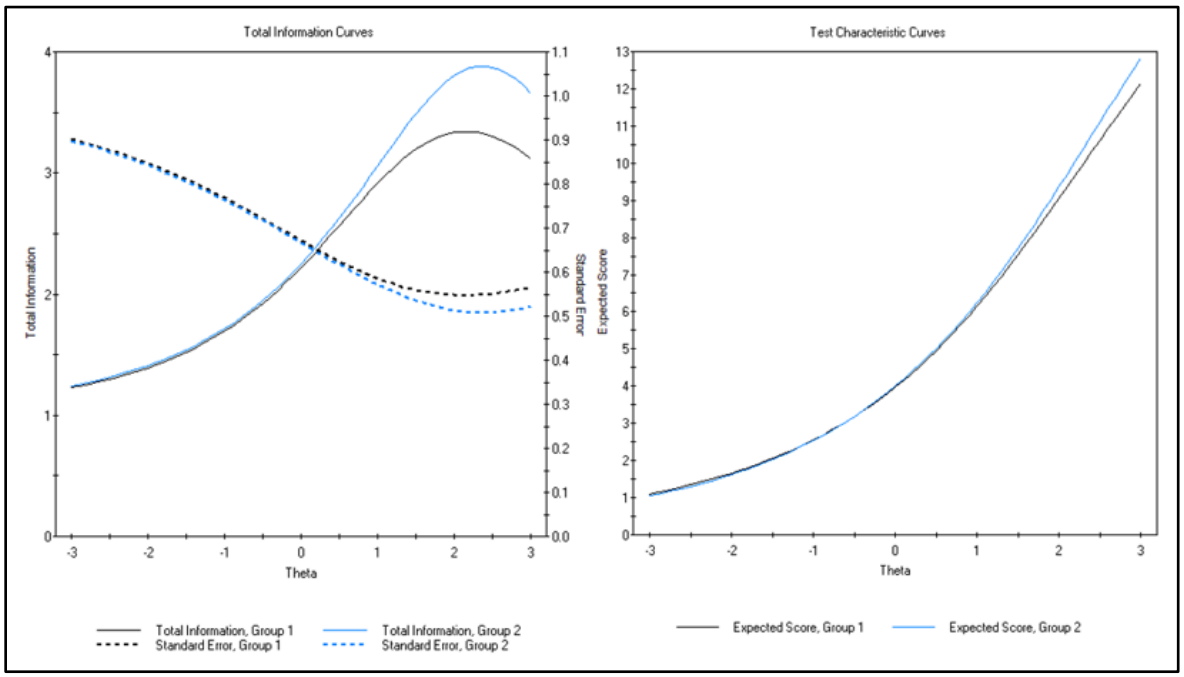
DIF statistics for Men-Women Comparisons

Item	*Total* X^2	*p*	X^2a	*p*	$X^2c\|a$	*p*
1. EXT2	8.1	0.017	0.1	0.804	8.1	0.005*
2. EXT9	2.8	0.245	1	0.323	1.8	0.176
3. EXT10	17	< 0.001*	2.3	0.134	14.7	< 0.001*
4. CI1	12.1	0.002*	3.8	0.052	8.3	0.004*
5. CI6	23.3	< 0.001*	1.5	0.218	21.8	< 0.001*
6. CI11	6.1	0.047	3.8	0.050	2.3	0.132
7. SS1	1.6	0.440	0.3	0.562	1.3	0.253
8. SS2	11.3	0.004*	7.2	0.007*	4.1	0.042
9. SS3	4.8	0.090	0.8	0.376	4	0.045
10. SS9	0.3	0.849	0	0.907	0.3	0.576
11. FLN1	11.1	0.004*	0.1	0.716	11	< 0.001*
12. FLN2	1.1	0.582	0	0.997	1.1	0.299
13. FLN10	7.4	0.024	6.9	0.009*	0.6	0.459
14. RSI1	8.6	0.014	1.2	0.276	7.4	0.007*
15. RSI6	1.8	0.414	0.3	0.587	1.5	0.226
16. RSI7	12.1	0.002*	0.3	0.563	11.8	0.001*
17. RSI8	2	0.363	0.1	0.777	2	0.163
18. RSI10	8.7	0.013	1	0.329	7.7	0.006*
19. RSI12	7.1	0.029	0	0.837	7.1	0.008*
20. INS2	0.6	0.724	0.3	0.561	0.3	0.578
21. INS3	38	< 0.001*	6.6	0.010*	31.4	< 0.001*
22. INS9	7.5	0.024	0.1	0.748	7.4	0.006*

Note. Total X^2 = Omnibus DIF, X^2a = DIF due to discrimination parameter, $X^2c|a$ = remaining DIF with discrimination parameter removed, indicating DIF due to difficulty parameter. * $p < .01$.

Figure S1

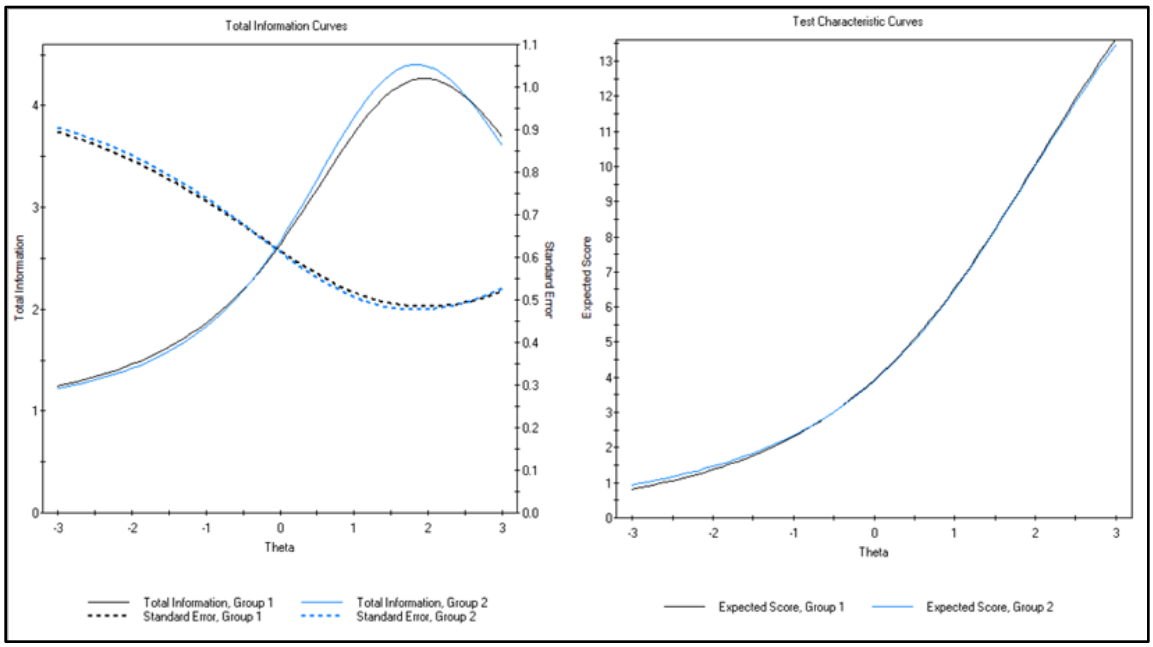
Male-Female DIF Total Information Curves and Test Characteristic Curve



Note. Male = Group 1 and black lines. Female = Group 2 and blue lines. CRT-WP information is higher for females from +1 to +3 levels of theta, and females are expected to score around 1 point higher at +3 theta.

Figure S2

White-Visible Minority DIF Total Information Curves and Test Characteristic Curves



Note. White/Caucasian = Group 1, black lines. Visible minority = Group 2, blue lines. As evidenced by both graphs, there is little difference in information and no difference in expected scores.

Table S2

DIF statistics for White-Visible Minority Comparisons

Item	*Total* X^2	*p*	X^2a	*p*	$X^2c\|a$	*p*
1. EXT2	8.4	.015	8.2	.004*	0.2	.619
2. EXT9	0.9	.635	0.4	.504	0.5	.496
3. EXT10	21.7	< .001*	0	.930	21.7	< .001*
4. CI1	6	.051	2	.162	4	.046
5. CI6	7.1	.029	2.6	.109	4.5	.034
6. CI11	7.7	.022	1.3	.254	6.4	.012
7. SS1	11.5	.003*	6.6	.010*	4.9	.026
8. SS2	0.3	.869	0.1	.799	0.2	.642
9. SS3	1.2	.559	0.1	.802	1.1	.294
10. SS9	4.3	.117	0.1	.799	4.2	.040
11. FLN1	1	.607	0.4	.517	0.6	.447
12. FLN2	3	.222	0	.857	3	.085
13. FLN10	70.7	< .001*	1.2	.271	69.5	< .001*
14. RSI1	0.8	.686	0.7	.387	0	.943
15. RSI6	2.7	.260	0	.845	2.7	.104
16. RSI7	0.5	.769	0.4	.505	0.1	.775
17. RSI8	3.7	.156	0.6	.455	3.2	.075
18. RSI10	8.3	.016	3.9	.047	4.3	.037
19. RSI12	5	.084	1.6	.214	3.4	.065
20. INS2	5	.081	0.1	.760	4.9	.026
21. INS3	26.2	< .001*	2.2	.139	24	< .001*
22. INS9	57.6	< .001*	2	.156	55.6	< .001*

Note. Total X^2 = Omnibus DIF, X^2a = DIF due to discrimination parameter, $X^2c|a$ = remaining DIF with discrimination parameter removed, indicating DIF due to difficulty parameter. * $p < .01$.

Table S3

Welch's t-Tests for Indigenous to Non-Indigenous and Disabled to Non-Disabled Comparisons

	Indigenous vs Non-Indigenous					*Disability vs No Disability*				
Item	*t*	*df*	*p*	*Nagelkerke's* ΔR^2	*DIF Effect Size*	*t*	*df*	*p*	*Nagelkerke's* ΔR^2	*DIF Effect Size*
1. EXT2	4.423	1, 338.20	.036*	.001	Negligible	.004	1, 143.78	.952	.000	Negligible
2. EXT9	.330	1, 372.77	.566	.001	Negligible	2.674	1, 145.31	.104	.006	Negligible
3. EXT10	.341	1, 341.05	.560	.000	Negligible	.021	1, 144.03	.886	.004	Negligible
4. CI1	2.136	1, 350.97	.145	.000	Negligible	3.775	1, 155.67	.054	.002	Negligible
5. CI6	5.162	1, 321.25	.024*	.001	Negligible	2.010	1, 137.65	.159	.002	Negligible
6. CI11	2.156	1, 357.80	.143	.000	Negligible	4.470	1, 146.26	.036*	.002	Negligible
7. SS1	1.717	1, 365.88	.191	.000	Negligible	2.297	1, 149.05	.132	.001	Negligible
8. SS2	3.539	1, 355.94	.061	.001	Negligible	.738	1, 148.73	.392	.003	Negligible
9. SS3	1.463	1, 356.44	.227	.001	Negligible	1.555	1, 152.59	.214	.000	Negligible
10. SS9	.817	1, 346.15	.367	.000	Negligible	.019	1, 146.13	.891	.002	Negligible
11. FLN1	.326	1, 347.26	.568	.001	Negligible	8.169	1, 151.20	.005*	.003	Negligible
12. FLN2	.111	1, 350.50	.739	.001	Negligible	5.523	1, 150.49	.020*	.003	Negligible
13. FLN10	.281	1, 355.81	.596	.001	Negligible	3.258	1, 150.52	.073	.003	Negligible
14. RSI1	.020	1, 366.53	.887	.002	Negligible	.372	1, 148.06	.543	.002	Negligible
15. RSI6	.274	1, 372.99	.601	.001	Negligible	.190	1, 151.95	.663	.000	Negligible
16. RSI7	.017	1, 352.89	.896	.001	Negligible	.102	1, 148.95	.750	.000	Negligible
17. RSI8	.230	1, 346.21	.632	.000	Negligible	2.185	1, 147.79	.141	.001	Negligible
18. RSI10	.658	1, 361.06	.418	.000	Negligible	.003	1, 145.66	.958	.002	Negligible
19. RSI12	8.238	1, 329.56	.004**	.002	Negligible	.882	1, 147.37	.349	.001	Negligible
20. INS2	1.259	1, 338.02	.263	.000	Negligible	.433	1, 138.29	.512	.002	Negligible
21. INS3	.517	1, 353.07	.473	.000	Negligible	1.303	1, 144.85	.255	.001	Negligible
22. INS9	1.439	1, 354.79	.231	.000	Negligible	2.182	1, 142.12	.142	.003	Negligible
Total	7.127	1, 365.75	.008**			1.35	1, 152.04	.247		

Note. Total = overall CRT-WP score. * $p < .05$, ** $p < .01$. Nagelkerke's ΔR^2 based on DIF analyses using logistic regression. DIF Effect Size Band based on practical significance analyses using the ranges taken from Jodoin & Gierl (2001). Negligible ($\Delta R^2 < .035$), Moderate ($.035 \leq \Delta R^2 < .070$), and Large ($\Delta R^2 \geq .070$).

Table S4

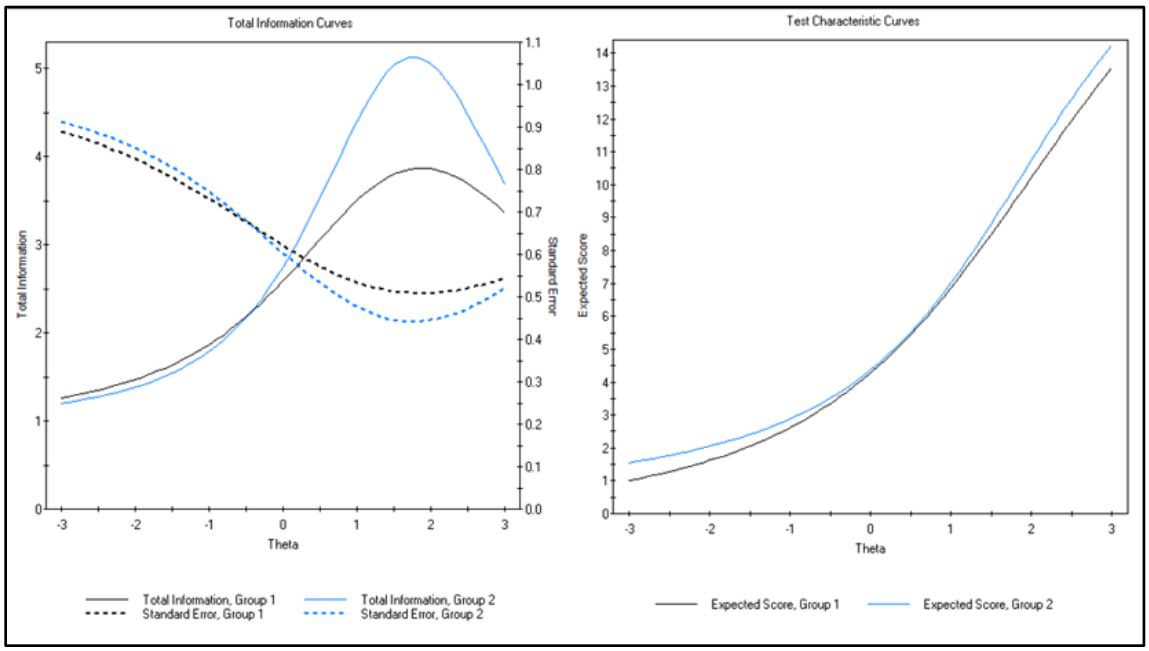
DIF statistics for English-French language version Comparisons

Item	*Total* X^2	*p*	X^2a	*P*	$X^2c\|a$	*p*
1. EXT2	0.4	.816	0.4	.527	0	.942
2. EXT9	9	.011	0	.905	9	.003*
3. EXT10	51.2	< .001*	5.8	.017	45.4	< .001*
4. CI1	1.7	.434	0.9	.339	0.8	.386
5. CI6	0.4	.823	0.1	.756	0.3	.588
6. CI11	112.2	< .001*	0	.835	112.2	< .001*
7. SS1	22.8	< .001*	7.1	.008*	15.7	< .001*
8. SS2	4.7	.095	3.3	.070	1.4	.234
9. SS3	53.1	< .001*	2.3	.130	50.8	< .001*
10. SS9	12.2	.002*	0.2	.620	12	< .001*
11. FLN1	2.3	.317	2.1	.150	0.2	.641
12. FLN2	0.3	.852	0	.989	0.3	.571
13. FLN10	19.5	< .001*	2.1	.145	17.4	< .001*
14. RSI1	6.9	.031	0.6	.427	6.3	.012
15. RSI6	23.8	< .001*	8	.005*	15.8	< .001*
16. RSI7	49.2	< .001*	15.5	< .001*	33.6	< .001*
17. RSI8	3	.227	0.2	.622	2.7	.099
18. RSI10	14.1	.001*	1.1	.294	13	< .001*
19. RSI12	11.7	.003*	1.7	.189	10	.001*
20. INS2	104.5	< .001*	15.5	< .001*	89.1	< .001*
21. INS3	526	< .001*	48.4	< .001*	477.6	< .001*
22. INS9	121.3	< .001*	12.8	< .001*	108.5	< .001*

Note. Total X^2 = Omnibus DIF, X^2a = DIF due to discrimination parameter, $X^2c|a$ = remaining DIF with discrimination parameter removed, indicating DIF due to difficulty parameter. * $p < .01$.

Figure S3

English-French DIF Total Information Curves and Test Characteristic Curves



Note. English = Group 1, black lines. French = Group 2, blue lines. As shown, CRT-WP information is higher for French from 0 to +3 levels of theta, and French are expected to score around 0.5 to 1 point higher at -3 and +3 theta.

Table S5

Descriptive Statistics, Reliability Coefficients, and Correlations for English (lower diagonal) and French (upper) datasets

	M	*SD*	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. CRT-WP	-12.12/-10.25	5.60/5.38	(.82/.76)	-.14**	-.21**	-.13**	-.20**	.03	-.13**	.16**	.01	.05	.02
2. Verbal	0.70/0.69	0.17/0.16	-.18**	(.65/.64)	.27**	.38**	.64**	.18**	.02	-.06**	-.01	.01	.20**
3. Visuospatial	0.73/0.73	0.23/0.23	-.16**	.39**	(.82/.82)	.49**	.77**	-.21**	.05*	-.35**	.02	.01	-.11**
4. Problem Solving	0.62/0.67	0.16/0.18	-.14**	.37**	.51**	(.78/.83)	.87**	-.01	-.06*	-.03	-.02	-.02	.15**
5. Cog Ability Total	0.67/0.69	0.15/0.15	-.20**	.69**	.81**	.86**	(.87/.89)	-.03	-.01	-.18**	-.01	-.01	.10**
6. Age	28.06/27.72	9.67/9.87	-.03	.06**	-.12**	.06**	.01	-	.01	.26**	-.07**	.09**	.47**
7. Gender (M/F)	0.31/0.28	0.46/0.45	-.08**	-.03*	-.06**	-.10**	-.09**	.03*	-	-.05	.01	.01	.03
8. Minority (N/Y)	0.33/0.26	0.47/0.44	.13**	-.11*	-.10**	.10**	-.03*	.04**	-.03*	-	-.06**	.02	.34**
9. Indigenous (N/Y)	0.06/0.02	0.24/0.13	.04**	-.05**	-.02	-.10**	-.08**	-.06**	.02	-.14**	-	-.01	-.02
10. Disability (N/Y)	0.03/0.01	0.16/0.05	-.02	.01	.01	-.02	-.01	-.02	.04**	-.01	.03*	-	.01
11. Education	0.31/0.25	0.46/0.44	-.03*	.06**	.01	.20**	.12**	.31**	.04**	.22**	-.06**	-.01	-

Note. English data presented on bottom diagonal and first value when means, *SD*, and reliability coefficients presented side-by-side. French data presented on upper diagonal and second values when means, *SD*s, and reliability coefficients presented side-by-side. Pairwise *N*s for English data = 5196 to 5849. Pairwise Ns for French data = 1795 to 1939. N/Y = No/Yes. Education scored as 0 for non-university, 1 for any university. As noted in participants section, many people either did not complete demographics section or were coded missing, and these participants would be excluded from these analyses. * $p < .05$, ** $p < .01$.

Supplementary Material A – Example CRT-WP Item in English and French

(**Bold** = psychopathic, *Italics* = non-psychopathic option)

Some people in leadership positions consider their subordinates as pawns that are used to get things done for more important people, similar to the pawns in a game of chess. This means that these leaders think that it is best to use, control, and manipulate all of their subordinates to achieve the goals of the organization in any way that they see fit. This leadership style can be a very effective one.

However, what is the biggest issue with comparing subordinate employees to pawns?

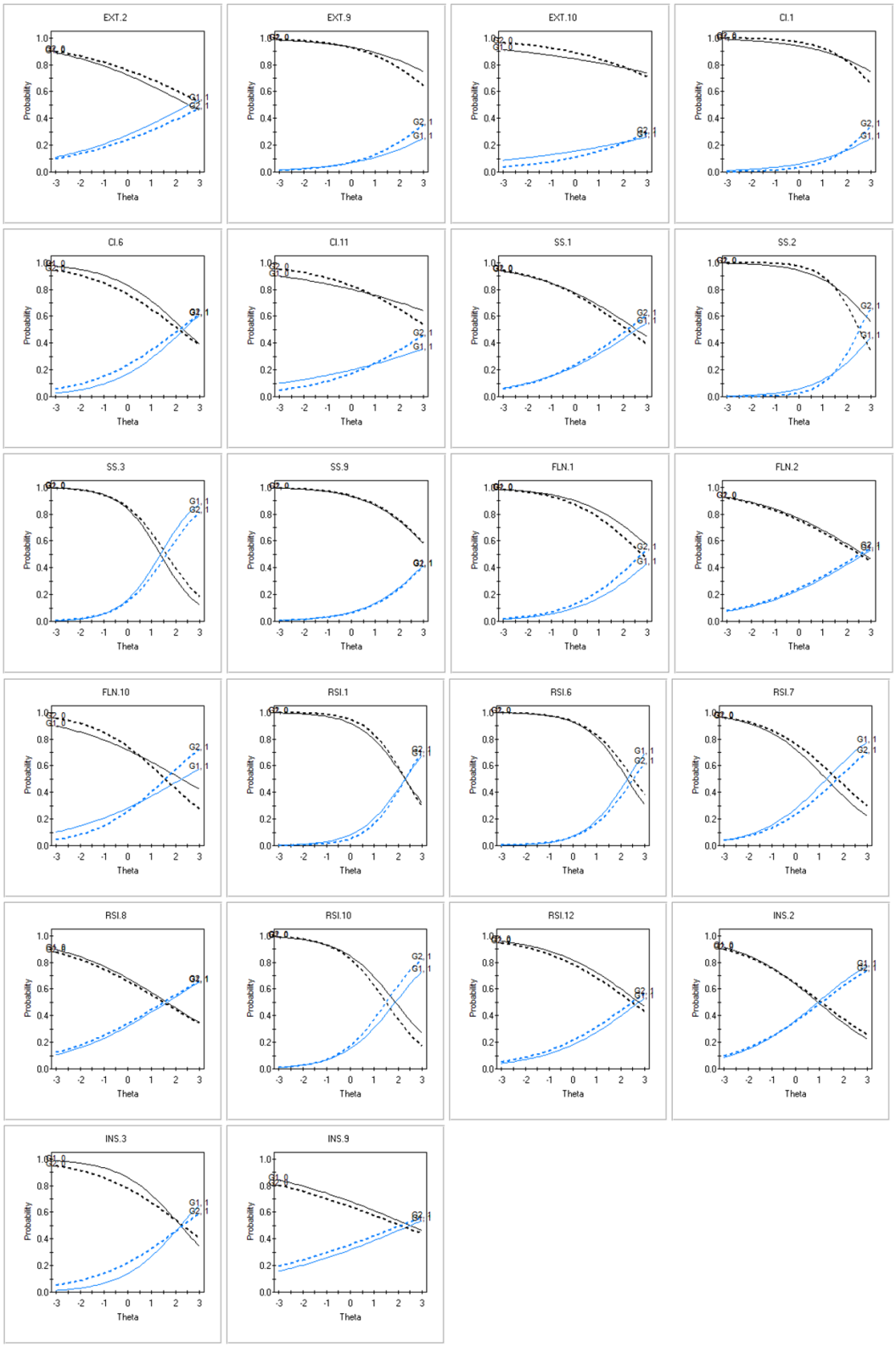
A) Unlike chess pieces, subordinate employees do not always do what you tell them to do
B) It bridges the gap between fellow organizations
C) It is not a viable strategy in workplaces with no internet connection
D) All employees should be treated with respect and consideration

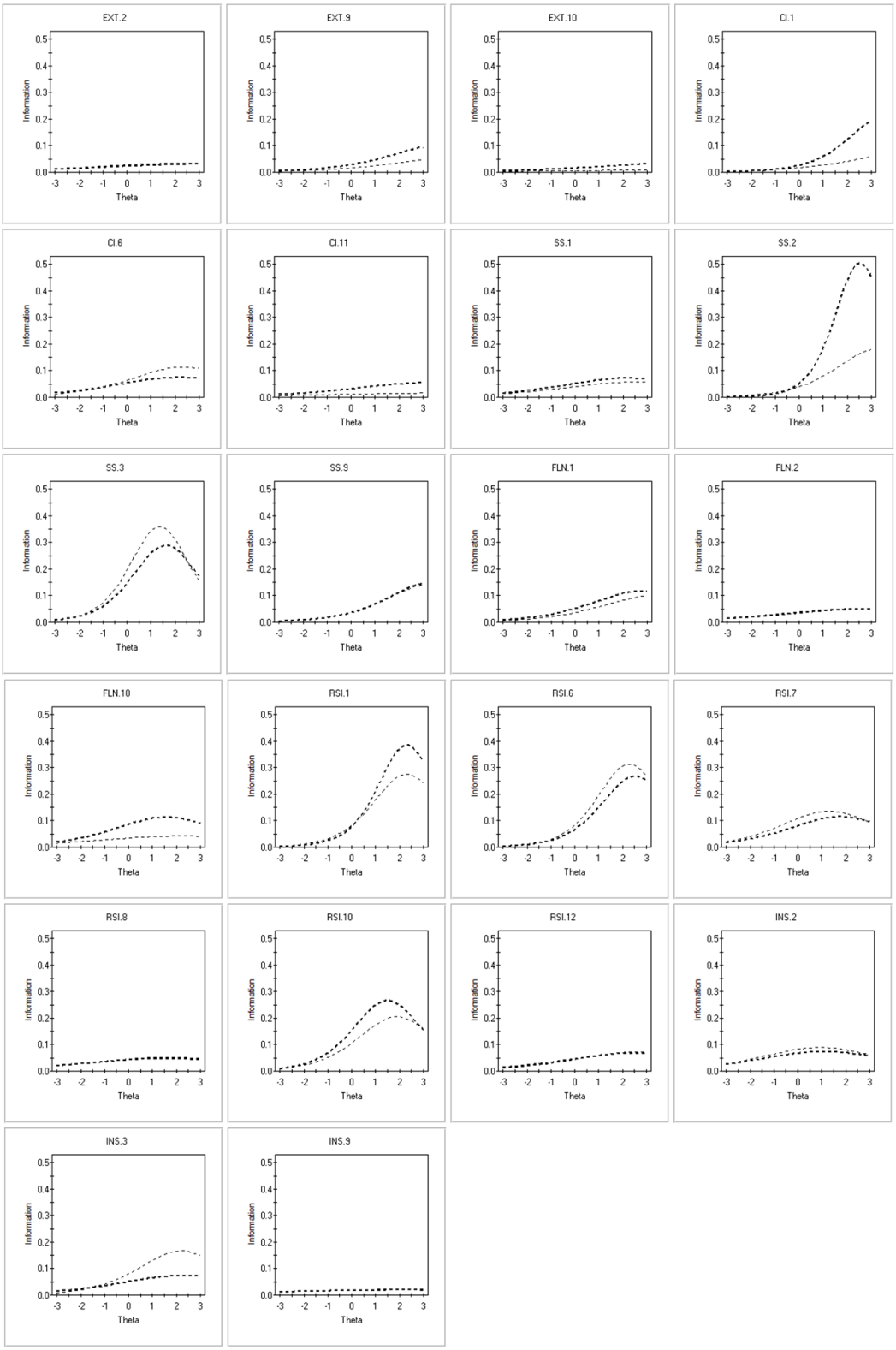
Certaines personnes occupant des postes de direction considèrent leurs subordonnés comme des pions utilisés pour faire avancer les choses pour des personnes plus importantes, comme les pions dans une partie d'échecs. Cela signifie que ces dirigeants pensent qu'il est préférable d'utiliser, de contrôler et de manipuler tous leurs subordonnés pour atteindre les objectifs de l'organisation de la manière qu'ils jugent appropriée. Ce style de leadership peut être très efficace.

Cependant, quel est le plus gros problème avec la comparaison des employés subordonnés à des pions?

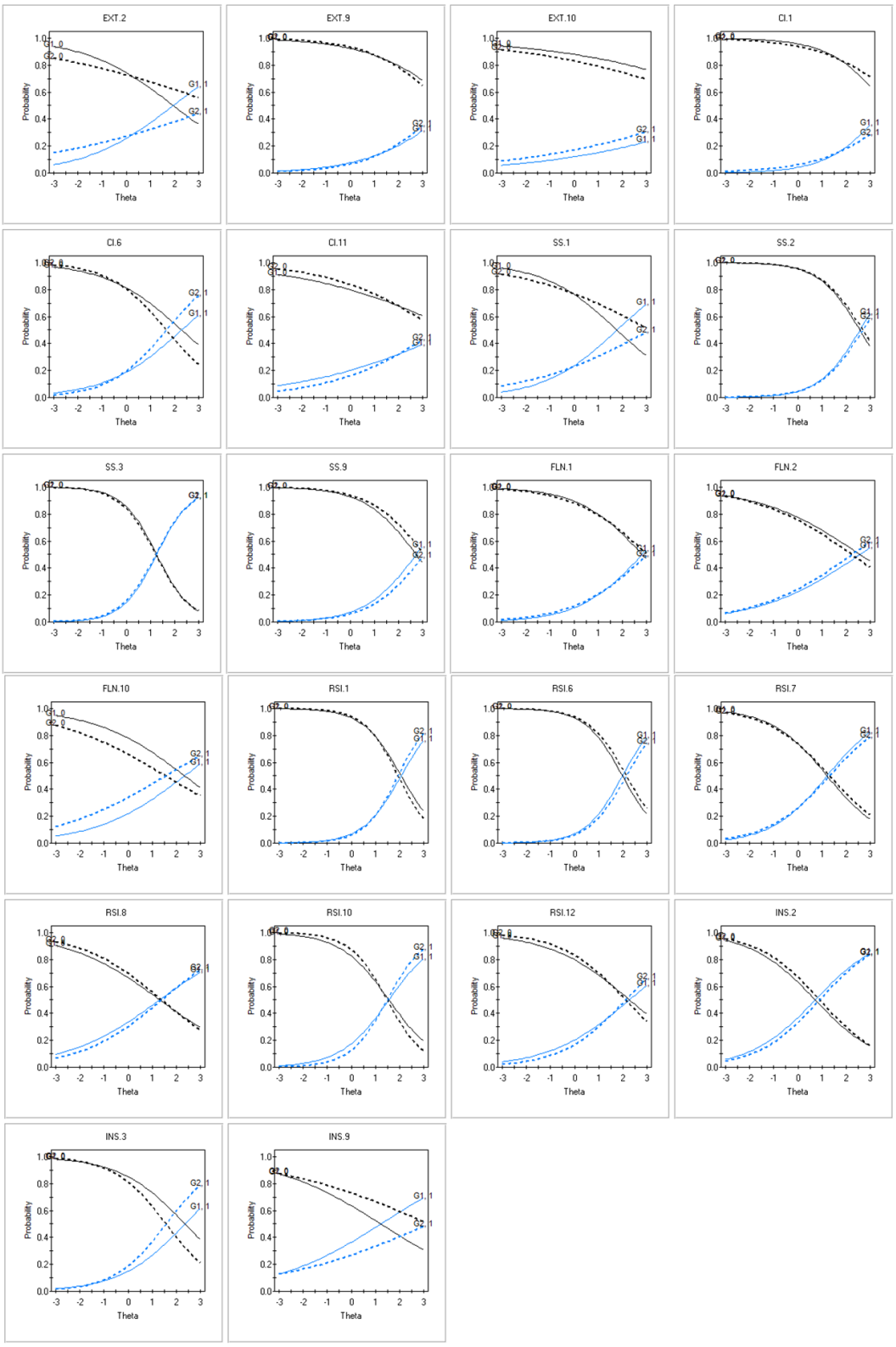
A) Contrairement aux pièces d'échecs, les subordonnés ne font pas toujours ce que vous leur dites de faire
B) Cela comble le fossé entre les organisations aux profils similaires
C) Ce n'est pas une stratégie viable dans les lieux de travail sans connexion Internet
D) Tous les employés doivent être traités avec respect

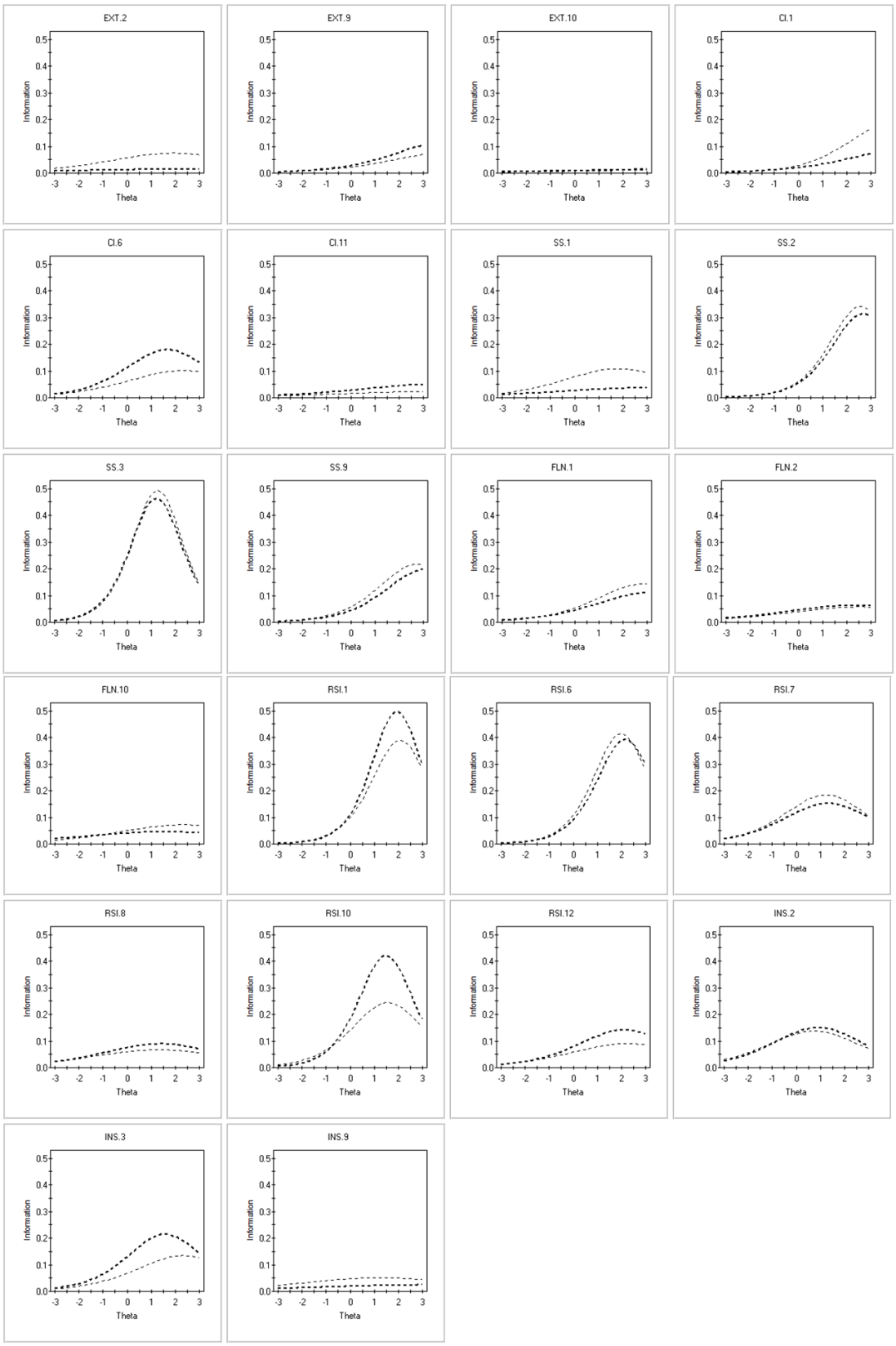
Supplementary Material B – Item Characteristic Curves, Item Information Curves, and Overall Test Curves for Gender DIF Comparisons (Man = G1, Woman = G2)



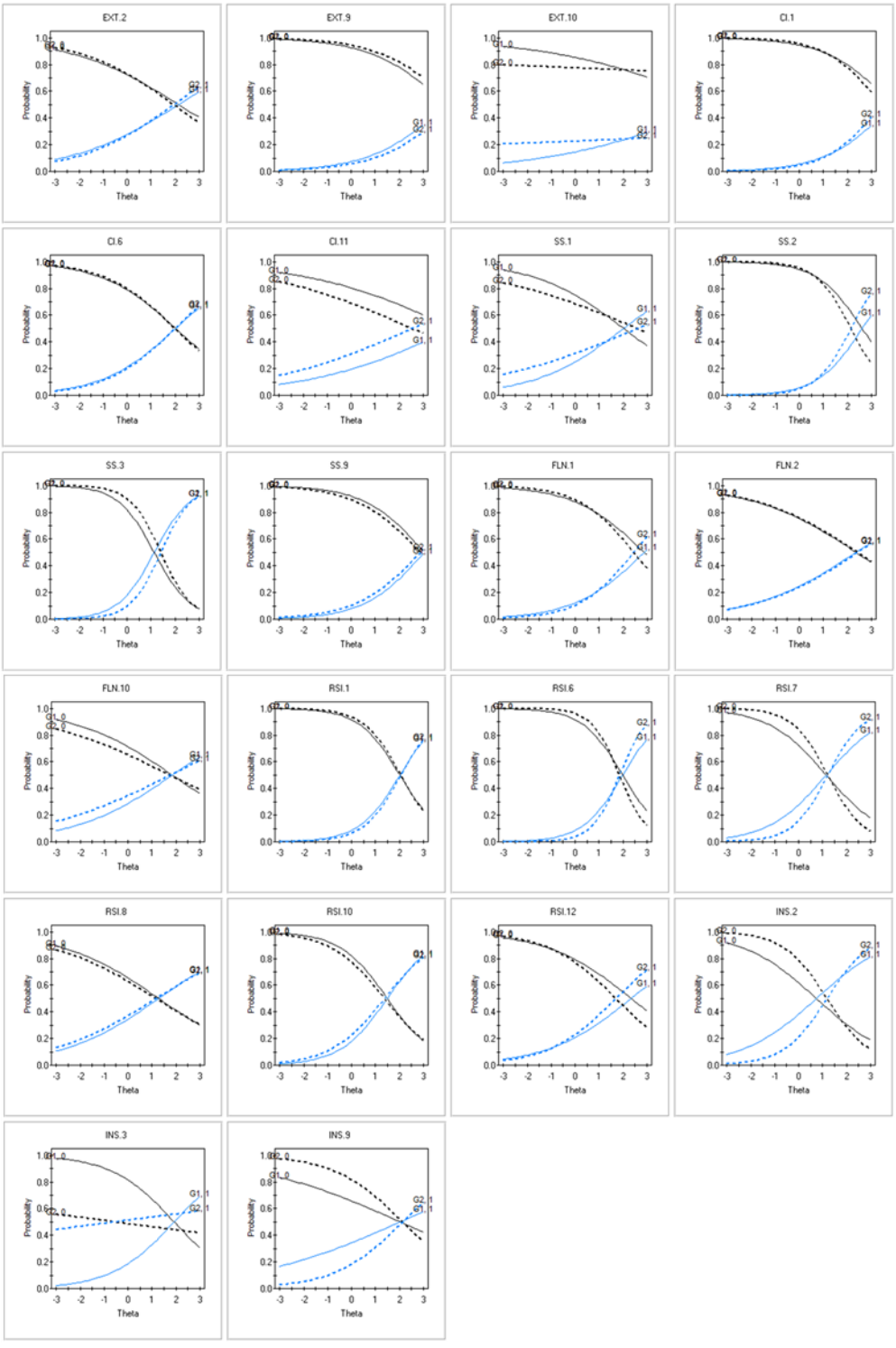

EXT.2
EXT.9
EXT.10
CI.1
CI.6
CI.11
SS.1
SS.2
SS.3
SS.9
FLN.1
FLN.2
FLN.10
RSI.1
RSI.6
RSI.7
RSI.8
RSI.10
RSI.12
INS.2
INS.3
INS.9
Information
Theta

Supplementary Material C – Item Characteristic Curves, Item Information Curves, and Overall Test Curves for Ethnicity DIF Comparisons (White = G1, Visible Minority = G2)




EXT.2
EXT.9
EXT.10
CI.1
CI.6
CI.11
SS.1
SS.2
SS.3
SS.9
FLN.1
FLN.2
FLN.10
RSI.1
RSI.6
RSI.7
RSI.8
RSI.10
RSI.12
INS.2
INS.3
INS.9
Information
Theta

Supplementary Material D – Item Characteristic Curves, Item Information Curves, and



Overall Test Curves for Language/Version DIF Comparisons (English = G1, French = G2)